

Topologische Datenbanken für Architektonische Räume

Genehmigte Dissertation
an der
Fakultät für Architektur
der
Universität Karlsruhe (TH)

von

Dipl. Ing. Norbert Paul
aus
München

Tag der Prüfung:
10. März 2008

Referent:
Korreferent:

Prof. Dr. ès. sc. techn. Niklaus Kohler
Prof. Dr. Frank Herrlich

Vorwort

„Junger Mann, in der Mathematik brauchen Sie die Dinge nicht zu verstehen, Sie gewöhnen sich einfach an sie.“

JOHN VON NEUMANN – [Zuk81, S. 240, Fussnote]

Die Anregung zu dieser Arbeit stammt von einer Idee, die im Institut für Industrielle Bauproduktion (ifib) herumgeisterte: Gebäudegrundrisse mit einem Geoinformationssystem (GIS) zu zeichnen und zu untersuchen, welche Möglichkeiten dies für die Planung bietet. Vermutlich war FILIZ OZEL, ehemalige Gastprofessorin des ifib, Urheberin dieser Idee.

Es ist ein interessantes Erlebnis, mit GIS-Erfahrungen im Hintergrund Topologie zu lernen, denn dabei entsteht zunächst der Eindruck, dass der Topologiebegriff der Mathematik nicht viel mit einer „Topologie“ im GIS zu tun hat. Erst nach einiger Gewöhnung sieht man ein, dass es sich dabei in der Tat um die selbe Sache handelt.

Die Ideen und Konzepte im Hauptteil dieser Arbeit werden in etwa in der Reihenfolge behandelt, wie sie entstanden sind: Zunächst wird Topologie in der Architektur bestimmt, dann werden die relationalen Kettenkomplexe behandelt und erst danach die topologischen Datenbanken. Diese enthalten die Topologie der relationalen Kettenkomplexe. Die Erkenntnis, dass mit den topologischen Datenbanken im Wesentlichen die topologische Datenmodellierung abgedeckt ist, war sehr aufregend. Weitere Recherchen ergaben dann, dass die entsprechenden Topologien bereits 1937 durch ALEXANDROFF bekannt waren, deren Übertragung auf das relationale Datenmodell scheint jedoch neuartig zu sein.

Die Arbeit richtet sich an alle, die mit der Modellierung von Gebäuden am Rechner befasst sind — insbesondere an Architekten, die für Tätigkeiten außerhalb des traditionellen Berufsbildes aufgeschlossen sind. Daher wird hier elementare Mathematik besonders ausführlich behandelt — die Arbeit umfasst also auch einen Einführungskurs in die hier entwickelte Theorie.

Ich danke allen Personen, die diese Arbeit unterstützt haben, insbesondere die Prüfer Prof. Dr. Niklaus Kohler und Prof. Dr. Frank Herrlich. Herrn Prof. Dr. Kohler danke ich dafür, dass er meine Arbeit über mehrere Jahre wohlwollend unterstützt hat. Meinem Kollegen Dr. Patrick Erik Bradley danke ich für geduldigen Mathematikunterricht und für die stete Ermahnung, diese Arbeit zügig zu Ende zu bringen. Viele Ideen darin stammen aus Diskussionen mit ihm, insbesondere die zur kategoriellen Beschreibung des relationalen Datenmodells. Als erster Korrekturleser sorgte er auch dafür, dass Herr Prof. Dr. Herrlich fast nur noch Kommafehler vorfand — davon aber um so mehr. Frau cand. math. Anita Höschele danke ich für den Beweis von Satz 4.12 und für das Beispiel in Abschnitt 4.4.3. Jutta Mülle vom Institut für Programmstrukturen und Datenorganisation danke ich für den Zugang zu einem geeigneten Datenbankserver, um die hier formulierten Abfragen testen zu können, genauso wie den Mitarbeitern im Rechenzentrum, die ebenfalls stets bemüht waren, mir einen solchen Zugang zu verschaffen.

Die Arbeit wurde mit dem XEmacs und mit L^AT_EX unter Debian-Linux angefertigt und ihre Versionsgeschichte mit CVS verwaltet. Die Abbildungen sind, bis auf die drei entsprechend gekennzeichneten Ausnahmen, vom Autor selbst überwiegend mit dem Zeichenprogramm Draw von OpenOffice erstellt.

Zusammenfassung

Alle Systeme zur rechnergestützten Gebäudeplanung müssen in irgendeiner Weise berücksichtigen, wie die Einzelelemente eines Gebäudes — z.B. dessen Räume und die sie begrenzenden Bauelemente — miteinander in Beziehung stehen.

Derartige Eigenschaften bezeichnet man als topologische Eigenschaften, und jeder architektonische Raum ist damit auch ein topologischer Raum. Da ein Gebäude nur aus endlich vielen Elementen besteht, handelt es sich hierbei um einen endlichen topologischen Raum. Mit einem solchen Raum hat es nun die Bewandnis, dass sich alle seine topologischen Eigenschaften verlustfrei in einer relationalen Datenbank nach einem sehr einfachen Schema speichern lassen. Somit werden diese Datenbanken ebenfalls zu topologischen Räumen, und deren Topologie unterscheidet sich nicht von der Topologie des zugrunde liegenden Gebäudes.

Diese Beobachtung wird zu einer Theorie der Datenhaltung von topologischen Informationen — eben den topologischen Datenbanken — ausgebaut. Dies geschieht durch eine Erweiterung des relationalen Datenmodells, die sich wiederum mit den Mitteln herkömmlicher relationaler Datenbankmanagementsysteme¹ realisieren lässt. Die topologischen Datenbanken sind auch eine gemeinsame Verallgemeinerung der Ansätze zur räumlichen Modellierung von Gebäuden, zur Modellierung von geographischen Räumen und von räumlichen (insbesondere raumzeitlichen) Modellen ganz allgemein, beschränkt auf topologische Informationen. Mit Hinweisen zum Entwurf eines Planungsinformationssystems wird schließlich eine Einsatzmöglichkeit für diese Datenbanken zur Gebäudeproduktmodellierung gezeigt.

¹plus einem Körnchen Salz, genannt „transitive Hülle“.

Inhaltsverzeichnis

1 Fragestellung und Grundlagen	1
1.1 Einleitung	1
1.2 Fragestellung	2
1.3 Überblick über die Arbeit	3
1.4 Stand der Technik	4
1.4.1 STEP	4
1.4.2 COMBINE	6
1.4.3 RATAS	7
1.4.4 Industry Foundation Classes	8
1.4.5 Das DFG Schwerpunktprogramm 1103	9
1.4.6 Abstrakte Modelle	10
1.4.7 Geoinformationssysteme	10
1.4.8 Volumenmodellierer	13
1.4.9 G-Maps	19
2 Mathematische Grundlagen	22
2.1 Mengen, Relationen und Abbildungen	22
2.1.1 Mengen	22
2.1.2 Teilmenge, Gleichheit von Mengen	27
2.1.3 Mengenoperatoren	28
2.1.4 Relationen	29
2.1.5 Das Relationenprodukt	35
2.1.6 Abbildungen	40
2.2 Metrische Räume	44
2.3 Topologische Räume	48
2.4 Stetige Abbildungen	52
2.5 Quotientenbildung	57
2.6 Algebraische Topologie	60
2.6.1 Moduln und Lineare Abbildungen	62
2.6.2 Lineare Algebra an einem Graphen	66
2.6.3 Komplexe	71
2.6.4 CW-Komplexe	74
2.7 Kategorien und Funktoren	77
2.8 Das relationale Datenmodell	80
2.8.1 Relationen	80
2.8.2 Operatoren für Relationen	82
2.8.3 Relationaler Entwurf	93

3	Die Topologischen Postulate	98
3.1	Die Architektonische Zerlegung	98
3.2	Die Zerlegung zu einem Komplex	99
4	Topologische Datenbanken	101
4.1	Das provisorische Modell $\mathcal{DKetKomp}^{\text{tot}}$	101
4.1.1	Dynamische Dimension	107
4.1.2	Informationsverluste	107
4.2	Das Datenmodell für Komplexe $\mathcal{DKetKomp}$	108
4.3	Das Datenmodell für topologische Räume \mathcal{DTop}	114
4.3.1	Topologische Datenbanken	114
4.3.2	Stetige Datenbankabbildungen	118
4.3.3	Die Kategorie \mathcal{DTop}	121
4.4	Topologische Konstruktionen	127
4.4.1	Vergleichen von Topologien	128
4.4.2	Wahl der besten Topologie	129
4.4.3	Verfeinerung und transitive Hülle	135
4.4.4	Verallgemeinerte topologische Konstruktion	140
4.4.5	Topologische Konstruktionen in \mathcal{DTop}	144
4.4.6	Einstellige Abfragen	144
4.4.7	Verbinden von Räumen	148
4.4.8	Versionen von Räumen	159
4.4.9	Diskussion der Ergebnisse	160
4.4.10	Ausblick	161
4.5	Der Entwurf topologischer Datenbanken	161
4.5.1	Normalformen	162
4.5.2	Trivialität und Monotonie	163
4.5.3	Wege in topologischen Datenbanken	170
4.5.4	Entscheiden von Zusammenhang und Monotonie	173
4.5.5	Monotonie in Bauteildatenbanken	178
5	Geometrie	193
5.1	Vektorräume und affine Räume	193
5.2	Darstellung geometrischer Körper	194
5.2.1	Punkte	194
5.2.2	Linien	194
5.2.3	Flächen	195
5.2.4	Volumen	195
5.3	Zusammenfassung	195
6	Anwendung	196
6.1	Planungsbegriff	196
6.2	Planungsaufgabe	198
6.2.1	Das Raum- und Funktionsprogramm	199
6.2.2	Planungskontext	202
6.2.3	Planungsobjekt	202
6.3	Beteiligte	204
6.4	Dokumente	204
6.5	Versionen	205
6.6	Semantik	207

6.6.1	Detailbezeichner	207
6.6.2	Detailklassen	208
6.6.3	Dimensionsklassen	209
6.7	Zusammenstellung	211
7	Zusammenfassung und Ausblick	213
A	Endliche Wege	215
A.1	Der Begriff des endlichen Wegs	215
A.2	Grundlagen	218
A.2.1	Reelle Zahlen	218
A.2.2	Folgen	218
A.2.3	Kompaktheit	221
A.3	Der Beweis	223
B	Monotone Mengen	233
B.1	Algorithmus zur Entscheidung von Monotonie.	233

Kapitel 1

Fragestellung und Grundlagen

„The advent of computer methods has meant that architectural elements and relationships need to be given mathematical representation of some kind.“

MARCH und STEADMAN [MS74, Vorwort]

1.1 Einleitung

Bereits zur Anfangsphase der computergestützten Planung in der Architektur wurden die Systeme zum Zeichnen von Plänen (CAD — Computer Aided Design) zu besonders wichtigen Planungswerkzeugen. Es trat dabei bald der Wunsch auf, nicht nur graphische Objekte, wie etwa Linien oder Schraffuren, anzufertigen, sondern eine angemessene Repräsentation des zu planenden Objekts selbst im Rechner zu erstellen und zu bearbeiten. Es war also ein allgemeines Modell von Gebäuden zu entwickeln, so dass ein Einzelgebäude im Rechner als Ausprägung eben dieses allgemeinen Modells dargestellt werden konnte.

Es erscheint jedoch relativ schwierig, ein Gebäudemodell zu entwickeln, das die gegenseitigen Abhängigkeiten von Gebäudeelementen angemessen berücksichtigt. So werden Räume von Wänden begrenzt und durch Türen verbunden. Die Türen wiederum sitzen in diesen Wänden, die zudem Verbindungen zu Böden und Decken haben und in denen oft Medien wie Wasserrohre oder Stromleitungen verlaufen. Diesem relativ komplizierten Gefüge von Bauelementen scheint die enorme Komplexität mancher der entwickelten Modelle, wie etwa die Industrie Foundation Classes (IFC), angemessen zu sein.

Auf der anderen Seite gibt es abstraktere räumliche Datenmodelle, die für einfachere Anwendungsfälle entwickelt wurden. Diese haben aber mit Eigenschaften zu tun, die dem Gefüge von Bauelementen sehr ähneln, sind dabei aber sehr einheitlich und vor allem einfach strukturiert. Dies sind z.B. die Datenstrukturen der Volumenmodellierer und die Modelle der Geoinformationssysteme (GIS). Diese Modelle behandeln ebenfalls räumliche Objekte, wie etwa Parzellen, und deren Verbindungen, wie z.B. gemeinsame Grundstückskanten und Grenzpunkte. Dabei wird dort stets unterschieden zwischen der abstrakten räumlichen Struktur auf der einen Seite und der spezifischen Interpretation

dieser Räume durch die Anwender auf der anderen Seite — eine Abstraktion, die bei den meisten Gebäudemodellen fehlt. Zwar sind GIS traditionellerweise auf zweidimensionale Räume beschränkt, aber sie eignen sich dennoch gut als Vorbilder für eine vereinheitlichte räumliche Modellierung von Gebäuden, zumal höherdimensionale GIS bereits in Entwicklung sind [Bre01].

All diesen Modellen ist gemeinsam, dass sie ganz explizit topologische Eigenschaften modellieren und dass sich dabei deren Autoren auf topologische Konzepte berufen. Die Verbindungen der Teile eines Gebäudes untereinander erzeugen nun ebenfalls eine topologische Struktur dieser Teile. Ganz in Analogie zu den GIS kann man das Gefüge der Räume in einem Gebäude als eine Art von „Landkarte“ mit dreidimensionalen „Regionen“ betrachten. Deren Grenzen sind die zweidimensionalen Oberflächen der Wände, Böden und Decken, und anstelle der Grenzpunkte auf der Landkarte haben wir die Verbindungskanten dieser Boden- und Deckenelemente. Diese Kanten wiederum münden — ebenfalls ähnlich wie bei Landkarten — in Eckpunkte. Wir haben also eine sehr starke Analogie zu Landkarten, und tatsächlich werden wir sehen, dass es Ansätze gibt, Gebäude unter Verwendung einer Technik, die als „höherdimensional verallgemeinerte Karten“ (G-Maps) bezeichnet wird, zu modellieren.

1.2 Fragestellung

Der Rechneinsatz in der Architektur erfordert also eine Darstellung von Gebäuden¹ im Computer. In jeder derartigen Darstellung werden gewisse Eigenschaften des dargestellten realen Gebäudes imitiert, sie ist also ein Modell des realen Gebäudes. Wegen der Notwendigkeit zur Zusammenarbeit bei der Planung müssen derartige Modelle zwischen den Planungsbeteiligten ausgetauscht werden können. Dies hat zu Bestrebungen geführt, die Modellierung von industriellen Produkten ganz allgemein zu standardisieren, insbesondere sind dabei auch viele Vorschläge zur Standardisierung der Gebäudeproduktmodelle entwickelt worden[Eas99]. Jedes dieser Modelle ist nun gezwungen, in irgendeiner Form topologische Eigenschaften von Gebäuden zu berücksichtigen. Dies sind Eigenschaften wie etwa

- Welche Bauteile sind miteinander verbunden?
- Welche Räume grenzen an Bauteile oder aneinander?
- Wie werden einzelne Räume zu Zonen oder Geschossen zusammengefasst?

In Kapitel 1.4 wird dies anhand einiger Beispiele weiter erläutert.

Die Frage ist, ob es möglich ist — ganz in Analogie zu den GIS-Systemen — ein abstraktes, einheitliches Datenmodell für topologisch strukturierte Gebilde zu entwickeln, und erst dann, auf der Basis dieses Modells, ein Gebäude- bzw. ein Planungsinformationssystem aufzubauen. Damit wäre dann auch eine gemeinsame Vereinheitlichung der Geoinformationssysteme mit anderen räumlichen Planungsinformationssystemen geschaffen. In dieser Arbeit wird also folgende Frage untersucht:

¹„Gebäude“ bezeichnet hier stets Bauwerke im umgangssprachlichen Wortsinn

Wie lässt sich die Modellierung der topologischen Eigenschaften von Gebäuden verallgemeinern, und kann diese Verallgemeinerung als Grundlage eines Modells für Gebäude, deren Elemente und Verbindungen dienen?

Zur Beantwortung dieser Frage wird zunächst ein abstraktes, topologisches Datenmodell entwickelt. Dieses wird dann verwendet, um ein Gebäudeproduktmodell als eine spezielle fachspezifische Erweiterung des abstrakten Modells darauf aufzubauen. Damit wäre dieses abstrakte Modell auch für andere Disziplinen außerhalb der Architektur anwendbar.

1.3 Überblick über die Arbeit

Zunächst wird auf den Stand der Technik bei der Modellierung von Gebäuden am Computer eingegangen, wobei naturgemäß besonderes Gewicht auf die Modellierung von topologischen Informationen gelegt wird. Danach wird auf abstraktere räumliche Modelle, wie die Volumenmodellierer, die Geoinformationssysteme und die G-Maps, eingegangen. Letztere versprechen, GIS-Modelle auf beliebige höhere Dimensionen zu verallgemeinern.

Danach folgt eine Einführung in die mathematischen Grundlagen, auf denen diese Arbeit aufbaut. Dieser Teil ist sehr ausführlich und richtet sich insbesondere an Leserinnen² mit weniger mathematischen Kenntnissen. Dies trägt erheblich zum Umfang dieser Arbeit bei, da hier sehr viele mathematische Grundkenntnisse erst eingeführt werden, die sich so auch in einschlägigen Lehrbüchern finden. Es wird dabei weitestgehend versucht, Begriffe möglichst anschaulich zu erklären und überwiegend an Beispielen aus dem Bauwesen zu illustrieren, um die Anschauung von Anfang an in die richtige Richtung zu lenken. Dennoch kann es zum Verständnis oft hilfreich (für Anfängerinnen sogar notwendig) werden, die Argumente selbst handschriftlich in der Form von eigenen Notizen nachzuvollziehen.

Am Ende des einführenden Abschnittes steht das, was in dieser Arbeit als die topologischen Postulate für die Architektur bezeichnet wird. Diese bieten die formale Rechtfertigung, Gebäude topologisch aufzufassen. Der Einführungsteil sollte die Anschauung bereits so weit vorbereitet haben, dass diese Postulate in der Regel sofort einsichtig sind. Sie definieren, wie die architektonische Gegenstände der realen Welt mit den hier vorgestellten, rein logischen Konstrukten der Mathematik in Zusammenhang gebracht werden.

Danach wird die eigentliche Theorie entwickelt. Dies ist zuerst die Theorie der topologischen Datenbanken, die wir hier als die Kategorien $\mathcal{DKetKomp}$ und \mathcal{DTop} bezeichnen. Diese bilden den zentralen Teil dieser Arbeit, wobei auch eine Anregung von RÖSCH aufgegriffen [Rös98, S.46ff] und weiterentwickelt wird. Dabei zeigt sich, dass topologische Informationen sehr einfach in einer relationalen Datenbank gespeichert und von dort auch wieder abgefragt werden können. Zu diesem Zweck werden die topologischen Äquivalente zu den aus der relationalen Algebra bekannten Abfrageoperatoren entwickelt und umgekehrt Datenbankabfragen für bekannte topologische Konstruktionen formuliert. Dabei wird sich zeigen, dass nach diesem Modell im allgemeinen Fall auf die Möglichkeit zur Berechnung der sogenannten transitiven Hülle einer Relation nicht verzichtet

²Zur besseren Lesbarkeit stets mit kleinem „i“ geschrieben.

werden kann. Im besonderen Fall der Modellierung von Bauwerken hingegen besteht diese Notwendigkeit jedoch meist nicht.

Auf dieser Theorie aufbauend, wird dann erläutert, wie ein Planungsinformationssystem für die Gebäudeplanung aussehen könnte. Zuvor wird kurz noch auf einige allgemein bekannte Grundlagen der Geometrie eingegangen. Topologie kann man sich ja vorstellen als das, was übrig bleibt, wenn die geometrische Form als solche verschwindet — wie ein Grinsen ohne Katze. Für eine räumliche Planung müssen wir diese Form aber wieder zurückgewinnen. Wie dies genau geschehen kann, wird in dieser Arbeit aber nicht weiter vertieft werden.

1.4 Stand der Technik

Diese Arbeit hat zwei Aspekte: Einerseits behandelt sie die Gebäudeproduktmodellierung, andererseits soll dazu zunächst ein abstraktes räumliches Modell entwickelt werden, auf dem das Gebäudeproduktmodell dann erst aufbaut. Entsprechend wird nun der Stand der Technik einerseits bei der Gebäudeproduktmodellierung und andererseits bei den allgemeineren räumlichen Modellen betrachtet.

Als allgemeines Produktmodell wird zunächst STEP vorgestellt. Die weiteren in dieser Arbeit vorgestellten Gebäudeproduktmodelle bauen nämlich alle auf STEP auf. Diese sind das Integrated Data Model (IDM) des COMBINE II Projekts [Aug95, Eas99] und die Industry Foundation Classes (IFC) [LAF⁺]. Die Beschreibung von STEP, COMBINE II und IFC folgt im Wesentlichen EASTMAN [Eas99]. Zudem wird das Modell aus dem SFB 1103 „Relationale Prozessmodellierung in kooperativer Gebäudeplanung“ [DK02a, Dam00] vorgestellt, da es topologische Eigenschaften ausdrücklich thematisiert.

Abstraktere räumliche Modelle sind die Geoinformationssysteme und die Volumenmodellierer. Verwandte Ansätze finden sich auch bei GIELSDORF [Gie05] und BREUNIG [Bre01]. STEP hat zudem einen Abschnitt `topology_resources`, für die Volumenmodellierung. Auf diese wird nicht gesondert eingegangen, da sie den anderen Volumenmodellierern sehr ähnlich sind.

Ein sehr interessantes Modell ist das G-Map-Konzept von Pascal LIENHARDT [Lie94] — ein Datenmodell für topologische Räumen beliebiger Dimension. Bemerkenswerterweise verwenden FRADIN und MENEVEAUX aus der Gruppe von LIENHARDT die Bezeichnung „Architektonischer Komplex“ [FML02]. Die Beziehung zwischen den elementaren Objekten einer G-Map zu den Punkten des zugehörigen topologischen Raums erscheint allerdings etwas umständlich, was die Handhabung des Modells erschwert.

1.4.1 STEP

STEP ist die Abkürzung von STandard for the Exchange of Product model data. Es ist eine Norm (DIN EN ISO 10303), die den Datenaustausch von Produktdaten regelt. Sie ist nach drei Aspekten der Modellierung gegliedert:

physikalisch Dieser Modellaspekt beschreibt den allgemeinen Aufbau der Daten. Dies ist insbesondere die Struktur der SPF (Step Physical Files). Diese Dateien sind besonders formatierte Textdateien.

logisch Die logische Beschreibung des Modells ist noch von den Anwendungsprogrammen, die dieses Modell bearbeiten, unabhängig. Das logische Modell legt auch fest, wie Modellinformationen physikalisch repräsentiert werden.

anwendungsbezogen Hier wird festgelegt, welche Informationen des logischen Modells für gewisse Anwendungen, wie etwa Architektur oder Maschinenbau, von Belang sind.

Diese Modellierungsaspekte entsprechen also den drei Schichten der bekannten ANSI/X3/SPARC Referenzarchitektur für Datenbanken [TK78]. Auch diese unterscheidet eine physikalische Schicht für die Datenhaltung, eine konzeptuelle („logische“) Schicht des Gesamtmodells und eine externe („anwendungsbezogene“) Schicht.

Die physikalische Struktur

STEP bietet mehrere Möglichkeiten, Daten abzulegen: Speicherung als Textdatei oder Speicherung in einer Datenbank. Zur Speicherung als Textdatei war zunächst das STEP Physical File Format [ISO03] vorgesehen³. Gegenwärtig wird dazu übergegangen, STEP-Modelle im XML-Format zu speichern. Zudem gibt es auch Programmierbibliotheken zur Speicherung von STEP-Daten in relationalen Datenbanken.

Die logische Struktur

STEP bietet eine Auswahl von Beschreibungssprachen an, um die logische Struktur von Produktdaten festzulegen. Die populärsten dieser Beschreibungssprachen sind EXPRESS und EXPRESS-G. Die Sprache EXPRESS ähnelt den Klassendeklarationen einer objektorientierten Programmiersprache, EXPRESS-G hingegen ist eine graphische Darstellung von EXPRESS-Modellen und etwa vergleichbar mit der Unified Modeling Language (UML), die wir weiter unten noch kennen lernen werden.

Die fachspezifischen Modellteile

Die selben Sprachen, mit denen die logische Struktur des Modells beschrieben wird, sind auch zu verwenden, um bestimmten Anwendungsbereiche zu spezifizieren. Als Beispiel soll die EXPRESS-Beschreibung einer Türe im IFC-Modell dienen. Die IFC sind zwar nicht Teil der STEP-Norm, lehnen sich jedoch eng daran an und eignen sich somit als Beispiel. Türen in den IFC sind Elemente vom Typ `IfcDoor`:

```
ENTITY IfcDoor
  SUBTYPE OF (IfcBuildingElement);
  OverallHeight : OPTIONAL IfcPositiveLengthMeasure;
  OverallWidth  : OPTIONAL IfcPositiveLengthMeasure;
END_ENTITY;
```

³Beispieldateien findet man auf <http://www.steptools.com/>

Jede `IfcDoor` ist gleichzeitig vom Typ `IfcBuildingElement` mit dessen Eigenschaften. Man sagt auch: „`IfcDoor` erbt von `IfcBuildingElement`.“ Eine `IfcDoor` hat die zusätzlichen Eigenschaften (Attribute) `OverallHeight` und `OverallWidth`, also Breite und Höhe, vom Typ `IfcPositiveLengthMeasure`. Breite und Höhe müssen aber nicht immer angegeben werden, wie das Schlüsselwort `OPTIONAL` anzeigt.

Abbildung 1.1 zeigt das selbe Element in EXPRESS-G-Notation (ohne Verweis auf `IfcBuildingElement`).



Abbildung 1.1: Die `IfcDoor` in EXPRESS-G-Notation. (Bildquelle: [LAF⁺])

Folgende Zeile aus einer SPF-Datei (bzw. IFC-Datei) stellt eine Tür dar:

```
#8015 = IFCDOOR('wrlmbrmft', #7, $,$,$, #2478, #8014, $,$,$);
```

Alle hier angegebenen Attributwerte, bis auf die letzten beiden, sind die Werte, die diese `IfcDoor` vermöge ihrer Zugehörigkeit zu `IfcBuildingElement` hat. So ist z.B. 'wrlmbrmft' der Wert des Attributs `IfcObject.GlobalId`, das in `IfcObject` definiert ist und von dem `IfcDoor` erbt. Der Wert #8014 ist die Referenz auf das Geometrie-Objekt, welches die Form dieser Türe definiert (vom Typ `IfcProductDefinitionShape`). Umgekehrt haben Referenzen auf diese Türe den Wert #8015. Die \$-Einträge sind Platzhalter für nicht definierte Attribute und nur bei `OPTIONAL` Attributen zugelassen. Die zwei letzten \$ stehen für Breite und Höhe der Türe. Diese müssen also in unserem Beispiel aus der geometrischen Form ermittelt werden.

Die STEP-Norm umfasst auch einen einen Teil (Part 42), der ein `topology_schema` definiert. Dieses beschreibt vor allem den topologischen Aspekt der Volumenmodellierung. In der Einführung zum `topology_schema` steht zwar ausdrücklich, dass dieses Modell auch ganz allgemein verwendet werden kann, um Informationen über Verbindungen darzustellen. Diese Möglichkeit wird jedoch in der Praxis kaum genutzt.

1.4.2 COMBINE

Das COMBINE-Projekt der EU lief von 1990 bis 1995 und hatte das erklärte Ziel, Werkzeuge zur energieeffizienten Planung zu verbinden („integrieren“). Dazu wurde ein Gebäudemodell, das sog. integrierte Datenmodell (IDM), entwickelt. EASTMAN erwähnt, dass dieses Model selbst so umfangreich war, dass es nur mit Werkzeugunterstützung entworfen werden konnte [Eas99, S.221].

Ein Blick auf das Kernmodell von COMBINE (Abb. 1.2) zeigt, dass dort topologische Informationen modelliert sind, wie z.B.

```
building_space.bounded_by_bldg_enclosing_element[1:?]
```

bzw. Spezialisierungen davon, wie

```
space_subsystem.bounded_by_subsystem_enclosing_element[1:?].
```

Eine andere topologische Information ist etwa

```
enclosing_element.adjacent_to_enclosing_elements[0:?].
```

COMBINE hat zudem das `topology_schema` von STEP überarbeitet, und jedes `physical_object` verweist auf ein `representation_item` dieses Schemas, was die Komplexität des Modells noch erhöht.

Der Grund für diese enorme Komplexität des topologischen Modells dürfte vor allem darin liegen, dass topologische Struktur und die architektonische Semantik nicht unabhängig modelliert wurden. Aus topologischer Sicht ist es nämlich unerheblich, ob eine Verbindung die Beziehung eines Verbindungsmittels (`joint`) zu Bauteilen (`building_element`) oder die einer Wand (`elementary_space_enclosing_element`) zu den durch sie verbundenen (bzw. getrennten) Räumen (`elementary_space`) ist.

1.4.3 RATAS

RATAS ist die finnische Abkürzung für „Computergestützter Gebäudeentwurf“. Das finnische RATAS-Modell des Technischen Forschungszentrums von Finnland (VTT) hatte eine umfassendere Zielsetzung als COMBINE: Unterstützung des Austauschs von Gebäudedaten zwischen Computerprogrammen allgemein. Wir erinnern uns, dass sich COMBINE auf die energetische Planung konzentrierte.

Im Verlauf der Arbeiten verwendete auch das VTT die Sprache EXPRESS. In RATAS sind ebenfalls topologische Informationen von Gebäuden modelliert. Im Vergleich zu COMBINE fällt aber auf, dass die topologischen Beziehungen auf einer höheren Abstraktionsebene definiert werden. Es gibt im wesentlichen Gebäudeelemente, Verbindungen und Räume und es gibt Relationen vom Typ `part_of` und `connected_to`. Wir werden später sehen, dass man zum topologischen Modellieren mit diesen zwei Relationen im Wesentlichen auskommt, wobei wir `part_of` später als *stetige Abbildung* und `connected_to` als *Rand* bezeichnen werden.

Eigenständiges räumliches Modell

Beim RATAS-Modell findet sich bereits eine weitgehende Trennung der topologischen Informationen von den sonstigen planungsrelevanten Sachverhalten. Das räumliche Modell enthält dennoch Bezüge zu der fachspezifischen Interpretation dieser räumlichen Struktur. So gibt es z.B. den Typ `hole` und die Relation `contains_holes` zwischen diesem und dem Typ `physical_space_boundary`. Ein `hole` steht in Beziehung `filled_by` mit Objekten des Typs `opening_component` wie z.B. `window` oder `door` [Eas99, S.259].

Genau betrachtet ist jedoch jedes Wandelement oder jedes Deckenelement selbst ebenfalls eine Füllung für ein Loch — für das Loch das entsteht, wenn wir dieses Element entfernen. Wir werden später einen topologischen Raum behandeln, der als *Komplex* bezeichnet wird. Dies ist ein Raum, der ausschließlich aus Löchern (Zykeln) und deren Füllungen (Zellen) aufgebaut ist. Der Beziehung „gefüllt mit“ zwischen Loch und Zelle entspricht dort die Beziehung „ist

ebenfalls nach einem Schichtenmodell aufgebaut:

Resource Layer. Dies ist die unterste Modellschicht, in der die allgemeinen Grundtypen definiert sind. Diese werden dann von den anderen Definitionen verwendet. Beispiele sind etwa Zeit, Datum oder Materialien. In dieser Schicht findet sich auch eine Kopie des `topology_schema` von STEP, angepasst an die Namenskonventionen der IFC als `IfcTopologyResource`.

Core Layer. Diese Modellschicht spezifiziert abstrakte Typen, wie Objekt (ein physikalisch existierendes Ding) oder Root (ein allgemeines Ding wie z.B. Objekt aber auch eine Beziehung), in einem Paket namens „Kernel“. Auf diesen Definitionen baut dann das weitere Modell auf. Das Gebäudemodell ist in der `ProductExtension` dieser Modellschicht definiert. Hier setzt es sich zunächst aus nicht weiter spezifizierten Gebäudeelementen vom Typ `IfcBuildingElement` zusammen.

Interoperability Layer. Im Interoperability Layer werden die abstrakten Typen des Core Layer so weit spezialisiert, wie es alle beteiligten Fachplanungen benötigen. So spezialisiert dieses Layer die bisher noch allgemein gehaltenen Gebäudeelemente zu Träger (`IfcBeam`), Stütze (`IfcColumn`), und die uns bereits bekannte Türe (`IfcDoor`) u.s.w. Typen dieser Schicht sind nur so weit spezialisiert, dass sie zwischen unterschiedlichen Anwendungsprogrammen verschiedener Fachplaner noch ausgetauscht werden können.

Domain Layer. Diese Modellschicht definiert Typen, die für spezielle Anwendungen bzw. Fachplanungen zugeschnitten sind. So gibt es etwa ein Paket für die Haustechnik (`IfcHVACDomain`) mit Typen, wie z.B. Ventilator (Typ `IfcFanType`), oder ein Paket für die architektonische Planung mit dem Typ Raumprogramm (`IfcSpaceProgram`). Für die Gebäudeverwaltung existiert das Paket `IfcFacilitiesMgmtDomain` mit Typen, wie etwa der Umzug von einem Raum in einen anderen (`IfcMove`).

Topologische Informationen werden in den IFC z.B. mit den Typen `IfcSpace` und `IfcSpaceBoundary` und der Beziehung `BoundedBy` definiert. Für die topologischen Beziehungen der `IfcSpaceBoundary`-Objekte untereinander steht ein Beziehungstyp namens `IfcRelConnectsElements` zur Verfügung. Die `IfcTopologyResources` werden hingegen nur für die Volumenmodellierung der einzelnen Gebäudeelemente verwendet. Wir sehen also ein Beispiel, wie das weitergehende Angebot des `topology_schema` eben nicht genutzt wird. Stattdessen werden topologische Beziehungen, die über die reine Volumenmodellierung hinaus gehen, jeweils neu modelliert.

1.4.5 Das DFG Schwerpunktprogramm 1103

Das Schwerpunktprogramm 1103 der Deutschen Forschungsgemeinschaft „Vernetzt - kooperative Planungsprozesse im Konstruktiven Ingenieurbau“ hat die Zielsetzung, die verteilte Planung mit Datenaustausch über Netzwerke zu fördern.

Das Projekt baut auf den IFC auf. Diese werden um einen eigenen Ansatz zur Beschreibung der Topologie erweitert: „Eine relationale Gebäudestruktur

beschreibt die Topologie eines Gebäudes. Sie besteht aus Komponenten und Verbindungen sowie deren Beziehungen.“[DK02a, 3.2.2] Danach wird festgestellt, dass sich dieses Modell mit den IFC Version 2x umsetzbar ist[DK02a, 3.2.4]. Auf Abbildung 1.3 ist eine Ausprägung eines solchen Modells zu sehen.

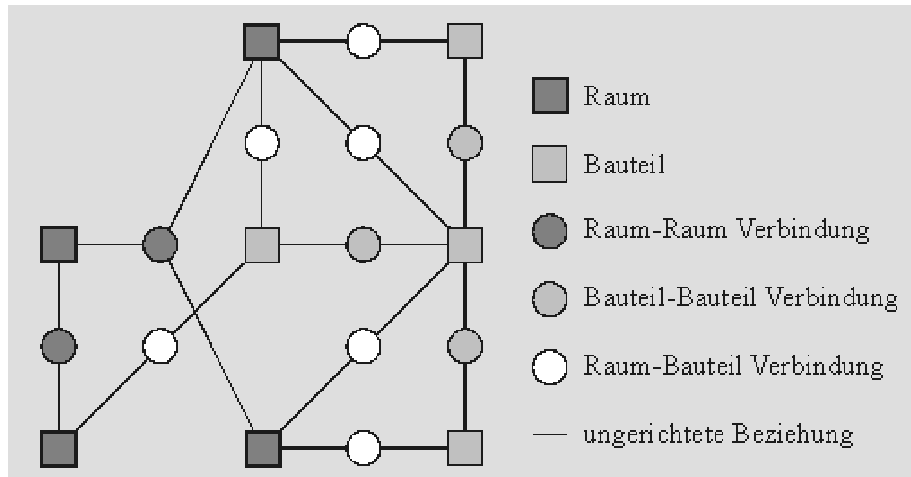


Abbildung 1.3: *Ein Gebäudegrundriss im relationalen Modell der Gebäudetopologie des SPP 1103 nach DAMRATH. (Bildquelle: [DK02b])*

Die Modellierung der topologischen Information erfolgt hier über die drei Relationen Raum - Raum, Raum - Bauteil und Bauteil - Bauteil.

1.4.6 Abstrakte Modelle

Unter abstrakten Modellen werden hier Modelle verstanden, die eine räumliche Struktur unabhängig von der Interpretation dieser Räume als „geographisch“ oder „architektonisch“ modellieren. So könnte man etwa die „Raum“ - „Raum“ - Beziehung oder die „Raum“ - „Bauteil“ - Beziehung abstrakt als eine Beziehung von Volumenkörpern auffassen. Die Interpretation der Volumenkörper als Raum oder Bauteil lässt sich dann durch zusätzliche Attribute realisieren, die dann einer anderen Abstraktionsschicht des Modells zugeordnet werden. In Anlehnung an die Sprechweise in der Geoinformatik könnten wir derartige räumliche Pläne mit zusätzlichen fachspezifischen Informationen als „thematische Pläne“ bezeichnen.

In dieser Kategorie werden hier unter anderem die Geoinformationssysteme (GIS), die Volumenmodellierer und die G-Maps gesehen.

1.4.7 Geoinformationssysteme

Ein Geoinformationssystem ist ein Informationssystem zur Verarbeitung geographischer Daten. Als Informationssystem versteht man im allgemeinen eine Kombination aus Hard- und Software und die „Verarbeitung“ von Daten umfasst unter anderem die Tätigkeiten Erfassen, Speichern, Abfragen, Bearbeiten, Darstellen und Löschen.

Eine ähnliche Definition finden wir auch in der Literatur, wie etwa bei BILL:

Definition 1.1 (geographisches Informationssystem (GIS)). *Ein Geo-Informationssystem ist ein rechnergestütztes System [...] ⁴. Mit ihm können raumbezogene Daten digital erfasst und redigiert, gespeichert und reorganisiert, modelliert und analysiert sowie alphanumerisch und graphisch präsentiert werden. [BF],*

Dabei ist „Raum“ oder „raumbezogen“ die branchenübliche Bezeichnung für geographische Räume. Diese stellt man sich, bis auf geringfügige Abweichungen, traditionellerweise zweidimensional vor.

Die Geoinformatik ist die einschlägige wissenschaftliche Disziplin, die sich mit den Grundlagen zur Entwicklung und Anwendung von GIS beschäftigt. Dort ist es üblich, stets streng zwischen dem räumlichen Modell als solchem und dessen Anreicherung mit fachspezifischen Zusatzinformationen zu unterscheiden. Derart angereicherte Modelle heißen dann *thematisch*.

Ein beliebtes traditionelles Datenmodell der GIS ist das *Dual Independent Map Encoding* (DIME), das zur Auswertung der Volkszählung 1970 in den U.S.A. entwickelt wurde [The97]. Dieses Modell ist noch heute in Gebrauch und soll hier kurz vorgestellt werden.

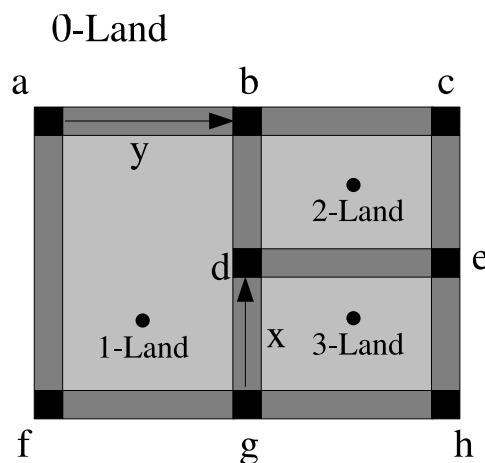


Abbildung 1.4: Eine „Region“ untergliedert in drei „Länder“. Die Grenzlinie x hat links das 1-Land, rechts das 3-Land und Anfangspunkt g und Endpunkt d . Man könnte auch „Stube“ anstelle von „1-Land“ sagen.

Ein geographischer Raum wird als zweidimensionale Fläche betrachtet, die in mehrere Regionen gegliedert ist. Dabei ist es freigestellt, was diese Region darstellen soll: Es kann sich dabei um Länder, um Parzellen in einer Kommune oder um die Zellen auf der mikroskopischen Aufnahme eines Zwiebelhäutchens handeln. In jedem Fall entstehen bei dieser Untergliederung zweidimensionale Flächen, eindimensionale Flächengrenzen und die nulldimensionalen Knotenpunkte, an denen mehrere Flächengrenzen zusammentreffen. Nun hat jeder Abschnitt einer Grenzlinie einen Anfangs- und einen Endpunkt und je ein Land

⁴Auslassung durch den Autor

auf der linken und ein Land auf der rechten Seite. DIME besteht nun aus je einer Liste der Flächen, der Grenzabschnitte und der Knotenpunkte (z.B. in der Form von Referenznummern) und einer Zuordnung, die jedem Grenzabschnitt jeweils den Anfangs- und Endpunkt und die linke und die rechte Fläche zuordnet. Abbildung 1.4 zeigt ein Beispiel einer Region mit einer naheliegenden Interpretation als Geschossgrundriss. Die entsprechenden Datensätze in DIME für die „Ländergrenzen“ x und y sind:

Kante	links	rechts	von	zu
x	1	3	g	d
y	0	1	a	b

Wir können uns nun gut vorstellen, dass nach Entfernen eines der „Länder“ ein Loch im Plan entsteht. Die Grenzlinien dieses Lochs bilden dessen Rand und laufen als Zyklus um dieses Loch herum. Wenn wir z.B. entgegen dem Uhrzeigersinn um 1-Land laufen, dann laufen wir entlang der Grenzlinie x von g nach d und irgendwann später entlang von y in „verkehrter“ Richtung der Grenzlinie von b nach a . Wir können aber auch die Richtung von y umkehren und erhalten so den modifizierten Datensatz $-y | 1 | 0 | b | a$ durch Tauschen von Anfangs- und Endpunkt und von der linken und rechten Seite. Das vorangestellte negative Vorzeichen soll diese Vertauschung symbolisieren. Insbesondere ist $-(-y) = y$, da wir durch erneutes Tauschen der beiden Punkte und der beiden Seiten den alten Datensatz wieder hergestellt haben.

An der DIME-Struktur fällt nun eine Symmetrie auf: Wenn wir bei den Attributbezeichnern „links“ mit „von“ und „rechts“ mit „zu“ tauschen, dann erhalten wir ebenfalls eine DIME-Struktur zu einer gültigen Landkarte, die einen entsprechenden dualen Raum codiert. In jedes Land können wir nämlich genau einen Punkt setzen und dann die Punkte zweier benachbarter Länder durch eine Linie so verbinden, dass die neue Verbindungslinie die gemeinsame Ländergrenze genau an einem Punkt schneidet. Dabei entsteht um jeden Eckpunkt herum ein neues Land und insgesamt eine zu der ursprünglichen Landkarte duale Landkarte. Zu jeder DIME-Struktur gibt es also eine *duale* Struktur, und diese ist ebenfalls ein DIME.

Wir erinnern uns an unseren Kommentar zum RATAS-Modell, wonach „Loch sein“ eine fundamentale Eigenschaft einer Klasse von topologischen Räumen ist, die als *Komplexe* bezeichnet werden. DIME ist nun ein Datenmodell für ebensolche Komplexe. Da dieses Modell jedoch stark auf die Besonderheiten eines zweidimensionalen Raums zugeschnitten ist, lässt es sich in dieser Form schlecht auf Räume höherer Dimension verallgemeinern.

Es gibt noch weitere Datenmodelle für GIS, wie etwa die PLA-Strukturen, die einen inversen Blickpunkt haben und bei den Punkten beginnen. Das „Loch“ entsteht hier durch die Entfernung eines Knotenpunktes, und der Zyklus ist die Abfolge der Kanten, die sternförmig von diesem Punkt ausgehen [Hid95]. Die PLA-Struktur lässt sich nun leicht auf höhere Dimension verallgemeinern, wie von Autor dieser Arbeit gezeigt wurde [PB03].

Die Beschränkung auf zweidimensionale Räume sind natürlich für die Geoinformatik mit ihrer intensiven Forschungstätigkeit bald Geschichte. Aktuelle Forschungen zielen ebenfalls auf eine höherdimensionale Verallgemeinerung des räumlichen Modells ab[Bre01, OS01]. Diese räumlichen Modelle basieren dann überwiegend auf den topologischen Komplexen. Es ist daher zu vermuten, dass

die Zukunft eine Annäherung von GIS und Gebäudeinformationssystemen bringen wird. Eventuell werden sich dann künftige Gebäudemodelle am abstrakteren Zugang der Geoinformatiker beim räumlichen Modellieren orientieren. Erste Schritte dieser Annäherung geht auch OZEL mit ihrem Ansatz, Grundrisse mit einem GIS zu zeichnen und dessen räumliche Analysemöglichkeiten für die Gebäudeplanung zu nutzen[Oze00].

1.4.8 Volumenmodellierer

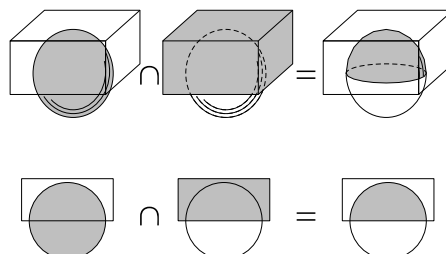
Die ursprünglichen CAD-Systeme waren eine Imitation der Zeichenbretter am Computer und sind insofern mit den ersten Automobilen vergleichbar. Diese hatten zunächst ebenfalls das Aussehen ihrer Vorgänger und glichen Kutschen, denen die Pferde irgendwie abhanden kamen. Danach folgte eine Phase, in der die neuen Möglichkeiten dieser Innovation ausgelotet wurden und sich diese zu eigenständigen Lösungen weiterentwickeln. Eine neue Möglichkeit der CAD-Systeme war, realistischere Modelle von dreidimensionalen Objekten zu erzeugen, als dies auf Papier möglich war. Die Entwicklung eines dreidimensionalen CAD setzte beispielsweise sofort mit der Entwicklung des anerkannt ersten CAD-Systems, SKETCHPAD, ein:

The methods outlined in this report generalize nicely to three dimensional drawing. In fact, work has already begun to make a complete "Sketchpad Three" which will let the user communicate solid objects to the computer.

(SUTHERLAND [Sut63, S.138f], in Buchform: [Sut80])

Konstruktive Volumengeometrie vs. Randdarstellung

Es gibt gegenwärtig u.A. zwei gängige Methoden, Volumenmodelle zu erstellen: Konstruktive Volumengeometrie (Constructive Solid Geometry CSG) und Randdarstellung (Boundary Representation B-Rep). In der Konstruktiven Volumengeometrie werden vom System primitive Volumenkörper, wie Quader oder Würfel, zur Verfügung gestellt. Komplexere Volumenkörper lassen sich dann durch mengentheoretische Operationen an diesen Grundkörpern konstruieren. Eine Halbkugel HK ist z.B. aus dem Schnitt Kugel K mit einem geeigneten Quader Q darstellbar, also $HK = K \cap Q$. Größe und Lage der Grundkörper im Raum wird durch die Angabe entsprechender Parameter wie z.B. Kugelmitelpunkt und Radius festgelegt. Folgende Abbildung zeigt dies in 2D und in 3D:



Volumenkörper werden dann als entsprechende algebraische Ausdrücke gespeichert.

Randdarstellung in der Winged-Edge-Struktur

Die andere Alternative der Modellierung von Volumenkörpern ist die Modellierung der Oberflächen eben dieser Körper. Die Oberfläche eines geometrischen Volumens⁵ im dreidimensionalen Raum wird in der Topologie als *Rand* dieses Volumens bezeichnet. Dieser Rand selbst ist zweidimensional, kann also mit Mitteln, die zu denen der GIS analog sind, modelliert werden. Daher stammt dann auch die Bezeichnung dieses Modellansatzes: Randdarstellung oder B-REP vom Englischen „boundary representation“.

Wenn wir uns nun einen durch ebene Flächen begrenzten Volumenkörper vorstellen, dann haben wir auf dessen Oberfläche etwas ähnliches wie eine Landkarte: flächige Facetten, begrenzt von Kanten und diese Kanten wiederum münden in Ecken. Es überrascht daher nicht, dass eine besonders populäre Datenstruktur für Volumenkörper, die sogenannte Winged-Edge-Struktur [Bau75] [Män88, S. 106ff] eine gewisse formale Ähnlichkeit mit der DIME-Struktur des GIS hat.

Winged-Edge hat ebenfalls je eine Liste von Flächen, Kanten und Knoten. Jedem Knoten ist seine Lage im Raum als x , y und z -Koordinate zugeordnet. Die Zuordnung von Start- und Endknoten wird gleich gehandhabt wie beim DIME. Nur bei der Zuordnung der Fläche zu den Kanten wird etwas anders verfahren. Aus allen Kanten, die den Rand einer Fläche A bilden, wird *genau eine* beliebig ausgewählte Kante $\pm e_A$ der Fläche zugeordnet. Dabei gibt eine Markierung zusätzlich an, ob die Fläche von außen gesehen rechts oder links von der Kante e_A liegt. Da wir bei der Erläuterung der DIME-Struktur einen Richtungs- und Seitenwechsel durch einen Vorzeichenwechsel angezeigt haben, wollen wir dies auch hier tun und rechts ($+e_f$) oder links ($-e_f$) durch ein entsprechendes Vorzeichen angeben.

Nun hat jede der Kanten jeweils eine Fläche auf der linken und eine Fläche auf der rechten Seite. Für jede dieser Flächen wird nun vereinbart, dass (immer von außen gesehen) der Zyklus der Kanten, die diese Fläche umlaufen, im Uhrzeigersinn verläuft. Damit zeigt die Kante einmal in die „richtige“ Richtung und die darauf folgende Kante im Zyklus ist „im Uhrzeigersinn“ und einmal, von der benachbarten Fläche aus gesehen, zeigt die Kante in die „falsche“ Richtung und die darauf folgende Kante zeigt „gegen den Uhrzeigersinn“. Genau diese zwei Kanten werden als Referenzen mit der Bezeichnung *ncw* (next clockwise) und *nccw* (next counterclockwise) der einen Kante zugeordnet. Wir setzen hier $ncw(e) := nccw(-e)$ und damit ist natürlich auch $nccw(e) = ncw(-e)$.

Auf Abbildung 1.5 wird die Winged-Edge-Struktur eines Hausdaches gezeigt. Die Dachfläche A referenziert den Dachfirst $-e_A$ als Anfangskante des Zyklus mit negativem Vorzeichen, da der Dachfirst bezüglich der Fläche A gegen den Uhrzeigersinn orientiert ist (denn er verläuft an deren unteren Rand von links nach rechts). Die nächste Kante im Uhrzeigersinn ist wegen des negativen Vorzeichens $nccw(e_A) = ncw(-e_A)$, die Gratkante links von A . Da $ncw(-e_A)$ im First beginnt und nicht etwa dort endet, ist dessen Vorzeichen positiv, und wir haben

$$-e_A \rightarrow +ncw(-e_A) \rightarrow \dots$$

Die nun folgende Kante $ncw(ncw(-e_A))$ ist die Traufflinie. Da diese nun nicht in $ncw(-e_A)$ beginnt, sondern dort endet, ist ein negatives Vorzeichen zu wählen

⁵„Körper“ bedeutet in der Mathematik etwas völlig anderes.

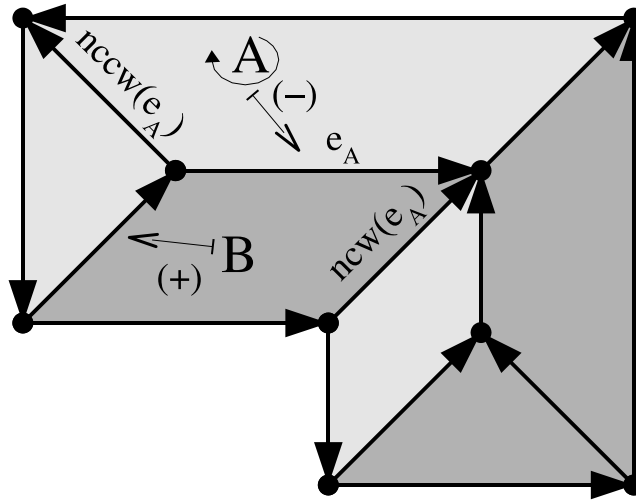


Abbildung 1.5: Die Winged-Edge-Struktur eines Dachfirstes e_A in einem Hausdach

und der Zyklus geht weiter mit

$$-e_A \rightarrow +ncw(-e_A) \rightarrow -ncw(+ncw(-e_A)) \rightarrow \dots$$

Dies wird nun so lange fortgesetzt, bis die ursprüngliche Kante $-e_A$ wieder erreicht ist, und damit der Zyklus vollständig vorliegt als

$$-e_A \rightarrow +ncw(-e_A) \rightarrow -ncw(+ncw(-e_A)) \rightarrow \dots \rightarrow -e_A.$$

Die Vorzeichen geben dabei jeweils an, ob die Kante „mit“ oder „gegen“ den Zyklus orientiert ist. Ohne Vorzeichen lautet die Folge der Randkanten:

$$e_A \rightarrow nccw(e_A) \rightarrow ncw(nccw(e_A)) \rightarrow \dots \rightarrow e_A.$$

Die Fläche B hingegen referenziert den Dachgrat auf ihrer linken Seite mit positivem Vorzeichen, denn bezüglich B verläuft die Orientierung dieses Dachgrates nun mit dem Uhrzeigersinn. Danach setzt sich der Rand von B im First weiter. Dieser verläuft nun bezüglich B ebenfalls im Uhrzeigersinn, erhält positives Vorzeichen $+e_A$ und somit ist die Dachkehle $ncw(e_A)$ die nächste Randkante und daher kommt

$$\dots \rightarrow +e_A \rightarrow -ncw(+e_A) \rightarrow \dots$$

im Zyklus der Randkanten von B vor.

Aus einem DIME ist nun eine Winged-Edge-Struktur leichter durch eine entsprechende relationale Datenbankabfrage erzeugt als umgekehrt. Zur Erzeugung eines DIME aus einer Winged-Edge-Struktur ist es nämlich notwendig, die Kette der Referenzen auf die jeweils nächste Kante (ncw oder $nccw$) eines Zyklus zu verfolgen, und somit – in Begriffen des relationalen Datenmodells ausgedrückt – die transitiven Hüllen der durch ncw und $nccw$ definierten Relationen zu berechnen. Im DIME hingegen lassen sich die Kanten als Vektoren betrachten,

und deren Vektorsumme ergibt den Zyklus. Bei einer Vektoraddition kommt es aber auf die Reihenfolge der Summanden nicht an. Dies bezeichnet man als *Homologie* und wird uns bei den algebraischen Komplexen wieder begegnen.

Bei den Winged-Edge-Strukturen hingegen werden die Kanten wie aufeinanderfolgende Wegabschnitte betrachtet: Das Ende eines Wegabschnittes muss aber mit dem Anfang des nächsten Abschnittes übereinstimmen und damit wird die Reihenfolge der Randelemente wichtig. Dies heißt dann *Homotopie* (von Wegen) und wird in dieser Arbeit ebenfalls behandelt.

Ganz analog zum DIME, könnte man auch bei Winged-Edge die zweidimensionalen Elemente (Flächen) mit den nulldimensionalen Elementen (Kanten) vertauschen und erhielte dann eine zu Winged-Edge duale Datenstruktur, die wir hier als co-Winged-Edge bezeichnen könnten: Die Kanten erhalten anstelle der Referenz auf den Start- und Endpunkt eine Referenz auf die linke und die rechte Fläche. Jeder Eckpunkt v (vertex) hat dann eine vorzeichenbehaftete Referenz auf eine ausgewählte Kante $\pm e_v$ (edge). Dabei bedeutet $+e_v$, die Kante e_v beginnt in v , und $-e_v$ heißt, die Kante e_v endet in v . Die Referenzen *ncw* und *nccw* bleiben unverändert. Ein Kantenzyklus setzt sich dann aus den „Strahlen“ zusammen, die von dem einen Knoten wie Uhrzeiger ausgehen und dann buchstäblich im Uhrzeigersinn um diesen Knoten herumlaufen. Eine solche duale Variante gibt es zu jeder Datenstruktur, die einen Komplex modelliert, und sie modelliert dann den sogenannten co-Komplex. Diese Dualität existiert für eine ganze Klasse von topologischen Räumen, und für genau diese wird hier das Modell der topologischen Datenbanken vorgestellt.

Auch Winged-Edge modelliert einen zweidimensionalen topologischen Raum: die Oberfläche des Volumens. So ist z.B. bei Winged-Edge nicht vorgesehen, dass *drei* Flächen an einer gemeinsamen Kante zusammenstoßen. Zwar gibt es auch Datenstrukturen, die diese Situation beherrschen, wie etwa die Radial-Edge-Struktur [Wei85], dies wollen wir hier aber nicht vertiefen.

Euler-Operatoren

Ein wichtiges Konzept der Volumenmodellierung sind die sogenannten *Euler-Operatoren*, die zusammen mit der Winged-Edge-Struktur entwickelt wurden. Die Euler-Operatoren sollen sicherstellen, dass nur zulässige Veränderungen an den Daten des Volumenmodells vorgenommen werden können, und sind auf jede Datenstruktur anwendbar, die einen Komplex modelliert. Es gibt nämlich gewisse Änderungen in der Datenstruktur eines Volumenmodells, die nicht erlaubt werden dürfen, da sie dann Daten zurücklassen, zu denen es gar keine entsprechende geometrische Figur geben kann, die also inkonsistent sind. Jeder Komplex hat nun eine notwendige Eigenschaft, die sich auch in der Datenstruktur finden muss — die sogenannte Euler-Poincaré-Formel, die wir nun an unserem Dach erläutern wollen: Zählen wir alle Eckpunkte der Dachfläche zusammen, dann haben wir sechs Ecken an der Traufe und drei Ecken im First, also insgesamt neun Eckpunkte. Wir unterstellten hier, dass die Basisfläche unseres Dachvolumens nicht weiter untergliedert ist. Die Anzahl der Kanten ist die Summe von sechs Trauflinien plus fünf Grate plus eine Kehle plus zwei Firste, und dies ergibt 14 Kanten. Wir blicken von oben auf das Dach und sehen sechs schräge Dachflächen, wissen, dass es noch die Basisfläche gibt, und erhalten so

sieben Flächen. Diese Flächen umhüllen ein Volumen. Insgesamt haben wir also

9	0D-Objekte	Eckpunkte
14	1D-Objekte	Kanten
7	2D-Objekte	Flächen
1	3D-Objekt	Volumeninneres

Wenn wir nun jeder Anzahl ein positiven Vorzeichen geben, falls deren entsprechende Dimensionszahl gerade ist, und ein negatives Vorzeichen bei ungerader Dimensionszahl und diese dann addieren, so erhalten wir eine charakteristische Zahl für unseren massiven Volumenkörper:

$$\chi_{\text{Dach}} = 9 - 14 + 7 - 1 = 1$$

Bei Volumenmodellierern wird üblicherweise das Volumeninnere nicht als eigenständiges Element gesehen und nur die Schale um diese Volumen herum betrachtet [Män88, S.43]. Wir erinnern uns, dass unser Modellansatz als „Randdarstellung“ bezeichnet wird. Wir wollen nun diese Schale eines Volumens V mit ∂V bezeichnen, es den Anderen gleich tun und das Volumeninnere vergessen. Dann erhalten wir

$$\chi_{\partial\text{Dach}} = 9 - 14 + 7 = 2$$

als charakteristischen Zahl (gr. $\chi\alpha\rho\alpha\kappa\tau\acute{\eta}\rho$ = Charakter) für die Schale, die unser Volumen umgibt. Entfernen wir nun eine beliebige Fläche, dann erhalten wir eine defekte Dachschale dD ohne Hohlraum. Deren charakteristische Zahl ist nun

$$\chi_{dD} = 9 - 14 + 6 = 1$$

Wenn wir weitere Flächen entfernen, dann entsteht mit jeder entfernten Fläche ein neues „Loch“ (Schleife) im Komplex und wenn wir weitere Kanten hinzufügen, dann erhalten wir ebenfalls je eine neue Schleife. Sind alle Dachflächen (inklusive Basisfläche) entfernt, dann haben wir ein Kantengerüst mit 6 Löchern. Durch Entfernen einer weiteren an einer Schleife beteiligten Kante reduziert sich die Anzahl dieser Löcher. Ist die Kante jedoch nicht Teil einer solchen Schleife und entfernen wir diese, dann trennen wir einen Teil vom Modell ab, die Zahl der Teile erhöht sich dann also um eins. Entfernen wir etwa die sechs Kanten entlang der Traufe, dann haben wir nur die acht Grat- und Firstkanten und keine weiteren Löcher. Wenn wir dann noch eine Kante entfernen, erhalten wir zwei isolierte Teile unseres Daches. Wollen wir nun einen zusammenhängenden Teil eines Komplexes als 0-dimensionalen „Zyklus“ bezeichnen, eine Schleife als 1-dimensionalen Zyklus und eine Schale, die sich um einen Hohlraum legt als 2-dimensionalen Zyklus. Dann hat unsere Schale des Dachvolumens insgesamt

1	0D-Zyklus	Die Schale besteht aus einem zusammenhängenden Teil
0	1D-Zyklen	Es gibt keine Löcher
1	2D-Zyklus	Es gibt einen Hohlraum

Auch hier geben wir diesen Zahlen je nach Dimension entweder ein positives oder ein negatives Vorzeichen und es gilt:

$$\chi_{\partial\text{Dach}} = 1 - 0 + 1 = 2$$

Die Euler-Charakteristik kann also entweder mit den Anzahlen der Löcher oder mit den Anzahlen der Elemente (Flächen, Kanten, Knoten) berechnet werden, denn stets gilt

$$\chi_C = \text{Ecken} - \text{Kanten} + \text{Flächen} = \text{Teile} - \text{Schleifen} + \text{Hohlräume}.$$

Jede Veränderung an der Datenstruktur, die ein Element hinzufügt oder entfernt, verändert auch die Zyklenzahlen im Modell, und jede Veränderung, welche die Zyklenzahlen unverändert lässt, besteht darin, Zellen entweder zu teilen (split) oder sie zusammenzufassen (join). So kann z.B. die Dachfläche in einzelne Dachziegel zerlegt werden (split) oder unser Dach in ein Tonnendach verwandelt werden, indem die obere Firstlinie mit den zwei jeweils angrenzenden Flächen zu einer durchgehenden Fläche vereinigt wird (join).

Wir erlauben also nur diejenigen Veränderungen am Modell, welche die Eigenschaft

$$\text{Ecken} - \text{Kanten} + \text{Flächen} = \text{Teile} - \text{Schleifen} + \text{Hohlräume}$$

erhalten.

So gilt z.B. für eine leere Datenbasis, noch bevor irgendwelche Daten enthalten sind: Sie hat 0 Ecken, 0 Kanten, 0 Flächen, besteht aus 0 Teilen mit 0 Schleifen und 0 Hohlräumen, und es gilt in der Tat

$$0 - 0 + 0 = 0 - 0 + 0.$$

Wir können nun die Gleichungsausdrücke der Form

$$E - K + F = T - S + H$$

mit Eckenanzahl E , Knotenzahl K etc. unserer Datenbasis zuordnen. Wenn wir diese durch Hinzufügen einer Kante verändern, dann kann dabei entweder eine neue Schleife entstehen:

$$0 - 1 + 0 = 0 - 1 + 0$$

oder es werden zwei bisher unverbundene Teile zu einem gemeinsamen Teil verbunden:

$$0 - 1 + 0 = (-1) - 0 + 0$$

Die erste Gleichung beschreibt den Euler-Operator „Make Edge Make Loop“ oder kurz MEML, die zweite beschreibt „Make Edge Kill Ring“ oder kurz MEKR. „M“ entspricht also einem positiven Vorzeichen und „K“ entsprechend einem negativen Vorzeichen. Die Anwendung eines Euler-Operators auf die Datenbasis entspricht der Addition der entsprechenden Gleichungen und es ist

$$\begin{aligned} \partial\text{Dach} : & \quad (E - K + F = T - S + H) \\ +\text{MEML} : & \quad (0 - 1 + 0 = 0 - 1 + 0) \\ =\partial\text{Dach} + \text{MEML} : & \quad (E - (K + 1) + F = T - (S + 1) + H). \end{aligned}$$

Fügen wir unserem Dach eine weitere Ecke hinzu, dann entsteht auch ein neuer zusammenhängender Teil (Make Edge Make Shell), und es gilt

$$\begin{aligned} (9 - 14 + 6 = 1 - 0 + 1) \\ +(1 - 0 + 0 = 1 - 0 + 0) \\ = (10 - 0 + 0 = 2 - 0 + 0). \end{aligned}$$

Das Entfernen einer Dachfläche hingegen beseitigt einen Hohlraum:

$$\begin{aligned} &(9 - 14 + 6 = 1 - 0 + 1) \\ &-(0 - 0 + 1 = 0 - 0 + 1) \\ &= (9 - 14 + 5 = 1 - 0 + 0), \end{aligned}$$

wobei die letzte 0 in der letzten Zeile für den durch Entfernen einer Fläche verschwundenen Hohlraum steht. Das Entfernen einer weiteren Fläche geht nur durch Erzeugen einer Schleife, denn wir werden negative Anzahlen nicht erlauben und müssen folglich „Kill Face Make Ring“ verwenden:

$$\begin{aligned} &(9 - 14 + 5 = 1 - 0 + 0) \\ &-(0 - 0 + 1 = 1 - 0 + 0) \\ &= (9 - 14 + 4 = 0 - 0 + 0), \end{aligned}$$

Offensichtlich können wir mit den Euler-Poincaré-Formeln rechnen, als ob es sich um Vektoren handelt und sie bilden in der Tat etwas Vektorraum-Ähnliches, das als *Modul* bezeichnet wird. Euler-Operatoren sind nun elementare Veränderungen des Volumenmodells, die, in geeigneter Weise hintereinander ausgeführt, jede zulässige Veränderungen am Modell ermöglichen und jede unzulässige Veränderung verhindern. Sie bilden gewissermaßen ein Koordinatensystem (genauer: eine Basis) für den Raum der Euler-Poincaré-Formeln unseres Volumenmodells und sind so eine Grundlage der Zugriffe auf dessen Daten.

1.4.9 G-Maps

Einen interessanten Ansatz, räumliche Daten für beliebige Dimension zu modellieren, verfolgt LIENHARDT mit seinen mehrdimensional verallgemeinerten Karten (generalized Maps oder kurz: „G-Maps“). Die Grundobjekte sind sogenannte „Darts“ (Stachel). Zunächst wird durch eine beliebige Zahl n angegeben, welche Dimension das Modell haben soll. Damit haben wir schon eine wesentliche Eigenschaft der G-Maps angedeutet: Das Modell ist bezüglich seiner Dimension nicht nach oben beschränkt. Dann wird für jede Dimensionszahl i von 0 bis n eine Paarung α_i von Stacheln definiert, die jedem Stachel x seinen „Partner“ $\alpha_i(x)$ zuordnet. Insbesondere ist stets $\alpha_i(\alpha_i(x)) = x$, jeder Partner ist selbst Partner seines Partners und „geht“ also bezüglich der Dimension i „nicht fremd“. Eine solche Paarung heißt *Involution*. Es werden also für ein n -dimensionales Modell $n + 1$ Involutionen auf den Stacheln definiert. Wenn wir nun eine bestimmte Dimensionszahl i und einen Stachel d festhalten, dann könnte die Frage lauten: „Wer sind denn nun die Partner $\alpha_j(d)$ und deren Partner-Partner $\alpha_k(\alpha_j(d))$ u.s.w. bezüglich aller *anderen* Dimensionszahlen $j, k, \text{etc.}$ “ Die Menge all dieser Stachel wird als die i -Bahn⁶ dieses Stachels bezeichnet, und diese Menge bildet das i -dimensionale *Element*, dem dieser Stachel angehört. Damit sind mehrere sogenannte Äquivalenzrelationen auf der Stachelmenge definiert, die für jede Dimensionszahl i die Menge aller Stachel in ein sogenannte Äquivalenzklasse zerlegt. Die 0-Bahnen entsprechen dann den Eckpunkten, die 1-Bahnen den Kanten, die 2-Bahnen sind die Flächen und die 4-Bahnen könnten wir als Hypervolumen bezeichnen. Eine geeignete Bezeichnung für die 3-Bahnen ist naheliegend.

⁶Englisch: Orbit

Zwei derartige Bahnen sind nun „direkt“ miteinander verbunden⁷, wenn sie mindestens einen gemeinsamen Stachel haben.

Formal ist die G-Map-Struktur eine konsequente Fortsetzung der sog. Half-Edge oder Half-Plane Strukturen: Jede (orientierbare) Fläche hat im dreidimensionalen Raum eine Vorder- und eine Rückseite. Damit kann diese Fläche als ein Paar, bestehend aus deren Vorderseite v und deren Rückseite r , betrachtet werden. Dieses Paar bildet dann zwei miteinander verbundene Halbflächen. Deren Verbindung wird durch die Involution α_3 definiert. Es ist $\alpha_3(v) = r$ und umgekehrt $\alpha_3(r) = v$. Die Randkanten und Randecken werden dabei ebenfalls mithalbiert. Wenden wir dies auf alle Flächen an, dann ist jede Halbkante nur noch Verbindungs- bzw. Grenzkante von höchstens zwei Halbflächen. Ist nun aber eine Kante gemeinsame Kante von zwei Flächen, dann kann diese ebenfalls in zwei Halbkanten (oder besser: „Viertelkanten“) mit der entsprechenden Paarung α_2 aufgeteilt werden. Aus jedem Punkt, der zwei Kanten verbindet, wird danach ein Paar Halbpunkte: je ein Halbpunkt pro Kante, verbunden durch die Paarung α_1 . Die letzte Paarung α_0 assoziiert jeweils diejenigen zwei Halbpunkte, die sich entlang einer Halbkante gegenüber liegen. Diese verbleibenden Halbpunkte bilden dann unsere Stachel. Auf Abbildung 1.6 ist dies an unserem Beispieldach illustriert. Insbesondere zeigt sich dort, dass eine derartige Bahn relativ groß werden kann. Der mittlere Firstpunkt ist z.B. eine 0-Bahn, die aus acht Stacheln besteht.

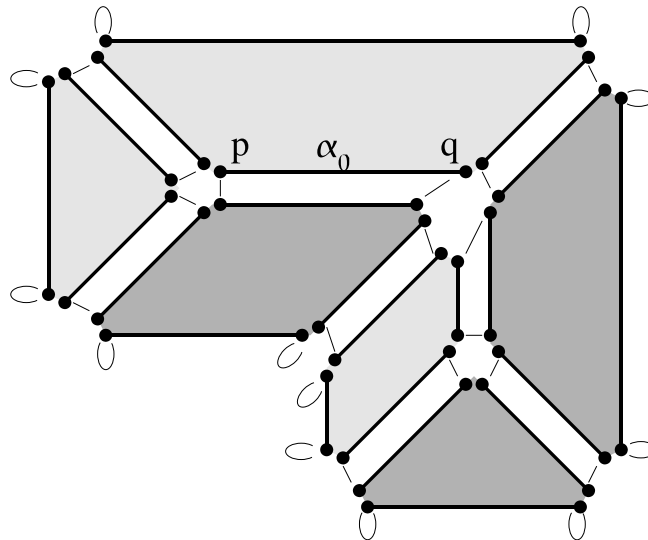


Abbildung 1.6: Die obere Fläche unseres Beispieldachs in G-Map-Darstellung. Die kleinen Knötchen sind die Stachel. Die Paarung α_2 verbindet Flächenelemente und ist durch dünne Linien dargestellt. In der Traufe gilt gelegentlich $\alpha_2(t) = t$, dargestellt durch kleine dünne Schleifen. Die Paarung α_1 verbindet jeweils die in der Ecke eines Flächenelements nahe beieinander liegenden Stachel. α_0 ist durch die dick gezeichneten Linien entlang der (Halb-)Kanten dargestellt. Auf dem First gilt z.B. $\alpha_0(p) = q$ und $\alpha_0(q) = p$.

⁷Wir werden später die Sprechweise „Ein Element berührt das andere.“ verwenden

Einschätzung Die Besonderheit der G-Maps ist, dass sie einerseits sehr ausführlich theoretisch begründet werden und dass andererseits die Dimension des Modells nicht beschränkt ist. Auf der anderen Seite machen G-Maps aber einen etwas umständlichen Eindruck. Insbesondere erscheint es etwas mühsam, einem Element, das als *Menge* modelliert ist, zusätzliche Sachdaten wie Farbe, Spannung, Temperatur etc. zuzuordnen.

In den Publikationen zur Volumenmodellierung findet sich sehr häufig ein eigentümlicher Jargon um topologische Begriffe herum, der relativ weit von den entsprechenden Sprechweise in der Mathematik entfernt ist. So wird z.B. selten erklärt, wie in der Datenstruktur die offenen oder die abgeschlossenen Mengen definiert sind. LIENHARDT ist hier eine bemerkenswerte Ausnahme. Er legt großen Wert auf eine mathematische Begründung seiner Datenstruktur. Insbesondere definiert z.B. abgeschlossene Mengen ausdrücklich in [Lie94, Seite 279]. Er erwähnt zudem, dass es zu eine n -dimensionalen G-Map mit Involutionen $\alpha_0, \dots, \alpha_n$ eine dazu duale n -dimensionale G-Map mit Involutionen β_0, \dots, β_n gibt, indem einfach $\beta_i := \alpha_{n-i}$ gesetzt wird, also durch einfaches Umkehren der Nummerierung der Paarungen [Lie94, Abs. 5.1].

Bemerkenswert ist zudem, dass in der Gruppe um LIENHARDT Gebäude als G-Maps modelliert und diese Modelle als *Architektonische Komplexe*⁸ bezeichnet werden [FML02]. Wir erinnern uns, dass „Komplex“ ein topologischer Begriff ist. Zudem entspricht diese Vorgehensweise genau der Zielsetzung der vorliegenden Arbeit, räumliche Struktur und Semantik jeweils unabhängig zu modellieren.

⁸FRADIN: complexes architecturaux

Kapitel 2

Mathematische Grundlagen

Nachdem wir gesehen haben, dass viele der vorgestellten Modelle topologische Eigenschaften darstellen, wollen wir hier erläutern, was unter dem Begriff „Topologie“ eigentlich zu verstehen ist. Bevor wir jedoch auf die Begriffe wie „Topologie“ oder „topologisch“ eingehen, werden wir die dazu notwendigen mathematischen Grundbegriffe einführen. Zudem wird hier das Konzept der relationalen Datenbanken und der Begriff „Kategorie“ vorgestellt. Diese sind zum Verständnis der vorliegenden Arbeit notwendig. Es werden dazu überwiegend Beispiele aus der Architektur verwendet, um die Anschauung der Leserinnen von Anfang an in die gewünschte Richtung zu lenken. Nun hat es sich mittlerweile in der Mathematik eingebürgert, die Anschauung zwar möglichst oft zu Hilfe zu nehmen, um das Verständnis zu erleichtern, grundsätzlich jedoch nie anschaulich sondern stets formal zu argumentieren. Derart formale Argumente sind aber oft durch reine Lektüre nicht nachvollziehbar. Vielmehr ist es zum Verständnis oft erforderlich, sich beim Lesen begleitende Notizen zu machen und die Argumentation selbst handschriftlich nachzuvollziehen.

Dies ist in gewisser Weise die Fortsetzung der Dokumentation zum Stand der Technik — hier jedoch beschränkt auf die Mathematik. Die Einführung in die Topologie beginnt mit dem Kapitel 2.2 „Metrische Räume“.

2.1 Mengen, Relationen und Abbildungen

Mengen, Relationen und Abbildungen sind die fundamentalen Begriffe in der Mathematik. Wir fassen Dinge zu Mengen zusammen, stellen derartige Mengen zueinander in Beziehung. Eine derartige Beziehung einer Menge X zu einer Menge Y kann dabei manchmal auch als eine Y -Eigenschaft der Elemente in X aufgefasst werden.

2.1.1 Mengen

Definition 2.1 (Menge). *„Unter einer ‚Menge‘ verstehen wir jede Zusammenfassung M von bestimmten wohlunterschiedenen Objecten m unserer Anschauung oder unseres Denkens (welche die ‚Elemente‘ von M genannt werden) zu einem Ganzen.“*

Ist m ein Element der Menge M schreiben wir $m \in M$, ansonsten schreiben wir $m \notin M$.

Anstelle von $m \in M$ kann man auch $M \ni m$ schreiben und entsprechend sagen „ M enthält m “. Analoges gilt für $m \notin M$, das man auch in der Form $M \not\ni m$ schreiben kann. Mengen bezeichnen wir i.d.R. mit Großbuchstaben wie A, B, X . Gelegentlich sagen wir statt „ m ist Element von M “ auch „ m ist ein Punkt in M “ — gemeint ist jedes Mal das Selbe.

Der erste Satz der Definition ist die originale Definition von Georg Cantor (1845-1918) [Can95, Can62]. Nach gegenwärtigem Standard ist dies streng genommen als Definition nicht präzise genug, aber für unsere Zwecke ist sie ausreichend. Lässt man nun zu, dass die Elemente einer Menge selbst wieder Mengen sind, muss man aufpassen, da sich aus einem Ding, wie „die Menge aller Mengen“ Widersprüche konstruieren lassen. Andererseits benötigen wir – insbesondere für die Topologie – häufig Mengen, deren Elemente wiederum Mengen sind. Eine Möglichkeit, widersprüchliche Konstruktionen auszuschließen, ist der sogenannte Stufenkalkül: Wir beginnen mit einer Menge U – das Universum – die Menge aller Objekte, die selbst keine Elemente enthalten. Diese haben Stufe 0 und U hat Stufe 1. Jede weitere Menge hat nun ebenfalls eine bestimmte Stufe $n + 1$ und ihre Elemente dürfen dann höchstens Stufe n haben. Zudem gibt es für jede natürliche Zahl n ein Universum $U^{(n)}$ der Stufe n . Es ist üblich Objekte der Stufe 0 in Kleinbuchstaben zu schreiben (z.B. x, y), Objekte der Stufe 1 (Mengen) in Großbuchstaben (z.B. A, B, X) und Mengen von Mengen in Schönschrift (z.B. $\mathcal{A}, \mathcal{B}, \mathcal{T}$). Mengen von Mengen bezeichnet man auch als *Mengensysteme*.

Für unsere Zwecke reicht es, sich mit der obigen Definition von „Menge“ zu begnügen. Eine derartige Mengenlehre heißt „naive Mengenlehre“. Später werden wir noch den etwas allgemeineren Begriff der *Klasse* verwenden, der uns erlaubt z.B. von der „Klasse aller Mengen“ zu sprechen.

Beispiele 2.1 (Mengen). *Es folgen einige Beispiele für Mengen:*

- Die Einrichtungsgegenstände im Raum 207 dieses Gebäudes
- Die Menge, die aus den Elementen 1, 2 und 7 besteht.
- Die leere Menge, also die Menge die kein Element enthält. In Zeichen: \emptyset oder $\{\}$.
- Die Menge mit den Elementen Zimmer, Türe und Flur.

Mengen kann man definieren durch explizite Angabe der Elemente:

$$\{\text{Zimmer, Türe, Flur}\}$$

oder durch Angabe einer Eigenschaft, die entscheidet ob ein Element einer Menge angehört:

$$E = \{x \mid x \text{ ist Einrichtungsgegenstand und } x \text{ befindet sich in Raum 207}\}.$$

Dann ist die Aussage $x \in E$ genau dann wahr, wenn x ein Einrichtungsgegenstand im Raum 207 ist.

Eine wichtige Menge in unserem Zusammenhang ist die Menge der natürlichen Zahlen $\mathbb{N} = \{0, 1, 2, \dots\}$. Wir wollen hier festlegen $0 \in \mathbb{N}$, dass also Null

eine natürliche Zahl ist. Für natürliche Zahlen gilt das Induktionsprinzip: Wenn in zwei Sonderfällen eine bestimmte Eigenschaft für natürliche Zahlen gilt, dann gilt diese Eigenschaft für alle natürlichen Zahlen.

Fall 1: Die Zahl 0 hat diese Eigenschaft.

Fall 2: Der Nachfolger einer beliebigen Zahl mit dieser Eigenschaft hat ebenfalls diese Eigenschaft.

Dann hat auch 1 diese Eigenschaft, da es Nachfolger von 0 ist und 0 diese Eigenschaft ja hat. Dann hat auch 2 diese Eigenschaft, da es Nachfolger von 1 mit dieser Eigenschaft ist. Dann hat auch 3 als Nachfolger von 2 diese Eigenschaft und so weiter.

Dieses Prinzip wird im Verlauf dieser Arbeit sehr oft als Beweismethode verwendet. Es leuchtet z.B. ein, dass man durch (ausreichend ausdauerndes) Zählen im Prinzip jede Zahl erreichen kann. Dabei kann niemals eine Zahl „verpasst“ sondern höchstens noch nicht erreicht worden sein:

Fall 1: (Induktionsanfang) Ich beginne mit „0“ und kann also bis 0 zählen.

Fall 2: (Induktionsschritt) Wenn ich bis x zählen kann und zudem noch Ausdauer habe, dann kann ich auch noch eins weiter zählen bis $x + 1$.

Weitere Zahlenmengen. Zu den natürlichen Zahlen benötigen wir noch die Menge \mathbb{Z} der ganzen Zahlen, also

$$\mathbb{Z} := \{a - b \mid a, b \in \mathbb{N}\},$$

die Menge \mathbb{Q} der rationalen Zahlen

$$\mathbb{Q} := \left\{ \frac{a}{b} \mid a, b \in \mathbb{Z}, b \neq 0 \right\}$$

und die Menge \mathbb{R} der reellen Zahlen

$$\mathbb{R} := \left\{ \lim_{i \rightarrow \infty} a_i \mid (a_i) \text{ ist Cauchyfolge in } \mathbb{Q} \right\}.$$

Jede dieser zusätzlichen Zahlenmengen ist eine Erweiterung der vorhergehenden mit dem Zweck, gewisse bis dahin nicht immer mögliche Berechnungen zu ermöglichen. Die ganzen Zahlen erweitern die natürlichen Zahlen derart, dass Subtraktion immer möglich ist. Die rationalen Zahlen erweitern die ganzen Zahlen entsprechend für die Division (außer Division durch 0) und die reellen Zahlen ermöglichen, jeder Cauchyfolge von Rationalzahlen einen Grenzwert zu geben. Dabei ist eine *Cauchyfolge* eine Folge von Elementen, hier Rationalzahlen, die sich auf zunehmend kleiner werdende Umgebungen beschränken, so dass der „Radius“ ε dieser Umgebung beliebig klein, nicht jedoch gleich Null wird. Bei den Rationalzahlen heißen derartige Umgebungen „Intervalle“ und der Radius eines Intervalls $[a, b]$ mit $a < b$ ist einfach dessen halbe Breite $\frac{b-a}{2}$.

Wir wollen hier kurz andeuten, wie diese Zahlenmengen aufeinander aufbauend konstruiert werden. Zudem geben wir an, wie eine Zahl der einen Zahlenmenge als Konstruktion in der nächstgrößeren Zahlenmenge betrachtet wird. Als Beispiel, wie sich Addition, Subtraktion, Multiplikation etc. in diesen Erweiterungen fortsetzen, wird jeweils nur die Definition der Addition mit angegeben:

Ganze Zahlen. Für zwei natürliche Zahlen n, m ist zwar die Addition $n + m$ definiert, aber nicht die dazu inverse Verknüpfung, die Subtraktion $n - m$, denn ist $m > n$, dann hat die Formel $n - m$ kein sinnvolles Ergebnis in den natürlichen Zahlen. Eine ganze Zahl a ist zunächst einfach nur eine solche Formel $[n - m]$, die „Differenz“ $n - m$ zweier natürlicher Zahlen n und m . Die eckigen Klammern sollen andeuten, dass dies zunächst nicht das Ergebnis irgend einer Subtraktion ist, sondern nur ein formelhafter Ausdruck „ $n - m$ “: lediglich zwei Zahlen, nebeneinander geschrieben und mit einem Zeichen „ $-$ “ dazwischen. Für diese formalen Differenzen definieren wir nun Gleichheit und Inverses wie folgt:

$$\begin{aligned} [n - m] &= [a - b] && \text{genau dann wenn} && n + b = m + a \\ -[n - m] &= [m - n]. \end{aligned}$$

Eine positive ganze Zahl $+a$ können wir dann definieren als $[a - 0]$ und eine negative Zahl $-a$ als $[0 - a]$. Die Zahl $+0$ ist die formale Differenz $[0 - 0]$ und daher gilt $+0 = -0$. Jede andere Formel $[n - m]$ mit natürlichen Zahlen n und m ist dann zu einer Differenz $[a - 0]$ oder $[0 - a]$ gleich im Sinne der obigen Definition. Die Addition ganzer Zahlen ist definiert durch $[n - m] + [a - b] := [(n + a) - (m + b)]$. Dabei ist zu überprüfen, ob das Ergebnis dieser Addition „gleich“ bleibt im Sinne der obigen Definition, wenn wir entsprechend „gleiche“ Zahlen addieren. Man sagt dazu auch, die Addition ist *wohldefiniert*, da ihr Ergebnis von der Wahl der sogenannten Repräsentanten a, b bzw. n, m unabhängig ist. Gelte etwa $[a - b] = [x - y]$ im obigen Sinne, also $a + y = b + x$. Dann ist

$$[n - m] + [x - y] = [(n + x) - (m + y)].$$

Dann gilt $a + y = b + x$, denn dies hatten wir vorausgesetzt. Folglich gilt auch $a + y + (n + m) = b + x + (n + m)$ und durch Umstellen der Summanden erhalten wir $(n + x) + (m + b) = (m + y) + (n + a)$ und damit auch

$$[n - m] + [a - b] = [(n + a) - (m + b)] = [(n + x) - (m + y)] = [n - m] + [x - y].$$

Wir dürfen also in der Tat „gleiche“ Summanden auf der *rechten* Seite des Pluszeichens austauschen und erhalten dabei stets auch ein „gleiches“ Ergebnis der Addition. Für den Summanden auf der linken Seite zeigt man dies analog. Besser ist jedoch, sich zu überlegen, dass bei unserer Definition der Addition die Summanden in der Tat vertauscht werden können (Kommutativität). Dann kann die Wohldefiniertheit der Addition mit den linken Summanden $[n - m]$ auf die oben bereits gezeigte Wohldefiniertheit bezüglich des rechten Summanden reduziert werden.

Eine natürliche Zahl n betrachten wir nun als positive ganze Zahl $+n = [n - 0]$. Nehmen wir nun zwei natürliche Zahlen a und b als ganze Zahlen $+a$ und $+b$, dann ist deren Differenz

$$+a - (+b) = [a - 0] - [b - 0] = [a - 0] + [0 - b] = [(a + 0) - (0 + b)] = [a - b]$$

in der Tat gleich ihrer der bloß formalen Differenz.

Rationalzahlen. Analog definiert man die rationalen Zahlen als formale Division $[a : b]$ zweier ganzer Zahlen a und b und fordert noch $b \neq 0$. Statt $[a : b]$ schreiben wir $\frac{a}{b}$ und definieren die Gleichheit derartiger Ausdrücke, indem wir

sagen $\frac{a}{b} = \frac{c}{d}$ gilt genau dann, wenn $a \cdot d = b \cdot c$ gilt. Die Addition ist hier definiert als $\frac{a}{b} + \frac{c}{d} := \frac{(d \cdot a) + (b \cdot c)}{b \cdot d}$ und offensichtlich kommutativ. Ersetzen wir nun $\frac{c}{d}$ durch ein „gleiches“ $\frac{x}{y}$, dann gilt:

$$\frac{a}{b} + \frac{x}{y} = \frac{(y \cdot a) + (b \cdot x)}{b \cdot y}.$$

Da wir $\frac{c}{d}$ und $\frac{x}{y}$ als „gleich“ voraussetzen, gilt $c \cdot y = d \cdot x$ und damit auch

$$\begin{aligned} ((y \cdot a) + (b \cdot x)) \cdot (b \cdot d) &= (y \cdot a) \cdot (b \cdot d) + (b \cdot x) \cdot (b \cdot d) \\ &= (y \cdot a) \cdot (b \cdot d) + (b \cdot b) \cdot (d \cdot x) \\ &= (y \cdot a) \cdot (b \cdot d) + (b \cdot b) \cdot (c \cdot y) \\ &= (a \cdot d) \cdot (b \cdot y) + (b \cdot c) \cdot (b \cdot y) \\ &= ((a \cdot d) + (b \cdot c)) \cdot (b \cdot y). \end{aligned}$$

Unsere Ergebnisse sind also auch „gleich“, und die Addition von Rationalzahlen ist ebenfalls wohldefiniert. Dass auch $\frac{a}{b}$ auf der linken Seite durch beliebige „gleiche“ Brüche ausgetauscht werden kann, überlegt man sich wieder anhand der Kommutativität. Die Leserin kann sich leicht davon überzeugen, dass dann der Sonderfall $\frac{a}{b} + \frac{c}{b} = \frac{a+c}{b}$ gilt.

Jede ganze Zahl z fassen wir nun als Rationalzahl $\frac{z}{1}$ auf.

Reelle Zahlen. Für die reellen Zahlen \mathbb{R} ist die Konstruktion etwas komplizierter. Ordnen wir jeder natürlichen Zahl $i \in \mathbb{N}$ eine Rationalzahl $q_i \in \mathbb{Q}$ zu, dann heißt eine solche Zuordnung (oder Abbildung) auch *Folge* und wir schreiben sie als (q_0, q_1, \dots) oder $(q_i)_{i \in \mathbb{N}}$ oder kurz (q_i) . Eine Folge heißt *Cauchyfolge*, wenn es für jeden „Radius“ $r \in \mathbb{Q}$ mit $r > 0$ eine feste rationale Zahl c_r gibt, so dass irgendwann ab einer natürlichen Zahl n_0 alle weiteren Folgenglieder q_{n_0+i} mit $i \in \mathbb{N}$ in der Nähe dieses c_r bleiben, wenn also für alle $i \in \mathbb{N}$ gilt: $c_r - r < q_{n_0+i} < c_r + r$. Gibt es nun ein festes c , das als ein solches c_r unabhängig von r fest gewählt werden kann, dann heißt c auch Grenzwert oder *Limes* von (q_i) und man schreibt $c = \lim(q_i)$ bzw. – etwas dekorativer – auch $c = \lim_{i \rightarrow \infty} q_i$ und sagt (q_i) konvergiert in \mathbb{Q} gegen c . In den reellen Zahlen soll nun jede Cauchyfolge (a_i) einen solchen Limes besitzen. Dies ist nicht selbstverständlich, denn es gibt z.B. eine Cauchyfolge (r_i) von Rationalzahlen, so dass die Folge der Quadrate (r_i^2) gegen 2 konvergiert. Allerdings hat dann (r_i) keinen Limes in \mathbb{Q} , denn sei $r \in \mathbb{Q}$ dieser Limes, dann würde $r^2 = 2$ gelten. Bekanntermaßen existiert eine solche rationale Zahl r aber nicht. Wir setzen zunächst nur formal den Limes $[\lim(r_i)]$, indem wir einfach „lim“ vor die Cauchyfolge schreiben und den Unterschied zum „echten“ Limes wieder durch eckige Klammern kennzeichnen. Dann definieren wir die Gleichheit derartiger formaler Limes-Ausdrücke, indem wir sagen $[\lim(r_i)] = [\lim(s_i)]$ gilt genau dann, wenn die Folge der gliedweisen Differenzen $(r_i - s_i)$ gegen 0 konvergiert, also $\lim(r_i - s_i) = 0$. Hat z.B. eine Folge $\lim(r_i^2)$ den Grenzwert 2, und sind alle $r_i \geq 0$, dann bezeichnet man $[\lim(r_i)]$ auch mit $\sqrt{2}$. Zur Konstruktion von (r_i) gibt es mehrere Möglichkeiten, wie etwa die aus der Schule bekannte Intervallschachtelung, das Newton-Verfahren oder das Regula-Falsi-Verfahren. Auch wenn sich die so konstruierten Folgen in ihren Folgengliedern unterscheiden, so sind sie doch als Limes paarweise gleich im obigen Sinne. Die reellen Zahlen \mathbb{R}

sind nun nichts anderes als diese formalen Grenzwerte von Cauchyfolgen mit der oben definierten Gleichheit. Die Addition von reellen Zahlen ist definiert durch Addition der Folgenglieder $[\lim(a_i)] + [\lim(b_i)] := [\lim(a_i + b_i)]$. Auch hier ist die Wohldefiniertheit dieser Verknüpfung noch zu zeigen: Ist $[\lim(b_i)] = [\lim(c_i)]$, dann ist $\lim(b_i - c_i) = 0$. Wegen $((a_i + b_i) - (a_i + c_i)) = (b_i - c_i)$ folgt gleich $\lim((a_i + b_i) - (a_i + c_i)) = 0$ und somit $[\lim(a_i + b_i)] = [\lim(a_i + c_i)]$.

Jede rationale Zahl $q \in \mathbb{Q}$ ist auch Grenzwert der Folge (q, q, q, \dots) , der sogenannten konstanten Folge, also ist $q = \lim(q, q, q, \dots)$, und damit entspricht q der reellen Zahl mit $q := [\lim(q, q, q, \dots)]$.

Um genauer anzugeben, was dieses „nur formal“ bedeutet und wieso wir dann so großzügig einfach eine Gleichheit definieren dürfen, benötigen wir bisher nicht eingeführte Begriffe wie „Äquivalenzrelation“ und „Quotientenmenge“, auf die aber erst später eingegangen wird. Zudem muss natürlich noch bewiesen werden, dass derart großzügiges Stiften von Gleichheit nichts kaputt macht und das Rechnen mit derart erweiterten Zahlenmengen wie bisher funktioniert. Wir werden hier darauf verzichten.

2.1.2 Teilmenge, Gleichheit von Mengen

Die Gleichheit von Mengen ist durch die Teilmengeneigenschaft definiert. Eine Menge ist Teilmenge einer anderen Menge, wenn diese mindestens alle Elemente von jener enthält.

Definition 2.2 (Teilmenge, Gleichheit von Mengen). *Seien A und B Mengen. Dann heißt A Teilmenge von B , wenn jedes Element x von A auch Element von B ist, wenn also für alle x , für die $x \in A$ gilt, stets auch $x \in B$ gilt. Für „ A ist Teilmenge von B “ schreiben wir $A \subseteq B$ bzw. $B \supseteq A$.*

Zwei Mengen A und B sind gleich, wenn jeweils eine Teilmenge der anderen ist:

$$A = B \quad \text{genau dann wenn} \quad A \subseteq B \quad \text{und} \quad \text{zugleich} \quad B \subseteq A.$$

Dies deutet bereits an, wie man die Gleichheit von Mengen zeigt: Man behauptet $x \in A$ und versucht daraus zu folgern: $x \in B$. Genau dann gilt $A \subseteq B$. Dann geht man den umgekehrten Weg, behauptet $x \in B$, und wenn es gelingt daraus $x \in A$ zu folgern gilt $B \subseteq A$ und somit $A = B$.

Ist A Teilmenge von B und B Teilmenge von C , dann ist A auch Teilmenge von C (Transitivität). Damit kann man bei mehr als zwei Mengen, z.B. A , B und C , deren Gleichheit zeigen, indem man die „Ringinklusion“ $A \subseteq B \subseteq C \subseteq A$ zeigt.

Ist A Teilmenge von B dann sagen wir auch „ B umfasst A “ oder „ A liegt in B “. Wir sagen in diesem Fall jedoch nicht „enthält“, denn dieses Wort soll der Elementeigenschaft vorbehalten bleiben.

Beispiele 2.2 (Teilmengen). *Es folgen einige Beispiele für Teilmengen:*

- $\{x \mid x \text{ ist Stuhl in Raum } 207\} \subseteq \{x \mid x \text{ ist Gegenstand in Raum } 207\}$
- $\{7, 1, 2, 7, 1, 2, 7\} = \{1, 2, 7\} \subseteq \mathbb{N} = \text{Menge der natürlichen Zahlen. Es gibt also keine „Reihenfolge“, „Häufigkeit eines Elements“ etc. in einer Menge.}$
- *Für jede beliebige Menge A gilt: $\emptyset \subseteq A$. Wenn wir behaupten, dass $x \in \emptyset$ gelte, dann ist diese Behauptung falsch. Aus einer falschen Annahme kann*

aber Beliebiges geschlossen werden, insbesondere auch $x \in A$. Dies mag nun wahr oder falsch sein, die Schlussfolgerung „Wenn $x \in \emptyset$ gilt, dann gilt $x \in A$ “ insgesamt ist in jedem Fall wahr.

- Nehmen wir zwei beliebige leere Mengen \emptyset_1 und \emptyset_2 , dann gilt: $\emptyset_1 \subseteq \emptyset_2$ und $\emptyset_2 \subseteq \emptyset_1$ und folglich $\emptyset_1 = \emptyset_2$. Es gibt also nur eine einzige leere Menge \emptyset .
- Für jede Menge M gibt es die Menge $\mathcal{P}(M) = \{A \mid A \subseteq M\}$ — die Menge der Teilmengen von M bzw. die Potenzmenge von M . Manchmal schreibt man auch 2^M . Wir können nämlich die Zahl 2 durch die zweielementige Menge $\{0, 1\}$ oder auch $\{\text{false}, \text{true}\}$ repräsentieren. Wir werden später die Menge Y^X der Abbildungen von einer Menge X zu einer Menge Y kennen lernen. Durch jede Teilmenge $A \subseteq M$ kann nun eine derartige Abbildung x_A von M nach $\{0, 1\}$ definiert werden, indem wir $x_A(m) = 1$ genau dann setzen, wenn m Element von A ist und ansonsten $x_A(m) = 0$. Umgekehrt entspricht jeder derartige Abbildung f eine Teilmenge $M_f := \{m \in M \mid f(m) = 1\}$ von M . Daher können wir $\{0, 1\}^M$ bzw. 2^M mit $\mathcal{P}(M)$ gleichsetzen.
- Für die Zahlenmengen \mathbb{N} , \mathbb{Z} , \mathbb{Q} und \mathbb{R} gilt $\mathbb{N} \subseteq \mathbb{Z}$, $\mathbb{Z} \subseteq \mathbb{Q}$ und $\mathbb{Q} \subseteq \mathbb{R}$.
- Für Zahlenmengen definieren wir auch Teilmengen durch Angabe einer unteren Grenze a und einer oberen Grenze b als

$$\begin{array}{ll} \text{offenes Intervall} &]a, b[:= \{x \mid a < x, x < b\}, \\ \text{halboffene Intervalle} &]a, b] := \{x \mid a < x, x \leq b\} \\ & [a, b[:= \{x \mid a \leq x, x < b\} \end{array}$$

und

$$\text{abgeschlossenes Intervall} \quad [a, b] := \{x \mid a \leq x, x \leq b\}.$$

Die untere Grenze schreiben wir dabei immer links, die obere rechts. Ist $a > b$, dann ist also $[a, b] = \emptyset$.

2.1.3 Mengenoperatoren

Für Mengen sind Verknüpfungen definiert, die es ermöglichen mit Mengen zu „rechnen“ wie mit Zahlen. Dies macht die Mengenlehre so praktisch.

Definition 2.3 (Vereinigung, Schnitt, Rest). Seien A und B Mengen. Dann heißt die Menge $A \cup B := \{x \mid x \in A \text{ oder } x \in B\}$ die Vereinigung von A und B . Die Menge $A \cap B := \{x \mid x \in A \text{ und } x \in B\}$ heißt Schnitt von A und B und $A \setminus B := \{x \mid x \in A \text{ und } x \notin B\}$ der Rest von A ohne B .

Dabei ist „A oder B“ zu verstehen als „entweder A oder B oder beides“ und nicht im umgangssprachlichen Sinne des „entweder – oder“.

Man kann sich leicht klarmachen, dass folgende Beziehungen gelten:

$$\begin{aligned} A \cup B &= B \cup A \\ A \cap B &= B \cap A \\ A &\subseteq A \cup B \\ A \cap B &\subseteq A \\ A \setminus B &\subseteq A \\ (A \cup B) \cup C &= A \cup (B \cup C) \\ (A \cap B) \cap C &= A \cap (B \cap C). \end{aligned}$$

Daher können wir auch beim Schnitt (und bei der Vereinigung) mehrerer Mengen $A_1 \cap A_2 \cap \dots \cap A_n$ die Klammern weglassen und die Reihenfolge beliebig abändern. Dies ermöglicht uns ebenfalls für eine Menge $\mathcal{A} = \{A_1, A_2, \dots, A_n\}$ von Mengen kurz

$$\begin{aligned} \bigcup \mathcal{A} &:= A_1 \cup A_2 \cup \dots \cup A_n \\ \bigcap \mathcal{A} &:= A_1 \cap A_2 \cap \dots \cap A_n \end{aligned}$$

zu schreiben. Auch Schnitt und Vereinigung von beliebigen (auch unendlich großen) Mengensystemen ist so möglich:

$$\begin{aligned} \bigcup \mathcal{A} &= \{x \mid \text{Es gibt ein } A \in \mathcal{A} \text{ mit: } x \in A\} \\ \bigcap \mathcal{A} &= \{x \mid \text{Für jedes } A \in \mathcal{A} \text{ gilt: } x \in A\} \end{aligned}$$

Dies wiederum ermöglicht, Schnitt und Vereinigung von nur einer oder sogar von gar keiner Menge zu definieren:

$$\begin{aligned} \bigcup \{A\} &= A \\ \bigcap \{A\} &= A \\ \bigcup \emptyset &= \emptyset \\ \bigcap \emptyset &= U. \end{aligned}$$

Wenn wir U als Universum voraussetzen.

2.1.4 Relationen

Mengen können nun zueinander in Beziehung gesetzt werden. Nehmen wir etwa eine Menge G , die Menge aller Bauelemente eines bestimmten Gebäudes, und F die Menge aller Firmen, die Bauleistungen erbringen. Dann könnte eine Relation etwa lauten „Bauelement x wurde durch Firma y bearbeitet“. Z.B. „Innenwand w wurde von m bearbeitet“ (m ist Rohbaufirma) und „Innenwand w wurde von p bearbeitet“ (p ist Maler) u.s.w.

Dabei kann ein Bauelement g von mehreren Firmen bearbeitet worden sein und eine Firma f kann mehrere Bauelemente bearbeiten. Zuerst aber wollen wir die Prosa „Bauelement x wurde durch Firma y bearbeitet.“ beseitigen:

Definition 2.4 (geordnetes Paar, Tupel). Die Menge $(a, b) := \{\{a\}, \{a, b\}\}$ heißt geordnetes Paar oder 2-Tupel. $() := \emptyset$ ist ein 0-Tupel und $(a) := \{\{a\}\}$ ein 1-Tupel und für eine beliebige natürliche Zahl $n \in \mathbb{N}$ mit $n > 2$ ist $(a_1, \dots, a_n) := (a_1, (a_2, \dots, a_n))$ ein n -Tupel. a_i heißt i -te Komponente des Tupels $(a_1, \dots, a_i, \dots, a_n)$.

Bei dieser Definition gilt, dass zwei Paare (a, b) und (c, d) dann und nur dann gleich sind, wenn $a = c$ und $b = d$. Insbesondere gilt für $a \neq b$ immer $(a, b) \neq (b, a)$. Analoges gilt für n -Tupel. Die Komponenten dürfen also, im Gegensatz zu den Mengen, im Allgemeinen nicht vertauscht werden und eine Komponente kann auch mehrfach in einem Tupel vorhanden sein.

Nun schreiben wir also kurz (x, y) und interpretieren es als „Bauelement x wurde durch Firma y bearbeitet“. Eine Relation ist einfach eine Menge von derartigen Tupeln. Allerdings ist es sinnvoll anzugeben, aus welchen Mengen die Komponenten stammen:

Definition 2.5 (Relation). Seien A_1 und A_2 u.s.w. bis A_n Mengen. Sei zudem G eine Menge, die ausschließlich aus Tupeln (a_1, a_2, \dots, a_n) mit der Eigenschaft $a_i \in A_i$ für alle $i \in \{1, 2, \dots, n\}$ besteht. Das Paar $R = ((A_1, A_2, \dots, A_n), G)$ heißt dann n -stellige Relation auf A_1, \dots, A_n . Für $n = 2$ sagen wir Relation zwischen A_1 und A_2 und bezeichnen A_1 als Vorbereich und A_2 als Nachbereich von R . Ist dann auch noch $A_1 = A_2$ dann heißt R auch Relation auf A_1 . Ist $R = ((A_1, \dots, A_n), G)$ eine Relation, dann heißt das Tupel (A_1, \dots, A_n) auch Schema von R und G heißt Graph von R , bezeichnet mit $\text{Graph}(R)$.

In diesem Abschnitt werden wir nur auf die zweistelligen Relationen eingehen. Ist R eine solche Relation, dann schreiben wir statt $(a, b) \in \text{Graph}(R)$ auch kurz $(a, b) \in R$, $a R b$ oder $R(a, b)$.

Gilt $a R b$ und $b S c$ und $c T d$ für Relationen R , S und T , dann schreiben wir auch einfach $a R b S c T d$. Ist das Schema einer Relation aus dem Kontext klar, dann werden wir einfach den Graph der Relation als Relation bezeichnen und das Schema weglassen. Dies entspricht auch der häufigen Praxis zur Definition von „Relation“ in den meisten Lehrbüchern.

Für die Theorie der relationalen Datenbanken sind die Relationen beliebiger Stelligkeit wichtig. Eine solche Relation heißt dann auch *Tabelle*, ihre Elemente, die Tupel, heißen auch *Zeilen* oder *Datensätze*.

Sind (T, R) und (T, S) zwei Relationen mit gleichem Schema T , dann werden (T, R) und (T, S) unter Datenbanktheoretikern als *vereinigungsverträglich* bezeichnet. Für zwei vereinigungsverträgliche Relationen (T, R) und (T, S) können wir auch die Mengenoperatoren Vereinigung mit $(T, R) \cup (T, S) := (T, R \cup S)$, Schnitt mit $(T, R) \cap (T, S) := (T, R \cap S)$ und Restmenge mit $(T, R) \setminus (T, S) := (T, R \setminus S)$ definieren, und wir sagen Teilrelation $(T, R) \subseteq (T, S)$ genau dann wenn $R \subseteq S$ gilt. Es sei uns hier jedoch noch vergönnt etwas weniger streng zu sein und auch dann die Vereinigung zuzulassen, wenn $((A_1, \dots, A_n), R)$ und $((B_1, \dots, B_n), S)$ zwei Relationen gleicher Stelligkeit sind. Denn dann können wir setzen

$$\begin{aligned} ((A_1, \dots, A_n), R) \cup ((B_1, \dots, B_n), S) &:= ((A_1 \cup B_1, \dots, A_n \cup B_n), R \cup S) \\ ((A_1, \dots, A_n), R) \cap ((B_1, \dots, B_n), S) &:= ((A_1 \cap B_1, \dots, A_n \cap B_n), R \cap S). \end{aligned}$$

Beispiele 2.3 (Relationen). Hier wollen wir einige Beispiele für Relationen auflisten, wobei wir hier meist nur die Graphen angeben, da das Schema jeweils klar ist:

- Für zwei Mengen A und B ist die Relation

$$A \times B := \{(a, b) \mid a \in A, b \in B\}$$

die Universalrelation zwischen A und B oder das kartesische Produkt von A und B . Jede Relation zwischen A und B ist Teilrelation des kartesischen Produkts $A \times B$. Die Universalrelation ist die mathematische Form der Haltung „alles hängt mit allem irgendwie zusammen“. Diese sagt oft nicht viel aus, so dass die interessanten Relationen meist diejenigen sind, die Stellung beziehen und gelegentlich auch zwei Elemente nicht in Beziehung setzen.

- Für jede Menge A ist $\Delta_A := \{(a, a) \mid a \in A\}$ die Diagonale oder die Gleichheitsrelation in A .
- Ist $R = ((A, B), G)$ eine zweistellige Relation, dann heißt die Relation $R^T := ((B, A), \{(b, a) \mid a R b\})$ die Transponierte von R . Sie entsteht durch Vertauschen der Komponenten. Die Verallgemeinerung der Transposition im relationalen Datenmodell heißt Umbenennung, denn die erste Komponente heißt ja nun „zweite“ und die zweite ist in „erste“ umbenannt worden.
- Sei Z die Menge aller Räume (Zimmer) eines beliebigen Gebäudes H und $G \subseteq Z \times Z$ folgende Relation in Z :

$$G := \{(a, b) \in Z \times Z \mid a \text{ grenzt in } H \text{ an } b\}.$$

z.B. gilt hier $Z = \{A, B, D, C\}$ und

$$H : \begin{array}{|c|c|} \hline & B \\ \hline A & C \\ \hline & D \\ \hline \end{array} \quad G = \left\{ \begin{array}{cccc} (A, B), & (B, A), & (A, C), & (C, A), \\ (A, D), & (D, A), & (B, C), & (C, B), \\ (C, D), & (D, C) & & \end{array} \right\}.$$

- Sei P_V die Menge aller Punkte des Volumens, das ein Gebäude im dreidimensionalen Raum einnimmt inklusive aller Raumvolumen („Zimmer“). Dann können wir die Relation $R_Z \subseteq P_V \times P_V$, die „Zimmerrelation“ folgendermaßen definieren:

$$R_Z := \{(p_1, p_2) \mid p_1 \text{ und } p_2 \text{ liegen in einem gemeinsamen „Zimmer“}\}$$

Gewisse Eigenschaften von Relationen sind besonders wichtig und haben daher besondere Namen:

Definition 2.6 (Eigenschaften von Relationen). Sei A eine Menge und $R \subseteq A \times A$ Relation auf A . Dann heißt R

reflexiv, wenn für alle $a \in A$ gilt: $a R a$ (äquivalent dazu ist: $\Delta_A \subseteq R$),

irreflexiv, wenn es kein $a \in A$ mit $a R a$ gibt, also: $R \setminus \Delta_A = R$,

symmetrisch, wenn für alle $(a, b) \in R$ auch gilt: $(b, a) \in R$,

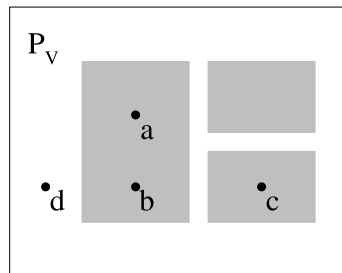


Abbildung 2.1: Die Zimmerrelation $R_Z \subseteq P_V \times P_V$. Es gilt $a R_Z b$ und $c R_Z c$, aber weder gelten $b R_Z c$ noch gibt es ein x aus P_V , so dass $d R_Z x$ gilt.

antisymmetrisch, wenn für alle $a, b \in A$ gilt: falls $a R b$ und $b R a$ dann $a = b$,

transitiv, wenn für alle $a, b, c \in A$ gilt: falls $a R b$ und $b R c$ dann auch $a R c$.

Betrachten wir die obigen Beispiele von Relationen, dann gilt:

- $A \times A$ ist reflexiv, transitiv und symmetrisch.
- Die leere Relation $((A, A), \emptyset)$ ist irreflexiv, symmetrisch, antisymmetrisch und transitiv. Irreflexiv ist klar, aber die anderen Eigenschaften sollen kurz begründet werden, da sie ein wichtiges Prinzip der mathematischen Logik betreffen: Die Definitionen dieser Eigenschaften sagen immer, dass *alle* Paare (a, b) einer Relation R eine gewisse Eigenschaft haben. Dies gilt natürlich immer für die leere Menge, denn da diese überhaupt keine Elemente hat, gibt es natürlich auch keine Elemente, die diese Eigenschaft nicht haben. Sogar die Aussage „Für jedes Element x der leeren Menge gilt: $x \neq x$ “ ist eine wahre Aussage, obwohl $x \neq x$ natürlich stets ein Widerspruch ist. Insbesondere ist die leere Menge wunderschön, denn jedes Element darin riecht nach Bananen.
- Δ_A , die Gleichheitsrelation, ist reflexiv, transitiv und symmetrisch. Grund (Schreibweise: $a = b$ statt $a \Delta_A b$): reflexiv: $a = a$, symmetrisch: gilt $a = b$ dann gilt auch $b = a$, transitiv: gelten $a = b$ und $b = c$ dann gilt auch $a = c$. Die Leserin kann sich überlegen, dass es in der Tat auch eine irreflexive Gleichheitsrelation gibt.
- G aus unserem Beispiel (Raum grenzt an Raum) ist irreflexiv, symmetrisch und nicht transitiv.
- R_Z aus unserem Beispiel (a im selben Zimmer wie b) ist symmetrisch und transitiv aber weder reflexiv noch irreflexiv. Sie ist nicht reflexiv, da es auch Punkte in Bauelementen gibt, die in überhaupt keinem „Zimmer“ liegen. Sie ist auch nicht irreflexiv, denn es gibt einen Punkt p in einem „Zimmer“, dann gilt aber: p liegt im gleichen „Zimmer“ wie p also $p R_Z p$. Allerdings kann R_Z zu einer reflexiven, symmetrischen und transitiven Relation (sogenannte Äquivalenzrelation) erweitert werden, doch davon später.

Die wichtigsten Relationen sind die Äquivalenzrelationen — die Verallgemeinerung der Gleichheitsrelation:

Definition 2.7. (*Äquivalenzrelation*) Sei M eine Menge. Eine Relation $R \subseteq M \times M$ heißt Äquivalenzrelation, wenn R reflexiv, symmetrisch und transitiv ist.

Für Äquivalenzrelationen verwenden wir häufig das Symbol \sim . Die Gleichheit ist ebenfalls eine Äquivalenzrelation mit der einen zusätzlichen Eigenschaft, dass wenn a gleich b ist und a eine gewisse Eigenschaft P hat, dann hat auch b die Eigenschaft P .

Um aus unserer Zimmerrelation R_Z im obigen Beispiel eine Äquivalenzrelation auf P_V zu machen, können wir z.B. die Gleichheitsrelation Δ_{P_V} mit R_Z vereinigen. Die Relation $\Delta_{P_V} \cup R_Z$ ist dann eine Äquivalenzrelation auf P_V . Eine andere Alternative zeigt die Abbildung 2.2. Eine Relation zu einer Äquivalenzrelation zu „machen“ geht auch ganz allgemein:

Definition 2.8 (Erzeugte Äquivalenzrelation). Sei R eine Relation in X , dann existiert eine eindeutig bestimmte minimale Äquivalenzrelation $R^\sim \subseteq X \times X$, die R umfasst, also $R \subseteq R^\sim$ und für alle anderen Äquivalenzrelationen $S \subseteq X \times X$ mit $R \subseteq S$ gilt $R^\sim \subseteq S$. R^\sim bezeichnen wir dann als die von R erzeugte Äquivalenzrelation.

Wir erhalten R^\sim z.B., indem wir die Menge aller Äquivalenzrelationen

$$\mathcal{R} := \{A \subseteq X \times X \mid R \subseteq A \text{ und } A \text{ ist Äquivalenzrelation}\},$$

die R umfassen, schneiden. Dann behaupten wir $R^\sim = \bigcap \mathcal{R}$. Zum Beweis hilft die Überlegung, dass wenn R^\sim existiert, dann muss es Element von \mathcal{R} sein. Zudem ist auch $X \times X$ eine Äquivalenzrelation, die R umfasst, und folglich ist \mathcal{R} nicht leer. Zudem ist der Schnitt beliebiger Äquivalenzrelationen auf einer Menge selbst wieder eine Äquivalenzrelation auf dieser Menge. Später werden wir noch eine konstruktivere Definition sehen, wenn wir die reflexive und transitive Hülle einer Relation kennen. Die von einer Relation erzeugte Äquivalenzrelation werden wir später benötigen, um Räume aneinanderzuheften, und sie wird wichtig werden, wenn wir uns mit Wegen in Räumen und mit der Modellierung von Detailbibliotheken befassen.

Bei den Zahlenmengen hatten wir nun ganz freizügig eine Gleichheit von Elementen definiert und es wurde versprochen zu zeigen, wie dies bewerkstelligt wird. Zumindest war jede so definierte „Gleichheit“ eine Äquivalenzrelation, wie sich leicht nachprüfen lässt. Nun wollen wir sehen, wie umgekehrt aus einer Äquivalenzrelation eine Gleichheit wird:

Definition 2.9 (Äquivalenzklasse, Quotientenmenge). Sei M eine Menge und \sim Äquivalenzrelation auf M . Für jedes $x \in M$ heißt die Menge $[x] := \{y \in M \mid x \sim y\}$ Äquivalenzklasse von x bezüglich R . Wir bezeichnen x auch als einen Repräsentant für $[x]$, und $M/\sim := \{[x] \mid x \in M\}$, die Menge aller Äquivalenzklassen bezüglich \sim heißt Quotientenmenge von M nach \sim oder kurz „ M modulo \sim “.

Der Gebrauch der Wörter „Klasse“ und „Äquivalenzklasse“ ist in der Mathematik leider irreführend: „Klasse“ und „Äquivalenzklasse“ sind zwei grundverschiedene Dinge die nicht verwechselt werden dürfen.

Nun sind zwei Äquivalenzklassen $[a]$ und $[b]$ genau dann gleich, wenn a äquivalent zu b ist. Wir haben also bei der Definition der Menge der ganzen Zahlen zunächst für Paare (n, m) von natürlichen Zahlen n und m eine Äquivalenzrelation $(a+n, a+m) \sim (b+n, b+m)$ für alle $a, b \in \mathbb{N}$ definiert. Was wir als „formale Differenz“ $[n - m]$ bezeichnet haben ist die Äquivalenzklasse $[(n, m)]$ von (n, m) bezüglich dieser Relation, und damit haben wir einfach $\mathbb{Z} := (\mathbb{N} \times \mathbb{N})/\sim$ gesetzt.

Ganz nebenbei stellen wir fest, dass $-[(n, m)] = [(m, n)]$ gilt, dass es also zwischen einem Minus und einer Vertauschung eine Beziehung gibt, mit der wir uns bereits die Erläuterung von DIME und Winged-Edge erleichtert haben.

Angenommen, unsere „Zimmerrelation“ R_Z sei wie oben zu einer Äquivalenzrelation R erweitert. Wenn wir nun einen Punkt p des Gebäudevolumens P_V nehmen und die Äquivalenzklasse $[p]$ betrachten, dann ist $[p]$ die Menge aller Punkte desjenigen Zimmers in dem auch p liegt, also der Raum, den dieses Zimmer einnimmt, also das Zimmer selbst. Insbesondere entspricht jedes Zimmer unseres Gebäudes einer Äquivalenzklasse in P_V/R . Alle anderen Äquivalenzklassen befinden sich irgendwo in den Bauelementen, aus denen sich das Gebäude zusammensetzt. Idealerweise ist nun R so definiert, dass sie in diesem Fall der Konstruktion des Gebäudes entspricht.

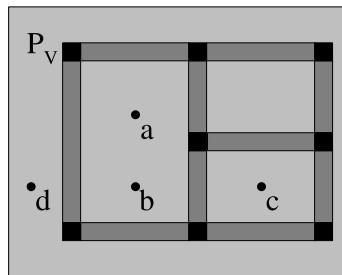


Abbildung 2.2: Die Fortsetzung der Zimmerrelation $R_Z \subseteq P_V \times P_V$ zu einer Äquivalenzrelation R — der Zimmer- und Bauteilerrelation. Die Äquivalenzklassen sind die „Zimmer“, die Wände (dunkelgrau) inklusive der Wandoberflächen (schwarz) und Wandverbindungen (schwarz). Hier gilt auch: $[a] = [b]$.

Umgekehrt ist durch ein Gebäude ein Gesamtvolumen in einzelne Bereiche untergliedert:

Definition 2.10 (Zerlegung). Sei M eine Menge und \mathcal{M} eine Menge von Teilmengen von M (Also $\mathcal{M} \subseteq \mathcal{P}(M)$). Wenn gilt:

$$\emptyset \notin \mathcal{M}$$

$$M = \bigcup \mathcal{M}$$

und für alle $A, B \in \mathcal{M}$:

$$A \cap B = \emptyset \text{ oder } A = B$$

dann heißt \mathcal{M} Zerlegung von M .

Zwischen Zerlegung und Äquivalenzrelation gibt es nun einen engen Zusammenhang:

Satz 2.1 (Äquivalenzrelation und Zerlegung). *Sei M eine Menge. Jede Zerlegung von M ist Quotientenmenge von M bezüglich einer Äquivalenzrelation und jede Quotientenmenge M/R ist eine Zerlegung von M .*

Der Beweis kann in entsprechenden Lehrbüchern nachgeschlagen werden und ist nicht schwierig. Insbesondere ergibt sich daraus, dass zwei verschiedene Äquivalenzklassen kein gemeinsames Element haben.

Nun kann durch geeignete Wahl einer Äquivalenzrelation R auf P_V , der „Gesamtraum“ P_V in eine Menge P_V/R zerlegt werden, die genau den Räumen, Raumgrenzen, deren Verbindungen und Knotenpunkte des Gebäudes entspricht. Dies ist jedoch keine mathematische Aussage sondern eine physikalische Behauptung.

2.1.5 Das Relationenprodukt

Wenn wir etwa sagen „Ich kenne Herrn A.“ dann entspricht dies der Aussage, dass das Paar (Ich, Herr A) in der Relation „kennt“ \subseteq Mensch \times Mensch liegt. Sagen wir hingegen „Herr A kann Entwerfen.“ heißt das wiederum, das Paar (Herr A, Entwerfen) ist Element der Relation „kann“ \subseteq Mensch \times Fertigkeit. Dann gilt ebenfalls: „Ich kenne jemanden, der kann Entwerfen.“ Dies ist ebenfalls ein Element (Ich, Entwerfen) einer Relation „kennt jemand der ... kann“ \subseteq Mensch \times Fertigkeit. Das „jemand, der“ ist also eine Verknüpfung, mit der aus den Relationen „kennt“ und „kann“ eine neue Relation hervorgeht, ganz analog wie mit „ \cdot “ aus den Zahlen 2 und 5 eine neue Zahl 10 (gleich $2 \cdot 5$) hervorgeht:

Definition 2.11 (Relationenprodukt). *Seien $R \subseteq X \times Y$ und $S \subseteq Y \times Z$ Relationen. Dann ist die Relation*

$$R \circ S := \{(a, c) \mid \text{Es gibt ein } j \in Y \text{ mit } a R j \text{ und } j S c\} \subseteq X \times Z$$

das Relationenprodukt von R und S . (Das j entspricht dem obigem „jemand“.)

Das Relationenprodukt hat folgende Eigenschaften: Sind $R \subseteq X \times Y$, $S \subseteq Y \times Z$ und $T \subseteq Z \times M$ Relationen, dann gilt:

$$(R \circ S) \circ T = R \circ (S \circ T) \quad \text{„}\circ\text{“ ist assoziativ}$$

$$\Delta_X \circ R = R = R \circ \Delta_Y \quad \text{Diagonalen sind links- und rechtsneutral}$$

Derartige „Übergänge“ (hier die Relationen) von einem mathematischen Objekt X zu einem weiteren mathematischen Objekt Y mit einer entsprechenden Verknüpfung \circ , die obige Eigenschaften hat, heißen *Morphismen*.

Beispiel 2.4 (Relationenprodukt).

Seien

$$\begin{aligned} \text{istVaterVon} &\subseteq \text{Mensch} \times \text{Mensch} \\ \text{istMutterVon} &\subseteq \text{Mensch} \times \text{Mensch} \\ \text{istElternteilVon} &= \text{istVaterVon} \cup \text{istMutterVon} \end{aligned}$$

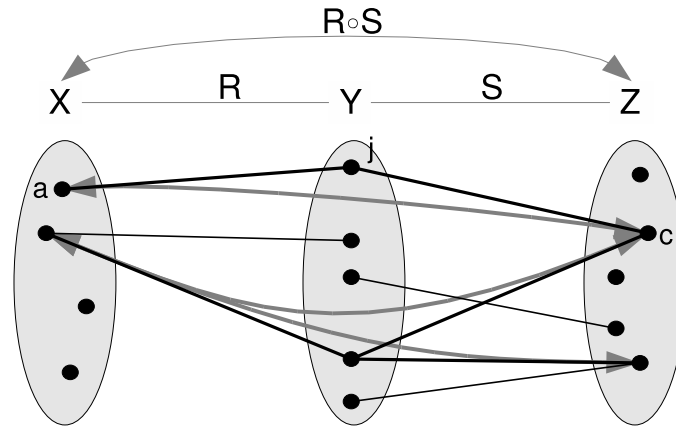


Abbildung 2.3: Relationenprodukt: Für jede „Verbindung“ $a R j S c$ von X über Y nach Z nehmen wir die „Abkürzung“ $a (R \circ S) c$ von X nach Z .

Relationen. Dann gilt:

$$\begin{aligned} \text{istOpaVon} &= \text{istVaterVon} \circ \text{istElternteilVon} \\ \text{istOmaVon} &= \text{istMutterVon} \circ \text{istElternteilVon} \\ \text{istGroßElternteilVon} &= \text{istElternteilVon} \circ \text{istElternteilVon}. \end{aligned}$$

Wir können nun auch Potenzen von Relationen auf einer Menge definieren: Ist $R \subseteq A \times A$ eine Relation, dann ist für alle natürlichen Zahlen n :

$$\begin{aligned} R^0 &:= \Delta_A \\ R^{n+1} &:= R^n \circ R \\ R^+ &:= \bigcup_{n \in \mathbb{N}} R^{n+1} \\ R^* &:= \bigcup_{n \in \mathbb{N}} R^n \end{aligned}$$

Zudem gilt: Eine Relation R ist genau dann transitiv, wenn $R^2 \subseteq R$ gilt, denn ist R transitiv, und gelte $(a, b) \in R^2$ (also $a R^2 b$), für beliebige a und b . Dann gibt es j mit $a R j R b$ und, da R transitiv ist, gilt $a R b$, also $(a, b) \in R$ und somit $R^2 \subseteq R$. Gelte umgekehrt diese Eigenschaft und $a R b R c$ für beliebige a, b, c . Dann gilt $a R^2 c$ und wegen $R^2 \subseteq R$ auch $a R c$ und somit ist R transitiv.

Wir bezeichnen R^+ als die transitive Hülle von R und R^* die transitive und reflexive Hülle von R . „Hülle“ einer Menge bedeutet hier, „die kleinste Menge, die diese Menge umfasst und zudem gewisse Eigenschaften hat“. Im Gegensatz zum Sprachgebrauch im Bauwesen ist ein Objekt stets Teil seiner eigenen Hülle. Die „Hülle“ eines Raumes in Sinne der Architektur werden wir später als „Rand“ des Raumes bezeichnen.

Diese Hüllen (transitiv bzw. transitiv und reflexiv) spielen eine zentrale Rolle für die vorliegende Arbeit und verdienen daher besondere Aufmerksamkeit:

Satz 2.2 (Eigenschaften der Hüllen $*$ und $+$). *Seien $R, S \subseteq M \times M$ Relationen in M . Dann gilt:*

1. $R^n \circ R^m = R^{n+m}$.
2. $R^+ = R \circ R^* = R^* \circ R$ und $R^* = R^+ \cup \Delta_M$.
3. R^+ ist transitiv und R^* ist transitiv und reflexiv.
4. Gilt $R \subseteq S$, dann gelten $R^+ \subseteq S^+$ und $R^* \subseteq S^*$.
5. $R \subseteq R^+ \subseteq R^*$.
6. $(R^+)^+ = R^+$ und $(R^*)^* = R^*$.
7. R^+ ist die minimale transitive Relation, die R umfasst. D.h. jede transitive Relation T , die R umfasst, umfasst auch R^+ .
8. R^* ist die minimale transitive und reflexive Relation, die R umfasst. D.h. jede transitive und reflexive Relation T , die R umfasst, umfasst auch R^* .

Wegen der Bedeutung dieser Eigenschaften für die vorliegende Arbeit sollen sie hier bewiesen werden. Der Beweis ist einfach, wird hier aber aus didaktischen Gründen sehr ausführlich geführt.

Beweis. Wir setzen voraus, dass M eine Menge ist und R, S und T Relationen auf M sind.

Zuerst beweisen wir die Assoziativität $(R \circ S) \circ T = R \circ (S \circ T)$ für $R \subseteq A \times B$, $S \subseteq B \times C$ und $T \subseteq C \times D$:

Sei $(a, d) \in (R \circ S) \circ T$.

Dann existiert ein $c \in C$ mit $(a, c) \in R \circ S$ und $(c, d) \in T$.

Dann existiert ein $b \in B$ mit $(a, b) \in R$ und $(b, c) \in S$.

Dann gilt $(b, d) \in S \circ T$.

Also gilt $(a, d) \in R \circ (S \circ T)$.

folglich $(R \circ S) \circ T \subseteq R \circ (S \circ T)$. Die andere Richtung der Teilmengenbeziehung zeigt man entsprechend.

1. $(R^n \circ R^m = R^{n+m})$:

Wir benutzen nun die eben gezeigte Assoziativität für einen Beweis durch vollständige Induktion über m :

Sei $m = 0$. (Induktionsanfang)

Dann gilt: $R^{n+m} = R^{n+0} = R^n$.

Zu zeigen ist also: $R^n = R^n \circ R^0$

„ \subseteq “: Zuerst zeigen wir $R^n \circ R^0 \subseteq R^n$.

Sei $(a, b) \in R^n \circ R^0$.

Dann gibt es ein $j \in M$ mit $(a, j) \in R^n$ und $(j, b) \in R^0 = \Delta_M$.

Dann gilt $j = b$ also $(a, b) = (a, j) \in R^n$.

„ \supseteq “: Nun zeigen wir $R^n \subseteq R^n \circ R^0$.

Sei $(a, b) \in R^n$.

Dann ist $(b, b) \in R^0$ und somit $(a, b) \in R^n \circ R^0$.

Also gilt $R^{n+0} = R^n \circ R^0$.

Sei m beliebig, fest und gelte $R^{n+m} = R^n \circ R^m$ (Induktionsschritt).

Dann gilt $R^{n+(m+1)} = R^{(n+m)+1}$

(Definition) $= R^{n+m} \circ R$

(Induktionsbehauptung) $= (R^n \circ R^m) \circ R$

(Assoziativität) $= R^n \circ (R^m \circ R) = R^n \circ R^{m+1}$.

Also gilt die Behauptung auch für $m + 1$.

Wir erinnern uns, dass damit nach dem Induktionsprinzip jede natürliche Zahl m die Eigenschaft $R^n \circ R^m = R^{n+m}$ hat.

2. Die zweite Eigenschaft, also $R^+ = R^+ \cup \Delta_M$, ergibt sich direkt aus der Definition von R^* .

Wir zeigen hier also nur noch $R^+ = R \circ R^* = R^* \circ R$.

Dazu beweisen wir $R^+ \subseteq R \circ R^* \subseteq R^* \circ R \subseteq R^+$:

Sei $(a, b) \in R^+$, dann existiert ein n mit $(a, b) \in R^{n+1}$.

Also gibt es ein j mit $(a, j) \in R$ und $(j, b) \in R^n$.

Dann ist $(j, b) \in R^*$ und somit $(a, b) \in R \circ R^*$.

Also gilt: $R^+ \subseteq R \circ R^*$.

Sei nun $(a, b) \in R \circ R^*$ beliebig.

Dann existiert j mit $(a, j) \in R$ und $(j, b) \in R^*$.

Dann ist $(j, b) \in R^n$ für irgend ein n von dem wir wissen, dass es existiert.

Dann ist $(a, b) \in R^1 \circ R^n = R^{1+n} = R^{n+1} = R^n \circ R^1$.

Dann existiert i mit $(a, i) \in R^n$ und $(i, b) \in R$.

Dann ist $(a, i) \in R^*$ und somit $(a, b) \in R^* \circ R$.

Also gilt: $R \circ R^* \subseteq R^* \circ R$.

Sei nun $(a, b) \in R^* \circ R$.

Dann existiert j mit $(a, j) \in R^*$ und $(j, b) \in R$.

Dann ist $(a, j) \in R^n$ für ein n .

Dann ist $(a, b) \in R^n \circ R^1 = R^{n+1}$.

Dann ist $(a, b) \in R^+$.

Also gilt: $R^* \circ R \subseteq R^+$.

Dann sind jedoch die drei Mengen gleich.

3. Seien (a, b) und (b, c) Elemente von R^* (bzw. R^+).

Dann ist $(a, b) \in R^n$ und $(b, c) \in R^m$ für nat. Zahlen n und m . (Bei R^+ sind n und m größer 0.)

Dann ist $(a, c) \in R^n \circ R^m = R^{n+m}$

Somit gilt: $(a, c) \in R^*$ (bzw. R^+). (transitiv)

Wegen $\Delta_M \subseteq R^*$ ist R^* reflexiv.

4. Wir setzen voraus, dass R Teilrelation von S ist: $R \subseteq S$

Wir beweisen zuerst $R^n \subseteq S^n$ für alle $n \in \mathbb{N}$ per vollständiger Induktion über n :

Sei $n = 0$: Ist $(a, b) \in R^0 = R^0$, dann ist $a = b$ und somit $(a, b) \in S^0 = S^0$.

Sei nun n beliebig und gelte $R^n \subseteq S^n$:

Sei $(a, b) \in R^{n+1}$ dann existiert $j \in M$ mit $(a, j) \in R^n$ und $(j, b) \in R$.

Dann gilt $(a, j) \in S^n$ laut Induktionsbehauptung und $(j, b) \in S$ laut Annahme $R \subseteq S$.

Also gilt $(a, b) \in S^n \circ S = S^{n+1}$.

Somit ist $R^{n+1} \subseteq S^{n+1}$.

Also haben wir $R^n \subseteq S^n$ per vollständige Induktion bewiesen.

Sei nun $(a, b) \in R^+$ (bzw. $(a, b) \in R^*$).
 Dann ist auch $(a, b) \in R^n$ für $n \in \mathbb{N} \setminus \{0\}$ (bzw. $n \in \mathbb{N}$).
 Dann ist $(a, b) \in S^n$, wie wir oben gesehen haben.
 Dann ist $(a, b) \in S^+$ (bzw. $(a, b) \in S^*$).
 Folglich gilt die Behauptung.

5. R^+ ist die Vereinigung der R^{n+1} für $n \in \mathbb{N}$, umfasst also auch $R^1 := R$.
 $R^* = R^+ \cup \Delta_M$, also $R^+ \subseteq R^*$.
6. Es wurde eben gezeigt, dass auch $R^+ \subseteq (R^+)^+$ und $R^* \subseteq (R^*)^*$.
 Also bleibt nur noch zu zeigen: $R^+ \supseteq (R^+)^+$ und $R^* \supseteq (R^*)^*$.
 Sei nun $(a, b) \in (R^+)^+$.
 Dann existiert $n \in \mathbb{N} \setminus \{0\}$ mit $(a, b) \in (R^+)^n$.
 Zu zeigen $(R^+)^n \subseteq R^+$ durch Induktion über n (mit $n > 0$):
 Sei $n = 1$, dann gilt $(R^+)^1 = R^+$ also $(R^+)^1 \subseteq R^+$. Gelte $(R^+)^n \subseteq R^+$
 für ein $n \in \mathbb{N}$.
 Sei $(a, b) \in (R^+)^{n+1} = (R^+)^n \circ R^+$.
 Dann existiert $j \in M$ mit $(a, j) \in (R^+)^n$ und $(j, b) \in R^+$.
 Dann gilt $(a, j) \in R^+$ laut Induktionsbehauptung.
 Dann gilt $(a, b) \in R^+$, denn R^+ ist transitiv.
 Also gilt $(R^+)^n \subseteq R^+$ für alle nat. Zahlen $n > 0$.
 $(R^*)^* \subseteq R^*$ zeigt man ganz analog.
7. Hier haben wir zu zeigen, dass R^+ Teilmenge jeder transitiven Relation T ist, die R umfasst.
 Sei also $T \subseteq M \times M$ eine transitive Relation mit $R \subseteq T$.
 Sei $(a, b) \in R^+$, dann ist $(a, b) \in R^n$ für ein $n \in \mathbb{N} \setminus \{0\}$.
 Auch hier zeigen wir $R^n \subseteq T$ für beliebige nat. Zahl $n > 0$ durch Induktion.
 Sei $n = 1$:
 Ist $(a, b) \in R^1$, dann ist $(a, b) \in T$ laut Voraussetzung $R \subseteq T$.
 Gelte für ein beliebiges n : $R^n \subseteq T$.
 Ist $(a, b) \in R^{n+1}$, dann ist $(a, j) \in R^n$ und $(j, b) \in R$ für ein $j \in M$.
 Dann ist $(a, j) \in T$ wegen der Induktionsbehauptung und $(j, b) \in T$ laut Voraussetzung.
 Dann ist ebenfalls $(a, b) \in T$ denn T ist transitiv.
8. Dies zeigt man ganz analog, indem man den Induktionsanfang bei $n = 0$ macht.

□

Mit den Abfragesprachen relationaler Datenbanksysteme ist es oft nicht möglich, die transitive Hülle einer Relation zu berechnen [Mai83, S.529]. Auf dieses Problem wird im Abschnitt über relationale Datenbanken und besonders im Teil über die topologischen Datenbanken \mathcal{DTop} genauer eingegangen.

Nun wollen wir hier noch die konstruktive Variante der von einer Relation erzeugten Äquivalenzrelation R^{\sim} zeigen:

Lemma 2.3 (Konstruktion der erzeugten Äquivalenzrelation). *Sei $R \subseteq X \times X$ eine Relation in X , dann ist $(R \cup R^T)^*$ die von R erzeugte Äquivalenzrelation. Es ist also $R^{\sim} = (R \cup R^T)^*$.*

Wir erinnern uns, dass R^T die transponierte Relation von R ist, indem einfach die linke mit der rechten Seite vertauscht wird. Es gilt also $a R^T b$ genau dann, wenn $b R a$ gilt. R^\sim ist also die transitive und reflexive Hülle von $R \cup R^T$.

Beweis. Es ist zu zeigen, dass $(R \cup R^T)^*$ eine Äquivalenzrelation ist und dass für jede Äquivalenzrelation S mit $R \subseteq S$ auch $(R \cup R^T)^* \subseteq S$ gilt. Die Transitivität und Reflexivität von $(R \cup R^T)^*$ ist klar. Zu zeigen ist also noch die Symmetrie: Gelte $a (R \cup R^T)^* b$, dann gilt $a (R \cup R^T)^n b$ für eine natürliche Zahl n . Dann zeigen wir $b (R \cup R^T)^n a$ durch Induktion über n :

Ist $n = 0$, dann gilt $a = b$ und somit $b (R \cup R^T)^n a$.

Gelte die Induktionsbehauptung, dass aus $a (R \cup R^T)^n b$ auch $b (R \cup R^T)^n a$ folgt für eine natürliche Zahl n und zudem $a (R \cup R^T)^{n+1} b$. Dann existiert ein $j \in X$, so dass $a (R \cup R^T)^n j (R \cup R^T) b$ gilt. Dann gilt $j (R \cup R^T)^n a$ wegen der Induktionsbehauptung und es gilt $j R b$ oder $j R^T b$. Dann gilt $b R^T j$ oder $b R j$ also $b (R \cup R^T) j$. Es gilt also $b (R \cup R^T) j (R \cup R^T)^n a$ und somit $b (R \cup R^T)^{n+1} a$. Also ist $(R \cup R^T)^*$ symmetrisch.

Es bleibt noch zu zeigen, dass $(R \cup R^T)^*$ minimal ist: Sei $S \subseteq X \times X$ eine Äquivalenzrelation mit $R \subseteq S$. Dann ist $R^T \subseteq S^T$. Da jedoch S als Äquivalenzrelation symmetrisch ist, gilt $S^T = S$ und somit $R^T \subseteq S$. Folglich gilt auch $R \cup R^T \subseteq S$. Damit gilt $(R \cup R^T)^* \subseteq S^*$. Da S jedoch zudem transitiv und reflexiv ist haben wir $S^* = S$ und somit $(R \cup R^T)^* \subseteq S$. \square

Nachdem wir nun mit dem Begriff der Relation vertraut sind, wollen wir uns nun mit dem wichtigsten Sonderfall der Relationen beschäftigen.

2.1.6 Abbildungen

Wenn wir von „der Nutzung“ eines Raums sprechen, dann haben wir einerseits die Vorstellung von einer Menge von Räumen, einer Menge von Nutzungen und jedem Raum eine zugeordnete Nutzung. Wir haben also auch hier eine Relation „Nutzung“ zwischen Räumen und Nutzungen, jedoch sind dabei oft folgende Einschränkungen sinnvoll:

1. Jedem Raum ist eine Nutzung zugeordnet.
2. Keinem Raum sind mehrere Nutzungen zugeordnet.

Relationen mit diesen Eigenschaften heißen Abbildungen:

Definition 2.12 (Abbildung). *Seien X und Y Mengen, und $f = ((X, Y), \mapsto)$ eine Relation zwischen X und Y . Dann heißt f (totale) Abbildung von X nach Y , wenn sie folgende Eigenschaften hat:*

(F1) f ist linkstotal, also für jedes $x \in X$ gibt es ein $y \in Y$ mit $x f y$.

(F2) f ist rechtseindeutig, also für alle $x f y$ und $x f z$ gilt $y = z$.

Statt $f = ((X, Y), \mapsto)$ schreiben wir $f : X \rightarrow Y$. Wir bezeichnen dann den Vorbereich X auch als Definitionsbereich von f , der Nachbereich Y heißt auch Wertebereich von f und „ \mapsto “ heißt weiterhin Graph von f . Statt $x f y$ schreiben wir $f(x) = y$. Die Menge aller Abbildungen von X nach Y bezeichnen wir mit $\text{Abb}(X, Y)$ oder auch Y^X .

Diese Definition umfasst auch mehrstellige Abbildungen $f : X_1 \times \dots \times X_n \rightarrow Y$. Dies sind Abbildungen, deren Definitionsbereich X das kartesische Produkt der X_1 bis X_n ist. Für ein n -Tupel (x_1, \dots, x_n) schreiben wir einfach $f(x_1, \dots, x_n)$ anstelle von $f((x_1, \dots, x_n))$.

Wird die Zuordnung „ \mapsto “ durch eine Regel festgelegt, dann schreiben wir auch

$$f : X \rightarrow Y, x \mapsto f(x) := \text{Regel für die Auswahl des } y$$

Eine Abbildung $f : X \rightarrow Y$ heißt *surjektiv*, wenn für jedes $y \in Y$ mindestens ein $x \in X$ mit $f(x) = y$ existiert, *injektiv*, wenn für jedes $y \in Y$ höchstens ein solches $x \in X$ mit $f(x) = y$ gibt und *bijektiv*, wenn f surjektiv und injektiv ist, wenn also für jedes $y \in Y$ genau ein $x \in X$ mit $f(x) = y$ existiert.

Die Schreibweise Y^X erklärt sich durch die Tatsache, dass bei endlichen Mengen X und Y gilt: $|X^Y| = |X|^{|Y|}$.

Dass wir Abbildungen auch als „totale“ Abbildungen bezeichnen liegt daran, dass wir auch noch den allgemeineren Begriff der partiellen Abbildung benötigen werden. Wir wollen diesen hier gleich einführen:

Definition 2.13 (partielle Abbildung). *Seien X und Y Mengen, und $f = ((X, Y), \mapsto)$ eine Relation zwischen X und Y . Dann heißt f partielle Abbildung von X nach Y , wenn sie rechtseindeutig ist. Für partielle Abbildungen $f = ((X, Y), \mapsto)$ schreiben wir $f : \subseteq X \rightarrow Y$. Die größte Teilmenge $X_f \subseteq X$, mit der $((X_f, Y), \mapsto)$ eine totale Abbildung ist, heißt Definitionsbereich von f .*

Eine totale Abbildung ist nach dieser Definition also ein Sonderfall einer partiellen Abbildung. Bei einer partiellen Abbildung kann es ein $x \in X$ geben, dem gar kein Wert zugeordnet ist. Dann schreiben wir $f(x) = \uparrow$ und sagen „ f ist in x nicht definiert“. Wichtig ist dabei, \uparrow nicht falsch als einen sonderbaren Wert von f an der Stelle x zu interpretieren, sondern als symbolische Schreibweise dafür, dass f an der Stelle x überhaupt keinen Wert hat. Bei den partiellen Abbildungen $f : \subseteq X \rightarrow Y$ ist also der Vorbereich X vom Definitionsbereich X_f zu unterscheiden.

Da wir (partielle) Abbildungen als Tupel definiert haben, ist sofort klar, dass zwei Abbildungen $f : A \rightarrow X$ und $g : B \rightarrow Y$ genau dann gleich sind, wenn $A = B$, $X = Y$ und $\text{Graph}(f) = \text{Graph}(g)$ gelten.

Beispiele 2.5 (Abbildungen). *Hier sollen einige Abbildungen vorgestellt werden. Es sind vor allem diejenigen Abbildungen, die wir im Verlauf dieser Arbeit noch benötigen werden.*

(1) *Sei X eine beliebige Menge, dann ist*

$$((X, X), \Delta_X) = \text{id}_X : X \rightarrow X, x \mapsto \text{id}_X(x) := x$$

die identische Abbildung in X .

(2) *Für eine Teilmenge A von B ist*

$$i : A \rightarrow B, a \mapsto i(a) := a$$

die Inklusion von A in B . Es ist dann auch üblich den Abbildungspfeil mit einem kleinen Häkchen zu dekorieren und

$$i : A \hookrightarrow B$$

zu schreiben. Dieses Häkchen erinnert an das Zeichen „ \subset “.

- (3) In obigem Beispiel kann A auch die leere Menge sein. Dann bezeichnen wir $i : \emptyset \hookrightarrow B$ als leere Abbildung. Dies ist die einzige Abbildung in B^{\emptyset} . Man beachte, dass für ein endliches B gilt: $|B^{\emptyset}| = |B|^{|\emptyset|} = |B|^0 = 1$. Insbesondere ist auch $|\emptyset^{\emptyset}| = 1$.

- (4) Für die Mengen X und Y und ein beliebiges $b \in Y$ ist die konstante Abbildung

$$c_b : X \rightarrow Y, x \mapsto c_b(x) := b.$$

- (5) Für zwei Abbildungen $f : X \rightarrow Y$ und $g : Y \rightarrow Z$ ist

$$g \circ f : X \rightarrow Z, x \mapsto g \circ f(x) := g(f(x))$$

die Verkettung oder Komposition der Abbildungen f und g . Diese ist nichts anderes als das Relationenprodukt der Relationen $f = ((X, Y), F)$ und $g = ((Y, Z), G)$ mit vertauschten Operanden: Bezeichnen wir hier das im vorhergehenden Abschnitt definierte Relationenprodukt mit \bowtie , um es nicht mit der Abbildungskomposition \circ hier zu verwechseln, dann ist $g \circ f = f \bowtie g$. Dieser „Dreher“ in den Schreibweisen ist leider etwas verwirrend und manche Autoren gleichen daher die Schreibweise für Relationenprodukt und Abbildungskomposition einander an.

- (6) Für zwei Abbildungen $f : A \rightarrow X$ und $g : B \rightarrow Y$ ist das direkte Produkt von f und g

$$f \times g : A \times B \rightarrow X \times Y, (a, b) \mapsto (f \times g)(a, b) := (f(a), g(b)).$$

- (7) Für zwei Abbildungen $f : A \rightarrow X$ und $g : A \rightarrow Y$ mit gleichem Definitionsbereich A ist die Produktabbildung

$$(f, g) : A \rightarrow X \times Y, a \mapsto (f, g)(a) := (f(a), g(a)).$$

- (8) Für Mengen X_1, \dots, X_n und eine natürliche Zahl i mit $1 \leq i \leq n$ ist die i -te Projektion definiert durch

$$\pi_i : X_1 \times \dots \times X_i \times \dots \times X_n \rightarrow X_i, (x_1, \dots, x_i, \dots, x_n) \mapsto x_i.$$

- (9) Jedes n -Tupel $t = (a_1, \dots, a_n) \subseteq D_1 \times \dots \times D_n$ ordnet einer Zahl $i \in \{1, \dots, n\}$ ein a_i aus $D_1 \cup \dots \cup D_n$ zu. Also ist dieses Tupel eine Abbildung

$$t : \{1, \dots, n\} \rightarrow D_1 \cup \dots \cup D_n, i \mapsto t(i).$$

Im relationalen Datenmodell entspricht dies einem Datensatz. Nur verwendet man dort statt der Zahlen $1, \dots, n$ aussagekräftigere Symbole wie „id“, „name“, ..., „telefonnummer“ und man schreibt dort üblicherweise $t.i$ statt $t(i)$.

- (10) Bei einer Abbildung $f : X \rightarrow Y$ und einem $x \in X$ suggeriert der Ausdruck $f(x)$ die Vorstellung von einer Variablen x , die die Menge X durchläuft und dabei in ein festes f eingesetzt wird. Wir können jedoch diese Rollen auch tauschen, ein $x \in X$ festhalten und die Abbildungen $f \in Y^X$ als

veränderlich betrachten. Dann erhalten wir für jedes $x \in X$ die sogenannten Einsetzungsabbildung an der Stelle x .

$$e_x : \text{Abb}(X, Y) \rightarrow Y, f \mapsto e_x(f) := f(x).$$

Die i -te Projektion π_i aus (8) ist eine solche Einsetzungsabbildung e_i für Abbildungen aus (9), bei der ein festes i die Tupel „durchläuft“.

- (11) In der Programmiersprache LISP [Ste90] gibt es die obige Einsetzungsabbildung als Funktion

$$\text{funcall} : \text{function} \times T \rightarrow T, (f, x) \mapsto \text{funcall}(f, x) := f(x)$$

Dabei ist T der universelle Datentyp in LISP und `function` der Datentyp der Abbildungen $\text{Abb}(T, T)$.

- (12) Ist M Menge und \sim eine Äquivalenzrelation in M , dann ist die natürliche Abbildung (auch: „natürliche Projektion“, „Quotientenabbildung“ oder „Restklassenabbildung“)

$$\pi_{\sim} : M \rightarrow M/\sim, x \mapsto [x].$$

- (13) Ist $f : X \rightarrow Y$ Abbildung und $A \subseteq X$, dann ist $f|_A$ die Einschränkung von f auf A :

$$f|_A : A \rightarrow Y, a \mapsto f(a).$$

Umgekehrt heißt f dann Fortsetzung von $f|_A$. Es gilt: $f|_A = f \circ i$ Für die Inklusion $i : A \hookrightarrow X$ aus (2).

- (14) Für jede Abbildung $f : X \rightarrow Y$ haben wir auch die Abbildungen

$$f : \mathcal{P}(X) \rightarrow \mathcal{P}(Y), A \mapsto f[A] := \{f(x) \mid x \in A\}$$

und sagen $f[A]$ ist das Bild von A unter f , und wir haben das Urbild

$$f^{-1} : \mathcal{P}(Y) \rightarrow \mathcal{P}(X), B \mapsto f^{-1}[B] := \{x \mid f(x) \in B\}.$$

Wir wollen im Folgenden die Schreibweise $f[A]$ mit eckigen Klammern für das Bild einer Menge und $f(a)$ mit runden Klammern für das Bild eines Punktes verwenden. Das Urbild ist besonders wichtig in der Topologie. $f^{-1}[B]$ ist die Menge aller Punkte $x \in X$, deren Bild $f(x)$ in B liegt.

Zwischen Bild und Urbild bestehen nun folgende Beziehungen:

$$f[f^{-1}[B]] \subseteq B \quad \text{und} \quad A \subseteq f^{-1}[f[A]].$$

Sei nämlich $b \in f[f^{-1}[B]]$, dann gibt es ein $a \in f^{-1}[B]$ mit $b = f(a)$. Dann ist aber $f(a) \in B$, wegen der Definition von f^{-1} und somit $b \in B$. Sei, für die andere Inklusion, $a \in A$, dann ist $f(a) \in f[A]$ und somit $a \in f^{-1}[f[A]]$.

- (15) Das obige Beispiel von Bild und Urbild lässt sich auch auf beliebige zweistellige Relationen verallgemeinern. Ist $R \subseteq X \times Y$ eine Relation, dann haben wir die Abbildung

$$\text{im}_R : \mathcal{P}(X) \rightarrow \mathcal{P}(Y), A \mapsto \text{im}_R[A] := \{y \in Y \mid \exists x \in A : x R y\}.$$

Die Schreibweise „ $\exists x \in A$: Eigenschaft von x “ ist eine Kurzschreibweise für „Es gibt zumindest ein $x \in A$ mit der angegebenen Eigenschaft“.

(16) Für $f : A \rightarrow Y$ und $g : B \rightarrow Y$ mit $f|_{A \cap B} = g|_{A \cap B}$, ist

$$f \cup g : A \cup B \rightarrow Y, a \mapsto (f \cup g)(a) := \begin{cases} f(a) & : a \in A \\ g(a) & : a \notin A. \end{cases}$$

die gemeinsame Fortsetzung oder Verklebung von f und g . Sind f und g Datensätze einer relationalen Datenbank, dann heißt $f \cup g$ auch natürlicher Verbund von f und g .

(17) Eine Abbildung der Form $\Delta : M \times M \rightarrow M$ heißt auch Verknüpfung in M . Statt $\Delta(a, b)$ schreiben wir meist $a \Delta b$, es sei denn, das was wir schreiben ist ein LISP-Programm. Zum Beispiel ist die Addition der ganzen Zahlen eine Abbildung $+$: $\mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{Z}$ und es ist z.B. $+(3, 7) = 3 + 7 = 10$.

(18) Haben wir eine derartige Verknüpfung $\Delta : M \times M \rightarrow M$ und zwei Abbildungen $f : X \rightarrow M$ und $g : X \rightarrow M$, dann überträgt sich diese Verknüpfung auch auf f und g in der Form

$$f \Delta g : X \rightarrow M, x \mapsto f \Delta g(x) := f(x) \Delta g(x)$$

als Verknüpfung in M^X . Es ist $f \Delta g = \Delta \circ (f, g)$.

Wir sehen, dass es oft alternative Bezeichnungen für Abbildungen gibt, je nachdem, welchem Zweck die Abbildung dient. Wenn etwa eine Abbildung $x : I \rightarrow M$ nur bezweckt, mit Elementen aus I auf Elemente aus M zu verweisen, dann schreiben wir auch x_i statt $x(i)$ und bezeichnen die Menge I als *Indexmenge*, die Abbildung x als *Familie* und schreiben sie in der Form $(x_i)_{i \in I}$ oder $\{x_i\}_{i \in I}$. Eine solche Familie heißt *Folge*, wenn I die Menge der natürlichen Zahlen ist.

Abbildungen sind mit der Komposition „ \circ “ Morphismen von Mengen, denn es gilt für $f : A \rightarrow B$, $g : B \rightarrow C$ und $h : C \rightarrow D$:

1. $f \circ (g \circ h) = (f \circ g) \circ h$
2. $\text{id}_A \circ f = f = f \circ \text{id}_B$

Ein weiteres Synonym für „Abbildung“ ist auch *Operator*, insbesondere, wenn die beteiligten Mengen aus einheitlich strukturierten Elementen, wie z.B. Instanzen von Datenstrukturen bestehen. Bezeichnen wir $f : X \rightarrow Y$ als Operator, dann ist die Zuordnung $x \mapsto f(x) = y$ eine *Operation*.

Dies sind also die allgemeinen mathematischen Grundlagen für diese Arbeit. Weitere Definitionen und Eigenschaften werden bei Bedarf eingeführt. Nun wollen wir noch auf die Grundbegriffe der Topologie und auf die relationalen Datenbanken eingehen. Diese sind für das Konzept der topologischen Datenbanken und deren Einsatz für Gebäudeinformationssysteme Voraussetzung.

2.2 Metrische Räume

Topologie kann man als Erweiterung der Mengenlehre auffassen. Wenn wir in der Mengenlehre etwa sagen $x \in M$, also „ x ist Element der Menge M “, dann gibt es in der Topologie die verstärkte Version $x \in \text{int } M$, das heißt „ x ist sogar

im Inneren der Menge M “ und die abgeschwächte Version $x \in \text{cl}M$, mit der Bedeutung „ x ist zumindest in der Nähe der Menge M “.

Diese Begriffe gingen aus der Untersuchung von Räumen hervor, für die ein Abstand definiert ist. Architektur findet im euklidischen Raum statt und dort kann stets für je zwei Punkte ein Abstand angegeben werden. Die Festlegung eines Abstands je zweier Punkte heißt *Metrik*:

Definition 2.14 (Metrik, metrischer Raum). *Sei M eine Menge. Eine Abbildung $d : M \times M \rightarrow \mathbb{R}$, die je zwei Punkten $a, b \in M$ eine reelle Zahl $d(a, b)$, „der Abstand von a nach b “ zuordnet, heißt Metrik für M , wenn d folgende Eigenschaften für alle $a, b, c \in M$ erfüllt:*

(M1) (Nichtnegativität) $d(a, b) > 0$ falls $a \neq b$, und $d(a, b) = 0$ falls $a = b$.

(M2) (Symmetrie) $d(a, b) = d(b, a)$

(M3) (Dreiecksungleichung) $d(a, b) + d(b, c) \geq d(a, c)$.

Das Paar (M, d) heißt dann metrischer Raum.

Umgangssprachlich heißt das: Der Abstand eines Punktes zu sich selbst ist 0 und zu jedem anderen Punkt stets größer als 0. Der Abstand zweier Punkte ist unabhängig davon, in welche Richtung gemessen wird, und bei abschnittweiser Messung ist die Summe der Abstände für die Teilabschnitte niemals kleiner als der gesamte Abstand. Die drei Axiome umschreiben also lediglich Eigenschaften, die man allgemein einem „Abstand“ zuschreiben würde.

Beispiel 2.6 (Metriken). *Die folgenden Beispiele für Metriken sollen unter anderem auch zeigen, dass eine Menge mehrere Metriken haben kann. Die Leserin kann sich davon überzeugen, dass es sich hierbei in der Tat um Metriken handelt, dass also jede der drei Eigenschaften (Axiome) der Definition erfüllt ist.*

- Für eine beliebige Menge M heißt

$$t : M \times M \rightarrow \mathbb{R}, (a, b) \mapsto \begin{cases} 1 & \text{falls } a \neq b \\ 0 & \text{falls } a = b \end{cases}$$

triviale Metrik. Offensichtlich erlaubt die Definition auch „Abstände“, die unserer Intuition zuwiderlaufen.

- Ist \mathbb{R}^n der n -dimensionale reelle Vektorraum, dann heißt

$$d : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$$

mit

$$d((a_1, \dots, a_n), (b_1, \dots, b_n)) := \sqrt{(a_1 - b_1)^2 + \dots + (a_n - b_n)^2}$$

die euklidische Metrik.

- Obiges Beispiel gilt auch für \mathbb{R} (also $n = 1$): Der Betrag der Differenz zweier reeller Zahlen ist auch ein Abstand:

$$d : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}, (x, y) \mapsto |x - y|.$$

- Alternativ können wir für \mathbb{R}^n die Metrik

$$d_{MN} : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$$

mit

$$d_{MN}((a_1, \dots, a_n), (b_1, \dots, b_n)) := |a_1 - b_1| + \dots + |a_n - b_n|,$$

definieren – die Manhattan-Metrik (auch: „Mannheim-Metrik“). Es wird hier abschnittsweise parallel zu den Koordinatenachsen gemessen, im Fall $n = 2$ z.B. zuerst waagrecht und dann senkrecht.

Nun wollen wir schrittweise die topologischen Grundbegriffe für metrische Räume aufbauen, bevor wir uns mit der Topologie selbst befassen. Dazu müssen wir zunächst festlegen, was wir als „Nähe“ verstehen: Wenn wir einen Punkt $p \in X$ und eine positive reelle Zahl ε nehmen, dann sagen wir, alle Punkte, deren Abstand zu p kleiner als ε ist, bilden zusammen eine *Umgebung des Punktes*:

Definition 2.15 (offene Vollkugel, (metrische) Umgebung). Sei (X, d) ein metrischer Raum. Für einen Punkt $p \in X$ und eine reelle positive Zahl $\varepsilon \in \mathbb{R}$ mit $\varepsilon > 0$ ist

$$B_\varepsilon(p) := \{x \in X \mid d(x, p) < \varepsilon\}$$

die offene ε -(Voll)Kugel um p in (X, d) . Eine Menge $U_p \subseteq X$ heißt Umgebung von p in (X, d) , wenn es eine offene ε -Kugel um p gibt, die vollständig in U_p liegt, wenn also $B_\varepsilon(p) \subseteq U_p$ gilt.

Wichtig ist, dass der Radius ε der Kugel größer als Null sein soll und dass die Kugel alle Punkte enthält, deren Abstand zu p *echt kleiner* als ε ist. Wir sagen auch *Vollkugel*, um darauf hinzuweisen, dass das Kugelinere mit zur Kugel gehören soll und *offen*, da wir die Kugeloberfläche nicht mehr dabei haben wollen. Wenn wir nur die Kugeloberfläche meinen, dann sagen wir *Sphäre*, wenn wir uns für Kugelinneres inklusive Oberfläche interessieren, dann sagen wir *abgeschlossene Vollkugel*.

Ein Sonderfall der offenen Kugeln sind die offenen Intervalle $]a, b[$ in \mathbb{R} . Die Intervallgrenzen bilden als Menge $\{a, b\}$ dessen „Oberfläche“, das Mittel von a und b ist der Kugelmittelpunkt und der Radius ε ist die halbe Intervallbreite. Insbesondere ist diese „Oberfläche“ des offenen Intervalls ebenfalls nicht Teil dieses Intervalls.

Nun wollen wir einige Beobachtungen zu Umgebungen machen, diese als Axiome festhalten, um so schrittweise zum Begriff der Topologie zu gelangen:

- Zunächst beobachten wir, dass jede Umgebung eines Punktes diesen Punkt als Element enthält, denn ist U_p Umgebung von p , dann existiert die Kugel $B_\varepsilon(p)$ für ein ε größer 0 als Teilmenge von U_p . Da $d(p, p) = 0$ also kleiner ε ist, ist $p \in B_\varepsilon(p)$ und somit auch Element von U_p .
- Zudem hat jeder Punkt $p \in X$ zumindest eine Umgebung, nämlich X selbst. Grund: Für ein beliebiges $\varepsilon > 0$ ist $B_\varepsilon(p)$ nach Definition Teilmenge von X .
- Ist U_p Umgebung von p und Teilmenge von V , dann ist auch V Umgebung von p , denn die Kugel in U_p liegt ja dann auch in V .

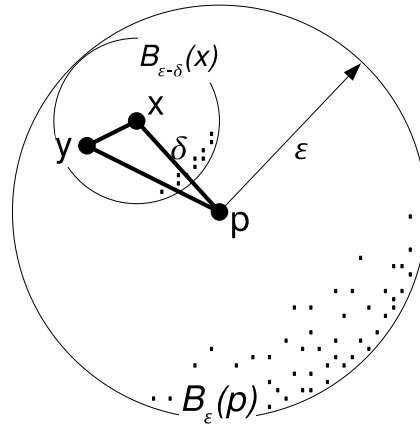


Abbildung 2.4: Eine offene Kugel $B_\varepsilon(p)$ umgibt jeden beliebigen ihrer Punkte (z.B. x), denn die kleinere Kugel $B_{\varepsilon-d(p,x)}(x)$ liegt vollständig in der größeren. Jeder Punkt y in der kleineren Kugel hat nämlich wegen der Dreiecksungleichung einen geringeren Abstand zu p als der größere Kugelradius ε .

- Sind U_p und V_p Umgebungen von p , dann ist auch deren Schnittmenge $U_p \cap V_p$ Umgebung von p .
- Jede offene Kugel ist Umgebung von allen in ihr enthaltenen Punkten.

Beweis. Wir zeigen zunächst nur die vorletzte der obigen Eigenschaften. Sind U_p und V_p Umgebungen von p , dann existieren jeweils eine Kugel $B_\varepsilon(p) \subseteq U_p$ und eine Kugel $B_\delta(p) \subseteq V_p$. Dann liegt die kleinere der beiden Kugeln $\min\{B_\varepsilon(p), B_\delta(p)\} = B_{\min\{\varepsilon, \delta\}}(p)$ in beiden Kugeln somit auch in beiden Umgebungen U_p und V_p und somit auch in deren Schnittmenge. Also ist diese eine Umgebung von p . \square

Damit ist auch der Schnitt $\bigcap \mathcal{V}$ jeder endlichen Menge von Umgebungen von p eine Umgebung von p . Die Voraussetzung der Endlichkeit ist dabei wichtig, denn die Menge $\{B_\varepsilon(p) \mid \varepsilon > 0\}$ aller ε -Kugeln um p (in \mathbb{R}^n mit $n > 0$) ist ebenfalls eine Menge von Umgebungen von p . Der Schnitt all dieser Umgebungen ist jedoch im allgemeinen *keine* Umgebung von p , denn es gilt $\bigcap \{B_\varepsilon(p) \mid \varepsilon > 0\} = \{p\}$. Ist nämlich ein Punkt q in dieser Schnittmenge, dann muss für den Abstand $d(q, p)$ gelten $0 \leq d(q, p) < \varepsilon$ für alle $\varepsilon > 0$. Also gilt $d(q, p) = 0$ und damit $q = p$. Jedoch ist $\{p\}$ im Allgemeinen keine Umgebung für p .

Zu zeigen bleibt noch, dass eine offene Kugel alle enthaltenen Punkte auch umgibt.

Beweis. Wir wählen einen beliebigen Punkt $x \in B_\varepsilon(p)$ und zeigen, dass diese Kugel um p auch Umgebung von x ist. Die Abbildung 2.4 veranschaulicht die nun folgende Argumentation:

x hat den Abstand $\delta = d(x, p) < \varepsilon$ zu p . Dann ist die Differenz $\varepsilon - \delta > 0$. Betrachten wir nun die kleinere Kugel $B_{\varepsilon-\delta}(x)$, dann behaupten wir, dass diese vollständig in der größeren Kugel $B_\varepsilon(p)$ liegt, denn für jeden Punkt $y \in B_{\varepsilon-\delta}(x)$

haben wir $d(y, x) < \varepsilon - \delta$. Nun gilt ebenfalls $d(x, p) < \varepsilon$ und da d eine Metrik ist haben wir die Dreiecksungleichung $d(y, x) + d(x, p) \geq d(y, p)$. Mit $d(y, x) + d(x, p) < (\varepsilon - \delta) + \delta = \varepsilon$ erhalten wir $d(y, p) < \varepsilon$. Damit liegt y in der großen Kugel $B_\varepsilon(p)$. Da y beliebig aus der kleinen Kugel gewählt wurde gilt $B_{\varepsilon-\delta}(x) \subseteq B_\varepsilon(p)$, also ist $B_\varepsilon(p)$ Umgebung von x . \square

Wenn wir nun diese Eigenschaften als Axiome festlegen, dann wiederholen wir im wesentlichen¹ die historische Definition des Begriffs *topologischer Raum* von Felix HAUSDORFF [Hau78]:

Definition 2.16 (Umgebungssystem). Sei M eine Menge und für jedes $x \in M$ eine Menge \mathcal{U}_x von Teilmengen von M mit folgenden Eigenschaften:

(U1) $x \in U$ für alle $U \in \mathcal{U}_x$. Und es gilt $M \in \mathcal{U}_x$.

(U2) Ist $U \in \mathcal{U}_x$ und $V \in \mathcal{U}_x$, dann ist auch $U \cap V \in \mathcal{U}_x$.

(U3) Ist $U \in \mathcal{U}_x$, dann existiert ein $B \in \mathcal{U}_x$ so dass für alle $b \in B$ gilt: $B \in \mathcal{U}_b$.

(U4) Ist $U \in \mathcal{U}_x$ und $U \subseteq V \subseteq M$, dann ist auch $V \in \mathcal{U}_x$.

Dann heißt $\mathcal{U} := \{\mathcal{U}_x \mid x \in M\}$ ein Umgebungssystem für M .

Wir fordern also: Jede Umgebung von x enthält x als Element. M ist Umgebung von jedem $x \in M$. Der Schnitt zweier Umgebungen von x ist selbst Umgebung von x . In jeder Umgebung von x liegt eine Umgebung von x , die auch Umgebung aller ihrer Punkte ist. Und jede Erweiterung einer Umgebung von x ist selbst Umgebung von x .

2.3 Topologische Räume

Die Angabe eines Umgebungssystems für eine Menge X war der historische Vorgänger zur modernen Definition des topologischen Raums. Heutzutage definiert man topologische Räume durch Angabe von sogenannten offenen Mengen. Das sind diejenigen Mengen, die alle ihre Punkte umgeben. Drei derartige Mengen sind durch die Umgebungsaxiome bereits gegeben:

- Die Gesamtmenge des metrischen Raums umgibt alle ihre Punkte. (U1)
- Zu jeder Umgebung eines Punktes ist gefordert, dass eine derartige offene Menge um diesen Punkt als Teilmenge existiert (U3). Im Fall des metrischen Raums ist dies z.B. eine offene Kugel.
- Die leere Menge ist Umgebung von allen ihren Punkten. Wir erinnern uns, dass eine Aussage der Form „Jeder Punkt der leeren Menge hat F00-Eigenschaft“ immer wahr ist, egal welches nun F00-Eigenschaft sein mag. So sind z.B. alle Marsmenschen grün, haben Antennen auf dem Kopf und riechen — wie wir auf Seite 32 bereits erfahren haben — nach Bananen. Hier ist F00 gleich „wird von \emptyset umgeben“.

¹HAUSDORFF forderte noch eine weitere Eigenschaft, die sogenannte Trennbarkeit, auf die heute in der allgemeinen Definition von Topologie verzichtet wird. Derart trennbare topologische Räume bezeichnet man heute auch als *Hausdorff-Räume*

Beobachten wir nun diejenigen Eigenschaften von offenen Mengen, die zur Definition von topologischen Räumen verwendet werden:

Satz 2.4 (Eigenschaften von offenen Mengen). *Sei X eine Menge und \mathcal{U} ein System von Umgebungen $\{\mathcal{U}_x\}_{x \in X}$. Dann gilt folgendes*

(O1) \emptyset ist offen und X ist offen.

(O2) Sind A und B offen, dann ist auch deren Schnittmenge $A \cap B$ offen.

(O3) Ist \mathcal{A} eine Menge von offenen Mengen, dann ist auch deren Vereinigung $\bigcup_{A \in \mathcal{A}} A$ eine offene Menge.

Jedes System von Umgebungen für X definiert genau ein System \mathcal{T} von offenen Mengen für X und umgekehrt.

Der Beweis steht in Lehrbüchern über Topologie. JÄNICH [Jän01, S.10] überlässt ihn der Anfängerin zur Übung. Als Beispiel soll das Axiom O2 gezeigt werden: Sind A und B offen und $p \in A \cap B$ beliebig, dann ist $p \in A$ und $p \in B$ und es sind A und B Umgebungen von p bzw. $A \in \mathcal{U}_p$ und $B \in \mathcal{U}_p$. Dann gilt aber auch wegen (U2): $A \cap B \in \mathcal{U}_p$. Da wir p beliebig aus $A \cap B$ gewählt haben, ist $A \cap B$ Umgebung von jedem seiner Punkte, und damit offen.

Umgebungen lassen sich also durch offene Mengen und diese wiederum durch Umgebungen definieren. Wie bereits gesagt ist es heute üblich, topologische Räume mit offenen Mengen zu definieren. Hier also die Definition des Begriffs Topologie:

Definition 2.17 (Topologie, topologischer Raum). *Sei X eine Menge und \mathcal{T} eine Menge von Teilmengen von X . Dann nennen wir \mathcal{T} eine Topologie für X , falls \mathcal{T} folgende Eigenschaften hat:*

(O1) $\emptyset \in \mathcal{T}$ und $X \in \mathcal{T}$.

(O2) Ist $A \in \mathcal{T}$ und $B \in \mathcal{T}$ dann ist auch $A \cap B \in \mathcal{T}$.

(O3) Ist $\mathcal{A} \subseteq \mathcal{T}$ so ist auch $(\bigcup_{A \in \mathcal{A}} A) \in \mathcal{T}$.

Das Paar (X, \mathcal{T}) heißt dann topologischer Raum und eine Menge $A \in \mathcal{T}$ heißt offen in (X, \mathcal{T}) . Die Komplementmenge $X \setminus A$ einer offenen Menge heißt abgeschlossen in (X, \mathcal{T}) .

Die nun folgende Beobachtung wird später noch wichtig werden. Zu Axiom (O2) ist äquivalent:

(O2') Ist $\mathcal{A} \subseteq \mathcal{T}$ und \mathcal{A} endlich, dann ist $(\bigcap_{A \in \mathcal{A}} A) \in \mathcal{T}$.

Das zweite Axiom sieht also auf dem ersten Blick dem dritten Axiom sehr ähnlich und unterscheidet sich nur durch eine Endlichkeitseinschränkung.

Beweisskizze:

(O2') \Rightarrow (O2): Angenommen, es gelte (O2'). Ist nun $A \in \mathcal{T}$ und $B \in \mathcal{T}$, dann ist $\{A, B\} \subseteq \mathcal{T}$ endlich. Dann ist $A \cap B = \bigcap \{A, B\} \in \mathcal{T}$, also gilt (O2).

(O2') \Leftarrow (O2): Dies beweist man durch Induktion über die Anzahl n von Elementen in \mathcal{A} .

Dabei setzt man für den Induktionsanfang: $(\bigcap_{A \in \emptyset} A) := X$. Der Induktionsschritt ist einfach und wird hier nicht ausgeführt.

Wir wollen diesen Begriff nun anhand einiger Beispiele illustrieren. Das letzte Beispiel soll schon andeuten, wo wir Topologie in der Architektur vorfinden. Die „Umgebung“ einer Türe ist bereits ein topologischer Raum für sich.

Beispiele 2.7 (Topologische Räume).

- $\{\emptyset\}$ ist die einzige Topologie für \emptyset . Der Raum $(\emptyset, \{\emptyset\})$ heißt leerer Raum.
- Für jede beliebige Menge X ist $\mathcal{P}(X)$, die Menge aller Teilmengen von X , eine Topologie für X . $\mathcal{P}(X)$ heißt diskrete Topologie und der Raum $(X, \mathcal{P}(X))$ diskreter Raum. Jede Topologie für X ist Teilmenge der diskreten Topologie.
- Auf der anderen Seite ist für X auch die Menge $\{\emptyset, X\}$ eine Topologie für X . Diese heißt indiskrete Topologie (oder auch „Klumpentopologie“ bzw. „triviale Topologie“). Der Raum $(X, \{\emptyset, X\})$ heißt entsprechend indiskreter Raum. Jede Topologie für X umfasst die indiskrete Topologie.
- Ist $\underline{X} = (X, \mathcal{T})$ ein topologischer Raum und $A \subseteq X$ eine beliebige Teilmenge von X , dann ist $\mathcal{T}|_A := \{A \cap U \mid U \in \mathcal{T}\}$ eine Topologie für A , die sogenannte Spurtopologie von A in (X, \mathcal{T}) . Der Raum $(A, \mathcal{T}|_A)$ heißt Teilraum von A in \underline{X} . Ein Teilraum von \mathbb{R} , von dem wir häufig Gebrauch machen werden, ist das abgeschlossene Einheitsintervall $[0, 1] \subseteq \mathbb{R}$. Auch ein Grundriss eines Gebäudes ist der Teilraum aller Punkte, die 1.20m über dem Rohfußboden des entsprechenden Geschosses liegen.
- Ist (X, d) ein metrischer Raum, dann heißt die in unserem einführenden Beispiel durch offenen Kugeln definierte Topologie auch metrische Topologie und wird mit $\mathcal{T}(d)$ bezeichnet. Ist d die euklidische Metrik $d : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$, dann heißt $\mathcal{T}(d)$ auch natürliche Topologie von \mathbb{R} . Wenn wir also vom „topologischen Raum“ \mathbb{R} sprechen, ohne eine Topologie anzugeben, dann ist damit stets der Raum $(\mathbb{R}, \mathcal{T}(d))$ mit der euklidischen Metrik gemeint. Entsprechendes gilt für kartesische Produkte \mathbb{R}^n und deren Teilmengen, wie z.B. das Einheitsintervall $[0, 1]$.

Umgekehrt heißt eine Topologie metrisierbar, wenn es eine Metrik gibt, mit der die Topologie definiert werden kann. Ist X endlich und \mathcal{T} metrisierbar, dann ist \mathcal{T} diskret, denn wenn d Metrik für X ist, dann gibt es einen kleinsten Abstand $r > 0$ zwischen zwei verschiedenen Punkten. Dann ist auch für jeden Punkt $p \in X$ die Kugel $B_r(p) = \{p\}$ und diese ist ja laut Definition offen. Da jede Teilmenge von X auch Vereinigung von solchen Kugeln ist, ist jede Teilmenge offen.

- Für $\{\text{Innen, Türe, Außen}\}$, ist

$$\{\emptyset, \{\text{Innen}\}, \{\text{Außen}\}, \{\text{Innen, Außen}\}, \{\text{Innen, Türe, Außen}\}\}$$

eine Topologie.

Wir haben also zu einer gegebenen Menge mehrere Möglichkeiten eine Topologie zu definieren. Diese liegen alle zwischen den beiden Extremfällen der diskreten und der indiskreten Topologie.

Wir benötigen noch weitere Grundbegriffe der Topologie:

Definition 2.18 (Inneres, Rand, Hülle). Sei $\underline{X} = (X, \mathcal{T})$ ein topologischer Raum, $A \subseteq X$ Teilmenge von X und $p \in X$ ein beliebiger Punkt (nicht notwendig in A). Dann heißt p ...

innerer Punkt von A , falls es eine offene Menge $U \in \mathcal{T}$ gibt, mit $p \in U \subseteq A$.

Ist p innerer Punkt von A , dann heißt A Umgebung von p . Die Menge aller inneren Punkte von A heißt auch Inneres (interior) von A (manchmal auch „der offene Kern von A “) und wird bezeichnet mit $\text{int}_{\underline{X}} A$.

Berührungspunkt von A , falls jede offene Menge $U \in \mathcal{T}$, die p enthält, die Menge A schneidet. p ist also genau dann Berührungspunkt von A , falls aus $p \in U \in \mathcal{T}$ folgt, dass $U \cap A \neq \emptyset$ gilt. Die Menge aller Berührungspunkte von A bezeichnen wir als abgeschlossene Hülle (closure) von A oder $\text{cl}_{\underline{X}} A$.

Randpunkt von A , falls p Berührungspunkt aber kein innerer Punkt von A ist.

Die Menge aller Randpunkte von A bezeichnen wir als Rand von A oder $\text{fr}_{\underline{X}} A$ (engl. „frontier“ oder „boundary“).

äußerer Punkt von A , falls p kein Berührungspunkt von A ist. Ein äußerer Punkt von A ist ein innerer Punkt der Komplementmenge $X \setminus A$.

Ist \underline{X} aus dem Kontext klar, dann schreiben wir auch $\text{int} A$ statt $\text{int}_{\underline{X}} A$ und sinngemäß für cl und fr . In der Literatur sind auch die Bezeichnungen \bar{A} für die abgeschlossene Hülle von A , $\overset{\circ}{A}$ für den offenen Kern von A und ∂A für den Rand von A üblich. Die hier verwendete Schreibweise folgt [Her86b].

Die Abbildung 2.5 veranschaulicht diese Definition und nun folgen einige Beispiele für diese Begriffe:

Beispiel 2.8 (Türumgebung). Am Beispiel der „Türumgebung“ ergibt sich: Türe ist Randpunkt (also auch Berührungspunkt) der Mengen $\{\text{Innen}\}$, $\{\text{Außen}\}$ oder auch $\{\text{Innen}, \text{Türe}\}$.

$\{\text{Innen}\}$ ist das Innere von $\{\text{Innen}, \text{Türe}\}$.

$\{\text{Innen}, \text{Türe}\}$ ist wiederum die abgeschlossene Hülle von $\{\text{Innen}\}$.

Beispiel 2.9 (Vollkugel). Betrachten wir nun die Kugel $B_\varepsilon(p) \subseteq \mathbb{R}^n$:

Zur offenen Vollkugel $B_\varepsilon(p)$ gibt es die abgeschlossene Hülle

$$\text{cl} B_\varepsilon(p) = \{q \in X \mid d(p, q) \leq \varepsilon\},$$

die auch abgeschlossene Vollkugel heißt. Der Rand dieser Kugel heißt

$$\text{fr} B_\varepsilon(p) = S_\varepsilon(p) = \{q \in X \mid d(p, q) = \varepsilon\}$$

die ε -Sphäre um p .

Ohne Beweis wollen wir noch festhalten:

Satz 2.5 (abgeschlossene Hülle, offener Kern). Sei $\underline{X} = (X, \mathcal{T})$ topologischer Raum, $A \subseteq X$ beliebig. Dann gilt:

1. Die abgeschlossene Hülle $\text{cl}_{\underline{X}} A$ von A ist eine abgeschlossene Menge. Jede abgeschlossene Menge, die A umfasst, umfasst auch die abgeschlossene Hülle von A . $\text{cl}_{\underline{X}} A$ ist also die kleinste abgeschlossene Menge, die A umfasst und insbesondere gilt daher $\text{cl}_{\underline{X}} \text{cl}_{\underline{X}} A = \text{cl}_{\underline{X}} A$.
2. Der offene Kern $\text{int}_{\underline{X}} A$ von A ist eine offene Menge. Jede offene Teilmenge von \underline{X} , die in A liegt ist auch Teilmenge von $\text{int}_{\underline{X}} A$. Diese ist also die größte in \underline{X} offene Teilmenge von A und insbesondere gilt daher $\text{int}_{\underline{X}} \text{int}_{\underline{X}} A = \text{int}_{\underline{X}} A$.

Offene Kerne und abgeschlossene Hüllen sind demnach in der Tat offen bzw. abgeschlossen und heißen nicht bloß so und die Hülle der Hülle einer Menge ist einfach die Hülle dieser Menge und entsprechendes gilt auch für den Kern. Beweise finden sich bei OSSA [Oss92, Satz 2.1.6] und bei SCHUBERT [Sch75, Abschnitt 2.4f].

2.4 Stetige Abbildungen

In der Analysis heißt eine Abbildung $f : \mathbb{R} \rightarrow \mathbb{R}$ stetig, wenn sie an keiner Stelle „einen Sprung“ macht. Dabei sollen $f(x)$ und $f(y)$ immer dann „nahe beieinander“ liegen, wenn x und y „nahe beieinander“ sind. Auch hier taucht offenbar der Begriff der „Nähe“ auf. Ersetzen wir „nahe beieinander“ durch den oben definierten Begriff des Berührungspunktes, dann können wir stetige Abbildungen als diejenigen Abbildungen definieren, die Berührungspunkte erhalten. Ist also p Berührungspunkt einer Menge M , dann ist $f(p)$ Berührungspunkt der Menge $f[M]$.

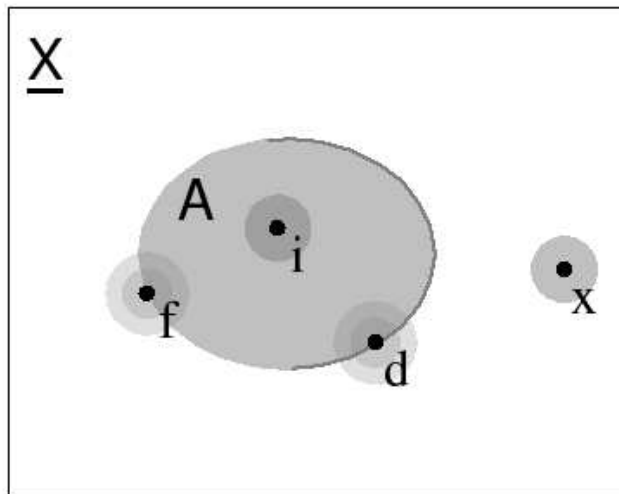


Abbildung 2.5: Innerer Punkt i , Randpunkte d und f , und äußerer Punkt x einer Menge A . Eine Scheibe um einen Punkt steht für eine Umgebung dieses Punktes. Jede Umgebung eines Randpunktes (kleinere und größere Scheibe um d und f) liegt teilweise in A und teilweise außerhalb von A .

Mengen werden also von stetigen Abbildungen nicht „zerrissen“ (nach [Her86a, S.29]).

Zu dieser sehr anschaulichen Definition ist die folgende, üblichere Definition von Stetigkeit äquivalent:

Definition 2.19 (Stetige Abbildungen). *Seien (X, \mathcal{T}_X) und (Y, \mathcal{T}_Y) topologische Räume und $f : X \rightarrow Y$ eine Abbildung von der Punktmenge X in die Punktmenge Y . Dann heißt f stetig, wenn das Urbild jeder offenen Menge in (Y, \mathcal{T}_Y) offen in (X, \mathcal{T}_X) ist, wenn also gilt:*

$$f^{-1}[B] \in \mathcal{T}_X \text{ für alle } B \in \mathcal{T}_Y.$$

Für stetige Abbildungen schreiben wir dann $f : (X, \mathcal{T}_X) \rightarrow (Y, \mathcal{T}_Y)$.

Wir wollen hier zeigen, wie man beweist, dass diese Definition äquivalent zu der Forderung ist, dass die Abbildung alle Berührungspunkte erhält. Dies zeigt, wie in der Topologie argumentiert wird, und kann auch in den meisten Lehrbüchern nachgelesen werden.

Beweis. Wir müssen einerseits zeigen, dass stetige Abbildungen Berührungspunkte erhalten (\implies) und andererseits, dass die Umkehrung ebenfalls gilt und Abbildungen, die Berührungspunkte erhalten, auch stetig sind (\impliedby):

\implies Sei $f : X \rightarrow Y$ stetig, $p \in X$ Berührungspunkt von $M \subseteq X$ und U eine beliebige Umgebung von $f(p)$. Dann existiert eine offene Menge $V \subseteq U$ mit $f(p) \in V$. Dann ist $f^{-1}[V]$ offen nach Definition der Stetigkeit und wegen $f(p) \in V$ ist $p \in f^{-1}[V]$, also ist $f^{-1}[V]$ Umgebung von p . Da p Berührungspunkt von M ist, gibt es einen Punkt $m \in M \cap f^{-1}[V]$. Es ist $f(m) \in f[M]$ und $f(m) \in V$ und somit auch $f(m) \in U$, also ist $f(m) \in f[M] \cap U$. Dann ist $f[M] \cap U$ nicht leer. Da U beliebig gewählt wurde, gilt dies für jede Umgebung von $f(p)$, also ist $f(p)$ Berührungspunkt von $f[M]$.

\impliedby Habe $f : X \rightarrow Y$ die Eigenschaft alle Berührungspunkte zu erhalten und sei V offen in Y . Ist nun p Berührungspunkt von $X \setminus f^{-1}[V]$, dann ist $f(p)$ Berührungspunkt von $f(X \setminus f^{-1}[V])$. Es gilt $f[X \setminus f^{-1}[V]] \subseteq Y \setminus V$ und somit ist $f(p)$ ebenfalls Berührungspunkt von $Y \setminus V$. Da V offen ist, ist $Y \setminus V$ abgeschlossen, enthält also alle Berührungspunkte. Daher gilt sogar: $f(p)$ ist Element von $Y \setminus V$. Damit ist $f(p)$ kein Element von V , also p kein Element von $f^{-1}[V]$ und somit ist p Element der Komplementmenge $X \setminus f^{-1}[V]$. Wir haben aber p nur als Berührungspunkt dieser Menge vorausgesetzt, also enthält $X \setminus f^{-1}[V]$ alle Berührungspunkte, ist somit abgeschlossen und damit ist sein Komplement $f^{-1}[V]$ offen. Also ist f stetig.

Damit sind beide Eigenschaften äquivalent. □

Es folgen Beispiele von stetigen Abbildungen. Dabei soll bereits die architektonische Anwendung angedeutet werden. Später werden wir sehen, wie Gebäude und die dazugehörigen Pläne als topologische Räume betrachtet werden können. Dass dies möglich ist, setzen wir hier zunächst einfach voraus.

Beispiel 2.10 (stetige Abbildungen). *Seien \underline{X} und \underline{Y} topologische Räume.*

- Die identische Abbildung $\text{id}_{\underline{X}} : \underline{X} \rightarrow \underline{X}$ ist stetig. Denn das Urbild einer offenen Menge A unter $\text{id}_{\underline{X}}$ ist eben diese Menge A .
- Für jedes $y \in Y$ ist die konstante Abbildung $c_y : \underline{X} \rightarrow \underline{Y}$, $x \mapsto y$ stetig. Denn ist B eine offene Teilmenge von Y , dann gilt entweder $y \in B$ oder $y \notin B$. Im ersten Fall gilt $c_y^{-1}[B] = X$ und im zweiten Fall $c_y^{-1}[B] = \emptyset$. Damit ist das Urbild von B eine offene Menge.
- Eine stetige Abbildung $w : [0, 1] \rightarrow \underline{X}$ heißt auch Weg in \underline{X} . Dann bezeichnet man $w(0)$ als Startpunkt von w und $w(1)$ als Endpunkt von w .

- Ist (A, \mathcal{T}_A) Teilraum von (X, \mathcal{T}_X) also $A \subseteq X$ und $\mathcal{T}_A = \mathcal{T}_X|_A$, dann ist die Inklusion

$$i : A \hookrightarrow X, a \mapsto a$$

eine stetige Abbildung, denn für jede Teilmenge $U \subseteq X$ ist $i^{-1}[U] = A \cap U$. Ist nun U offen in (X, \mathcal{T}_X) , dann ergibt sich $A \cap U \in \mathcal{T}_X|_A$ unmittelbar aus der Definition der Spurtopologie.

- Ist \underline{X} ein diskreter Raum, dann ist jede Abbildung $f : \underline{X} \rightarrow \underline{Y}$ stetig.
- Ist \underline{Y} ein trivialer Raum, dann ist ebenfalls jede Abbildung $f : \underline{X} \rightarrow \underline{Y}$ stetig.

- **Beziehung Plan und geplantes Objekt.** Sei Z eine Planzeichnung (Grundriss, Schnitt) eines architektonischen Objekts A . Dann gibt es zu jedem Symbol s in Z ein zugeordnetes Element $\sigma(s)$ in A . Damit haben wir eine Abbildung $\sigma : Z \rightarrow A$, $s \mapsto \sigma(s)$, die Bedeutung von Z . Ist der Plan Z korrekt, dann ist die Bedeutung von Z , also σ , eine stetige Abbildung.

- **Beziehung von Plänen unterschiedlichen Maßstabs.** Seien $Z_{1:100}$ ein Entwurfsplan und $Z_{1:50}$ ein Werkplan eines Ausschnitts von $Z_{1:100}$, der Objekte in $Z_{1:100}$ weiter detailliert. Wenn es zu jeder symbolischen Darstellung d im Werkplan ein Objekt $f(d)$ im Entwurfsplan gibt, dem d angehört, dann ist $f : Z_{1:50} \rightarrow Z_{1:100}$ eine stetige Abbildung. Ist z.B. $d \in Z_{1:50}$ ein im Entwurf noch nicht vorgesehener Deckendurchbruch der Deckenplatte $D \in Z_{1:50}$, dann ist $f(d)$ die entsprechende Decke in $Z_{1:100}$, denn der Durchbruch d berührt die durchbrochene Decke D , also muss gelten $f(d)$ berührt $\{f(D)\}$. Also gilt $f(d) \in \text{cl}\{f(D)\}$. Ist aber $f(d) \neq f(D)$, dann ist ja $f(d)$ ein Element, das die Stelle des vorgesehenen Deckendurchbruchs markiert. Wir haben aber vorausgesetzt, dass ein solcher Deckendurchbruch in $Z_{1:100}$ nicht vorgesehen war. Da aus $f(d) \neq f(D)$ also ein Widerspruch folgt, kann es nicht richtig sein und es muss gelten $f(d) = f(D)$. Diese Argumentation ist allerdings nicht mathematisch streng, sondern verwendet auch anschauliche Überlegungen. So müsste etwa definiert werden, was „nicht vorgesehen“ bedeutet.

Stetige Abbildungen dienen auch dazu, topologische Räume miteinander zu vergleichen. Sie haben für diesen Zweck die uns bereits bekannten nützlichen Eigenschaften:

Satz 2.6 (Eigenschaften stetiger Abbildungen). Seien \underline{X} , \underline{Y} und \underline{Z} topologische Räume und $f : \underline{X} \rightarrow \underline{Y}$, $g : \underline{Y} \rightarrow \underline{Z}$ und $h : \underline{Z} \rightarrow \underline{M}$ beliebige stetige Abbildungen. Dann gilt immer

1. $g \circ f : \underline{X} \rightarrow \underline{Z}$ ist ebenfalls eine stetige Abbildung.
2. Die Verkettung stetiger Abbildungen ist assoziativ: $(f \circ g) \circ h = f \circ (g \circ h)$
3. $f \circ \text{id}_{\underline{X}} = f = \text{id}_{\underline{Y}} \circ f$

Stetige Abbildungen sind also Morphismen. Der Beweis ist trivial, denn ist das Urbild jeder offenen Menge offen, dann ist das Urbild des Urbildes einer offenen Menge ebenfalls offen und damit $g \circ f$ stetig. Die weiteren Eigenschaften vererben sich von den entsprechenden Eigenschaften der Verkettung von Abbildungen.

Diese Eigenschaften werden wir nun verwenden, um festzulegen, wann wir zwei topologische Räume als „gleich“ betrachten. Sie sind es dann, wenn jeder ein stetiges Bild des jeweils anderen ist und sich die dabei beteiligten Abbildungen gegenseitig „aufheben“. Abbildungen (oder Morphismen allgemein), die wir auf diese Art rückgängig machen können, heißen *Isomorphismen*. Die topologischen Isomorphismen heißen auch *Homöomorphismen*.

Definition 2.20 (Homöomorphismus, homöomorph). Seien \underline{X} und \underline{Y} topologische Räume. Eine stetige Abbildung $f : \underline{X} \rightarrow \underline{Y}$ heißt Homöomorphismus, wenn es eine stetige Abbildung $g : \underline{Y} \rightarrow \underline{X}$ zurück gibt, so dass $g \circ f = \text{id}_{\underline{X}}$ und $f \circ g = \text{id}_{\underline{Y}}$. Zwei topologische Räume \underline{X} und \underline{Y} heißen homöomorph, wenn es einen Homöomorphismus $f : \underline{X} \rightarrow \underline{Y}$ gibt.

Diese Definition von Homöomorphismus weicht geringfügig von der üblichen Definition in den Lehrbüchern ab, ist aber dazu äquivalent. Die obige, sogenannte kategorielle Definition wollen wir hier bevorzugen, da sie einer allgemeineren Definition von Isomorphie in der Mathematik entspricht.

Die Abbildung g zurück ist eindeutig bestimmt, denn ist h ebenfalls eine solche Abbildung zurück dann gilt

$$g = \text{id}_{\underline{X}} \circ g = (h \circ f) \circ g = h \circ (f \circ g) = h \circ \text{id}_{\underline{Y}} = h$$

und somit $g = h$. Die Abbildung g heißt auch Umkehrmorphismus von f und wird mit f^{-1} bezeichnet. Diese Eindeutigkeit gilt für jeden Isomorphismus.

Wir wollen nun eine Kaffeetasse und einen Doughnut als „gleich“ betrachten:

Satz 2.7. *Homöomorphie ist eine Äquivalenzrelation zwischen topologischen Räumen.*

Beweis. Wir müssen die Eigenschaften reflexiv, symmetrisch und transitiv zeigen:

- **Reflexivität.** Für jeden topologischen Raum \underline{X} ist die identische Abbildung $\text{id}_{\underline{X}} : \underline{X} \rightarrow \underline{X}$ ein Homöomorphismus, denn sie ist stetig und gleichzeitig ihr eigener Umkehrmorphismus: $\text{id}_{\underline{X}} \circ \text{id}_{\underline{X}} = \text{id}_{\underline{X}}$. Also ist \underline{X} homöomorph zu sich selbst.
- **Symmetrie.** Ist \underline{X} zu \underline{Y} homöomorph mit Homöomorphismus $f : \underline{X} \rightarrow \underline{Y}$, dann sieht man leicht, dass man in der obigen Definition g mit f tauschen kann.

- **Transitivität.** Sind $f : \underline{X} \rightarrow \underline{Y}$ und $g : \underline{Y} \rightarrow \underline{Z}$ Homöomorphismen. Dann existieren Umkehrmorphisme, $f^{-1} : \underline{Y} \rightarrow \underline{X}$ und $g^{-1} : \underline{Z} \rightarrow \underline{Y}$ wie oben. Dann ist auch $g \circ f : \underline{X} \rightarrow \underline{Z}$ ein Homöomorphismus, denn $f^{-1} \circ g^{-1}$ ist sein Umkehrmorphismus:

$$\begin{aligned} (g \circ f) \circ (f^{-1} \circ g^{-1}) &= (g \circ (f \circ f^{-1}) \circ g^{-1}) \\ &= (g \circ \text{id}_{\underline{Y}} \circ g^{-1}) \\ &= (g \circ g^{-1}) = \text{id}_{\underline{Z}} \\ (f^{-1} \circ g^{-1}) \circ (g \circ f) &= (f^{-1} \circ (g^{-1} \circ g) \circ f) \\ &= (f^{-1} \circ \text{id}_{\underline{Y}} \circ f) \\ &= (f^{-1} \circ f) = \text{id}_{\underline{X}} \end{aligned}$$

□

Bei diesem Beweis haben wir ausschließlich Isomorphie-Eigenschaften benutzt. Wir haben also gezeigt, dass Isomorphie im Allgemeinen eine Äquivalenzrelation ist. Dies ist der Vorteil der allgemeinen Definition von Isomorphie.

Beispiele 2.11 (Homöomorphe Räume).

- Die Menge der reellen Zahlen ist homöomorph zu dem offenen Intervall $]0, 1[\subseteq \mathbb{R}$, denn die Abbildungen

$$f :]-1, 1[\rightarrow \mathbb{R}, x \mapsto \frac{x}{1-x^2}$$

und

$$g :]0, 1[\rightarrow]-1, 1[, x \mapsto 2x - 1$$

sind Homöomorphismen, also ist auch $f \circ g :]0, 1[\rightarrow \mathbb{R}$ ein Homöomorphismus.

- Nehmen wir ein Gebäude, bestehend aus Elementen A und einer Topologie \mathcal{T}_A . Nun wollen wir ein Modell dieses Raumes $\text{code}(A)$ im Computer erzeugen, das alle topologischen Informationen von (A, \mathcal{T}_A) enthält. Dann ist in diesem Modell auch eine Topologie \mathcal{T}_c definiert, so dass $(\text{code}(A), \mathcal{T}_c)$ homöomorph zu (A, \mathcal{T}_A) ist. Insbesondere sollte jedem architektonischen Element a ein Objekt $\text{code}(a)$ entsprechen. Also wollen wir, dass code eben dieser Homöomorphismus ist. Dies zu ermöglichen ist ja genau das Ziel dieser Arbeit.

Im Allgemeinen ist es sehr schwierig zu entscheiden, ob zwei gegebene topologische Räume homöomorph sind. Denn gäbe es ein solches allgemeines Entscheidungsverfahren (einen Algorithmus) dafür, dann würde dies auch das sogenannte Graphen-Isomorphismus-Problem (oder kurz GI) lösen. Wir werden später sehen, wie wir einem Graphen einen topologischen Raum so zuordnen, dass zwei Graphen genau dann isomorph sind, wenn die entsprechenden topologischen Räume homöomorph sind. Damit lässt sich GI auf das Entscheiden von Homöomorphismus reduzieren. Also ist das Entscheiden von Homöomorphismus (nicht notwendigerweise echt) schwerer als GI. Da aber für das Graphen-Isomorphismus-Problem kein effizienter Algorithmus bekannt ist und ein solcher möglicherweise gar nicht existiert [Hof82], kennt man auch keinen effizienten Algorithmus zur Entscheidung von Homöomorphismus.

2.5 Quotientenbildung

Es gibt einen Zoo von topologischen Konstruktionen, die jedoch erst im Zusammenhang mit den topologischen Datenbanken eingeführt werden sollen. Wir hatten aber die Konstruktion des Teilraums bereits vorweggenommen und wollen dies nun auch mit dem Quotientenraum tun, da er grundlegend für diese Arbeit ist.

Architektur ist nach gängiger Lehrmeinung (unter anderem) eine „Organisation des Raums“. Auf den zu organisierenden Raum sind wir bereits eingegangen — es ist der euklidische n -dimensionale reelle Raum \mathbb{R}^n mit seiner metrischen Topologie. Als seine „Organisation“ wollen wir nun eine Zerlegung in endlich viele Teile betrachten. Natürlich gibt dies noch nicht adäquat die Vielschichtigkeit einer architektonischen Gliederung wieder. Zumindest erhalten wir so jeweils *eine* solche Schicht als Ausgangspunkt weiterer Überlegungen.

Wir werden nun schrittweise Gebäudeelemente aus dem Raum „herausschneiden“. Der Raum bekommt dabei „Löcher“ und es verbleibt stets ein Restraum, den wir als Gerüst (oder Skelett) bezeichnen. Aus diesem Gerüst schneiden wir weitere Elemente heraus bis es vollständig aufgebraucht ist. Unser Gerüst wird dabei immer kleiner und die Anzahl unserer Elemente immer größer. Am Ende haben wir den gesamten Raum in derartige Elemente zerlegt und der Restraum ist leer. Dabei beginnen wir mit den dreidimensionalen Objekten und gehen dann schrittweise je eine Dimensionszahl hinunter, bis wir bei den nulldimensionalen Objekten angelangt sind. Wir nehmen es dabei mit „Dimension“ nicht so streng und betrachten z.B. ein Wandelement oder ein Türblatt als zweidimensional, obwohl es doch eigentlich ein Volumen bildet. Wichtig ist dabei nur, dass wir jeweils im Gerüst als Teilraum offene Mengen als Elemente herausschneiden.

Volumenobjekte. Sei \mathcal{G} (wie „Gebäude“) ein Bauwerk mit Bauelementen und seinen nutzbaren Räumen, im dreidimensionalen Raum \mathbb{R}^3 realisiert. Bezeichnen wir diesen Raum mit X^3 , das Volumengerüst in dem es bisher noch keine Löcher gibt. Dann ordnen wir jedem Punkt $(x, y, z) \in \mathbb{R}^3$ ein Element von \mathcal{G} zu. Betrachten wir zuerst alle nutzbaren Räume von \mathcal{G} , also das, was von X^3 übrig bleibt, wenn wir alle Punkte entfernen, die ein Bauelement berühren. Dann haben wir wieder die Relation R_Z in X^3 mit $R_Z(u, v)$ genau dann, wenn u im Inneren des selben nutzbaren Raum wie v liegt. Dabei sollen auch der Außenraum oberhalb des Geländes und das Erdreich unterhalb des Geländes als ein solcher „nutzbarer“ Raum betrachtet werden. Diese bilden unsere erste Menge E_3 von Gebäudeelementen, sind also die Volumen.

Flächige Objekte. Nun nehmen wir die verbleibenden Punkte aus X^3 , die wir mit X^2 oder, dem Flächengerüst bezeichnen. Darin liegen alle Bauelemente, die wir als „flächig“ betrachten, da ihre Dicke (in \mathbb{R}^3) gering ist im Vergleich zu den weiteren Abmessungen. Dies sind die Türöffnungen, Fensterscheiben oder auch Boden- und Deckenplatten und Öffnungen als Übergänge von einem Raum in einen anderen (gewissermaßen Bauelemente aus dem Material „Luft“). Wie die Räume von flächigen Elementen begrenzt werden, werden diese wiederum von „linienartigen“ Elementen begrenzt, wie etwa Tür- und Fensterrahmen, Schwellen, der obere (untere) Wandanschluss an eine Deckenplatte oder der

Anschluss mehrerer Wände an einer Raumecke. Auch hier definieren wir eine Äquivalenzrelation R_F in \mathbb{R}^3 mit $R_F(x, y)$ genau dann, wenn x und y in X^2 in einer gemeinsamen Facette eines flächigen Bauelements liegen und *nicht* deren Randlinie berühren. Insbesondere ist das Gelände ein solches flächiges Objekt. Diese flächigen Objekte bilden die Menge E_2 .

Kanten und Knoten. Die linienartigen Verbindungen bilden wiederum eine noch kleinere Restmenge X^1 , das Kantengerüst, wobei diese Linien selbst wiederum durch Punkte oder konstruktive Knoten begrenzt sind. Auch hier haben wir eine Relation R_L mit $R_L(x, y)$ genau dann, wenn x und y in einem gemeinsamen Abschnitt eines derartigen Elements liegen. Die nun noch verbleibenden Punkte bilden das Knotengerüst X^0 mit Relation R_K : $R_K(x, y)$ genau dann wenn x den selben Verbindungsknoten wie y berührt. Wir haben damit die Mengen E_1 der Kanten und die Menge E_0 der Verbindungsknoten und unser Raum wurde vollständig aufgebraucht. Das nächste Skelett X^{-1} ist also der leere Raum $(\emptyset, \{\emptyset\})$.

So haben wir eine Äquivalenzrelation $R_G := R_Z \cup R_F \cup R_L \cup R_K$ im euklidischen Raum \mathbb{R}^3 und damit eine Zerlegung des \mathbb{R}^3 in Äquivalenzklassen, deren jede einem architektonischen Element in \mathcal{G} entspricht. Diese Zerlegung ist die Vereinigung $E_3 \cup E_2 \cup E_1 \cup E_0$ der Mengen unserer Gebäudeelemente.

Abbildung 2.6 veranschaulicht dies für eine zweidimensionale Welt mit einer stark vereinfachten „Architektur“. Das Verfahren funktioniert aber in beliebiger Dimension n und bei beliebiger Komplexität des architektonischen Entwurfs. Ist n die Dimension des Raums, dann entstehen bei dieser Zerlegung $n + 1$ derartige Gerüste X^i und Elementemengen E_i .

Mit dieser Zerlegung haben wir zu dem unendlichen topologischen Raum \mathbb{R}^3 eine endliche Menge $\mathbb{R}^3/\mathcal{G} := \mathbb{R}^3/R_G$ erhalten. Für diese Menge suchen wir nun eine Topologie. Diese soll folgende Eigenschaften haben:

- Die natürliche Abbildung $\pi : \mathbb{R}^3 \rightarrow \mathbb{R}^3/\mathcal{G}$ soll stetig sein.
- Mit der indiskreten Topologie $\{\emptyset, \mathbb{R}^3/\mathcal{G}\}$ ist π selbstverständlich immer stetig, aber das ist für diese Zwecke zu grob. Die Topologie soll nämlich weitgehend von π abhängig sein.

Eine Präzision dieser Forderungen geben wir im Abschnitt über die topologischen Konstruktionen. In diesem Fall tut es die sogenannte Quotiententopologie: Eine Menge von Äquivalenzklassen ist offen im Quotientenraum, wenn deren Vereinigung offen im ursprünglichen Raum ist.

Definition 2.21 (Quotiententopologie, Quotientenraum). Seien (X, \mathcal{T}_X) ein topologischer Raum und \sim eine Äquivalenzrelation auf X . Dann heißt

$$\mathcal{T}_X/\sim := \left\{ \mathcal{A} \subseteq X/\sim \mid \bigcup \mathcal{A} \in \mathcal{T}_X \right\}$$

Quotiententopologie von \mathcal{T}_X bezüglich R . Das Paar

$$(X, \mathcal{T}_X)/\sim := (X/\sim, \mathcal{T}_X/\sim)$$

heißt entsprechend Quotientenraum .

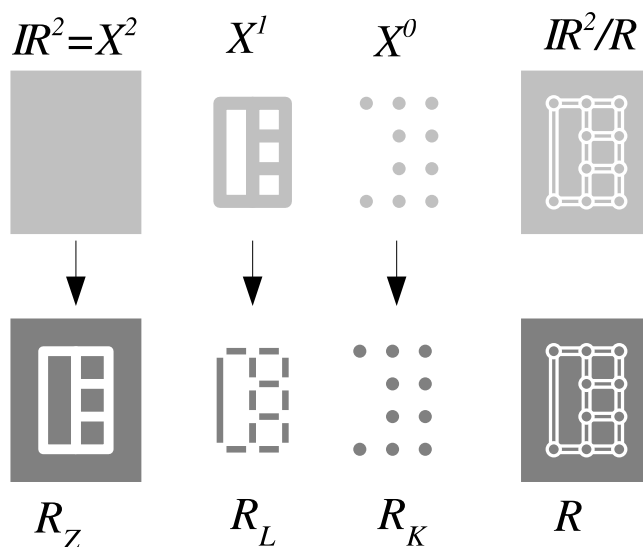


Abbildung 2.6: Die architektonische Zerlegung in einer zweidimensionalen Welt X^2 (Flatland). Die obere Zeile zeigt jeweils die verbleibenden Gerüste, wenn wir alle Bauelemente einer bestimmten Dimension entfernen. Die entfernten Elemente sind in der unteren Zeile gezeichnet. Diese Teilmengen des X^2 bilden dann den Quotientenraum X^2/R .

Der sehr einfache Beweis, dass \mathcal{T}_X/\sim tatsächlich eine Topologie ist, kann in einschlägigen Lehrbüchern nachgeschlagen werden (z.B. bei SCHUBERT [Sch75, S.29]).

Die Quotiententopologie hat folgende Eigenschaft: Jede Topologie \mathcal{T} für X/\sim mit der $\pi : (X, \mathcal{T}_X) \rightarrow (X/\sim, \mathcal{T})$ stetig ist, ist eine Teilmenge von \mathcal{T}_X/\sim . Die Quotiententopologie ist also eine maximale Topologie, in der π stetig ist, und als solche eindeutig.

Eine wichtige „Anwendung“ des Quotientenraums ist das sogenannte Zusammenschlagen: Man nehme einen topologischen Raum (X, \mathcal{T}_X) und eine Teilmenge $A \subseteq X$. Durch A ist eine Relation R_A in X wie folgt definiert: $a R_A b$ gilt genau dann, wenn $\{a, b\} \subseteq A$ gilt. Es ist also ein Sonderfall unserer „Zimmerrelation“, wenn es nur genau ein solches „Zimmer“ A in unserem Raum gibt. Diese Relation ist bekanntermaßen transitiv und symmetrisch, nicht jedoch reflexiv. Es ist jedoch dann R_A^* die von R_A erzeugte Äquivalenzrelation und der Quotientenraum $(X, \mathcal{T}_X)/A := (X, \mathcal{T}_X)/R_A^*$ heißt *Zusammenschlagen* der Menge A in (X, \mathcal{T}_X) . Dabei wird die Menge A zu einem einzigen Punkt zusammengefasst und ansonsten bleibt der Raum weitest möglich erhalten.

Mit der Bildung der Quotiententopologie bezüglich der durch Bauelemente erzeugten Zerlegung des Raums haben wir also sofort eine Topologie für eben diese Bauelemente, deren Verbindungen und die nutzbaren Räume in einem Gebäude, also eine Topologie für Architektur. Zudem haben wir aus dem unendlich großen topologischen Raum \mathbb{R}^3 einen endlichen Raum \mathbb{R}^3/\mathcal{G} gewonnen, denn diese Endlichkeit haben wir ja für \mathcal{G} vorausgesetzt. Damit ist auch dessen Topologie endlich und somit ist jede Menge \mathcal{U} von offenen Mengen in \mathbb{R}^3/\mathcal{G} end-

lich. Wir erinnern uns, dass dann auch deren Schnittmenge $\bigcap \mathcal{U}$ offen ist und könnten somit in diesem Fall in unserer Definition des topologischen Raums auf die Endlichkeitsbedingung verzichten. Mit diesem Verzicht erhalten wir nun eine besondere Klasse von topologischen Räumen:

Definition 2.22 (alexandroffscher Raum). *Ein topologischer Raum (X, \mathcal{T}_X) heißt alexandroffsch, wenn für jede beliebige Menge \mathcal{U} offener Mengen deren Schnitt $\bigcap \mathcal{U}$ ebenfalls offen ist, wenn also für jede beliebige Teilmenge $\mathcal{U} \subseteq \mathcal{T}_X$ gilt: $\bigcap \mathcal{U} \in \mathcal{T}_X$.*

ALEXANDROFF selbst bezeichnete diese Räume als „diskrete“ Räume [Ale37], worunter man heute jedoch etwas anderes versteht. Allerdings ist jeder diskrete Raum im heutigen Wortsinn auch alexandroffsch².

Es ist klar, dass ein endlicher topologischer Raum stets alexandroffsch ist. Umgekehrt gibt es jedoch unendliche alexandroffsche Räume, wie etwa der diskrete Raum $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ oder der indiskrete Raum $(\mathbb{R}, \{\emptyset, \mathbb{R}\})$. Wir werden später sehen, dass zwischen den alexandroffschen Räumen und Relationen eine enge Beziehung besteht, was wir ausnutzen werden, um Topologien für relationale Datenbanken zu definieren. Man bedenke, dass eine Tabelle einer relationalen Datenbank (in der Praxis) ebenfalls stets endlich ist und somit ist auch jede Topologie dafür alexandroffsch sein muss. Zudem bildet die Menge der abgeschlossenen Mengen eines alexandroffschen Raums ebenfalls eine Topologie für dessen Punktmenge und mit dieser den sogenannten *dualen Raum*.

2.6 Algebraische Topologie

Die algebraische Topologie beschäftigt sich mit der Untersuchung von topologischen Räumen unter Einsatz von algebraischen Mitteln. Sie ist auch eine mathematische Grundlage der rechnergestützten Volumenmodellierung und auch der „Topologien“ von Geoinformationssystemen.

Zunächst betrachten wir ein recht anschauliches Beispiel für eine derartige „Algebraisierung“ von topologischen Räumen: Wir wählen einen festen Basispunkt x in einem Raum und betrachten die Menge aller Rundwege, die von x nach x führen. Zwei Rundwege können auch nacheinander gegangen werden und ergeben so einen neuen Rundweg.

Beispiel 2.12 (Fundamentalgruppe). *Wir erinnern uns, dass ein Weg in einem topologischen Raum $\underline{X} = (X, \mathcal{T}_X)$ eine stetige Abbildung $w : [0, 1] \rightarrow \underline{X}$ mit Startpunkt $w(0)$ und Endpunkt $w(1)$ ist. Ein Rundweg mit Basispunkt x ist ein Weg w mit $x = w(0) = w(1)$. Sei $\text{Weg}(\underline{X}, x)$ die Menge aller Rundwege mit Basispunkt x , dann können wir für zwei Rundwege v, w eine Verknüpfung $v * w$ – die Komposition von Wegen – definieren, ein Rundweg, der zuerst den Weg v mit doppelter Geschwindigkeit von 0 bis 0.5 durchläuft und dann den Weg w ebenfalls mit doppelter Geschwindigkeit von 0.5 bis 1:*

$$v * w : [0, 1] \rightarrow X, t \mapsto v * w(t) := \begin{cases} v(2t) & : t \leq 0.5 \\ w(2t - 0.5) & : t \geq 0.5 \end{cases}$$

²Es gibt noch eine weitere Klasse von topologischen Räumen, die ebenfalls (gekrümmte) Alexandroffräume heißen, worunter jedoch etwas vollständig anderes verstanden wird. Von diesen Räumen ist hier *nicht* die Rede.

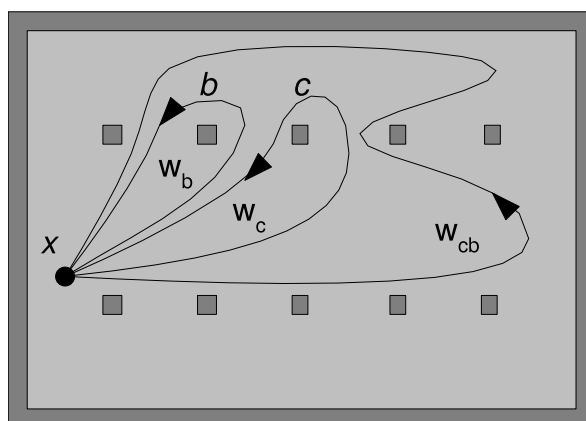


Abbildung 2.7: Die Fundamentalgruppe einer Säulenhalle. Für jede Säule s nehmen wir einen Rundweg w_s um diese Säule herum. Der Rundweg $w_c * w_b$ mit Zwischenstopp in x lässt sich dann in den großen Rundweg w_{cb} um beide Säulen deformieren, nicht jedoch der Rundweg $w_b * w_c$. Es gilt also $[w_c] \cdot [w_b] = [w_{cb}]$ jedoch auch $[w_b] \cdot [w_c] \neq [w_{cb}]$.

Dies ist eine algebraische Verknüpfung von Rundwegen und bei geeigneter Wahl von Äquivalenzrelationen in $\text{Weg}(\underline{X}, x)$ ergibt dies eine Verknüpfung $[v] \cdot [w] := [v * w]$ auf den Äquivalenzklassen $\text{Weg}(\underline{X}, x)/\sim$ und gewisse topologische Eigenschaften von \underline{X} übertragen sich dann auf algebraische Eigenschaften in $\text{Weg}(\underline{X}, x)/\sim$. Üblicherweise wird als Äquivalenzrelation die sogenannte Homotopieäquivalenz verwendet: Zwei Wege v und w sind homotop, wenn es eine Homotopie zwischen ihnen gibt, das ist ein „Weg von Wegen“ $h : [0, 1] \rightarrow \text{Weg}(\underline{X}, x)$, so dass für alle $i \in [0, 1]$ gilt: $h(0) = v$ und $h(1) = w$ und die Abbildung $H : [0, 1] \times [0, 1] \rightarrow X, (i, t) \mapsto H(i, t) := (h(i))(t)$ ist stetig. Dann gibt es ein neutrales Element der Verknüpfung (die zur konstanten Abbildung $c_x : [0, 1] \rightarrow \underline{X}$ homotopen Wege $[c_x]$) und zu jeder Wegklasse $[w]$ ein Inverses, die Klasse der Rückwege $[w^-]$ mit $w^-(t) := w(1 - t)$. Zudem ist es bei drei gegebenen Wegklassen $[u]$, $[v]$ und $[w]$ unerheblich, ob ich sie in der Form $([u] \cdot [v]) \cdot [w]$ oder $[u] \cdot ([v] \cdot [w])$ verknüpfe. Es entsteht dabei jedesmal die selbe Wegklasse, oder wie man sagt, die Verknüpfung ist assoziativ. Eine assoziative Verknüpfung mit neutralem und inverselem Element bildet nun eine sogenannte Gruppe. Die Gruppe der homotopen Rundwege heißt (erste) Fundamentalgruppe $\pi_1(\underline{X}, x)$.

Anschaulich kann man sich homotope Wege wie folgt vorstellen: Man nehme einen roten Faden aus einem ideal elastischen dehnbaren Material R (wie „Rubber“) fixiere ein Ende am Basispunkt x und gehe dann einen Rundweg r durch den Raum X , wobei man den roten Faden am anderen Ende fest halte und hinter sich her ziehe. Hat man x wieder erreicht, dann befestige man auch das andere Ende des Fadens an x . Danach nehme man einen grünen Faden G , der ebenfalls elastisch ist (z.B. Gummi), und verfare damit ebenso entlang eines Weges g . Dann sind die Wege g und r homotop, wenn der rote Faden durch Dehnen und Verschieben innerhalb des Raums so verzogen werden kann, dass er genau auf dem grünen Faden liegt (wobei die Richtungen übereinstimmen

müssen). Die Verknüpfung der Rundwege $r * g$ besteht darin, das Ende des roten Fadens an den Anfang des grünen Fadens zu befestigen, um einen rot-grünen Faden zu erhalten. $g * r$ hingegen erzeugt einen grün-roten Faden. Abbildung 2.7 soll dies an einem architekturnahen Beispiel veranschaulichen.

Wir wollen hier aber einen anderen Ansatz der algebraischen Topologie diskutieren: Lineare Algebra zur Untersuchung topologischer Räume. Ein Unterschied zur Fundamentalgruppe ist der, dass die Verknüpfungen stets kommutativ (bzw. abelsch) sind. Wenn wir beim obigen Beispiel bleiben, dann heißt das, dass der rot-grüne und der grün-rote Weg ebenfalls als äquivalent betrachtet werden [Sti80, 5.1.3].

2.6.1 Moduln und Lineare Abbildungen

Lineare Algebra ist die Theorie der Moduln (sprich: „Modul“ mit betontem „o“). Diese werden in der algebraischen Topologie üblicherweise eingesetzt. Moduln sind eine Verallgemeinerung von Vektorräumen. In Moduln ist die Multiplikation mit einem Skalar definiert, diese muss aber nicht invertierbar sein. Bei Vektorräumen wird diese Invertierbarkeit zusätzlich gefordert. Ein Beispiel für einen Modul wäre das Rechnen mit Geldbeträgen: Jeder Eurobetrag kann verdreifacht werden aber ein Euro kann nicht gleich auf drei Personen aufgeteilt werden, da es keine Drittelcent gibt.

In dieser Arbeit beschränken wir uns auf \mathbb{Z} -Moduln, der Skalar ist also \mathbb{Z} :

Definition 2.23 (\mathbb{Z} -Modul). Ein \mathbb{Z} -Modul ist eine Menge M mit einer Addition $+$ und einer Multiplikation $\cdot : \mathbb{Z} \times M \rightarrow M$ mit folgenden Eigenschaften:

- **Assoziativität der Addition:** $(u + v) + w = u + (v + w)$ für alle $u, v, w \in M$.
- **Kommutativität der Addition:** $u + v = v + u$ für alle $u, v \in M$.
- **Neutrales Element der Addition:** Es gibt ein $0 \in M$, so dass für alle $u \in M$ gilt: $u + 0 = u$
- **Inverses Element der Addition:** Zu jedem $u \in M$ existiert ein $u' \in M$ mit $u + u' = 0$.
- **linkes Distributivgesetz:** $(a + b) \cdot u = (a \cdot u) + (b \cdot u)$ für alle $a, b \in \mathbb{Z}$ und $u \in M$.
- **rechtes Distributivgesetz:** $a \cdot (u + v) = (a \cdot u) + (a \cdot v)$ für alle $a \in \mathbb{Z}$ und $u, v \in M$.
- **Assoziativität der Multiplikation:** $(a \cdot b) \cdot u = a \cdot (b \cdot u)$ für alle $a, b \in \mathbb{Z}$ und $u \in M$.
- **Neutrales Element der Multiplikation:** Für alle $u \in M$ gilt: $1 \cdot u = u$.

Verwenden wir in der obigen Definition anstelle von \mathbb{Z} die reellen Zahlen \mathbb{R} , dann erhalten wir einen reellen Vektorraum. Bei diesem ist zur Skalarmultiplikation auch eine Division durch Skalare definiert, denn für jede reelle Zahl $r \neq 0$ existiert eine Zahl r^{-1} , so dass für jeden Vektor v gilt: $r^{-1} \cdot (r \cdot v) = v$.

Die Multiplikation mit r^{-1} macht eine Multiplikation mit r rückgängig. Auf Vektorräume werde wir bei den Überlegungen zur Geometrie wieder eingehen.

Für manche \mathbb{Z} -Moduln gibt es genauso wie bei Vektorräumen Basis, Koordinaten und Dimension:

Definition 2.24 (Basis, Koordinaten, Dimension, freier \mathbb{Z} -Modul). Sei M ein \mathbb{Z} -Modul. Eine Teilmenge $B \subseteq M$ heißt Erzeugendensystem von M , wenn jedes $m \in M$ als Summe $\sum_{b \in B} \lambda_b \cdot b$ darstellbar ist. Ein Erzeugendensystem, bei dem diese Darstellung immer eindeutig ist, heißt Basis von M , und die λ_b heißen dann Koordinaten von m bezüglich B . Ist B eine endliche Basis, dann heißt die Anzahl $\dim(M) := |B|$ der Elemente in B auch Dimension von M . Ein \mathbb{Z} -Modul, der eine solche Basis, besitzt heißt auch freier \mathbb{Z} -Modul.

Wir wollen uns auf freie \mathbb{Z} -Moduln beschränken. Sprechen wir also in der Folge von \mathbb{Z} -Moduln, dann meinen wir stets die freien \mathbb{Z} -Moduln. Wenn wir eine endliche Basis B haben und die Basiselemente in B in der Form $B := \{b_1, \dots, b_n\}$ durchnummerieren, dann können wir $m \in M$ durch Angabe seiner Koordinaten waagrecht als n -Tupel $m = (\lambda_{b_1}, \dots, \lambda_{b_n})$ oder auch als senkrechte Spalte in der Form

$$m = \begin{pmatrix} \lambda_{b_1} \\ \vdots \\ \lambda_{b_n} \end{pmatrix}$$

schreiben.

Jedem Basiselement b_i entspricht dann das Tupel

$$b_i = (0, \dots, 0, \underset{\substack{\uparrow \\ \text{Stelle } i}}{1}, 0, \dots, 0).$$

„Zeilen“- und „Spalten“-Vektoren sind also nur spezielle Schreibweisen für ein allgemeineres Konzept. Auch eine uns eventuell vorliegende Wirtshausrechnung kann als Element eines \mathbb{Z} -Moduls betrachtet werden. Die Basiselemente entsprechen den Positionen in der Speisekarte und die „Bier-Koordinate“ unserer Rechnung könnte 2 – die Anzahl der berechneten Biere – betragen. Speisekartenpositionen, die auf der Rechnung nicht auftauchen (z.B. Kutteln) haben dann Koordinate 0.

Beispiel 2.13 (Beispiele für \mathbb{Z} -Moduln).

- \mathbb{Z} ist selbst ein \mathbb{Z} -Modul. Die Mengen $\{1\}$ und $\{-1\}$ sind jeweils eine Basis von \mathbb{Z} . Also hat \mathbb{Z} Dimension 1.
- Die Menge $\{0\}$ ist ein \mathbb{Z} -Modul mit Basis \emptyset und Dimension 0.
- $\mathbb{Z} \times \mathbb{Z}$ ist \mathbb{Z} -Modul mit Basen $\{(0, 1), (1, 0)\}$ oder $\{(1, 1), (-1, 1)\}$, u.s.w. und mit Dimension 2.
- Ist U Teilmenge eines Moduls M , dann heißt diese Untermodul von M , wenn die Verknüpfungen in M auch in U gelten. Das heißt für alle $u, v \in U$ ist $u + v$ ebenfalls in U und für jede ganze Zahl $a \in \mathbb{Z}$ ist $a \cdot u$ in U . Für einen Untermodul wollen wir hier $U \leq M$ statt $U \subseteq M$ schreiben.

- Mit einem Untermodul kann man auch einen Quotienten bilden: Sei U Untermodul eines Moduls M . Für ein beliebiges Element $m \in M$ heißt dann $m + U := \{m + u \mid u \in U\}$ eine Nebenklasse von U . Die Menge aller Nebenklassen von U ist eine Zerlegung von M , bezeichnet mit M/U (gesprochen „ M modulo U “), und heißt Faktormodul. Dies ist ein \mathbb{Z} -Modul, denn für Nebenklassen $a + U$ und $b + U$ in M/U ist die Addition mit $(a + U) + (b + U) := (a + b) + U$ und die skalare Multiplikation $a \cdot (b + U) := (a \cdot b) + U$ definiert. Abbildung 2.8 veranschaulicht dies.
- Für eine beliebige Menge B ist die Menge \mathbb{Z}^B der Abbildungen $f : B \rightarrow \mathbb{Z}$ ein Modul. Dieser heißt freier Modul über B . Ist B endlich, dann ist $|B|$ die Dimension von \mathbb{Z}^B . Für $B = \emptyset$ setzen wir $\mathbb{Z}^\emptyset = \{0\}$. Dieses Nullelement ist dann die leere Abbildung $0 : \emptyset \rightarrow \mathbb{Z}$. Ist $B = \{1, \dots, n\} \subseteq \mathbb{N}$, dann schreiben wir kurz \mathbb{Z}^n statt $\mathbb{Z}^{\{1, \dots, n\}}$.

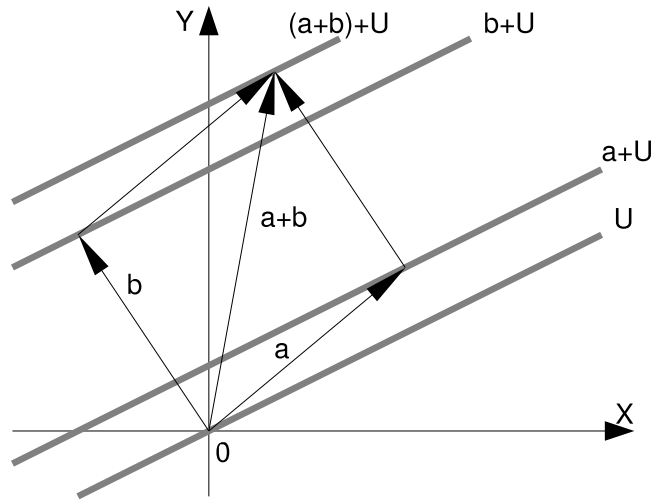


Abbildung 2.8: Ein Modul (Vektorraum) von Dimension 2 mit Untermodul (Untervektorraum) U von Dimension 1. Die Nebenklassen $a + U$ und $b + U$ sind dann ebenfalls Vektoren mit Addition $(a + U) + (b + U) = (a + b) + U$.

Ein \mathbb{Z} -Modul kann mehrere Basen haben, jedoch ist die Anzahl der Elemente jeder Basis eines Moduls stets gleich, die Dimension eines Moduls ist also wohldefiniert.

Zu einer beliebigen Menge M lässt sich nun der freie Modul $\mathbb{Z}M$ wie folgt definieren: Zu jedem Element $a \in M$ definiere man die Abbildung

$$\chi_a : M \rightarrow \mathbb{Z}, m \mapsto \chi_a(m) := \begin{cases} 1 & : m = a \\ 0 & : m \neq a \end{cases}$$

Diese χ_a bilden mit Addition

$$f + g : M \rightarrow \mathbb{Z}, m \mapsto (f + g)(m) := f(m) + g(m)$$

und mit Multiplikation mit $a \in \mathbb{Z}$

$$a \cdot f : M \rightarrow \mathbb{Z}, m \mapsto (a \cdot f)(m) := a \cdot f(m),$$

eine Basis für die Menge der Abbildungen $\text{Abb}(M, \mathbb{Z})$, die mit $\mathbb{Z}M$ bezeichnet wird und *freier \mathbb{Z} -Modul* oder *freie abelsche³ Gruppe* über M heißt. Damit können wir nun für beliebigen Elemente einer Menge, wie etwa bei der Menge $M = \{\text{Apfel, Birne, Tomate}\}$, auch formale Summen schreiben wie

$$\text{Apfel} + \text{Apfel} + \text{Birne} - 5 \text{ Apfel} = \text{Birne} - 3 \text{ Apfel} + 0 \text{ Tomate},$$

wobei hier „Apfel“ als Kurzschreibweise für die Abbildung $\chi_{\text{Apfel}} : M \rightarrow \mathbb{Z}$ steht. Eine derartige formale Summe heißt auch *Linearkombination* oder *Kette*. Man beachte, dass M sehr groß sein kann, wenn man z.B. mit M die Menge aller Räume eines Bauwerks meint.

Auch für Moduln gibt es – ganz analog zu den stetigen Abbildungen für die topologischen Räume – strukturerhaltende Abbildungen:

Definition 2.25 (Lineare Abbildungen, Bild und Kern). *Seien A und B \mathbb{Z} -Moduln. Ein Abbildung $f : A \rightarrow B$ heißt linear, wenn für alle $a, b \in A$ und $z \in \mathbb{Z}$ gilt:*

$$\begin{aligned} f(a + b) &= f(a) + f(b) \\ z \cdot f(a) &= f(z \cdot a). \end{aligned}$$

Ist f linear, dann heißen

$$\begin{aligned} \text{Bild}(f) &:= \{f(a) \mid a \in A\} \\ \text{Kern}(f) &:= \{a \mid f(a) = 0\} \end{aligned}$$

Bild und Kern von f .

Die Leserin kann sich selbst davon überzeugen, dass $\text{Kern}(f)$ und $\text{Bild}(f)$ wiederum \mathbb{Z} -Moduln sind, indem sie die 8 Eigenschaften in der Definition überprüft.

Wir wollen nun einige Eigenschaften von linearen Abbildungen festhalten:

Satz 2.8 (Eigenschaften von linearen Abbildungen). *Seien A, B, C \mathbb{Z} -Moduln und $f : A \rightarrow B$ und $g : B \rightarrow C$ lineare Abbildungen. Dann gilt stets:*

- *Die identische Abbildung $\text{id}_A : A \rightarrow A$ ist linear, denn $\text{id}_A(a+b) = a+b = \text{id}_A(a) + \text{id}_A(b)$.*
- *Die Komposition $g \circ f : A \rightarrow C$ ist linear, denn $(g \circ f)(a+b) = g(f(a+b)) = g(f(a) + f(b)) = g(f(a)) + g(f(b)) = (g \circ f)(a) + (g \circ f)(b)$.*
- *$f \circ \text{id}_A = f = \text{id}_B \circ f$ und $(g \circ f) \circ h = g \circ (f \circ h)$.*

Wir haben es also wieder mit Morphismen zu tun. Diese können wir nun verwenden, um Moduln zu vergleichen, wie wir es mit den topologischen Räumen bereits getan haben.

³eine Gruppe heißt abelsch, wenn die Verknüpfung kommutativ ist, wenn also immer $a+b = b+a$ gilt.

Definition 2.26 (Isomorphismus, isomorph). Seien X und Y \mathbb{Z} -Moduln. Eine lineare Abbildung $f : X \rightarrow Y$ heißt Isomorphismus, wenn es eine lineare Abbildung $g : Y \rightarrow X$ zurück gibt, so dass $g \circ f = \text{id}_X$ und $f \circ g = \text{id}_Y$. Zwei \mathbb{Z} -Moduln X und Y heißen isomorph, wenn es einen Isomorphismus $f : X \rightarrow Y$ gibt.

Es ist allgemein bekannt, dass zwei (freie) \mathbb{Z} -Moduln von endlicher Dimension genau dann isomorph sind, wenn sie von gleicher Dimension sind.

Der „Trick“ der algebraischen Topologie ist nun, zunächst eine Zerlegung, ähnlich der architektonischen Zerlegung in unserem obigen Beispiel zur Quotiententopologie, vorzunehmen. Dem so entstandenen Quotientenraum werden dann mehrere \mathbb{Z} -Moduln zugeordnet — jeweils einer pro Dimensionszahl der Elemente — und aus der Topologie ergeben sich lineare Abbildungen, die diese Moduln miteinander verbinden. Den Dimensionszahlen der entsprechenden Moduln entsprechen dann topologische Eigenschaften wie etwa Anzahl von Löchern und Hohlräumen.

2.6.2 Lineare Algebra an einem Graphen

Wir wollen zuerst am Beispiel von Graphen demonstrieren, wie topologische Räume mit den Mitteln der linearen Algebra untersucht werden, bevor wir das Konzept in seiner allgemeinen Form darstellen.

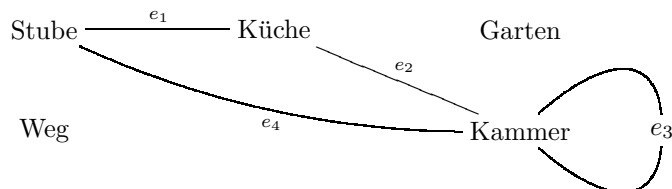
Ungerichtete Graphen. Legen wir zunächst fest, was wir hier unter einem Graphen verstehen:

Definition 2.27 (ungerichteter Graph). Ein ungerichteter Graph ist ein 3-Tupel $G = (K, N, d)$ mit einer Menge K und einer zu K disjunkten Menge N (also $K \cap N = \emptyset$) und einer Abbildung

$$d : K \rightarrow \{\{a, b\} \mid a, b \in N\}.$$

Die Menge K heißt die Kanten von G und die Menge N die Knoten von G . Für $\{a, b\} = d(k)$ sagen wir auch, „die Kante k verbindet Knoten a mit Knoten b “.

Ein Beispiel für einen solchen Graphen sind die Wegebeziehungen in einem Gebäude, wobei jeder Raumübergang (Türe, Durchgang etc.) eine Kante und jeder Raum ein Knoten ist. Für einen solchen Raumübergang t ist dann $d(t)$ die Menge der Räume, die „wechseln“ werden, wenn ich durch t trete. Die Abbildung d gibt offenbar wieder, auf welche Art die Räume von den Durchgängen verbunden sind. Folgendes Beispiel zeigt einen ungerichteten Graphen:



In diesem Beispiel gilt $d(e_1) = \{\text{Küche}, \text{Stube}\}$, also e_1 verbindet Küche und Stube.

Graphen sind ein nützliches Hilfsmittel für die architektonische Planung wie z.B. JOEDICKE über deren Einsatz zur Analyse von Raumprogrammen sagt:

„In einem solchen Stadium interessiert zunächst nicht die Größe der Räume, sondern allein ihre Zuordnung. Zur Analyse derartiger Raumbeziehungen ist eine Graphendarstellung nützlich.“
[Joe76, S. 45]

Eine ausführliche Diskussion der Graphentheorie für die Architektur findet sich bei MARCH und STEADMAN [MS74, Kap. 10].

Morphismen von Graphen $f : (K, N, d) \rightarrow (E, V, b)$ sind Paare $f = (f_K, f_N)$ von Abbildungen $f_K : K \rightarrow E$ und $f_N : N \rightarrow V$, so dass für alle Kanten $k \in K$ gilt: $f_N[d(k)] = b[f_K(k)]$. Das heißt, wenn eine Kante k die Knoten a und b verbindet, also $d(k) = \{a, b\}$, dann verbindet die Bildkante $f_K(k)$ die Bildpunkte $\{f_N(a), f_N(b)\} = f_K[\{a, b\}] = f_K[d(k)]$. Der Einheitsmorphismus $\text{id}_{(K, N, d)}$ ist natürlich das Paar der Identitäten $(\text{id}_K, \text{id}_N)$. Damit ist auch klar, wie Graphen-Isomorphismus definiert ist. Es wurde bereits erwähnt, dass es sehr schwer sein kann, zu entscheiden, ob ein solcher Isomorphismus für zwei gegebene endliche Graphen überhaupt existiert.

Einem ungerichteten Graphen kann nun ein topologischer Raum $(K \cup N, \mathcal{T}(d))$ zugeordnet werden. Das ist derjenige topologische Raum, in dem eine Kante k genau dann die Knoten a und b verbindet, wenn diese Knoten Berührungspunkte der Menge $\{k\}$ sind.

Dann ergibt sich folgende Topologie:

$$\mathcal{T}(d) := \{A \cup B \mid A \subseteq K, B \subseteq N, \text{ für alle } k \in K \text{ gilt: } d(k) \cap B \neq \emptyset \Rightarrow k \in A\}$$

Diese etwas umständliche Formel besagt Folgendes: Eine Menge A ist genau dann offen, wenn für jeden Knoten n in A auch alle Kanten in A sind, die n mit irgend einem anderen Knoten verbinden. Zwei derartige Räume sind genau dann homöomorph, wenn die zugrunde liegenden Graphen isomorph sind. Auf den Beweis verzichten wir hier, da dies im weiteren Verlauf der Arbeit für allgemeine Fälle noch gezeigt wird.

Für jede Kante k gilt dann $\text{fr}\{k\} = d(k)$ und d ist somit der auf einzelne Kanten eingeschränkte Randoperator des topologischen Raums. Also: $d = \text{fr}|_{\{\{k\} \mid k \in K\}}$, und jedes Element x dieses Raums hat eine kleinste offene Umgebung $\text{St}(x)$, den *Stern* um x — eine charakteristische Eigenschaft von alexandroffschen Räumen: Ist nämlich der Schnitt jeder beliebigen Menge von offenen Mengen selbst eine offene Menge, dann ist auch der Schnitt all jener offenen Mengen, die x enthalten eine offene Menge. Genau diese Schnittmenge ist der Stern um x . Der Stern um eine Kante k ist diese Kante selbst als Menge $\{k\}$ und der Stern um einen Knoten n ist diejenige Menge, die aus all den Kanten besteht, die diesen Knoten irgendwohin verbinden und der Knoten selbst:

$$\text{St}(n) = \{n\} \cup \{k \in K \mid n \in d(k)\}.$$

In unserem Beispielgraph gilt:

$$\text{St}(\text{Küche}) = \{\text{Küche}, e_1, e_2\}.$$

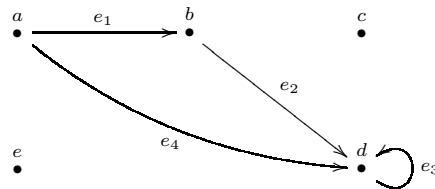
Gerichtete Graphen. Bei einem gerichteten Graphen wird nicht nur angegeben, welche Knoten durch eine Kante verbunden sind, sondern es wird zudem noch eine Richtung dieser Verbindung festgelegt.

Definition 2.28 (gerichteter Graph). Ein gerichteter Graph ist ein 3-Tupel $G = (K, N, f)$ mit einer Menge K (Kanten) und einer zu K disjunkten Menge N (Knoten) und einer Abbildung

$$f : K \rightarrow N \times N, k \mapsto f(k) := (o_k, t_k).$$

Für $(a, b) = d(k)$ sagen wir auch, „die Kante k geht von Knoten a zu Knoten b “.

Hier sind die Wegebeziehungen in Gebäuden sogar ein besseres Beispiel, da nun auch angegeben werden kann, in welche Richtung eine eventuell vorhandene Türe öffnet. Für Pendeltüren oder offene Durchgänge muss dann für jede Durchgangsrichtung je eine eigene Kante definiert sein. Hier unser obiges Beispiel als gerichteter Graph mit geänderten Bezeichnungen:



In diesem Beispiel gilt $f(e_1) = (a, b)$, also e_1 geht von a nach b . Der Knoten a hieß früher Küche, und b hieß Stube.

Morphismen für gerichtete Graphen $(K, N, f) \rightarrow (E, V, g)$ sind — ganz analog zu den ungerichteten Graphen — Paare (f_K, f_N) von Abbildungen $f_K : K \rightarrow E$ und $f_N : N \rightarrow V$ mit $g \circ f_K = (f_N \times f_N) \circ f$. Die Richtung muss also respektiert werden. Einheitsmorphismus und Isomorphismus definieren wir wie gewohnt.

Für jeden gerichteten Graphen $G = (K, N, f)$ existiert ein zugeordneter ungerichteter Graph $U = (K, N, d)$. Geht in G eine Kante k von a nach b dann sagen wir einfach: in U verbindet k die Knoten a und b . Wir definieren also

$$d(k) = \{\pi_1(f(k)), \pi_2(f(k))\}.$$

Umgekehrt ist jeder ungerichtete Graph (K, N, d) zumindest einem solchen gerichteten Graphen (K, N, f) , in dieser Weise zugeordnet. Diesen finden wir, indem wir einfach für jede Kante k mit $d(k) = \{a, b\}$ eine Orientierung in der Form $f(k) := (a, b)$ willkürlich festlegen.

Mit ungerichteten Graphen haben wir nun topologische Räume und wollen diese mit linearer Algebra untersuchen. Dazu wählen wir uns für unseren ungerichteten Graphen einen zugeordneten gerichteten Graphen beliebig aus. Dann definieren wir die zwei Moduln $C_1 := \mathbb{Z}K$ über den Kanten und $C_0 := \mathbb{Z}N$ über den Knoten und für die Verbindungen definieren wir die lineare Abbildung

$$\partial_f : C_1 \rightarrow C_0, k \mapsto \partial_f(k) := b - a, \text{ falls } f(e) = (a, b).$$

Im obigen gerichteten Beispielgraph ergibt sich die entsprechende lineare Abbildung $\partial_f : C_1 \rightarrow C_0$ wie folgt:

$$\begin{array}{lll} f(e_1) = (a, b) & \text{ergibt} & \partial_f(e_1) := b - a \\ f(e_2) = (b, d) & \text{ergibt} & \partial_f(e_2) := d - b \\ f(e_3) = (d, d) & \text{ergibt} & \partial_f(e_3) := d - d \\ f(e_4) = (a, d) & \text{ergibt} & \partial_f(e_4) := d - a. \end{array}$$

∂_f wird als *algebraischer Randoperator* bezeichnet. Der (algebraische) Rand einer Kante ist also die (formale) Differenz „Endpunkt minus Anfangspunkt“. Es dürfte nun auch klar sein, wieso wir bei der Diskussion der Datenstrukturen DIME und Winged-Edge die Richtungsumkehr einer Kante durch ein negatives Vorzeichen angezeigt haben. Da ∂_f linear ist gilt z.B. für die Kante e_1

$$\partial_f(-e_1) = -\partial_f(e_1) = -(b - a) = a - b.$$

Mit $-e_1$ erhalten wir also eine zu e_1 inverse Kante von b nach a .

Ketten von Kanten. Betrachten wir nun folgenden Rundweg oder Zyklus

$$a \xrightarrow{z} a := a \xrightarrow{+e_1} b \xrightarrow{+e_2} d \xrightarrow{-e_4} a$$

in diesem Graphen, der von a ausgehend e_1 und e_2 folgt und e_4 entgegengesetzt der Pfeilrichtung zurückkehrt. Das Vorzeichen zeigt uns an, ob das Wegstück entlang oder entgegen der Pfeilrichtung verläuft. Die formale Summe der Kanten dieses Rundwegs ist $z = e_1 + e_2 - e_4 \in C_1$. Dann gilt für den Rand von z :

$$\begin{aligned} \partial_f(z) &= \partial_f(e_1 + e_2 - e_4) \\ &= \partial_f(e_1) + \partial_f(e_2) - \partial_f(e_4) \\ &= (b - a) + (d - b) - (d - a) \\ &= b - a + d - b - d + a \\ &= b - b + a - a + d - d \\ &= 0 \end{aligned}$$

Es gilt allgemein $\partial_f(x) = 0$ genau dann, wenn x formale Summe der Wegstücke eines derartigen Zyklus ist. Wir erinnern uns an die Definition von $\text{Kern}(\partial_f)$: das sind eben diejenigen x mit $\partial_f(x) = 0$. Daher bezeichnen wir ein x mit $\partial_f(x) = 0$ als *Zykel* und bezeichnen $\text{Kern}(\partial_f)$ mit Z_1 , die Menge der 1-Zykeln im Graph.

Wir haben also zu topologischen Objekten „Rundwege im Graph“ ein algebraisches Objekt, die Menge Z_1 gefunden. Diese ist ebenfalls ein Modul und dessen Dimension ist genau die Anzahl von „Löchern“ und Schleifen im Graph. Eine Schleife ist eine Kante, die einen Knoten mit sich selbst verbindet.

Ketten von Knoten. Umgekehrt können wir sagen, eine Linearkombination von Knoten ist ein *Rand*, wenn es das Bild unter ∂_f einer Linearkombination von Kanten ist. So ist z.B. $a + b$ in unserem Beispiel kein Rand, denn sonst

würde es $x, y, w, z \in \mathbb{Z}$ geben mit:

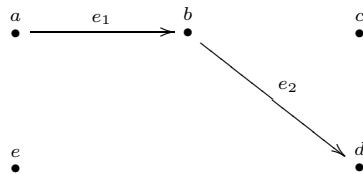
$$\begin{aligned}
 a + b &= \partial(x \cdot e_1 + y \cdot e_2 + w \cdot e_3 + z \cdot e_4) \\
 &= \partial(x \cdot e_1) + \partial(y \cdot e_2) + \partial(w \cdot e_3) + \partial(z \cdot e_4) \\
 &= x\partial(e_1) + y\partial(e_2) + w\partial(e_3) + z\partial(e_4) \\
 &= x(b - a) + y(d - b) + w(d - d) + z(d - a) \\
 &= xb - xa + yd - yb + zd - za \\
 &= (-x - z)a + (x - y)b + (y + z)d \\
 &= \lambda_a a + \lambda_b b + \lambda_d d
 \end{aligned}$$

Dann sind aber $\lambda_a = (-x - z)$, $\lambda_b = (x - y)$ und $\lambda_d = (y + z)$ Koordinaten von $a + b$ und es folgt:

$$\begin{aligned}
 \lambda_a &= (-x - z) &= 1 \\
 \lambda_b &= (x - y) &= 1 \\
 \lambda_d &= (y + z) &= 0 \\
 \lambda_a + \lambda_b &= (-x - z) + (x - y) \\
 &= -x + x - z - y \\
 &= -(y + z) \\
 &= -\lambda_d &= 0 \\
 \lambda_a + \lambda_b &= 1 + 1 \\
 &= 2 &= 0 \quad \text{Widerspruch, denn } 2 \neq 0
 \end{aligned}$$

Die Menge aller Ränder ist also $\text{Bild}(\partial_f)$ und die Koordinaten zweier verschiedener Knoten sind dann voneinander abhängig in $\text{Bild}(\partial_f)$, wenn sie durch eine Kante verbunden sind.

Beziehung zwischen Zykeln und Rändern. Wenn in einem Graphen mit Zykeln eine Kante entfernt wird, die einem Zykel angehört (z.B. die Kante e_4), dann hat sich die Anzahl der unabhängigen Zykeln um eins verringert, die Anzahl der zusammenhängenden Teile hingegen bleibt gleich. Dies kann schrittweise geschehen, bis keine Zykeln mehr vorhanden sind, falls wir Endlichkeit voraussetzen. Dann erhalten wir einen zykelfreien Graphen, bzw. einen sog. *Wald*. Ein *Wald* mit nur einem zusammenhängenden Teil heißt *Baum*. Die folgende Abbildung zeigt einen Wald mit drei Bäumen:



Das Bild des Randoperators (Menge der algebraischen Ränder) bezeichnen wir auch mit $B_0 := \text{Bild}(\partial_f)$ (B wie „Boundary“). Die Dimension von B_0 als \mathbb{Z} -Modul ist genau die Anzahl der Kanten des maximalen Teilwaldes unseres ursprünglichen Graphen und diese ist die Differenz der Knotenanzahl von der Anzahl der Bäume in diesem Wald.

Sei β_0 die Anzahl der Bäume eines Graphen, dann ist also

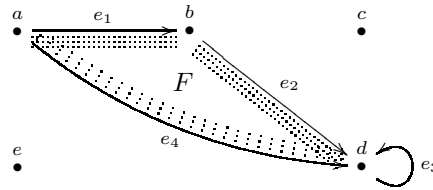
$$\dim Z_0 - \dim B_0 = \beta_0$$

In unserem Beispiel gilt $\beta_0 = 3$. Dies ist auch die Anzahl der Zusammenhangskomponenten unseres ursprünglichen ungerichteten Graphen — ebenfalls eine topologische Eigenschaft.

2.6.3 Komplexe

Komplexe verallgemeinern nun den Ansatz, Graphen mit linearer Algebra zu untersuchen, auf beliebige topologische Räume. Die Knoten eines Graphen wurden oben als 0-Zykeln aufgefasst. Wenn eine Kante zwei Knoten verbindet, dann können wir uns vorstellen, die Kante sei an die Knoten „angeheftet“. Nun sollen auch noch die Kanten durch Flächen verbunden werden. Dabei wollen wir jedoch fordern, dass die Kanten, an die eine Fläche angeheftet ist, einen Zykel bilden, also um die Fläche „herumlaufen“.

Wenn wir nun eine Fläche F in den Zykel $a \xrightarrow{e_4} d \xrightarrow{-e_2} b \xrightarrow{-e_1} a$ einfügen, dann erhalten wir:



Diese Anheftung einer Fläche an Randkanten lässt sich ebenfalls durch eine lineare Abbildung ausdrücken, wie z.B.:

$$\partial_2 : \mathbb{Z}\{F\} \rightarrow C_1, (\lambda_F \cdot F) \mapsto \lambda_F \cdot (e_4 - e_2 - e_1)$$

Bezeichnen wir den \mathbb{Z} -Modul über den Flächen mit $C_2 := \mathbb{Z}\{F\}$, dann haben wir die linearen Abbildungen

$$\begin{aligned} \partial_2 & : C_2 \rightarrow C_1 \\ \partial_1 & : C_1 \rightarrow C_0 \end{aligned}$$

oder auch als Diagramm:

$$C_2 \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0$$

Und wenn wir die beiden Randabbildungen verketteten erhalten wir:

$$\begin{aligned} \partial_1 \circ \partial_2(\lambda \cdot F) &= \lambda \cdot \partial_1 \circ \partial_2(F) \\ &= \lambda \cdot \partial_1(\partial_2(F)) \\ &= \lambda \cdot \partial_1(e_4 - e_2 - e_1) \\ &= \lambda \cdot (\partial_1(e_4) - \partial_1(e_2) - \partial_1(e_1)) \\ &= \lambda \cdot ((d - a) - (d - b) - (b - a)) \\ &= \lambda \cdot 0 \\ &= 0. \end{aligned}$$

Also ist $\partial_1 \circ \partial_2$ die konstante Nullabbildung. Diese Konstruktion lässt sich nun verallgemeinern:

Definition 2.29 ((algebraischer) Komplex). *Ein (endlicher algebraischer) Komplex ist eine Familie C_n, \dots, C_m von (freien) \mathbb{Z} -Moduln C_i für ganze Zahlen $m \leq i \leq n$ mit einer Familie $\partial_i : C_i \rightarrow C_{i-1}$ für $i = m+1, \dots, n$ von linearen Abbildungen, so dass jedes $\partial_{i-1} \circ \partial_i : C_i \rightarrow C_{i-2}$ für alle $i = m+2, \dots, n$ die Nullabbildung ist.*

Damit ist $\text{Bild}(\partial_{i+1}) \leq \text{Kern}(\partial_i) \leq C_i$ — alle Ränder sind Zykeln. Wir erinnern uns, dass $B_i := \text{Bild}(\partial_{i+1})$ und $Z_i := \text{Kern}(\partial_i)$ ebenfalls \mathbb{Z} -Moduln sind und dass durch $B_i \leq Z_i$ eine Zerlegung von Z_i , definiert ist: Ist $c_i \in Z_i$ ein Zykel, dann ist $c_i + B_i$ die Menge aller Summen $c_i + b_i$ dieses Zyklus mit irgendeinem Rand $b_i \in B_i$, also $c_i + B_i := \{c_i + b_i \mid b_i \in B_i\}$. Ist nun c_i aber selbst ein Rand (z.B. der Rahmen eines Fensters), dann ist ja $c_i \in B_i$ und damit $c_i + B_i \subseteq B_i$. Umgekehrt ist dann jeder Rand $b_i \in B_i$ auch in $c_i + B_i$, denn es gibt ja den zu c_i inversen Zykel $-c_i$, der ebenfalls in B_i ist. Dann ist aber auch $b_i - c_i$ in B_i und somit ist $c_i + (b_i - c_i)$ in $c_i + B_i$ aber $c_i + b_i - c_i = b_i$, also ist $b_i \in c_i + B_i$. Also gilt in diesem Fall $c_i + B_i = B_i$.

Ist c_i hingegen ein „Loch“, also ein Zykel, der nicht Rand eines anderen Elements ist (etwa ein Türrahmen mit Schwelle und ohne Türe), dann ist $c_i + \text{Bild}(\partial_i) \neq \text{Bild}(\partial_i)$. Damit haben wir die Menge

$$H_i := \{c_i + B_i \mid c_i \in Z_i\} := Z_i/B_i$$

die *i-te Homologiegruppe* des Komplexes.

Ist nun

$$C_n \xrightarrow{\partial_n} C_{n-1} \xrightarrow{\partial_{n-1}} \dots \xrightarrow{\partial_1} C_0$$

ein Komplex, dann ist

$$\{0\} \xrightarrow{c_0} C_n \xrightarrow{\partial_n} C_{n-1} \xrightarrow{\partial_{n-1}} \dots \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} \{0\}$$

mit den konstanten Nullabbildungen c_0 und ∂_0 ebenfalls ein Komplex.

Jeder Rand in einem Komplex ist zwar ein Zykel, aber nicht jeder Zykel ist notwendigerweise Rand eines Elements im Komplex. Auch die Zykeln, die keine Ränder sind, bilden ein Modul, dessen Dimension gleich der Anzahl der „Löcher“ im Komplex ist. Es gilt nun allgemein: Die Anzahl β_n dieser n -dimensionalen Löcher in einem Komplex ist gleich:

$$\beta_n = \dim Z_n - \dim B_n = \dim H_n$$

Dabei betrachten wir eine Zusammenhangskomponente als ein 0-dimensionales „Loch“, eine Schleife, die nicht durch Flächen ausgefüllt ist, als ein 1-dimensionales Loch und einen Hohlraum als ein 2-dimensionales Loch. Eine wichtige Erkenntnis der algebraischen Topologie ist, dass die alternierende Summe (also abwechselndes plus-minus-plus-minus etc.) der Löcherzahlen $\chi := \beta_0 - \beta_1 + \beta_2 - \beta_3 + \dots$ gleich der alternierenden Summe der Zahlen $\dim C_0 - \dim C_1 + \dim C_2 - \dim C_3 + \dots$ ist. Diese Zahl χ bezeichnet man als *Euler-Charakteristik* eines Komplexes — eine topologische Eigenschaft, die uns bei den Euler-Operatoren bereits begegnet ist. Für alle Datenstrukturen, die auf der Basis der algebraischen Topologie arbeiten, ist diese Eigenschaft eine wichtig Konsistenzregel [Män88].

Bezug zur Architektur. Wir wollen nun illustrieren, wie wir architektonische Räume in Komplexen wiederfinden. Für jedes C_n wählen wir uns als Basis E_n die n -dimensionalen Bauelemente: Volumen E_3 , Flächen E_2 , Kanten E_1 und Knotenpunkte E_0 . Die Randabbildungen ∂_n geben an, auf welche Weise Elemente in E_n mit Elementen in E_{n-1} , verbunden sind. Ist etwa $t \in E_2$ eine Türe wie in Abbildung 2.9 und sind $in, out \in E_3$ Innen- und Außenraum, die durch t verbunden werden, dann gilt: Die t -Koordinate von $\partial(in)$ ist gleich der negativen t -Koordinate von $\partial(out)$. Welches Vorzeichen dabei gewählt wird, ist beliebig.

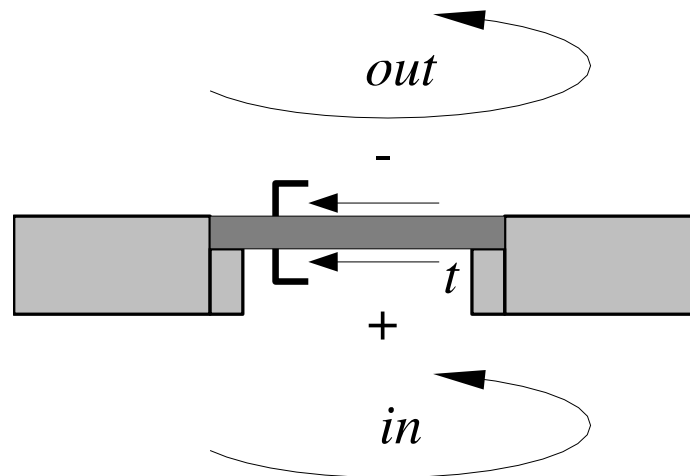


Abbildung 2.9: Die Türe t ist ein flächiges Bauelement, also $t \in E_2$. Es gilt $\partial(in) = \dots + t + \dots$ und $\partial(out) = \dots - t + \dots$.

Allerdings hat ein Komplex in dieser Form noch gewisse Nachteile. In einem algebraischen Komplex gehen nämlich manche topologische Eigenschaften verloren, wie wir später sehen werden.

Definition 2.30 (Komplexmorphismen). Sind $\dots C_i \xrightarrow{\partial_i} C_{i-1} \dots$ und $\dots K_i \xrightarrow{\delta_i} K_{i-1} \dots$ mit $i = a + 1, \dots, n$ Komplexe, dann heißt eine Familie $m_i : C_i \rightarrow K_i$ mit $i = a, \dots, n$ von linearen Abbildungen **Komplexmorphismus**, wenn für jedes $i = a + 1, \dots, n$ die Kompositionen $m_{i-1} \circ \partial_i$ und $\delta_i \circ m_i$ gleich sind.

Folgendes Diagramm veranschaulicht diese Regel: Die Abbildung, die wir erhalten, wenn wir dem Pfeil zuerst senkrecht und dann waagrecht folgen, muss die selbe sein, wie wenn zuerst waagrecht und dann senkrecht gegangen wird. Die Gleichheit dieser Kompositionen wird durch das „//“ im Inneren des Vierecks symbolisiert. Man sagt dann auch „das Diagramm kommutiert“.

$$\begin{array}{ccccccc}
\cdots & \xrightarrow{\partial_{i+1}} & C_i & \xrightarrow{\partial_i} & C_{i-1} & \xrightarrow{\partial_{i-1}} & \cdots \\
& & \downarrow m_i & & \downarrow m_{i-1} & & \\
& & & & \text{//} & & \\
& & & & & & \\
\cdots & \xrightarrow{\delta_{i+1}} & K_i & \xrightarrow{\delta_i} & K_{i-1} & \xrightarrow{\delta_{i-1}} & \cdots
\end{array}$$

Die Leserin kann sich davon überzeugen, dass Komplexmorphismen in der Tat Morphismen sind (Familie der identischen Abbildungen ist Komplexmorphismus, Komposition von Morphismen ist Morphismus, Identität ist neutrales Element bei Komposition etc.).

Es gibt zwar ein allgemeines Verfahren, jedem beliebigen topologischen Raum einen solchen Komplex zuzuordnen, indem man sogenannte singuläre Komplexe verwendet. Wir wollen uns jedoch hier nur mit topologischen Räumen befassen, bei denen dies auf sehr natürliche Weise möglich ist und die zudem eine naheliegende Entsprechung in der Architektur haben. Diese naheliegende Analogie hat die Mathematiker sogar dazu veranlasst, manche dieser Räume selbst als „Gebäude“ zu bezeichnen [Bro88].

2.6.4 CW-Komplexe

CW-Komplexe sind topologische Räume, die unserer oben erwähnten architektonischen Zerlegung sehr ähnlich sind. Gleichzeitig kann einem CW-Komplex in naheliegender Weise ein algebraischer Komplex zugeordnet werden. Sie sind hier wichtig, weil wir später behaupten werden, dass jeder architektonische Raum durch einen derartigen Komplex beschrieben werden kann.

Zuerst benötigen wir unsere Elemente, aus denen sich unser architektonischer Raum zusammensetzt:

Definition 2.31 (Zelle, Mannigfaltigkeit). *Eine n -Zelle ist ein topologischer Raum, der zum n -dimensionalen offenen Einheitswürfel $]0, 1[^n$ homöomorph ist. Eine n -Mannigfaltigkeit ist ein topologischer Raum (X, \mathcal{T}_X) , in dem es zu jedem Punkt $p \in X$ eine n -Zelle $C_n \subseteq X$ gibt, die eine offene Umgebung von p ist.*

Es ist auch üblich, für Zellen Homöomorphie mit \mathbb{R}^n oder mit der offenen Einheitskugel $B^n := B_1(0)$ zu fordern. Diese sind jedoch homöomorph zum Einheitswürfel, der unserer Vorstellung von Bauwerkselementen (Quader oder Rechtecke) näher kommt. Eine 0-Zelle ist ein einzelner Punkt, eine 1-Zelle eine Linie ohne ihrem Anfangs- und Endpunkt (und die sich nicht selbst schneidet), eine 2-Zelle ist ein Flächenstück (ohne dessen Randlinie) und eine 3-Zelle ein Volumen (ohne dessen Oberfläche). Die Dimension n ist aber nicht nach oben beschränkt.

Jede n -Zelle ist auch eine n -Mannigfaltigkeit. n -Mannigfaltigkeiten ähneln n -Zellen, dürfen aber Löcher haben. Ein Beispiel für eine 2-Mannigfaltigkeit wäre etwa die Oberfläche einer Lochfassade. Jeder Punkt darauf kann durch eine geeignet kleine „Briefmarke“ – eine 2-Zelle – überklebt werden. Die Randlinie eines Fensterlochs gehört bereits nicht mehr zu dieser Oberfläche sondern bildet eine 1-Mannigfaltigkeit.

Nun denken wir uns Architektur aus solchen Elementen zusammengesetzt:

Definition 2.32 (endlicher CW-Komplex). *Sei X ein topologischer Raum mit einer endlichen Zerlegung $\mathcal{X} = X/\sim$ und sei diese wiederum zerlegt in*

$\{\mathcal{X}_0, \dots, \mathcal{X}_n\}$. Dann setzen wir $\mathcal{X}^i := \mathcal{X}_0 \cup \dots \cup \mathcal{X}_i$ für $i = 0, \dots, n$ und $X^k := \bigcup_{i \leq k \leq n} \mathcal{X}^i$ für alle $k \in \mathbb{N}$. Dann heißt (X, \sim) endlicher CW-Komplex, wenn für jedes $e_i \in \mathcal{X}_i$ für $i = 0, \dots, n$ eine stetige Abbildung $\phi_i : [0, 1]^i \rightarrow X$ existiert, die folgende Eigenschaften hat:

- $\phi_i([0, 1]^i) = e_i$
- $\phi_i|_{[0, 1]^i} : [0, 1]^i \rightarrow e_i$ ist ein Homöomorphismus (in anderen Worten: e_i ist eine i -Zelle)

Die Zahl n ist die Dimension des CW-Komplexes und für alle $i = 0, \dots, n$ ist (X^i, \sim_i) (\sim_i ist die Einschränkung von \sim auf X^i) ebenfalls ein CW-Komplex, das i -Skelett von X .

Die Abbildungen ϕ_i heißen auch charakteristische Abbildungen oder Anheftungsabbildungen. Die Einschränkung auf Zellen wollen wir später fallen lassen, da z.B. Räume mit Stützen oder Fassaden mit Fensteröffnungen keine Zellen sind, denn sie haben Löcher. Üblicherweise fordert man noch, dass sich die Hülle einer Zelle aus nur endlich vielen Zellen zusammensetzt. Daher stammt auch das „C“ in CW-Komplex: C wie „closure-finite“⁴. In unserer Definition haben wir aber Endlichkeit generell vorausgesetzt und damit besteht die Hülle einer Zelle ebenfalls aus endlich vielen Zellen. Das „W“ steht für „weak topology“: Eine Menge $A \subseteq X$ ist genau dann abgeschlossen in der schwachen Topologie von (X, \sim) , wenn $A \cap \text{cl} e_i$ für alle Zellen e_i des Komplexes abgeschlossen in X ist.

Beispiele 2.14 (CW-Komplexe).

Würfel. Ein abgeschlossener Würfel $[0, 1]^3$ bildet mit folgender Zerlegung einen CW-Komplex

X_0 sind die 8 Eckpunkte des Würfels (0-Zellen)

X_1 sind die 12 Kanten des Würfels (1-Zellen)

X_2 sind die 6 Seitenflächen (2-Zellen)

X_3 ist das 1 Innere des Würfels (3-Zelle)

Die Dimensionen der entsprechenden \mathbb{Z} -Moduln seines algebraischen Komplexes sind dann

C_0 hat Dimension 8

C_1 hat Dimension 12

C_2 hat Dimension 6

C_3 hat Dimension 1

Damit haben wir 4 aufeinander folgende Moduln von Dimension ungleich 0 und folglich hat der Komplex selbst Dimension $3 = 4 - 1$. Die Charakteristik χ des abgeschlossenen Würfels ist $\chi = 8 - 12 + 6 - 1 = 1$. Die Bettizahlen sind

⁴Dass CW auch Teil der Initialien von John Henry Constantine Whitehead dem Erfinder der CW-Komplexe ist, hat sicher mit Zufall zu tun. Sie heißen ja auch nicht JHCW-Komplexe.

β_0 ist $8 - 7 = 1$, denn eine Basis der 0-Zykeln sind die 8 Eckpunkte: $\dim Z_0 = 8$. Die Ränder der Kanten werden von den 12 Kanten erzeugt und damit gilt $\dim B_0 \leq 12$. Allerdings bilden 7 Kanten davon einen Baum, mit dem man jeden Eckpunkt erreicht und damit ist $\dim B_0 = 7$. Also ist $\dim H_0 = \dim Z_0/B_0 = \dim Z_0 - \dim B_0 = 8 - 7$

β_1 ist $5 - 5 = 0$. Man sieht zwar auf den ersten Blick sechs Zykeln, die von den Kanten des Würfels gebildet werden. Jedoch kann immer einer dieser Zykeln als Kombination der anderen fünf erzeugt werden. Es ist $\dim Z_1 = 5$. Die Ränder der Flächen werden von den 6 Flächen erzeugt. Allerdings erzeugen bereits 5 Flächen alle 6 Ränder: $\dim B_1 = 5$.

β_2 ist $1 - 1 = 0$. $\dim B_2 = 1$, der Rand des einen und einzigen Volumens ist ungleich 0. $\dim Z_2 = 1$ die Flächen, bilden nur einen einzigen Hohlraum.

β_3 ist 0, denn da es keine vierdimensionale Zelle, und damit auch keinen Rand einer solchen gibt ist $\dim B_3 = 0$. Es ist auch $\dim Z_3 = 0$, denn die eine Volumenzelle im Inneren des Würfels bildet keinen Zykel.

Anhand der Bettizahlen ergibt sich also die selbe Charakteristik $\chi = \beta_0 - \beta_1 + \beta_2 - \beta_3 = 1 - 0 + 0 - 0 = 1$.

Viele Würfel. Wenn wir zwei gleich große Würfel an einer gemeinsamen Seitenfläche aufeinander setzen, haben wir 12 Eckpunkte, 20 Kanten, 11 Flächen (denn sie haben eine Fläche gemeinsam) und 2 Würfelinnere. Auch hier ist $\chi = 12 - 20 + 11 - 2 = 1$. Die LeserIn kann dies anhand von Abbildung 2.10 überprüfen. Ordnen wir 9 Würfel regelmäßig zu einem Quadrat erhalten wir ebenfalls $\chi = 1$. Nehmen wir aber den mittleren Würfel heraus, dann haben wir ein Loch und es gilt $\chi = 32 - 64 + 40 - 8 = 0$, die Charakteristik des Volltorus.

Graph. Ein CW-Komplex von Dimension 1 ist auch ein ungerichteter Graph mit

$$(\mathcal{X}_1, \mathcal{X}_0, \text{fr}|_{\mathcal{X}_1}).$$

Landkarte. Den „Polygontopologien“ in GIS-Systemen entsprechen CWKomplexe, wenn keine En- oder Eckklaven in den Polygonflächen liegen. Die 2-Zellen sind die Länder (bzw. Regionen, Parzellen), Die 1-Zellen die entsprechenden Grenzlinien und die 0-Zellen die Knoten an denen Grenzlinien aufeinander treffen.

Auch für CW-Komplexe haben wir strukturerhaltende Abbildungen, also Morphismen:

Definition 2.33 (zelluläre Abbildung). Es seien (X, \sim_X) und (Y, \sim_Y) CW-Komplexe. Eine stetige Abbildung $f : X \rightarrow Y$ heißt zellulär, wenn für alle $n \in \mathbb{N}$ gilt: $f(X^n) \subseteq Y^n$.

Insbesondere ist eine stetige Abbildung $f : X \rightarrow Y$, die jede n -Zelle in (X, \sim_X) auf eine m -Zelle in (Y, \sim_Y) mit $m \leq n$ abbildet, eine zelluläre Abbildung. Umgekehrt hat aber nicht jede zelluläre Abbildung diese Eigenschaft. Wir könnten diese Eigenschaft eventuell als „zellentreu“ bezeichnen. Diese zellentreuen Abbildungen sind ebenfalls Morphismen und sie haben zusätzlich die

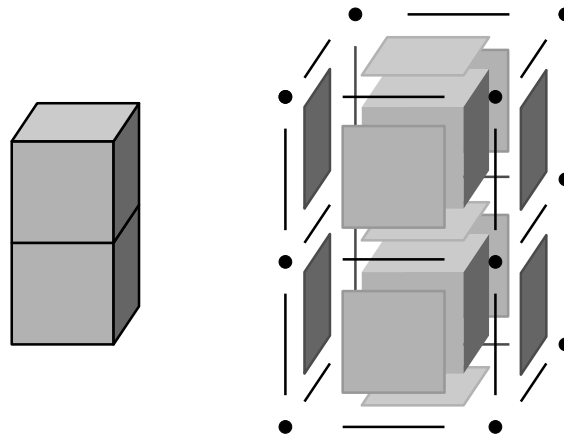


Abbildung 2.10: Links: Ein CW-Komplex bestehend aus zwei Würfeln. Rechts: Die einzelnen Zellen dieses CW-Komplexes.

Eigenschaft, dass durch eine zellentreue Abbildung stets auch eine stetige Abbildung

$$F : X/\sim_X \rightarrow Y/\sim_Y, A \mapsto F(A) := f[A]$$

auf den Quotientenräumen definiert ist.

2.7 Kategorien und Funktoren

Wir haben nun oft von den Morphismen als den strukturerhaltenden Abbildungen gehört. Aus der algebraischen Topologie stammt nun eine Begriffswiese, die das Konzept der strukturerhaltenden Abbildungen verallgemeinert und somit den Einsatz von Mitteln aus mehreren verschiedenen mathematischen Disziplinen systematisiert. Kategorien verallgemeinern den Gegenstand einer bestimmten mathematischen Disziplin (z.B. Topologie). Funktoren verallgemeinern den Ansatz, diesen Gegenstand mit den Mitteln einer anderen Disziplin (z.B. Algebra) zu betrachten.

Definition 2.34 (Kategorie). Eine Kategorie \mathcal{C} besteht aus einer Klasse von Objekten, die wir mit $\text{Ob}_{\mathcal{C}}$ bezeichnen, und für je zwei Objekte A, B aus $\text{Ob}_{\mathcal{C}}$ einer Menge $\text{Mor}_{\mathcal{C}}(A, B)$, den Morphismen von A nach B , so dass es für je zwei Morphismen $f \in \text{Mor}_{\mathcal{C}}(A, B)$ und $g \in \text{Mor}_{\mathcal{C}}(B, C)$ einen Morphismus $g \circ f \in \text{Mor}_{\mathcal{C}}(A, C)$ gibt, der folgenden Eigenschaften hat:

- Ist $h \in \text{Mor}_{\mathcal{C}}(C, D)$, dann ist $(h \circ g) \circ f = h \circ (g \circ f)$.
- Für jedes Objekt B existiert ein Einheitsmorphismus $\text{id}_B \in \text{Mor}_{\mathcal{C}}(B, B)$ mit $\text{id}_B \circ f = f$ und $g = g \circ \text{id}_B$.

Ist $\text{Ob}_{\mathcal{C}}$ eine Menge, dann heißt \mathcal{C} kleine Kategorie.

Wie anfangs erwähnt ist eine Klasse eine Verallgemeinerung des Begriffs Menge. Manches aus der Mengenlehre ist auch mit Klassen erlaubt, anderes nicht. Wir haben z.B. gesagt, dass Homöomorphie eine Äquivalenzrelation ist. Damit haben wir stillschweigend vorausgesetzt, dass es Äquivalenzrelationen in Klassen gibt, denn die topologischen Räume bilden eine Klasse und keine Menge. In jedem Fall kann es keine Klasse der Klassen geben, also auch nicht die Kategorie der Kategorien.

Für jedes Objekt X einer Kategorie ist sein Einheitsmorphismus id_X eindeutig bestimmt, denn seien id_1 und id_2 zwei beliebige Einheitsmorphisme für X , dann ist $\text{id}_1 \circ \text{id}_2 = \text{id}_1$ und es ist $\text{id}_1 \circ \text{id}_2 = \text{id}_2$ und somit $\text{id}_1 = \text{id}_2$.

Ist die Kategorie \mathcal{C} aus dem Zusammenhang klar, dann kann man auch $\text{Mor}(A, B)$ anstelle von $\text{Mor}_{\mathcal{C}}(A, B)$ schreiben. Für $f \in \text{Mor}(A, B)$ schreiben wir auch $f : A \rightarrow B$, wobei f aber nicht notwendig eine Abbildung sein muss.

Wir haben bereits einige Kategorien gesehen, wie etwa Set , mit den Mengen als Objekten, den Abbildungen mit der Abbildungskomposition als Morphismen und der Identität als Einheitsmorphismus. Es folgen die weiteren Kategorien, die bisher erwähnt wurden:

Beispiele 2.15 (Kategorien.).

- *\mathcal{Rels} ist die Kategorie der Mengen mit den zweistelligen Relationen als Morphismen und dem Relationenprodukt als Komposition. Die Gleichheitsrelation ist dann der Einheitsmorphismus.*
- *Top ist die Kategorie der topologischen Räume mit den stetigen Abbildungen.*
- *$\mathbb{Z} - \text{Mod}$ ist die Kategorie der \mathbb{Z} -Moduln mit den linearen Abbildungen.*
- *KetKomp ist die Kategorie der algebraischen Kettenkomplexe mit den Komplexmorphismen. Die Komposition von Komplexmorphismen geschieht in naheliegender Weise komponentenweise: Sind $m_i : C_i \rightarrow K_i$ und $n_i : K_i \rightarrow Z_i$ die linearen Abbildungen zweier Komplexmorphisme, dann ist $n_i \circ m_i : C_i \rightarrow Z_i$ die entsprechende lineare Abbildung im komponierten Komplexmorphismus.*
- *Graph ist die Kategorie der ungerichteten Graphen. Ihre Objekte sind Graphen (N, E, f) und ihre Morphisme $f : (N, E, g) \rightarrow (L, P, h)$ haben wir oben bereits erläutert.*
- *Kat ist die Kategorie der kleinen Kategorien. Ihre Objekte sind Kategorien, deren Klassen Mengen sind. Auf ihre Morphisme, die Funktoren, werden wir noch eingehen.*
- *Zu jeder Kategorie \mathcal{C} können wir eine weitere Kategorie \mathcal{C}^{op} durch „umdrehen“ der Morphisme definieren. Die Objekte von \mathcal{C}^{op} sind die selben wie in \mathcal{C} . Und für Objekte A, B ist $\text{Mor}_{\mathcal{C}^{\text{op}}}(A, B) := \text{Mor}_{\mathcal{C}}(B, A)$ mit der Komposition $f \circ^{\text{op}} g := g \circ f$. Dies Kategorie heißt opposite Kategorie.*
- *Für ein beliebiges Objekt \bullet können wir eine triviale Kategorie \mathcal{T} definieren: Die Objektklasse ist die Menge $\{\bullet\}$ und der einzige Morphismus ist $\text{id}_{\bullet} : \bullet \rightarrow \bullet$. Es handelt sich also in der Tat um eine kleine Kategorie.*

Mit Morphismen können Objekte einer Kategorie in Beziehung gesetzt werden. Sie legen auch fest, wann zwei Objekte einer Kategorie als „im Wesentlichen gleich“ zu betrachten sind.

Definition 2.35 (Isomorphismus, Isomorphie). *Seien A und B Objekte einer Kategorie \mathcal{C} . Ein Morphismus $f : A \rightarrow B$ heißt Isomorphismus, wenn es einen Morphismus $g : B \rightarrow A$ zurück gibt, so dass $g \circ f = \text{id}_A$ und $f \circ g = \text{id}_B$. Zwei Objekte A und B heißen isomorph in \mathcal{C} , wenn es einen Isomorphismus $f : A \rightarrow B$ gibt. Dann schreiben wir auch $A \cong B$.*

Wie wir bereits bei den stetigen Abbildungen gesehen haben, ist Isomorphie eine Äquivalenzrelation und der Morphismus zurück ist eindeutig bestimmt.

Beim Übergang von einer Kategorie in eine andere müssen wir darauf achten, dass wir die Morphismen respektieren. Wir haben also eine strukturerhaltende „Abbildung“ von einer Kategorie in die andere:

Definition 2.36 (Funktor). *Seien \mathcal{C} und \mathcal{D} Kategorien. Ein kovarianter Funktor F von \mathcal{C} nach \mathcal{D} ist eine Abbildung $F : \text{Ob}_{\mathcal{C}} \rightarrow \text{Ob}_{\mathcal{D}}$, die jedem Objekt A von $\text{Ob}_{\mathcal{C}}$ ein Objekt $F(A)$ (oder kurz FA) aus $\text{Ob}_{\mathcal{D}}$ zuordnet, und für je zwei Objekte A, B eine (gleich bezeichnete) Abbildung $F : \text{Mor}_{\mathcal{C}}(A, B) \rightarrow \text{Mor}_{\mathcal{D}}(FA, FB)$ mit folgenden Eigenschaften:*

(F1) $F(\text{id}_A) = \text{id}_{FA}$ für alle Objekte A .

(F2) $F(f \circ g) = F(f) \circ F(g)$ für alle Morphismen f und g , für die $f \circ g$ definiert ist.

Ein kovarianter Funktor F von \mathcal{C} nach \mathcal{D}^{op} heißt auch kontravarianter Funktor von \mathcal{C} nach \mathcal{D} . Kovariante Funktoren bezeichnen wir mit $F : \mathcal{C} \rightarrow \mathcal{D}$ und kontravariante Funktoren mit $F : \mathcal{C} \leftarrow \mathcal{D}$.

Ein Beispiel, das wir später benötigen, ist der identische Funktor $\text{ID} : \mathcal{C} \rightarrow \mathcal{C}$, mit $\text{ID}A = A$ für alle Objekte und $\text{ID}(f) = f$ für alle Morphismen.

Funktoren sind geeignet, Objekte einer Kategorie mit den Mitteln einer anderen Kategorie zu untersuchen. Denn wenn zwei Objekte A und B in einer Kategorie \mathcal{C} isomorph sind, und ein Funktor $F : \mathcal{C} \rightarrow \mathcal{D}$ existiert, dann ist ebenfalls FA isomorph zu FB in der Kategorie \mathcal{D} : Ist nämlich A isomorph zu B dann existiert ein Isomorphismus $f : A \rightarrow B$ mit Umkehrmorphismus $f^{-1} : B \rightarrow A$. Dann gilt:

$$\begin{aligned} F(f) \circ F(f^{-1}) &= F(f \circ f^{-1}) = F(\text{id}_B) = \text{id}_{FB} \\ F(f^{-1}) \circ F(f) &= F(f^{-1} \circ f) = F(\text{id}_A) = \text{id}_{FA} \end{aligned}$$

Also ist $F(f) : FA \rightarrow FB$ ein Isomorphismus in \mathcal{D} mit Umkehrmorphismus $F(f^{-1})$. Die Umkehrung gilt jedoch im allgemeinen nicht. Sind FA und FB isomorph, dann folgt daraus nicht notwendigerweise die Isomorphie von A und B , denn für jede Kategorie gibt es einen Funktor T in die triviale Kategorie unseres obigen Beispiels. Dann ist für alle Objekte A, B der Ausgangskategorie ihr Bild TA isomorph zu TB , denn es ist sogar $TA = TB = \bullet$.

Interessant sind nun diejenigen Funktoren, bei denen Isomorphie in der einen Kategorie immer mit Isomorphie in der anderen Kategorie übereinstimmt. In dieser Arbeit soll topologische Isomorphie (also Homöomorphie) in architektonischen Räumen mit Isomorphie ihrer Darstellung im Computer übereinstimmen:

Definition 2.37 (Äquivalenz von Kategorien). *Seien \mathcal{Arch} und \mathcal{Code} Kategorien. Ein Funktor $F : \mathcal{Arch} \rightarrow \mathcal{Code}$ heißt Äquivalenz von Kategorien, wenn es einen Funktor $G : \mathcal{Code} \rightarrow \mathcal{Arch}$ zurück gibt, so dass jedes Objekt A in \mathcal{Arch} isomorph zu $G \circ F(A)$ und jedes Objekt B in \mathcal{Code} isomorph zu $F \circ G(B)$ ist. Zwei Kategorien \mathcal{Arch} und \mathcal{Code} heißen äquivalent, wenn es eine Äquivalenz von Kategorien $F : \mathcal{Arch} \rightarrow \mathcal{Code}$ gibt.*

Wir wollen nun zeigen, dass die Äquivalenz von Kategorien das gewünschte tut. Sei also $F : \mathcal{C} \rightarrow \mathcal{D}$ eine Äquivalenz von Kategorien mit Umkehrfunktor $G : \mathcal{D} \rightarrow \mathcal{C}$. Für Objekte A und B in \mathcal{C} mit FA isomorph zu FB gilt GFA isomorph zu GFB . Dann ist ebenfalls GFA isomorph zu A und GFB isomorph zu B . Also $A \cong GFA \cong GFB \cong B$ und da Isomorphie transitiv ist gilt $A \cong B$. Ist umgekehrt X und Y in \mathcal{D} mit $GX \cong GY$, dann zeigt man genauso $X \cong Y$.

Also gibt es für alle Objekte einer Kategorie entsprechende Objekte in der anderen Kategorie und Isomorphie auf der einen Seite entspricht genau der Isomorphie auf der anderen Seite.

Wir wollen diese Sprechweise nun zunächst verwenden, um das relationale Datenmodell zu erläutern. Später soll dessen topologische Erweiterung mit kategoriellen Mitteln definiert und untersucht werden.

2.8 Das relationale Datenmodell

Ein Datenmodell ist eine Konvention, um Datentypen, also Mengen von Datenelementen, und die für die entsprechenden Datenmengen erlaubten Operatoren zu definieren. Das relationale Datenmodell basiert auf dem uns bereits bekannten Konzept der Relationen mit den üblichen mengentheoretischen Operatoren darauf.

2.8.1 Relationen

Im relationalen Datenmodell definieren wir eine Tabellenstruktur anhand vorgegebener Wertebereiche. Einem Datensatz entspricht eine Zeile in einer Tabelle mit benannten Spalten und setzt sich aus einzelnen elementaren Werten zusammen. Ein bestimmter Wert eines Datensatzes ist der Eintrag in dieser Zeile in einer entsprechend benannten Spalte. Jede Spalte darf nur Werte des vorher bestimmten Wertebereichs enthalten. Dieser Wertebereich bildet einen Datentyp, ist also eine Menge von zulässigen Werten (Zahlen, Texte etc.) und den jeweils erlaubten Operatoren wie Addition, Verkettung oder Vergleich.

Wir haben bereits gesehen, dass ein n -Tupel $t = (a_1, \dots, a_n)$ eine Abbildung

$$t : \{1, \dots, n\} \rightarrow D, \quad i \mapsto t(i) := a_i$$

ist. Für relationale Datenbanken sollen nun aus praktischen Gründen beliebige symbolische Bezeichner für die Komponenten eines Tupels verwendet werden, statt diese bloß durchnummerieren. Zudem ist für jede Komponente des Tupels dessen Datentyp aus einer vorher festgelegten Menge von Grunddatentypen, den *Domänen*, festzulegen.

Definition 2.38 (Relationenschema, Tupel, Relation). *Sei \mathcal{D} eine Menge von Mengen, $\mathcal{D} = \{D_1, \dots, D_k\}$, die wir als Domänen bezeichnen und*

$\mathbf{U} := \bigcup \mathcal{D}$ die universelle Domäne. Sei \mathcal{A} eine Menge von Attributsymbolen (z.B. Wörter über einem Alphabet). Ein Relationenschema zu \mathcal{D} ist eine Abbildung $\mathcal{R} : \mathcal{A} \rightarrow \mathcal{D}$. Ein Tupel zu einem Relationenschema \mathcal{R} ist eine Abbildung $t : \mathcal{A} \rightarrow \mathbf{U}$ mit der Eigenschaft $t(a) \in \mathcal{R}(a)$ für alle $a \in \mathcal{A}$. Sei $\text{TUP}(\mathcal{R})$ die Menge aller Tupel zu \mathcal{R} , dann heißt ein Paar (\mathcal{R}, R) (endliche) Relation zu \mathcal{R} , wenn R (endliche) Teilmenge $R \subseteq \text{TUP}(\mathcal{R})$ von $\text{TUP}(\mathcal{R})$ ist. R heißt dann die Ausprägung von (\mathcal{R}, R) . Die Menge aller Relationen zu \mathcal{R} bezeichnen wir mit $\text{REL}(\mathcal{R})$.

Eine *Relation* ist hier eine Tabelle mit ihren Daten. Ein *Relationenschema* ist ein Synonym für die Definition der Tabelle, unabhängig davon, welche Daten sie enthält. *Ausprägung* hingegen ist nur ein Synonym für Graph einer Relation. Wir erinnern uns, dass wir für eine Relation $R = (\mathcal{R}, G)$ definieren, dass $t \in R$ gleichbedeutend mit $t \in G$ ist.

Wenn wir von „dem Attribut aa“ sprechen, dann müssen wir streng genommen unterscheiden zwischen dem Attributsymbol „aa“, dem Attributwert $t(\text{aa})$ für jedes Tupel t und der Abbildung $\pi_{\text{aa}} : \text{TUP}(\mathcal{R}) \rightarrow \text{dom}_{\mathcal{R}}(\text{aa})$, die jedem Tupel t den Attributwert $\pi_{\text{aa}}(t) := t(\text{aa})$ zuordnet. Die Menge \mathcal{A} der Attributsymbole wird in der Regel als endlich angenommen. Wir wollen dies hier nicht generell fordern, werden uns in der Regel aber auf endliche Attributmengen beschränken. Ist auf den Domänen eine Verknüpfung definiert, dann ist z.B. in SQL erlaubt, entsprechend sinnvolle Verkettungen von Attributsymbolen mit Verknüpfungssymbolen vorzunehmen: Ist z.B. $\text{dom}_{\mathcal{R}}(a) = \text{dom}_{\mathcal{R}}(b) = \mathbb{Z}$, dann kann mit „+“ als Verknüpfungssymbol für die Addition das Wort „a+b“ ebenfalls als Attributsymbol und für ein Tupel t zu \mathcal{R} aufgefasst werden. Dessen Wert ist definiert durch

$$t(\text{„a+b“}) := t(\text{„a“}) + t(\text{„b“}).$$

Damit lassen sich dann unendlich viele sinnvolle „Attributsymbole“ durch fortgesetztes Verketteten erzeugen.

In der Literatur wird meist nur die Attributmenge \mathcal{A} als Relationenschema bezeichnet. Dann wird aber zusätzlich gefordert, dass dieser eine Abbildung $\text{dom} : \mathcal{A} \rightarrow \mathcal{D}$ zugeordnet wird. Es erscheint also gerechtfertigt, eben diese Abbildung als Relationenschema zu bezeichnen, denn die Attributmenge ist in dieser Abbildung ja als Definitionsbereich bereits enthalten. Wir wollen die Schreibweise $\text{dom}_{\mathcal{R}}$ verwenden, wenn wir besonderes Augenmerk auf die Zuordnung von Attributsymbolen zu den Domänen legen, ansonsten einfach \mathcal{R} .

Ein Relationenschema $\mathcal{R} : \{A_1, \dots, A_n\} \rightarrow \mathcal{D}, A_i \mapsto D_i$ schreiben wir auch in der Form $\mathcal{R}[A_1 : D_1, \dots, A_n : D_n]$, und für ein Tupel $t : \{A_1, \dots, A_n\} \rightarrow \mathbf{U}, A_i \mapsto a_i$ ist auch die Schreibweise $\langle a_1 : A_1, \dots, a_n : A_n \rangle$ üblich.

Beispiel 2.16 (Relationenschema). Für eine Liste von Adressen wäre das folgende Relationenschema geeignet (effektiv). Das heißt in einer Tabelle nach diesem Schema lassen sich in der Tat Firmenadressen ablegen:

```
Adresse[firmenname:TEXT, vorwahl:TEXT, telefonnummer:TEXT,
strasse:TEXT, hausnummer:ZAHL, hausnummerzusatz:TEXT,
plz:TEXT, ort:TEXT,
land:TEXT ]
```

Auf die Frage, ob eine Relation nach diesem Schema auch praxistauglich (effizient) ist, wird hier noch nicht eingegangen. (Sie ist es übrigens nicht.) Das Attribut `plz` ist hier nicht als Zahl dargestellt, weil sonst die Handhabung von Postleitzahlen mit führenden Nullen umständlicher würde. Standardgemäß werden nämlich bei Abfragen fast alle Ganzzahlen ohne führende Nullen ausgegeben.

Diese Definition von Relation ist eine Verallgemeinerung unseres bisherigen Relationenbegriffs, wie folgendes Beispiel zeigt:

Beispiel 2.17 (mathematische Relationen). Sei $R \subseteq M_1 \times \dots \times M_n$ eine Relation im traditionellen mathematischen Sinn, dann ist

$$\mathcal{R} : \{1, \dots, n\} \rightarrow \{M_1, \dots, M_n\}, i \mapsto M_i$$

sein Relationenschema und ein n -Tupel $(m_1, \dots, m_n) \in M_1 \times \dots \times M_n$ ist ebenfalls ein Tupel in $\text{TUP}(\mathcal{R})$ und $((M_1, \dots, M_n), R)$ ist eine Relation nach obiger Definition. Es ist $(m_1, \dots, m_n) = \langle m_1 : 1, \dots, m_n : n \rangle$ in der oben erwähnten Schreibweise.

Allerdings wird unter „Relation“ auch in der Mathematik oft dieser etwas allgemeinere Begriff verstanden, in dem nicht bloß Zahlen als Attributsymbole erlaubt sind.

2.8.2 Operatoren für Relationen

Mit Relationen und Relationenschemata haben wir die Möglichkeit, Datenstrukturen festzulegen. Die zugehörigen Daten müssen nun angelegt, abgefragt, geändert und gelöscht werden können. Im relationalen Datenmodell werden die dazu notwendigen Grundoperatoren bereitgestellt. Ein Beispiel für einen derartigen Operator haben wir bereits mit dem Relationenprodukt kennen gelernt:

Beispiel 2.18 (Relationenprodukt). Wir erinnern uns, dass das Relationenprodukt \circ für Relationen eine Verknüpfung von zwei Relationen $R \subseteq X \times Y$ und $S \subseteq Y \times Z$ ist, die eine Relation $R \circ S \subseteq X \times Z$ ergibt.

Mengentheoretische Verknüpfungen. Da die Graphen von Relationen Mengen sind, können wir auch die mengentheoretischen Verknüpfungen, wie Vereinigungs-, Schnitt- und Restmenge an Relationen vornehmen. Da das Ergebnis aber stets wieder Graph einer Relation sein soll, müssen wir diese Operatoren einschränken:

Definition 2.39 (Vereinigungsverträglichkeit). Zwei Relationen heißen vereinigungsverträglich, wenn sie Relationen zum gleichen Relationenschema sind.

Im relationalen Datenmodell ist diese Definition nun so streng, dass sie unsere Definition von Vereinigung, Schnitt und Differenz von Relationen aus 2.1.4 auf Seite 30 in dieser Allgemeinheit untersagt. Diese Strenge erleichtert aber die Implementierung von relationalen Datenbankmanagementsystemen (DBMS).

Für vereinigungsverträgliche Relationen haben wir nun unsere ersten Operatoren Vereinigung, Schnitt und Restmengen:

Definition 2.40 (Vereinigung, Schnitt und Rest). Sei \mathcal{R} ein Relationenschema und (\mathcal{R}, R) und (\mathcal{R}, S) Relationen dazu. Dann sind diese Relationen vereinigungsverträglich und es ist

- die Vereinigung in \mathcal{R}

$$\cup : \text{REL}(\mathcal{R}) \times \text{REL}(\mathcal{R}) \rightarrow \text{REL}(\mathcal{R})$$

definiert durch

$$(\mathcal{R}, R) \cup (\mathcal{R}, S) := (\mathcal{R}, \{t \in \text{TUP}(\mathcal{R}) \mid t \in R \text{ oder } t \in S\}).$$

- der Schnitt in \mathcal{R}

$$\cap : \text{REL}(\mathcal{R}) \times \text{REL}(\mathcal{R}) \rightarrow \text{REL}(\mathcal{R})$$

definiert durch

$$(\mathcal{R}, R) \cap (\mathcal{R}, S) := (\mathcal{R}, \{t \in \text{TUP}(\mathcal{R}) \mid t \in R \text{ und } t \in S\})$$

und

- die Restmenge in \mathcal{R}

$$\setminus : \text{REL}(\mathcal{R}) \times \text{REL}(\mathcal{R}) \rightarrow \text{REL}(\mathcal{R})$$

definiert durch

$$(\mathcal{R}, R) \setminus (\mathcal{R}, S) := (\mathcal{R}, \{t \in \text{TUP}(\mathcal{R}) \mid t \in R \text{ und } t \notin S\}).$$

Für nicht vereinigungsverträgliche Relationen sind diese Verknüpfungen nicht zugelassen.

Mit diesen Verknüpfungen können wir auch Datensätze verschiedener Tabellen zusammenführen (Vereinigung) oder einen Datensatz t zu R durch $R \cup \{t\}$ hinzufügen, durch $R \setminus \{t\}$ entfernen oder – falls $t \in R$ – zu t' durch $(R \setminus \{t\}) \cup \{t'\}$ ändern.

Selektion, Projektion und Umbenennung Wir werden hier die kategorielle Sprechweise verwenden und sehen, dass die drei wichtigsten Operatoren der relationalen Algebra sowohl Morphismen einer Kategorie \mathcal{RelSc} der relationalen Schemata als auch einer Kategorie \mathcal{RelDb} der Relationen sind. Die Ausführung dieser Abfrageoperatoren entspricht dann einem kontravarianten Funktor von den Relationenschemata zu den jeweiligen Relationen.

Selektion. Wenn wir eine Relation R haben und alle Tupel in R suchen, die eine gewisse Eigenschaft P haben, dann bezeichnen wir diesen Vorgang als Selektion von R anhand von P . Doch zuerst muss feststehen, was eine „Eigenschaft“ denn überhaupt ist.

Definition 2.41 (Prädikat). Sei $\mathcal{R}[a_1 : D_1, \dots, a_n : D_n]$ ein Relationenschema. Ein Prädikat Θ zu \mathcal{R} ist eine Formel, die aus konstanten Werten aus $D_1 \cup \dots \cup D_n$, Attributsymbolen aus $\{a_1, \dots, a_n\}$ und den in den D_1, \dots, D_n erlaubten algebraischen Operatoren und Relationensymbolen besteht, so dass für jedes Tupel $t \in \text{TUP}(\mathcal{R})$ durch Ersetzen von jedem Attributsymbol a_i in Θ durch seinen Wert $t(a_i)$ eine logische Aussage $\Theta(t)$ entsteht, die entweder wahr oder falsch ist.

Eine „Formel“ ist ein Wort über einem Alphabet, das durch gewisse Regeln konstruiert wird. Einzelheiten entnehme man einschlägiger Literatur über formale Sprachen oder dem kurzen Überblick in [Sed91, S.358ff].

Ein Beispiel für ein Prädikat ist " $x=5$ oder $y<7$ " — ein Prädikat zu dem Schema $\mathcal{R}[x : \mathbb{Z}, y : \mathbb{Z}]$. Haben wir ein Tupel $\langle 5 : x, 13 : y \rangle$, dann erhalten wir durch Ersetzen der Attributsymbole durch die entsprechenden Werte des Tupels den Ausdruck " $5=5$ oder $13<7$ ", der in diesem Fall wahr ist. Wir sagen dann auch, das Tupel *erfüllt* das Prädikat. Ein Ausdruck wie " $x+y-3$ " ist hingegen kein Prädikat, da seine Auswertung nicht „wahr“ oder „falsch“, sondern einen Zahlenwert zurückgibt.

Selektion ist die Auswahl derjenigen Tupel, für die die Auswertung des Prädikats wahr ist:

Definition 2.42 (Selektion). Sei \mathcal{R} ein Relationenschema und Θ ein Prädikat zu \mathcal{R} . Die Selektion mit Θ ist die Abbildung

$$\sigma_{\Theta} : \text{REL}(\mathcal{R}) \rightarrow \text{REL}(\mathcal{R}), R \mapsto \sigma_{\Theta}(R) := (\mathcal{R}, \{t \in R \mid \Theta(t) \text{ ist wahr.}\}).$$

Die Selektion $\sigma_{\Theta}(R)$ hat also das selbe Relationenschema wie R .

Beispiel 2.19 (Selektion einer IfcDoor). Angenommen, wir haben alle Türen eines Projekts als Relation `IfcLikeDoor` gespeichert. Das ist ein Datentyp, den wir erhalten, wenn wir bei `IfcDoor` aus Abschnitt 1.4.1 einige Attribute weglassen und ID für den Identifizierer von STEP-Objekten hinzufügen.

<i>IfcLikeDoor</i>				
<u>ID</u>	<i>GlobalId</i>	<i>Representation</i>	<i>OverallHeight</i>	<i>OverallWidth</i>
8015	'wrldmbrmft'	8014	\$	\$
8032	'wrldmdrmft'	8031	\$	\$
8049	'wrlbbrmft'	8014	\$	\$

Suchen wir nun alle Türen, welche die geometrische Repräsentation mit der Nummer 8014 haben, dann erhalten wir

$\sigma_{\text{Representation}=8014}(\textit{IfcLikeDoor})$				
<u>ID</u>	<i>GlobalId</i>	<i>Representation</i>	<i>OverallHeight</i>	<i>OverallWidth</i>
8015	'wrldmbrmft'	8014	\$	\$
8049	'wrlbbrmft'	8014	\$	\$

Es fehlt z.B. die Türe 'wrldmdrmft' mit der Nummer 8032.

Projektion. Selektion ist die Auswahl von Zeilen einer Tabelle, Projektion hingegen die Auswahl von Spalten. Mit der Projektion wird nur auf bestimmte Attribute einer Relation zugegriffen, andere hingegen ausgeblendet. So darf etwa ein Angestellter einer Firma zwar die Liste der dienstlichen Telefonnummern

seiner Kollegen, nicht jedoch deren Gehälter einsehen. Durch Projektion lässt sich dann z.B. die Spalte mit den Gehältern ausblenden. Wir erinnern zunächst an unsere, von der Standardliteratur etwas abweichende, Definition 2.38, wonach ein Relationenschema \mathcal{R} eine Abbildung $\mathcal{R} : \mathcal{A} \rightarrow \mathcal{D}$ von einer Menge \mathcal{A} von Attributbezeichnern in die Menge \mathcal{D} der Domänen ist. Diese sind eine vorher festgelegte Menge der verwendeten elementaren Datentypen und in der Regel durch ein Datenbankmanagementsystem vorgegeben.

Definition 2.43 (Projektion). Sei $\mathcal{R} : \mathcal{A} \rightarrow \mathcal{D}$ ein Relationenschema und $\mathcal{B} \subseteq \mathcal{A}$ eine Menge von Attributsymbolen in \mathcal{A} . Die Projektion auf \mathcal{B} ist die Abbildung

$$\pi_{\mathcal{B}} : \text{REL}(\mathcal{R}) \rightarrow \text{REL}(\mathcal{R}|_{\mathcal{B}}), R \mapsto \pi_{\mathcal{B}}(R) := (\mathcal{R}|_{\mathcal{B}}, \{t|_{\mathcal{B}} \mid t \in R\}).$$

Wir erinnern uns, dass zu einer Abbildung $f : X \rightarrow Y$ und einer Teilmenge $A \subseteq X$ die Einschränkung $f|_A : A \rightarrow Y, a \mapsto f(a)$ definiert ist. Ein Tupel t zu einem Relationenschema $\mathcal{R} : \mathcal{A} \rightarrow \mathcal{D}$ wird nun durch $t|_{\mathcal{B}}$ auf die Attribute in \mathcal{B} eingeschränkt und ist ein Tupel zum eingeschränkten Relationenschema $\mathcal{R}|_{\mathcal{B}} : \mathcal{B} \rightarrow \mathcal{D}$. Statt $t|_{\mathcal{B}}$ schreiben wir auch $\pi_{\mathcal{B}}(t)$.

Es ist immer $f|_A = f \circ i$ mit der Inklusionsabbildung $i : A \hookrightarrow X, a \mapsto a$ und daher ist $\pi_{\mathcal{B}}(R) = (\mathcal{R} \circ i, \{t \circ i \mid t \in R\})$ wobei i die Inklusion $i : \mathcal{B} \hookrightarrow \mathcal{A}$ ist. Umgekehrt geht $\pi_{\mathcal{B}}$ in die entgegengesetzte Richtung von den Relationen mit Attributen \mathcal{A} zu den Relationen mit Attributen \mathcal{B} .

Beispiel 2.20 (Projektion einer IfcDoor). Unser Beispieltyp `IfcLikeDoor` ist bereits eine Projektion von `IfcDoor` auf eine kleinere Attributmenge, wieder vorausgesetzt, ID ist der Identifizierer von STEP-Objekten. Wollen wir nun die Zuordnung von STEP-IDs auf die geometrische Repräsentation, dann erreichen wir dies durch folgende Projektion:

$\pi_{\{ID, Representation\}}(\text{IfcLikeDoor})$	
<u>ID</u>	Representation
	⋮
8015	8014
8032	8031
8049	8014
	⋮

Da wir hier auf einen sog. Superschlüssel projiziert haben, ist die Anzahl der Ergebnisdatensätze gleich der Anzahl der Ausgangsdatsätze.

Quotientenrelation. Durch die Projektion $\pi_{\mathcal{B}}$ haben wir auch für jede Relation (\mathcal{R}, R) auch eine Äquivalenzrelation $\sim_{\mathcal{B}}$ in R definiert durch $t \sim_{\mathcal{B}} s$ genau dann, wenn $t|_{\mathcal{B}} = s|_{\mathcal{B}}$ und somit eine Zerlegung von R in Äquivalenzklassen $R/\mathcal{B} := R/\sim_{\mathcal{B}}$. Jeder Klasse $[t] \in R/\mathcal{B}$ kann auch ein Tupel \bar{t} zu einem geeigneten Relationenschema zugeordnet werden. So ist für jedes Attributsymbol a in \mathcal{B} der Attributwert von $\bar{t}(a)$ gleich dem Wert von $t(a)$ für einen beliebigen Repräsentanten t . Weiterhin ist es sinnvoll, ein Attributsymbol für die Anzahl der Tupel in $[t]$, oder für Summe, Mittel- und Extremwerte von Attributwerten zu vereinbaren.

In SQL wird diese Zerlegung durch `group by`-Ausdrücke vorgenommen. Die entsprechenden Attributsymbole sind `count(*)` für die Anzahl von Tupeln in $[t]$. `max(a)` und `min(a)` sind das Maximum und Minimum aller Attributwerte von a in $[t]$ (falls die Domäne von a linear geordnet ist). Ist in der Domäne von a die Addition definiert, dann gibt es auch `sum(a)` und `mid(a)`. Der letztere ist gleichbedeutend zu `sum(a)/count(a)`. CODD nennt diese Operation „Framing“, eine Äquivalenzklasse entsprechend „Frame“ und er hat auch eine andere Konvention für die Attributsymbole für Anzahl, Summe, Extremwerte und Mittelwerte [Cod90, S.98ff].

Formal wäre gegen eine Projektion (bzw. Quotientenbildung) auf \emptyset als Attributmenge nichts einzuwenden. Die entsprechende Äquivalenzrelation ist die Universalrelation und die Projektion auf \emptyset ergibt die leere Relation (\emptyset, \emptyset) oder die Relation $(\emptyset, \{()\})$ mit genau einem leeren Tupel. Dies hätte sogar einen gewissen Nutzen in Kombination mit anderen Abfragen, wenn wir z.B. (\emptyset, \emptyset) als logisch „falsch“ und $(\emptyset, \{()\})$ als logisch „wahr“ interpretieren.

In SQL ist die Projektion auf \emptyset nicht ohne weiteres möglich. Dort kann aber ein Abfrageergebnis R wie ein Element aus R/\emptyset behandelt werden, da die Symbole `count`, `min`, `sum` etc. auch ohne Verwendung des `group by`-Operators verwendet werden dürfen.

Umbenennung. Manchmal müssen Attributsymbole ausgetauscht werden, ohne die in der Relation gespeicherten Informationen zu ändern. Hat eine Relation R zum Beispiel ein Attribut `nr` und S ein Attribut `nummer`, wobei `nummer` in S das selbe bedeutet wie `nr` in R , dann kann es sein, dass wir aus S eine Relation S' erzeugen wollen, in der der Attributbezeichner `nummer` ersetzt wird durch `nr`. Damit kann für R und S' Vereinigungsverträglichkeit hergestellt werden, um beispielsweise Daten aus S nach R zu überspielen.

Definition 2.44 (Umbenennung). Sei $\mathcal{R} : \mathcal{A} \rightarrow \mathcal{D}$ ein Relationenschema und $f : \mathcal{B} \rightarrow \mathcal{A}$ eine surjektive Abbildung von Attributsymbolen \mathcal{B} zu Attributsymbolen in \mathcal{A} . Die Umbenennung der Attribute in \mathcal{A} durch f ist die Abbildung

$$\beta_f : \text{REL}(\mathcal{R}) \rightarrow \text{REL}(\mathcal{R} \circ f), R \mapsto \beta_f(R) := (\mathcal{R} \circ f, \{t \circ f \mid t \in R\}).$$

Meist werden nur wenige Attribute umbenannt. Dann schreibt man f in der Form $a \leftarrow f(a), b \leftarrow f(b)$ oder $(a, b) \leftarrow (f(a), f(b))$. Man gibt dabei nur diejenigen Attribute a_i an, für die $f(a_i) \neq a_i$ gilt, die also durch f echt umbenannt werden. Für unser obiges Beispiel gilt dann $S' = \beta_{\text{nr} \leftarrow \text{nummer}}(S)$. Durch Umbenennung ist auch die Verdoppelung von Attributen möglich. Wenn wir die Zahl 1 als Attributsymbol verwenden, dann ist jede Relation R zum (sehr simplen) Schema $\mathcal{R} : \{1\} \rightarrow M$ im Wesentlichen eine Teilmenge von M , denn ein Tupel in R hat ja die Form (m) für ein $m \in M$. Daraus kann nun die Gleichheitsrelation $\Delta_R = \beta_{(1,2) \leftarrow (1,1)}(R)$ von R erzeugt werden. Sie besteht aus denjenigen Tupel (m, m) , für die es ein Tupel $(m) \in R$ gibt. In SQL wird die Umbenennung durch das Schlüsselwort `as` vorgenommen.

Wir haben hier ebenfalls eine Abbildung f von den Attributen des neuen Relationenschemas in die Attribute des ursprünglichen Relationenschemas und eine Abbildung β_f in umgekehrter Richtung von den Relationen des Ausgangsschemas in die Relationen des neuen Schemas.

Implizites Umbenennen. Wird eine Relation mit einem Symbol R bezeichnet und ist S ebenfalls ein Symbol, dann gibt es die Umbenennung $\beta_S(R)$, die jedes Attributsymbol a zu $S.a$ umbenennt und die implizite Umbenennung $\beta(R) := \beta_R(R)$, die dabei das Symbol im Argument selbst verwendet. Ist also a Attribut von R , dann wird dieses durch $\beta(R)$ zu $R.a$ umbenannt. Derartige Umbenennungen werden von SQL bei Bedarf automatisch vorgenommen.

Projektion und Umbenennung. Die Definition der Umbenennung fordert eine surjektive Abbildung $f : \mathcal{B} \rightarrow \mathcal{A}$ von den Attributsymbolen \mathcal{B} zu den Symbolen \mathcal{A} . Erlauben wir auch, dass f nicht surjektiv ist, dann erhalten wir eine gemeinsame Verallgemeinerung von Projektion und Umbenennung in der Form

$$\delta_f : \text{REL}(\mathcal{R}) \rightarrow \text{REL}(\mathcal{R} \circ f), \quad R \mapsto \delta_f(R) := (\mathcal{R} \circ f, \{t \circ f \mid t \in R\}).$$

Dann hat δ bereits auffallende Ähnlichkeit mit einem kovarianten Funktor, denn für jede Relation (\mathcal{R}, R) ist:

$$\begin{aligned} \delta_f \circ \delta_g(\mathcal{R}, R) &= \delta_f(\delta_g(\mathcal{R}, R)) \\ &= \delta_f(\mathcal{R} \circ g, \{t \circ g \mid t \in R\}) \\ &= ((\mathcal{R} \circ g) \circ f, \{(t \circ g) \circ f \mid t \in R\}) \\ &= (\mathcal{R} \circ (g \circ f), \{t \circ (g \circ f) \mid t \in R\}) \\ &= \delta_{g \circ f}(\mathcal{R}, R) \end{aligned}$$

Insbesondere ist jede Abbildung $f : \mathcal{X} \rightarrow \mathcal{Y}$ zerlegbar in ein surjektives $g : \mathcal{X} \rightarrow f[\mathcal{X}]$, $a \mapsto f(a)$ und die Inklusionsabbildung $i : f[\mathcal{X}] \hookrightarrow \mathcal{Y}$ mit $f = i \circ g$ und somit

$$\delta_f = \delta_{i \circ g} = \delta_g \circ \delta_i = \beta_g \circ \pi_{f[\mathcal{X}]}$$

Damit gibt es zu jeder beliebigen Verkettung von Projektionen und Umbenennungen in Datenbankabfragen eine gleichwertige Abfrage mit nur einer Projektion und einer darauf folgenden Umbenennung.

Einstellige Abfragen als Morphismen. Jede der drei Operatoren Selektion, Projektion und Umbenennung ist eine Abbildung $f : \text{REL}(\mathcal{R}) \rightarrow \text{REL}(\mathcal{S})$. Andererseits ist die identische Abbildung $\text{id}_{\mathcal{R}} : \text{REL}(\mathcal{R}) \rightarrow \text{REL}(\mathcal{R})$ sowohl Selektion, als auch Projektion oder Umbenennung. Sie ist eine Selektion mit der Tautologie „wahr“, dem Prädikat, das stets erfüllt ist. Die Identität ist eine Projektion auf alle Attribute des Relationenschemas und sie ist eine Umbenennung, da sie jedes Attribut a (nicht echt) in a „umbenennt“. Damit ist jeder dieser drei Abfrageoperatoren offensichtlich als eine Verkettung der Form $\beta_f \circ \pi_{\mathcal{A}} \circ \sigma_{\Theta}(R)$ darstellbar: Zuerst werden alle Datensätze von R ausgewählt, die Θ erfüllen, dann wird auf \mathcal{A} projiziert und schließlich werden die Attribute in \mathcal{A} vermittels $f : \mathcal{B} \rightarrow \mathcal{A}$ umbenannt. Setzen wir $\mathcal{B} = \{b_1, \dots, b_n\}$, dann entspricht dies dem SQL-Ausdruck

```
select f(b1) as b1, ..., f(bn) as bn
from R
where  $\Theta$ ;
```

Dass wir jeden einzelnen der drei Abfrageoperatoren wiederum als eine Verkettung dieser drei Operatoren in der oben gezeigten Form darstellen können, ist klar, denn haben wir einen dieser Operatoren vorliegen, dann wählen wir für die beiden anderen einfach diejenige Form, die die identische Abbildung ergibt.

Wir wollen nun umgekehrt zeigen, dass jede beliebige zulässige Komposition dieser drei Operatoren wiederum als Verkettung der Form $\beta_f \circ \pi_{\mathcal{A}} \circ \sigma_P$ mit geeigneten \mathcal{A} , f und P darstellbar ist:

Sei $(\beta_f \circ \pi_{\mathcal{A}} \circ \sigma_P) \circ (\beta_g \circ \pi_{\mathcal{B}} \circ \sigma_Q)$ eine beliebige erlaubte Komposition von zwei derartigen Dreierketten. Dann können wir die Klammern weglassen (Assoziativität der Komposition) und erhalten

$$\beta_f \circ \pi_{\mathcal{A}} \circ \sigma_P \circ \beta_g \circ \pi_{\mathcal{B}} \circ \sigma_Q$$

Da σ_P nach β_g ausgeführt wird, ist P ein Prädikat für ein umbenanntes Relationenschema $\mathcal{R} \circ g$ mit Ausgangsschema $\mathcal{R} : \mathcal{X} \rightarrow \mathcal{D}$ und einer Abbildung $g : \mathcal{Y} \rightarrow \mathcal{X}$ die jedem umbenannten Attribut $y \in \mathcal{Y}$ sein ursprüngliches Attribut $g(y) \in \mathcal{X}$ zuordnet. Dann wollen wir ein Prädikat gP folgendermaßen definieren: Tausche jedes Attributsymbol x , das in P vorkommt durch $g(x)$ aus. Für jedes Tupel t zu \mathcal{R} ist somit $gP(t)$ genau dann wahr, wenn $P(t \circ g)$ wahr ist. $t \circ g$ ist ja das Tupel mit den durch g umbenannten Attributen. Also können wir $\sigma_P \circ \beta_g$ durch $\beta_g \circ \sigma_{gP}$ ersetzen und erhalten

$$\beta_f \circ \pi_{\mathcal{A}} \circ \beta_g \circ \sigma_{gP} \circ \pi_{\mathcal{B}} \circ \sigma_Q.$$

Nun wird die Selektion σ_{gP} nach einer Projektion ausgeführt und damit ist gP ein Prädikat für ein Relationenschema $\mathcal{R}|_{\mathcal{B}}$ einer Projektion auf \mathcal{B} . Damit ist gP ebenfalls ein Prädikat für \mathcal{R} und genau dann wahr für t , wenn es wahr für das projizierte Tupel $t|_{\mathcal{B}}$ ist, da in gP nur Attribute aus \mathcal{B} vorkommen. Also können wir Projektion und Selektion ebenfalls vertauschen:

$$\beta_f \circ \pi_{\mathcal{A}} \circ \beta_g \circ \pi_{\mathcal{B}} \circ \sigma_{gP} \circ \sigma_Q.$$

Ein Tausch in die andere Richtung ist natürlich nicht möglich. Denn wird eine Projektion nach einer Selektion ausgeführt, dann könnten ja durch die Projektion Attribute ausgeblendet worden sein, die von der Selektion noch zu berücksichtigen wären. Nun ist die Verkettung zweier Selektionen $\sigma_P \circ \sigma_Q$ gleich der Selektion $\sigma_{(P \text{ und } Q)}$. Denn $\sigma_P \circ \sigma_Q$ wählt zunächst alle Datensätze, die Q erfüllen, und von diesen dann diejenigen Datensätze, die P erfüllen. Das sind jedoch genau die Datensätze, die P und Q gleichzeitig erfüllen. Damit erhalten wir

$$\beta_f \circ \pi_{\mathcal{A}} \circ \beta_g \circ \pi_{\mathcal{B}} \circ \sigma_{(gP \text{ und } Q)}.$$

Ist nun eines der Prädikate P oder Q die Tautologie „wahr“, dann lassen wir eine dieser Tautologien weg und ersetzen „ gP und Q “ jeweils durch gP oder durch Q .

Die Verkettung von $\pi_{\mathcal{A}} \circ \beta_g$ bedeutet, dass zunächst die Attribute in \mathcal{B} durch $g : \mathcal{X} \rightarrow \mathcal{B}$ zu \mathcal{X} umbenannt werden und dann auf einen Teil $\mathcal{A} \subseteq \mathcal{X}$ dieser Attribute projiziert wird, die restlichen Attribute werden somit irrelevant. Dann können wir aber zunächst auf die relevanten Attribute $g[\mathcal{A}] \subseteq \mathcal{B}$ projizieren und nur diese durch die eingeschränkte Abbildung $g' : \mathcal{A} \rightarrow g[\mathcal{A}]$, $a \mapsto g(a)$ umbenennen. Damit ist $\pi_{\mathcal{A}} \circ \beta_g = \beta_{g'} \circ \pi_{g[\mathcal{A}]}$ und wir erhalten

$$\beta_f \circ \beta_{g'} \circ \pi_{g[\mathcal{A}]} \circ \pi_{\mathcal{B}} \circ \sigma_{(gP \text{ und } Q)}.$$

Die Verkettung der Projektionen und die der Umbenennungen können wir nun jeweils zusammenfassen zu einer Projektion und einer Umbenennung. Damit erhalten wir schließlich

$$(\beta_f \circ \pi_A \circ \sigma_P) \circ (\beta_g \circ \pi_B \circ \sigma_Q) = \beta_{g' \circ f} \circ \pi_{g[A]} \circ \sigma_{(gP \text{ und } Q)},$$

das Grundschemata eines einfachen (einstelligen) SQL-select Ausdrucks.

Bei anderen Verkettungen von Abfragen $Q = q_1 \circ \dots \circ q_n$ der Länge n ersetzen wir jede Abfrage q_i durch eine entsprechende Dreierkette $(\beta_{f_i} \circ \pi_{A_i} \circ \sigma_{\Theta_i})$, erhalten so eine Verkettung der Länge $3n$ und zeigen durch Induktion über n , dass diese zu einer Abfrage der Form $(\beta_{f_Q} \circ \pi_{A_Q} \circ \sigma_{\Theta_Q})$ gleich ist.

Nun können wir diese drei einstelligen Abfragen verallgemeinern:

Definition 2.45 (Abfrageschema). *Seien $\mathcal{R} : \mathcal{A} \rightarrow \mathcal{D}$ ein Relationenschema und \mathcal{B} eine Menge von Attributsymbolen. Ein Abfrageschema ist ein Paar (f, P) , bestehend aus einer Abbildung $f : \mathcal{B} \rightarrow \mathcal{A}$ und einem Prädikat P zu \mathcal{R} .*

Obiger Beweis zeigt, wie wir Abfrageschemata komponieren können, und damit erhalten wir eine Kategorie \mathcal{RelSc} mit den Relationenschemata als Objekten und den Abfrageschemata als Morphismen. Zudem haben wir eine weitere Kategorie \mathcal{RelDb} mit den Mengen der Relationen zu einem Schema als Objekten und den Abfragen Projektion, Selektion und Umbenennung bzw. deren Kompositionen als Morphismen. Der Funktor REL ordnet jedem Relationenschema \mathcal{R} die Menge $\text{REL}(\mathcal{R})$ der Relationen zu \mathcal{R} und jedem Abfrageschema $(f, P) : \mathcal{R} \circ f \rightarrow \mathcal{R}$ die Abfrage

$$\text{REL}(f, P) : \text{REL}(\mathcal{R}) \rightarrow \text{REL}(\mathcal{R} \circ f), R \mapsto (\mathcal{R} \circ f, \{t \circ f \mid t \in \sigma_P(R)\})$$

zu. REL ist damit ein kontravarianter Funktor $\text{REL} : \mathcal{RelSc} \leftarrow \mathcal{RelDb}$, wovon sich misstrauische Leserinnen selbst überzeugen können. In beiden Kategorien sind die bijektiven Umbenennungen die Isomorphismen.

Was haben wir gewonnen? Setzen wir endliche Attributmengen voraus, dann können wir nun zeigen, dass die hier definierten Relationen des relationalen Datenmodells zu den traditionellen n -stelligen Relationen äquivalent sind. Bis auf bijektive Umbenennung (Isomorphie in \mathcal{RelDb}) ist jede Relation mit endlich vielen Attributen im relationalen Datenmodell eine n -stellige Relation im traditionellen mathematischen Sinne.

Die Möglichkeit, Attribute durch Symbole zu benennen, hat also die praktische Handhabung von Relationen zwar sehr erleichtert, das Datenmodell dadurch aber nicht prinzipiell leistungsfähiger gemacht.

Im Verlauf dieser Arbeit werden oft die n -stelligen Relationen auftauchen, ohne dass dabei das relationale Datenmodell stets ausdrücklich erwähnt wird. Diese Arbeit zielt jedoch auf eine Realisierung als relationale Datenbank ab. Es wird nur nicht jedesmal neu darauf hingewiesen, dass es zu einer Relation $R \subseteq A \times B \times C$ stets auch eine gleichwertige Relation mit Schema $\mathcal{R}[a : A, b : B, c : C]$ nach dem relationalen Datenmodell gibt und dass R in einer relationalen Datenbank gespeichert werden kann, wenn R endlich ist.

Kartesisches Produkt und Verbund. Das kartesische Produkt ist Voraussetzung, um mehrere Relationen bei Abfragen zu kombinieren. In der relationalen Algebra weicht nun die Definition des kartesischen Produkts etwas von

der in der Mengenlehre üblichen Definition ab. Sind nämlich R und S Relationen mit Tupeln $(r_1, \dots, r_n) \in R$ und $(s_1, \dots, s_m) \in S$, dann wäre ein Element $p \in R \times S$ ein verschachteltes Paar von Tupeln $p = ((r_1, \dots, r_n), (s_1, \dots, s_m))$. In der Praxis sind jedoch Tupel der Form $t = (r_1, \dots, r_n, s_1, \dots, s_m)$ ohne innere Klammern bevorzugt.

Zunächst erinnern wir uns, dass für zwei Abbildungen $f : X \rightarrow M$ und $g : Y \rightarrow M$ die gemeinsame Fortsetzung $f \cup g$ genau dann definiert ist, wenn $f|_{X \cap Y} = g|_{X \cap Y}$. Sind f und g Tupel mit Attributmengen X und Y , dann ist $f|_{X \cap Y} = g|_{X \cap Y}$ gleichbedeutend damit, dass für jedes Attributsymbol $a \in X$, das auch in Y vorkommt, $f(a) = g(a)$ gilt.

Definition 2.46 (Natürlicher Verbund und kartesisches Produkt). *Seien $\mathcal{R} : \mathcal{A} \rightarrow \mathcal{D}$ und $\mathcal{S} : \mathcal{B} \rightarrow \mathcal{D}$ Relationenschemata mit $\mathcal{R}|_{\mathcal{A} \cap \mathcal{B}} = \mathcal{S}|_{\mathcal{A} \cap \mathcal{B}}$, für welche also die gemeinsame Fortsetzung $\mathcal{R} \cup \mathcal{S} : \mathcal{A} \cup \mathcal{B} \rightarrow \mathcal{D}$ definiert ist. Dann ist die Verknüpfung*

$$\bowtie : \text{REL}(\mathcal{R}) \times \text{REL}(\mathcal{S}) \rightarrow \text{REL}(\mathcal{R} \cup \mathcal{S})$$

definiert durch

$$R \bowtie S := (\mathcal{R} \cup \mathcal{S}, \{r \cup s \mid r \in R, s \in S, r|_{\mathcal{A} \cap \mathcal{B}} = s|_{\mathcal{A} \cap \mathcal{B}}\})$$

und heißt natürlicher Verbund von (Relationen zu) \mathcal{R} mit (Relationen zu) \mathcal{S} . Ist $\mathcal{A} \cap \mathcal{B} = \emptyset$, dann heißt der natürliche Verbund \bowtie auch kartesisches Produkt und wird auch mit dem Symbol \times geschrieben.

Eine mit dem natürlichen Verbund eng verwandte Abfrage ist das kartesische Produkt mit darauf folgender Selektion, was auch als *Theta-Verbund* bezeichnet wird. Die Bezeichnung kommt daher, dass der griechische Buchstabe Theta Θ in der Literatur meist als Platzhalter für das Auswahlprädikat verwendet wird.

Definition 2.47 (Theta-Verbund und Gleichheitsverbund). *Seien \mathcal{R} und \mathcal{S} Relationenschemata $\mathcal{R} : \mathcal{A} \rightarrow \mathcal{D}$ und $\mathcal{S} : \mathcal{B} \rightarrow \mathcal{D}$ mit disjunkten Attributmengen, also $\mathcal{A} \cap \mathcal{B} = \emptyset$. Sei zudem Θ ein Prädikat für das Relationenschema $\mathcal{R} \cup \mathcal{S}$, die gemeinsame Fortsetzung von \mathcal{R} und \mathcal{S} . Dann ist die Verknüpfung*

$$\bowtie_{\Theta} : \text{REL}(\mathcal{R}) \times \text{REL}(\mathcal{S}) \rightarrow \text{REL}(\mathcal{R} \cup \mathcal{S})$$

definiert durch

$$R \bowtie_{\Theta} S := \sigma_{\Theta}(R \times S)$$

und wird als Theta-Verbund von \mathcal{R} mit \mathcal{S} durch Θ bezeichnet. Ist Θ ein Prädikat in dem nur auf Gleichheit von Attributwerten geprüft wird, dann bezeichnet man \bowtie_{Θ} auch als Gleichheitsverbund (*equi-join*).

Der Theta-Verbund vergleicht also jeden Datensatz der einen Relation mit jedem Datensatz der anderen Relation, indem jeweils beide Datensätze zu einem gemeinsamen größeren Datensatz zusammengesetzt werden. Dann wird das Prädikat Θ auf diesen zusammengesetzten Datensatz angewendet. Ist Θ für den Datensatz wahr, wird er in das Ergebnis aufgenommen, ansonsten wird er verworfen.

Mit dem Gleichheitsverbund können wir das Relationenprodukt $R \circ S$ mit den Standardsymbolen der relationalen Algebra definieren:

$$R \circ S = \pi_{\{1,2\}} \beta_{L.2 \leftarrow 2}(R) \bowtie_{L.2=R.1} \beta_{R.1 \leftarrow 1}(S)$$

Wir benennen also die zweite Komponente in R zu $L.2$ um und die erste Komponente in S zu $R.1$. Dann bilden wir das kartesische Produkt der so umbenannten Relationen in Form von Tupeln $\langle a:1, b:L.2, c:R.1, d:2 \rangle$ für alle $a R b$ und $c S d$. Die darauf folgende Selektion wählt nur Tupel der Form $\langle a:1, j:L.2, j:R.1, d:2 \rangle$ aus, also diejenigen Tupel, die aus $a R j$ und $j S c$ hervorgegangen sind. Dann gilt jedoch $(a, c) \in R \circ S$. Die Projektion auf $\{1, 2\}$ ergibt genau diese Paare $\langle a:1, c:2 \rangle = (a, c)$.

Beispiel 2.21. *Angenommen wir haben die Mengen **Projekt** aller Projekte eines Büros, **Kunde** die Menge der Kunden und **Telefon** eine Liste von Telefonnummern. Zudem sei **beauftragt** eine Relation zum Schema $[\text{prj: Projekt, auftraggeber: Kunde}]$ und **telefonliste** eine Relation zum Schema $[\text{pers: Kunde, tel: Telefon}]$. Dann erhalten wir mit dem Gleichheitsverbund $\text{beauftragt} \bowtie_{\text{auftraggeber=pers}} \text{telefonliste}$ eine Relation mit Relationenschema*

$$\begin{aligned} & [\text{prj: Projekt, auftraggeber: Kunde}] \cup [\text{pers: Kunde, tel: Telefon}] \\ &= [\text{prj: Projekt, auftraggeber: Kunde, pers: Kunde, tel: Telefon}] \end{aligned}$$

in der wir für jedes Projekt die Telefonnummer des Auftraggebers erhalten.

Hat **beauftr** hingegen das Schema $[\text{prj: Projekt, pers: Kunde}]$, dann erhalten wir mit $\text{beauftr} \bowtie \text{telefonliste}$ ein praktisch gleichwertiges Ergebnis mit dem einfacheren Relationenschema

$$\begin{aligned} & [\text{prj: Projekt, pers: Kunde}] \cup [\text{pers: Kunde, tel: Telefon}] \\ &= [\text{prj: Projekt, pers: Kunde, tel: Telefon}]. \end{aligned}$$

Dieses Ergebnis hätten wir auch aus dem vorhergehenden Ergebnis des Gleichheitsverbundes durch Projektion auf die Attribute **prj**, **pers** und **tel** erhalten können.

In SQL wird das kartesische Produkt $R \times S$ durch die Angabe im **from**-Teil der **select**-Abfrage erzeugt:

```
select *
from   R,S;
```

Für einen Θ -Verbund ergänze man diese Abfrage noch um **where** Θ . Alternativ kann für $R \bowtie_{\Theta} S$ auch **R inner join S on** (Θ) anstelle von **R,S** im **from**-Teil geschrieben werden. „ Θ “ ist dabei natürlich stets durch den entsprechenden Prädikattext zu ersetzen. Das „**inner**“ deutet bereits an, dass es auch andere Varianten des relationalen Verbunds gibt, den sogenannten **outer-join**. Dieser macht jedoch beim topologischen Modellieren Probleme und wird deshalb in einem eigenen Kapitel behandelt.

Transitive Hülle. Mit den Verbundoperatoren allein ist ein Verbund einer Relation mit sich selbst nicht sinnvoll möglich, denn der Theta-Verbund $R \bowtie_{\Theta} R$ einer Relation R mit sich selbst nach obiger Definition ist meist nicht definiert, da für das kartesische Produkt die Attributmengen der beiden beteiligten Relationen disjunkt sein müssen. Es ist jedoch nur die leere Menge zu sich selbst disjunkt und es gibt nur zwei Relationen mit leerer Attributmenge. Auch der natürliche Verbund $R \bowtie R$ einer Relation mit sich selbst ist nicht nützlich, denn

dieser ergibt eben diese Relation R selbst. Um also eine Relation R auf sinnvolle Weise mit sich selbst zu verbinden, muss man R mindestens auf einer Seite des Verbundoperators umbenennen. Wir wollen hier eine Schreibweise vorschlagen, welche die obige Formulierung des Relationenprodukts etwas verallgemeinert.

Definition 2.48 (Potenzen von Relationen). *Seien S und T Relation zu einem Relationenschema $\mathcal{T} : \mathcal{A} \rightarrow \mathcal{D}$ und Θ ein Prädikat zu \mathcal{T} , welches aus verschiedenen Ausdrücken der Form „ $\mathbf{b} = \mathbf{a}$ “ mit Attributsymbolen a und b besteht, die jeweils durch „und“ verknüpft sind und in dem jedes Attributsymbol a höchstens einmal entweder auf der linken oder auf der rechten Seite einer solchen Gleichheit vorkommt. Für jedes Vorkommen von „ $\mathbf{b} = \mathbf{a}$ “ in Θ gelte zudem $\text{dom}_{\mathcal{T}}(a) = \text{dom}_{\mathcal{T}}(b)$. Dann ist \mathcal{A} überdeckt durch \mathcal{L} , die Menge der Attribute, die nicht auf der linken Seite einer der Gleichheiten in Θ auftauchen, und der entsprechenden Menge \mathcal{R} für die jeweils rechten Seiten. Wir definieren ein Prädikat $L\Theta R$, indem wir alle in Θ vorkommenden $\mathbf{b} = \mathbf{a}$ durch $L.b = R.a$ ersetzen und für alle nicht in Θ vorkommenden Attribute s jeweils „ und $L.s = R.s$ “ hinzufügen.*

Dann ist das Produkt von S und T bezüglich Θ

$$S \circ_{\Theta} T := \pi_{\mathcal{A}} \circ \beta_{\{a \leftarrow L.a \mid a \in \mathcal{L} \setminus \mathcal{R}\}} \circ \beta_{\{b \leftarrow R.b \mid b \in \mathcal{R}\}} (\beta_L(S) \bowtie_{L\Theta R} \beta_R(T))$$

ebenfalls eine Relation zum Relationenschema \mathcal{T} . Mit $U_{\mathcal{T}} = (\mathcal{T}, \text{TUP}(\mathcal{T}))$ als Universalrelation zu \mathcal{T} definieren wir die Potenz von T bezüglich Θ :

$$\begin{aligned} T_{\Theta}^0 &:= \sigma_{\Theta}(U) \\ T_{\Theta}^1 &:= T \\ T_{\Theta}^{n+1} &:= T \circ_{\Theta} T_{\Theta}^n \\ T_{\Theta}^+ &:= \bigcup_{n \in \mathbb{N}} T_{\Theta}^{n+1} \\ T_{\Theta}^* &:= \bigcup_{n \in \mathbb{N}} T_{\Theta}^n \end{aligned}$$

als Vorschlag für eine mögliche Schreibweise für diese Operatoren.

MAIER weist in [Mai83, S.527ff] nach, dass die transitive Hülle einer Relation im Allgemeinen nicht mit einem Ausdruck in relationaler Algebra berechnet werden kann, wenn darin nur die Operatoren Selektion, Projektion, Umbenennung und Verbund bzw. kartesisches Produkt, Vereinigung, Schnitt und Rest und konstante Relationen, Tupel und Attributwerte vorkommen. Die Menge dieser Operatoren wird als „relationale Algebra“ bezeichnet. Wir wollen eine relationale Abfragesprache, die mindestens diese Operatoren hat, als *eine* relationale Algebra bezeichnen, um anzudeuten, dass es weitere mächtigere relationale Algebren geben kann. Die transitive Hülle ist z.B. eine echte Erweiterung dieser Operatoren, die dann „relationale Algebra mit transitiver Hülle“ heißen soll. CODD betrachtet die transitive Hülle bereits als Teil *der* relationalen Algebra, bezeichnet diese wahlweise als „rekursiven Verbund“ [Cod90, S.140ff] oder als CLOSE [Cod79, S.426f].

Wir werden später sehen, dass zum Modellieren topologischer Informationen auf die transitive Hülle im Allgemeinen nicht verzichtet werden kann. Dies gilt jedoch nur für topologisches Modellieren in aller Allgemeinheit, nicht jedoch im besonderen Fall der Modellierung architektonischer Räume, bei der wir meist auch ohne transitive Hülle auskommen werden.

2.8.3 Relationaler Entwurf

Bisher haben wir beliebige Relationen zu Relationenschemata erlaubt und für diese die Operatoren definiert. In der Praxis ist es jedoch zusätzlich notwendig, gewisse Einschränkungen für Relationen festzulegen. Der Grund für derartige Einschränkungen ist, dass diese Relationen Informationen über die reale Welt wiedergeben. Eine Menge von Relationen bezeichnet man als *konsistent*, wenn die in einer Relation gespeicherten Informationen in der realen Welt auch möglich sind. So ist eine Relation, in der eine Person mehrere (natürliche) Väter hat nicht konsistent.

Es gibt nun eine Hierarchie von Einschränkungen für relationale Datenbanken, die als *Normalformen* bezeichnet werden.

Die erste und grundlegende Einschränkung ergibt sich aus praktischen Überlegungen zur Verwirklichung von relationalen Datenbanksystemen. Es ist zwar durchaus denkbar, dass z.B. der Wert eines Attributs **kinder** eines Datensatzes **herrX** in einer Tabelle **Angestellter** eine Menge ist — die Menge der Kinder von **herrX**. CODD zeigt jedoch, dass auf mengen- und tupelwertige Domänen stets verzichtet werden kann [Cod83]. Dieser Verzicht heißt erste Normalform:

Definition 2.49 (erste Normalform). *Eine Relation R zum Schema $\mathcal{R} : \mathcal{A} \rightarrow \mathcal{D}$ ist in erster Normalform (oder NF1), wenn von jedem Tupel t in R und jedem Attribut a in \mathcal{A} der Wert $t(a)$ atomar ist, also durch Abfragen nicht weiter in Komponenten zulegt werden kann. [Cod90, S.6]*

In den gängigen Datenbankmanagementsystemen gibt es von vornherein keine tupel- oder mengenwertigen Domänen. Dennoch können sich in der Praxis Verstöße gegen die erste Normalform einschleichen. So kann ein Attribut **name** etwa den Wert „Pavel, Alexandroff“ haben. Dieser Attributwert könnte nun von einem Anwendungsprogramm als Tupel mit den Attributen **vorname** und **nachname** verwendet werden, wäre also im Kontext des Anwendungsprogramms nicht mehr atomar. Beliebte sind auch „Codierungen“ der Form RN20061205-112 mit der Bedeutung „Rechnung Nummer 112 vom 5. Dezember 2006“, und bilden damit streng genommen ein Tupel mit den Attributen **rechnungsnummer** und **datum**. Derartige Entwurfsmuster sind jedoch nur dann als Verstöße gegen NF1 zu werten, wenn eine Datenbankanwendung Attributwerte als Tupel „missbraucht“. Zum Beispiel kann eine Rechnungsnummer nach obigen Schema auch nur als Hilfestellung für den Benutzer gedacht sein. In einer Relation mit tupelwertigen Attributen können diese stets durch deren Attribute ersetzt werden. Im ersten Beispiel ersetzen wir **name** durch **vorname** und **nachname** und verteilen die Attributwerte für **name** entsprechend auf Vor- und Nachnamen.

Eine Relation mit mengenwertigen Attributen hingegen lässt sich stets in mehrere äquivalente Relationen mit atomaren Attributwerten aufteilen. Unterstellen wir z.B. für ein Datenbankmanagementsystem eine Domäne **PERSON** und die Möglichkeit, auch Mengen von Personen als Attributwerte zu speichern. Dann ist $\mathcal{P}(\text{PERSON})$ die Menge der Menge von Personen. Sei **Mitglied** Relation zum Schema

[id : PERSON, name : TEXT, kinder : $\mathcal{P}(\text{PERSON})$].

Dann ist folgende Aufteilung möglich:

MitgliedNF1[id : PERSON, name : TEXT]
MitgliedKinder[mitglied : PERSON, kind : PERSON]

Die Relation zu MitgliedNF1 ist einfach die entsprechende Projektion von Mitglied und für jedes Tupel $\langle n : \text{id}, w : \text{name}, K : \text{kinder} \rangle$ mit id n und Kindermenge K fügen wir die Datensätze $\{n\} \times K$ der Relation MitgliedKinder hinzu.

Für die weiteren Einschränkungen benötigen wir zunächst einige zusätzliche Grundbegriffe:

Funktionale Abhängigkeit und Schlüssel. Funktionale Abhängigkeit ist die präzise Definition einer Eigenschaft der Form „Wenn X bekannt ist, dann ist auch Y bekannt“. Ein Schlüssel in einer Tabelle hat eben diese Eigenschaft: Wenn der Wert für einen Schlüssel vorliegt, dann ist dies ausreichend, um den Datensatz auch aufzufinden. Dann ist aber der Wert aller Attribute dieses Datensatzes bekannt.

Definition 2.50 (Funktionale Abhängigkeit). Sei $\mathcal{R} : \mathcal{A} \rightarrow \mathcal{D}$ ein Relationenschema, R Relation zu \mathcal{R} und seien \mathcal{X} und \mathcal{Y} Teilmengen von \mathcal{A} . Dann heißt \mathcal{Y} funktional abhängig von \mathcal{X} in R , wenn für alle Tupel s und t in R gilt:

$$\text{Wenn } \pi_{\mathcal{X}}(s) = \pi_{\mathcal{X}}(t), \text{ dann auch } \pi_{\mathcal{Y}}(s) = \pi_{\mathcal{Y}}(t).$$

In Zeichen schreiben wir dann $\mathcal{X} \xrightarrow{\text{Fd}_R} \mathcal{Y}$.

Wichtig ist, dass funktionale Abhängigkeit im Allgemeinen eine Eigenschaft von Relationen ist. Eine Relation R zu einem Schema $\mathcal{R} : \mathcal{A} \rightarrow \mathcal{D}$ definiert also immer eine Relation $\text{Fd}_R \subseteq \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{A})$ in der Potenzmenge der Attributmengen. Mit funktionaler Abhängigkeit wird der Begriff des Schlüssels definiert:

Definition 2.51 (Superschlüssel, Schlüssel). Sei $\mathcal{R} : \mathcal{A} \rightarrow \mathcal{D}$ ein Relationenschema, R Relation zu \mathcal{R} und sei $\mathcal{X} \subseteq \mathcal{A}$ Teilmenge der Attribute. Dann heißt \mathcal{X} Superschlüssel in R , wenn \mathcal{A} funktional abhängig von \mathcal{X} ist. Ein Superschlüssel \mathcal{X} in R heißt Schlüssel in R , wenn \mathcal{X} minimal ist, wenn also keine echte Teilmenge von \mathcal{X} ein Superschlüssel in R ist.

Auch Schlüssel und Superschlüssel sind im Allgemeinen nur definiert, wenn die Relation bekannt ist. Eine Relation kann mehrere Schlüssel haben. So bildet z.B. für die Relation $\{(10, 20), (20, 40)\}$ jedes Attribut $i \in \{1, 2\}$ einen Schlüssel $\{i\}$.

Nun können wir den Begriff des relationalen Schemas erweitern und zum Relationenschema \mathcal{R} noch eine Menge Fd von funktionale Abhängigkeiten explizit angeben. Eine Relation R ist dann Relation zu dem erweiterten Schema (\mathcal{R}, Fd) , wenn sie Relation zu \mathcal{R} ist und mindestens die angegebenen funktionalen Abhängigkeiten in Fd auch in R gelten:

$$\text{Fd} \subseteq \text{Fd}_R.$$

Anhand dieser Fd-Menge können wir nun die minimalen Attributmengen bestimmen, die Superschlüssel in jeder Relation R zu (\mathcal{R}, Fd) sind. Diese minimalen Attributmengen bezeichnen wir dann als *Schlüssel* für das Relationenschema (\mathcal{R}, Fd) . Auch hier sind mehrere Schlüssel möglich. In diesem Fall

wählen wir einen Schlüssel aus und bezeichnen ihn als *Primärschlüssel*. Alle anderen Schlüssel bezeichnen wir als *Schlüsselkandidaten*. Wir werden uns in dieser Arbeit auf die Angabe des Primärschlüssels beschränken, indem wir alle Primärschlüsselattribute unterstreichen. Folgende erweiterte Relationenschemata sind also gleich:

$$\mathcal{R}[\underline{id} : \mathbb{N}, name : Text] := (\mathcal{R}[id : \mathbb{N}, name : Text], \{ \{id\} \xrightarrow{Fd} \{id, name\} \}).$$

Bei Entwürfen komplexer Datenbankschemata kann es vorkommen, dass diese vereinfachende Notation nicht ausreicht. Derart komplexe Entwürfe sind in dieser Arbeit jedoch nicht zu erwarten.

Für die weiteren Normalformen sei auf die Literatur verwiesen. Sie sind für die vorliegende Arbeit ohne Relevanz. Anzumerken ist hier nur, dass bei Verletzung einer dieser Normalformen durch eine Relation R diese in der Regel in Relationen $\pi_{\mathcal{X}_1}(R), \dots, \pi_{\mathcal{X}_n}(R)$ aufgeteilt wird, die jeweils die Normalform nicht verletzen. Dabei überdeckt $\{\mathcal{X}_1, \dots, \mathcal{X}_n\}$ die Attributmenge von R so, dass für alle Relationen R zum erweiterten Schema gilt:

$$\pi_{\mathcal{X}_1}(R) \bowtie \dots \bowtie \pi_{\mathcal{X}_n}(R) = R.$$

Referentielle Integrität. Bisher haben wir immer nur eine einzelne Relation betrachtet. Die eben gezeigte Zerlegung in Projektionen zeigt jedoch bereits, dass mehrere Relationen im Zusammenhang betrachtet werden müssen.

Oft ist es notwendig, dass Attribute einer Relation auf Datensätze einer anderen Relation verweisen. Habe ich z.B. eine Relation **Rechnungsposition** mit einem Attribut **rechnungsnummer**, dann muss es zu jeder Rechnungsposition auch eine Rechnung mit entsprechender Rechnungsnummer geben.

In SQL werden referentielle Integritäten durch **foreign key** Klauseln angegeben. Für unser Beispiel etwa durch folgende **create table** Formeln:

```
create table Rechnung
  ( jahr:integer not null
  , ID:integer not null
  ...
  , primary key(jahr, ID)
  ...
  );
create table Rechnungsposition
  ( rJahr integer not null
  , rNummer integer not null
  , rPosition integer not null
  ...
  , primary key(rJahr, rNummer, rPosition)
  , foreign key(rJahr, rNummer) references Rechnung(jahr, ID)
  ...
  );
```

(unter Berücksichtigung der gegenwärtigen Anforderung der Finanzämter, die Nummerierung von Rechnungen in jedem Jahr neu zu beginnen.) Jeder Datensatz einer Rechnungsposition muss zu einer entsprechenden Rechnung gehören.

Dies entspricht der Forderung, dass jedes Tupel der Projektion

$$\Pi_{\text{rJahr, rNummer}}(\text{Rechnungsposition})$$

in der entsprechenden Projektion und Umbenennung

$$\beta_{(\text{rJahr, rNummer}) \leftarrow (\text{jahr, ID})} \circ \Pi_{\text{jahr, ID}}(\text{Rechnung})$$

liegen muss. Das Schlüsselwort **references** definiert eine Teilmenge

$$\mathcal{X} = \{\text{rJahr, rNummer}\}$$

der Attributsymbole einer Rechnungsnummer und eine Abbildung

$$\text{ref} : \mathcal{X} \rightarrow \mathcal{Y}, \begin{cases} \text{rJahr} & \mapsto \text{jahr} \\ \text{rNummer} & \mapsto \text{ID} \end{cases}$$

in die Menge \mathcal{Y} der Attributsymbole der Rechnungen.

Definition 2.52 (Referentielle Integrität). *Seien R und S Relationen mit Relationenschemata $\mathcal{R} : \mathcal{A} \rightarrow \mathcal{D}$ und $\mathcal{S} : \mathcal{B} \rightarrow \mathcal{D}$ und sei $\mathcal{X} \subseteq \mathcal{A}$. Eine referentielle Integrität von R nach S ist eine Abbildung $\text{ref} : \mathcal{X} \rightarrow \mathcal{B}$ mit $\mathcal{R}|_{\mathcal{X}} = \mathcal{S} \circ \text{ref}$ und der Eigenschaft $\pi_{\mathcal{X}}(R) \subseteq \delta_{\text{ref}}(S)$.*

Man könnte alternativ referentielle Integritäten als partielle Abbildungen $\text{ref} : \subseteq \mathcal{A} \rightarrow \mathcal{B}$ definieren. Dann ist \mathcal{X} der Definitionsbereich von ref .

Wenn wir es nun mit mehreren Relationen zu tun haben, kommen wir zum *Datenbankschema*: ein Tupel von erweiterten Relationenschemata und eine Menge von referentiellen Integritäten, die in der Datenbank vorkommen müssen:

Definition 2.53 (Datenbankschema). *Sei \mathcal{N} eine Menge von Tabellennamen und sei $((\mathcal{R}_T, \text{Fd}_T))_{T \in \mathcal{N}}$ eine Familie von erweiterten Relationenschemata jeweils mit Attributsymbolen \mathcal{A}_T . Diese Familie der Schemata zusammen mit einer Menge FK von partiellen Abbildungen*

$$\text{ref}_{T,S} : \subseteq \mathcal{A}_T \rightarrow \mathcal{A}_S,$$

die von den Attributsymbolen \mathcal{A}_T eines Relationenschemas zu den Attributsymbolen \mathcal{R}_S des anderen Schemas gehen, heißt *Datenbankschema*.

Eine *Datenbank* zu diesem Schema ist dann eine Familie von Relationen zu den erweiterten Schemata mit den Tabellennamen \mathcal{N} als Index, so dass alle Abbildungen $\text{ref}_{T,S}$ des Datenbankschemas entsprechend referentielle Integritäten in der Datenbank sind.

Unter einfachen Umständen wollen wir folgende Notation für referentielle Integritäten verwenden: Ist $\text{Rechnung}[\underline{\text{nr}} : \mathbb{Z}, \dots]$ ein Relationenschema mit Primärschlüssel $\{\text{nr}\}$ und weiteren Attributen, dann geben wir referentielle Integrität an, indem wir anstelle der Domäne das referenzierte Attribut angeben. Also ist

$$\text{Rechnungsposition}[\underline{\text{rnNr}} : \text{Rechnung.nr}, \underline{\text{posNr}} : \mathbb{N}, \dots]$$

ein Relationenschema Rechnungsposition mit Primärschlüssel $\{\text{rnNr}, \text{posNr}\}$ und referentieller Integrität $\text{ref}(\text{rnNr}) := \text{nr}$.

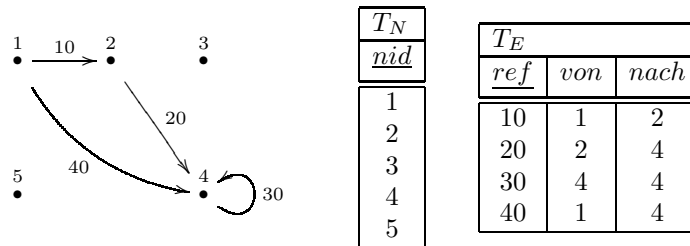
Beispiel 2.22 (Graph mit referentieller Integrität). Sei ein endlicher gerichteter Graph $(\text{Kanten}, \text{Knoten}, f : \text{Kanten} \rightarrow \text{Knoten})$ gegeben. Dann können wir ohne Einschränkung die Kanten und Knoten mit Zahlen identifizieren und erhalten ein Relationenschema $\text{Knoten}[\underline{\text{nid}} : \mathbb{Z}]$ für die Knoten. Für die Kanten nehmen wir den Graph von f mit Relationenschema $\text{Kanten}[\underline{\text{eid}} : \mathbb{Z}, \text{von} : \text{Knoten.id}, \text{nach} : \text{Knoten.id}]$, also ein Relationenschema mit Primärschlüssel $\{\text{eid}\}$ und referentieller Integrität

$$\text{ref} : \{\text{von}, \text{nach}\} \rightarrow \{\text{nid}\}.$$

Dann ist $((\text{Knoten}, \text{Kanten}), \{\text{ref}\})$ ein Datenbankschema und für jeden endlichen gerichteten Graph (N, E, f) gibt es eine Datenbank (T_N, T_E) zu diesem Schema, so dass der Graph (T_N, T_E, F) ein zu (N, E, f) isomorpher Graph ist. Dabei ist F die von T_E erzeugte Abbildung:

$$F : T_E \rightarrow T_N \times T_N, e \mapsto F(e) := (a, b) \text{ mit } a.\text{nid} = e.\text{von} \text{ und } b.\text{nid} = e.\text{nach}.$$

Die referentielle Integrität stellt dabei sicher, dass diese Abbildung sinnvoll definiert ist. Unser Beispielgraph hat dann folgende Realisierung als relationale Datenbank:



Damit stehen die formalen Grundlagen bereit, die nötig sind, um topologische Eigenschaften architektonischer Räume zu bestimmen und diese im Computer zu modellieren.

Kapitel 3

Die Topologischen Postulate

„Folglich sind alle mathematischen Begriffe für sich nicht Erkenntnisse, außer, sofern man voraussetzt, daß es Dinge gibt, die sich nur der Form jener reinen sinnlichen Anschauung gemäß uns darstellen lassen.“

KANT – Kritik der reinen Vernunft [Kan87, B-147]

Nach der Vorstellung der mathematischen Grundlagen wollen wir nun zum eigentlichen Teil dieser Arbeit übergehen: Modellierung von Architektur im Computer auf der Grundlage ihrer topologischen Eigenschaften. Dazu müssen wir natürlich zuerst voraussetzen, dass Architektur in der Tat topologische Eigenschaften hat.

Bisher waren alle unsere Überlegungen nur formaler, mathematischer Natur und nur gelegentlich durch Beispiele aus dem Bereich Architektur illustriert. Architektur ist aber ein Phänomen der realen Welt. Daher haben wir grundsätzlich zu überlegen, in welcher Beziehung die Realität „Architektur“ mit rein gedanklichen Konstruktionen wie „Topologie“ oder „relationale Datenbanken“ steht.

Dass die Realität auf irgendeine Art mathematische Eigenschaften besitzt ist zwar die wichtigste Rechtfertigung, warum Mathematik überhaupt betrieben wird, kann aber mit mathematischen Mitteln allein nicht bewiesen werden. Die Klärung dieser Frage ist Aufgabe der Philosophie, und wird unter anderem in [Kan87] behandelt. Wir wollen uns hier nur darauf beschränken, eine plausible Begründung des Zusammenhangs von Topologie und Architektur zu geben, und werden nun unsere weitere Arbeit auf zwei derart unbewiesene Behauptungen aufbauen. Wir bezeichnen diese als topologische Postulate für die Architektur.

3.1 Die Architektonische Zerlegung

Wir betrachten Gebäude in dieser Arbeit als eine Ansammlung von endlich vielen Gebäudeelementen im euklidischen Raum. Dies sind insbesondere alle Bauteile eines Gebäudes. Der Raum wird durch diese Bauteile in Raunteile untergliedert, wobei wir diese wiederum selbst als verallgemeinerte Gebäudeelemente ansehen. Einer dieser Raunteile ist der Außenraum, gegliedert in Luftvolumen, Erdreich und Gelände. Alle Gebäudeelemente zusammen füllen zudem den gesamten Raum aus.

Sei \mathcal{A} die Menge dieser endlich vielen Gebäudeelemente und für jedes Element $e \in \mathcal{A}$ haben wir den Raumteil $f(e) \subseteq \mathbb{R}^n$, die *Form von e* , die das Element im n -dimensionalen euklidischen Raum \mathbb{R}^n einnimmt. Wir wollen die Dimension n nicht beschränken, denn einerseits können wir für $n = 2$ etwa die Regionalplanung betrachten und $n = 4$ erlaubt auch die Betrachtung von Bauprozessen in der vierdimensionalen Raumzeit. Durch die Vereinbarung, dass die Gebäudeelemente zusammen den Raum ausfüllen, ist die Menge

$$f(\mathcal{A}) := \{f(e) \mid e \in \mathcal{A}\}$$

eine Überdeckung des \mathbb{R}^n , das heißt $\bigcup_{e \in \mathcal{A}} f(e) = \mathbb{R}^n$.

Wir wollen zudem vereinbaren, dass je zwei verschiedene Gebäudeelemente keinen gemeinsamen Punkt im euklidischen Raum haben. Es soll also für alle $d, e \in \mathcal{A}$ gelten: Ist $d \neq e$ dann ist $f(d) \cap f(e) = \emptyset$. Damit ist $f[\mathcal{A}]$ eine Zerlegung des euklidischen Raums, also eine Quotientenmenge \mathbb{R}^n/\sim zu einer Äquivalenzrelation \sim . Wie wir gesehen haben, kann damit auch eine Topologie für \mathcal{A} definiert werden: Eine Menge $A \subseteq \mathcal{A}$ ist genau dann offen, wenn $f(A)$ offen im Quotientenraum \mathbb{R}^n/\sim ist, wenn also $\bigcup_{a \in A} f(a)$ offen in \mathbb{R}^n ist:

Postulat 3.1 (architektonische Zerlegung). *Jedem Bauwerk entspricht eine endliche Menge von Bauwerkselementen \mathcal{A} , und deren Lage im euklidischen Raum in der Form einer surjektiven Abbildung $i : \mathbb{R}^n \rightarrow \mathcal{A}$, also einer Abbildung, die jedem Punkt p in \mathbb{R}^n genau ein Element $i(p) \in \mathcal{A}$ zuordnet und durch die jedes Element $e \in \mathcal{A}$ mindestens einem Punkt $q \in \mathbb{R}^n$ mit $i(q) = e$ zugeordnet ist.*

Wir bezeichnen Paar (\mathcal{A}, i) in dieser Arbeit als *architektonischen Raum*. Wir interessieren uns hier aber nur für die Elementmenge \mathcal{A} mit der durch i definierten Topologie. Alle weiteren Eigenschaften von architektonischen Räumen — insbesondere die künstlerischen — sind nicht Gegenstand dieser Arbeit. Wer nun mit dem Bezeichner „architektonischer Raum“ zu viele Affekte verbindet, kann stattdessen auch *architektonische Zerlegung* lesen.

Durch (\mathcal{A}, i) haben wir in der Tat eine Äquivalenzrelation \sim_i im euklidischen Raum, definiert durch $p \sim_i q$ genau dann, wenn $i(p) = i(q)$, und somit eine Zerlegung des euklidischen Raums in den Quotientenraum \mathbb{R}^n/\sim_i mit seiner Quotiententopologie. Damit können wir für \mathcal{A} eine Topologie \mathcal{T}_i definieren, die durch die Abbildung i eindeutig bestimmt ist und als *Finaltopologie* von i bezeichnet wird. Der Raum $(\mathcal{A}, \mathcal{T}_i)$ ist insbesondere homöomorph zum Quotientenraum \mathbb{R}^n/\sim_i . Insbesondere sind beide Räume endlich und damit alexandroffsch.

3.2 Die Zerlegung zu einem Komplex

Die obige Abbildung, mit der die Lage einer endlichen Menge im euklidischen Raum festgelegt wird, ist alleine noch zu allgemein für eine Beschreibung der Topologie von Gebäuden. Die Lage der architektonischen Elemente soll zudem noch gewisse Eigenschaften haben. Wir behaupten, dass ein Gebäude prinzipiell als endlicher Komplex dargestellt werden kann. Mit prinzipiell meinen wir, dass jedes Gebäudeelement nicht nur eine Lage und Ausdehnung im \mathbb{R}^n hat, sondern sogar in einem endlichen CW-Komplex (\mathbb{R}^n, \sim) nach Definition 2.32 liegt.

Postulat 3.2 (Architektonischer Komplex). *Jede architektonische Zerlegung (\mathbb{R}^n, \sim) hat eine Verfeinerung (\mathbb{R}^n, \simeq) zu einem endlichen CW-Komplex. Das heißt \simeq ist eine Äquivalenzrelation, \mathbb{R}^n/\simeq ist endlich und für alle $a \simeq b$ gilt auch $a \sim b$ und für jeden Punkt p in \mathbb{R}^n ist dessen Äquivalenzklasse $[p] \in (\mathbb{R}^n/\simeq)$ eine m -Zelle mit $m \leq n$.*

Damit ist \mathbb{R}^n/\sim im Wesentlichen (also „homöomorph zu einem“) Quotientenraum eines CW-Komplexes (\mathbb{R}^n, \simeq) .

Diese Postulate rechtfertigen es, Gebäude als endliche topologische Räume zu betrachten und bei Bedarf zu CW-Komplexen zu verfeinern. Die Endlichkeit ist natürlich Voraussetzung, um derartige Räume im Computer zu speichern.

Nun ist es auch möglich, dass ein und dieselbe Architektur auf verschiedenen Maßstabsebenen betrachtet wird. Dann ordnen wir jeder Maßstabsebene eine eigene architektonische Zerlegung zu, und wir setzen je zwei derartige Zerlegungen durch gewisse stetige Abbildungen in Beziehung. Unter einer solchen Detaillierungsabbildung sollen sich nun der detaillierte und der weniger detaillierte Raum weitgehend, bis auf die weggelassenen Details, ähneln. Dazu werden wir eine spezielle Klasse von stetigen Abbildungen, die monotonen Abbildungen, kennenlernen und auch sehen, wie die feinere Zerlegung auf einer detaillieren Maßstabsebene systematisch aus der weniger detaillierten gröberen Zerlegung gewonnen werden kann, wenn sich die Details wiederholen. Damit müssen diese Details nicht jedesmal explizit gespeichert werden.

Kapitel 4

Topologische Datenbanken

Wir haben aus Gebäuden endliche topologische Räume gemacht und kennen relationale Datenbanken. Nun soll gezeigt werden, wie topologische Informationen in relationalen Datenbanken gespeichert und abgefragt werden können. Wir hätten zwar auch objektorientierte Modelle oder STEP verwenden können, bevorzugen hier jedoch das relationale Datenmodell. Die Übertragung der hier vorgestellten Konzepte auf das objektorientierte Modell ist in naheliegender Weise einfach zu bewerkstelligen und wird hier nicht weiter diskutiert. Es ist zudem leichter, mit einer objektorientierten Sprache auf Daten einer relationalen Datenbank zuzugreifen, als umgekehrt, objektorientierte Datenstrukturen in relationalen Datenbanken zu speichern¹.

4.1 Das provisorische Modell $DKetKomp^{\text{tot}}$

Erinnern wir uns, dass ein Kettenkomplex \mathcal{C} eine Familie von \mathbb{Z} -Moduln ist, die durch lineare Abbildungen in folgender Weise verbunden sind:

$$\mathcal{C} : \dots \xrightarrow{\partial_{n+1}} C_n \xrightarrow{\partial_n} C_{n-1} \xrightarrow{\partial_{n-1}} \dots \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} \dots$$

Wir erinnern uns, dass wir hier nur die *freien* \mathbb{Z} -Moduln betrachten (S.63). Wir setzen zudem für den Komplex endliche Dimension n voraus. Dann sind alle C_i trivial (also $C_i = \{0\}$), wenn $i > n$ oder $0 > i$ gilt. Zudem setzen wir voraus, dass die C_i selbst ebenfalls von endlicher Dimension sind. Dann haben wir für jedes C_i eine endliche Basis E_i und wir können ohne Einschränkung annehmen, dass jedes E_i eine Tabelle (Relation) einer relationalen Datenbank ist. Für jedes i mit $0 \leq i \leq n$ benötigen wir dazu nur eine Tabelle $\mathbf{E}i$ mit Schema $[\mathbf{id} : \mathbb{Z}]$ und haben damit alle nötigen Informationen gespeichert, um die C_i erzeugen zu können.

Nun wollen wir die linearen Abbildungen $\partial_i : C_i \rightarrow C_{i-1}$ ebenfalls in einer relationalen Datenbank speichern:

Definition 4.1 (Matrix, Matrixprodukt). *Seien A und B endliche Mengen. Eine (ganzzahlige) $A \times B$ -Matrix M ist eine Abbildung*

$$M : A \times B \rightarrow \mathbb{Z}, (a, b) \mapsto M_{a,b}.$$

¹Nicht zuletzt deshalb, weil die „Erfinder“ dieser naheliegenden Idee nicht vor Patentklagen zurückschrecken.

Für Matrizen $M : A \times B \rightarrow \mathbb{Z}$ und $N : B \times C \rightarrow \mathbb{Z}$ ist die Matrix

$$M \cdot N : A \times C \rightarrow \mathbb{Z}, (a, c) \mapsto \sum_{b \in B} M_{a,b} \cdot N_{b,c}$$

das (Matrix-)produkt von M und N .

Die Menge A heißt dann *Zeilenindex* und B heißt *Spaltenindex* von M . Für jedes $a \in A$ heißt $M|_{\{a\} \times B}$ die *a-te Zeile* von M und für jedes $b \in B$ ist $M|_{A \times \{b\}}$ die *b-te Spalte* von M . Ist $M : A \times B \rightarrow \mathbb{Z}$ eine Matrix, dann ist $M^T : B \times A \rightarrow \mathbb{Z}, (b, a) \mapsto M_{a,b}$ die *transponierte Matrix* von M , also die Matrix, die aus M durch Vertauschen von Zeilen und Spalten entsteht. Eine konstante Abbildung $c_0 : A \times B \rightarrow \mathbb{Z}, (a, b) \mapsto 0$ heißt *Nullmatrix*.

Die Endlichkeitsbedingung in der Definition könnte man auch etwas abschwächen und eine Abbildung $M : A \times B \rightarrow \mathbb{Z}$ für beliebige Mengen A und B als Matrix bezeichnen, wenn in jeder Zeile und in jeder Spalte nur endlich viele von 0 verschiedene Einträge stehen.

Jede $A \times B$ -Matrix M ist als Relation isomorph in \mathcal{RelDb} zu einer Relation $M_{\mathcal{M}}$ zum Schema $\mathcal{M}[\underline{a} : A, \underline{b} : B, m : \mathbb{Z}]$ und somit haben wir sofort ein Relationenschema für Matrizen. Nun ist allgemein bekannt, dass jede lineare Abbildung $\partial_i : C_i \rightarrow C_{i-1}$ bezüglich einer Basis E_i von C_i und einer Basis E_{i-1} von C_{i-1} eine eindeutig bestimmte Darstellungsmatrix $M(\partial_i) : E_i \times E_{i-1} \rightarrow \mathbb{Z}$ hat.

Kategorie der Matrizen. Die $A \times B$ -Matrizen sind, zusammen mit dem Matrixprodukt als Komposition, Morphismen einer Kategorie, die wir als $\mathit{Matr}_{\mathbb{Z}}$ oder kurz als Matr bezeichnen wollen. Die Objekte sind die (endlichen) Mengen und die Identität für eine Menge A ist die Einheitsmatrix

$$\mathbb{1}_A : A \times A \rightarrow \mathbb{Z}, (a, b) \mapsto \begin{cases} 1 & : a = b \\ 0 & : a \neq b \end{cases}$$

zur Menge A . Diese Abbildung $\mathbb{1}(a, b)$ wird oft auch mit dem Kronecker-Symbol δ_{ab} geschrieben.

Als Komposition der Morphismen ist das Matrixprodukt in der Tat geeignet:

Satz 4.1 (Eigenschaften des Matrixprodukts). *Seien A, B und C Mengen und $M : A \times B \rightarrow \mathbb{Z}$ und $N : B \times C \rightarrow \mathbb{Z}$ Matrizen. Dann gilt immer*

1. $M \cdot N : A \times C \rightarrow \mathbb{Z}$ ist ebenfalls eine Matrix.
2. Das Matrixprodukt ist assoziativ: $(M \cdot N) \cdot K = M \cdot (N \cdot K)$ (für eine weitere Matrix $K : C \times D \rightarrow \mathbb{Z}$).
3. $M \cdot \mathbb{1}_B = M = \mathbb{1}_A \cdot M$

Den Beweis findet man in den meisten Lehrbüchern über lineare Algebra.

Nun können wir die Kategorie Matr der endlichen Mengen mit den Matrizen als Morphismen als praktisch als äquivalent zur Kategorie $\mathbb{Z}\mathit{Mod}$ der freien \mathbb{Z} -Moduln, mit den linearen Abbildungen als Morphismen betrachten. Der Beweis soll hier nur skizziert werden:

Kleine Unterkategorien Sei \mathcal{M} eine beliebige Teilmenge von Objekten aus $\mathbb{Z}\mathit{Mod}$ und \mathcal{B} die Menge aller Basen für die Elemente aus \mathcal{M} .

Freie \mathbb{Z} -Modul zu Matrizen. Es existiert ein Funktor mit einer Abbildung $F : \mathcal{M} \rightarrow \mathcal{B}$, die für jeden freien \mathbb{Z} -Modul $M \in \mathcal{M}$ eine Basis B_M auswählt und jeder linearen Abbildung $f : M \rightarrow N$ die Matrix

$$F_f : B_M \times B_N \rightarrow \mathbb{Z}, (m, n) \mapsto \text{die } n\text{-Koordinate vom } f(m) \text{ bezügl. } B_N$$

zuordnet. Wegen dieser *Auswahl* einer Basis ist es notwendig, sich hier auf eine *kleine* Kategorie zu beschränken.

Matrizen zu freien \mathbb{Z} -Moduln. Als Funktor wählen wir $G : \text{Matr} \rightarrow \mathbb{Z}\text{Mod}$, der jeder Menge B den freien \mathbb{Z} -Modul $\mathbb{Z}B$ und jeder $A \times B$ -Matrix $M : A \times B \rightarrow \mathbb{Z}$ die lineare Abbildung

$$G(M) : \mathbb{Z}A \rightarrow \mathbb{Z}B, \sum_{a \in A} \lambda_a \cdot a \mapsto \sum_{a \in A} \lambda_a \cdot \sum_{b \in B} M_{a,b} \cdot b$$

zuordnet.

Dann sind F und G tatsächlich Funktoren und es gilt $GFA \cong A$ als auch $FGB \cong B$ für alle entsprechenden \mathbb{Z} -Moduln A und Mengen B mit $\{GB, A\} \subseteq \mathcal{M}$.

Dies funktioniert zwar nur bei *kleinen* Kategorien, da „Auswählen“ bei Klassen im Allgemeinen nicht erlaubt ist. Damit ist aber jede kleine Unterkategorie von $\mathbb{Z}\text{Mod}$ äquivalent zu einer kleinen Unterkategorie von Matr und umgekehrt, was man auch als „lokale“ oder „praktische“ Äquivalenz von Kategorien bezeichnen könnte. Alternativ lässt sich anstelle von $\mathbb{Z}\text{Mod}$ die Kategorie der *markierten* \mathbb{Z} -Moduln verwenden. Deren Objekte sind Paare (M, B) , bestehend aus einem freien \mathbb{Z} -Modul M und einer Basis B von M . Der (entsprechend angepasste) Funktor F ist dann bei den Objekten einfach die Projektion von (M, B) auf B , hat entsprechend angepassten Umkehrfunktorkomplex G und muss nicht mehr auswählen.

Damit ist dann auch die Komposition von linearen Abbildungen gleichbedeutend mit dem Matrixprodukt ihrer Darstellungsmatrizen und die Komposition zweier linearer Abbildungen genau dann die konstante Nullabbildung, wenn das Matrixprodukt ihrer Darstellungsmatrizen eine Nullmatrix ist.

Matrizen in relationalen Datenbanken. Zeigen wir nun etwas ausführlicher, wie eine Matrix in einer relationalen Datenbank gespeichert wird. Eine Matrix $M : A \times B \rightarrow \mathbb{Z}$ ist eine Abbildung und damit auch eine Relation $M = ((A \times B, \mathbb{Z}), G_M)$. Der Graph G_M dieser Abbildung hat die Form $\{(a, b), M_{a,b} \mid a \in A, b \in B\}$ und ist damit als Relation nicht in erster Normalform, denn die erste Komponente (a, b) der Datensätze ist immer ein Tupel. Entfernen wir die inneren Klammern, dann erhalten wir ein Relationenschema $\mathcal{M}[\underline{a} : A, \underline{b} : B, \underline{x} : \mathbb{Z}]$ und können die Matrix als Relation in erster Normalform abspeichern als

$$M' := (\mathcal{M}, \{(i : \underline{a}, j : \underline{b}, M_{i,j} : \underline{x}) \mid a \in A, b \in B\})$$

Da M eine Abbildung ist, gilt in der Tat die funktionale Abhängigkeit

$$\{\underline{a}, \underline{b}\} \xrightarrow{\text{Fd}_{M'}} \{\underline{x}\}$$

und daher ist $\{\underline{a}, \underline{b}\}$ ein Schlüssel dieser Relation.

Beispiel 4.1 (Einheitsmatrix). Sei $A = \{1, 2\}$, dann ist die Einheitsmatrix zu A mit dem obigen relationalen Schema und in traditioneller Schreibweise.

$\mathbb{1}_A$		
<u>a</u>	<u>b</u>	<u>x</u>
1	1	1
1	2	0
2	1	0
2	2	1

$$\mathbb{1}_A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Matrizenmultiplikation. Die Multiplikation zweier Matrizen ist mit relationaler Algebra (mit Quotientenbildung) wie folgt möglich:

Seien $\mathcal{B}_m[\underline{B}_m.id : D, \dots]$, $\mathcal{B}_n[\underline{B}_n.id : D, \dots]$ und $\mathcal{B}_k[\underline{B}_k.id : D, \dots]$ Relationenschemata mit Relationen B_m zu \mathcal{B}_m , B_n zu \mathcal{B}_n und B_k zu \mathcal{B}_k . Sei zudem $\mathcal{M}_m[\underline{a} : \mathcal{B}_m.id, \underline{b} : \mathcal{B}_n.id, x : \mathbb{Z}]$ Relationenschema einer Matrix M_m und entsprechend $\mathcal{M}_n[\underline{c} : \mathcal{B}_n.id, \underline{d} : \mathcal{B}_k.id, y : \mathbb{Z}]$ Relationenschema von M_n .

Dann ergibt folgender Ausdruck in relationaler Algebra

$$\beta_{z \leftarrow \Sigma(x \cdot y)}(\pi_{a,d,\Sigma(x \cdot y)}((M_m \bowtie_{b=c} M_n) / \{a, d\}))$$

das Matrixprodukt $M_m \cdot M_n$. Dabei haben wir hier die Syntax für die Projektion stillschweigend um die Möglichkeit erweitert, auch auf algebraische Ausdrücke mit Attributsymbolen zu projizieren. Das Attribut „ $\Sigma(x \cdot y)$ “ ist dann die Summe der $s(x) \cdot s(y)$ für jedes Tupel s in einer Äquivalenzklasse $[t]$ der Quotientenrelation $(M_m \bowtie_{b=c} M_n) / \{a, d\}$.

Der obige Ausdruck entspricht der SQL-Abfrage

```
select  a, d, sum(x * y) as z
from    Mm inner join Mn on (b=c)
group by a, d;
```

Beispiel 4.2 (Multiplikation zweier Matrizen). Seien zum Beispiel die obigen Matrizen gegeben als

$$M_m = \begin{pmatrix} -3 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix} \quad M_n = \begin{pmatrix} 2 & -1 \\ 4 & -2 \\ -2 & 1 \end{pmatrix}$$

Dann sind diese als Relationen

M_m		
<u>a</u>	<u>b</u>	<u>x</u>
1	1	-3
1	2	2
1	3	1
2	1	0
2	2	1
2	3	2

M_n		
<u>c</u>	<u>d</u>	<u>y</u>
1	1	2
1	2	-1
2	1	4
2	2	-2
3	1	-2
3	2	1

Wir wollen nun den Ausdruck

$$\beta_{z \leftarrow \Sigma(x \cdot y)}(\pi_{a,d,\Sigma(x \cdot y)}((M_m \bowtie_{b=c} M_n) / \{a, d\}))$$

schrittweise berechnen. Zunächst der Verbund $M_m \bowtie_{b=c} M_n$ mit dem zusätzlichen Attribut „ $x \cdot y$ “:

$M_m \bowtie_{b=c} M_n$						
<u>a</u>	<u>b</u>	x	<u>c</u>	<u>d</u>	y	$x \cdot y$
1	1	-3	1	1	2	-6
1	1	-3	1	2	-1	3
1	2	2	2	1	4	8
1	2	2	2	2	-2	-4
1	3	1	3	1	-2	-2
1	3	1	3	2	1	1
2	1	0	1	1	2	0
2	1	0	1	2	-1	0
2	2	1	2	1	4	4
2	2	1	2	2	-2	-2
2	3	2	3	1	-2	-4
2	3	2	3	2	1	2

Diesen Verbund sortieren wir nun nach den Attributen **a** und **d**, trennen die Äquivalenzklassen von $(M_m \bowtie_{b=c} M_n)/\{\mathbf{a}, \mathbf{d}\}$ durch horizontale Linien voneinander ab und fügen jeweils die Summe $\Sigma(x \cdot y)$ über die Äquivalenzklasse als Attribut hinzu. Dann erhalten wir folgende Quotientenrelation:

$(M_m \bowtie_{b=c} M_n)/\{\mathbf{a}, \mathbf{d}\}$							
<u>a</u>	<u>b</u>	x	<u>c</u>	<u>d</u>	y	$x \cdot y$	$\Sigma(x \cdot y)$
1	1	-3	1	1	2	-6	0
1	2	2	2	1	4	8	
1	3	1	3	1	-2	-2	
1	1	-3	1	2	-1	3	0
1	2	2	2	2	-2	-4	
1	3	1	3	2	1	1	
2	1	0	1	1	2	0	0
2	2	1	2	1	4	4	
2	3	2	3	1	-2	-4	
2	1	0	1	2	-1	0	0
2	2	1	2	2	-2	-2	
2	3	2	3	2	1	2	

Durch Projektion auf $\{\mathbf{a}, \mathbf{d}, \Sigma(x \cdot y)\}$ und Umbenennung von $\Sigma(x \cdot y)$ in **z** erhalten wir

$$M_m \cdot M_n = \beta_{z \leftarrow \Sigma(x \cdot y)}(\pi_{\mathbf{a}, \mathbf{d}, \Sigma(x \cdot y)}((M_m \bowtie_{b=c} M_n)/\{\mathbf{a}, \mathbf{d}\}))$$

als Relation:

$M_m \cdot M_n$		
<u>a</u>	<u>d</u>	<u>z</u>
1	1	0
1	2	0
2	1	0
2	2	0

Dies ist offenbar eine Nullmatrix.

Sind nun M_m und M_n Matrizen der Randoperatoren ∂_i und ∂_{i-1} eines Kettenkomplexes, dann sind alle Matrixeinträge des Matrixprodukts gleich 0. Anders ausgedrückt ist dann die Selektion von Matrixeinträgen, die nicht Null sind (in SQL: `having sum(x * y) <> 0`), die leere Menge.

Das obige Beispiel zeigt bereits, wie ein Kettenkomplex als relationale Datenbank modelliert werden kann:

Definition 4.2 (totaler relationaler Kettenkomplex). Ein Tupel (oder eine „Sequenz“) $(B_n, D_n, \dots, D_1, B_0)$ von Mengen B_i für $i = 0, \dots, n$ und Relationen $D_j \subseteq B_j \times B_{j-1} \times \mathbb{Z}$ für $j = 1, \dots, n$, mit Schlüssel $\{1, 2\}$ und $\pi_{1,2}(D_j) = B_j \times B_{j-1}$ heißt totaler relationaler Kettenkomplex, wenn alle Matrixprodukte $D_k \cdot D_{k-1}$ für $k = 2, \dots, n$ Nullmatrizen sind.

Die Bezeichnung „total“ wählen wir, da die Matrizen totale Abbildungen sind und (vorerst) der Graph einer Matrix jeweils vollständig angegeben wird. Diese Definition ist nur formal zu verstehen und die naheliegende Implementierung als relationale Datenbank ist in dieser Form äußerst ineffizient, da die Matrizen meist dünn besetzt sind. Das heißt, die meisten Matrixeinträge sind Null und könnten im Prinzip auch weggelassen werden. Dann würden wir partielle Matrizen und eine effizientere Implementierung erhalten. Allerdings wollen wir erst später auf Effizienzbetrachtungen eingehen und uns vorerst damit begnügen, dass eine Realisierung von endlichen Komplexen mit einer relationalen Datenbank überhaupt existiert.

Zu den totalen relationalen Kettenkomplexen haben wir strukturerhaltende Abbildungen:

Definition 4.3 (totaler relationaler Komplexmorphismus). Ein totaler relationaler Komplexmorphismus $\mathcal{M} : \mathcal{B} \rightarrow \mathcal{K}$ von totalen relationalen Kettenkomplexen $\mathcal{B} = (B_n, D_n, \dots, D_1, B_0)$ und $\mathcal{K} = (K_n, R_n, \dots, R_1, K_0)$ ist ein Tupel $\mathcal{M} = (M_n, \dots, M_0)$ von (Graphen von) Matrizen $M_i : B_i \times K_i \rightarrow \mathbb{Z}$, so dass für alle $i = 1, \dots, n$ gilt:

$$D_i \cdot M_{i-1} = M_i \cdot R_i.$$

Das heißt für jedes $i = 1, \dots, n$ kommutiert das Diagramm:

$$\begin{array}{ccccc} \cdots & B_i & \xrightarrow{D_i} & B_{i-1} & \cdots \\ & \downarrow M_i & & \downarrow M_{i-1} & \\ \cdots & K_i & \xrightarrow{R_i} & K_{i-1} & \cdots \end{array}$$

\mathcal{M} ist mit der komponentenweisen Matrixmultiplikation

$$(M_n, \dots, M_0) \cdot (N_n, \dots, N_0) := (M_n \cdot N_n, \dots, M_0 \cdot N_0)$$

in der Tat ein Morphismus. Allerdings eignen sich diese Komplexmorphismen für den praktischen Einsatz wohl weniger, da die Einschränkung, Elemente stets Elementen gleicher Dimension zuzuordnen, ziemlich streng ist.

Sind die Sequenzen der Komplexe oder des Komplexmorphismus verschieden lang, dann werden sie entsprechend mit leeren Relationen und leeren Matrizen aufgefüllt. Die leeren Mengen sind dann jeweils Basis des trivialen 0-Moduls $\{0\}$

und die leeren Matrizen ($M_{\emptyset, K_i} : \emptyset \times K_i \rightarrow \mathbb{Z}$ bzw. $M_{B_i, \emptyset} : B_i \times \emptyset \rightarrow \mathbb{Z}$) sind die Darstellungsmatrizen der linearen Nullabbildungen $\{0\} \rightarrow \mathbb{Z}K_i$ und $\mathbb{Z}B_i \rightarrow \{0\}$.

Dann gilt, dass zwei totale relationale Kettenkomplexe genau dann isomorph sind, wenn ihre Kettenkomplexe isomorph sind. Wir bezeichnen die Kategorie der totalen relationalen Kettenkomplexe mit $\mathcal{DKetKomp}^{\text{tot}}$ und haben eine offensichtliche Äquivalenz der Kategorien $\mathcal{KetKomp}^{\text{fin}}$, der endlichen Kettenkomplexe, und $\mathcal{DKetKomp}^{\text{tot}}$.

4.1.1 Dynamische Dimension

Diejenigen Eigenschaften von Informationen in einem Anwendungsprogramm, die sich während der Laufzeit des Programms nicht ändern, bezeichnet man als *statisch*, die veränderlichen Eigenschaften hingegen als *dynamisch*. Statisch sind in der Regel die Datenstrukturen oder im Fall von Datenbankanwendungen (meist) die Relationenschemata. Die Daten selbst sind veränderlich, also dynamisch. Im obigen Modell der relationalen Kettenkomplexe haben wir ein Datenbankschema

$$(\mathcal{B}_n, \mathcal{D}_n, \dots, \mathcal{D}_1, \mathcal{B}_0),$$

mit einer eigenen Relation B_i für jede Dimension i , und für $i > 0$ jeweils eine Matrix D_i für die Randabbildung $\partial_i : \mathbb{Z}B_i \rightarrow \mathbb{Z}B_{i-1}$. Insbesondere ist hier die Dimension eines Komplexes zu diesem Schema stets durch n nach oben beschränkt.

Sind jedoch die B_i für $i = 0, \dots, n$ und die \mathcal{D}_j für $j = 1, \dots, n$ vereinigungsverträglich, dann ist es möglich, zu jeder i -Zelle in \mathcal{B}_i die Dimension i als Attributwert zu speichern: Das neue Relationenschema $\mathcal{B} := \mathcal{B}_n \cup [\underline{\text{dim}} : \mathbb{Z}]$ fügt einfach eine Spalte „dim“ hinzu, wobei wir noch festlegen wollen, dass dieses Attribut auch Schlüsselattribut von \mathcal{B} ist. Dann sind alle Relationen $B(i) := B_i \times \{(i : \underline{\text{dim}})\}$ disjunkt und in der Vereinigung $B = \bigcup_{i=0}^n B(i)$ ist der um $\underline{\text{dim}}$ erweiterte Schlüssel ein Superschlüssel.

Ganz analog verfahren wir mit den Randmatrizen D_i : Sei $\{\text{int}, \text{fr}\}$ Schlüssel in D_i , dann erweitern wir \mathcal{D}_n zu $\mathcal{D} := \mathcal{D}_n \cup [\underline{\text{dimInt}} : \mathbb{Z}, \underline{\text{dimFr}} : \mathbb{Z}]$ und jede Relation D_i wird zu $D(i) := D_i \times \{(i : \underline{\text{dimInt}}, i - 1 : \underline{\text{dimFr}})\}$. Deren Vereinigung ist dann $D' := \bigcup_{i=1}^n D(i)$, und wenn wir D' durch Auffüllen mit 0-Werten zu einer totalen $B \times B$ -Matrix D fortsetzen, dann erhalten wir eine Datenbank (B, D) mit Datenbankschema $(\mathcal{B}, \mathcal{D})$ für Komplexe ohne Beschränkung der Dimension. Auch hier nehmen wir vorerst noch in Kauf, dass dieses Datenmodell sogar noch ineffizienter ist als das vorhergehende.

4.1.2 Informationsverluste

Wir werden hier sehen, dass Kettenkomplexe nicht alle topologischen Informationen enthalten und zeigen, wie mehr topologische Information gespeichert werden kann.

Betrachten wir folgende zwei Graphen Γ_1 und Γ_2 als topologische Räume, wie oben bereits demonstriert:

$$\Gamma_1 : c \begin{array}{c} \circ \\ \curvearrowright \end{array} a \quad \bullet b \begin{array}{c} \circ \\ \curvearrowright \end{array} d \qquad \Gamma_2 : c \begin{array}{c} \circ \\ \curvearrowright \end{array} a \begin{array}{c} \circ \\ \curvearrowright \end{array} d \quad \bullet b$$

Dann ist in Γ_2 die Menge $\{b\}$ zugleich offen und abgeschlossen. Einen solchen Punkt bezeichnet man dann auch als *isolierten Punkt*. In Γ_1 gibt es jedoch

keinen isolierten Punkt. Daher können Γ_1 und Γ_2 als topologische Räume nicht homöomorph sein.

Dennoch sind ihre Kettenkomplexe isomorph. Die Moduln sind bei beiden Graphen jeweils $C_1 = \mathbb{Z}\{c, d\}$ und $C_0 = \mathbb{Z}\{a, b\}$. Der Randoperator $\partial_1 : C_1 \rightarrow C_0$ für Γ_1 ist $\partial_1(c) = a - a = 0$ und $\partial_1(d) = b - b = 0$ und somit ist ∂_1 die konstante Nullabbildung.

Ist $\delta_2 : C_1 \rightarrow C_0$ der Randoperator des Kettenkomplexes von Γ_2 , dann ist $\delta_2(c) = \delta_2(d) = a - a = 0$ ebenfalls die Nullabbildung. Damit sind beide Kettenkomplexe gleich, also isomorph, obwohl die beiden topologischen Räume nicht zueinander isomorph sind. Die Topologien lassen sich also nicht anhand der Komplexe unterscheiden, wir haben keine Äquivalenz von Kategorien und somit einen Informationsverlust.

4.2 Das Datenmodell für Komplexe $\mathcal{DKetKomp}$

Zur Handhabung der Informationsverluste wollen wir nun die partiellen Matrizen einführen. Sie eignen sich zudem, den Aufwand bei der Speicherung dünn besetzter Matrizen zu reduzieren. Ab hier tolerieren wir also nicht mehr die Ineffizienz der totalen relationalen Komplexe und suchen nach etwas Besserem:

Definition 4.4 (partielle Matrix, partielle Matrixmultiplikation). Eine partielle $A \times B$ -Matrix M ist eine partielle Abbildung $M : \subseteq A \times B \rightarrow \mathbb{Z}$. Für zwei partielle Matrizen $M : \subseteq A \times B \rightarrow \mathbb{Z}$ und $N : \subseteq B \times C \rightarrow \mathbb{Z}$ ist die partielle Abbildung

$$M \cdot N : \subseteq A \times C \rightarrow \mathbb{Z}$$

definiert durch

$$M \cdot N(a, c) := \sum_{\{(b, x, y) \mid ((a, b), x) \in M, ((b, c), y) \in N\}} x \cdot y$$

das partielle Matrixprodukt von M und N . Dabei setzen wir $\sum_{a \in \emptyset} f(a) := \uparrow$, das heißt, die Summe über die leere Menge ist nicht definiert.

Man beachte, dass b sowohl in $((a, b), x)$ als auch in $((b, c), y)$ auftaucht, dass also die Summe über den relationalen Gleichheitsverbund $M \bowtie_{M.b=N.b} N$ der Matrizen gebildet wird. Ist $M(a, b) = \uparrow$ oder $N(b, c) = \uparrow$ für ein $b \in B$, dann gibt es für dieses b auch kein $((a, b), x), ((b, c), y) \in M \times N$. Das hat nun die selbe Wirkung, wie wenn wir folgende Rechenregeln mit dem Symbol \uparrow vereinbart hätten:

$$\begin{aligned} \uparrow + x &= x = x + \uparrow \\ \uparrow \cdot x &= \uparrow = x \cdot \uparrow \end{aligned}$$

Damit gilt dann $0 \neq 0 \cdot \uparrow = \uparrow$. Das Symbol \uparrow könnte man als eine Art zweiter Null auffassen, die „nuller als Null ist“. Diese Rechenregeln unterscheiden sich übrigens von den für das relationale Datenmodell üblichen Rechenregeln mit NULL, einen Platzhalter für nicht definierte Werte in Tupeln². Dort ist im Gegensatz zu unseren Rechenregeln hier gefordert, dass jede Verknüpfung (Addition, Multiplikation) von NULL mit einem anderen Wert wiederum NULL ergibt.

²Wir erinnern uns an das OPTIONAL Schlüsselwort in STEP und den Platzhalter \$ für nicht definierte Werte

Ein Beispiel für eine partielle Matrix ist die partielle Einheitsmatrix

$$\mathbb{1}_{\subseteq, A} : \subseteq A \times A \rightarrow \mathbb{Z}, (a, b) \mapsto \begin{cases} 1 & : a = b \\ \uparrow & : a \neq b \end{cases} .$$

Die partiellen Matrizen sind die „platzsparenden“ Matrizen, aus denen sich überflüssige Nullen der dünn besetzten Matrizen entfernen lassen, indem wir $M_{\subseteq}(a, b) := \uparrow$ setzen, wenn für die totale Matrix $M(a, b) = 0$ gilt. Partielle Matrizen können jedoch noch mehr, wenn wir nicht alle Nullen entfernen. Zunächst können wir mit ihnen eine Kategorie bilden:

Satz 4.2 (partielle Matrizen als Morphismen). *Die partiellen Matrizen sind mit der partiellen Matrixmultiplikation als Komposition und der partiellen Einheitsmatrix als Identität Morphismen einer Kategorie, die wir hier als $\mathcal{P}\text{Matr}$ bezeichnen.*

Beweis. Zunächst beobachten wir, dass beim Rechnen mit \uparrow das Kommutativ-, das Distributiv- und das Assoziativgesetz eingehalten werden. Damit hätten wir das Produkt partieller Matrizen auch als gewöhnliches Matrixprodukt unter Berücksichtigung obiger Rechenregeln definieren können und dieses ist assoziativ. Wir hätten also einfach über die gesamte gemeinsame (Spalten- bzw. Zeilen-) Indexmenge aufsummieren können und uns nicht auf den Graph des Gleichheitsverbands beschränken müssen.

Sei $f : \subseteq X \rightarrow \mathbb{Z}$ partielle Abbildung und seien $a, b, c \in X$. Wenn f für jedes der a, b, c definiert ist, dann gelten die Rechenregeln selbstverständlich auch für $f(a)$, $f(b)$ und $f(c)$. Wir können also ohne Einschränkung annehmen, dass f an einer der drei Stellen nicht definiert ist.

Kommutativgesetz. Ist klar aus den obigen Rechenregeln.

Assoziativgesetz. Ist eines von $f(a)$, $f(b)$ oder $f(c)$ nicht definiert, dann ist deren Produkt gleich \uparrow . Bei der Addition gilt nun für den Fall $f(c) = \uparrow$

$$\begin{aligned} (f(a) + f(b)) + f(c) &= (f(a) + f(b)) + \uparrow \\ &= f(a) + f(b) \\ &= f(a) + (f(b) + \uparrow) \\ &= f(a) + (f(b) + f(c)). \end{aligned}$$

Ist $f(b) = \uparrow$, dann gilt:

$$\begin{aligned} (f(a) + f(b)) + f(c) &= (f(a) + \uparrow) + f(c) \\ &= f(a) + f(c) \\ &= f(a) + (\uparrow + f(c)) \\ &= f(a) + (f(b) + f(c)). \end{aligned}$$

Ist $f(a) = \uparrow$, dann gilt

$$\begin{aligned} (f(a) + f(b)) + f(c) &= (\uparrow + f(b)) + f(c) \\ &= f(b) + f(c) \\ &= \uparrow + (f(b) + f(c)) \\ &= f(a) + (f(b) + f(c)). \end{aligned}$$

Distributivgesetz. Ist $f(c) = \uparrow$, dann gilt

$$\begin{aligned}
 f(a) \cdot (f(b)) + f(c) &= f(a) \cdot (f(b) + \uparrow) \\
 &= f(a) \cdot f(b) \\
 &= f(a) \cdot f(b) + \uparrow \\
 &= f(a) \cdot f(b) + \uparrow \\
 &= f(a) \cdot f(b) + f(a) \cdot \uparrow \\
 &= f(a) \cdot f(b) + f(a) \cdot f(c).
 \end{aligned}$$

Für $f(b) = \uparrow$ nutzen wir das Kommutativgesetz aus und zeigen die Distributivität wie oben. Ist $f(a) = \uparrow$, dann gilt

$$\begin{aligned}
 f(a) \cdot (f(b)) + f(c) &= \uparrow \cdot (f(b) + f(c)) \\
 &= \uparrow \\
 &= \uparrow + \uparrow \\
 &= \uparrow \cdot f(b) + \uparrow \cdot f(c) \\
 &= f(a) \cdot f(b) + f(a) \cdot f(c).
 \end{aligned}$$

Seien nun W, X, Y und Z endliche Mengen und A, B, C partielle Matrizen $A : \subseteq W \times X \rightarrow \mathbb{Z}$, $B : \subseteq X \times Y \rightarrow \mathbb{Z}$ und $C : \subseteq Y \times Z \rightarrow \mathbb{Z}$. Dann gilt für das partielle Matrixprodukt

$$A \cdot B : \subseteq W \times Y \rightarrow \mathbb{Z}, (w, y) \mapsto \sum_{x \in X} A(w, x) \cdot B(x, y).$$

Sei $(w, y) \in W \times Y$ beliebig gewählt. Ist $A(w, y) = \uparrow$ für ein $y \in Y$, dann existiert kein $((w, y), a) \in A$ und damit auch kein $((w, y), a), ((y, z), b) \in A \times B$, selbst wenn es ein $((y, z), b) \in B$ geben sollte. Dies rechtfertigt die Rechenregel $\uparrow \cdot b = \uparrow$. Die Geltung der weiteren Rechenregeln überlegt man sich leicht selbst.

Wir können also über alle $x \in X$ summieren, müssen uns nicht auf die Graphen von A und B beschränken und erhalten so die selbe Formel für das Matrixprodukt, wie für die totalen Matrizen. Deren Produkt ist bekanntlich assoziativ. Damit ist die Multiplikation von partiellen Matrizen ebenfalls assoziativ. Sei nun $(A \cdot B) \cdot C : \subseteq W \times Z \rightarrow \mathbb{Z}$ gegeben und betrachten wir $(w, z) \in W \times Z$. Dann ist

$$\begin{aligned}
 ((A \cdot B) \cdot C)(w, z) &= \sum_{y \in Y} (A \cdot B)(w, y) \cdot C(y, z) \\
 &= \sum_{y \in Y} \left(\sum_{x \in X} A(w, x) \cdot B(x, y) \right) \cdot C(y, z)
 \end{aligned}$$

da Multiplikation auch mit \uparrow kommutativ ist haben wir

$$= \sum_{y \in Y} C(y, z) \cdot \sum_{x \in X} A(w, x) \cdot B(x, y)$$

Es gilt auch hier das Distributivgesetz, also

$$= \sum_{y \in Y} \sum_{x \in X} C(y, z) \cdot (A(w, x) \cdot B(x, y))$$

Da die Addition mit \uparrow kommutativ ist, können wir die Summenzeichen tauschen

$$= \sum_{x \in X} \sum_{y \in Y} C(y, z) \cdot (A(w, x) \cdot B(x, y))$$

Wegen dem Distributivgesetz können wir $A(w, x)$ ausklammern

$$\begin{aligned} &= \sum_{x \in X} \cdot A(w, x) \sum_{y \in Y} C(y, z) \cdot B(x, y) \\ &= \sum_{x \in X} \cdot A(w, x) \sum_{y \in Y} B(x, y) \cdot C(y, z) \\ &= (A \cdot (B \cdot C))(w, z). \end{aligned}$$

Also ist die Matrizenmultiplikation mit partiellen Matrizen ebenfalls assoziativ.

Zeigen wir noch, dass die partielle Einheitsmatrix die Identität unserer Kategorie ist:

Sei $A(w, x) \neq \uparrow$, dann gibt es ein Tupel $((w, x), a) \in A$ (bzw. $A(w, x) = a$). Dann gibt es ebenfalls das Tupel $((x, x), 1) \in \mathbb{1}_X$. Dies ist aber das einzige Tupel der Form $((x, y), n) \in \mathbb{1}_X$, denn für $x \neq y$ ist ja $\mathbb{1}_X(x, y)$ nicht definiert. Damit ist die Menge $\{((w, x), a), ((x, y), b)\} \subseteq A \times \mathbb{1}_X$ gleich $\{((w, x), a), ((x, x), 1)\}$ und damit ist

$$(A \cdot \mathbb{1}_X)(w, x) = \sum_{\{((w, x), a), ((x, x), 1)\}} a \cdot 1 = a \cdot 1 = a.$$

Außerdem gibt es genau ein Tupel in $\mathbb{1}_W$ der Form $((v, w), m)$ und das ist das Tupel $((w, w), 1)$ und damit ist $(\mathbb{1}_W \cdot A)(w, x) = 1 \cdot A(w, x) = A(w, x)$.

Ist hingegen $A(w, x) = \uparrow$, dann gibt es kein Tupel der Form $((w, x), a) \in A$ und damit ist die Menge

$$\{((w, x), a), ((x, y), b) \mid ((w, x), a) \in A \text{ und } ((x, y), b) \in \mathbb{1}_X\}$$

die leere Menge, und die Summe über die leere Menge ist nicht definiert. Damit gilt ebenfalls $(\mathbb{1}_W \cdot A)(w, x) = \uparrow$. Genauso gilt dann $(A \cdot \mathbb{1}_X)(w, x) = \uparrow$.

Also gilt

$$\mathbb{1}_W \cdot A = A = A \cdot \mathbb{1}_X.$$

Die partiellen Einheitsmatrizen sind links- und rechtsneutral.

Damit bilden die endlichen Mengen mit den partiellen Matrizen und der partiellen Matrixmultiplikation eine Kategorie \mathcal{PMatr} . \square

Für einen Komplex gilt stets folgendes: Wenn für beliebige Zellen c und b die c -Koordinate m von $\partial(b)$ ungleich Null ist, dann liegt c im Rand von $\{b\}$ (also $c \in \text{cl}\{b\}$). Dann ist auch das Tupel (b, c, m) im Graph der Darstellungsmatrix von ∂ selbst nach Beseitigung aller Datensätze $(x, y, 0)$ noch vorhanden.

Betrachten wir die Graphen Γ_1 und Γ_2 unseres obigen Beispiels. Im topologischen Raum \underline{G}_1 von Γ_1 gilt: $\text{fr}_{\underline{G}_1}\{d\} = \{b\}$. Im Raum \underline{G}_2 von Γ_2 gilt hingegen: $\text{fr}_{\underline{G}_2}\{d\} = \{a\}$. Also gilt $a \in \text{cl}_{\underline{G}_2}\{d\}$, nicht jedoch $a \in \text{cl}_{\underline{G}_1}\{d\}$. Wir wissen bereits, dass die Darstellungsmatrix der Randabbildungen in beiden Fällen die Nullmatrix, also gleich ist. Nun soll der hier beobachtete topologische Unterschied durch geeignete Behandlung von expliziten Nulleinträgen in der Matrix zu unterschiedlichen partiellen Matrizen führen.

Definition 4.5 (relationaler Kettenkomplex). Ein Tupel (oder eine „Sequenz“) $(B_n, D_n, B_{n-1}, D_{n-1}, \dots, D_1, B_0)$ von Mengen B_i für $i = 0, \dots, n$ und Relationen $D_j \subseteq B_j \times B_{j-1} \times \mathbb{Z}$ für $j = 1, \dots, n$, mit Schlüssel $\{1, 2\}$ bezeichnen wir als relationalen Kettenkomplex, wenn alle partiellen Matrixprodukte $D_k \cdot D_{k-1}$ für $k = 2, \dots, n$ partielle Nullmatrizen sind. Das heißt für alle $k = 2, \dots, n$ und alle $(a, b) \in B_n \times B_{n-2}$ gilt $D_k \cdot D_{k-1}(a, b) = \uparrow$ oder $D_k \cdot D_{k-1}(a, b) = 0$.

Wir sehen, dass nicht gefordert ist, alle Nulleinträge aus einer dünn besetzten Matrix zu entfernen, sondern lediglich erlaubt ist, dies zu tun. Damit ist jeder totale relationale Kettenkomplex ein relationaler Kettenkomplex.

Jedoch soll ein vorhandener Eintrag mit Wert 0 von einem fehlenden Eintrag genau unterschieden werden:

Definition 4.6 (relationaler Komplexmorphismus). Ein relationaler Komplexmorphismus $\mathcal{M} : \mathcal{B} \rightarrow \mathcal{K}$ von relationalen Kettenkomplexen $\mathcal{B} = (B_n, D_n, B_{n-1}, \dots, D_1, B_0)$ und $\mathcal{K} = (K_n, R_n, K_{n-1}, \dots, R_1, K_0)$ ist ein Tupel $\mathcal{M} = (M_n, M_{n-1}, \dots, M_0)$ von (Graphen von) partiellen Matrizen $M_n : B_n \times K_n \rightarrow \mathbb{Z}$, so dass für alle $i = 1, \dots, n$ gilt:

$$D_i \cdot M_{i-1} = M_i \cdot R_i.$$

Wir erinnern uns, dass partielle Abbildungen $f : \subseteq A \rightarrow X$ und $g : \subseteq B \rightarrow Y$ genau dann gleich sind, wenn $A = B$ und $X = Y$ und der Graph von f gleich dem Graph von g ist. Das heißt insbesondere, dass für jedes a aus A gilt, dass f genau dann in a definiert ist, wenn g in a definiert ist. Damit fordert die obige Definition von relationalen Komplexmorphismen insbesondere, dass explizite Matrixeinträge mit Wert 0 von „fehlenden“ Einträgen stets genau zu unterscheiden sind.

Wir wollen zeigen, dass die relationalen Kettenkomplexe topologische Eigenschaften besser unterscheiden:

Lemma 4.3. Die partiellen Kettenkomplexe von \underline{G}_1 und \underline{G}_2 sind nicht isomorph.

Beweis. Betrachten wir nun noch einmal die beiden Graphen

$$\Gamma_1 : \quad c \begin{array}{c} \circ \\ \curvearrowright \end{array} \bullet a \quad \bullet b \begin{array}{c} \circ \\ \curvearrowleft \end{array} d \qquad \Gamma_2 : \quad c \begin{array}{c} \circ \\ \curvearrowright \end{array} \bullet a \begin{array}{c} \circ \\ \curvearrowleft \end{array} d \quad \bullet b$$

Die Kantenmenge ist jeweils $\{c, d\}$, die Knotenmenge ist $\{a, b\}$. Die partiellen Randoperatoren D_1 und D_2 von Γ_1 und Γ_2 sind

$$D_1 = \begin{array}{c} c \\ d \end{array} \begin{array}{cc} a & b \\ \left(\begin{array}{cc} 0 & \uparrow \\ \uparrow & 0 \end{array} \right) \end{array} \qquad D_2 = \begin{array}{c} c \\ d \end{array} \begin{array}{cc} a & b \\ \left(\begin{array}{cc} 0 & \uparrow \\ 0 & \uparrow \end{array} \right) \end{array}$$

Dann kann es keinen Isomorphismus in $\mathcal{DKetKomp}$ von Γ_1 nach Γ_2 geben. Denn aus der Behauptung $\mathcal{M} = (M_{\{cd\}}, M_{\{ab\}})$ sei ein solcher Isomorphismus, folgt ein Widerspruch: $M_{\{cd\}}$ ist eine $\{c, d\} \times \{c, d\}$ -Matrix und M_{ab} entsprechend eine $\{a, b\} \times \{a, b\}$ -Matrix:

$$M_{ab} = \begin{matrix} & a & b \\ \begin{matrix} a \\ b \end{matrix} & \begin{pmatrix} m_{aa} & m_{ab} \\ m_{ba} & m_{bb} \end{pmatrix} \end{matrix} \quad M_{cd} = \begin{matrix} & c & d \\ \begin{matrix} c \\ d \end{matrix} & \begin{pmatrix} m_{cc} & m_{cd} \\ m_{dc} & m_{dd} \end{pmatrix} \end{matrix}$$

Da \mathcal{M} ein Morphismus ist, gilt: $D_1 \cdot M_{ab} = M_{cd} \cdot D_2$.

$$\begin{aligned} D_1 \cdot M_{ab} &= \begin{matrix} & a & b \\ \begin{matrix} c \\ d \end{matrix} & \begin{pmatrix} 0 & \uparrow \\ \uparrow & 0 \end{pmatrix} \end{matrix} \cdot \begin{matrix} & a & b \\ \begin{matrix} a \\ b \end{matrix} & \begin{pmatrix} m_{aa} & m_{ab} \\ m_{ba} & m_{bb} \end{pmatrix} \\ &= \begin{matrix} & a & b \\ \begin{matrix} c \\ d \end{matrix} & \begin{pmatrix} 0 \cdot m_{aa} + \uparrow \cdot m_{ba} & 0 \cdot m_{ab} + \uparrow \cdot m_{bb} \\ \uparrow \cdot m_{aa} + 0 \cdot m_{ba} & \uparrow \cdot m_{ab} + 0 \cdot m_{bb} \end{pmatrix} \\ &= \begin{matrix} & a & b \\ \begin{matrix} c \\ d \end{matrix} & \begin{pmatrix} 0 \cdot m_{aa} & 0 \cdot m_{ab} \\ 0 \cdot m_{ba} & 0 \cdot m_{bb} \end{pmatrix} = 0 \cdot M_{ab} \end{matrix} \\ M_{cd} \cdot D_2 &= \begin{matrix} & c & d \\ \begin{matrix} c \\ d \end{matrix} & \begin{pmatrix} m_{cc} & m_{cd} \\ m_{dc} & m_{dd} \end{pmatrix} \end{matrix} \cdot \begin{matrix} & a & b \\ \begin{matrix} c \\ d \end{matrix} & \begin{pmatrix} 0 & \uparrow \\ 0 & \uparrow \end{pmatrix} \\ &= \begin{matrix} & a & b \\ \begin{matrix} c \\ d \end{matrix} & \begin{pmatrix} m_{cc} \cdot 0 + m_{cd} \cdot 0 & m_{cc} \cdot \uparrow + m_{cd} \cdot \uparrow \\ m_{dc} \cdot 0 + m_{dd} \cdot 0 & m_{dc} \cdot \uparrow + m_{dd} \cdot \uparrow \end{pmatrix} \\ &= \begin{matrix} & a & b \\ \begin{matrix} c \\ d \end{matrix} & \begin{pmatrix} m_{cc} \cdot 0 + m_{cd} \cdot 0 & \uparrow \\ m_{dc} \cdot 0 + m_{dd} \cdot 0 & \uparrow \end{pmatrix} \end{matrix} \end{aligned}$$

Damit gilt $0 \cdot m_{ab} = \uparrow$ und $0 \cdot m_{bb} = \uparrow$ und daraus folgt $m_{ab} = \uparrow$ und $m_{bb} = \uparrow$. Da nun \mathcal{M} ein Isomorphismus ist, gibt es auch den Umkehrmorphismus $\mathcal{M}^{-1} = (M_{cd}^{-1}, M_{ab}^{-1})$, so dass

$$\text{id}_{\Gamma_1} = \mathcal{M} \cdot \mathcal{M}^{-1} = (M_{cd} \cdot M_{cd}^{-1}, M_{ab} \cdot M_{ab}^{-1}) = (\mathbb{1}_{\{c,d\}}, \mathbb{1}_{\{a,b\}}).$$

Insbesondere ist

$$\begin{aligned} M_{ab} \cdot M_{ab}^{-1} &= M_{ab} \cdot \begin{matrix} & a & b \\ \begin{matrix} a \\ b \end{matrix} & \begin{pmatrix} m_{aa}^- & m_{ab}^- \\ m_{ba}^- & m_{bb}^- \end{pmatrix} \\ &= \begin{matrix} & a & b \\ \begin{matrix} a \\ b \end{matrix} & \begin{pmatrix} m_{aa} \cdot m_{aa}^- + m_{ab} \cdot m_{ba}^- & m_{aa} \cdot m_{ab}^- + m_{ab} \cdot m_{bb}^- \\ m_{ba} \cdot m_{aa}^- + m_{bb} \cdot m_{ba}^- & m_{ba} \cdot m_{ab}^- + m_{bb} \cdot m_{bb}^- \end{pmatrix} \end{matrix} \end{aligned}$$

Durch Streichen der Summanden der Form $m_{ab} \cdot x$ und $m_{bb} \cdot x$ erhalten wir

$$M_{ab} \cdot M_{ab}^{-1} = \begin{array}{c} a \\ b \end{array} \begin{pmatrix} m_{aa} \cdot m_{aa}^- & m_{aa} \cdot m_{ab}^- \\ m_{ba} \cdot m_{aa}^- & m_{ba} \cdot m_{ab}^- \end{pmatrix} = \begin{array}{c} a \\ b \end{array} \begin{pmatrix} 1 & \uparrow \\ \uparrow & 1 \end{pmatrix} .$$

Da $m_{aa} \cdot m_{aa}^- = 1$ ist, muss gelten: $m_{aa} \neq \uparrow$. Wegen $m_{aa} \cdot m_{ab}^- = \uparrow$ folgt daraus $m_{ab}^- = \uparrow$. Dann gilt jedoch $m_{ba} \cdot m_{ab}^- = \uparrow$. Das steht jedoch im Widerspruch zu $m_{ba} \cdot m_{ab}^- = 1$, was uns ein Blick auf die obige Matrix zeigt. Also folgt aus der Annahme \mathcal{M} sei Isomorphismus ein Widerspruch und damit kann \mathcal{M} kein Isomorphismus sein. Also kann es keinen Isomorphismus zwischen den partiellen Komplexen zu Γ_1 und Γ_2 geben, diese sind also nicht isomorph. \square

Jede partielle Matrix lässt sich zu einer totalen Matrix fortsetzen, indem wir einfach die fehlenden Einträge mit 0 auffüllen. Dann sind die Komplexe unserer Beispielgraphen gleich. Da die partiellen relationalen Komplexe jedoch nicht isomorph sind, haben wir keine Äquivalenz der Kategorien $\mathcal{DKetKomp}$ und der Kategorie der endlichen Kettenkomplexe. Das heißt $\mathcal{DKetKomp}$ unterscheidet topologische Eigenschaften besser als die endlichen Kettenkomplexe. Wir wollen nun untersuchen, welche topologischen Eigenschaften auf diese Weise unterschieden werden können.

4.3 Das Datenmodell für topologische Räume

\mathcal{DTop}

In Abschnitt 4.1.1 haben wir gesehen, dass jeder relationale Komplex in zwei Relationen gespeichert werden kann, wenn wir die Dimension der Zellen als weiteres Schlüsselattribut definieren. Nehmen wir also ohne Einschränkung an, dass relationale Komplexe in dieser Form (B, D) vorliegen: eine Datenbank mit einer Menge B und einer partiellen $B \times B$ Matrix $D : \subseteq B \times B \rightarrow \mathbb{Z}$, so dass D^2 eine partielle Nullmatrix ist. Dann gilt für alle a, b in B : wenn $D(a, b) \neq \uparrow$ gilt, dann ist b im Rand von a (also $b \in \text{fr}\{a\}$) und somit wird a von b berührt (bzw. $b \in \text{cl}\{a\}$). Die Umkehrung gilt jedoch nicht, denn es ist zwar $a \in \text{cl}\{a\}$, aber stets $D(a, a) = \uparrow$.

4.3.1 Topologische Datenbanken

Wir werden nun die Relation $R \subseteq B \times B$, definiert durch $a R b$ gilt genau dann, wenn $D(a, b) \neq \uparrow$ gilt, verwenden, um eine Topologie für B zu definieren. R ist also nichts anderes als der Definitionsbereich der partiellen Matrix D . Ist D Tabelle einer relationalen Datenbank, dann erhalten wir R durch die Projektion $\pi_{\mathcal{K}}(D)$ von D auf den Primärschlüssel \mathcal{K} .

Definition 4.7 (Stern). Sei X eine Menge, $R \subseteq X \times X$ Relation auf X und $A \subseteq X$ beliebige Teilmenge von X . Dann ist

$$\text{St}_R(A) := \{p \in X \mid p R a, a \in A\}$$

der (unmittelbare) Stern um A bezüglich R .

Der unmittelbare Stern um eine Menge A ist also die Menge aller Elemente, die mit einem Element in A in einer gegebenen Relation stehen. Ist z.B. R eine Äquivalenzrelation, dann ist der Stern um einen Punkt einfach dessen Äquivalenzklasse. Damit haben wir eine Abbildung

$$\text{St}_R : \mathcal{P}(X) \rightarrow \mathcal{P}(X), A \mapsto \text{St}_R(A).$$

Wir erinnern uns, dass für jede Relation $S \subseteq A \times B$ eine Abbildung

$$\text{im}_S : \mathcal{P}(A) \rightarrow \mathcal{P}(B)$$

definiert ist. St_R ist nun nichts anderes als im_{R^T} .

Für diese Abbildung können wir nun Potenzen definieren:

$$\begin{aligned} \text{St}_R^0 &:= \text{id}_{\mathcal{P}(X)} \\ \text{St}_R^1 &:= \text{St}_R \\ \text{St}_R^{n+1} &:= \text{St}_R^n \circ \text{St}_R \\ \text{St}_R^*(A) &:= \bigcup_{i \in \mathbb{N}} \text{St}_R^i(A) \\ \text{St}_R^+(A) &:= \bigcup_{i \in \mathbb{N}} \text{St}_R^{i+1}(A) \end{aligned}$$

Die Komposition von Sternen ist nun im Wesentlichen der Stern der Kompositionen der entsprechenden Relationen:

Lemma 4.4 (Eigenschaften des Sterns). *Seien X eine Menge, R und S Relationen in X . Dann gilt: $\text{St}_R \circ \text{St}_S = \text{St}_{R \circ S}$, $\text{St}_R^n = \text{St}_{R^n}$, $\text{St}_R^* = \text{St}_{R^*}$ und $\text{St}_R^+ = \text{St}_{R^+}$. Zudem gilt $\text{St}_R^* \circ \text{St}_R^* = \text{St}_R^*$.*

Beweis. Sei A beliebige Teilmenge von X .

$$\begin{aligned} \text{St}_R \circ \text{St}_S(A) &= \text{St}_R(\text{St}_S(A)) \\ &= \{p \in X \mid p R j, j \in \text{St}_S(A)\} \\ &= \{p \in X \mid p R j, j S a, a \in A\} \\ &= \{p \in X \mid p (R \circ S) a, a \in A\} \\ &= \text{St}_{R \circ S}(A) \end{aligned}$$

Wir beweisen damit die zweite Gleichung durch Induktion über n :

Induktionsanfang ($n = 0$):

$$\begin{aligned} \text{St}_R^0(A) &= \text{id}(A) = A \\ &= \{a \in X \mid a \in A\} \\ &= \{a \in X \mid a = b, b \in A\} \\ &= \{a \in X \mid a \Delta_X b, b \in A\} \\ &= \text{St}_{\Delta_X}(A) = \text{St}_{R^0}(A) \end{aligned}$$

also gilt $\text{St}_R^n(A) = \text{St}_{R^n}(A)$ für $n = 0$.

Induktionsschritt($n \rightarrow n + 1$):

Gelte nun $\text{St}_R^n = \text{St}_{R^n}$ für ein $n \in \mathbb{N}$. $\text{St}_R^{n+1} = \text{St}_R^n \circ \text{St}_R = \text{St}_{R^n \circ R} = \text{St}_{R^{n+1}}$.

Damit ist die Behauptung $\text{St}_R^n = \text{St}_{R^n}$ bewiesen.

Wir zeigen nur noch die dritte Gleichung $\text{St}_R^* = \text{St}_{R^*}$, denn die vierte zeigt man ganz analog.

Sei $A \subseteq X$ beliebig. Ist $p \in \text{St}_R^*(A)$, dann existiert ein $n \in \mathbb{N}$ mit $p \in \text{St}_R^n(A)$. Dann folgt $p \in \text{St}_{R^n}(A)$ also $p R^n a$ für ein $a \in A$. Dann gilt auch $p R^* a$ und somit $p \in \text{St}_{R^*}(A)$. Es gilt also $\text{St}_R^*(A) \subseteq \text{St}_{R^*}(A)$.

Umgekehrt folgt aus $p \in \text{St}_{R^*}(A)$, dass $p R^* a$ für ein $a \in A$ gilt und somit existiert ein $n \in \mathbb{N}$ mit $p R^n a$. Damit ist $p \in \text{St}_{R^n}(A) = \text{St}_R^n(A) \subseteq \text{St}_R^*(A)$. Es gilt also ebenfalls $\text{St}_{R^*}(A) \subseteq \text{St}_R^*(A)$ und somit $\text{St}_{R^*}(A) = \text{St}_R^*(A)$.

Also sind die Abbildungen St_{R^*} und St_R^* gleich.

Damit gilt auch $\text{St}_R^* \circ \text{St}_R^* = \text{St}_{R^* \circ R^*}$. Da R^* transitiv ist, gilt $R^* \circ R^* \subseteq R^*$, und da R^* reflexiv ist, gilt $R^* \subseteq R^* \circ R^*$. Ist nämlich $(a, b) \in R^*$, dann ist auch $(b, b) \in R^*$ und somit $(a, b) \in R^* \circ R^*$. Folglich gilt $R^* = R^* \circ R^*$ und damit auch $\text{St}_{R^* \circ R^*} = \text{St}_{R^*}$, woraus die Behauptung $\text{St}_R^* \circ \text{St}_R^* = \text{St}_R^*$ folgt. \square

Nun wollen wir diese Abbildung verwenden, um die Topologie eines relationalen Komplexes zu definieren.

Satz 4.5 (Relation und Topologie). *Sei X eine Menge und R eine Relation auf X . Dann ist*

$$\mathcal{T}_R := \{\text{St}_R^*(A) \mid A \subseteq X\} = \text{St}_R^*[\mathcal{P}(X)]$$

eine Topologie für X , die wir als die von R erzeugte Topologie bezeichnen. Es gilt $A \in \mathcal{T}_R$ genau dann, wenn $\text{St}_R(A) \subseteq A$.

Wir erinnern uns, dass die Relation R bei einem relationalen Komplex genau der Definitionsbereich der partiellen Matrix $D : \subseteq X \times X \rightarrow \mathbb{Z}$ dieses Komplexes ist.

Beweis. Zuerst zeigen wir, dass \mathcal{T}_R in der Tat eine Topologie ist: $\text{St}_R^*(\emptyset) = \{p \in X \mid p R^* j, j \in \emptyset\}$. Da es ein $j \in \emptyset$ nicht gibt, gibt es auch kein p mit $p R^* j$ und damit ist $\text{St}_R^*(\emptyset) = \emptyset$ und somit $\emptyset \in \text{St}_R^*[\mathcal{P}(X)]$.

$\text{St}_R^*(X) = \{p \in X \mid p R^* j, j \in X\}$. Sei $p \in X$ beliebig. Da R^* reflexiv ist, gilt stets $p R^* p$ und somit ist p eben dieses j . Daher gilt $X \subseteq \text{St}_R^*(X)$. Laut Definition von St_R gilt auch $\text{St}_R^*(X) \subseteq X$ und folglich $\text{St}_R^*(X) = X$. Also ist $X \in \text{St}_R^*[\mathcal{P}(X)]$.

Da jede Menge $A \in \mathcal{T}_R$ das Bild $\text{St}_R^*(M)$ einer Teilmenge M von X ist gilt

$$A = \text{St}_R^*(M) = \text{St}_R^* \circ \text{St}_R^*(M) = \text{St}_R^*(\text{St}_R^*(M)) = \text{St}_R^*(A).$$

Gilt umgekehrt $\text{St}_R^*(M) = M$, dann ist M Element von \mathcal{T}_R , also sind die offenen Mengen genau die Teilmengen $M \subseteq X$ mit der Eigenschaft $\text{St}_R^*(M) = M$. Aus der Definition von St_R^* folgt sofort $\text{St}_R^0(M) \subseteq \text{St}_R^*(M)$ und damit gilt die eine Seite der Inklusion $M \subseteq \text{St}_R^*(M)$ für jede Menge M und nicht nur für offene Mengen. Damit ist $\text{St}_R^*(M) \subseteq M$ eine charakterisierende Eigenschaft für die offenen Mengen.

Sei nun \mathcal{A} eine beliebige Teilmenge von \mathcal{T}_R und $B := \bigcup \mathcal{A}$ die Vereinigung aller Mengen in \mathcal{A} . Für ein beliebiges $p \in \text{St}_R^*(B)$ gilt $p R^* b$ für ein b der

Menge B . Dann gilt $b \in A$ für ein $A \in \mathcal{A}$. Und für dieses A gilt $\text{St}_R^*(A) = A$. Wegen $p R^* b$ gilt $p \in \text{St}_R^*(A)$ und somit $p \in A$ und damit auch $p \in B$. Also gilt $\text{St}_R^*(B) \subseteq B$ und damit erfüllt B die oben gezeigte charakteristische Eigenschaft offener Mengen und ist somit offen.

Seien $C := \bigcap \mathcal{A}$ der Durchschnitt aller Mengen in \mathcal{A} und $p \in \text{St}_R^*(C)$, dann existiert ein $c \in C$, so dass $p R^* c$ gilt. Dann ist c Element jeder Menge $A \in \mathcal{A}$ und somit gilt $p \in A$ für alle $A \in \mathcal{A}$, denn für jedes dieser A gilt $\text{St}_R^*(A) = A$. Dann ist p Element von C . Also gilt $\text{St}_R^*(C) \subseteq C$. Damit ist C ebenfalls offen. Insbesondere ist damit der Schnitt jeder beliebigen Teilmenge offener Mengen offen.

Wir wollen nun noch zeigen, dass eine Menge $A \subseteq X$ genau dann offen ist, wenn $\text{St}_R(A) \subseteq A$.

„ \Rightarrow “: Sei A offen, dann gilt $A = \text{St}_R^*(A)$. Aus der Definition von St_R^* ergibt sich $\text{St}_R(A) \subseteq \text{St}_R^*(A)$.

„ \Leftarrow “: Gelte nun $\text{St}_R(A) \subseteq A$. Dann müssen wir $\text{St}_R^*(A) \subseteq A$ zeigen. Liegt $p \in \text{St}_R^*(A)$, dann gibt es ein n mit $p \in \text{St}_R^n(A)$. Durch Induktion über n zeigen wir, dass $\text{St}_R^n(A) \subseteq A$ gilt. Damit liegt $p \in A$ und daraus wiederum folgt $\text{St}_R^*(A) \subseteq A$:

Für $n = 0$ gilt dies wegen $\text{St}_R^0(A) = \text{id}_{\mathcal{P}X}(A) = A \subseteq A$.

Gelte nun $\text{St}_R^n(A) \subseteq A$ für ein n und sei $p \in \text{St}_R^{n+1}(A) = \text{St}_R(\text{St}_R^n(A))$.

Dann gibt es ein $q \in \text{St}_R^n(A)$ mit $p R q$. Wegen der Induktionsannahme gilt $\text{St}_R^n(A) \subseteq A$ und damit ist $q \in A$. Damit ist aber $p \in \text{St}_R(A)$ und wegen der Voraussetzung auch in A . Also gilt $\text{St}_R^{n+1}(A) \subseteq A$.

□

Wir haben also eine Topologie, die nicht nur den Schnitt von endlich vielen offenen Mengen sondern sogar von beliebig vielen offenen Mengen enthält. Also haben wir hier eine alexandroffsche Topologie bzw. einen alexandroffschen Raum.

Damit haben wir gezeigt, dass zu jedem gerichteten simplen Graphen (X, R) ein topologischer Raum (X, \mathcal{T}_R) existiert. Die offenen Mengen sind zwar unter Verwendung der transitiven (und reflexiven) Hülle einer Relation definiert, das zuletzt angegebene Kriterium für offene Mengen kann jedoch mit einer relationalen Algebra ohne transitive Hülle entschieden werden.

Für eine beliebige Menge A lässt sich deren unmittelbarer Stern $\text{St}_R(A)$ als Abfrage leicht realisieren, wie z.B. durch folgenden SQL-Ausdruck:

```
create view StRA as
select a as id
from   A, R
where  R.b = A.id;
```

Dabei ist vorausgesetzt, dass R Relationenschema $[a : X.\text{id}, b : X.\text{id}]$ hat. X habe Primärschlüssel $\{\text{id}\}$ und A sei Teilmenge von X mit dessen Relationenschema.

Wir werden jedoch noch sehen, dass viele sinnvolle topologische Eigenschaften im Allgemeinen ohne Verwendung der transitiven Hülle nicht abgefragt werden können.

Definition 4.8 (topologischer Datentyp, topologische Datenbank). Ein Paar (X, R) bestehend aus einer Menge X und einer Relation $R \subseteq X \times X$ bezeichnen wir auch als topologischen Datentyp. Eine Familie $((X_i, R_i))_{i \in I}$ von topologischen Datentypen bezeichnen wir als topologische Datenbank.

Dies ist offenbar nur ein anderes Wort für „simpler gerichteter Graph“. Allerdings wählen wir die Bezeichnung topologischer Datentyp dann, wenn wir den topologischen Raum (X, \mathcal{T}_R) betrachten. Wir bezeichnen dies als „Datenbank“, da diese Arbeit stets auf die naheliegende Realisierung von X und R als relationale Datenbank abzielt. Allerdings kann der größte Teil dieser Theorie allgemeiner dargestellt werden, ohne dass dazu jedesmal explizit auf das relationale Datenmodell verwiesen werden muss.

4.3.2 Stetige Datenbankabbildungen

Erst durch die Definition von stetigen Abbildungen für topologische Datentypen ist es gerechtfertigt, ein neues Wort für eine alte Sache zu verwenden, denn dann haben wir eine andere Kategorie, eben \mathcal{DTop} .

Satz 4.6 (stetige Abbildungen topologischer Datenbanken). Seien (X, R) und (Y, S) topologische Datentypen und $f : X \rightarrow Y$ eine Abbildung. Dann ist $f : (X, \mathcal{T}_R) \rightarrow (Y, \mathcal{T}_S)$ genau dann stetig, wenn $(f \times f)[R] \subseteq S^*$.

Später werden wir sehen, dass zur Entscheidung der Stetigkeit einer Abbildung auf die transitive Hülle nicht verzichtet werden kann. Zunächst aber der Beweis des obigen Satzes:

Beweis. Wir erinnern uns zunächst an die Definition von $f \times f$ als

$$f \times f : X \times X \rightarrow Y \times Y, (x_1, x_2) \mapsto (f(x_1), f(x_2)).$$

Seien (X, R) und (Y, S) topologische Datentypen und $f : X \rightarrow Y$ eine Abbildung.

\implies : Sei $f : (X, \mathcal{T}_R) \rightarrow (Y, \mathcal{T}_S)$ stetig.

Sei $(fa, fb) \in (f \times f)[R]$ beliebig gewählt. Dann existiert ein $(a, b) \in R$ mit $f(a) = fa$ und $f(b) = fb$. Dann ist $f(b) \in Y$ und $B := \text{St}_{S^*}(\{f(b)\})$ offen in (Y, \mathcal{T}_S) . Dann ist $f^{-1}[B]$ offen in (X, \mathcal{T}_R) , da f laut Voraussetzung stetig ist. Wegen $f(b) \in B$ ist $b \in f^{-1}[B]$. Dann ist $a \in f^{-1}[B]$, denn wegen $a R b$ gilt $a R^* b$ und somit $a \in \text{St}_{R^*}(f^{-1}[B]) = f^{-1}[B]$. Ist $a \in f^{-1}[B]$, dann ist $f(a) \in B$ und somit $f(a) S^* f(b)$, also $(fa, fb) \in S^*$ bzw. $(f \times f)[R] \subseteq S^*$.

\impliedby : Habe f die Eigenschaft $(f \times f)[R] \subseteq S^*$.

Sei $B \in \mathcal{T}_S$ eine beliebige offene Menge und sei $A := f^{-1}[B]$, also die Menge aller $x \in X$ mit $f(x) \in B$. Wir betrachten nun einen Punkt $a \in \text{St}_R(A)$. Dann gilt $a R b$ für ein $b \in A$. Laut Annahme gilt dann $f(a) S^* f(b)$. Also ist $f(a) \in \text{St}_{S^*}(B)$, denn $f(b)$ ist ja Element von B . Da B offen ist, gilt $\text{St}_{S^*}(B) = B$ und somit gilt $f(a) \in B$. Damit gilt $a \in f^{-1}[B] = A$. Also ergibt sich $\text{St}_R(A) \subseteq A$ und somit ist A offen in (X, \mathcal{T}_R) nach Satz 4.5. Da $A = f^{-1}[B]$ das Urbild einer beliebigen offenen Menge B ist, folgt die Stetigkeit von f .

□

Damit haben wir stetige Datenbankabbildungen:

Definition 4.9 (stetige Datenbankabbildung). *Seien (X, R) und (Y, S) topologische Datentypen. Eine Abbildung $f : X \rightarrow Y$ heißt stetige Datenbankabbildung, wenn gilt $(f \times f)[R] \subseteq S^*$.*

Stetige Datenbankabbildungen sind etwas allgemeiner als Graphenmorphis-men. (X, R) und (Y, S) sind zwar auch (simple gerichtete) Graphen. Ein Morphismus von simplen gerichteten Graphen ist eine Abbildung $f : X \rightarrow Y$ mit $(f \times f)[R] \subseteq S$. Wegen $S \subseteq S^*$ ist ein Graphenmorphismus ebenfalls stetig. Es ist jedoch umgekehrt nicht jede stetige Datenbankabbildung ein Graphenmorphi-smus.

Beispiele 4.3 (Stetige Datenbankabbildungen). *Sei $\{(X, R), (Y, S)\}$ eine topologische Datenbank.*

- *Jede konstante Abbildung $c_y : (X, R) \rightarrow (Y, S), x \mapsto y$ für ein festes y in Y ist eine stetige Datenbankabbildung, denn dann gilt für alle $(a, b) \in R$: $(f(a), f(b)) = (y, y) \in S^0 \subseteq S^*$.*
- *Ist $Q \subseteq R^*$, dann ist die identische Abbildung id_X eine stetige Daten-bankabbildung $\text{id}_X : (X, Q) \rightarrow (X, R)$, denn dann gilt für alle $(a, b) \in Q$: $(\text{id}_X(a), \text{id}_X(b)) = (a, b) \in R^*$.*
- *Ist $\emptyset \subseteq R \subseteq \Delta_X$ dann ist $R^* = \Delta_X$ und jede Abbildung $f : X \rightarrow Y$ ist ein stetige Datenbankabbildung $f : (X, R) \rightarrow (Y, S)$ bei beliebigem $S \subseteq Y \times Y$.*
- *Ist umgekehrt $S^* = Y \times Y$ dann ist ebenfalls jede Abbildung $f : X \rightarrow Y$ eine stetige Datenbankabbildung $f : (X, R) \rightarrow (Y, S)$ bei beliebigem $R \subseteq X \times X$.*

Da die stetigen Datenbankabbildungen genau den stetigen Abbildungen ent-sprechen, ist es nicht überraschend, dass stetige Datenbankabbildungen mit der Abbildungskomposition als Verknüpfung Morphismen einer Kategorie \mathcal{DTop} sind:

Satz 4.7 (Eigenschaften stetiger Datenbankabbildungen). *Seien (X, R) , (Y, S) , (Z, T) topologische Datentypen $f : (X, R) \rightarrow (Y, S)$ und $g : (Y, S) \rightarrow (Z, T)$ stetige Datenbankabbildungen. Dann gilt:*

- $\text{id}_{(X, R)} : (X, R) \rightarrow (X, R)$ ist stetig.
- $g \circ f : (X, R) \rightarrow (Z, T)$ ist stetig.

Die weiteren Morphismeneigenschaften (Assoziativität und Neutralität von id) vererben sich von den entsprechenden Eigenschaften für Abbildungen. Die Stetigkeit von $\text{id}_{(X, R)}$ wurde oben bereits gezeigt, denn es ist ja $R \subseteq R^*$. Man beachte den Unterschied von $\text{id}_{\underline{X}} = \text{id}_{(X, R)}$ – die identische Abbildung von einer topologischen Datenbank zu sich selbst – und id_X , die lediglich die identische Abbildung auf der Punktmenge ist, zu der es jedoch verschiedene Topologien geben kann.

Die Komposition $g \circ f$ zweier stetiger Datenbankabbildungen ist in der Tat ebenfalls eine stetige Datenbankabbildung. Es wurde bereits gezeigt, dass die Komposition zweier stetiger Abbildungen stetig ist. Genau dann gilt aber

$$((g \circ f) \times (g \circ f))[R] \subseteq T^*.$$

Dies zeigt bereits die Nützlichkeit der kategoriellen Sichtweise: Betrachte Morphismen in \mathcal{DTop} als Morphismen in \mathcal{Top} , und schließe aus deren Eigenschaften in \mathcal{Top} wiederum auf Eigenschaften in \mathcal{DTop} . Wir wollen diese Eigenschaft dennoch erneut beweisen, bei diesem Beweis aber die Kategorie \mathcal{DTop} nicht verlassen:

Beweis. Wenn g und f stetig sind und (a, b) in R liegt, dann ist $(f(a), f(b))$ in S^* . Ist jedoch $(f(a), f(b))$ in S^* , dann ist $(g(f(a)), g(f(b)))$ in T^* , denn ist $(f(a), f(b))$ in S^* so ist $f(a)S^n f(b)$ für eine natürliche Zahl n . Durch Induktion über n zeigen wir $(g(f(a)), g(f(b)))$ ist in T^* :

$n = 0$:

Ist $n = 0$, dann ist $f(a) = f(b)$, also $g(f(a)) = g(f(b))$, also

$$(g(f(a)), g(f(b))) \in T^0 \subseteq T^*.$$

$n \rightarrow n + 1$:

Gelte die Induktionsannahme, dass aus $aS^n b$ stets $g(a)T^*g(b)$ folgt. Gelte zudem $f(a)S^{n+1}f(b)$, dann existiert ein j mit $f(a)S^n j S f(b)$. Dann ist laut Induktionsannahme $g(f(a))T^*g(j)$ und wegen der Stetigkeit von g gilt $g(j)T^*g(f(b))$. Da T^* transitiv ist, gilt $g(f(a))T^*g(f(b))$ bzw.

$$(g(f(a)), g(f(b))) \in T^*.$$

Damit ist $(g(f(a)), g(f(b))) = (g \circ f(a), g \circ f(b)) = (g \circ f) \times (g \circ f)(a, b) \in T^*$. Es ist also $((g \circ f) \times (g \circ f))[R] \subseteq T^*$ und damit $g \circ f$ stetig. \square

Bei der Entscheidung von Stetigkeit mit relationaler Algebra ist im Allgemeinen eine relationale Algebra mit transitiver Hülle erforderlich. Wir wissen bereits, dass die transitive Hülle eine in der Literatur viel diskutierte Operation ist [Mai83, Cod90, Cod79] und zudem auch nicht von jedem relationalen Datenbankmanagementsystem angeboten wird.

Satz 4.8 (Stetigkeit und transitive Hülle). *Eine relationale Algebra, mit der allgemein entschieden werden kann, ob eine Abbildung $f : (X, R) \rightarrow (Y, S)$ stetig ist, kann für jede beliebige zweistellige Relation $T \subseteq M \times M$ entscheiden, ob ein gegebenes Paar (a, b) in der transitiven Hülle T^+ von T liegt.*

Erinnern wir uns, dass eine relationale Algebra eine Abfragesprache für relationale Datenbanken ist, mit der mindestens die Operationen Vereinigung, Schnitt und Rest, Selektion, Projektion und kartesisches Produkt bzw. Verbund formuliert werden können und mit der konstante Relationen wie z.B. $\{ \langle a : int, b : fr \rangle \}$ und $\{ \langle a : id, \langle b : id \rangle \}$, konstante Tupel und konstante Werte aus den Domänen angegeben werden können.

„Allgemein Entscheiden“ soll heißen, dass allein anhand der Relationenschemata von f , X , R , Y und S ein Ausdruck $Q(f, X, R, Y, S)$ in relationaler Algebra konstruiert werden kann, so dass f genau dann stetig ist, wenn $Q(f, X, R, Y, S) \neq \emptyset$ gilt. Die Abfrage ist also nicht von der Ausprägung der Datenbank abhängig.

Das Prinzip der Beweise, mit denen wir hier die Notwendigkeit der transitiven Hülle zeigen werden, ist stets gleich: Wir behaupten, es sei mit einer

relationalen Algebra möglich, eine Abfrage Q zu formulieren, die gewisse topologische Eigenschaften abfragen oder eine gewisse topologische Konstruktion vornehmen kann. Dann konstruieren wir aus dieser Abfrage eine weitere Abfrage $C(Q)$, die das Ergebnis von Q verwendet, um daraus die transitive Hülle einer gegebenen Relation abzufragen bzw. zu konstruieren. Also kann mit dieser relationalen Algebra die transitive Hülle einer Relation abgefragt werden:

Beweis. Seien $M[\text{id}]$, $T[\mathbf{x}:M.\text{id}, \mathbf{y}:M.\text{id}]$ Relationen einer relationalen Datenbank und die Datensätze $\langle a:\text{id} \rangle, \langle b:\text{id} \rangle \in M$ gegeben. Dann ist (M, T) ein topologischer Datentyp. Mit $X := \{\langle a:\text{id} \rangle, \langle b:\text{id} \rangle\}$ und $R := \{\langle a:\mathbf{x}, b:\mathbf{y} \rangle\}$ ist das Paar (X, R) ebenfalls ein topologischer Datentyp. Sei i die Inklusionsabbildung $i : X \rightarrow M$, $p \mapsto p$. Als Relation erhalten wir i z.B. durch die Abfrage $\beta_{(p, \text{id}) \leftarrow (\text{id}, \text{id})}(X)$. Dann ist i genau dann stetig, wenn $(i \times i)(R) \subseteq T^*$. Es ist jedoch $(i \times i)(R) = R$, also $R \subseteq T^*$ bzw. $\langle a:\mathbf{x}, b:\mathbf{y} \rangle \in T^*$. Also entscheidet jeder Ausdruck Q mit der Stetigkeit von i auch, ob $(a, b) \in T^*$ gilt.

Ohne Einschränkung enthalte das Relationenschema von Q nicht die Attribute \mathbf{x} und \mathbf{y} , denn ansonsten benennen wir die Attribute in Q einfach um.

Es gilt nun $(a, b) \in T^+$ genau dann, wenn $(a, b) \in T$ gilt, oder wenn sowohl $(a \neq b)$ als auch $(a, b) \in T^*$ gilt.

Wir geben nun einen Ausdruck $C(Q)$ in relationaler Algebra an, der Q verwendet und als Ergebnis die Relation $\{\langle a:\mathbf{x}, b:\mathbf{y} \rangle\}$ zurückgibt, wenn $(a, b) \in T^+$ liegt, und ansonsten die leere Menge.

Der Ausdruck $\sigma_{a=\mathbf{x} \text{ und } b=\mathbf{y}}(T)$ entscheidet die erste Bedingung $(a, b) \in T$. Die zweite Bedingung entscheidet der Ausdruck $\sigma_{\mathbf{x} \neq \mathbf{y}}(\pi_{\mathbf{x}, \mathbf{y}}(\{\langle a:\mathbf{x}, b:\mathbf{y} \rangle\} \times Q))$. Die Vereinigung dieser Ausdrücke ergibt den gewünschten Ausdruck $C(Q)$

$$\sigma_{a=\mathbf{x} \text{ und } b=\mathbf{y}}(T) \cup \sigma_{\mathbf{x} \neq \mathbf{y}}(\pi_{\mathbf{x}, \mathbf{y}}(\{\langle a:\mathbf{x}, b:\mathbf{y} \rangle\} \times Q)).$$

□

Wir werden später sehen, dass es bei den topologischen Konstruktionen auch notwendig sein wird, die gesamte transitive Hülle zu berechnen und nicht bloß ein fest vorgegebenes Paar darin zu suchen.

4.3.3 Die Kategorie \mathcal{DTop}

Wir definieren nun folgende Kategorie der topologischen Datentypen:

Definition 4.10 (\mathcal{DTop}). *Die Kategorie \mathcal{DTop} hat als Objektklasse die topologischen Datentypen (X, R) und für je zwei topologische Datentypen (X, R) und (Y, S) sind die Morphismen die Menge $\text{Mor}_{\mathcal{DTop}}((X, R), (Y, S))$ der stetigen Datenbankabbildungen $f : (X, R) \rightarrow (Y, S)$.*

Wir wissen bereits, dass \mathcal{DTop} eine Kategorie ist. Wir können vorerst offen lassen, ob die topologischen Datentypen endlich sein müssen — in der Praxis sind sie es natürlich. Allerdings wird in dieser Arbeit Endlichkeit nicht grundsätzlich vorausgesetzt. Sollte von dieser Regel abgewichen werden, dann wird explizit darauf hingewiesen.

Nun betrachten wir folgende Kategorie von topologischen Räumen:

Definition 4.11 (*Alex*). *Die Kategorie *Alex* hat als Objektklasse die alexandroffschen Räume (X, \mathcal{T}_X) und für je zwei alexandroffsche Räume (X, \mathcal{T}_X)*

und (Y, \mathcal{T}_Y) ist $\text{Mor}_{\text{Alex}}((X, \mathcal{T}_X), (Y, \mathcal{T}_Y))$ die Menge der stetigen Abbildungen $f : (X, \mathcal{T}_X) \rightarrow (Y, \mathcal{T}_Y)$.

Laut Definition 2.22 auf Seite 60 sind alexandroffsche Räume diejenigen topologischen Räume, bei denen die Schnittmenge jeder beliebigen Menge von offenen Mengen wiederum offen ist. Dann ist offensichtlich jeder endliche topologische Raum ein alexandroffscher Raum. Denn ist (X, \mathcal{T}) ein topologischer Raum und X endlich, dann ist auch dessen Potenzmenge $\mathcal{P}(X)$ endlich. Dann ist \mathcal{T} als Teilmenge von $\mathcal{P}(X)$ ebenfalls endlich und somit auch jede Teilmenge von \mathcal{T} . Der Schnitt einer endlichen Teilmenge einer Topologie ist wiederum Element dieser Topologie. Allerdings gibt es auch nichtendliche alexandroffsche Räume wie z.B. $(\mathbb{N}, \mathcal{T}_{\leq})$, die von der „kleiner-gleich“-Relation im Sinne von Satz 4.5 (S. 116) erzeugte Topologie auf den natürlichen Zahlen.

Es gilt nun folgender Satz

Satz 4.9 ($\mathcal{D}\text{Top}$ und Alex). *Die Kategorien $\mathcal{D}\text{Top}$ und Alex sind äquivalent.*

Dieser Satz ist sehr wichtig, denn er besagt nicht weniger, als dass jeder beliebige endliche topologische Raum ohne Verlust an topologischen Eigenschaften als topologischer Datentyp gespeichert werden kann.

Beweis. Als Beweis müssen wir einen Funktor $F : \mathcal{D}\text{Top} \rightarrow \text{Alex}$ und einen Funktor $G : \text{Alex} \rightarrow \mathcal{D}\text{Top}$ konstruieren und zeigen, dass für alle topologischen Datentypen (X, D) gilt: $G \circ F((X, D)) \cong (X, D)$, und für alle alexandroffschen Räume (Y, \mathcal{T}_Y) muss gelten $F \circ G(Y, \mathcal{T}_Y) \cong (Y, \mathcal{T}_Y)$.

Der Funktor F wurde oben bereits gegeben: Jedem Datentyp (X, R) haben wir den topologischen Raum (X, \mathcal{T}_R) zugeordnet und gezeigt, dass jede stetige Datenbankabbildung f auch eine stetige Abbildung zwischen den entsprechenden topologischen Räumen ist. Also ordnet F einer stetigen Datenbankabbildung $f : (X, R) \rightarrow (Y, S)$ die stetige Abbildung $f : (X, \mathcal{T}_R) \rightarrow (Y, \mathcal{T}_S)$ zu, bzw. $F(f) = f$. Damit gelten die Eigenschaften $F(f \circ g) = f \circ g = F(f) \circ F(g)$ und $F(\text{id}_{(X,R)}) = \text{id}_{F(X,R)}$ und F ist somit ein Funktor.

Wir wollen nun $G : \text{Alex} \rightarrow \mathcal{D}\text{Top}$ angeben: Sei (X, \mathcal{T}_X) ein alexandroffscher Raum. Dann definieren wir folgende Relation „ \geq “ in X : $a \geq b$ gilt genau dann, wenn $b \in \text{cl}\{a\}$ gilt [Ale37]. Dann ist (X, \geq) ein topologischer Datentyp. Die Relation $a \geq b$ ist für jeden topologischen Raum definiert und nicht auf die alexandroffschen Räume beschränkt. Sie kann gelesen werden als „Der Punkt b berührt den Punkt a “ und wird manchmal auch als *Spezialisierungspräordnung* bezeichnet. Insbesondere ist „ \geq “ transitiv und reflexiv: Wegen $a \in \text{cl}\{a\}$ gilt stets $a \geq a$. und damit ist \geq reflexiv. Und ist $c \geq b \geq a$, dann gilt $a \in \text{cl}\{b\}$ und $b \in \text{cl}\{c\}$ bzw. $\{b\} \subseteq \text{cl}\{c\}$. Folglich gilt laut Satz 2.5 auch $\text{cl}\{b\} \subseteq \text{cl}\{c\} = \text{cl}\{c\}$. Es folgt $a \in \text{cl}\{c\}$ bzw. $c \geq a$ und damit ist „ \geq “ transitiv.

Ebenso wie F übernimmt G jede stetige Abbildung g unverändert, also $G(g) := g$. Damit gelten sofort die Eigenschaften $G(f \circ g) = G(f) \circ G(g)$ und $G(\text{id}_X) = \text{id}_{GX}$.

Seien nun $\underline{X} := (X, \mathcal{T}_X)$ und $\underline{Y} := (Y, \mathcal{T}_Y)$ alexandroffsche Räume und $g : \underline{X} \rightarrow \underline{Y}$ eine stetige Abbildung. Seien zudem (X, \geq) und (Y, \geq) die entsprechenden topologischen Datentypen. Sei nun $b \in \text{cl}_X\{a\}$ für $a, b \in X$, also $a \geq b$. Dann gilt $g(b) \in \text{cl}_Y g[\{a\}]$, denn eine stetige Abbildung erhält ja die Berührungspunkte von Mengen und somit $g(a) \geq g(b)$. Also ist jede stetige Abbildung g eine stetige Datenbankabbildung $G(g) : (X, \geq) \rightarrow (Y, \geq)$ und somit ist G ein Funktor.

Sei nun (X, R) ein topologischer Datentyp und $F(X, R) = (X, \mathcal{T}_R)$, der zugeordnete alexandroffsche Raum und $G \circ F(X, R) = G(X, \mathcal{T}_R) = (X, \geq)$ wiederum der topologische Datentyp von (X, \mathcal{T}_R) . Dann ist zu zeigen, dass (X, R) isomorph zu (X, \geq) ist. Naheliegenderweise werden wir zeigen, dass $\text{id}_X : X \rightarrow X$ der gesuchte Isomorphismus ist.

Sei also $(a, b) \in R$ und sei U_b eine beliebige Umgebung um b in (X, \mathcal{T}_R) . Dann existiert eine offene Menge V_b mit $b \in V_b \subseteq U_b$. Dann gilt aber $\text{St}_R(V_b) \subseteq V_b$ und somit ist $a \in V_b$ und wegen $V_b \subseteq U_b$ auch $a \in U_b$. Also enthält jede Umgebung U_b um b den Punkt a und somit gilt: $b \in \text{cl}\{a\}$. Dann gilt aber $a \geq b$ und somit ist R Teilrelation der Relation \geq ($R \subseteq \text{„}\geq\text{“}$) und somit ist wegen $R = (\text{id}_X \times \text{id}_X)[R]$ die Abbildung $\text{id}_X : (X, R) \rightarrow (X, \geq)$ eine stetige Datenbankabbildung.

Nun ist zu zeigen, dass auch die Umkehrung $\text{id}_X^{-1} : (X, \geq) \rightarrow (X, R)$ stetig ist: Sei also $a \geq b$, dann ist $b \in \text{cl}\{a\}$, also enthält jede offene Menge $U_b \in \mathcal{T}_R$, die b enthält, auch a . Insbesondere ist $\text{St}_R^*(\{b\})$ eine solche offene Menge. Also gilt $a \in \text{St}_R^*(\{b\})$ und somit aR^*b . Also ist die Relation „ \geq “ Teilrelation von R^* und damit ist id_X^{-1} ebenfalls eine stetige Datenbankabbildung und damit auch ein Isomorphismus in \mathcal{DTop} .

Nun ist noch zu zeigen, dass auch auf der anderen Seite ein alexandroffscher Raum isomorph bleibt, wenn wir ihn in einer topologischen Datenbank „zwischen speichern“. Nehmen wir nun einen beliebigen alexandroffschen Raum (X, \mathcal{T}_X) und definieren die Relation $a \geq b$ durch $b \in \text{cl}\{a\}$. Dann ist $G(X, \mathcal{T}_X)$ der topologische Datentyp (X, \geq) und $F(G(X, \mathcal{T}_X))$ ist der topologische Raum (X, \mathcal{T}_{\geq}) . Auch hier ist $\text{id}_X : (X, \mathcal{T}_X) \rightarrow (X, \mathcal{T}_{\geq})$ ein Isomorphismus.

Es ist nämlich die Menge A genau dann offen in (X, \mathcal{T}_{\geq}) , wenn sie offen in (X, \mathcal{T}_X) ist. Zunächst gilt wegen der Transitivität und Reflexivität von \geq , dass $\text{St}_{\geq} = \text{St}_{\geq}^*$. Sei A offen in (X, \mathcal{T}_{\geq}) . Dann ist $A = \text{St}_{\geq}(A)$. Sei $p \in A$ beliebig, dann hat p die offene Umgebung $U_p := \text{St}_{\geq}(\{p\})$. Da $\{p\} \subseteq A$ gilt, ist $\text{St}_{\geq}(\{p\}) \subseteq \text{St}_{\geq}(A) = A$. Also ist A die Vereinigung der minimalen Umgebungen U_p von allen seinen Punkten $p \in A$. Damit reicht es zu zeigen, dass U_p offen in (X, \mathcal{T}_X) ist, denn dann ist auch A als Vereinigung offener Mengen offen in (X, \mathcal{T}_X) . Wir betrachten $V_p := \bigcap \{U \in \mathcal{T}_X \mid p \in U\}$. Dies ist eine offene Menge, denn es ist Schnittmenge von offenen Mengen eines alexandroffschen Raums. Da es Schnittmenge aller offenen Mengen ist, die p enthalten, ist V_p minimal und Teilmenge jeder offenen Menge, die p enthält, also auch Teilmenge von jeder Umgebung von p . Ist $q \in V_p$, dann ist q in jeder Umgebung von p und somit gilt $p \in \text{cl}_{(X, \mathcal{T}_X)}\{p\}$ und das ist eben $q \geq p$. Also ist $V_p = \text{St}_{\geq}(\{p\}) = U_p$. Damit ist A offen in (X, \mathcal{T}_X) . Umgekehrt ist jedes offene $B \in \mathcal{T}_X$ Vereinigung der V_p für jedes $p \in B$ und damit offen in (X, \mathcal{T}_{\geq}) . Also gilt: Eine Teilmenge ist genau dann offen in (X, \mathcal{T}_X) , wenn sie offen in (X, \mathcal{T}_{\geq}) ist. Damit ist die identische Abbildung id_X ein Isomorphismus.

Also bleibt ein topologischer Datentyp im wesentlichen unverändert, wenn wir sie durch die Kategorie \mathcal{Alex} „hindurchschleifen“ und umgekehrt bleibt ein alexandroffscher Raum unverändert, wenn er als topologischer Datentyp zwischengespeichert wird. Damit sind beide Kategorien äquivalent. \square

Also haben wir mit \mathcal{DTop} ein Datenmodell gefunden, mit dem jeder endliche topologische Raum verlustfrei abgespeichert werden kann. Es handelt sich also um ein *effektives* Datenmodell. Wir werden später noch überlegen, wie *effizient* dieses Modell ist, wie „sparsam“ es also mit Speicherplatz umgeht. Zudem können wir Datenbankabfragen angeben, die entscheiden, ob Abbildungen stetig

und Mengen offen sind. Für die Entscheidung der Stetigkeit ist im Allgemeinen die Entscheidung der transitiven Hülle von Relationen Voraussetzung. In den meisten Sonderfällen wie z.B. bei Komplexen mit beschränkter Dimension d ist dies jedoch auch mit Abfragesprachen ohne transitive Hülle möglich, denn dann ist $R^+ = \cup_{i=1}^d R^i$.

Beispiel 4.4. *Angenommen wir haben eine Datenbank mit einer Tabelle $\text{OBJ}[\underline{\text{id}}, \dots]$ der räumlichen Objekte eines architektonischen Entwurfs. Sei zudem $\text{BEGRENZTVON}[\text{wen:OBJ.id}, \text{wer:OBJ.id}]$ eine Tabelle, welche die Topologie dieses Entwurfs erzeugt. Sind a und b Objekte in OBJ , dann gilt $a \text{ BEGRENZTVON } b$ genau dann wenn*

$$\sigma_{a.\text{id}=\text{wen} \text{ und } b.\text{id}=\text{wer}}(\text{BEGRENZTVON}) \neq \emptyset.$$

Z.B. gilt für eine Türe t , die den Raum z mit einem Flur f verbindet: $z \text{ BEGRENZTVON } t$ und $f \text{ BEGRENZTVON } t$. Für ein Element r des Türrahmens von t gilt $t \text{ BEGRENZTVON } r$, also der Rahmen berührt die Türe (nicht andersherum: Die Türe berührt nicht den Rahmen) und somit gilt auch $z \text{ BEGRENZTVON}^2 r$, also berührt der Türrahmen r auch das Zimmer z .

Formal erhalten wir diese Tabellen folgendermaßen: Zunächst nehmen wir den euklidischen Raum \mathbb{R}^3 und zerlegen ihn vermöge eines architektonischen Entwurfs in eine Menge \mathcal{A} von endlich vielen Teilen: die dreidimensionalen Außenräume „Luft“ und „Erdreich“, die Räume des Gebäudes, die flächigen Bauteile und die zweidimensionalen Außenräume „Gelände“, die linienförmigen Verbindungen dieser flächigen Teile und die Knotenpunkte, in denen diese Verbindungen zusammenlaufen. Dann haben wir eben diese Äquivalenzrelation \sim , so dass $\mathcal{A} = \mathbb{R}^3 / \sim$. Und daher ist die Topologie $\mathcal{T}_{\mathcal{A}}$ eben die Quotiententopologie $\mathcal{T}_{\mathbb{R}^3} / \sim$. Da \mathcal{A} endlich ist, haben wir einen alexandroffschen Raum mit endlicher Relation \geq . Für jedes Objekt a legen wir einen entsprechenden Datensatz $t(a)$ in OBJ an, und für die Relation BEGRENZTVON wählen wir eine Relation, deren reflexive und transitive Hülle BEGRENZTVON^ gleich der Relation „ \geq “ des alexandroffschen Raums ist. Idealerweise ist BEGRENZTVON minimal, d.h. nicht transitiv und irreflexiv. In unserem Beispiel ist diese minimale Relation sogar eindeutig bestimmt, da der topologische Raum trennbar und die Relation BEGRENZTVON^* antisymmetrisch ist.*

Als praktisches Beispiel für Stetigkeit nehmen wir an, wir hätten ein Raumprogramm (R, V) , also eine Menge R der geforderten Raumnutzungen und eine Menge V der geforderten Raumverbindungen als ungerichteten Graph. Anfangs hatten wir gesehen, dass eine derartige Graphendarstellung von Entwurfstheoretikern für nützlich gehalten wird. Wir haben auch gesehen, dass wir diesem Graphen einen topologischen Raum $(R \cup V, \mathcal{T})$ zuordnen können. Sei (A, \mathcal{T}) ein architektonischer Entwurf mit Räumen $Z \subseteq A$ und Raumverbindungen L , welche die geforderten Raumnutzungen R bzw. die Raumverbindungen V realisieren. Ist $f : R \cup V \rightarrow Z \cup L$, diejenige Abbildung, die angibt, welches Entwurfselement $f(a)$ eine Anforderung $a \in R \cup V$ erfüllt, dann ist die Abbildung $f : R \cup V \rightarrow Z \cup L$ stetig.

Ausführlichere Beispiele werden im Abschnitt zur Anwendung von topologischen Datenbanken vorgestellt. Es folgt nun noch die versprochene Überlegung, ob diese denn auch effizient sind.

Speicherplatzkomplexität von \mathcal{DTop} . Wie wollen uns hier nun überlegen, ob es bezüglich des benötigten Speicherplatzes noch eine effizientere Datenstruktur als \mathcal{DTop} geben kann. Dabei werden wir nur Überlegungen zur Speicherplatzkomplexität anstellen und nicht auf die Zeitkomplexität von Algorithmen eingehen. Bei Komplexitätsüberlegungen allgemein ist es üblich, nur das relative Wachstum einer untersuchten Größe einer Datenstruktur D (bzw. eines Algorithmus) in Abhängigkeit von der Größe der Eingabedaten zu betrachten. Dazu wird die Funktion $g_D : \mathbb{N} \rightarrow \mathbb{N}$ betrachtet, die angibt, dass die Datenstruktur D zum Abspeichern von n Elementen die Größe $g_D(n)$ annimmt. Dabei werden multiplikative und additive Konstanten vernachlässigt. Es ist also gleichgültig, ob eine Datenstruktur $g_D(n)$ oder $a \cdot g_D(n)$ oder $a \cdot g_D(n) + b$ Speicherplatz benötigt, falls a und b Konstanten sind und a größer als Null ist. Es interessiert also nur, ob eine Funktion f mit einer anderen Funktion g noch „Schritt halten“ kann, egal welchen „Vorsprung“ b und „Beschleunigungsfaktor“ a die Funktion g erhält. Dies gilt genau dann, wenn die Folge

$$\left(\frac{|g(i)| + 1}{|f(i)| + 1} \right)_{i \in \mathbb{N}}$$

eine obere Schranke hat. Man überlegt sich leicht, dass durch Addition und Multiplikation von Konstanten diese Schranke nur verschoben, jedoch kein Einfluss auf die Existenz einer solchen Schranke genommen werden kann. Multiplikation mit 0 ist dabei natürlich untersagt. Dann schreibt man $g \in \mathcal{O}(f)$ und sagt g wächst asymptotisch nicht schneller als f . Der Buchstabe \mathcal{O} steht für „Ordnung“ und die Notation $\mathcal{O}(f)$ bezeichnet man auch als Landau-Symbol. Es gibt noch weitere Landau-Symbole, die wir hier aber nicht verwenden. Ist f eine Polynomfunktion der Form $f(n) = a \cdot n^2 + b \cdot n + c$ (natürlich mit $a \neq 0$), dann gilt $\mathcal{O}(f) = \mathcal{O}(a \cdot n^2 + b \cdot n + c) = \mathcal{O}(n^2)$.

Wir wollen hier als „Größe der Eingabedaten“ die Größe der Punktmenge zugrundelegen und nehmen dabei an, dass diese lineare Speicherkomplexität hat, dass also die Komplexität zum Speichern eines einzelnen Punktes konstant ist. Da wir nun jede Topologie für eine Menge X mit n Elementen als Relation $R \subseteq X \times X$ speichern können, ist zunächst der Speicherbedarf für die Topologie durch $\mathcal{O}(n^2)$ nach oben beschränkt. Eine solche obere Schranke wird in der Graphentheorie bereits als effizient akzeptiert, ist praktisch jedoch kaum vertretbar.

Jedoch ist laut ERNÉ die, von ihm mit $A_0(n)$ bezeichnete Anzahl der T_0 -Topologien, also der trennbaren Topologien, für X durch $2^{\frac{1}{2}n^2}$ nach unten beschränkt [Ern74, S.244], womit wir auch eine untere Schranke für die Anzahl aller Topologien für X erhalten.

Diese untere Schranke ist leicht einzusehen: Als Beweis können wir zunächst ohne Einschränkung annehmen, dass X nicht leer und $n = |X|$ eine gerade Zahl ist. Dann zerlegen wir X in zwei Mengen E wie Edge (Kante) und N wie Node (Knoten) und betrachten nur die Relationen

$$\mathcal{R}_0 = \{R_0 \subseteq X \times X \mid \text{Graph } R_0 \subseteq E \times N\}.$$

Jede derartige Relation ist transitiv und irreflexiv, wie man sich leicht überlegt: Eine Relation $R_0 \in \mathcal{R}_0$ ist transitiv, denn es gilt $R_0^2 = \emptyset \subseteq R_0$ und irreflexiv, da aus $e R_0 m$ folgt, dass $e \in E$ und $m \in N$ gilt. Nun sind diese Mengen eine Zerlegung von X und somit folgt $e \neq m$. Damit gilt auch für je zwei

verschiedene Relationen $R_0, S_0 \in \mathcal{R}_0$, dass deren transitive und reflexive Hüllen ebenfalls verschieden sind — es gilt dann also $R_0^* \neq S_0^*$, was leicht aus $R_0^* = R_0 + \Delta_X$ folgt. Jede dieser Relationen erzeugt somit auch eine andere Topologie. Da es nun $|\mathcal{R}_0| = 2^{|E \times N|}$ derartiger Relationen gibt, existieren auch genauso viele entsprechende Topologien. Wegen $|E \times N| = \frac{1}{2}|X| \cdot \frac{1}{2}|X| = \frac{1}{4}n^2$ haben wir in der Tat mindestens diese $2^{\frac{1}{4}n^2}$ Topologien für X . Die entsprechenden Räume werden übrigens auch als *Hypergraphen* bezeichnet. Dies ist eine Verallgemeinerung der ungerichteten Graphen, wobei eine (Hyper-)Kante $e \in E$ beliebig viele Knoten verbinden kann.

Wenn wir nun von einer binären Darstellung einer Datenstruktur am Rechner ausgehen, und diese in der Lage sein soll, $2^{\frac{1}{4}n^2}$ verschiedenen Topologien zu unterscheiden, dann muss sie auch $2^{\frac{1}{4}n^2}$ verschiedenen Zustände annehmen können. Doch dazu ist mindestens ein Datenwort mit einer Länge von $\frac{1}{4}n^2$ bit notwendig, das diese Topologien nach irgend einem Verfahren „durchnummeriert“. Keine andere Datenstruktur benötigt also im ungünstigsten Falle weniger Speicherplatz als $\mathcal{O}(\frac{1}{4}n^2) = \mathcal{O}(n^2)$ und damit ist dies die bestmögliche obere Schranke für den Platzbedarf zum Speichern von Topologien. Wir haben also mit den topologischen Datenbanken bereits die bezüglich Speicherplatzbedarf effizienteste Datenstruktur für endliche topologische Räume gefunden.

Nun wäre dies ein sehr unbefriedigendes Ergebnis, da quadratische Komplexität für praktische Zwecke bereits untragbar ineffizient ist. Wenn wir uns jedoch an das erinnern, was wir mit den topologischen Datenbanken bezwecken wollen, dann ist es ja gar nicht jede denkbar mögliche Topologie, die wir abzuspeichern wünschen. Wir werden nun sehen, dass wir in der Architektur nur eine wesentlich kleinere Auswahl von Topologien vorfinden, und dass diese bezüglich des Speicherplatzbedarfs sehr viel gutmütiger sind:

Definition 4.12 (Grad und Dimension einer Relation). Sei $R \subseteq X \times X$ eine azyklische Relation, also eine Relation mit der Eigenschaft $R^d = \emptyset$ für eine natürliche Zahl d (dann ist \mathcal{T}_R trennbar). Dann bezeichnen wir d als die Dimension von R , falls d mit dieser Eigenschaft minimal ist und schreiben $\dim R = d$. Für $X = \emptyset$ setzen wir $\dim R = -1$. Des Weiteren bezeichnen wir

$$\deg R := \max \{n \in \mathbb{N} \mid n = |\text{St}_R(x)| \text{ und } x \in X\}$$

als den Grad von R .

Es gibt eine eigene Theorie, die sogenannte Dimensionstheorie, die sich mit der Möglichkeit eines sinnvollen Dimensionsbegriffs für topologischen Räume befasst (zB. [Eng78]). Im Lichte dieser Theorie ist unser Dimensionsbegriff vergleichsweise naiv. Zumindest stimmt er aber mit dem entsprechenden Dimensionsbegriff für Komplexe überein.

Der Grad einer Relation gibt die maximale Anzahl von Elementen an, die mit einem Element in dieser Relation stehen. Ist nun ein dreidimensionaler architektonischer Raum (natürlich bis auf Homöomorphie) „rechtwinklig“, dann berührt ein Eckpunkt höchstens sechs Kanten, eine Kante höchstens vier Flächen und eine Fläche höchstens zwei Volumen. Der Grad der Relation „berührt“ ist in diesem Falle also gleich sechs, bzw. \deg „berührt“ = 6, und die Dimension von „berührt“ ist drei. In höherdimensionalen „rechtwinkligen“ Räumen kann der Grad zwar ebenfalls höher werden, ist dann aber immer noch durch eine Konstante nach oben beschränkt.

Wenn wir nun den Grad einer Relation $R \subseteq X \times X$ durch eine feste Zahl g nach oben beschränken, dann umfasst diese höchstens $g \cdot |X|$ Tupel und hat somit Speicherplatzkomplexität $\mathcal{O}(g \cdot n) = \mathcal{O}(n)$. Wir haben also in diesem Fall lineare Komplexität und damit eine Effizienz, die nicht nur die Graphentheoretiker, sondern auch die Praktiker zufriedenstellt: Unter üblichen Bedingungen ist also die Größe einer topologischen Datenbank durch die Anzahl ihrer Elemente linear nach oben beschränkt. Treten vereinzelt komplexere Situationen auf, wie etwa eine Kante, die mehr als nur vier Wände verbindet, dann steigt damit auch der entsprechende Speicheraufwand. Quadratische Komplexität wird dadurch aber praktisch nie erreicht, und sollte dies unter irgendwelchen sonderbaren Umständen dennoch geschehen, dann liegt die Ursache dafür im architektonischen Raum selbst und nicht am Datenmodell dafür, und muss daher vom Entwerfer derartiger Räume stets in Kauf genommen werden.

Wer hingegen die topologischen Datenbanken nicht nur in Planungswerkzeugen einsetzen möchte, sondern, ganz allgemein, beliebige Topologien abzuspeichern wünscht, etwa für theoretische Untersuchungen, muss sich mit quadratischer Komplexität zufriedustellen, so wie es ihm die Graphentheoretiker bereits vormachen.

Duale Räume

Bei der Beschreibung der Volumenmodellierer wurde erwähnt, dass zu jeder Datenstruktur, die topologische Eigenschaften modelliert, eine dazu duale Struktur existieren muss. Das liegt nun daran, dass es bei den alexandroffschen Räumen eine charakteristische Dualität gibt. So ist für jeden dieser Räume (X, \mathcal{T}_X) die Menge $co\text{-}\mathcal{T}_X := \{X \setminus A \mid A \in \mathcal{T}_X\}$ der abgeschlossenen Mengen wiederum eine alexandroffsche Topologie, die zu \mathcal{T}_X *duale* Topologie. Der Raum $(X, co\text{-}\mathcal{T}_X)$ heißt entsprechend der zu (X, \mathcal{T}_X) *duale* Raum. Zudem fällt eine starke Ähnlichkeit des Sternes um eine Menge mit der abgeschlossenen Hülle einer Menge auf: Der Stern um eine Menge ist die kleinste offene Menge, die diese Menge umfasst. Und dieser ist die abgeschlossene Hülle im dualen Raum. Ist nun eine Topologie \mathcal{T}_R für X durch eine Relation $R \subseteq X \times X$ definiert, dann ist die dazu duale Topologie $co\text{-}\mathcal{T}_R$ von der transponierten Relation erzeugt, es gilt also

$$co\text{-}\mathcal{T}_R = \mathcal{T}_{R^T}.$$

Damit können wir leicht die abgeschlossene Hülle $cl A$ einer Menge berechnen:

Definition 4.13 (abgeschlossene Hülle in \mathcal{DTop}). Sei (X, R) eine topologische Datenbank, $A \subseteq X$, dann definieren wir

$$Bd_R(A) := \text{im}_R(A) = \text{St}_{R^T}(A).$$

Und es ist $Bd_R^*(A)$ die abgeschlossene Hülle von A in (X, \mathcal{T}_R) .

Wir werden diese Schreibweise später bei der Betrachtung der monotonen Mengen wieder verwenden.

4.4 Topologische Konstruktionen

In diesem Abschnitt wollen wir nun die Abfragen von topologischen Datenbanken betrachten und dazu die relationale Algebra auf die topologischen Datenbanken übertragen.

Unter den Grundoperatoren der relationalen Algebra gibt es Selektion, Projektion, Umbenennung und kartesisches Produkt oder Verbund. Jeder dieser Operatoren f erhält eine oder mehrere Relationen R bzw. R und S als Eingabe und erzeugt wiederum Relationen $f(R)$ oder $f(R, S)$ als Ergebnis. Topologische Konstruktionen verfahren ähnlich: Man nehme topologische Räume $(X_1, \mathcal{T}_1), \dots, (X_n, \mathcal{T}_n)$, setze deren Punktmenge X_1, \dots, X_n mit einer Menge Y in Beziehung, suche dann für Y eine am besten geeignete Topologie \mathcal{T}_Y , und erhalte so einen topologischen Raum (Y, \mathcal{T}_Y) . Die Punktmenge X_1, \dots, X_n sind nun die Eingaberelationen der Abfrage f , und Y ist die Ergebnisrelation $Y = f(X_1, \dots, X_n)$. Diese „Beziehung“ zwischen den Eingabemengen und der Ergebnismenge besteht aus Abbildungen und die Topologie wird so gewählt, dass diese Abbildungen gerade noch stetig sind. Zu jedem der Grundoperatoren der relationalen Algebra gibt es nun derart „typische“ Abbildungen, welche die Topologie für das Abfrageergebnis eindeutig festlegen.

4.4.1 Vergleichen von Topologien

Zunächst benötigen wir eine Möglichkeit eine „beste“ Topologie zu bestimmen. Je größer eine Topologie ist, desto feiner „unterteilt“ sie den Raum. Eine kleinere Topologie hingegen ergibt eine gröbere Unterteilung des Raums. Daher sind folgende Bezeichnungen üblich:

Definition 4.14 (Vergleich von Topologien). *Seien \mathcal{G} und \mathcal{F} Topologien für eine Menge X . Dann heißt \mathcal{F} feiner als \mathcal{G} , falls $\mathcal{G} \subseteq \mathcal{F}$. Dann heißt \mathcal{G} auch gröber als \mathcal{F} . Ist $\mathcal{F} \neq \mathcal{G}$ dann sagen wir auch echt feiner bzw. echt gröber.*

Daraus folgt dann sofort, dass eine Topologie \mathcal{F} genau dann feiner als \mathcal{G} ist, wenn die identische Abbildung $\text{id}_X : (X, \mathcal{F}) \rightarrow (X, \mathcal{G})$ stetig ist, und wir haben damit sofort ein Kriterium, feiner bzw. gröber bei topologischen Datenbanken zu entscheiden. Daher wollen wir folgende Sprechweise vereinbaren:

Definition 4.15 (Vergleich von Relationen in \mathcal{DTop}). *Sind F und G Relationen in X , dann bezeichnen wir F als feiner (gröber) als G , wenn $F \subseteq G^*$ ($G \subseteq F^*$) ist. F bezeichnen wir dann als echt feiner (gröber) als G , wenn zudem $F^* \neq G^*$ ist. Ist F feiner als G und G feiner als F , dann bezeichnen wir F und G als äquivalent.*

Nun suchen wir für bestimmte Abbildungen die jeweils „beste“ Topologie bezüglich unseres obigen Vergleichskriteriums aus. Wir betrachten zunächst erneut das Beispiel des Quotientenraums aus Abschnitt 2.5, um dieses Vorgehen zu erläutern:

Beispiel 4.5 (Quotientenraum). *Sei (X, \mathcal{T}_X) ein topologischer Raum und \sim eine Äquivalenzrelation in X . Dann haben wir die folgende Abbildung $\pi : X \rightarrow X/\sim$, die natürliche Projektion, die jedem Element x aus X seine Äquivalenzklasse $[x]$ zuordnet. Die Topologie \mathcal{T}_X/\sim von X/\sim hatten wir in Definition 2.21 so definiert, dass eine Menge $A \subseteq X/\sim$ genau dann offen in $(X, \mathcal{T}_X)/\sim$ ist, wenn die Vereinigung aller Äquivalenzklassen in A offen in (X, \mathcal{T}_X) ist. Die Vereinigung dieser Äquivalenzklassen in A ist aber die Menge der $x \in X$ mit $[x] \in A$, also $\pi^{-1}(A)$. Damit ist π eine stetige Abbildung. Umgekehrt ist jede Topologie \mathcal{S} für X/\sim , mit der die Abbildung π stetig von (X, \mathcal{T}_X) nach $(X/\sim, \mathcal{S})$ geht, Teilmenge von \mathcal{T}_X/\sim , denn ist $A \in \mathcal{S}$, dann ist $\pi^{-1}(A)$ in \mathcal{T}_X , denn π ist*

ja stetig, also ist $A \in \mathcal{T}_X/\sim$. Also gilt $\mathcal{S} \subseteq \mathcal{T}_X/\sim$ und somit ist \mathcal{S} gröber als \mathcal{T}_X/\sim . Folglich ist \mathcal{T}_X/\sim die feinste Topologie, so dass die natürliche Projektion noch stetig ist.

In diesem Beispiel hatten wir eine Abbildung von einem topologischen Raum (X, \mathcal{T}_X) in eine Menge X/\sim hinein. Umgekehrt kann aber auch die Abbildung aus einer Menge heraus in einen topologischen Raum gehen:

Beispiel 4.6 (Teilraum). Sei $\underline{X} := (X, \mathcal{T}_X)$ topologischer Raum und A eine Teilmenge von X . Dann ist die folgende Menge $\{A \cap U \mid U \in \mathcal{T}_X\}$ eine Topologie für A , wie man leicht überprüft. Sie heißt Teilraumtopologie oder Spurtopologie von \mathcal{T}_X für A und wird mit $\mathcal{T}_X|_A$ bezeichnet. Es ist $(X, \mathcal{T}_X)|_A := (A, \mathcal{T}_X|_A)$. Hier ist die Inklusionsabbildung $i : A \rightarrow X, a \mapsto a$ eine stetige Abbildung. Denn ist U offen in X , dann ist

$$i^{-1}(U) = \{a \in A \mid i(a) \in U\} = \{a \mid a \in A, a \in U\} = A \cap U \in \mathcal{T}_X|_A.$$

Ist umgekehrt eine Topologie \mathcal{T} auf A so gewählt, dass i stetig ist, dann muss zu jeder offene Menge $U \in \mathcal{T}_X$ die Menge $i^{-1}(U) = A \cap U$ in \mathcal{T} sein. Damit ist $\mathcal{T}_X|_A$ Teilmenge von \mathcal{T} . Also ist die Teilraumtopologie die gröbste Topologie, so dass die Inklusion noch stetig ist.

Offensichtlich bevorzugen Topologen bei Abbildungen, die in eine Menge hineingehen, die feinste Topologie und im anderen Fall, wenn die Abbildung aus der Menge heraus geht, die gröbste Topologie.

4.4.2 Wahl der besten Topologie

Im allgemeinen gibt es nun bei den topologischen Konstruktionen mehrere derart „typische“ Abbildungen, die dann alle stetig sein müssen. So hat z.B. das kartesische Produkt zweier Mengen $A \times B$ zwei solche Abbildungen: Die beiden Projektionen $\pi_1 : A \times B \rightarrow A, (a, b) \mapsto a$ und $\pi_2 : A \times B \rightarrow B, (a, b) \mapsto b$. Sind nun A und B die Punktmengen topologischer Räume, dann ist die Topologie für $A \times B$ so zu wählen, dass beide Abbildungen gerade noch stetig sind. Dies kann auch schrittweise geschehen: Zunächst wird je eine Topologie \mathcal{T}_{π_1} und \mathcal{T}_{π_2} gewählt, und in einem zweiten Schritt wird aus diesen eine Topologie $\mathcal{T}_{\pi_1, \pi_2}$ erzeugt, die die oben geforderte Eigenschaft hat. Wollen wir zunächst mit diesem zweiten Schritt beginnen:

Satz 4.10 (gemeinsame Verfeinerung, Vergrößerung). Sei X eine Menge, I eine Indexmenge und $(\mathcal{T}_i)_{i \in I}$ eine Familie von Topologien. Dann existiert eine gröbste Topologie $\sup(\mathcal{T}_i)_{i \in I}$, die feiner ist als jede der \mathcal{T}_i , und es existiert eine feinste Topologie $\inf(\mathcal{T}_i)_{i \in I}$, die gröber ist als jede der \mathcal{T}_i . $\sup(\mathcal{T}_i)_{i \in I}$ heißt gemeinsame Verfeinerung und $\inf(\mathcal{T}_i)_{i \in I}$ gemeinsame Vergrößerung der $(\mathcal{T}_i)_{i \in I}$. Beide Topologien sind durch $(\mathcal{T}_i)_{i \in I}$ eindeutig bestimmt.

Die gemeinsame Vergrößerung ist einfach die Schnittmenge aller angegebenen Topologien. Die gemeinsame Verfeinerung hingegen bekommen wir nicht so billig, denn anders als bei der Schnittmenge ist die Vereinigung von Topologien im allgemeinen keine Topologie. Wir können aber stattdessen die Schnittmenge all jener Topologien nehmen, die diese Vereinigung enthalten. Dies geht auch ganz allgemein:

Definition 4.16 (Erzeugte Topologie). Sei X eine Menge und $\mathcal{S} \subseteq \mathcal{P}(X)$ eine Menge von Teilmengen von X . Dann heißt

$$\mathcal{T}(\mathcal{S}) := \inf \{ \mathcal{T} \mid \mathcal{S} \subseteq \mathcal{T}, \mathcal{T} \text{ ist Topologie für } X \}$$

die von \mathcal{S} erzeugte Topologie. Ist eine Topologie \mathcal{E} von \mathcal{S} erzeugt, dann heißt \mathcal{S} auch Subbasis von \mathcal{E} .

Die gemeinsame Verfeinerung von Topologien ist nun die aus der Vereinigung dieser Topologien erzeugte Topologie. Näheres dazu findet sich u.A. bei SCHUBERT [Sch75, S.22.] und OSSA [Oss92, S.45]. Insbesondere wird jede Topologie von sich selbst erzeugt und damit gilt $\mathcal{T}(\mathcal{T}(\mathcal{S})) = \mathcal{T}(\mathcal{S})$.

Ein Beispiel für eine gemeinsame Verfeinerung ist die Struktur, die sich auf einer Geschossplatte durch das Geschoss oberhalb und das Geschoss unterhalb dieser Platte abzeichnet, wie z.B. auf Abbildung 4.1 zu sehen ist. Das Muster auf der Geschossplatte ist sowohl feiner als dasjenige des Grundrisses darüber, als auch feiner als der Grundriss darunter. Gleichzeitig ist jedes andere Muster, das feiner ist als beide angrenzenden Geschosse (etwa bei Hinzufügen von Durchbrüchen oder Unterzügen), auch feiner als die gemeinsame Verfeinerung der Geschosse. In GIS-Systemen wird die gemeinsame Verfeinerung von Topologien als „Vereinigungsüberlagerung“ bezeichnet.

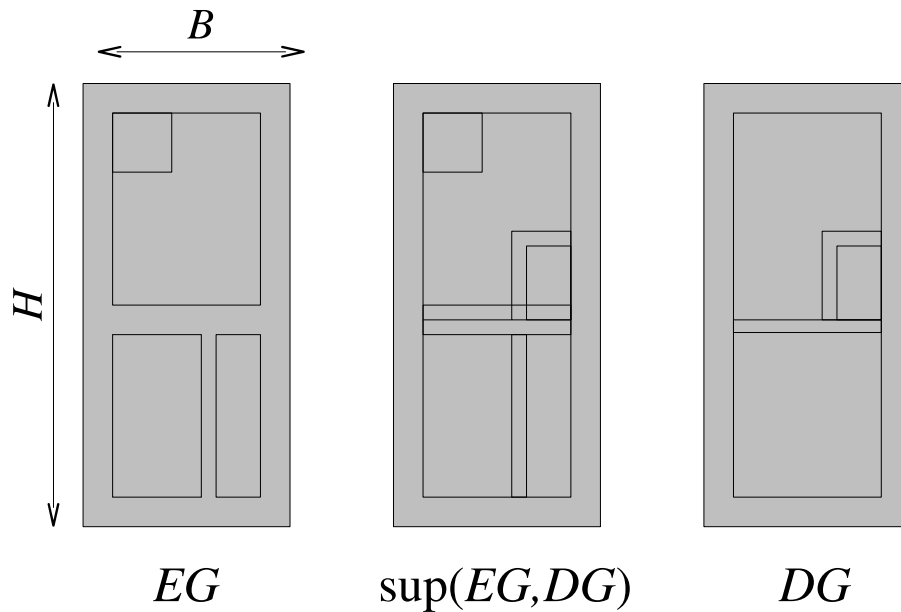


Abbildung 4.1: *Gemeinsame Verfeinerung:* Die gemeinsame Verfeinerung $\sup(EG, DG)$ der zwei Topologien EG und DG für eine Punktmenge $B \times H$ (graues Rechteck), kann man sich als das Muster auf einer Geschossplatte zwischen EG und DG vorstellen. $\sup(EG, DG)$ unterteilt $B \times H$ sowohl feiner als EG als auch feiner als DG .

Wir werden später noch folgende Eigenschaft der gemeinsamen Verfeinerung benötigen:

Lemma 4.11 (Assoziativität der Verfeinerung (Vergrößerung)). *Seien \mathcal{Q} , \mathcal{R} und \mathcal{S} Topologien auf einer Menge X , dann gilt*

$$\sup \{ \sup \{ \mathcal{Q}, \mathcal{R} \}, \mathcal{S} \} = \sup \{ \mathcal{Q}, \sup \{ \mathcal{R}, \mathcal{S} \} \}.$$

Sinngemäßes gilt auch für die gemeinsame Vergrößerung.

Bei der gemeinsamen Vergrößerung ist dies klar, da diese einfach der Schnitt der Topologien ist, und Schnittmengenbildung ist assoziativ. Bei der Verfeinerung gilt nun folgendes: Ist $\mathcal{B} \subseteq \mathcal{C}$, dann gilt für die erzeugten Topologien $\mathcal{T}(\mathcal{B}) \subseteq \mathcal{T}(\mathcal{C})$, denn jede Topologie, die \mathcal{C} umfasst, die umfasst auch \mathcal{B} als dessen Teilmenge. Sei nun $\{\mathcal{T}_{\mathcal{B}}\}$ Menge der Topologien, die \mathcal{B} umfassen und $\{\mathcal{T}_{\mathcal{C}}\}$ die entsprechende Menge für \mathcal{C} , dann gilt also $\{\mathcal{T}_{\mathcal{B}}\} \supseteq \{\mathcal{T}_{\mathcal{C}}\}$, die Teilmengenrelation hat sich also umgekehrt. Gilt jedoch für Mengensysteme $\mathcal{X} \subseteq \mathcal{Y}$, dann gilt für deren Schnitt $\bigcap \mathcal{X} \supseteq \bigcap \mathcal{Y}$, die Teilmengenrelation hat sich also erneut umgekehrt. Die von \mathcal{B} erzeugte Topologie haben wir aber als einen solchen Schnitt von Topologien definiert. Diese Beobachtung wollen wir nun zum Beweis der Assoziativität ausnutzen:

Beweis. Wir schreiben hier $\mathcal{T}(\mathcal{Q} \cup \mathcal{R})$ für $\sup \{ \mathcal{Q}, \mathcal{R} \}$.

Es reicht zu zeigen, dass $\mathcal{T}(\mathcal{R} \cup \mathcal{S}) = \mathcal{T}(\mathcal{T}(\mathcal{R}) \cup \mathcal{S})$ gilt, denn wenn wir dann \mathcal{R} durch $\mathcal{Q} \cup \mathcal{R}$ ersetzen erhalten wir

$$\mathcal{T}(\mathcal{Q} \cup \mathcal{R} \cup \mathcal{S}) = \mathcal{T}((\mathcal{Q} \cup \mathcal{R}) \cup \mathcal{S}) = \mathcal{T}(\mathcal{T}(\mathcal{Q} \cup \mathcal{R}) \cup \mathcal{S})$$

woraus wiederum sofort die Behauptung folgt.

Nun ist $\mathcal{R} \subseteq \mathcal{R} \cup \mathcal{S}$ und folglich gilt $\mathcal{T}(\mathcal{R}) \subseteq \mathcal{T}(\mathcal{R} \cup \mathcal{S})$. Damit gilt ebenfalls $\mathcal{T}(\mathcal{R}) \cup \mathcal{S} \subseteq \mathcal{T}(\mathcal{R} \cup \mathcal{S}) \cup \mathcal{S}$. Wegen $\mathcal{S} \subseteq \mathcal{T}(\mathcal{R} \cup \mathcal{S})$ gilt aber $\mathcal{T}(\mathcal{R} \cup \mathcal{S}) \cup \mathcal{S} = \mathcal{T}(\mathcal{R} \cup \mathcal{S})$ und folglich $\mathcal{T}(\mathcal{R}) \cup \mathcal{S} \subseteq \mathcal{T}(\mathcal{R} \cup \mathcal{S})$. Damit gilt $\mathcal{T}(\mathcal{T}(\mathcal{R}) \cup \mathcal{S}) \subseteq \mathcal{T}(\mathcal{T}(\mathcal{R} \cup \mathcal{S}))$. Da aber eine Topologie sich selbst erzeugt gilt $\mathcal{T}(\mathcal{T}(\mathcal{R} \cup \mathcal{S})) = \mathcal{T}(\mathcal{R} \cup \mathcal{S})$ und wir haben

$$\mathcal{T}(\mathcal{T}(\mathcal{R}) \cup \mathcal{S}) \subseteq \mathcal{T}(\mathcal{R} \cup \mathcal{S}).$$

Umgekehrt gilt $\mathcal{R} \subseteq \mathcal{T}(\mathcal{R})$ und folglich $\mathcal{R} \cup \mathcal{S} \subseteq \mathcal{T}(\mathcal{R}) \cup \mathcal{S}$ und schließlich

$$\mathcal{T}(\mathcal{R} \cup \mathcal{S}) \subseteq \mathcal{T}(\mathcal{T}(\mathcal{R}) \cup \mathcal{S}).$$

Damit erhalten wir

$$\mathcal{T}(\mathcal{T}(\mathcal{R}) \cup \mathcal{S}) = \mathcal{T}(\mathcal{R} \cup \mathcal{S}).$$

□

Bevor wir nun topologische Konstruktionen mit Datenbanken vornehmen, müssen wir die gemeinsame Verfeinerung und Vergrößerung bei topologischen Datentypen definieren.

Satz 4.12 (Verfeinerung, Vergrößerung in \mathcal{DTop}). *Sei X eine Menge und $\{R_i\}_{i \in I}$ eine Familie von Relationen in X . Seien \mathcal{T}_{R_i} die jeweils von den R_i erzeugten Topologien. Dann erzeugt $\inf \{R_i\}_{i \in I} := \bigcup_{i \in I} R_i$ die gemeinsame Vergrößerung der R_i und ist I endlich, dann erzeugt $\sup \{R_i\}_{i \in I} := \bigcap_{i \in I} R_i^+$ die gemeinsame Verfeinerung der Topologien \mathcal{T}_{R_i} . Es gilt also*

$$\begin{aligned} \mathcal{T}_{\inf \{R_i\}_{i \in I}} &= \inf \{ \mathcal{T}_{R_i} \}_{i \in I} \\ \mathcal{T}_{\sup \{R_i\}_{i \in I}} &= \sup \{ \mathcal{T}_{R_i} \}_{i \in I} \text{ falls } I \text{ endlich} \end{aligned}$$

Zur Bestimmung der gemeinsamen Verfeinerung ist im allgemeinen die Berechnung der transitiven Hüllen R_i^+ der Relationen R_i für $i \in I$ notwendig.

Im Falle unendlich vieler Relationen kann deren gemeinsame Verfeinerung zu fein werden, denn es gibt Familien $\{R_i\}_{i \in I}$ von Relationen mit

$$\mathcal{T}_{\sup \{R_i\}_{i \in I}} \not\supseteq \sup \{\mathcal{T}_{R_i}\}_{i \in I}.$$

Bei den gemeinsamen Verfeinerungen und später bei den Initialtopologien werden wir uns also auf endlich viele Topologien bzw. endlich viele Abbildungen beschränken.

Im obigen Satz zeigt sich — ganz analog zu den Topologien —, dass die Konstruktion der gemeinsamen Vergrößerung einfacher ist als die der gemeinsamen Verfeinerung: Bei der gemeinsamen Verfeinerung von Relationen muss im allgemeinen noch die transitive Hülle von jeder der beteiligten Relationen bestimmt werden. Dies ist eine auffällige Parallele zu der Tatsache, dass die gemeinsame Verfeinerung von Topologien aus der Vereinigung dieser Topologien erst noch erzeugt werden muss, während wir die gemeinsame Vergrößerung bereits mit deren Schnittmenge erhalten. Im Gegensatz zu den Topologien ist eine gemeinsame Verfeinerung oder Vergrößerung von Relationen im Allgemeinen nicht eindeutig bestimmt. Doch alle gemeinsamen Verfeinerungen oder Vergrößerungen einer Familie von Relationen sind äquivalent und erzeugen die gleiche Topologie. Ist diese T_0 -trennbar [Ale37, S.503] und endlich, dann existiert jeweils sogar eine eindeutig bestimmte minimale Relation, die auch mit dem von CODD vorgeschlagenen OPEN-Operator [Cod79, S.427] berechnet werden kann. An dieser Stelle wird übrigens auch ein dazu dualer Operator CLOSE zur Berechnung der transitiven Hülle vorgeschlagen.

Beweis (von Satz 4.12). Es ist klar, dass $\inf \{R_i\}_{i \in I}$ gröber als jedes R_i ist. Damit bleibt zu zeigen, dass jede andere Relation, die gröber als jedes R_i ist, auch gröber als $\inf \{R_i\}_{i \in I}$ ist. Diese ist dann bis auf Äquivalenz die feinste Relation mit dieser Eigenschaft. Klar ist auch, dass $\sup \{R_i\}_{i \in I}$ feiner als jedes R_i ist, und es muss auch hier nur noch gezeigt werden, dass jede andere Relation mit dieser Eigenschaft auch feiner als $\sup \{R_i\}_{i \in I}$ ist.

Sei eine Menge A offen in allen (X, \mathcal{T}_{R_i}) , dann gilt für alle $i \in I$: $\text{St}_{R_i}(A) \subseteq A$. Ist nun $a \in A$ beliebig, dann gibt es ein R_i und ein $b \in A$, so dass $a R_i b$ gilt. Die Relation $\inf \{R_i\}_{i \in I}$ ist Vereinigung aller R_i und damit gilt: $a \inf \{R_i\} b$ und somit auch $a \in \text{St}_{\inf \{R_i\}}(A)$, also $\text{St}_{\inf \{R_i\}}(A) \subseteq A$ und damit ist A offen in der von $\inf \{R_i\}$ erzeugten Topologie. Sei nun umgekehrt A offen in dieser Topologie, dann gilt $\text{St}_{\inf \{R_i\}}(A) \subseteq A$. Wählen wir ein beliebiges R_i aus und sei $a \in \text{St}_{R_i}(A)$. Dann gilt $a R_i b$ für ein $b \in B$. Dann ist jedoch auch $a \inf \{R_i\} b$ damit $a \in \text{St}_{\inf \{R_i\}}(A)$ und folglich $a \in A$. Damit ist A offen in (X, \mathcal{T}_{R_i}) , und da R_i beliebig gewählt wurde, gilt dies für alle R_i . Damit ist A genau dann Element von $\inf \{R_i\}_{i \in I}$ erzeugten Topologie, wenn A Element jeder von einem R_i erzeugten Topologie ist.

$\inf \{R_i\}_{i \in I}$ ist Vereinigung aller R_i und somit ist jedes R_i Teilmenge von $\inf \{R_i\}_{i \in I}$ und damit auch Teilmenge von $(\inf \{R_i\}_{i \in I})^*$. Also ist jedes R_i feiner als $\inf \{R_i\}_{i \in I}$. Sei nun ein beliebiges G gröber als jedes R_i , also $R_i \subseteq G^*$. Ist nun $(a, b) \in \inf \{R_i\}_{i \in I}$, dann ist $(a, b) \in R_i$ für ein $i \in I$ und somit $(a, b) \in G^*$. Also gilt $\inf \{R_i\}_{i \in I} \subseteq G^*$ und daher ist G gröber als die gemeinsame Vergrößerung.

$\sup \{R_i\}_{i \in I}$ ist Schnitt der transitiven Hüllen R_i^+ der R_i . Damit ist $\sup \{R_i\}_{i \in I} \subseteq R_i^+ \subseteq R_i^*$ und somit feiner als jedes R_i . Sei nun eine Relation F ebenfalls feiner als jedes R_i , also $F \subseteq R_i^*$. Dann ist F feiner als $\sup \{R_i\}_{i \in I}$,

denn sei $(a, b) \in F$, dann können wir ohne Einschränkung $a \neq b$ annehmen, denn sonst ist sowieso $(a, b) \in \sup \{R_i\}_{i \in I}^*$. Dann ist (a, b) in jedem der R_i^* und wegen $a \neq b$ sogar in jedem der R_i^+ . Somit ist $(a, b) \in \sup \{R_i\}_{i \in I}$ und daher gilt: $F \subseteq \sup \{R_i\}_{i \in I} \subseteq \sup \{R_i\}_{i \in I}^*$. Also ist F feiner als $\sup \{R_i\}_{i \in I}$, diese also die grösste Relation, die feiner ist als alle R_i .

Die Relation $\sup \{R_i\}_{i \in I}$ erzeugt also die grösste alexandroffsche Topologie, die feiner ist als jede von R_i erzeugte Topologie. Jedoch ist nicht jede gemeinsame Verfeinerung von alexandroffschen Topologien ebenfalls alexandroffsch.

Es ist also noch zu zeigen, dass bei endlichem I die von $\sup \{R_i\}_{i \in I}$ erzeugte Topologie tatsächlich die gemeinsamer Verfeinerung der von den R_i erzeugten Topologien ist. Dies zeigen wir durch Induktion über die Anzahl $|I|$ der Elemente in I :

Sei $|I| = 0$, dann ist $\sup \{R_i\}_{i \in I} = \bigcap \{R_i\}_{i \in \emptyset} = \bigcap \emptyset = X^2$. Die Verfeinerung ist die Universalrelation und diese erzeugt die triviale Topologie $\{\emptyset, X\}$. Sie ist die grösste Topologie überhaupt und sie ist feiner als R_i für alle $i \in \emptyset$. (Wir erinnern uns, dass alle Marsmenschen nach Banane riechen).

Sei $|I| = 1$, dann ist $\{R_i\}_{i \in I} = \{R\}$ und $\sup \{R_i\}_{i \in I} = \bigcap R^+ = R^+$ erzeugt die selbe Topologie wie R und diese ist auch die „gemeinsame“ Verfeinerung von \mathcal{T}_R (mit sich selbst).

Sei $|I| = 2$, dann ist $\{R_i\}_{i \in I} = \{R_1, R_2\}$. Sei nun \mathcal{T} feiner als \mathcal{T}_{R_1} und \mathcal{T}_{R_2} . Sei zudem $A \in \mathcal{T}_{R_1^+ \cap R_2^+}$. Dann ist

$$\begin{aligned} A &= \text{St}_{R_1^+ \cap R_2^+}^*[A] \\ &= \bigcup_{a \in A} \{x \in X \mid x(R_1^+ \cap R_2^+)^* a\} \\ &= \bigcup_{a \in A} \{x \in X \mid x R_1^* a, x R_2^* a\} \\ &= \bigcup_{a \in A} (\{x \in X \mid x R_1^* a\} \cap \{x \in X \mid x R_2^* a\}) \\ &= \bigcup_{a \in A} (\text{St}_{R_1}^*({a}) \cap \text{St}_{R_2}^*({a})). \end{aligned}$$

Die Gleichung $(R_1^+ \cap R_2^+)^* = R_1^* \cap R_2^*$ überlegt man sich leicht selbst. Nun ist jedes der $\text{St}_{R_1}^*({a})$ Element von \mathcal{T}_{R_1} und jedes $\text{St}_{R_2}^*({a})$ ist Element von \mathcal{T}_{R_2} . Da \mathcal{T} feiner ist als diese beiden Topologien, sind diese Sterne auch Element von \mathcal{T} . Da \mathcal{T} selbst eine Topologie ist, enthält sie auch deren Schnittmenge $\text{St}_{R_1}^*({a}) \cap \text{St}_{R_2}^*({a})$. Dann ist auch A als Vereinigung dieser Schnittmengen ebenfalls Element von \mathcal{T} . Folglich ist \mathcal{T} feiner als $\mathcal{T}_{R_1^+ \cap R_2^+}$. Diese ist somit die grösste Topologie, die feiner ist als \mathcal{T}_{R_1} und \mathcal{T}_{R_2} .

Gelte nun, für eine natürliche Zahl n , dass aus $\mathcal{R} = \{R_1, \dots, R_n\}$ folgt, dass $\sup \mathcal{R}$ die gemeinsame Verfeinerung der $\{\mathcal{T}_{R_1}, \dots, \mathcal{T}_{R_n}\}$ erzeugt, sei aber $\mathcal{S} = \{R_1, \dots, R_{n+1}\}$. Dann ist die gemeinsame Verfeinerung erzeugt von

$$\begin{aligned} \sup \mathcal{S} &= \bigcap \{R_1^+, \dots, R_{n+1}^+\} \\ &= (\bigcap \{R_1^+, \dots, R_n^+\})^+ \cap R_{n+1}^+ \\ &= \sup \{\sup \{R_1, \dots, R_n\}, R_{n+1}\} \end{aligned}$$

unter Ausnutzung der Assoziativität von „sup“. Denn $\sup \{R_1, \dots, R_n\}$ erzeugt die gemeinsame Verfeinerung der von den jeweiligen Relationen erzeugten To-

pologien, die laut Induktionsbehauptung alexandroffsch ist, und der äußere sup-Operator erzeugt wiederum deren Verfeinerung mit $\mathcal{T}_{R_{n+1}}$, die ebenfalls alexandroffsch ist. Wir haben im Fall $n = 2$ gesehen, dass wir damit ebenfalls eine alexandroffsche Topologie erhalten. \square

Als Beispiel für eine Verfeinerung unendlich vieler alexandroffscher Topologien, die keine alexandroffsche Topologie ergibt, sei für jede natürliche Zahl $n > 0$ die Menge \mathbb{R} der reellen Zahlen mit offenen Intervallen (Kugeln) der Form

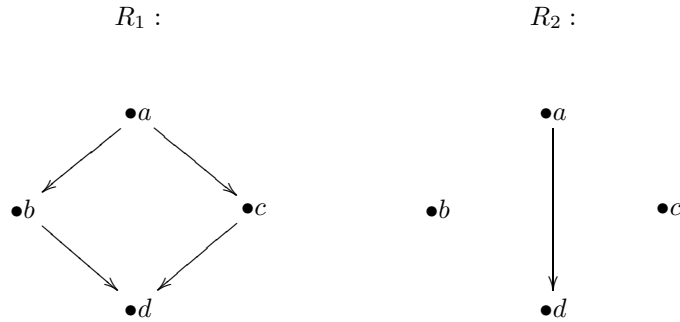
$$\mathcal{B}_n := \left\{ \left] \frac{z-1}{n}, \frac{z+1}{n} \left[\mid z \in \mathbb{Z} \right\}$$

überdeckt. Jede dieser Intervallmengen erzeugt eine alexandroffsche Topologie, die wir hier mit $\mathcal{T}_n := \mathcal{T}(\mathcal{B}_n)$ bezeichnen wollen. Die gemeinsame Verfeinerung all dieser alexandroffschen Topologien $\sup \{ \mathcal{T}_n \mid n \in \mathbb{N}, n > 0 \}$ ergibt hingegen die natürliche Topologie der reellen Zahlen und diese ist wiederum nicht alexandroffsch.

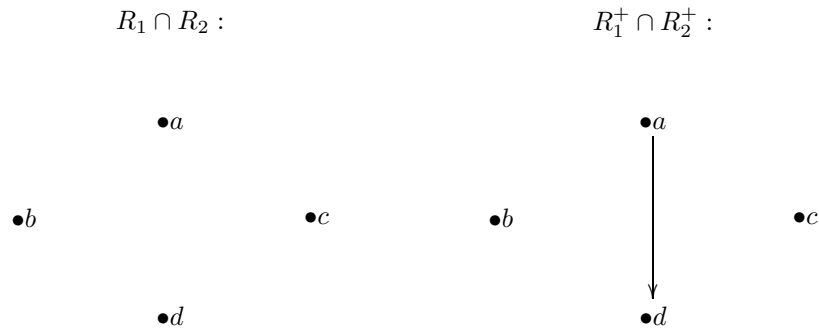
Dem Beweis, dass für die Bestimmung der gemeinsamen Verfeinerung die transitive Hülle im allgemeinen unverzichtbar ist, wollen wir nun ein eigenes Kapitel widmen.

4.4.3 Verfeinerung und transitive Hülle

Betrachten wir zunächst folgendes Beispiel mit den Relationen R_1 und R_2 :



Dann ist $R_1 \cap R_2$ echt feiner als $R_1^+ \cap R_2^+$:



Damit ist $R_1 \cap R_2$ zwar eine Verfeinerung von R_1 und R_2 , aber es ist nicht die größte Verfeinerung. Nun könnte man vermuten, dass die Berechnung der transitiven Hülle doch irgendwie umgangen werden kann, wenn denn nur die Datenbankabfrage geschickt genug formuliert wäre. Dies ist aber nicht möglich:

Satz 4.13 (Verfeinerung und transitive Hülle). *Eine relationale Algebra, mit der die gemeinsame Verfeinerung $\sup\{R_1, R_2\}$ von zwei beliebigen Relationen R_1 und R_2 berechnet werden kann, ist eine relationale Algebra mit transitiver Hülle.*

Sei also $R \subseteq X \times X$ Relation auf X , von der R^+ zu berechnen ist. Wir setzen wieder $[a:X, b:X]$ als Relationenschema für R voraus.

Zunächst wollen wir die Beweisidee skizzieren: Wir betrachten zunächst ein beliebiges Paar $(a, b) \in X \times X$. Dieses Paar bildet eine Relation $\{(a, b)\} \subseteq X \times X$, und diese ist transitiv. Sind nämlich (x, y) und (y, z) Elemente von $\{(a, b)\}$, dann gilt ja $(x, y) = (a, b) = (y, z)$ Folglich ist $x = a = y$ bzw. $x = y$ und damit $(x, z) = (y, z)$ und da $(y, z) \in \{(a, b)\}$ vorausgesetzt wurde, gilt $(x, z) \in \{(a, b)\}$.

Wenn also ein Ausdruck einer Abfragesprache in der Lage ist, die gemeinsame Verfeinerung zweier beliebiger Relationen R_1 und R_2 zu berechnen, dann kann dieser Ausdruck auch eine Relation V_{ab} erzeugen, die zu $\sup\{R, \{(a, b)\}\}$ äquivalent ist, also $V_{ab}^* = (R^+ \cap \{(a, b)\}^+)^*$. Da $\{(a, b)\}$ aber transitiv ist, gilt $V_{ab}^* = (R^+ \cap \{(a, b)\})^*$. Ist nun $a \neq b$, dann ist $(a, b) \in V_{ab}$ genau dann, wenn $(a, b) \in R^+$ gilt, und damit kann die Abfrage für jedes feste $(a, b) \in X^2$ eine

Relation V_{ab} erzeugen, die für $a \neq b$ das Paar (a, b) genau dann enthält, wenn $a R^+ b$ gilt. Wenn wir nun für alle $(a, b) \in X^2$ diese Relation V_{ab} berechnen, dann ergibt deren Vereinigung $V := \bigcup_{a,b \in X} V_{ab}$ eine Relation, die bis auf Diagonalelemente die transitive Hülle von R ist, also $V \setminus \Delta_X = R^+ \setminus \Delta_X$. Dann ist jedoch $(V \setminus \Delta_X) \cup R = R^+$. Der Ausdruck auf der linken Seite dieser Gleichung entspricht dem Ausdruck $\sigma_{a \neq b}(V) \cup R$ in relationaler Algebra. Diese Erzeugung und Vereinigung der einzelnen V_{ab} ist nun ebenfalls mit relationaler Algebra möglich, indem wir X^2 disjunkte Kopien von R erzeugen, jede dieser Kopien mit (a, b) markieren und jeder mit (a, b) markierten Kopie von R die mit (a, b) markierte Kopie von $\{(a, b)\}$ zuordnen. Die Vereinigung dieser Kopien von R ist jedoch nichts anderes als das kartesische Produkt $X \times X \times R$ und die markierten (a, b) erhalten wir durch geeignete Selektion aus $X \times X \times X \times X$. Beides kann durch Ausdrücke in relationaler Algebra berechnet werden. Kann eine Abfragesprache mit einer Abfrage V die Verfeinerung dieser beiden Relationen berechnen, dann ist diese genau die disjunkte Summe der mit (a, b) markierten Verfeinerungen von $\{(a, b)\}$ mit R . Die Vereinigung dieser markierten Einzelergebnisse erhalten wir dann, indem wir die Markierungen durch Projektion wieder beseitigen.

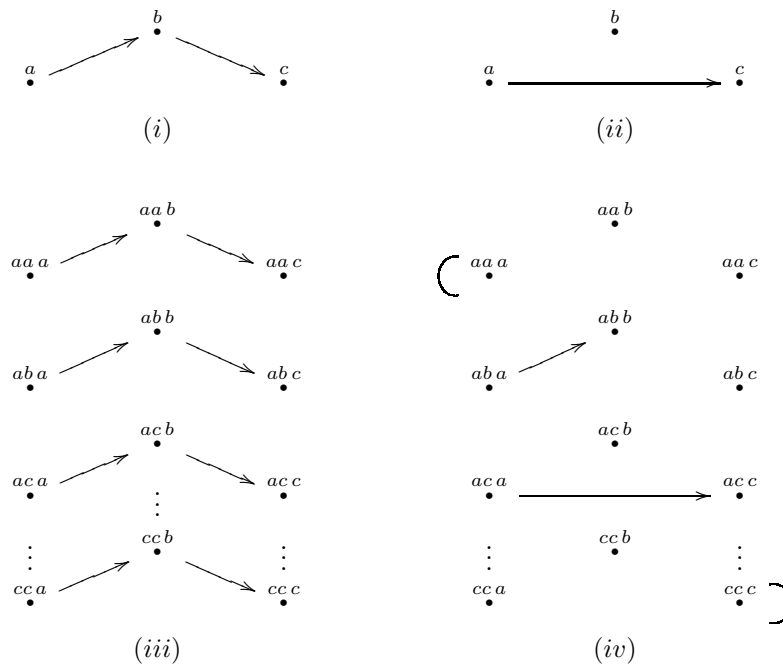


Abbildung 4.2: Konstruktion der im Beweis verwendeten Relationen: (i): Die Beispielrelation $R = \{(a \rightarrow b), (b \rightarrow c)\} \subseteq X \times X$ auf der Menge $X = \{a, b, c\}$. (ii): Die einzelne Kante $a \rightarrow c$ als Relation V_{ac} auf X . Diese Kante muss in der gemeinsamen Verfeinerung mit R enthalten sein. (iii): Die Relation $\Delta_{X^2} \otimes R$ auf X^3 . (iv): Die Relation $\{(i,j) \rightarrow (i,j)\} \subseteq X^3 \times X^3$. Jede Kante $i \rightarrow j$ wird mit i, j markiert. Es sind jeweils nur vier der insgesamt neun disjunkten Kopien von X angegeben.

Die Stelligkeit der Relation ist dabei keine Einschränkung, denn ist eine Abfrage V zur gemeinsamen Verfeinerung nur für Relationen mit einem einfachen Schema wie etwa $[a:X.id, b:X.id]$ formuliert, dann können wir aus V eine Abfrage V' erzeugen, die dies entsprechend für mehrstellige Relationen tut. Wir müssen dafür zunächst jedes Attribut a , das in V vorkommt, durch das entsprechende Tupel von Attributen wie etwa (a_1, a_2, a_3) ersetzen. Da V die gemeinsame Verfeinerung ganz allgemein berechnet, können wir von V annehmen, dass als Vergleichsoperator in den Prädikaten ausschließlich „=“ oder „ \neq “ vorkommt (siehe dazu auch MAIER [Mai83, S. 529]). Jedes so entstandene $(a_1, a_2, a_3) = (b_1, b_2, b_3)$ ersetzen wir durch $(a_1=b_1 \text{ und } a_2=b_2 \text{ und } a_3=b_3)$ und jedes $(a_1, a_2, a_3) \neq (b_1, b_2, b_3)$ durch $(a_1 \neq b_1 \text{ oder } a_2 \neq b_2 \text{ oder } a_3 \neq b_3)$ und bei den Projektionen ersetzen wir die Tupel (a_1, a_2, a_3) einfach durch a_1, a_2, a_3 . Entsprechend verfahren wir auch bei Umbenennungen.

Wollen wir zunächst eine Notation einführen, um das direkte Produkt von Abbildungen auf Relationen zu verallgemeinern.:

Definition 4.17 (Direktes Produkt von Relationen). *Seien $R \subseteq A \times B$ und $S \subseteq X \times Y$ Relationen, dann ist die Relation $R \otimes S \subseteq (A \times X) \times (B \times Y)$ definiert durch*

$$(a, x) R \otimes S (b, y) \text{ genau dann, wenn } a R b \text{ und } x S y$$

das direkte Produkt von R und S ,

Dies ist also nichts anderes als das kartesische Produkt der Graphen von R und S interpretiert als Relation von $A \times X$ nach $B \times Y$, was wir durch Umstellung der Tupel $((a, b), (x, y)) \in R \times S$ zu $((a, x), (b, y)) \in R \otimes S$ erreichen. Offensichtlich ist $R \otimes S$ lediglich eine Umbenennung von $R \times S$ und bei relationalen Datenbanken unterscheiden sich kartesisches Produkt und direktes Produkt überhaupt nicht.

Zunächst benötigen wir eine einfache Eigenschaft des direkten Relationenprodukts:

Lemma 4.14 (Produkte transitiver und reflexiver Relationen). *Seien $R \subseteq A \times A$ und $S \subseteq B \times B$, Relationen. Wenn R und S reflexiv (transitiv) sind, dann ist $R \otimes S \subseteq (A \times B) \times (A \times B)$ ebenfalls reflexiv (transitiv). Ist $R \otimes S$ nicht leer und reflexiv, dann sind auch R und S reflexiv.*

Beweis. Seien R und S die Relationen laut obigem Lemma.

Reflexiv: Seien R und S reflexiv. Ist $(a, b) \in A \times B$, dann ist $a \in A$, $b \in B$ und damit gilt $a R a$ und $b S b$ und somit auch $(a, b) R \otimes S (a, b)$ und damit ist $R \otimes S$ reflexiv. Sei nun $R \otimes S$ nicht leer und reflexiv. Dann sind auch R und S nicht leer und auch A und B nicht leer. Seien nun $a \in A$ (bzw. $b \in B$) dann gibt es ein $b \in B$ (bzw. ein $a \in A$) und damit ist $(a, b) \in A \times B$ und wegen der Reflexivität gilt $(a, b) R \otimes S (a, b)$. Dann gelten $a R a$ und $b S b$ und somit ist R (S) reflexiv.

Transitiv: Seien R und S transitiv und gelten $(a, b) R \otimes S (i, j) R \otimes S (c, d)$. Dann gelten auch $a R i R c$ und $b S j S d$ und damit auch $a R c$ und $b S d$ und folglich $(a, b) R \otimes S (c, d)$, also ist $R \otimes S$ transitiv. \square

Nehmen wir nun das kartesische Produkt $X \times X$ der Punktmenge X als Indexmenge. Dann erzeugen wir $|X \times X|$ disjunkte Kopien von X , indem wir aus

jedem $(i, j) \in X \times X$ und jedem $x \in X$ das Tupel (i, j, x) erzeugen. Wir berechnen also lediglich das kartesische Produkt $X^3 = X \times X \times X$ und betrachten es als Familie $\{X_i\}_{i \in X \times X}$ von Kopien von X . Dann machen wir uns ebenfalls $X \times X$ disjunkte Kopien von R . Die Vereinigung dieser Kopien von R hat die Form $\{((i, j, a), (i, j, b)) \mid (i, j) \in X \times X, a R b\}$ und kann wiederum als Relation auf der Vereinigung der Kopien von X betrachtet werden. Dies ist also die Relation $\Delta_{X \times X} \otimes R \subseteq X^3 \times X^3$. Zudem erzeugen wir für jedes $(i, j) \in X \times X$ die Relation $\{((i, j, i), (i, j, j)) \mid (i, j) \in X \times X\}$, ebenfalls eine Relation in X^3 . All diese Relationen können mit relationaler Algebra aus X und R leicht erzeugt werden. Wenn nun eine Relation $V \subseteq X^3 \times X^3$ gemeinsame Verfeinerung dieser beiden Relationen ist, dann ist die Projektion $\pi(V) := \{(x, y) \mid (i, j, x) V (n, k, y)\}$ eine Relation in X , mit $(\pi(V) \setminus \Delta_X) \cup R = R^+$ bzw. $\pi(V) \cup \Delta_X = R^*$. Wenn also eine Abfrage existiert, die ein solches V berechnet, existiert auch eine Abfrage für $\pi(V)$ und somit auch für R^+ und R^* . Die Abbildung 4.2 veranschaulicht die Konstruktion der beiden Relationen an einem Beispiel.

Wir nehmen also den Datentyp (X, R) und erzeugen daraus $(X^3, \Delta_{X^2} \otimes R)$. Dabei ist $\Delta_{X^2} \otimes R := \{((i, j, a), (i, j, b)) \mid (i, j) \in X^2, (a, b) \in R\}$. Sei zudem $S := \{((i, j, i), (i, j, j)) \mid i, j \in X\}$. Dann ist (X^3, S) ebenfalls ein topologischer Datentyp. Angenommen wir haben V als gemeinsame Verfeinerung von $\Delta_{X^2} \otimes R$ und S , dann ist $V^* = (S^+ \cap (\Delta_{X^2} \otimes R)^+)^*$. Zunächst behaupten wir einige Eigenschaften, die wir in unserem Beweis benötigen:

Lemma 4.15 (Eigenschaften von T , S und $\Delta_{X^2} \otimes R$). *Für die oben definierten Relationen S und $\Delta_{X^2} \otimes R$ gilt:*

1. S ist transitiv, und somit gilt $S = S^+$.
2. Für zwei beliebige Relationen P und Q und für jede natürliche Zahl n gilt: $P^n \otimes Q^n = (P \otimes Q)^n$. Somit gilt auch $\Delta_{X^2} \otimes R^+ = (\Delta_{X^2} \otimes R)^+$.
3. Sind zwei Relationen P^+ und Q^+ transitiv, dann ist auch der Schnitt $P^+ \cap Q^+$ transitiv. Insbesondere gilt dann $S \cap (\Delta_{X^2} \otimes R^+)$ ist transitiv.
4. V ist transitiv!
5. Gilt $a \neq b$ und $a Q^* b$ für eine beliebige Relation Q auf X und $a, b \in X$, dann gilt auch $a Q^+ b$. Also gilt: $Q^* \setminus \Delta_X \subseteq Q^+ \subseteq Q^*$.

Wir werden zunächst diese Eigenschaften benutzen, um unseren Satz zu beweisen:

Beweis.

$R^+ \setminus \Delta_X \subseteq \pi(V)$: Sei $(a, b) \in R^+ \setminus \Delta_X$, gelte also $a R^+ b$ und $a \neq b$. Dann ist wegen $(a, b) \Delta_{X^2} (a, b)$ auch $((a, b, a), (a, b, b)) \in \Delta_{X^2} \otimes R^+$. Es ist wegen der Definition von S auch $(a, b, a) S (a, b, b)$. Damit ist $((a, b, a), (a, b, b)) \in S \cap (\Delta_{X^2} \otimes R^+)$. Wegen $S = S^+$ und $(\Delta_{X^2} \otimes R^+) = (\Delta_{X^2} \otimes R)^+$ gilt $(S \cap (\Delta_{X^2} \otimes R^+)) = (S^+ \cap (\Delta_{X^2} \otimes R)^+)$. Damit ist $((a, b, a), (a, b, b)) \in (S^+ \cap (\Delta_{X^2} \otimes R^+))$. Da V Verfeinerung der Relationen S und $\Delta_{X^2} \otimes R$ ist, gilt: $V^* = (S^+ \cap (\Delta_{X^2} \otimes R)^+)^*$ und folglich gilt $((a, b, a), (a, b, b)) \in V^*$. Wegen $a \neq b$ ist auch $(a, b, a) \neq (a, b, b)$, also $((a, b, a), (a, b, b)) \in V^+$. Da V aber transitiv ist, gilt $V^+ = V$ und damit $((a, b, a), (a, b, b)) \in V$ und somit $(a, b) \in \pi(V)$.

$\pi(V) \subseteq R^*$: Sei nun $(a, b) \in \pi(V)$. Ist $a = b$, dann gilt trivialerweise $a R^* b$. Wir müssen also nur noch den Fall $a \neq b$ untersuchen. Ist $(a, b) \in \pi(V)$, dann existieren $x, y, u, v \in X$ mit $(x, y, a) V (u, v, b)$. Wegen $V^* = (S \cap (\Delta_{X^2} \otimes R^+))^*$ ist $((x, y, a), (u, v, b)) \in (S \cap (\Delta_{X^2} \otimes R^+))^*$, und wegen $a \neq b$ ist auch $(x, y, a) \neq (u, v, b)$ und damit auch $((x, y, a), (u, v, b)) \in (S \cap \Delta_{X^2} \otimes R^+)^+$. Die Relation $(S \cap \Delta_{X^2} \otimes R^+)$ ist jedoch transitiv, da sie Schnitt zweier transitiver Relationen ist, und damit gilt: $((x, y, a), (u, v, b)) \in (S \cap \Delta_{X^2} \otimes R^+)$. Daraus folgt $(x, y, a) \Delta_{X^2} \otimes R^+ (u, v, b)$ und damit auch $a R^+ b$. Daraus folgt nun $a R^* b$.

□

Damit haben wir gezeigt, dass eine Abfrage, die V erzeugt, ein Ergebnis liefert, aus dem sich mit den Standardoperatoren der relationalen Algebra die transitive Hülle von R berechnen lässt, denn V unterscheidet sich nur durch Diagonalelemente von R^+ .

Wir zeigen nun noch im einzelnen die Aussagen in unserem obigen Lemma.

S ist transitiv: Jedes Paar $((a, b, c), (d, e, f))$ der Relation S liegt in der Form $((a, b, a), (a, b, b))$ vor. Gelte nun $(a, b, x) S (a, b, y) S (a, b, z)$. Dann gilt $a = x$ und $b = y$ wegen $(a, b, x) S (a, b, y)$ und aus $(a, b, y) S (a, b, z)$ folgt $a = y$ und $b = z$. Insbesondere ist $x = a = y$ und folglich auch $(a, b, x) = (a, b, y)$ und damit $(a, b, x) S (a, b, z)$.

$P^n \otimes Q^n = (P \otimes Q)^n$: Der Beweis durch Induktion über n ist einfach:

Ist $n = 0$, dann ist $(P \otimes Q)^n$ die Gleichheitsrelation. Gelte $(a, x) (P^0 \otimes Q^0) (b, y)$ dann ist $a P^0 b$ und $x Q^0 y$, also $a = b$ und $x = y$ und somit $(a, x) = (b, y)$ bzw. $(a, x) (P \otimes Q)^0 (b, y)$.

Gilt die Behauptung für ein n , so gilt $(a, x) (P \otimes Q)^{n+1} (b, y)$ genau dann, wenn ein (c, z) existiert mit $(a, x) (P \otimes Q)^n (c, z) (P \otimes Q) (b, y)$. Genau dann gilt wegen der Behauptung $(a, x) (P^n \otimes Q^n) (c, z) (P \otimes Q) (b, y)$. Also $a P^n c P b$ und $x Q^n z Q y$ und damit $a P^{n+1} b$ und $x Q^{n+1} y$ oder $(a, x) (P^{n+1} \otimes Q^{n+1}) (b, y)$.

$P^+ \cap Q^+$ ist transitiv: Seien P und Q Relationen auf X . Dann sind P^+ und Q^+ transitiv. Gilt $a (P^+ \cap Q^+) b (P^+ \cap Q^+) c$, dann auch $a P^+ b P^+ c$ und $a Q^+ b Q^+ c$. Da P^+ und Q^+ transitiv sind, gilt $a P^+ c$ und $a Q^+ c$ und somit auch $a (P^+ \cap Q^+) c$.

V ist transitiv: Es gilt $V \subseteq S^*$ wegen $V \subseteq V^* = (S^+ \cap (\Delta_X \otimes R)^+)^*$ und $S^+ \cap (\Delta_X \otimes R)^+ \subseteq S^+ = S$. Gelte $(a, b, c) V (d, e, f)$ und $(d, e, f) V (g, h, i)$. Dann ist $(a, b, c) S^* (d, e, f)$ und $(d, e, f) S^* (g, h, i)$.

Ist $(a, b, c) = (d, e, f)$ dann gilt auch $(a, b, c) V (g, h, i)$, denn es wurde ja $(d, e, f) V (g, h, i)$ angenommen.

Ist jedoch $(a, b, c) \neq (d, e, f)$ dann ist $(a, b, c) S^+ (d, e, f) S^* (g, h, i)$.

Wegen $S^+ = S$, gilt also $(a, b, c) S (d, e, f) S^* (g, h, i)$ und daher $(a, b, c) S^+ (g, h, i)$ und somit $(a, b, c) S (g, h, i)$. Wegen $(a, b, c) S (d, e, f)$ und zugleich $(a, b, c) S (g, h, i)$ gilt $(d, e, f) = (g, h, i)$, denn laut Definition haben alle Paare in S die Form $((a, b, a), (a, b, b))$. Also gilt auch hier $(a, b, c) V (g, h, i)$ und somit ist V transitiv.

$Q^* \setminus \Delta_X \subseteq Q^+$: Dies ist elementare Mengenlehre, denn $Q^* = Q^+ \cup \Delta_X$.

Wir haben nun gesehen, dass zur Berechnung der gemeinsamen Verfeinerung im allgemeinen auf die transitive Hülle nicht verzichtet werden kann. Diese Einschränkung ist jedoch nur theoretisch. In der Praxis werden wir es mit Topologien zu tun haben, die von Räumen mit einer festen oberen Dimensionsschranke stammen, die zudem noch relativ klein ist. In diesem Sonderfall kann die transitive Hülle auch mit konventionellen Abfragesprachen berechnet werden, denn ist $R[a, b]$ Relation eines n -dimensionalen Komplexes, dann ist

$$\begin{aligned} R^1 &:= \beta_{R^1.a, R^1.b \leftarrow R.a, R.b}(R) \\ R^m &:= \beta_{R^m.a, R^m.b \leftarrow R^{m-1}.a, R.b} \pi_{R^{m-1}.a, R.b}(R^{m-1} \bowtie_{R^{m-1}.b=R.a} R) \\ R^+ &= R^n \cup R^{n-1} \cup \dots \cup R, \end{aligned}$$

wobei wir die impliziten Umbenennungen der Form $\beta(S) := \beta_{S.a, S.b \leftarrow a, b}(S)$ stillschweigend weggelassen haben. Diese Dimensionsschranke ist dann aber eine Eigenschaft der gespeicherten Daten und nicht des Datenbankschemas. Unser Beweis bezog sich jedoch auf das Schema.

Da wir bisher stets behauptet haben, bestimmte Konstruktionen ließen sich „leicht mit relationaler Algebra berechnen“, geben wir hier der Vollständigkeit halber noch die entsprechenden Abfragen zur Konstruktion dieser Relationen an: Mit $[\text{id}:D]$ als Relationenschema von X und $[\mathbf{a}:X.\text{id}, \mathbf{b}:X.\text{id}]$ als Relationenschema von R erhalten wir

$$\begin{aligned} X^3 &:= \beta_{\mathbf{i} \leftarrow \text{id}}(X) \times \beta_{\mathbf{j} \leftarrow \text{id}}(X) \times X \\ \Delta_{X^2} \otimes R &:= \beta_{\mathbf{i}, \mathbf{j} \leftarrow \text{id}, \text{id}}(X) \times R \\ S &:= \beta_{\mathbf{i}, \mathbf{a} \leftarrow \text{id}, \text{id}}(X) \times \beta_{\mathbf{j}, \mathbf{b} \leftarrow \text{id}, \text{id}}(X). \end{aligned}$$

4.4.4 Verallgemeinerte topologische Konstruktion

Gemeinsame Verfeinerung und Vergrößerung wird benötigt, um aus gegebenen topologischen Räumen neue Räume zu konstruieren. Wir wollen nun unsere Beispiele Quotiententopologie und Teilraumtopologie verallgemeinern:

Definition 4.18 (Initial- und Finaltopologie). Seien X und Z Mengen und sei $\{(Y_i, \mathcal{T}_i)\}_{i \in I}$ und eine Familie von topologischen Räumen.

Zu einer Familie $f = \{f_i : X \rightarrow (Y_i, \mathcal{T}_i)\}_{i \in I}$ von Abbildungen von der Menge X in diese Räume gibt es eine grösste Topologie \mathcal{I}_f für X , so dass alle Abbildungen f_i stetig sind.

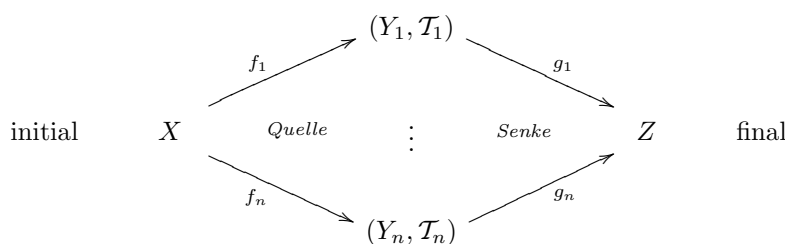
Zu einer Familie $g = \{g_i : (Y_i, \mathcal{T}_i) \rightarrow Z\}_{i \in I}$ von Abbildungen von den Räumen in die Menge Z gibt es eine feinste Topologie \mathcal{F}_g , so dass alle Abbildungen g_i stetig sind.

\mathcal{I}_f heißt dann Initialtopologie für X bezüglich f und \mathcal{F}_g heißt Finaltopologie für Z bezüglich g . Die Abbildungsfamilie f heißt auch Quelle und g wird als Senke bezeichnet.

Die Topologien \mathcal{I}_f und \mathcal{F}_g sind durch f und g eindeutig bestimmt.

Die Bezeichnung *initial* oder *final* bezieht sich auf das Ende des Abbildungspfeils, an dem die Topologie „sitzt“. Ist diese Topologie im Definitionsbereich der Abbildung, dann sitzt sie am Anfangspunkt des Abbildungspfeils und heißt

demzufolge „initial“. Die Topologie im Wertebereich am Endpunkt des Abbildungspfeils heißt entsprechend „final“. OSSA bezeichnet die Initialtopologie auch als *induzierte* und die Finaltopologie als *Identifizierungstopologie* [Oss92, §4] (Dort findet sich auch ein Beweis der Eindeutigkeit). Die Quotiententopologie ist also die Finaltopologie bezüglich der natürlichen Projektion und die Teilraumtopologie die Initialtopologie bezüglich der Inklusion. Folgendes Diagramm veranschaulicht die Definition.



Ohne Beweis stellen wir folgendes fest (nachzulesen bei OSSA Satz 2.2.8 und Satz 2.2.14 [Oss92]): Man kann bei einer Familie $(f_i)_{i \in I}$ von Abbildungen die Initial- bzw. Finaltopologie \mathcal{T}_i für jedes i einzeln bestimmen und dann jeweils die gemeinsame Verfeinerung oder Vergrößerung bestimmen. Denn die gemeinsame Verfeinerung der einzelnen Initialtopologien $\{\mathcal{T}_i \mid i \in I\}$ ist die Initialtopologie der Abbildungsfamilie. Umgekehrt ist die gemeinsame Vergrößerung der einzelnen Finaltopologien bezüglich der Abbildungen einer Familie $(g_j)_{j \in J}$ auch die Finaltopologie bezüglich dieser Abbildungsfamilie. Damit ist es möglich, sich zunächst auf die Bestimmung der einzelnen Topologien zu beschränken und erst dann entsprechend zu verfeinern oder zu vergrößern.

Satz 4.16 (Initial- und Finaltopologien in $\mathcal{D}Top$). *Seien (Y, R) ein topologischer Datentyp, X und Z Mengen und $f : X \rightarrow (Y, R)$ und $g : (Y, R) \rightarrow Z$ Abbildungen (z.B. Fremdschlüssel oder Projektion). Dann erzeugt $(g \times g)(R)$ die Finaltopologie für Z bezüglich g und $(f \times f)^{-1}(R^+)$ die Initialtopologie für X bezüglich f . Die Berechnung der transitiven Hülle von R ist zur Berechnung der Initialtopologie im Allgemeinen erforderlich. Ist jedoch f surjektiv, dann erzeugt auch $(f \times f)^{-1}(R)$ die Initialtopologie bezüglich f .*

Beispiel 4.7 (Initial und Finaltopologien mit SQL). *Wir werden die Bezeichnungen des obigen Satzes beibehalten und die beteiligten Räume und Abbildungen als relationale Datenbank modellieren und die entsprechenden Abfragen in SQL angeben. Dies soll insbesondere zeigen, dass diese Arbeit in der Tat von Datenbanken handelt. Wir nehmen an, es gäbe Domänen D_X , D_Y und D_Z , die sich nicht notwendigerweise unterscheiden müssen. Wir bezeichnen mit X , Y , Z und R sowohl das Relationenschema als auch die entsprechenden Relationen.*

Sei $Z[\underline{idz} : D_Z]$ Relationenschema einer Relation Z . Sei zudem $Y[\underline{idY} : D_Y, g : Z.\underline{idz}]$ Schema einer Relation Y mit Fremdschlüssel g nach Z und entsprechend $X[\underline{idX} : D_X, f : Y.\underline{idY}]$ Schema der Relation X mit Fremdschlüssel f nach Y . Zudem habe R Relationenschema $R[\underline{a} : Y.\underline{idY}, \underline{b} : Y.\underline{idY}]$.

Die Fremdschlüsselattribute f und g dürfen dabei keine sogenannten null-Werte haben, denn sonst wären die entsprechenden Abbildungen keine totalen Abbildungen. Nur für diese haben wir aber die initialen und finalen Relationen

definiert. Auf die Problematik von Fremdschlüsseln mit `null`-Werten wollen wir erst im Abschnitt 4.5.5 eingehen.

Dann ist

```
select  A.g as a, B.g as b
from    Y as A, R, Y as B
where   A.idy=R.a and B.idy=R.b ;
```

eine finale Relation für Z bezüglich g und es ist

```
select      distinct A.idx as a, B.idx as b
from        X A, X B
where exists (select *
              from R
              where R.b=B.f
              start with A.f=R.a
              connect by nocycle prior R.b=R.a);
```

die Relation $(f \times f)^{-1}(R^+)$ – eine initiale Relation für X bezüglich f . Dabei werden die `start with`- und `connect by`-Klauseln von Oracle-SQL [LG03, Kap.9, S.3-9] benutzt, um die transitive Hülle von R abzufragen. (In Oracle SQL darf zum Umbenennen der gesamten Relationen im `from`-Teil das Schlüsselwort `as` nicht verwendet werden.)

Ist f surjektiv, dann können diese Klauseln auch weggelassen werden, und wir erhalten eine Abfrage, die der Abfrage der Finaltopologie stark ähnelt. Nur Funktionsbezeichner f und Schlüsselattribute `idx` sind vertauscht:

```
select  distinct A.idx as a, B.idx as b
from    X A, R, X B
where   A.f=R.a and B.f=R.b;
```

Damit ist insbesondere die Initial- und Finaltopologie bezüglich einer einzelnen Abbildung mit einem alexandroffschen Raum ebenfalls alexandroffsch.

Geben wir den Relationen, welche die Initial- und Finaltopologien erzeugen, noch geeignete Bezeichnungen:

Definition 4.19 (Initiale und finale Relationen). Sei Y eine Menge und seien $\{(X_i, R_i)\}_{i \in I}$ und $\{(Z_j, T_j)\}_{j \in J}$ topologischen Datenbanken und $f = \{f_i : X_i \rightarrow Y\}_{i \in I}$ und $g = \{g_j : Y \rightarrow Z_j\}_{j \in J}$ Familien von Abbildungen. Eine Relation $F \subseteq Y \times Y$ heißt final bezüglich f , wenn sie die entsprechende Finaltopologie erzeugt. Eine Relation $I \subseteq Y \times Y$ heißt initial bezüglich g , wenn sie die entsprechende Initialtopologie erzeugt.

Dann werden wir nun den vorhergehenden Satz 4.16 beweisen, der uns ja die grundlegenden Datenbankabfragen für topologische Konstruktionen liefert. Der Beweis bezüglich der Finaltopologie ist einfach:

Beweis. Finaltopologie: $(f \times f)[R]$ ist final, denn $(f \times f)[R] \subseteq (f \times f)[R]^*$ ist offensichtlich und damit ist $f : (X, R) \rightarrow (Y, (f \times f)[R])$ stetig. Zudem ist $(f \times f)[R]$ die feinste Relation mit dieser Eigenschaft, denn ist $f : (X, R) \rightarrow (Y, T)$ stetig, dann gilt ja $(f \times f)[R] \subseteq T^*$ und somit ist $(f \times f)[R]$ feiner als T .

Bei der Initialtopologie ist die Sache wieder ein bisschen aufwändiger: Wir wissen bereits, dass für jede Abbildung $m : X \rightarrow Y$ und für jede Menge $B \subseteq Y$

gilt: $m[m^{-1}[B]] \subseteq B$. Damit gilt auch $(g \times g)[(g \times g)^{-1}[S^+]] \subseteq S^+ \subseteq S^*$ und damit ist g stetig.

Wir zeigen nun noch, dass diese Relation die größte ist, so dass g stetig wird. Sei $T \subseteq Y \times Y$ eine Relation, so dass $g : (Y, T) \rightarrow (Z, S)$ stetig ist. Dann gilt $(g \times g)(T) \subseteq S^*$. Wir betrachten nun ein beliebiges Paar $(a, b) \in T$ mit $a \neq b$. Dann ist $(g(a), g(b)) \in S^*$. Ist $g(a) = g(b)$ dann ist $(g(a), g(b)) \in (g \times g)^{-1}(S^+)^*$ und ist $g(a) \neq g(b)$, dann ist $(g(a), g(b)) \in S^+$ bzw. $(g \times g)(a, b) \in S^+$ und damit $(a, b) \in (g \times g)^{-1}[S^+] \subseteq (g \times g)^{-1}[S^+]^*$. Also ist $T \subseteq (g \times g)^{-1}[S^+]^*$ und somit ist T feiner als $(g \times g)^{-1}[S^+]$. \square

Auch hier können wir auf die transitive Hülle im Allgemeinen nicht verzichten, denn eine relationale Algebra, mit der die Initialtopologie berechnet werden kann, kann auch die transitive Hülle einer Relation berechnen. Der Beweis funktioniert ähnlich wie in allen anderen Fällen: Wir unterstellen, eine Abfrage Q könne für jeden topologischen Datentyp (Y, S) und jede (als Relation in einer Datenbank definierte) Abbildung $f : X \rightarrow Y$ eine initiale Relation I_f für X bezüglich f berechnen. Dann konstruieren wir daraus eine Abfrage $A(Q)$ zur Berechnung der transitiven Hülle einer gegebenen Relation R :

Beweis. Sei $R \subseteq X \times X$ eine Relation, deren transitive Hülle zu berechnen ist, und (X, R) der entsprechende topologische Datentyp. (X könnte etwa auch durch die Abfrage $\pi_1(R) \cup \pi_2(R)$ gegeben sein.)

Die Konstruktion der Abfrage von R^+ erfolgt hier ähnlich wie bei der gemeinsamen Verfeinerung ausgehend von folgender Beobachtung:

Sei $\{a, b\} \subseteq X$ beliebig und $i : \{a, b\} \rightarrow X$ die Inklusionsabbildung. Wir interessieren uns nur für den Fall $a \neq b$. Wenn eine Abfrage eine initiale Relation I bezüglich i erzeugt, dann gilt $a I b$ genau dann, wenn $a R^+ b$ gilt.

Wir erzeugen nun wieder die Vereinigung von $X \times X$ disjunkten Kopien von (X, R) und die Vereinigung S von X^2 mit (a, b) markierten disjunkten Kopien von $\{a, b\}$, berechnen dann die initiale Relation I für die Inklusion mit einer hypothetischen Abfrage und zeigen, dass für eine Projektion $\pi(I)$ gilt: $\pi(I) \setminus \Delta_X \cup R = R^+$

Die X^2 disjunkten Kopien von (X, R) sind $(X^3, (\Delta_{X^2} \otimes R))$ und die Teilmenge davon ist $S = \{(a, b, x) \in X \times X \times X \mid x = a \text{ oder } x = b\} \subseteq X^3$.

Ist nun I initial bezüglich der Inklusion $i : S \rightarrow (X^3, \Delta_{X^2} \otimes R)$, $(a, b, x) \mapsto (a, b, x)$, dann ist die Projektion

$$\pi(I) := \{(x, y) \mid (a, b, x) I (c, d, y)\}$$

die gesuchte Relation. Der Beweis $R^+ \setminus \Delta_X \subseteq \pi(I) \subseteq R^*$ ist ganz analog zum entsprechenden Beweis zur gemeinsamen Verfeinerung und wird hier nicht weiter ausgeführt. \square

Wir wollen nur noch zeigen, dass beim Sonderfall von surjektiven Abbildungen keine transitive Hülle berechnet werden muss. Anschaulich ist klar, dass es dann im Urbild der Relation keine „Lücken“ gibt, die die initiale Relation „überbrücken“ muss:

Beweis. Ist $g : X \rightarrow (Y, R)$ eine surjektive Abbildung, dann ist $(g \times g)^{-1}[R]^* = (g \times g)^{-1}[R^+]$. Die eine Richtung der Inklusion ist klar, denn wegen $R \subseteq R^+$ ist auch $(g \times g)^{-1}[R] \subseteq (g \times g)^{-1}[R^+]$ und damit gilt auch für deren Hüllen $(g \times g)^{-1}[R]^* \subseteq (g \times g)^{-1}[R^+]$.

Ist nun $(a, b) \in (g \times g)^{-1}[R^+]^*$, dann gilt $(a, b) \in (g \times g)^{-1}[R^+]^n$ für eine natürliche Zahl n . Dann zeigt man (wieder) durch Induktion über n , dass $(a, b) \in (g \times g)^{-1}[R]^*$ gilt:

Ist $n = 0$, dann ist $a = b$ und der Rest klar.

Ist $n = 1$, dann ist $(g(a), g(b)) \in R^+$. Hier werden wir später ebenfalls durch Induktion sehen, dass dann $(a, b) \in (g \times g)^{-1}[R]^+$ gilt.

Folgt für ein gewähltes n aus $(a, b) \in (g \times g)^{-1}[R^+]^n$ stets $(a, b) \in (g \times g)^{-1}[R]^+$ und sei $(a, b) \in (g \times g)^{-1}[R^+]^{n+1}$, dann ist $(a, j) \in (g \times g)^{-1}[R^+]^n$ und $(j, b) \in (g \times g)^{-1}[R^+]$ (für ein $j \in X$). Dann ist (a, j) laut Induktionsannahme, und (j, b) ist laut bisher unbewiesener Behauptung für den Fall $n = 1$ auch Element von $(g \times g)^{-1}[R]^+$. Diese Relation ist transitiv, also ist (a, b) ebenfalls in ihr enthalten.

Dieser Induktionsbeweis ist noch unvollständig, denn die Induktionsbehauptung für den Fall $n = 1$ wurde ja ohne Beweis verwendet:

Ist $(g(a), g(b)) \in R^+$ dann existiert eine Zahl i , so dass $(g(a), g(b)) \in R^{i+1}$.

Ist i gleich 0, dann ist $(g(a), g(b)) \in R$. und somit $(a, b) \in (g \times g)^{-1}(R)$.

Folgt jedoch aus $(g(a), g(b)) \in R^{i+1}$ stets $(a, b) \in (g \times g)^{-1}[R]^+$ und ist $(g(a), g(b)) \in R^{(i+1)+1}$, dann existiert ein j in Y mit $(g(a), j) \in R^{i+1}$ und $(j, g(b)) \in R$. Da g surjektiv ist, existiert ein $c \in X$ mit $g(c) = j$. Damit ist $(g(a), g(c)) \in R^{i+1}$ und $(g(c), g(b)) \in R$ und folglich auch $(a, c) \in (g \times g)^{-1}[R]^+$ laut Induktionsannahme und $(c, b) \in (g \times g)^{-1}[R]$. Also liegen beide Paare in $(g \times g)^{-1}[R]^+$ und da diese Relation transitiv ist enthält sie auch (a, b) . \square

Wir haben nun alle Grundlagen zusammen, um aus den Abfragen einer relationalen Algebra mit transitiver Hülle topologische Konstruktionen vorzunehmen. Dazu wollen wir zunächst die Abfragen der relationalen Algebra selbst zu topologischen Konstruktionen ausbauen.

4.4.5 Topologische Konstruktionen in \mathcal{DTop}

Das Prinzip, mengentheoretische Operationen mit den Punktmengen topologischer Räume durchzuführen und dann für das Ergebnis eine Topologie auszuwählen, wollen wir nun mit den Operatoren der relationalen Algebra durchführen. Diese sind zunächst die einstelligen Abfragen wie Selektion, Projektion und Umbenennung. Danach wollen wir entsprechend mit den mehrstelligen Operatoren wie Vereinigung, Schnitt, Rest, kartesisches Produkt und relationaler Verbund in entsprechender Weise verfahren. Es sind zudem einige besondere Konstruktionen in der Topologie sehr verbreitet. Wir werden einige davon mit den bis dahin entwickelten Abfrageoperatoren für topologische Datenbanken realisieren.

4.4.6 Einstellige Abfragen

Wir wollen also zunächst die Topologien für die Abfrageergebnisse von den einstelligen Datenbankabfragen Selektion, Projektion und Umbenennung bestimmen. Als Eingabe erhält eine solche Operation ϕ einen topologischen Datentyp (X, R) und erzeugt damit einen topologischen Datentyp $\underline{\phi}(X, R) = (\phi(X), R_\phi)$, genauso, wie in der relationalen Algebra aus einer gegebenen Relation R eine neue Relation $\phi(R)$ entsteht. Um zu zeigen, dass wir die topologische Variante des Abfrageoperators meinen, werden wir diesen unterstreichen, analog zu der

Konvention, einen topologischen Raum (X, \mathcal{T}_X) kurz mit \underline{X} zu bezeichnen und so von seiner Punktmenge X zu unterscheiden.

Selektion und Spurtopologie

Wir suchen zunächst eine Konstruktion, die der Selektion in der relationalen Algebra entspricht. Das Ergebnis der Selektion ist eine Teilmenge der Ausgangsrelation, der wir naheliegender Weise die Spurtopologie zuordnen. Damit erhalten wir einen topologischen Teilraum. Wir wissen bereits, dass für die Bestimmung des Teilraums im allgemeinen auf die Berechnung der transitiven Hülle nicht verzichtet werden kann.

Definition 4.20 (Selektion in $\mathcal{D}Top$). Sei (X, R) ein topologischer Datentyp, Θ ein Prädikat für die Datensätze in X und $\sigma_\Theta(X) := \{t \in X \mid t \text{ erfüllt } \Theta\}$. Dann ist

$$\underline{\sigma}_\Theta(X, R) := (\sigma_\Theta(X), (\sigma_\Theta(X) \times \sigma_\Theta(X)) \cap R^+)$$

die topologische Θ -Selektion von (X, R) .

Ist nun X eine Tabelle einer relationalen Datenbank, dann wird man natürlich kaum die Relation R in der Form $R \subseteq X \times X$ als Menge von Paaren (x_1, x_2) von Datensätzen x_1 und x_2 aus X modellieren. Die beiden Komponenten wären dann nämlich nicht atomar und R wäre nicht in erster Normalform. Hat X nun auch noch Nichtschlüsselattribute, was in der Regel der Fall sein wird, dann verstößt R zudem noch gegen die zweite Normalform. In der Praxis wird R also nur Schlüsselattributwerte aufweisen, die dann als zwei Fremdschlüssel jeweils einen³ Schlüssel in X referenzieren. Dies ist jedoch keine echte Einschränkung, da wir dann einfach das Ergebnis des Ausdrucks $(\sigma_\Theta(X) \times \sigma_\Theta(X))$ entsprechend durch Umbenennung und Projektion an das Schema von R anpassen müssen. Um die *Definition* zu vereinfachen, werden wir stets in der *formalen Definition* der Abfragen auf die Einhaltung der Normalformen verzichten. Bei guter Datenbankentwurfs-*Praxis* werden diese jedoch stets eingehalten. Daher wollen wir jeweils an einem Beispiel sehen, wie diese Abfragen bei normalisierten Datenbanken aussehen. Wir wollen dies nun am Beispiel nachholen:

Beispiel 4.8 (topologische Theta-Selektion). Sei das Schema

```
OuterSpace[id:integer, dim:integer, in:integer, out:integer, ...]
```

Relationenschema einer Tabelle OuterSpace und

```
T[a:integer, dimA:integer, b:integer, dimB:integer]
```

Schema einer Relation T mit Fremdschlüssel

```
(a, dimA) ↦ (id, dim) und (b, dimB) ↦ (id, dim).
```

Dann ist $(OuterSpace, T)$ ein topologischer Datentyp. Nehmen wir an, diese enthalte eine Familie $(plan_i)_{i=1, \dots, 9}$ von Plänen für einen Außenraum (inklusive topologischer Eigenschaften von Planänderungen). Für jedes Element ist das Attribut *in* die Planversion, zu der das Element eingeführt wurde, und *out* die Planversion, zu der es aufgegeben wurde. Sei $\Theta = in \leq 9 \leq out$. Das heißt,

³Meist wird dies der Primärschlüssel sein.

ein Planelement p war genau dann für Planversion 9 gültig, wenn $\Theta(p)$ also $p.in \leq 9 \leq p.out$ gilt. Dann ist $\text{Plan9} = \underline{\sigma}_{\Theta}(\text{OuterSpace}, T)$ der topologische Datentyp des Plans Nummer 9 für den Außenraum:

In SQL ist der Ausdruck

```
create view Plan9 as
select *
from OuterSpace
where in <= 9 and 9 <= out;
```

die Selektion Plan9 der Elemente des Planes Nummer 9. Folgender Ausdruck berechnet dessen Spurtopologie $T9$ unter Verwendung der `connect by` Klausel von Oracle-SQL:

```
create view T9 as
select distinct P9A.id as a, P9A.dim as dimA
, P9B.id as b, P9B.dim as dimB
from Plan9 P9A, Plan9 P9B
where exists( select *
from T
where T.b=P9B.id and T.dimB=P9B.dim
start with T.a=P9A.id and T.dimA=P9A.dim
connect by nocycle prior T.b=T.a
and prior T.dimB=T.dimA );
```

Dann ist $(\text{Plan9}, T9)$ der topologische Datentyp des Plans Nummer 9.

Wir werden auch weiterhin so verfahren, dass wir auf die Regeln guter Entwurfspraxis, also auf die Normalformen, bei der formalen Definition der Abfrageoperatoren verzichten. Damit vereinfacht sich die Definition, denn wir vermeiden, Fremdschlüssel dereferenzieren zu müssen. In den darauf folgenden Beispielen werden wir dies korrigieren und zeigen, wie Abfragen von topologischen Datenbanken bei üblicher Entwurfspraxis aussehen könnten.

Projektion und Quotiententopologie

Dem Ergebnis einer Projektion geben wir die Finaltopologie der Projektionsabbildung. Zur Vereinfachung der formalen Definition verzichten wir zunächst wieder auf die Normalformen.

Definition 4.21 (Projektion in \mathcal{DTop}). Sei (X, R) ein topologischer Datentyp und \mathcal{A} die Menge der Attributsymbole von X und $\mathcal{B} \subseteq \mathcal{A}$ Teilmenge der Attribute von X . Ohne Einschränkung verzichten wir auf Einhaltung der ersten und zweiten Normalform (NF1) und (NF2) für R und nehmen an, dass jeder Datensatz in R jeweils ein Paar mit zwei vollständigen Kopien von Datensätzen in X ist. Dann ist

$$\underline{\pi}_{\mathcal{B}}(X, R) := (\pi_{\mathcal{B}}(X), (\pi_{\mathcal{B}} \times \pi_{\mathcal{B}})(R))$$

die Projektion von (X, R) auf \mathcal{B} . Durch \mathcal{B} ist eine Äquivalenzrelation \sim auf X definiert: Zwei Datensätze t und s in X sind genau dann äquivalent, wenn alle

Attributwerte in \mathcal{B} für t und s gleich sind. Es gilt also $t \sim s$ genau dann, wenn $\pi_{\mathcal{B}}(t) = \pi_{\mathcal{B}}(s)$ gilt. Dann ist

$$(X, R)/\mathcal{B} := (X/\mathcal{B}, (\pi_{\mathcal{B}} \times \pi_{\mathcal{B}})(R))$$

der Quotientenraum von (X, R) bezüglich \mathcal{B} . Wobei wir daran erinnern, dass wir oben $X/\mathcal{B} := X/\sim$ definiert hatten.

Der Quotientenraum ist die topologische Variante des group by-Operators in SQL und die Relation $(\pi_{\mathcal{B}} \times \pi_{\mathcal{B}})(R)$ erzeugt sowohl die Finaltopologie der Projektion $\pi_{\mathcal{B}}$ als auch die Finaltopologie der natürlichen Projektion $\pi_{\mathcal{B}} : X \rightarrow X/\mathcal{B}$. Die Relation des Quotientenraums enthält nicht, wie bei den anderen Definitionen, Paare von ganzen Tupeln sondern — fast im Sinne der zweiten Normalform — nur Paare von Superschlüsselwerten, denn \mathcal{B} ist Superschlüssel von X/\mathcal{B} . Allerdings verstoßen Paare von Tupeln dann immer noch gegen die erste Normalform, was durch entsprechende Umbenennungen jedoch leicht behoben ist.

Beispiel 4.9 (Projektion). Angenommen wir haben für ein Projekt einen topologischen Datentyp (Elemente, Begrenzung) mit Relationenschemata

```
Elemente[id,detail:Details.id,...]
```

für die Elemente und

```
Begrenzung[a:Elemente.id,b:Elemente.id]
```

für deren Topologie. Das detail-Attribut referenziert die jeweils verwendete Detail aus einer Bauteilbibliothek. Dann ist $\pi_{\text{detail}}(\text{Elemente})$ die Menge der im Projekt verwendeten Details und die Abfrage

$$\pi_{A.\text{detail}, B.\text{detail}}(\beta_A(\text{Elemente}) \bowtie_{A.\text{id}=a} \text{Begrenzung} \bowtie_{B.\text{id}=b} \beta_B(\text{Elemente}))$$

erzeugt die entsprechende Relation dazu. Diese gibt an, in welcher Weise die verwendeten Details im Projekt untereinander verbunden sind. In SQL lauten die Abfragen:

```
create view DetailVerwendung as
select    detail
from      Elemente;
```

für die verwendeten Details und

```
create view DetailVerbindungen as
select    distinct A.detail as a, B.detail as b
from      Elemente as A, Begrenzung, Elemente as B
where     A.id = Begrenzung.a and B.id = Begrenzung.b;
```

für die Relation. Es ist also

$$\pi_{\text{detail}}(\text{Elemente}, \text{Begrenzung}) \cong (\text{DetailVerwendung}, \text{DetailVerbindung}).$$

Umbenennung als Homöomorphismus

Die Umbenennung erfordert eigentlich fast keine eigenständige Definition, da sie so einfach ist. Die charakteristische Abbildung der Umbenennung $\beta_f(X)$ einer Relation X ist die Umbenennung der einzelnen Datensätze $\beta_f(t) : X \rightarrow \beta_f[X]$, $t \mapsto t \circ f$ und erhält folglich die Finaltopologie bezüglich dieser Abbildung. Da diese Abbildung eine Bijektion ist, erhalten wir einen Homöomorphismus. Wir erinnern uns, dass Isomorphismen von Relationen ebenfalls nichts anderes als Umbenennungen sind.

Der Vollständigkeit halber wollen wir diese ebenfalls formal definieren:

Definition 4.22 (Umbenennung \mathcal{DTop}). Sei (X, R) topologischer Datentyp mit Relationenschema \mathcal{X} und sei \mathcal{A} eine Attributmenge und $f : \mathcal{A} \rightarrow \mathcal{X}$ eine surjektive Abbildung. Ohne Einschränkung verzichten wir auf Einhaltung der ersten und zweiten Normalform (NF1) und (NF2) für R und nehmen an, dass jeder Datensatz in R jeweils ein Paar mit zwei vollständigen Kopien von Datensätzen in X ist. Dann ist

$$\underline{\beta}_f(X, R) := (\beta_f(X), (\beta_f \times \beta_f)(R))$$

die Umbenennung der Attribute von (X, R) durch f .

In der Praxis wird eine topologische Datenbank (X, R) wieder normalisiert sein. Dann ist jedoch jede Referenz eines Datensatzes in R zu einem Datensatz in X auch eine Referenz zu $\beta_f(X)$ und deshalb kann R auch unverändert als Relation für $\beta_f(X)$ übernommen werden: $\underline{\beta}_f(X, R) \cong (\beta_f(X), R)$.

4.4.7 Verbinden von Räumen

Wenden wir uns nun den Abfragen zu, die mehrere Relationen kombinieren. In Lehrbüchern zur Topologie finden sich zudem weitere Beispiele für topologische Konstruktionen zum Verbinden von Räumen. Es werden dort Räume z.B. disjunkt vereinigt (topologische Summe), verklebt und es gibt auch eine Konstruktion, die als „topologischer Verbund“ bezeichnet wird. Bis auf den Produktraum und gewisse Sonderfälle der Vereinigung finden sich dort jedoch kaum Konstruktionen, die den mehrstelligen Operatoren der relationalen Algebra entsprechen. Wir werden hier nun kartesisches Produkt, Verbund, Schnitt, Vereinigung und Rest für topologische Datenbanken definieren. Danach werden wir einen Blick auf das Standardrepertoire aus den Topologie-Lehrbüchern werfen.

Kartesisches Produkt und Produktraum

Dem kartesischen Produkt von Relationen entspricht natürlich der topologische Produktraum. Beim kartesischen Produkt $A_1 \times \cdots \times A_n$ haben wir als typische Abbildungen jeweils die Projektionen

$$\pi_i : A_1 \times \cdots \times A_i \times \cdots \times A_n \rightarrow A_i, (a_1, \dots, a_i, \dots, a_n) \mapsto a_i \quad i = 1 \dots n.$$

Sind die A_i topologische Räume mit Topologien \mathcal{T}_{A_i} , dann ist die Topologie \mathcal{T}_π für $A_1 \times \cdots \times A_n$ die Initialtopologie bezüglich $\{\pi_1, \dots, \pi_n\}$. Dies ist die gröbste Topologie, in der alle „offenen Kästchen“ $U_1 \times \cdots \times U_n$ mit $(U_1, \dots, U_n) \in \mathcal{T}_{A_1} \times \cdots \times \mathcal{T}_{A_n}$ offen sind.

Beim kartesischen Produkt sind wir in der angenehmen Lage, dass hier eine Initialtopologie vorliegt, die auch ohne transitive Hülle berechnet werden kann. Für jede der einzelnen Projektionen ist dies klar, da Projektion eine surjektive Abbildung ist. Es gilt aber auch für die Initialtopologie insgesamt:

Satz 4.17 (Produkttopologie in \mathcal{DTop}). Sei $\{(X_1, R_1), \dots, (X_n, R_n)\}$ eine topologische Datenbank. Sei $\Delta_i(R_i)$ die „Einbettung“ von R_i in die Diagonale des kartesischen Produkts $X_1 \times \dots \times X_n$, definiert durch

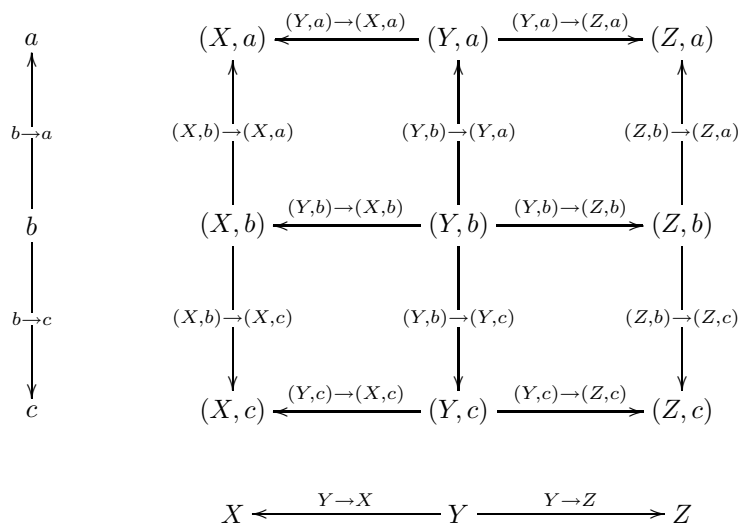
$$\Delta_i(R_i) := \Delta_{X_1} \otimes \dots \otimes \Delta_{X_{i-1}} \otimes R_i \otimes \Delta_{X_{i+1}} \otimes \dots \otimes \Delta_{X_n}.$$

Dann ist

$$(X_1, R_1) \underline{\times} \dots \underline{\times} (X_n, R_n) := (X_1 \times \dots \times X_n, \Delta_1(R_1) \cup \dots \cup \Delta_n(R_n))$$

der topologische Datentyp des Produktraums. Wir wollen die Relation dieser Datenbank auch als Produktrelation bezeichnen.

Wir erinnern uns, dass wir die Standardoperatoren der relationalen Algebra unterstreichen, wenn wir die daraus abgeleiteten topologischen Konstruktionen meinen. Hier bezeichnet also der Operator „ \times “ das kartesische Produkt von Mengen oder Relationen, der Operator „ $\underline{\times}$ “ hingegen die entsprechende Konstruktion des Produktraums. Von dieser ist noch zu zeigen, dass sie in der Tat durch die obige Definition der Produktrelation erzeugt wird. Bevor wir dies tun, wollen wir die Konstruktion der Produktrelation an einem Beispiel mit zwei Relationen illustrieren:



Dabei handelt es sich um zwei topologische Datentypen $(\{a, b, c\}, R)$ und $(\{X, Y, Z\}, S)$ und deren Produktraum $(\{a, b, c\}, R) \underline{\times} (\{X, Y, Z\}, S)$. Wenn wir die beiden Datentypen als Graphen betrachten, dann ist dies offenbar nichts anderes als der sogenannte Produktgraph [Har74, S. 31]. Bei den waagrechten Pfeilen im Produkt bleibt die zweite Komponente gleich und bei den senkrechten ist es die erste Komponente, wie z.B. bei $(X, b) \rightarrow (X, a)$. Es gilt $X = X$ beziehungsweise $(X, X) \in \Delta_{\{X, Y, Z\}}$ und zudem $a R b$ und folglich

$$((X, b), (X, a)) \in \Delta_{\{X, Y, Z\}} \otimes R.$$

Die Produktrelation ist nun die Vereinigung der Menge der senkrechten Kanten $\Delta_{\{X,Y,Z\}} \otimes R$ mit den waagrechten Kanten $S \otimes \Delta_{\{a,b,c\}}$.

Wir wollen nun sehen, dass unsere Definition in der Tat den Produktraum der beteiligten Räume erzeugt:

Beweis. Es ist zu zeigen, dass die Produktrelation $\Delta_1(R_1) \cup \dots \cup \Delta_n(R_n)$ die Initialtopologie der Projektionsabbildungen $\pi_j : X_1 \times \dots \times X_n \rightarrow X_j$ erzeugt. Vereinbaren wir zunächst Kurzschreibweisen und schreiben einfach (a_i) statt (a_1, \dots, a_n) und bezeichnen die Produktrelation mit Π .

Die π_j sind in der Tat stetig, denn ist ein $(x, y) := ((a_i), (b_i)) \in \Delta_1(R_1) \cup \dots \cup \Delta_n(R_n)$, dann ist $(x, y) \in \Delta_k(R_k)$ für ein $k \in \{1, \dots, n\}$. Dieses (x, y) hat die Form

$$((a_1, \dots, a_{k-1}, a_k, a_{k+1}, \dots, a_n), (a_1, \dots, a_{k-1}, b_k, a_{k+1}, \dots, a_n))$$

mit $(a_k, b_k) \in R_k$. Dann gilt entweder $j = k$ und $(\pi_j(x), \pi_j(y)) = (a_k, b_k) \in R_k = R_j$ und somit ist auch $(\pi_j \times \pi_j)(x, y)$ Element von R_j^* , oder es ist $j \neq k$, dann ist $(\pi_j(x), \pi_j(y)) = (a_j, a_j) \in \Delta_j = R_j^0$ und somit ebenfalls in R_k^* .

Wir müssen nun noch zeigen, dass jede andere Relation R auf $X_1 \times \dots \times X_n$, mit der die Projektionsabbildungen stetig sind, auch feiner ist als die angegebene Produktrelation. Ist also mit R jede Projektion stetig, dann gilt für jedes $(x, y) \in R$ und jede Projektion π_j , dass $(\pi_j(x), \pi_j(y)) \in R_j^*$ liegt. Dann müssen wir zeigen, dass $(x, y) \in \Pi^*$ ist.

Sei $(x, y) = ((a_i), (b_i)) \in R$. Wir können wieder voraussetzen, dass $x \neq y$ gilt, da sonst $(x, y) \in \Pi^*$ trivialerweise gilt. Ist aber $x \neq y$, dann existiert ein maximales j so dass $a_j \neq b_j$ ist. Durch Induktion über j zeigen wir dann, dass $(x, y) \in \Pi^*$ gilt.

Ist $j = 1$, dann ist $(x, y) = ((a_1, a_2, \dots, a_n), (b_1, a_2, \dots, a_n))$. Also nur die beiden ersten Komponenten a_1 und b_1 unterscheiden sich. Dann gilt auch $(a_1, b_1) \in R_1^*$ und damit $(x, y) \in \Delta_1(R_1^*)$. Wir behaupten nun, dass für alle $i = 1, \dots, n$ gilt: $\Delta_i(R_i^*) = (\Delta_i(R_i))^*$. Dann gilt wegen $\Delta_i(R_i) \subseteq \Pi$ auch $(x, y) \in \Delta_i(R_i^*) = (\Delta_i(R_i))^* \subseteq \Pi^*$.

Gelte nun die Induktionsbehauptung für eine natürliche Zahl j , sei aber $j+1$ die größte Zahl, so dass $a_{j+1} \neq b_{j+1}$ gilt. Seien also alle Tupel-Komponenten a_k und b_k rechts von $j+1$ (mit $k > j+1$) gleich.

Dann gilt auch, wenn wir auf der rechten Seite b_{j+1} durch a_{j+1} ersetzen

$$((a_1, \dots, a_j, a_{j+1}, \dots, a_n), (b_1, \dots, b_j, a_{j+1}, \dots, a_n)) \in \Pi^*,$$

denn für jede Projektion π_i galt vorher $(a_i, b_i) \in R_i^*$; nach Ersetzen von b_{j+1} durch a_{j+1} gilt immer noch

$$(\pi_{j+1} \times \pi_{j+1})((a_1, \dots, a_j, a_{j+1}, \dots, a_n), (b_1, \dots, b_j, a_{j+1}, \dots, a_n)) = (a_{j+1}, a_{j+1})$$

und das ist Element von R_{j+1}^* . Zudem gilt für das Paar

$$((b_1, \dots, b_j, a_{j+1}, \dots, a_n), (b_1, \dots, b_j, b_{j+1}, \dots, a_n)) \in \Pi^*,$$

denn es ist in $\Delta_{j+1}(R_{j+1}^*)$, da ja gilt $(a_{j+1}, b_{j+1}) \in R_{j+1}^*$. Damit ist dieses Paar auch in Π^* .

Da Π^* transitiv ist, ist ebenfalls

$$((a_1, \dots, a_j, a_{j+1}, a_{n+2}, \dots, a_n), (b_1, \dots, b_j, b_{j+1}, a_{n+2}, \dots, a_n)) \in \Pi^*.$$

Also gilt $(x, y) \in \Pi^*$ und folglich $R \subseteq \Pi^*$ und somit ist R feiner als Π . Damit ist Π eine größte Relation, so dass alle Projektionsabbildungen stetig sind und folglich initial.

Zeigen wir noch die oben verwendete Behauptung $\Delta_i(R_i)^* = \Delta_i(R_i^*)$: Wegen $R_i \subseteq R_i^*$ ist auch $\Delta_i(R_i) \subseteq \Delta_i(R_i^*)$, wie man anhand der Definition von Δ_i leicht einsieht und damit auch $\Delta_i(R_i)^* \subseteq \Delta_i(R_i^*)^*$. Zudem ist $\Delta_i(R_i^*)$ transitiv und reflexiv, denn es ist das direkte Produkt von transitiven und reflexiven Relationen und damit ist $\Delta_i(R_i^*)^* = \Delta_i(R_i^*)$, und damit gilt $\Delta_i(R_i)^* \subseteq \Delta_i(R_i^*)$.

Sei nun $((a_1, \dots, a_i, \dots, a_n), (a_1, \dots, b_i, \dots, a_n)) \in \Delta_i(R_i^*)$, dann ist auch $(a_i, b_i) \in R_i^*$ und somit $(a_i, b_i) \in R_i^m$ für eine natürliche Zahl m . Durch Induktion über m zeigen wir, dass dann $((a_1, \dots, a_i, \dots, a_n), (a_1, \dots, b_i, \dots, a_n)) \in \Delta_i(R_i)^m$ gilt. Der Induktionsanfang $m = 0$ ist bereits erledigt, da $\Delta_i(R_i^*)$ reflexiv ist. Gelte also die Induktionsbehauptung für eine beliebige natürliche Zahl m und sei $(a_i, b_i) \in R_i^{m+1}$, dann existiert ein c_i mit $(a_i, c_i) \in R_i^m$ und $(c_i, b_i) \in R_i$. Wegen der Induktionsbehauptung gilt

$$((a_1, \dots, a_i, \dots, a_n), (a_1, \dots, c_i, \dots, a_n)) \in \Delta_i(R_i)^m$$

und wegen $(c_i, b_i) \in R_i$ gilt

$$((a_1, \dots, c_i, \dots, a_n), (a_1, \dots, b_i, \dots, a_n)) \in \Delta_i(R_i)$$

und somit

$$((a_1, \dots, a_i, \dots, a_n), (a_1, \dots, b_i, \dots, a_n)) \in \Delta_i(R_i)^{m+1}.$$

Da $\Delta_i(R_i)^{m+1} \subseteq \Delta_i(R_i)^*$ gilt, haben wir $\Delta_i(R_i^*) \subseteq \Delta_i(R_i)^*$. Insgesamt gilt also $\Delta_i(R_i^*) = \Delta_i(R_i)^*$. \square

Wir wollen nun noch ein Beispiel zur Erzeugung des Produkts von topologischen Datentypen mit SQL zeigen:

Seien $X[\text{id}, \dots]$ und $Y[\text{nr}, \dots]$ Relation einer Datenbank mit id und nr als Primärschlüsselattributen und gegebenenfalls weitere Attributen. Seien zudem $R[\text{a}:X.\text{id}, \text{b}:X.\text{id}]$ und $S[\text{p}:Y.\text{nr}, \text{q}:Y.\text{nr}]$ Relationen in X und Y , dann sind (X, R) und (Y, S) topologische Datentypen.

Mit dem kartesischen Produkt

```
create view XtimesY as
select * from X, Y;
```

und der Produktrelation

```
create view RprodS as
select a, b, nr as p, nr as q
from R, Y
union
select id as a, id as b, p, q
from X, S;
```

ist $(X \times Y, R \times S)$ der topologische Datentyp des Produktraums $(X, \mathcal{T}_R) \times (Y, \mathcal{T}_S)$.

Das kartesische Produkt ist auch deshalb eine wichtige Abfrage, da sich mit ihm viele weitere Verbundoperationen definieren lassen.

Verbundoperatoren

Mit den obigen Definitionen von Produktraum und mit der topologischen Selektion haben wir sofort eine topologische Entsprechung für den Theta-Verbund $X \bowtie_{\Theta} Y$ zweier Relationen X und Y anhand eines Prädikats Θ , denn dieser ist gleich dem kartesischen Produkt mit darauf folgender Selektion. Damit haben wir für zwei topologische Datentypen (X, R) und (Y, S) eine naheliegende Definition für den Theta-Verbund

$$(X, R) \bowtie_{\Theta} (Y, S) := \sigma_{\Theta}((X, R) \times (Y, S))$$

und für den natürlichen Verbund

$$(X, R) \bowtie (Y, S) := (X \bowtie Y, RS).$$

Dabei ist RS die initiale Relation der Projektionen $\pi_X : X \bowtie Y \rightarrow X$ und $\pi_Y : X \bowtie Y \rightarrow Y$, von der wir ja bereits wissen, wie sie zu berechnen ist.

Der Gleichheitsverbund ist bei der relationalen Modellierung besonders wichtig. Auch hier müssen wir festhalten, dass eine relationale Algebra, die im Allgemeinen einen (Theta-, Gleichheits- oder natürlichen) Verbund zweier topologischer Datentypen erzeugen kann, eine relationale Algebra mit transitiver Hülle ist. Der Beweis ist einfach: Haben wir eine Relation X , dann ist deren natürlicher Verbund mit sich selbst $X \bowtie X = X$. Sind nun R und S Relationen auf X , dann ist $(X, R) \bowtie (X, S) = (X, \sup\{R, S\})$, denn eine Relation V von $(X, V) = (X, R) \bowtie (X, S)$ ist initial bezüglich der Abbildungen $\text{id}_R : X \rightarrow (X, R)$ und $\text{id}_S : X \rightarrow (X, S)$. Also ist V die gemeinsame Verfeinerung zweier initialer Relationen bezüglich id_R und id_S . Es sind aber R und S selbst jeweils initial bezüglich id_R und id_S und damit erzeugt der natürliche Verbund der Räume (X, R) und (X, S) die gemeinsame Verfeinerung von R und S . Wie wir jedoch gesehen haben, ist dazu im Allgemeinen die Berechnung der transitiven Hülle erforderlich.

Wir werden z.B. bei der Modellierung von Detailbibliotheken häufig den Gleichheitsverbund benötigen, damit Bauelemente ihre detaillierte Ausführung referenzieren können.

Der topologische Verbund In der Topologie gibt es ebenfalls eine Konstruktion, die als „Verbund“ bezeichnet wird, jedoch etwas ganz anderes bedeutet. Auf ihn soll nur kurz eingegangen werden, zum Einen, um den Unterschied zum relationalen Verbund zu zeigen, zum Anderen aber auch, um endliche Alternativen zu Konstruktionen anzubieten, bei denen das reelle Einheitsintervall $\mathbb{I} = [0, 1] := \{x \in \mathbb{R} \mid 0 \leq x \leq 1\}$ beteiligt ist, wie Kegel, Einhängung, Abbildungszylinder und später auch Wege.

Der topologische Verbund $X * Y$ zweier Räume X und Y ist wie folgt definiert (siehe auch HATCHER [Hat02, S. 9]: Man erstellt zunächst den Produktraum $X \times \mathbb{I} \times Y$ und schlägt dann erst den Teilraum $X \times \{0\} \times Y$ an der 0-Seite wieder zu X zusammen und schließlich den Teilraum $X \times \{1\} \times Y$ auf der 1-Seite zu Y . Sind

z.B. X und Y selbst das Einheitsintervall \mathbb{I} , dann ist $X \times \mathbb{I} \times Y$ der Einheitswürfel und $X * Y$ ergibt ein Tetraeder. Ist P ein Raum mit nur einem Punkt und B ein beliebiger anderer Raum, dann heißt $P * B$ *Kegel auf B* . Wählen wir stattdessen für P einen diskreten Raum mit genau zwei Punkten, dann entsteht durch $P * B$ ein „Doppelkegel“, der als *Einhängung* von B bezeichnet wird. Abbildung 4.3 veranschaulicht diese Konstruktionen.

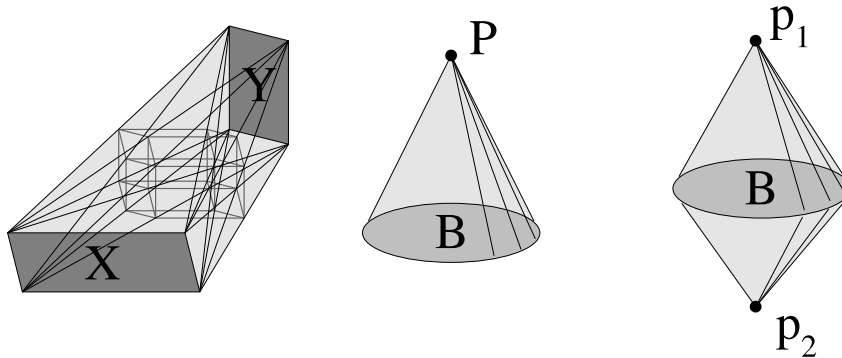


Abbildung 4.3: *Beispiele für topologischen Verbund. Links: Der topologische Verbund $X * Y$ zweier Räume X und Y . Der Schnitt an der Stelle 0.5, also der Teilraum $X \times \{0.5\} \times Y$ ist darüber gelegt. Mitte: Besteht P aus nur einem Punkt, dann heißt $B * P$ auch Kegel über B . Rechts: Der Doppelkegel (Einhängung oder Suspension), $B * \{p_1, p_2\}$, wobei mit $\{p_1, p_2\}$ der entsprechende diskrete Raum gemeint ist.*

Anstelle des reellen Einheitsintervalls können wir den topologischen Datentyp eines ungerichteten Graphen W (wie „Weg“) mit genau einer Kante u und Anfangs- und Endpunkt a und e verwenden: $W = (\{a, u, e\}, \{(u, a), (u, e)\})$. Dies entspricht dem Einheitsintervall $[0, 1] /]0, 1[$, zerlegt in drei Teilmengen: Den Startpunkt $a = \{0\}$, das offene Intervall $u =]0, 1[$, den Endpunkt $e = \{1\}$. Die Topologie, die von $\{(u, a), (u, e)\}$ erzeugt wird, ist dann die Quotiententopologie dieser Zerlegung. Bei endlichen Wegen müssen wir das Einheitsintervall in mehrere derartige Abschnitte zerlegen. Wir werden später darauf zurückkommen.

Nun können wir für zwei topologische Datentypen (X, R) und (Y, S) eine endliche Variante des topologischen Verbundes definieren, indem wir $I = (\{a, u, e\}, \{(u, a), (u, e)\})$ setzen und den Produktraum $(X, R) \times I \times (Y, S)$ mit

den beiden Projektionsabbildungen

$$\begin{aligned}\pi_X &: (X, R) \times_{I|_{\{a\}}} \times (Y, S) \rightarrow (X, R) \\ \pi_Y &: (X, R) \times_{I|_{\{e\}}} \times (Y, S) \rightarrow (Y, S)\end{aligned}$$

jeweils an (X, R) und (Y, S) „anheften“. Zum Anheften sind jedoch (X, R) und (Y, S) erst mit $(X, R) \times_{I} \times (Y, S)$ vereinigungsverträglich zu machen, indem wir je einen Punkt $x \in X$ und $y \in Y$ auswählen. Anstelle von (X, R) nehmen wir dann $(X, R) \times_{I|_{\{a\}}} \times ((Y, R)|_{\{y\}})$ und anstelle von Anstelle von (Y, S) setzen wir entsprechend $((X, R)|_{\{x\}}) \times_{I|_{\{a\}}} \times (Y, R)$.

Die Konstruktion des Produkts topologischer Datentypen kennen wir bereits, und die Abfrage zum Anheften werden wir später noch sehen. Die Formulierung der vollständigen Abfrage zur Konstruktion des topologischen Verbundes wollen wir uns hier aber ersparen.

Vereinigung, Schnitt und Restmenge

Auch für die Vereinigung $A \cup B$ und den Schnitt $A \cap B$ zweier topologischer Räume (A, \mathcal{T}_A) und (B, \mathcal{T}_B) existiert jeweils eine sinnvolle neue Topologie: Die Initialtopologie der Inklusionen

$$i_{\cap, A} : A \cap B \hookrightarrow (A, \mathcal{T}_A) \quad \text{und} \quad i_{\cap, B} : A \cap B \hookrightarrow (B, \mathcal{T}_B)$$

als Topologie für die Schnittmenge und die Finaltopologie der Inklusionen

$$i_{\cup, A} : (A, \mathcal{T}_A) \hookrightarrow A \cup B \quad \text{und} \quad i_{\cup, B} : (B, \mathcal{T}_B) \hookrightarrow A \cup B$$

als Topologie für die Vereinigung. Wir wollen die so entstehenden Räume im Sinne unserer Konvention mit $(A, \mathcal{T}_A) \sqcap (B, \mathcal{T}_B)$ und $(A, \mathcal{T}_A) \sqcup (B, \mathcal{T}_B)$ bezeichnen. Die Restmenge $A \setminus B$ erhält natürlich die Spurtopologie $\mathcal{T}_A|_{A \setminus B}$:

Definition 4.23 (Vereinigung, Schnitt und Rest in \mathcal{DTop}). *Seien (X, R) und (Y, S) vereinigungsverträgliche topologische Datentypen, dann ist*

$$(X, R) \sqcup (Y, S) := (X \cup Y, R \cup S)$$

die Vereinigung von (X, R) mit (Y, S) und

$$(X, R) \sqcap (Y, S) := (X \cap Y, R^+ \cap S^+)$$

der Schnitt von (X, R) mit (Y, S) und

$$(X, R) \setminus (Y, S) := (X \setminus Y, R^+ \cap (X \setminus Y)^2)$$

der Rest von (X, R) ohne (Y, S) .

Auch hier benötigen wir für Schnitt und Rest die transitive Hülle.

Ist $\mathcal{T}_A|_{A \cap B}$ feiner als $\mathcal{T}_B|_{A \cap B}$, dann heißt diese Vereinigung topologischer Räume auch *Anheften* von (A, \mathcal{T}_A) an (B, \mathcal{T}_B) . Der Teilraum $(A, \mathcal{T}_A)|_{A \cap B}$ ist dann gewissermaßen die „Klebelasche“ die beim Anheften vollständig im anderen Raum aufgeht. Die Vereinigung hier ist allerdings nur ein Sonderfall der Verklebung, auf die wir später noch eingehen werden, denn diese Konstruktion

ist sehr wichtig. Das zeigt sich nicht nur darin, dass sie in praktisch jedem Lehrbuch auftaucht — durch Verkleben ist es auch möglich, komplexere topologische Räume aus einfacheren Räumen aufzubauen und diese anhand der einfacheren Räume zu untersuchen. Wir werden später auch eine praktische Anwendung des Verklebens für die Versionskontrolle kennen lernen: Schneide eine veraltete Stelle aus dem Plan und überklebe sie mit der neuen Version (patch).

Die topologische Summe und Verklebung

Die topologische Summe ist ein Sonderfall der oben erwähnten Vereinigung von topologischen Räumen zu einem weiteren Raum mit der Finaltopologie der Inklusionen. Hier wird zusätzlich eingeschränkt, dass die Räume disjunkt sind.

Um nun zwei Räume allgemein aneinanderzuheften, erzeugt man zunächst die topologische Summe. Dann legt man eine Beziehung zwischen diesen Räumen fest, indem man für einen Teil eines Raumes (gewissermaßen dessen „Klebelasche“) angibt, wo dieser im anderen Raum liegen soll. Diese Beziehung ergibt unseren „Klebstoff“ und wird als Anheftungsabbildung bezeichnet. Der so entstandene Raum heißt *Verklebung* oder Anheftung.

Zunächst folgt also die disjunkte Summe:

Definition 4.24 (topologische Summe). *Sind (X, \mathcal{T}_X) und (Y, \mathcal{T}_Y) disjunkte topologische Räume, also $X \cap Y = \emptyset$, dann ist*

$$(X, \mathcal{T}_X) + (Y, \mathcal{T}_Y) := (X \cup Y, \{A \cup B \mid A \in \mathcal{T}_X, B \in \mathcal{T}_Y\})$$

die topologische Summe von (X, \mathcal{T}_X) und (Y, \mathcal{T}_Y) . Sind X und Y nicht disjunkt, dann erhalten wir disjunkte Räume X' und Y' , indem wir $X' = \{0\} \times X$ und $Y' = \{1\} \times Y$ setzen, die Topologien entsprechend zu $\mathcal{T}_{X'} = \{\{0\} \times A \mid A \in \mathcal{T}_X\}$ und $\mathcal{T}_{Y'} = \{\{1\} \times B \mid B \in \mathcal{T}_Y\}$ abändern und dann deren Summe definieren als: $(X, \mathcal{T}_X) + (Y, \mathcal{T}_Y) = (X', \mathcal{T}_{X'}) + (Y', \mathcal{T}_{Y'})$.

Die topologische Summe $\underline{X} + \underline{Y}$ trägt in der Tat die Finaltopologie bezüglich der Inklusionsabbildungen $i_X : \underline{X} \rightarrow \underline{X} + \underline{Y}$, $x \mapsto x$ und $i_Y : \underline{Y} \rightarrow \underline{X} + \underline{Y}$, $y \mapsto y$, wovon sich die Leserin leicht überzeugt.

Mit dem „Trick“, nicht disjunkte Räume disjunkt zu machen, können wir sogar die Summe eines topologischen Raums mit sich selbst erzeugen. Im Falle relationaler Datenbanken wollen wir nun ganz analog die disjunkte Vereinigung zweier Tabellen definieren:

Definition 4.25 (disjunkte Vereinigung von Relationen). *Sind X und Y vereinigungsverträgliche Relationen und \mathbf{ref} ein Attribut, das im Relationenschema von X und Y nicht vorkommt. Dann ist die disjunkte Summe*

$$X +_{\mathbf{ref}} Y := (\{\langle 0 : \mathbf{ref} \rangle\} \times X) \cup (\{\langle 1 : \mathbf{ref} \rangle\} \times Y).$$

Wir fügen also ein Attribut \mathbf{ref} an das Datenbankschema an, um jeden Datensatz entsprechend seiner Herkunft zu „markieren“. Wollen wir mehrere Attribute $\mathbf{ref}1$ und $\mathbf{ref}2$ verwenden, dann schreiben wir entsprechend $X +_{\mathbf{ref}1, \mathbf{ref}2} Y$. In SQL erhalten wir $X +_{\mathbf{ref}} Y$ durch die Abfrage:

```

create view XplusY as
select 0 as ref, *
from X
union all
select 1 as ref, *
from Y;

```

Dabei ist `union all` eine „beschleunigte“ Variante von `union`, die sich insbesondere zur Vereinigung von disjunkten Relationen anbietet. Diese disjunkte Vereinigung ist auch mit einer größeren Familie von Mengen möglich. Dann verwenden wir einfach eine größere Indexmenge. Im Beweis, dass die gemeinsame Verfeinerung zweier Relationen im Allgemeinen die Berechnung der transitiven Hülle voraussetzt, haben wir z.B. die topologische Summe von $X \times X$ Kopien des Raums (X, R) erzeugt und $X \times X$ dazu als Indexmenge verwendet. Hier die Definition der Summe topologischer Datentypen:

Definition 4.26 (Disjunkte Summe in $\mathcal{D}Top$). *Seien (X, R) und (Y, S) zwei vereinigungsverträgliche topologische Datentypen, und `ref` kein Attribut in X und in R , dann ist $(X, R) +_{\text{ref}} (Y, S) := (X +_{\text{ref}} Y, R +_{\text{ref}} S)$ die topologische Summe von (X, R) und (Y, S) .*

Wir werden also stets die Räume mit unserem Trick disjunkt machen, selbst wenn sie zufälligerweise bereits disjunkt sein sollten. Das zusätzliche Attribut `ref` muss natürlich bei der Summe der Relationen nur einmal angegeben werden. Ein Tripel (ref, a, b) fassen wir dann als Paar $((\text{ref}, a), (\text{ref}, b))$ auf.

Die disjunkte Summe hat auch eine wichtige praktische Bedeutung, da Planungsinformationssysteme sinnvollerweise mehrere Projekte in einer gemeinsamen Datenbank speichern. Diese können dann z.B. durch eine Projektnummer ausgewählt werden. Damit ist eine solche Projektdatenbank eine disjunkte Summe von Einzelprojekten der Form:

$$\sum_{\text{prjNr} \in \text{Projekt}} (X_{\text{prjNr}}, R_{\text{prjNr}}).$$

Mit der disjunkten Summe haben wir zunächst nur zwei Räume als einen zusammengefassten Raum betrachtet. Um jedoch die Räume aneinanderzuheften, müssen wir noch angeben, welche Punkte des einen Raums wo im anderen Raum liegen sollen: Wir benötigen also zu den Räumen X und Y noch eine „Klebelasche“ X_0 , die zu X gehört, also $X_0 \subseteq X$. Und wir müssen zudem angeben, an welcher Stelle in Y diese Lasche zu kleben ist. Dies geschieht durch eine Abbildung $\phi : X_0 \rightarrow Y$, die für jeden Punkt $p \in X_0$ auf unserer Klebelasche einen Punkt $\phi(p) \in Y$ angibt, auf den p zu kleben ist. Den verklebten Raum erhalten wir, indem wir sagen, der Punkt p soll mit dem Punkt $\phi(p)$ gleichgesetzt werden. Dadurch erhalten wir eine Äquivalenzrelation auf $X + Y$, und zwar ist dies die kleinste Äquivalenzrelation \sim in der $\phi(p) \sim p$ gilt. Wir betrachten ϕ einfach als Relation $R_\phi \subseteq (X + Y) \times (X + Y)$, definiert durch: $a R_\phi b$ gilt genau dann, wenn $a \in X_0$ und $\phi(a) = b$ gilt. Die von R_ϕ erzeugte Äquivalenzrelation $(R_\phi \cup R_\phi^T)^*$ bezeichnen wir als \sim_ϕ .

Zudem wird sinnvollerweise gefordert, dass ϕ die Punkte von X_0 nicht wild irgendwo nach Y hineinklebt, sondern stetig ist. Dann können wir X nämlich wieder von Y abtrennen, ohne dass Y „beschädigt“ wurde.

Definition 4.27 (Verklebung). Seien $\underline{X} = (X, \mathcal{T}_X)$ und $\underline{Y} = (Y, \mathcal{T}_Y)$ topologische Räume, $X_0 \subseteq X$ und $\phi : \underline{X}_0 \rightarrow \underline{Y}$ eine stetige Abbildung, wobei wir $\underline{X}_0 := \underline{X}|_{X_0}$ setzen. Dann ist

$$\underline{Y} \cup_{\phi} \underline{X} := (\underline{X} + \underline{Y}) / \sim_{\phi}$$

die Anheftung von \underline{X} an \underline{Y} mit Anheftungsabbildung ϕ . Dabei ist \sim_{ϕ} die von ϕ in $X + Y$ erzeugte Äquivalenzrelation.

Verkleben ist im Allgemeinen nicht kommutativ, und wir schreiben $\underline{Y} \cup_{\phi} \underline{X}$ mit dem Vorbereich von ϕ auf der rechten Seite des Verklebungsoperators \cup_{ϕ} und dem Wertebereich auf der linken. Wir sprechen hier vom „Vorbereich“, da ϕ auch als partielle Abbildung

$$\phi : \subseteq X \rightarrow Y$$

mit Definitionsbereich X_0 betrachtet werden kann.

Durch die Stetigkeit von ϕ wird sichergestellt, dass \underline{Y} Teilraum von $\underline{Y} \cup_{\phi} \underline{X}$ ist. Denn ist ϕ nicht stetig, dann ist die Spurtopologie von Y in $\underline{Y} \cup_{\phi} \underline{X}$ im allgemeinen echt gröber als \mathcal{T}_Y .

Offensichtlich ist unsere oben definierte mengentheoretische Vereinigung von zwei Räumen $\underline{X} \cup \underline{Y}$ mit der Finaltopologie der Inklusionen genau dann eine Anheftung $\underline{X} \cup_i \underline{Y}$, wenn die Inklusion $i : \underline{Y}|_{X \cap Y} \rightarrow \underline{X}$ stetig ist, also genau dann, wenn $\mathcal{T}_Y|_{X \cap Y}$ feiner ist als $\mathcal{T}_X|_{X \cap Y}$.

Die Leserin kann sich überlegen, dass die disjunkte Summe $X + Y$ zweier topologischer Räume X und Y bereits eine Verklebung $Y \cup_f X$ dieser Räume ist, indem sie sich überlegt, wie dann die entsprechende stetige Anheftungsabbildung $f : X_0 \rightarrow Y$ beschaffen ist. Insbesondere ist dazu X_0 geeignet zu wählen.

Verkleben in topologischen Datenbanken Datenbankabfragen zum Verkleben von topologischen Datentypen sind etwas umständlich, was daran liegt, dass für die nur implizit durch die Anheftungsabbildung definierte Äquivalenzrelation eine Abfrage formuliert werden muss. Wir wollen nun die Verklebung in topologischen Datenbanken mit SQL angeben. Dabei setzen wir voraus, dass \mathbf{X} und \mathbf{Y} vereinigungsverträgliche Relationen sind. Zudem sollen \mathbf{R} und \mathbf{S} ebenfalls vereinigungsverträglich sein und (\mathbf{X}, \mathbf{R}) und (\mathbf{Y}, \mathbf{S}) topologische Datentypen bilden. Zur Vereinfachung habe \mathbf{X} und \mathbf{Y} jeweils einen einstelligen Schlüssel $\{\text{id}\}$ und \mathbf{R} und \mathbf{S} die Attribute $\{\mathbf{a}, \mathbf{b}\}$. Wir benötigen dann noch eine Relation $\text{PHI}[\underline{x}:\mathbf{X}.\text{id}, \underline{fx}:\mathbf{Y}.\text{id}]$, die den Graph einer Anheftungsabbildung $\phi : X_0 \rightarrow Y$ darstellt.

Die Äquivalenzrelation \sim_{ϕ} wollen wir mit einer Fortsetzung

$$c\phi : (X + Y) \rightarrow (X + Y)$$

von $\phi : X_0 \rightarrow Y$ realisieren. Dabei soll jede Faser $c\phi^{-1}[\{c\phi(x)\}]$ (also jedes nichtleere Urbild eines Bildpunktes) genau der Äquivalenzklasse $[x]$ bezüglich \sim_{ϕ} entsprechen. Dies erreichen wir, indem wir ϕ wie folgt fortsetzen:

$$c\phi : (X + Y) \rightarrow (X + Y), p \mapsto \begin{cases} p & : p \notin X_0 \\ \phi(p) & : p \in X_0 \end{cases}$$

Dann ist $c\phi(t) = c\phi(s)$ genau dann, wenn $t \sim_\phi s$ gilt. Genau genommen unterscheidet $c\phi$ drei Fälle, denn wenn $p \notin X_0$ liegt, dann muss noch unterschieden werden, ob $p \in X$ oder $p \in Y$ gilt, wie wir gleich sehen werden. In SQL erhalten wir diese Fortsetzung durch folgende Abfrage CPHI, die den Graph für jede der drei Fallunterscheidungen mit einer eigenen `select`-Abfrage erzeugt und dann diese Fälle disjunkt vereinigt:

```
create view CPHI as
select 0 as ref, x, 1 as cPhiRef, fx as cPhix
from   PHI
union all
select 0 as ref, id as x, 0 as cPhiRef, ref as cPhix
from   X
where  not exists (select * from PHI where PHI.x=X.id)
union all
select 1 as ref, id as x, 1 as cPhiRef, ref as cPhix
from   Y;
```

Die erste `select`-Abfrage bettet $\phi : X_0 \rightarrow Y$ in $c\phi : X_0 +_{ref} Y \rightarrow X +_{ref} Y$ ein, indem sie das direkte Produkt $i \times \phi$ der Abbildungen $i : \{0\} \rightarrow \{1\}$ und ϕ erzeugt. Das zweite und dritte `select` Statement erzeugt jeweils die identischen Abbildungen

$$\text{id}_{\{0\} \times (X \setminus X_0)} \quad \text{und} \quad \text{id}_{\{1\} \times Y}.$$

Die (disjunkte) Vereinigung dieser drei Relationen erzeugt den Graph von $c\phi$.

Für die Menge $X +_{ref} Y$ haben wir oben bereits ein SQL Statement angegeben. Die Quotientenmenge einer Tabelle erhalten wir durch `group by` Klauseln:

```
create view YpastefX as
select CPHI.cPhiRef,CPHI.cPhix
from   XplusY inner join CPHI
       on (XplusY.ref=CPHI.ref and XplusY.id=CPHI.x)
group  by CPHI.cPhiRef,CPHI.cPhix;
```

Die Topologie für die Verklebung `YpastefX` ist durch die Relation `SpastefR` definiert, die wir durch folgende Abfragen erhalten:

Zunächst ergibt folgende Abfrage

```
create view RplusS as
select 0 as ref, a, b
from   R
union all
select 1 as ref, a, b
from   S;
```

die Topologie des Summenraums. Dessen Finaltopologie bezüglich `cntF` erzeugt die Topologie der Verklebung:

```

create view SpastefR as
select  FA.cPhiRef as refA ,FA.cPhix as a
        ,FB.cPhiRef as refB ,FB.cPhix as b
from    CPHI as FA, RplusS, CPHI as FB
where   FA.ref = RplusS.ref and
        FA.x = RplusS.a and
        FB.ref = RplusS.ref and
        FB.x = RplusS.b;

```

Auf einen Beweis, dass diese Abfragen korrekt die Verklebung konstruieren, wird hier verzichtet. Im wesentlichen muss dabei nur gezeigt werden, dass CPHI der Graph von $c\phi$ ist und dass die Fasern von $c\phi$ in der Tat die Äquivalenzklassen von \sim_ϕ sind. Verkleben ist offenbar eine Konstruktion, die ohne transitive Hülle vorgenommen werden kann.

4.4.8 Versionen von Räumen

In UNIX gibt es die Befehle `diff` und `patch` zum Vergleichen und zum Aktualisieren von Textdateien. Der Befehl `diff` vergleicht zwei Dateien `x` und `y` zeilenweise und gibt diejenigen Abschnitte mit Zeilennummer paarweise aus, die sich unterscheiden. Das Ergebnis `d` von `diff x y` kann nun verwendet werden, um mit dem Befehl `patch -o y x d` die Datei `y` anhand der Datei `x` und den Unterschieden in `d` zu rekonstruieren. Man beachte, dass `d` nicht eindeutig durch `x` und `y` bestimmt ist. Insbesondere erlaubt der Befehl `diff` die Auswahl mehrerer Algorithmen zur Bestimmung der Unterschiede.

Der Zweck derartiger Befehle besteht darin, Werkzeuge zur Versionierung von Textdateien zu unterstützen und insbesondere, um mehrere Versionen eines Textes platzsparend abzuspeichern. Beliebt sind z.B. die Versionskontrollsysteme RCS, CVS und Subversion und diese werden in der Softwareentwicklung häufig eingesetzt. Auch die hier vorliegende Arbeit wurde unter CVS-Versionskontrolle erstellt.

Da die räumliche Planung ebenfalls mit Versionen von Räumen arbeitet, macht es Sinn, dieses Konzept der Versionskontrolle auf topologische Räume zu übertragen. Wir wollen hier eine topologische Variante für die Differenz zweier topologischer Räume angeben:

Definition 4.28 (diff und patch). *Seien $\underline{X} := (X, \mathcal{T}_X)$ und $\underline{Y} := (Y, \mathcal{T}_Y)$ zwei topologische Räume und sei A Teilraum von \underline{X} und B Teilraum von \underline{Y} . Dann sagen wir, das Paar (A, B) steht in diff-Relation zu dem Paar $(\underline{X}, \underline{Y})$, wenn gilt*

$$(\underline{X} \setminus A) \sqcup B = \underline{Y}.$$

Dann schreiben wir:

$$(A, B) \in \mathbf{diff}(\underline{X}, \underline{Y}).$$

Den Ausdruck $(\underline{X} \setminus A) \sqcup B$ bezeichnen wir als Patch von \underline{X} mit (A, B) oder $\mathbf{patch}(\underline{X}, (A, B))$.

Der Teil A ist also aus dem Raum zu entfernen, so dass $\underline{X} \setminus A$ entsteht und in diesen ist der Teil B hineinzukleben. Damit sind B und $\underline{X} \setminus A$ im allgemeinen nicht disjunkt. Dies macht die Bezeichnung *Patch* bzw. „Flicken“ anschaulich, da der eingeklebte Teil ganz analog zu einem realen Flecken im Allgemeinen

noch einen Heftrand haben muss. Dass es immer ein \mathbf{diff} gibt, dieses aber nicht eindeutig bestimmt ist, sieht man leicht ein, denn für jedes beliebige Paar (X, Y) von Räumen gilt $(X, Y) \in \mathbf{diff}(X, Y)$ und zudem gilt $(\emptyset, \emptyset) \in \mathbf{diff}(X, X)$. Die Kunst besteht in der Abwägung zwischen der Anforderung nach möglichst kleinen \mathbf{diff} s, um Speicherplatz zu sparen, und der Anforderung nach möglichst effizienten Algorithmen zu deren Berechnung, die dann aber nicht immer ein minimales Ergebnis liefern.

Diese Definition von \mathbf{diff} und \mathbf{patch} umfasst auch Textdateien, wenn man diese als topologische Räume betrachtet. Dies ist möglich, indem die Zeilen zunächst nummeriert werden. Diese nummerierten Zeilen und deren Zeilenzwischenräume bilden dann die Punktmenge eines Raums, dessen Topologie dadurch definiert ist, dass jede Zeile Berührungspunkt des vorhergehenden und des nachfolgenden Zwischenraums ist.

Mit diesen Vergleichs- und Aktualisierungsoperationen ist es nun möglich, ein Versionensystem von topologischen Räumen ganz in Analogie zu den entsprechenden Werkzeugen [Ves03] aus der Softwareentwicklung zu erstellen.

4.4.9 Diskussion der Ergebnisse

Wir haben gesehen, dass es zu jedem Operator der relationalen Algebra eine entsprechende Konstruktion für die topologischen Datenbanken gibt. Wie wir ebenfalls gesehen haben, kann diese wiederum mit einer relationalen Algebra mit transitiver Hülle ausgeführt werden, im Allgemeinen aber nicht mit einer relationalen Algebra ohne transitive Hülle. Diese Notwendigkeit besteht zwar für den allgemeinen Fall, aber in der Praxis der Modellierung von Gebäuden besteht diese Notwendigkeit wohl nicht. Denn wir können für Entwürfe von topologischen Datenbanken folgende Vereinbarung treffen: Legen wir für jeden Typ (X, R) eine natürliche Zahl n fest und fordern für die Relation $R \subseteq X \times X$ als Konsistenzregel, dass $R^n \subseteq \Delta_X$ gilt: die n -te Potenz von R darf nur Datensätze der Form (a, a) enthalten. Dann gilt für die transitive Hülle:

$$R^+ = R^{1..n} := R^1 \cup R^2 \cup \dots \cup R^n.$$

Denn für jede höhere Potenz R^{n+i} gilt

$$R^{n+i} = R^n \circ R^i \subseteq \Delta_X \circ R^i = R^i,$$

und somit $R^{n+i} \subseteq R^{1..n}$. Also haben wir $R^{k+1} \subseteq R^{1..n}$ für jede natürliche Zahl k . Die Zahl n (zB. $n = 42$) ist dann auch eine obere Schranke der Dimension eines relationalen Komplexes. Wir bezeichnen hier deshalb die kleinste Zahl n mit $R^n \subseteq \Delta_X$ als *Dimension* von R . Dies ist jedoch nur als eine informelle Bezeichnung zu verstehen, da eine formale Definition von „Dimension“ für die topologischen Datenbanken einer ausführlichen Diskussion bedarf. Insbesondere sollte dann die Relation azyklisch, sein, also $(R \setminus \Delta_X)^+ \cap \Delta_X = \emptyset$. Alexandroffsche Räume mit dieser Eigenschaft bezeichnet man als T_0 -trennbar.

In der Praxis dürfte eine Dimension von fünf ausreichend sein: Drei Dimensionen verwenden wir für die Topologie eines Gebäudes zu einem festen Zeitpunkt bei einer bestimmten Version. Eine weitere Dimension benötigen wir zur Modellierung von Bau- und Fertigungsprozessen über die Zeit und eine weitere Dimension dient der Modellierung einer Versionsgeschichte der laufenden Planung. Allerdings gibt es keinen Grund, sich von Anfang an auf eine derartige

obere Schranke festzulegen. Es erscheint sinnvoller, erst am Ende des Entwurfs eines topologischen Datenbankschemas zu fragen, ob die Dimension der Daten nach oben beschränkt ist und von welcher Dimension die Daten dann höchstens sein werden.

4.4.10 Ausblick

Über die Grundoperatoren der relationalen Algebra hinaus gibt es noch viele weitere Möglichkeiten der Abfrage von topologischen Datenbanken. So bietet sich für die Kategorie der relationalen Kettenkomplexe eine Verallgemeinerung der Euleroperatoren an, wie sie in den einleitenden Kapiteln über die Volumenmodellierer vorgestellt wurden. Es reicht hier aus, für einen Komplex (X, D) die Operation `makeCell(cell, cycle)` anzubieten. Dabei ist `cell` eine neue, noch nicht in X vorkommende Zelle und `cycle` ist eine Linearkombination von Zellen, deren Rand in D ein (partieller) Nullvektor ist. Ist diese bereits ein Rand, dann entsteht durch `makeCell` ein neuer Zyklus, was den Euler Operatoren MEMR (make edge make ring), MFMH (make face make hole) oder MVMS (make vertex make solid) entspricht, oder er ist kein Rand und entspricht den Euler Operatoren MEJS (make edge join solid) oder MFMH (make face make hole). Damit ist dann die Anheftung der Zelle an den Komplex definiert. Die entsprechend inverse Operation `killCell(cell)` hingegen ist nur zulässig, wenn `cell` nicht im Rand einer anderen Zelle liegt, und ändert ebenfalls entsprechend die Bettizahlen.

Weitere topologische Abfragen, wie die gemeinsame Verfeinerung bezüglich der geometrischen Realisierung erfordern die Erweiterung dieses Modells um geometrische Eigenschaften. Dies wird hier jedoch in dieser Arbeit nicht geleistet. Eine solche geometrische gemeinsame Verfeinerung ist diejenige Operation, die in Geoinformationssystemen als „Überlagerung“ bezeichnet wird. Werden z.B. zwei Geschossgrundrisse übereinander „gestapelt“, dann zeichnet sich auf deren gemeinsamer Geschossplatte ein Muster ab, das die gemeinsame Verfeinerung dieser beiden Geschossgrundrisse darstellt. Damit entsteht ein neuer topologischer Raum, den man als „Geschossstapel“ bezeichnen könnte.

Derartiges Stapeln ist auch in abstrakterer Form sinnvoll, denn auch dieses kann zum Vergleich einer älteren mit einer neueren Version eines Gebäudeentwurfs eingesetzt werden und ist eine mögliche Alternative zum oben vorgestellten `diff`, die dann aber mehr Speicherplatz benötigt. Die Abbildung 4.4 zeigt einen Geschossstapel und eine Idee, wie diese Variante für ein `diff` in etwa aussehen könnte.

4.5 Der Entwurf topologischer Datenbanken

Die Theorie der relationalen Datenbanken lässt sich grob in zwei Aspekte unterteilen: Zum Einen gibt es die Theorie des relationalen Datenmodells mit den Relationenschemata und den Abfragen und zum Anderen die Theorie des Datenbankentwurfs. Diese Theorie behandelt sinnvolle Einschränkungen von Relationen falls damit Sachverhalte der realen Welt dargestellt werden sollen. Wir wollen hier nun einige Eigenschaften topologischer Datenbanken diskutieren, die als Konsistenzregeln für den Entwurf von Planungssystemen sinnvoll erscheinen. Dabei soll vor allem die Konzeption von Detailbibliotheken diskutiert werden.

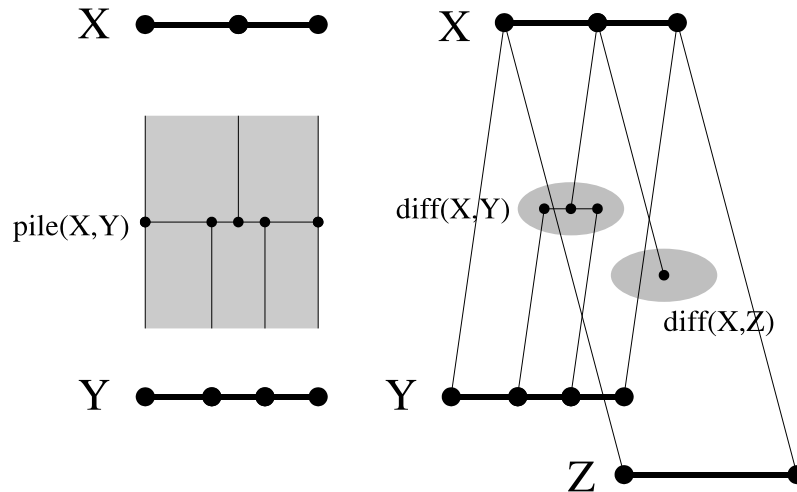


Abbildung 4.4: Links: Ein „Geschossstapel“ des topologischen Raums X auf dem Raum Y enthält einen Teilraum, der eine offensichtliche Ähnlichkeit mit der gemeinsamen Verfeinerung hat. Rechts: Ein hypothetischer diff -Operator zum Vergleich zweier topologischer Räume und dessen Einsatz in einer Versionsverwaltung. diff ist ein minimaler Ausschnitt aus dem „Stapel“ mit dem X in Y (bzw. Z) überführt werden kann.

Um Missverständnissen vorzubeugen, sei hier klargestellt, dass es hier *nicht* um den architektonischen Gebäudeentwurf geht.

4.5.1 Normalformen

Wenn wir für jede Detaillierungsstufe eines Plans eines Gebäudes eine entsprechende Verfeinerung vornehmen, dann erhalten wir sehr viel redundante Information, denn es wiederholen sich viele gleichartige Teile des Raums. In einem mehrstöckigen Wohnhaus liegen z.B. gleichartige Geschosse übereinander und in einem Geschoss wird oft ein gleiches Bauteil mehrfach verwendet. Derartige redundante Daten sollen jedoch nicht explizit in der Datenbank abgespeichert werden, denn einerseits will man die Daten mit möglichst wenig Speicheraufwand ablegen, andererseits führt Redundanz auch zur Gefahr inkonsistenter Daten. Diese Inkonsistenz entsteht z.B., wenn ein Element mehrmals in der Datenbank abgelegt wurde. Soll dieses Element geändert werden, dann ist darauf zu achten, dass diese Änderung bei jedem Auftreten dieses Elements in der Datenbasis vorgenommen werden muss. Ansonsten können zwei verschiedene Ausprägungen ein- und desselben Elements in der Datenbasis existieren. So etwas ist zwar auch in der realen Welt durchaus möglich, wie z.B. bei Schrödingers Katze, die gleichzeitig tot und lebendig ist, oder bei Kunden von Telefongesellschaften, die nach einem Umzug gleichzeitig sowohl unter der alten als auch unter der neuen Adresse geführt werden und so die Kundenstatistik verbessern. In der überwiegenden Zahl der Fälle sind derartige Ausprägungen von Datenbanken jedoch

unerwünscht und werden als Änderungsanomalie bezeichnet. Weitere Anomalien von Datenbanken entstehen beim Einfügen und Löschen von Datensätzen und heißen entsprechend Lösch- und Einfügeanomalien.

Es ist Aufgabe einer Theorie des Datenbankentwurfs, Regeln anzugeben, mit denen sich derartige Anomalien und Redundanzen vermeiden lassen. Von Datenbanken, die diese Regeln einhalten, sagt man, sie erfüllen die sogenannten *Normalformen*. Zur Definition der Normalformen sei, wie gesagt, auf die Literatur verwiesen. Die (neben NF1) wichtigsten Normalformen haben ebenfalls wenig fantasievolle Namen und heißen zweite und dritte Normalform (NF2 und NF3). Als Entitätstyp bezeichnet man ein Objekt (des Denkens oder der Anschauung), das überwiegend als „für sich selbst“ existierend gedacht wird. Dem gegenüber steht der Beziehungstyp, der von der Existenz von Entitätstypen abhängig ist und nicht „für sich selbst“ existiert. Diese Unterscheidung ist jedoch nicht streng formal zu verstehen sondern eher als eine pragmatische Entwurfshilfe gedacht und wird z.B. von CODD abgelehnt [Cod90, S.477]. Stehen nun zwei Entitätstypen in einer $m : n$ -Beziehung und werden diese zusammen mit deren Beziehung in einer einzigen Relation gespeichert, dann verstößt diese Relation gegen die zweite Normalform. Wird dieser Entwurfsfehler bei einer $1 : n$ -Beziehung gemacht, dann haben wir einen Verstoß gegen die dritte Normalform. Eine topologische Datenbank, die zu jedem Gebäudeelement ein mehrmals verwendetes Detail jeweils als Kopie abspeichert, verstößt also gegen die dritte Normalform, da der Entitätstyp „architektonisches Element“ und der Entitätstyp „Detailelement“ in einer $n : m$ Beziehung steht: Das Detail zu einem architektonischen Element besteht i.A. aus mehreren Detailelementen und jedes Detailelement kann natürlich von mehreren architektonischen Elementen verwendet werden. Die Beseitigung eines solchen Entwurfsfehlers ist stets *Normalisierung*: Die „Verfeinerung“⁴ der ungeeigneten Relation R in mehrere Teilrelationen R_1, \dots, R_n . Diese sind dann jeweils in NF3 und deren Verbund $R_1 \bowtie \dots \bowtie R_n$ ergibt wiederum R . Details hierzu können der Literatur entnommen werden [Mai83, S.101ff].

Wir wollen hier nun einen topologischen Datentyp eines zu stark detaillierten Entwurfs (D, R) zu zwei topologischen Datentypen (E, T) und (Y, S) verfeinern. Dabei wollen wir (E, T) als Entwurfsskizze und (Y, S) als Detaildatenbank bezeichnen und hier topologische Eigenschaften derartiger Zerlegungen untersuchen.

4.5.2 Trivialität und Monotonie

Da wir uns hier mit der künstlerischen Beurteilung von Bauwerken grundsätzlich zurückhalten, meint keiner der beiden Begriffe ein ästhetisches Urteil über Architektur. Trivialität und Monotonie sind vielmehr sehr nützliche Eigenschaften von stetigen Abbildungen. Monotonie ist durch die monoton steigenden bzw. fallenden Abbildungen allgemein bekannt. Hier werden wir topologische Verallgemeinerungen dieses Monotoniebegriffs und seinen praktischen Nutzen kennenlernen. Eine Abbildung heißt trivial, wenn sie im wesentlichen (also bis auf Homöomorphie) nur eine Projektionsabbildung aus einem Produktraum ist.

⁴In der Literatur ist hierfür die Bezeichnung „Zerlegung“ gebräuchlich, obwohl es sich im strengen Sinne gar nicht um eine Zerlegung handelt. Bezüglich der Relationenschemata wird eine *feinere Überdeckung* der Attributmenge vorgenommen und bezüglich der Relationen eine *Projektion*, an keiner Stelle jedoch eine echte Zerlegung.

Definition 4.29 (Trivialität). Sind X , Y und Z topologische Räume und ist $h : Z \rightarrow X \times Y$ ein Homöomorphismus und $\pi : X \times Y \rightarrow Y$ die Projektion $\pi(x, y) := y$, dann sagt man zu der Abbildung $\pi \circ h : Z \rightarrow Y$ auch Z ist trivial über X .

Bei Gebäuden finden wir viele derartige triviale Teilräume. Wenn z.B. Z ein Wandabschnitt mit einem festen Schichtaufbau ist, dann ist dieses dreidimensionale Sandwichpaket der Produktraum einer zum Wandabschnitt parallelen zweidimensionalen Fläche X und einer eindimensionalen Abfolge Y von Schichten.

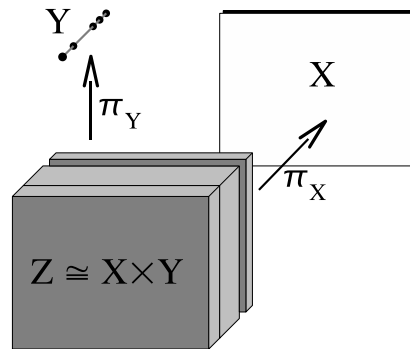


Abbildung 4.5: *Trivialität eines Wandabschnitts. Ein Wandabschnitt Z ist homöomorph zum Produkt einer zweidimensionalen Fläche X und eines eindimensionalen Schichtaufbaus Y .*

Wenn es nun zu jeder dieser Stellen x sogar eine triviale Umgebung U_x von x gibt, dann bezeichnet man dies als „lokal trivial“. Allerdings ist die Forderung, dass immer auch eine Umgebung trivial sein muss, für unsere Zwecke zu streng, denn dann müsste sich z.B. ein Detail für einen Fensterrahmen auch als Detail für die Fensterfläche, für die angrenzende Wand, für den zugehörigen Innenraum und sogar für den Außenraum eignen.

Es liegt jedoch nahe, auszunutzen, dass Architektur an vielen Stellen trivial ist. Dies erreichen wir, indem wir anstelle trivialer Teilräume der Form $X \times Y$ nur die Faktoren X und Y speichern und bei Bedarf $X \times Y$ neu berechnen. X kann dann Teil des Entwurfs und Y ein Element einer Bauteilbibliothek sein. Damit kann das Detail Y auch mehrmals wiederverwendet werden, etwa wenn verschiedene Wandabschnitte den gleichen Aufbau haben. Dazu verwenden wir einen topologischen Datentyp (E, T) mit den Entwurfselementen und einen weiteren Datentyp (Y, S) mit den zugehörigen Details und ein Prädikat Θ , das angibt, welches Bauelement in A durch welches Detail in D realisiert wird. Wir haben also einen praktischen Anwendungsfall des Theta-Verbundes $(A, R) \bowtie_{\Theta} (D, S)$.

Allerdings ist dies nicht ganz so einfach, wie dies auf den ersten Blick wirkt. An einem einfachen Beispiel werden wir nun sehen, dass die naheliegende Verwirklichung mit einer Familie von konstruktiven Details und einem Fremdschlüssel d , der für jedes Element x angibt, welches Detail $d(x)$ verwendet werden soll, nicht ohne weiteres funktioniert.

Betrachten wir z.B. eine Wandvorlage w mit Stirnseite s , dann ist die Stirnseite Berührungspunkt der Wandvorlage und folglich $V = (\{w, s\}, \{(w, s)\})$ der entsprechende topologische Datentyp. Andererseits könnte $W = (\{a, m, b\}, \{(m, a), (m, b)\})$ ein topologischer Datentyp für den Schichtaufbau einer massiven Wand mit Innerem m und Oberflächen a, b sein. Für die Wandvorlage wollen wir ein viereckiges Profil S vorsehen. Dieses Profil habe ein Inneres P , Seitenflächen d, e, f, g und Eckpunkte i, j, k, l . Ein topologischer Datentyp dafür ist

$$S = (\{i, d, j, e, P, f, k, g, l\}, \{(i, i), (d, i), (d, d), (d, j), (j, j), (e, i), (P, d), (f, j), (e, e), (P, e), (P, P), (P, f), (f, f), (e, k), (P, g), (f, l), (k, k), (g, k), (g, g), (g, l), (l, l)\}).$$

Nun können wir die beiden Details Schichtaufbau und Profil disjunkt vereinigen, indem wir jedem Detail einen Index zuordnen, z.B. 0 für den Schichtaufbau und 1 für das Viereckprofil, und dann ist $B := W +_{\text{ref}} S$ eine einfache Detailbibliothek. Nun könnte $V' := (\{(w, 0), (s, 1)\}, \{((w, 0), (s, 1))\})$ die selbe Wandvorlage sein, wobei die zweite Komponente jeweils auf ein Detail der Detailbibliothek B verweist. Die Wand $(w, 0)$ hat also das Detail W und die Stütze $(s, 1)$ das Detail S . Betrachten wir nun den relationalen Verbund der topologischen Datentypen

$$V' \bowtie_{\pi_2=\text{ref}} B := \underline{\sigma}_{(\pi_2=\text{ref})}(V' \times B).$$

Auf Abbildung 4.6 sehen wir, dass der so detaillierte Raum dabei „zerrissen“ wurde. Die Leserin kann sich davon überzeugen, dass auf dieser Abbildung in der Tat $V' \bowtie_{\pi_2=\text{ref}} B$ zu sehen ist.

Wir beobachten, dass die Stirnseite der Wandvorlage durch den Verbund von der Wandvorlage abgetrennt wurde und der anfangs zusammenhängende Raum nun aus zwei unzusammenhängenden Teilen besteht. Wollen wir zunächst genau sagen, was „zusammenhängend“ eigentlich bedeutet.

Zusammenhang

Ein Raum ist zusammenhängend, wenn jeder Teil dieses Raumes mit dem restlichen Raum „verbunden“ ist, diesen also berührt oder von ihm berührt wird:

Definition 4.30 ((Weg-) Zusammenhang). *Ein topologischer Raum $\underline{X} = (X, \mathcal{T}_X)$ heißt zusammenhängend, wenn es keine zwei nichtleere offenen Mengen A, B gibt mit $A \cap B = \emptyset$ und $A \cup B = X$.*

\underline{X} heißt wegzusammenhängend, wenn für alle Punkte $a, b \in X$ ein Weg von a nach b in \underline{X} existiert.

Eine Teilmenge $A \subseteq X$ heißt (weg-) zusammenhängend in \underline{X} , wenn \underline{A} als Teilraum von \underline{X} (also $\underline{A} = \underline{X}|_A$) (weg-) zusammenhängend ist.

Es ist allgemein bekannt, dass zwischen Zusammenhang und Wegzusammenhang ein kleiner Unterschied besteht. Wegzusammenhängende Räume sind stets zusammenhängend, denn ansonsten gäbe es offene Mengen A und B mit $A \cup B = X$ und $A \cap B = \emptyset$ und für jedes $a \in A$ und $b \in B$ einen Weg

$w : [0, 1] \rightarrow \underline{X}$ von a nach b . Dann läge ein Teil dieses Weges in A und ein anderer Teil in B . Dann wäre aber das Einheitsintervall $[0, 1]$ zerlegt in zwei nichtleere Teilmengen $w^{-1}[A]$ und $w^{-1}[B]$, die zudem noch offen wären, da w stetig ist und A, B offen sind. Damit wäre aber $[0, 1]$ nicht zusammenhängend. Es ist jedoch allgemein bekannt, dass $[0, 1]$ zusammenhängend ist, und somit kann ein wegzusammenhängender Raum, der nicht zusammenhängend ist, gar nicht existieren. Umgekehrt gibt es aber sehr wohl zusammenhängende Räume, die nicht wegzusammenhängend sind (siehe z.B. JÄNICH [Jän01, S.19f]). Der Zusammenhang von $[0, 1]$ ist leicht gezeigt, denn die Behauptung, es gelte das Gegenteil, führt zu einem Widerspruch: Diese würde nämlich bedeuten, es gäbe eine Zerlegung von $[0, 1]$ in zwei offene Teilmengen A und B . Dann sind A und B nicht leer und disjunkt. Wählen wir je einen Punkt $a \in A$ und $b \in B$ aus diesen Mengen aus und definieren die Folgen $(a_i)_{i \in \mathbb{N}}$ und $(b_i)_{i \in \mathbb{N}}$ mit $a_0 := a$

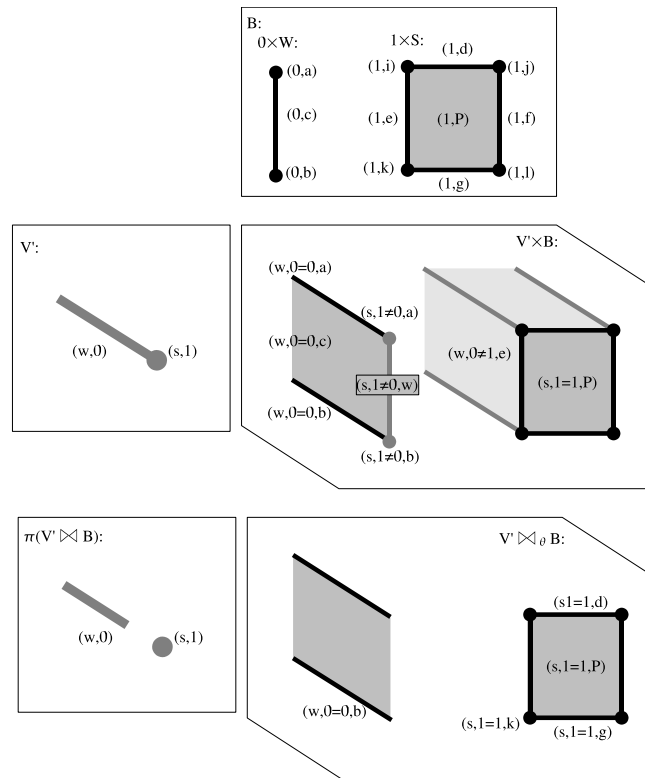


Abbildung 4.6: Der relationale Verbund $V' \times_{\Theta} B$ des topologischen Datentyps V' mit der Bauteilbibliothek B ist der Teilraum aller Tupel (a, i, j, b) des Produkt-raums $V' \times B$ mit $i = j$, wie z.B. $(s, 1, 1, p)$. Um (Un-) Gleichheit anzuzeigen, wurde das mittlere Komma jeweils durch das entsprechende (Un-) Gleichheitszeichen ersetzt. Offensichtlich ist das Ergebnis nicht das erwünschte, denn die Projektion auf die ersten Komponenten ergibt einen ganz anderen topologischen Raum $\pi(V' \times_{\Theta} B)$, in dem die Stütze $(s, 1)$ vom Wandabschnitt $(w, 0)$ abgetrennt wurde.

und $b_0 := b$ und für alle $i \in \mathbb{N}$

$$a_{i+1} := \begin{cases} a_i & \text{falls } \frac{a_i+b_i}{2} \in B \\ \frac{a_i+b_i}{2} & \text{falls } \frac{a_i+b_i}{2} \notin B \end{cases}, \quad b_{i+1} := \begin{cases} b_i & \text{falls } \frac{a_i+b_i}{2} \in A \\ \frac{a_i+b_i}{2} & \text{falls } \frac{a_i+b_i}{2} \notin A \end{cases}.$$

Dann liegen alle Folgenglieder von (a_i) in A und die Folgenglieder von (b_i) liegen in B . Eine dieser Folgen ist monoton steigend und nach oben beschränkt, die andere fallend und nach unten beschränkt und daher konvergieren beide Folgen gegen einen Grenzwert. Da sich die Differenz der Folgenglieder mit jedem Schritt halbiert, konvergiert die Folge der Differenzen nach 0. Also ist $\lim(a_i) = \lim(b_i)$, beide Folgen haben den gleichen Grenzwert. Nun muss dieser entweder in A oder in B liegen. Doch jede Umgebung um diesen Grenzwert enthält sowohl Folgenglieder aus der Menge A , als auch Glieder aus der Menge B . Also ist der Grenzwert ein Randpunkt von beiden Mengen und damit kein innerer Punkt irgend einer dieser Mengen. Damit ist diejenige Menge, die den Grenzwert dieser Folgen enthält, nicht offen.

Nun braucht uns hier der subtile Unterschied zwischen Zusammenhang und Wegzusammenhang nicht weiter zu interessieren, denn bei topologischen Datenbanken gibt es diesen Unterschied nicht:

Satz 4.18 (Zusammenhang in Alex und DTop). *Ein topologischer Datentyp (bzw. ein alexandroffscher Raum) ist genau dann zusammenhängend, wenn sie (bzw. er) wegzusammenhängend ist.*

Da wir bereits wissen, dass jeder wegzusammenhängende Raum auch zusammenhängend ist und dies dann natürlich auch für jeden wegzusammenhängenden alexandroffschen Raum gilt, ist nur noch zu zeigen, dass ein nicht wegzusammenhängender alexandroffscher Raum auch nicht zusammenhängend ist:

Beweis. Sei also ein alexandroffscher Raum (X, \mathcal{T}_X) mit topologischem Datentyp (X, R) nicht wegzusammenhängend. Dann gibt es zwei Punkte $a, b \in X$, für die kein Verbindungsweg von a nach b existiert. Sei nun Ca die Menge aller Punkte x , die durch irgend einen Weg w_{ax} mit a verbunden sind und Cb die Menge aller Punkte für die kein Weg nach a existiert. Damit sind Ca und Cb disjunkt und jeder Punkt $x \in X$ ist entweder in Ca oder in Cb . Dann ist auch b Element von Cb und somit ist Cb nicht leer. Auch Ca ist nicht leer, da ein Weg von a nach a existiert, und zwar der konstante Weg $c_a : [0, 1] \rightarrow X, t \mapsto a$. Damit gilt $a \in Ca$. Zudem ist Ca offen, denn ist ein Punkt $x \in \text{St}_R(Ca)$, dann gibt es einen Punkt y in Ca mit $x R y$. Dann behaupten wir, dass die Abbildung

$$w_{xy} : [0, 1] \rightarrow X, t \mapsto \begin{cases} x & : t < 1 \\ y & : t = 1 \end{cases}$$

ein Weg von x nach y in $\{x, y\}$ ist. Da es aber auch von y einen Weg w_{ya} nach a gibt, existiert ebenfalls der komponierte Weg $w_{xy} * w_{ya}$ von x über y nach a und damit ist x in Ca . Wir hatten aber nur vorausgesetzt, dass x in $\text{St}_R(Ca)$ liegt und damit haben wir $\text{St}_R(Ca) \subseteq Ca$, unser Kriterium für offene Mengen in topologischen Datenbanken, und deshalb ist Ca offen.

Auf der anderen Seite ist auch Cb offen, denn liegt x in $\text{St}_R(Cb)$, dann existiert wegen $x R y$ ein Weg w_{xy} von x nach y für ein $y \in Cb$, wie wir eben bei Ca behauptet haben. Gäbe es nun einen Weg von a nach x , dann gäbe es

auch einen Weg von a über x nach y und damit wäre auch $y \in Ca$. Das kann jedoch nicht sein, da Ca und Cb disjunkt sind und y ja bereits in Cb liegt. Also gibt es keinen Weg von a nach x und somit ist $x \in Cb$. Wir haben also auch hier $\text{St}_R(Cb) \subseteq Cb$ und damit ist Cb ebenfalls offen. Also gibt es zwei offene nichtleere Teilmengen Ca und Cb mit $Ca \cap Cb = \emptyset$ und $Ca \cup Cb = X$ und damit ist (X, \mathcal{T}_X) nicht zusammenhängend.

Also folgt aus nicht wegzusammenhängend bereits nicht zusammenhängend und damit ist Zusammenhang und Wegzusammenhang bei alexandroffschen Räumen bzw. in topologischen Datenbanken äquivalent. \square

Die oben verwendete Behauptung, dass w_{xy} ein Weg ist, werden wir später in der etwas allgemeineren Form für $x R^* y$ noch beweisen.

Wir wollen nun für unsere Detaildatenbank fordern, dass in einem detaillierteren Plan all die Teile zusammenhängend sind, die auch in der zugeordneten weniger detaillierten Sicht des selben Raums zusammenhängen. Die Umkehrung, dass zusammenhängende Teile des detaillierteren Plans auch im weniger detaillierten Plan zusammenhängen ergibt sich von alleine:

Satz 4.19 (stetige Bilder zusammenhängender Mengen). *Für jede stetige Abbildung $f : (X, \mathcal{T}_X) \rightarrow (Y, \mathcal{T}_Y)$ gilt: Ist $A \subseteq X$ zusammenhängend in (X, \mathcal{T}_X) , dann ist $f[A]$ zusammenhängend in (Y, \mathcal{T}_Y) .*

Dies ist wohlbekannt und einfach einzusehen, denn ist $f[A]$ nicht zusammenhängend in (Y, \mathcal{T}_Y) , dann ist der Raum $(f[A], \mathcal{T}_Y|_{f[A]})$ nicht zusammenhängend und besteht aus zwei disjunkten offenen nichtleeren Mengen U, V . Deren Urbilder $f^{-1}[U]$ und $f^{-1}[V]$ sind dann offen in $(A, \mathcal{T}_X|_A)$, denn die Einschränkung $f|_A : (A, \mathcal{T}_X|_A) \rightarrow (f[A], \mathcal{T}_Y|_{f[A]})$ ist ebenfalls stetig. Diese Urbilder sind nicht leer, denn sowohl U als auch V enthalten Punkte $f(a) \in U$ und $f(b) \in V$ mit $a, b \in A$. U und V sind ja nichtleere Teilmengen von $f[A]$. Damit ist A ebenfalls nicht zusammenhängend. Die Stetigkeit von $f|_A$ überlegt man sich, indem man eine offene Menge B in $f[A]$ betrachtet. Dann gibt es eine offene Menge B' in Y mit $B = B' \cap f[A]$. Dann ist $f^{-1}[B']$ offen in X und $f^{-1}[B'] \cap A$ somit offen in $(A, \mathcal{T}_X|_A)$. Man muss sich nur noch überlegen, dass $f^{-1}[B'] \cap A$ die selbe Menge ist wie $f^{-1}[B]$.

Monotonie

Es ist also sichergestellt, dass unsere Datenbankabfragen zum Erstellen von detaillierteren Räumen einen derartigen Raum nicht „verklumpen“ und dort mehr Zusammenhang stiften als der weniger detaillierte Entwurf vorsieht. Umgekehrt müssen wir jedoch explizit fordern, dass unsere Details den Entwurf nicht auch noch „zerreißen“ und erhalten somit folgende Konsistenzregel:

Definition 4.31 (Monotonie). *Eine stetige Abbildung $f : (X, \mathcal{T}_X) \rightarrow (Y, \mathcal{T}_Y)$ heißt monoton, wenn das Urbild $f^{-1}[B]$ von jeder zusammenhängenden Menge $B \subseteq Y$ ebenfalls zusammenhängend ist.*

Man findet in der Literatur oft eine schwächere Definition von Monotonie, wonach nur das Urbild jedes Punktes (sogenannte Faser) zusammenhängend sein muss, wie z.B. [AF90, S. 40]. Die hier verwendete Definition von Monotonie stammt von KURATOWSKI [Kur68, Kap. V, § 46. I] und passt besser auf die Bedürfnisse einer Detaildatenbank. Wollen wir fürs erste eine Abbildung,

deren Fasern zusammenhängend sind, als *fast monoton* bezeichnen. Dann ist jede monotone Abbildung fast monoton, denn jeder Punkt y bildet ja auch die zusammenhängende Menge $\{y\}$ und hat somit ein zusammenhängendes Urbild $f^{-1}[\{y\}]$. Die Umkehrung, dass eine fast monotone Abbildung monoton sei, gilt jedoch nicht. So ist z.B. für eine Menge X mit mindestens zwei verschiedenen Elementen die „Identität“

$$\text{id}_X : (X, \mathcal{P}(X)) \rightarrow (X, \{\emptyset, X\}), x \mapsto x$$

vom diskreten Raum in den trivialen Raum zwar stetig und fast monoton, denn wegen $\text{id}_X^{-1}(a) = \{a\}$ ist das Urbild eines Punktes immer zusammenhängend. Dennoch ist id_X nicht monoton (im Sinne von KURATOWSKI), denn X ist zwar zusammenhängend im trivialen Raum, $\text{id}_X^{-1}[\{X\}] = X$ aber nicht zusammenhängend im diskreten Raum.

KURATOWSKIS Definition von Monotonie hat neben dem praktischen Vorteil als Korrektheitskriterium für Bauteildatenbanken auch noch den theoretischen Vorteil, dass diese monotonen Abbildungen Morphismen sind, denn ist das Urbild einer zusammenhängenden Menge zusammenhängend, dann ist dessen Urbild unter einer weiteren monotonen Abbildung ebenfalls zusammenhängend und somit sind Kompositionen monotoner Abbildungen ebenfalls monoton. Bei der schwächeren Definition gilt dies nicht, denn es ist z.B. die Komposition $c_\bullet \circ \text{id}_X$ der Identität id_X im obigen Beispiel mit der fast monotonen konstanten Abbildung $c_\bullet : (X, \{\emptyset, X\}) \rightarrow (\{\bullet\}\{\emptyset, \{\bullet\}\})$ nicht fast monoton, denn das Urbild des Punktes \bullet unter dieser Komposition ist der gesamte diskrete Raum $(X, \mathcal{P}(X))$, der jedoch nicht zusammenhängend ist.

Wenn wir nun für Datenbankabfragen Monotonie fordern können, dann lässt sich der relationale Verbund mit der Detailbibliothek nach unserem obigen Beispiel verbieten, denn die Wandvorlage w mit dem abschließenden Ende s ist ein zusammenhängender Raum, aber die Projektion des detaillierten Plans $W' \times B$ zurück nach W' ist nicht monoton. Zwischen der ausdetaillierten Wand und der ausdetaillierten Stütze gab es keine Verbindung. Damit verletzt die ungeeignete Detailbibliothek B unser Monotoniekriterium.

Wir wollen also Monotonie als Konsistenzregel bereitstellen. Das heißt wir ermöglichen, gewisse Fremdschlüssel als monoton zu deklarieren, etwa durch Erweiterung der `foreign key` Klausel in SQL bei der Definition von Relationen:

```
create table Entwurfselemente(
    id      integer not null
    , detail integer not null
    ...
    , primary key(id)
    , monotonic foreign key(detail)
      references Detailbibliothek
    ...
);
```

Allerdings ist diese Erweiterung so noch unvollständig, denn es sind dazu auch noch die Relationen anzugeben, welche die jeweiligen Topologien definieren.

4.5.3 Wege in topologischen Datenbanken

Bevor wir uns eingehender mit Monotonie beschäftigen, werden wir uns noch mit einigen Grundlagen dazu befassen. Wir werden insbesondere den Begriff des „Weges“ vertiefen und hier einen damit eng verwandten Begriff des „Pfades“ einführen.

Es ist anschaulich klar, dass durch die Wege in einem Raum X eine Äquivalenzrelation „durch einen Weg verbunden“ in X definiert ist. Ein Punkt a ist mit b durch einen Weg verbunden, wenn ein Weg von a nach b existiert:

Lemma 4.20 (Wegzusammenhangskomponenten). *Sei \underline{X} ein topologischer Raum. Dann ist die Relation \sim in X , definiert durch $a \sim b$ gilt genau dann, wenn ein Weg von a nach b existiert, eine Äquivalenzrelation. Ein Element $[a]$ der Quotientenmenge X/\sim heißt auch Wegzusammenhangskomponente.*

Dies ist allgemein bekannt und wird üblicherweise wie folgt bewiesen:

Beweis. Wir überlegen uns die drei Eigenschaften Reflexivität, Symmetrie und Transitivität: Für jeden Punkt x in X ist die konstante Abbildung

$$c_x : [0, 1] \rightarrow \underline{X}, t \mapsto x$$

ein Weg von x nach x und damit gilt $x \sim x$, also die Reflexivität.

Ist $w_{ab} : [0, 1] \rightarrow \underline{X}$ ein Weg von a nach b , gilt also $a \sim b$, dann existiert auch der Rückweg w_{ab}^- , ein Weg von b nach a . Dieser wird wie folgt konstruiert: Die Abbildung

$$r : [0, 1] \rightarrow [0, 1], t \mapsto r(t) := 1 - t$$

ist offensichtlich eine stetige Abbildung und damit ist $w_{ab}^- := w_{ab} \circ r$ ebenfalls stetig und somit ein Weg in \underline{X} . Dieser geht von $w_{ab}^-(0)$ nach $w_{ab}^-(1)$ und dies ist

$$\begin{aligned} w_{ab}^-(0) &= w_{ab} \circ r(0) = w_{ab}(r(0)) = w_{ab}(1 - 0) = w_{ab}(1) = b \\ w_{ab}^-(1) &= w_{ab} \circ r(1) = w_{ab}(r(1)) = w_{ab}(1 - 1) = w_{ab}(0) = a. \end{aligned}$$

Damit haben wir einen Weg von b nach a und es gilt $b \sim a$, also gilt die Symmetrie.

Fehlt noch die Transitivität: haben wir einen Weg w_{ab} von a nach b und einen Weg w_{bc} von b nach c , dann erinnern wir uns an die Komposition dieser Wege $w_{ab} * w_{bc}$. Diese ist ein Weg von a nach c , also existiert ein solcher Weg und es folgt $a \sim c$. \square

Wir werden nun sehen, in welcher Beziehung die Relation R eines topologischen Datentyps (X, R) mit den Wegen im entsprechenden topologischen Raum (X, \mathcal{T}_R) steht. Dazu benötigen wir ein Klasse von Wegen, die in höchstens drei Schritten gegangen werden:

Definition 4.32 (elementarer Weg). *Sei $\underline{X} = (X, \mathcal{T}_X)$ ein topologischer Raum. $w : [0, 1] \rightarrow \underline{X}$ ein Weg in \underline{X} . Dann bezeichnen wir w als elementar, wenn die Einschränkung $w|_{]0,1[}$ von w konstant ist.*

Zu jedem Paar $a R^* b$ eines topologischen Datentyps (X, R) gibt es einen charakteristischen elementaren Weg w_{ab} , dessen Definition wir vom Beweis zu Satz 4.18 bereits gesehen haben:

Lemma 4.21 (charakteristischer elementarer Weg). Sei (X, R) ein topologischer Datentyp und $a R^* b$. Dann existieren die elementaren Wege

$$w_{ab} : [0, 1] \rightarrow (X, R), t \mapsto w_{ab}(t) := \begin{cases} a & : t < 1 \\ b & : t = 1 \end{cases}$$

von a nach b und dessen Rückweg

$$w_{ab}^- : [0, 1] \rightarrow (X, R), t \mapsto w_{ab}^-(t) := \begin{cases} a & : t > 0 \\ b & : t = 0 \end{cases}$$

von b nach a .

Beweis. Es ist klar, dass die Einschränkungen $w_{ab}|]0, 1[$ und $w_{ab}^-|]0, 1[$ den konstanten Wert a haben.

Wir haben zu zeigen, dass w_{ab} eine stetige Abbildung von $[0, 1]$ nach (X, R) ist. Ist A eine offene Menge in (X, \mathcal{T}_R) , dann ist entweder $b \in A$ oder $b \notin A$.

Fall $b \in A$: Ist $b \in A$, dann ist wegen $a R^* b$ auch $a \in \text{St}_R^*[A]$, und da A offen ist, gilt $\text{St}_R^*[A] = A$, und somit $a \in A$. Damit folgt $w_{ab}^{-1}(A) = [0, 1]$ und diese Menge ist offen in $[0, 1]$.

Fall $b \notin A$: Ist $b \notin A$, dann kann entweder $a \in A$ sein und es ist $w_{ab}^{-1}(A) = [0, 1[$, oder $a \notin A$ und $w_{ab}^{-1}(A) = \emptyset$. Beide Mengen sind offene Teilmengen von $[0, 1]$ und somit offen.

In jedem Fall ist das Urbild der offenen Menge A eine offene Teilmenge von $[0, 1]$ und damit ist w_{ab} stetig, also ein Weg von a nach b , denn $w_{ab}(0) = a$ und $w_{ab}(1) = b$.

Wegen $w_{ab}^-(t) = w_{ab}(1 - t)$ ist w_{ab}^- in der Tat der Rückweg von w_{ab} und somit ebenfalls ein Weg in (X, \mathcal{T}_R) . \square

Ist nun $\underline{X} = (X, \mathcal{T}_R)$ ein alexandroffscher Raum mit topologischem Datentyp (X, R) , dann ist die von R erzeugte Äquivalenzrelation $R^\sim = (R \cup R^T)^*$ genau die von den Wegen in \underline{X} definierte Äquivalenzrelation.

Lemma 4.22 (Zusammenhangskomponenten). Sei R^\sim die von $R \subseteq X \times X$ erzeugte Äquivalenzrelation, also $R^\sim = (R \cup R^T)^*$. Dann ist R^\sim genau die durch die Wege im topologischen Datentyp (X, R) definierte Äquivalenzrelation und X/R^\sim die Menge der (Weg)Zusammenhangskomponenten. Damit gibt es auch für jeden Weg w_{ab} in (X, R) von a nach b eine endliche „Abkürzung“ $a(R \cup R^T)^n b$.

Es erscheint ausreichend zu sein, für elementare Wege w_{ab} bzw. deren Komposition $w_{ab} * w_{bc} * \dots * w_{xy}$ die entsprechenden Tupel mit den Wegpunkten anzugeben, also (a, b) statt w_{ab} und (a, b, c, \dots, x, y) statt $w_{ab} * w_{bc} * \dots * w_{xy}$. Daher wollen wir zunächst einen alternativen Begriff für *Weg* einführen:

Definition 4.33 ((gerichteter) Pfad). Sei (X, R) ein topologischer Datentyp. Ein $n + 1$ -Tupel (Wort) $p_{a_0 a_n} = (a_0, a_1, \dots, a_n)$ heißt Pfad von a_0 nach a_n der Länge n in (X, R) , falls $a_{i-1} R^* a_i$ oder $a_i R^* a_{i-1}$ für alle $i = 1, \dots, n$ gilt. Ein Pfad heißt gerichtet, falls $a_{i-1} R^* a_i$ für alle $i = 1, \dots, n$ oder falls $a_j R^* a_{j-1}$ für alle $j = 1, \dots, n$ gilt.

Sind (a_0, \dots, a_n) und (a_n, \dots, a_{n+m}) Pfade der Länge n und m , dann ist

$$(a_0, \dots, a_n) * (a_n, \dots, a_{n+m}) := (a_0, \dots, a_n, \dots, a_{n+m})$$

die Komposition dieser Pfade und wir bezeichnen

$$(a_0, \dots, a_n)^- := (a_n, \dots, a_0)$$

als inversen Pfad oder Pfad zurück.

Wir bemerken die Ähnlichkeit von Pfadkomposition und Relationenprodukt bzw. inversem Pfad und Transponieren von Relationen.

Zwischen Pfaden und Wegen gibt es nun einen engen Zusammenhang: Zu jedem Pfad p_{ab} gibt es auch einen Weg w_p von a nach b , der die Punkte (a_0, \dots, a_n) genau in dieser Reihenfolge nacheinander besucht. Diesen Weg erhalten wir, wenn wir für jedes (a_i, a_{i+1}) entweder den elementaren Weg $w_{a_i a_{i+1}}$ im Falle $a_i R^* a_{i+1}$, oder den Rückweg $w_{a_{i+1} a_i}^-$ im Falle $a_{i+1} R^* a_i$ nehmen. w_p ist dann die Komposition dieser Einzelwege.

Umgekehrt ist jeder Weg w_{ab} zu einem solchen Weg w_p eines solchen Pfades p_{ab} homotop. Der erste Teil dieser Behauptung wird hier klar werden, mit der Definition des endlichen Weges. Der zweite Teil ist eine alternative Formulierung zum Satz über die endlichen homotopen Wege, den wir erst später formulieren und erst im im Anhang beweisen werden, da er zwar interessant, für diese Arbeit aber nicht wesentlich ist.

Wollen wir nun zeigen, dass $(R \cup R^T)^*$ die Äquivalenzrelation „ist durch einen Weg verbunden“ ist:

Beweis. Sei $R^\sim := (R \cup R^T)^*$. Nach Lemma 2.3 auf Seite 39 ist $(R \cup R^T)^*$ in der Tat eine Äquivalenzrelation. Eine Richtung der Inklusion leuchtet sofort ein, denn gilt $a (R \cup R^T)^* b$, dann existiert eine natürliche Zahl n mit $a (R \cup R^T)^n b$. Durch Induktion über n zeigen wir, dass dann auch ein Weg von a nach b existiert. Ist $n = 0$, dann ist $a = b$ und somit existiert der konstante Weg von a nach b . Existiert für alle $x (R \cup R^T)^n y$ ein Weg w_{xy} und gilt $a (R \cup R^T)^{n+1} b$, dann existiert ein j („jemand“) mit $a (R \cup R^T)^n j (R \cup R^T) b$. Damit haben wir einen Weg von a nach j laut Induktionsvoraussetzung und weiter oben haben wir bereits gesehen, dass für $j (R \cup R^T) b$ ebenfalls ein Weg von j nach b existiert und somit gibt es auch einen Weg von a über j nach b .

Es ist noch zu zeigen, dass für jeden Weg w_{ab} von a nach b in \underline{X} gilt: $a (R \cup R^T)^* b$. Behaupten wir das Gegenteil, dass also für einen Weg w_{ab} von a nach b eben nicht $a (R \cup R^T)^* b$ gilt. Dann gilt ebenfalls $b \notin \text{St}_{(R \cup R^T)^*}^*(a)$. Die Menge $Ca := \text{St}_{(R \cup R^T)^*}^*(a)$ ist offen in \underline{X} , denn sei $x \in \text{St}_R(Ca)$, dann gilt $x R c (R \cup R^T)^* a$ für ein $c \in Ca$. Wegen $R \subseteq R \cup R^T \subseteq (R \cup R^T)^*$ gilt also $x (R \cup R^T)^* c$ und wegen der Transitivität auch $x (R \cup R^T)^* a$ bzw. $x \in Ca$. Damit erfüllt Ca das Kriterium für offene Mengen: $\text{St}_R(Ca) \subseteq Ca$.

Aber auch die Komplementmenge $X \setminus Ca$, die wir hier mit Cb bezeichnen wollen ist offen: Sei $x \in \text{St}_R(Cb)$ beliebig gewählt. Dann existiert ein Element $j \in Cb$ mit $x R j$. Wäre x Element von Ca , dann würde gelten $x (R \cup R^T)^* a$ und wegen $x R j$ folgte $x (R \cup R^T)^* j$, bzw. wegen der Symmetrie auch $j (R \cup R^T)^* x$ und $x (R \cup R^T)^* a$ und aus der Transitivität folgte $j (R \cup R^T)^* a$. Dann folgte $j \in Ca$ und somit $j \notin Cb$, ein Widerspruch zu $j \in Cb$. Also kann x kein Element von Ca sein und ist somit Element von Cb . Also gilt auch hier

$\text{St}_R(Cb) \subseteq Cb$ und damit ist Cb offen. Zudem ist a sowohl Element der Spur des Weges $\text{Bild}(w_{ab})$ als auch Element von Ca . Hingegen ist b Element von Cb und ebenfalls ein Punkt auf der Spur des Weges. Damit lässt sich die Spur in zwei nichtleere offene Mengen $Ca \cap \text{Bild}(w_{ab})$ und $Cb \cap \text{Bild}(w_{ab})$ zerlegen und ist damit nicht zusammenhängend. Dies ist jedoch ein Widerspruch, denn die Spur eines Weges ist als $\text{Bild } w_{ab}([0, 1])$ der zusammenhängenden Menge $[0, 1]$ unter einer stetigen Abbildung nach Satz 4.19 immer zusammenhängend. Da dieser Widerspruch aus der Behauptung folgt, es gebe einen Weg w_{ab} von a nach b , ohne dass zugleich $a(R \cup R^T)^*b$ gilt, muss sie falsch sein. \square

Da nun Punkte a und b genau dann durch einen Weg verbunden sind, wenn $a(R \cup R^T)^*b$ gilt, gibt es stets eine natürliche Zahl n mit $a(R \cup R^T)^n b$ und damit stets auch einen endlichen Pfad $(a, a_1, \dots, a_{n-1}, b)$ von a nach b in (X, R) . Wegen des engen Zusammenhangs von Wegzusammenhang und Existenz von Pfaden können wir unsere Sprechweise ein wenig lockern und erlauben uns, vom Weg (a_0, \dots, a_n) von a_0 nach a_n zu sprechen.

Es gilt nämlich sogar, dass jeder Weg w_{ab} in einem alexandroffschen Raum eine homotope endliche Abkürzung hat, also einen endlichen Weg, der alle Hindernisse in \underline{X} auf die selbe Weise „umkurvt“ wie w_{ab} . Dieser muss ja nicht notwendigerweise endlich sein. Wir wollen nun die Definition des elementaren Weges leicht verallgemeinern und sagen, was wir unter einem endlichen Weg verstehen:

Definition 4.34 (endlicher Weg). *Wir bezeichnen den Weg w als endlich, wenn ein Homöomorphismus $h : [0, 1] \rightarrow [0, 1]$ existiert, so dass der Weg $w \circ h : [0, 1] \rightarrow \underline{X}$ die Komposition $w_1 * \dots * w_n$ von endlich vielen elementaren Wegen ist.*

Damit ist eine solche Komposition $w = w_1 * \dots * w_n$ von endlich vielen elementaren Wegen ebenfalls endlich. Ein endlicher Weg wechselt also in seinem Verlauf nur endlich oft seinen Ort. Nun ist in jedem alexandroffschen Raum ein Weg bis auf Homotopie endlich.

Satz 4.23 (Homotope endliche Wege). *Zu jedem Weg w_{ab} in einem alexandroffschen Raum $\underline{X} = (X, \mathcal{T}_R)$ gibt es einen zu w_{ab} homotopen endlichen Weg $v_{ab} : [0, 1] \rightarrow \underline{X}$.*

Dieser Satz ist jedoch für die weitere Arbeit nicht wichtig, da wir in der Praxis nur mit endlichen Wegen zu tun haben werden. Ein Beweis mit einer Einführung in die dazu benötigten Grundbegriffe findet sich im Anhang. Für diesen Beweis benötigen wir z.B. die sogenannte *Kompaktheit*, eine Endlichkeitseigenschaft mancher topologischer Räume. Insbesondere ist die Spur eines Weges stets kompakt.

4.5.4 Entscheiden von Zusammenhang und Monotonie

Wenn wir Monotonie als Konsistenzregel für Datenbankentwürfe ermöglichen wollen, müssen wir auch in der Lage sein, Zusammenhang zu entscheiden. Bevor wir dies aber tun, wollen wir noch eine weitere einfache Operation mit Relationen einführen — die Einschränkung einer Relation in Analogie zur Einschränkung einer Abbildung:

Definition 4.35 (Einschränkung von Relationen). Sei $R \subseteq X \times Y$ eine Relation und $A \subseteq X$ und $B \subseteq Y$. Dann ist die Relation

$$R|_{A \times B} := (A, B, \{(a, b) \in A \times B \mid a R b\})$$

die Einschränkung von R auf $A \times B$. Ist $X = Y$ und $A = B$, dann schreiben wir auch $R|_A := R|_{A \times B}$ und sagen Einschränkung von R auf A .

Eine Menge A ist nun zusammenhängend, wenn sie aus höchstens einer Zusammenhangskomponente besteht. Dann ist die Zusammenhangsrelation gleich dem kartesischen Produkt, was uns ein Kriterium zur Entscheidung von Zusammenhang in einer topologischen Datenbank gibt:

Lemma 4.24 (Entscheiden von Zusammenhang). Sei A Teilmenge eines topologischen Datentyps (X, R) . Dann ist A genau dann zusammenhängend, wenn $(R^+|_A \cup (R^+)^T|_A)^+ \supseteq (A \times A) \setminus \Delta_A$ gilt.

Offensichtlich muss für dieses Kriterium sehr häufig die transitive reflexive Hülle R^+ einer Relation R berechnet werden. Auf die transitive Hülle im Inneren der Klammern kann nicht verzichtet werden, da wir diese ja zur Berechnung der Spurtopologie benötigen. Ist $R^+|_A$ aber berechnet, dann haben wir sofort auch $(R^+)^T|_A$. Für die äußere Hülle (bzw. die Berechnung von Zusammenhangskomponenten in Graphen) ist nun folgendes bekannt: Ist E die Anzahl der Paare einer Relation $R \subseteq X \times X$, dann kann die Hülle $(R \cup R^T)^*$ dieser Relation mit dem sogenannten *Algorithmus zur Vereinigungssuche* in praktisch linearer Zeit bezüglich E berechnet werden und es existiert kein Algorithmus, der bei ungünstigen Eingabewerten wesentlich schneller ist [Sed91, S. 510]. Praktisch linear meint hier, dass es noch einen von E abhängigen kleinen Korrekturfaktor $\alpha(E)$ gibt, der aber nur bei unrealistisch großen Relationen einen spürbaren Einfluss auf die Rechenzeit hat und ansonsten als konstanter Faktor angesehen werden kann. Das bedeutet, dass es ein Programm P gibt, das bei Eingabe einer Relation R die Zusammenhangskomponenten ermittelt. Zudem ist die von R abhängige Rechenzeit $t_P(R)$ dieses Programms folgendermaßen nach oben beschränkt: Es gibt eine feste natürliche Zahl p , so dass stets $t_P(R) \leq p \cdot (1 + (|R| \cdot \alpha(|R|)))$ gilt. In Landau-Symbolen ist dies $\mathcal{O}(n \cdot \alpha(n))$. Dieser Algorithmus wählt für jede Zusammenhangskomponente $[a]$ genau einen Repräsentanten $r(a) \in [a]$ aus und erzeugt so aus der Relation R eine Abbildung $r : X \rightarrow X$ derart, dass für jede Faser $r^{-1}\{r(x)\} = [x]$ gilt. Ähnliches haben wir mit unserer Datenbankabfrage zum Verkleben von topologischen Datentypen bereits getan. Dort hatte die Abbildung $c\phi : X + Y \rightarrow X + Y$ genau die selbe Eigenschaft.

Wenn wir nun die Zusammenhangskomponenten einmal berechnet haben, dann können wir anstelle der Relation $(R \cup R^T)^*$ diese Abbildung r verwenden. Dann existiert ein Weg von p nach q in \underline{X} genau dann, wenn $r(p) = r(q)$ gilt.

Ein Monotoniekriterium. Zur Entscheidung von Monotonie einer stetigen Abbildung $f : \underline{X} \rightarrow \underline{Y}$ können wir uns manchmal auf die Untersuchung von zwei Fällen beschränken:

- Das Urbild von jedem Punkt $p \in Y$ ist zusammenhängend.
- Das Urbild aller Mengen $\{p, q\}$ mit $p S^+ q$ ist zusammenhängend.

Die erste dieser Eigenschaften ist diejenige notwendige Voraussetzung von Monotonie, die gelegentlich schon als hinreichend akzeptiert wird. Diese lässt sich auch schreiben als

Das Urbild aller Mengen $\{p, q\}$ mit $p S^0 q$ ist zusammenhängend.

Und somit lassen sich beide Eigenschaften wie folgt zusammenfassen:

Das Urbild aller Mengen $\{p, q\}$ mit $p S^* q$ ist zusammenhängend.

Es ist klar, dass eine monotone Abbildung auch diese Eigenschaft bezüglich $p S^+ q$ hat, denn dann existiert ein Weg von p nach q innerhalb von $\{p, q\}$ und damit ist diese Menge zusammenhängend. Dann ist auch $f^{-1}[\{p, q\}]$ zusammenhängend, da f monoton ist. Wir erinnern uns, dass für jeden Punkt p auch die Menge $\{p\}$ zusammenhängend ist und damit ein zusammenhängendes Urbild unter einer monotonen Abbildung f hat. Wir haben hier also eine notwendige Bedingung für Monotonie. Zu zeigen ist also nur noch, dass diese Bedingung unter gewissen Voraussetzungen auch hinreichend ist:

Sei $f : \underline{X} \rightarrow \underline{Y}$ eine *surjektive* Abbildung mit obigen Eigenschaften und $B \subseteq Y$ zusammenhängend. Wählen wir nun zwei beliebige Punkte a und b in $f^{-1}[B]$. Dann ist zu zeigen, dass ein Weg von a nach b in $f^{-1}[B]$ existiert. Ohne Einschränkung sei $a \neq b$, da ansonsten ein solcher Weg trivialerweise existiert. Ist jedoch $a \neq b$, dann existiert ein Pfad (y_0, y_1, \dots, y_n) der Länge n von $f(a)$ nach $f(b)$, also $f(a) = y_0$ und $f(b) = y_n$ und $y_{i-1} R^* y_i$ oder $y_i R^* y_{i-1}$ für alle $i = 1, \dots, n$. Damit gibt es zu jedem Pfadpunkt y_i einen Punkt $x_i \in f^{-1}(y_i) \subseteq f^{-1}[B]$, denn f ist ja surjektiv. Wir wählen $x_0 := a$ und $x_n := b$. Wegen $y_{i-1} R^* y_i$ oder $y_i R^* y_{i-1}$ ist das Urbild $f^{-1}[\{y_{i-1}, y_i\}]$ stets eine zusammenhängende Teilmenge von $f^{-1}[B]$, die somit auch einen Pfad (x_{i-1}, \dots, x_i) von x_{i-1} nach x_i enthält. Damit existiert auch der Pfad

$$(a = x_0, \dots, x_1) * (x_1, \dots, x_2) * \dots * (x_{i-1}, \dots, x_i) * \dots * (x_{n-1}, \dots, x_n = b)$$

als Komposition dieser Einzelpfade von a nach b in $f^{-1}[B]$. Damit ist $f^{-1}[B]$ zusammenhängend und folglich f monoton. Folglich gilt:

Satz 4.25 (monotone surjektive Abbildungen). *Seien (X, R) und (Y, S) topologische Datentypen und $f : (X, R) \rightarrow (Y, S)$ eine surjektive stetige Abbildung. Dann ist f genau dann monoton, wenn für alle $(a, b) \in S^*$ die Menge $f^{-1}[\{a, b\}]$ zusammenhängend in (X, R) ist.*

Ist jedoch f nicht surjektiv, dann ist diese Eigenschaft nicht hinreichend. Sei nämlich folgende Abbildung gegeben:

$$\bullet : \{a, c\} \rightarrow \{\bullet_a \rightarrow \bullet_b \leftarrow \bullet_c\}, \bullet(x) = \bullet_x.$$

Dabei sollen die Pfeile $\bullet_a \rightarrow \bullet_b$ und $\bullet_b \leftarrow \bullet_c$ jeweils die Relation angeben und $\{a, c\}$ sei diskret. Die angegebene Relation „ \rightarrow “ ist transitiv. Da \bullet injektiv ist, erfüllt sie die erste und wegen $\bullet^{-1}[\{\bullet_a, \bullet_b\}] = \{a\}$ und $\bullet^{-1}[\{\bullet_b, \bullet_c\}] = \{c\}$ auch die zweite Eigenschaft.

Allerdings ist zwar der Raum $\{\bullet_a \rightarrow \bullet_b \leftarrow \bullet_c\}$ zusammenhängend, nicht jedoch dessen Urbild $\{a, c\}$ mit der diskreten Topologie. Für nicht surjektive Abbildungen benötigen wir also ein anderes Kriterium. Es ist nun allgemein bekannt, dass zu jeder Abbildung $f : X \rightarrow Y$ eine surjektive Abbildung $g : X \rightarrow Z$

und eine injektive Abbildung $i : Z \rightarrow Y$ existieren, so dass $f = i \circ g$ gilt. Dies bezeichnet man auch als *Faktorisierung* von f .

Ist nun die obige Abbildung $f : (X, R) \rightarrow (Y, S)$ stetig, dann sind die Abbildungen

$$\begin{aligned} g : (X, R) &\rightarrow (f[X], (f \times f)[R]), x \mapsto f(x) \\ i : (f[X], (f \times f)[R]) &\rightarrow (Y, S), f_x \mapsto f_x \end{aligned}$$

eine derartige Faktorisierung mit stetigen Faktoren. Die Stetigkeit von g und i sieht man sofort, denn $(f \times f)[R] \subseteq (f \times f)[R]^*$ ist offensichtlich und damit ist g stetig. Da f stetig ist gilt $(f \times f)[R] \subseteq S^*$ und wegen $(i \times i)[(f \times f)[R]] = (f \times f)[R]$ ist folglich i stetig. Wir können also stetige Abbildungen effizient in je eine surjektive und eine injektive stetige Abbildung faktorisieren.

Für monotone Abbildung gilt dabei nun folgendes:

Satz 4.26 (Faktoren monotoner Abbildungen). *Seien $f : (X, R) \rightarrow (Y, S)$ und $g : (Y, S) \rightarrow (Z, T)$ stetige Abbildungen und sei f surjektiv und g injektiv. Dann sind f und g genau dann monoton, wenn deren Komposition $g \circ f$ monoton ist.*

Damit können wir jede stetige Abbildung in derartige Faktoren zerlegen und den surjektiven und den injektiven Faktor jeweils gesondert auf Monotonie untersuchen.

Beweis. Die eine Richtung der Folgerung ist klar, denn sind f und g monotone Abbildungen, dann ist deren Komposition $g \circ f$ ebenfalls eine monotone Abbildung.

Zu zeigen ist also nur noch, dass umgekehrt bei der Faktorisierung einer monotonen Abbildung in einen surjektiven und einen injektiven Teil diese Teile monoton sind. Seien also $f : (X, R) \rightarrow (Y, S)$ und $g : (Y, S) \rightarrow (Z, T)$ stetige Abbildungen, und sei f surjektiv, g injektiv und $g \circ f$ monoton. Sei $C \subseteq Z$ zusammenhängend in (Z, T) . Dann ist $(g \circ f)^{-1}[C]$ zusammenhängend in (X, R) , da $g \circ f$ ja monoton ist. Dann ist $f[(g \circ f)^{-1}[C]]$ zusammenhängend in (Y, S) , da f stetig ist. Es ist nun $f[(g \circ f)^{-1}[C]] = g^{-1}[C]$ und damit ist g monoton:

\subseteq : Es ist allgemein $(g \circ f)^{-1}[C] = f^{-1}[g^{-1}[C]]$, denn die Menge auf der linken Seite enthält alle Elemente $a \in A$ mit $g(f(a)) \in C$ und auf der rechten Seite sind dies alle $a \in A$ mit $f(a)$ in $g^{-1}[C]$, also mit $g(f(a)) \in C$, also genau die selben Elemente. Damit ist $f[(g \circ f)^{-1}[C]] = f[f^{-1}[g^{-1}[C]]]$ und folglich gilt ebenfalls $f[(g \circ f)^{-1}[C]] \subseteq g^{-1}[C]$.

\supseteq : Sei nun umgekehrt $y \in g^{-1}[C]$ bzw. $g(y) \in C$. Dann existiert ein $x \in X$, so dass $f(x) = y$ gilt, denn f ist ja surjektiv. Damit ist $g(f(x)) \in C$ und $g(f(x)) = (g \circ f)(x)$. Damit gilt $(g \circ f)(x) \in C$ bzw. $x \in (g \circ f)^{-1}[C]$ und folglich gilt $f(x) \in f[(g \circ f)^{-1}[C]]$. Wegen $f(x) = y$ haben wir $y \in f[(g \circ f)^{-1}[C]]$ und damit $f[(g \circ f)^{-1}[C]] \supseteq g^{-1}[C]$.

Somit haben wir $f[(g \circ f)^{-1}[C]] = g^{-1}[C]$ und damit ist g monoton.

Sei nun $B \subseteq Y$ zusammenhängend in (Y, S) , dann ist $g[B]$ zusammenhängend in (Z, T) , denn g ist stetig. Damit ist auch $(g \circ f)^{-1}[g[B]]$ zusammenhängend in (X, R) , denn $g \circ f$ ist monoton. Auch hier zeigen wir ganz ähnlich wie oben mit $(g \circ f)^{-1}[g[B]] = f^{-1}[B]$ die Monotonie von f :

\subseteq : Ist x Element von $(g \circ f)^{-1}[g[B]]$, dann bedeutet dies $(g \circ f)(x)$ ist Element von $g[B]$. Also ist $g(f(x))$ Element von $g[B]$ und damit gibt es ein $b \in B$ mit $g(f(x)) = g(b)$. Da g injektiv ist, gilt $f(x) = b$, also $f(x) \in B$ bzw. $x \in f^{-1}[B]$. Also haben wir die Teilmengenbeziehung $(g \circ f)^{-1}[g[B]] \subseteq f^{-1}[B]$.

\supseteq : Sei nun $x \in f^{-1}[B]$ beliebig gewählt, dann gilt $f(x) \in B$. Dann gilt $g(f(x))$ in $g[B]$, also $(g \circ f)(x) \in g[B]$ bzw. $x \in (g \circ f)^{-1}[g[B]]$. Also gilt $(g \circ f)^{-1}[g[B]] \supseteq f^{-1}[B]$.

Damit gilt die Gleichheit $(g \circ f)^{-1}[g[B]] = f^{-1}[B]$ und damit ist auch f monoton. \square

Für eine injektive Abbildung i ist das erste notwendige Kriterium, dass das Urbild $i^{-1}(p)$ eines Punktes p zusammenhängend sein muss, stets gegeben, denn dieses Urbild hat höchstens einen Punkt. Für den injektiven Fall wollen wir anstelle eines Kriteriums einen Algorithmus angeben, der entscheidet, ob eine Abbildung monoton ist. Wir gehen dabei ohne Einschränkung davon aus, dass es sich bei der Abbildung um die Inklusion $i : (A, R^+|_A) \hookrightarrow (X, R)$ von einem Teilraum von (X, R) handelt. Dazu wollen wir folgende naheliegende Definition vereinbaren:

Definition 4.36 (Monotone Menge). *Seien (X, R) ein topologischer Datentyp und $A \subseteq X$ Teilmenge von X . Dann bezeichnen wir A als monoton in (X, R) , wenn der Schnitt $B \cap A$ jeder zusammenhängenden Menge $B \subseteq X$ ebenfalls zusammenhängend in (X, R) ist.*

Diese Definition zeigt bereits, dass der nicht-surjektive Fall wohl wenig praktische Bedeutung haben wird. Man braucht sich nur zu überlegen, wie eine monotone geometrische Figur im \mathbb{R}^3 aussehen müsste.

Eine Menge A ist genau dann monoton, wenn die Inklusionsabbildung $i : (A, \mathcal{T}_R|_A) \hookrightarrow (X, \mathcal{T}_R)$ eine monotone Abbildung ist. Wir benötigen also nur noch ein Kriterium zur Entscheidung der Monotonie einer Teilmenge, da wir jede Abbildung in einen surjektiven Faktor und eben diese Inklusion faktorisieren können. Allerdings suchen wir hier noch nicht nach einem effizienten Algorithmus, denn wir werden im Falle von Detaildatenbanken ein strengeres Monotoniekriterium kennen lernen. Dieses wird zwar manche monotone Abbildungen nicht akzeptieren, kann dafür aber effizienter entschieden werden, was wir wiederum gerne akzeptieren wollen.

Algorithmus zur Entscheidung von Monotonie. Wir behaupten nun, dass folgender Algorithmus Monotonie einer endlichen Teilmenge A in einem endlichen topologischen Datentyp (X, R) entscheidet und mit polynomieller Laufzeit bezüglich der Größe der angegebenen Parameter A , X und R implementiert werden kann:

Algorithmus: monoton
 Eingabe: Menge A , Menge X , Relation R .
 Vorbedingung: X endlich, $A \subseteq X$ und $R \subseteq X \times X$.
 Ausgabe: *wahr*, wenn A monotone Teilmenge von (X, R) ist,
 ansonsten *falsch*.
 1: Berechne $coA := X \setminus A$.

- 2: Für jede Zusammenhangskomponente $[p]$ aus $(X, \mathcal{T}_R)|_{coA}$:
- 3: Berechne $BdStp := (Bd_R^*([p]) \cup St_R^*([p])) \cap A$.
- 4: Solange einen Punkt $p \in BdStp$ mit $BdStp \subseteq St_R^*(p)$ existiert:
- 5: Entferne p aus $BdStp$.
- 6: Weiter mit Zeile 4.
- 8: Wenn nun $BdStp$ die leere Menge ist:
- 9: Wähle nächste Zusammenhangskomponente $[p]$.
- 10: Weiter mit Zeile 3.
- 11: Wenn hingegen $BdStp$ nicht die leere Menge ist:
- 12: Halt mit Ergebnis *falsch*.
- 13: Halt mit Ergebnis *wahr*.

Wir erinnern an Definition 4.13, wonach $Bd_R^*(A) = St_{R^T}^*(A)$ die abgeschlossene Hülle von A ist, und diese ist der Stern um A im dualen Raum.

Eine ausführliche Darstellung dieses Algorithmus mit Korrektheitsbeweis und Überlegungen zur Rechenzeit (Komplexität) finden sich im Anhang. Wir werden uns erst dann um einen effizienten Algorithmus für dieses Problem bemühen, wenn wir ihn benötigen — es sei denn, diese Arbeit ist bereits vorher zu Ende.

4.5.5 Monotonie in Bauteildatenbanken

Nun wollen wir sehen, was uns der neu gewonnene Begriff der Monotonie gebracht hat. Dazu modifizieren wir unser altes einfaches Beispiel einer Wandvorlage so, dass unsere Forderung nach Monotonie erfüllt ist. Dann soll diese Beobachtung auf komplexere Situationen ausgedehnt werden.

Vergrößerung der Bauteildatenbank

Gehen wir also wieder zurück zu unserer Detaildatenbank. Wir hatten zum einen für einen Index $i \in I$ eine Familie von Details (D_i, R_i) als topologische Datenbank, die wir ursprünglich als disjunkte Summe

$$(D, R) = \sum_{i \in I} (D_i, R_i)$$

in einem einzigen Datentyp abspeichern wollten. Nun wissen wir, dass dieser Summenraum noch zu fein ist, und wir benötigen eine Vergrößerung von R , die angibt, wie ein Bauteil in (D_i, R_i) mit einem anderen Bauteil (D_j, R_j) verbunden werden kann. Wir erinnern uns, dass wir dabei großzügig sein dürfen, denn bei Anwendung von Details kann es nicht geschehen, dass im detaillierteren Plan irgend etwas verklebt wird, was vorher unverbunden war. Es ist auch klar, dass wir bei derartigen Verbindungen zwischen Bauteilen D_i und D_j das Paar der Indizes i, j in der Relation mit abspeichern müssen, und erhalten so (bis auf Anzahl und Bezeichnung der Schlüsselattribute) das Relationenschema von R der Form

$$R[a : D.id, ai : D.ref, b : D.id, bj : D.ref]$$

mit den entsprechend naheliegenden Fremdschlüsseln (referentiellen Integritäten) $(a, ai) \mapsto (D.id, D.ref)$ und $(b, bj) \mapsto (D.id, D.ref)$. Die Tupel der Form $(a, i, b, j) \in R$ definieren die Topologie des Bauteils D_i und die Tupel der Form

(a, i, b, j) mit $i \neq j$ legen fest, dass die Stelle b des Bauteils D_j das Bauteil D_i an der Stelle a berührt. Die Abbildung 4.7 zeigt anschaulich, was wir durch Monotonie als Konsistenzregel gewonnen haben. Dies ist ein Beispiel, in dem auf Monotonie einer surjektiven Abbildung zu testen ist, da jedes Element $e \in V'$ im detaillierten Plan $V' \bowtie B$ als Teilmenge (Faser) $\pi^{-1}(e) \subseteq V' \bowtie B$ auftaucht. Insbesondere müssen wir uns bisher noch nicht um einen effizienten Algorithmus für den nichtsurjektiven Fall bemühen.

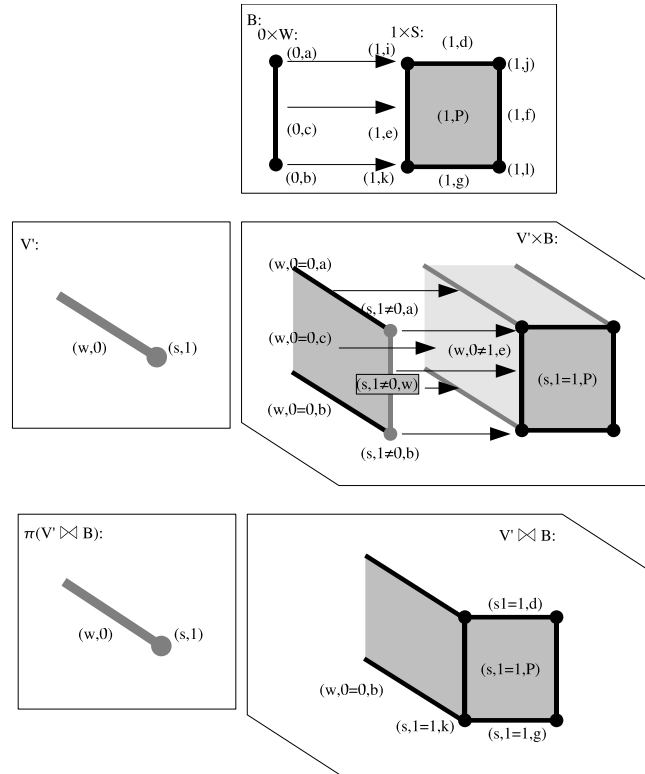


Abbildung 4.7: Der selbe Verbund topologischer Datentypen wie in Abbildung 4.6. Allerdings sind hier die erlaubten Bauteilverbindungen als Vergrößerung der Bauteilbibliothek angegeben (Pfeile in B nach rechts). Dann liefert $V' \bowtie B$ das gewünschte Ergebnis (bis auf einen Homöomorphismus, der den Knick zwischen Stütze und Wand beseitigt). Die Projektion auf die ersten Komponenten ist ebenfalls monoton.

Partielle Detaillierung

Es ist nun in der Praxis nicht so, dass ein Plan auf einen Schlag detailliert vorliegt. Üblicherweise werden in den ersten Planungsphasen nur grobe skizzenhafte Pläne erstellt und diese erst später weiter detailliert. Es kommt also vor, dass nur für einen Teil des Entwurfsplans die Details festliegen, für einen anderen Teil jedoch nicht.

Für diesen Fall gibt es nun im relationalen Datenmodell die sogenannten

null-Werte — eine Möglichkeit, die bisher nicht behandelt wurde. Wir erinnern uns, dass ein Datensatz oder *Tupel* eine Abbildung von einer Attributmenge in eine Wertemenge ist. Es ist nun auch möglich, für ein relationales Datenbankschema $\mathcal{R} : \mathcal{A} \rightarrow \mathcal{D}$ Tupel mit null-Werten zu erlauben. In einem derartigen Tupel gibt es Attribute $\mathbf{a} \in \mathcal{A}$, deren Wert $t(\mathbf{a})$ nicht immer definiert ist, etwa $t(\text{telefonnummer}) = \text{null}$, die Telefonnummer zu Datensatz t ist unbekannt. Das Symbol „null“ ist also eine alternative Schreibweise für \uparrow und das entsprechende Tupel t eine *partielle* Abbildung $t : \subseteq \mathcal{A} \rightarrow \bigcup \mathcal{D}$. Bei der Definition von Schemata, zu denen partielle Tupel erlaubt sind, werden üblicherweise diejenigen Attribute $\mathcal{N} \subseteq \mathcal{A}$ besonders ausgezeichnet, die stets einen Wert haben müssen. In SQL geschieht eine derartige Auszeichnung durch Hinzufügen von „not null“ an die Attributdeklaration.

Betrachten wir zum Beispiel folgendes Schema für eine Adressenliste:

```
create table Adressen(
    personnummer integer not null
    , strasse      char(32) not null
    , hausnummer  char(8)  not null
    , plz         char(8)  not null
    , ort         char(32) not null
    , telefonnummer char(32)
    , handynummer char(32)
    , primary key(personnummer)
    ...
);
```

Nach diesem Schema muss in jedem Datensatz die vollständige Adresse, nicht jedoch die Telefon- und Handynummer angegeben werden. Wir erinnern uns an die Rechenregeln, wie $a + \uparrow = a$, bei den partiellen Matrizen. Bei den partiellen *Tupeln* ist diese Regel hingegen nicht sinnvoll, da null in diesem Fall als „unbekannt“ interpretiert wird und somit $a + \text{null} = \text{null}$ eine sinnvollere Konvention ist: Wenn ich nicht weiß, welchen Wert b hat, dann kann ich auch nicht wissen, was der Wert von $a + b$ sein soll. Die partiellen Tupel sind also sinnvoll, wenn gewisse Informationen nicht immer bekannt sind.

Stellen wir uns nun einen idealisierten Ablauf einer Planung vor: Zunächst wird ein Konzept entwickelt, bei dem die detaillierte Ausarbeitung noch offen gelassen wird. Erst dann werden sukzessive die Detaillösungen festgelegt — eine Vorgehensweise, die in jeder Ingenieursdisziplin üblich ist. Insbesondere gehen auch Architekten in der Regel so vor.

Dazu benötigen wir eine Verallgemeinerung des relationalen Verbundes, die durch partielle Tupel nun ermöglicht ist (siehe auch: [Cod90, S.106ff]).

Definition 4.37 ((Links- , Rechts-) äußerer Verbund). Seien $\mathcal{L} : \mathcal{A} \rightarrow \mathcal{D}$ und $\mathcal{R} : \mathcal{B} \rightarrow \mathcal{D}$ Relationenschemata, für welche die gemeinsame Fortsetzung

$$\mathcal{L} \cup \mathcal{R} : \mathcal{A} \cup \mathcal{B} \rightarrow \mathcal{D}$$

existiert, für die also $\mathcal{L}|_{\mathcal{A} \cap \mathcal{B}} = \mathcal{R}|_{\mathcal{A} \cap \mathcal{B}}$ gilt. (Dies gilt insbesondere für $\mathcal{A} \cap \mathcal{B} = \emptyset$.) Für ein Tupel t zu \mathcal{L} definieren wir die partielle Fortsetzung t^\uparrow in $\mathcal{A} \cup \mathcal{B}$ als partielle Abbildung der Form

$$t^\uparrow : \subseteq \mathcal{A} \cup \mathcal{B} \rightarrow \bigcup \mathcal{D}, \mathbf{a} \mapsto \begin{cases} t(\mathbf{a}) & \text{falls } \mathbf{a} \in \mathcal{A} \\ \uparrow & \text{sonst.} \end{cases}$$

Für zwei Relationen L, R zu den Schemata \mathcal{L}, \mathcal{R} ist nun

$$L \bowtie_{\Theta} R := (L \bowtie_{\Theta} R) \cup \{t^{\uparrow} \mid t \in L \text{ und } t \notin \pi_{\mathcal{A}}(L \bowtie_{\Theta} R)\}$$

der links äußere Verbund von L mit R . Es heißt dann auch

$$L \bowtie_{\Theta} R := R \bowtie_{\Theta} L$$

rechts äußerer Verbund und

$$L \bowtie_{\Theta} R := (L \bowtie_{\Theta} R) \cup (L \bowtie_{\Theta} R)$$

heißt (symmetrischer) äußerer Verbund von L mit R .

Wir bilden also den üblichen Theta-Verbund und fügen alle Datensätze hinzu, die uns dabei „verloren gegangen“ sind. Allerdings werden wir sehen, dass ein äußerer Verbund für topologische Datenbanken in dieser Form nicht sinnvoll ist und wir eine für die Praxis äquivalente Alternative suchen müssen.

Wollen wir nun untersuchen, wie ein äußerer Verbund für unsere Detailbibliothek aussehen kann, und betrachten dazu einen Entwurf (E, R) und eine Bauteilbibliothek (B, S) mit Bauteilen $(B_i, S_i)_{i \in I}$ über einer Indexmenge I . Wir wissen bereits, dass dann S gröber ist als $\sum_{i \in I} S_i$.

Die Verwendung eines Bauteils (B_i, S_i) durch ein Entwurfselement $e \in E$ soll nun durch eine partielle Abbildung $b : \subseteq E \rightarrow I$ gegeben sein (etwa durch ein Attribut \mathbf{b}). Diese gibt an, dass ein Entwurfselement e durch ein Bauteil $b(e)$ realisiert sei. „Stetigkeit“ für b zu fordern ist uns nach bisherigem Kenntnisstand noch gar nicht möglich, da wir bisher nicht definiert haben, was wir unter einer stetigen partiellen Abbildung verstehen sollen. Um nun die topologischen Konstruktionen der äußeren Verbände zu definieren, benötigen wir zunächst einen sinnvollen Begriff von einer stetigen partiellen Abbildung:

Definition 4.38 (Stetigkeit bei partiellen Abbildungen). *Seien (X, R) und (Y, S) topologische Datentypen und sei $f : \subseteq X \rightarrow Y$ eine partielle Abbildung. Dann bezeichnen wir f als stetig, wenn gilt:*

$$(f \times f)[R^+] \subseteq S^*.$$

Dann schreiben wir $f : \subseteq (X, R) \rightarrow (Y, S)$.

Dabei setzen wir für $f(a) = \uparrow$ oder $f(b) = \uparrow$ sinnvollerweise auch $(f \times f)(a, b) = \uparrow$. Denn seien $f : \subseteq A \rightarrow X$ und $g : \subseteq B \rightarrow Y$ partielle Abbildungen, dann sind diese ja Relationen $f := (A, X, G_f)$ und $g := (B, Y, G_g)$ mit Graphen G_f und G_g . Dann ist $f \times g = (A \times B, X \times Y, G_f \otimes G_g)$. Wir erinnern uns, dass für ein $a \in A$ die Schreibweise $f(a) = \uparrow$ nicht bedeutet, dass $f(a)$ einen seltsamen mit „ \uparrow “ bezeichneten Wert hat, sondern dass kein $x \in X$ existiert, für das $(a, x) \in G_f$ gilt. Damit kann es auch kein $((a, x), (b, y)) \in G_f \times G_g$ geben und folglich auch nicht $((a, b), (x, y)) \in G_f \otimes G_g$, denn das direkte Produkt \otimes unterscheidet sich bei Relationen vom kartesischen Produkt \times nur durch Umstellen der Komponenten im Tupel. Also existiert kein $(x, y) \in X \times Y$ mit $(f \times f)(a, b) = (x, y)$ und somit $(f \times f)(a, b) = \uparrow$.

Es ist klar, dass die obige Definition ohne Verwendung der transitiven Hülle R^+ falsch wäre, denn $(f \times f)[R] \subseteq S^*$ ist bei den partiellen Abbildungen im

Gegensatz zu den totalen Abbildungen gar keine topologische Eigenschaft. Es lassen sich nämlich leicht ein topologischer Datentyp (Y, S) , zwei zueinander homöomorphe topologische Datentypen (X, R) und (X, R^+) und eine partielle Abbildung $b : \subseteq X \rightarrow Y$ konstruieren, so dass zwar $(f \times f)[R] \subseteq S^*$ aber auch $(f \times f)[R^+] \not\subseteq S^*$ gelten. Damit kann $(f \times f)[R] \subseteq S^*$ gar keine topologische Eigenschaft sein, da sie Homöomorphismen nicht respektiert. Ein derartiger Kandidat für X ist etwa die Menge $\{1, 2, 3\}$ mit der Nachfolgerrelation $R = \{(1, 2), (2, 3)\}$ und für Y die Teilmenge $\{1, 3\}$ mit der Relation $1S3$ und der naheliegenden partiellen Abbildung $i : \subseteq X \rightarrow Y$ definiert durch $i(1) = 1$, $i(2) = \uparrow$ und $i(3) = 3$.

Nun kennen wir von der Verklebung die Anheftungsabbildungen und damit bereits partielle Abbildungen mit einer Stetigkeitseigenschaft. Diese sind genau die partiellen stetigen Abbildungen:

Lemma 4.27 (Partielle stetige Abbildungen und Anheftungsabbildungen). *Sei $\{(X, R), (Y, S)\}$ eine topologische Datenbank und $f : \subseteq X \rightarrow Y$ eine partielle Abbildung mit Definitionsbereich $X_f = f^{-1}[Y]$ und (X_f, R_f) Teilraum von (X, R) . Dann ist f genau dann stetig, wenn die totale Abbildung $f|_{X_f} : X_f \rightarrow Y$ stetig von (X_f, R_f) nach (Y, S) ist.*

Damit hätten wir auch eine sinnvolle Definition für Stetigkeit von partiellen Abbildungen zwischen beliebigen topologischen Räumen und nicht nur den alexandroffschen Räumen und ihren topologischen Datentypen.

Beweis. Sei $\{(X, R), (Y, S)\}$ topologische Datenbank und $f : \subseteq X \rightarrow Y$ eine partielle Abbildung und (X_f, R_f) der Definitionsbereich X_f von f als Teilraum von (X, R) . Dann reicht es zu zeigen, dass $(f|_{X_f} \times f|_{X_f})[R^+|_{X_f}] = (f \times f)[R^+]$ gilt. Denn es ist bekanntermaßen $R_f^* = R^+|_{X_f}^*$, da $R^+|_{X_f}$ die Topologie des Teilraums erzeugt. Wenn also die obige Gleichheit gilt, dann sind die Aussagen $(f|_{X_f} \times f|_{X_f})[R_f] \subseteq S^*$ und $(f \times f)[R^+] \subseteq S^*$ äquivalent. Diese sind jedoch einerseits das Stetigkeitskriterium für die totale Abbildung $f|_{X_f}$ und andererseits für die partielle Abbildung f .

\subseteq : Sei $(x, y) \in (f|_{X_f} \times f|_{X_f})[R^+|_{X_f}]$ beliebig gewählt. Dann existiert ein (a, b) in $R^+|_{X_f}$ mit $(f|_{X_f} \times f|_{X_f})(a, b) = (x, y)$. Damit gilt einerseits $(a, b) \in R^+$ und andererseits sind $f(a) \neq \uparrow$ und $f(b) \neq \uparrow$. Folglich gelten $f(a) = f|_{X_f}(a)$ bzw. $f(b) = f|_{X_f}(b)$ und schließlich $(f \times f)(a, b) = (x, y)$. Wegen $(a, b) \in R^+$ also $(x, y) \in (f \times f)[R^+]$.

\supseteq : Sei nun $(x, y) \in (f \times f)[R^+]$. Dann existieren (a, b) in R^+ mit $(f \times f)(a, b) = (x, y)$. Doch dann sind zum Einen $a, b \in X_f$ und damit ist $(a, b) \in R^+|_{X_f}$ und zum Anderen gelten $f|_{X_f}(a) = f(a)$ und $f|_{X_f}(b) = f(b)$. Somit ist $(x, y) = (f|_{X_f} \times f|_{X_f})(a, b) \in (f|_{X_f} \times f|_{X_f})[R^+|_{X_f}]$.

Also gilt $(f|_{X_f} \times f|_{X_f})[R^+|_{X_f}] = (f \times f)[R^+]$. □

Mit den partiellen stetigen Abbildungen haben wir sogar Morphismen für eine Kategorie mit den topologischen Räumen als Objekten. Sind nämlich $f : \subseteq (X, \mathcal{T}_X) \rightarrow (Y, \mathcal{T}_Y)$ und $g : \subseteq (Y, \mathcal{T}_Y) \rightarrow (Z, \mathcal{T}_Z)$ partielle stetige Abbildungen, dann ist deren Komposition $g \circ f : \subseteq (X, \mathcal{T}_X) \rightarrow (Z, \mathcal{T}_Z)$ ebenfalls eine partielle stetige Abbildung. Dabei ist selbstverständlich $(g \circ f)(x) = \uparrow$ falls $f(x) = \uparrow$

gilt. Insbesondere sind damit die totalen stetigen Abbildungen auch besondere partielle stetige Abbildungen.

Zunächst vereinbaren wir wieder als Schreibweise für die Definitionsbereiche $X_f := f^{-1}[Y]$ und $Y_f := g^{-1}[Z]$ und betrachten diese jeweils als Teilräume von X bzw. Y . Sei nun $C \in \mathcal{T}_Z$ eine offene Menge, dann ist $B := g^{-1}[C]$ offen in Y_f . Dann gibt es eine offene Menge $U_B \in \mathcal{T}_Y$ mit $U_B \cap Y_f = B$. Da f ebenfalls stetig ist, ist $A := f^{-1}[U_B]$ offen in X_f . Damit existiert eine offene Menge $U_A \in \mathcal{T}_X$ mit $U_A \cap X_f = A$.

Sei nun $x \in U_A \cap (g \circ f)^{-1}[Z]$. Dann gilt $g(f(x)) \neq \uparrow$ und somit auch $f(x) \neq \uparrow$ und folglich $f(x) \in B$. Daraus folgt $g(f(x)) \in C$ oder auch $(g \circ f)(x) \in C$ bzw. $x \in (g \circ f)^{-1}[C]$.

Sei hingegen $x \in (g \circ f)^{-1}[C]$, dann ist $g(f(x)) \in C$ also $f(x) \in g^{-1}[C]$ bzw. $f(x) \in B$. Daraus folgt $f(x) \in U_B$ bzw. $x \in A$ und damit auch $x \in U_A$. Zudem ist $x \in (g \circ f)^{-1}[Z]$ und somit $x \in U_A \cap (g \circ f)^{-1}[Z]$.

Also ist $U_A \cap (g \circ f)^{-1}[Z] = (g \circ f)^{-1}[C]$ und somit offen in $(g \circ f)^{-1}[Z]$ als Teilraum von (X, \mathcal{T}_X) , denn U_A ist ja offen in (X, \mathcal{T}_X) .

Damit ist die Komposition von stetigen partiellen Abbildungen ebenfalls eine stetige partielle Abbildung. Die weiteren Eigenschaften von Morphismen wie Existenz des Einheitsmorphisms und Assoziativität der Komposition überlegt man sich leicht und werden hier nicht gezeigt. Nun werden wir sehen, wo die partiellen Abbildungen auftreten können:

Äußerer Verbund in topologischen Datenbanken Ganz in Analogie zu den bisher vorgestellten topologischen Konstruktionen wollen wir uns nun überlegen, wie die Topologie der äußeren Verbünde $L \bowtie R$, $L \bowtie_{\Theta} R$ und $L \bowtie_{\Theta} R$ definiert sein soll. Dazu werden wir zunächst versuchen, das Instrumentarium Initialtopologie und Finaltopologie auf die partiellen Abbildungen auszudehnen. Dies ist zwar bei einer Finaltopologie leicht möglich, doch die Topologie des Verbundes $L \bowtie_{\Theta} R$ ist die *Initialtopologie* bezüglich der Projektionsabbildungen $\pi_L : L \bowtie_{\Theta} R \rightarrow L$ und $\pi_R : L \bowtie_{\Theta} R \rightarrow R$.

Beim links äußeren Verbund $L \bowtie_{\Theta} R$ ist jedoch die (rechte) Projektion $\pi_R : \subseteq L \bowtie_{\Theta} R \rightarrow R$ eine partielle Abbildung und somit wäre die Topologie für $L \bowtie_{\Theta} R$ initial bezüglich einer partiellen Abbildung. Analoges gilt für die anderen äußeren Verbünde.

Wir versuchen nun, eine sinnvolle Verallgemeinerung des Begriffs Initialtopologie für die partiellen Abbildungen zu definieren, werden dann aber sehen, dass uns dies vor Probleme stellt. Insbesondere ist diese Topologie im Allgemeinen nicht eindeutig bestimmt. Unsere Initialtopologie muss nun folgende Eigenschaften haben:

- Sie muss die Initialtopologie der totalen Abbildungen als Sonderfall umfassen.
- Mit einer Initialtopologie muss eine partielle Abbildung natürlich stetig sein.
- Eine Initialtopologie soll weitgehend eindeutig sein — die Menge der Initialtopologien soll also möglichst gering sein.

Außer der letzten Forderung nach Minimalität erfüllt folgende Definition diese Bedingungen:

Definition 4.39 (Initiale Topologie partieller Abbildungen). Sei $f : \subseteq X \rightarrow (Y, \mathcal{T}_Y)$ eine partielle Abbildung von einer Menge X in einen topologischen Raum (Y, \mathcal{T}_Y) (also eine partielle Quelle) und $X_f := f^{-1}[Y]$ der Definitionsbereich von f . Dann bezeichnen wir eine Topologie \mathcal{I} für X als initial bezüglich f , falls sie die Initialtopologie einer totalen Fortsetzung $cf : X \rightarrow (Y, \mathcal{T}_Y)$ von f ist.

Eine totale Fortsetzung einer partiellen Abbildung $f : \subseteq X \rightarrow Y$ mit Definitionsbereich X_f ist eine totale Abbildung $cf : X \rightarrow Y$ mit $cf|_{X_f} = f|_{X_f}$. Die einzige totale Fortsetzung einer totalen Abbildung ist diese Abbildung selbst, und somit ist dies in der Tat eine Verallgemeinerung des Begriffs „Initialtopologie“ auf die partiellen Abbildungen.

Ist nun \mathcal{I} eine Initialtopologie bezüglich einer partiellen Abbildung f laut obiger Definition, und $A \subseteq X$ eine Teilmenge, dann ist $\mathcal{I}|_A$ eine Initialtopologie bezüglich $f|_A$. Liegt A auch im Definitionsbereich von f , also $A \subseteq X_f$, dann ist $\mathcal{I}|_A$ eindeutig bestimmt.

Beweis. Sei $f : \subseteq X \rightarrow (Y, \mathcal{T}_Y)$ eine partielle Abbildung mit einer Initialtopologie \mathcal{I} und $A \subseteq X$ eine Teilmenge des Vorbereichs von f . Dann ist \mathcal{I} die Initialtopologie einer totalen Fortsetzung $cf : X \rightarrow (Y, \mathcal{T}_Y)$. Dann ist $\mathcal{I}|_A$ die Initialtopologie von $cf|_A$ und diese Abbildung ist eine totale Fortsetzung von $f|_A$.

Dass die Initialtopologie für A bezüglich einer Komposition

$$A \xrightarrow{i} X \xrightarrow{f} (Y, \mathcal{T}_Y)$$

stets auch faktorweise berechnet werden kann, überlegt man sich leicht selbst. Zunächst nehme man die Initialtopologie \mathcal{T}_f bezüglich f und bilde damit die Abbildung $i : A \hookrightarrow (X, \mathcal{T}_f)$. Deren Initialtopologie ist dann die Initialtopologie der Komposition $f \circ i$. \square

Die Formulierung soll bereits andeuten, dass eine partielle Abbildung im Allgemeinen mehrere Initialtopologien hat. Nur ihre Einschränkung auf deren Definitionsbereich ist eindeutig bestimmt. Die Topologie eines äußeren Verbundes $(X, R) \underline{\bowtie} (Y, S)$ wäre damit also nur für $((X, R) \underline{\bowtie} (Y, S))|_{X \bowtie Y}$ eindeutig.

Wir wollen dies an einem sehr einfachen Beispiel demonstrieren: Betrachten wir eine Menge $X := Y := \{a, b, c, d\}$ mit vier Elementen und sei die Topologie des Raumes (Y, \mathcal{T}_Y) definiert durch eine Relation R mit $a R b R c R d$. Dann ist

$$\mathcal{T}_Y = \{\emptyset, \{a\}, \{a, b\}, \{a, b, c\}, \{a, b, c, d\}\}.$$

Unsere partielle Abbildung $f : \subseteq X \rightarrow (Y, \mathcal{T}_Y)$ soll definiert sein durch $f(a) = a$, $f(b) = b$ und Definitionsbereich $X_f = \{a, b\}$ (also $f(c) = \uparrow$ und $f(d) = \uparrow$). Dann ist die Initialtopologie für X_f bezüglich $f|_{X_f}$ $\mathcal{I}_f = \{\emptyset, \{a\}, \{a, b\}\}$. Nehmen wir zwei totale Fortsetzungen f_1 und f_2 von f mit

$$\begin{aligned} f_1|_{\{c, d\}} &: \{c, d\} \rightarrow Y, x \mapsto d \\ f_2|_{\{c, d\}} &: \{c, d\} \rightarrow Y, x \mapsto a \end{aligned}$$

Dann erhalten wir die folgenden zwei Initialtopologien \mathcal{T}_{f_1} und \mathcal{T}_{f_2} bezüglich f_1 und f_2 :

$$\begin{aligned} \mathcal{T}_{f_1} &:= \{\emptyset, \{a\}, \{a, b\}, \{a, b, c, d\}\} \\ \mathcal{T}_{f_2} &:= \{\emptyset, \{a, c, d\}, \{a, b, c, d\}\}. \end{aligned}$$

Und damit sind beide Topologien initial bezüglich f . Wir haben also mehr als eine Initialtopologie. Dies liegt auch nicht daran, dass wir auf die Minimalitätseigenschaft verzichtet haben, denn selbst die folgende echt größere Topologie

$$\mathcal{G} := \{\emptyset, \{a\}, \{a, b, c, d\}\} \subsetneq \mathcal{T}_{f_1}$$

ist eine minimale initiale Topologie für X bezüglich f und unterscheidet sich von der ebenfalls minimalen Topologie \mathcal{T}_{f_2} . Die Fortsetzung zur Erzeugung von \mathcal{G} schickt c und d nach b . Diese Topologien sind in der Tat minimal, denn jede weitere echte Vergrößerung von \mathcal{G} und \mathcal{T}_{f_2} würde die triviale Topologie für X ergeben. Deren Einschränkung auf $X_f = \{a, b\}$ ist aber ebenfalls trivial, somit nicht die Initialtopologie von $f|_{X_f}$ und verletzt damit Lemma 4.27.

Betrachten wir nun eine weitere partielle Abbildung $g : \subseteq X \rightarrow (Y, \mathcal{T}_Y)$ mit $g(c) = c$ und $g(d) = d$ und Definitionsbereich $X_g = \{c, d\}$. Auch hier betrachten wir zwei Topologien $\mathcal{T}_{g,1}$ und $\mathcal{T}_{g,2}$ für X

$$\begin{aligned} \mathcal{T}_{g,1} &:= \{\emptyset, \{a, b\}, \{a, b, c\}, \{a, b, c, d\}\} \\ \mathcal{T}_{g,2} &:= \{\emptyset, \{a, b, c\}, \{a, b, c, d\}\} \end{aligned}$$

und diese sind initial bezüglich g mit jeweils den Fortsetzungen $g_1(a) = g_1(b) = a$ und $g_2(a) = g_2(b) = c$. Die beiden partiellen Abbildungen g und f haben eine gemeinsame Fortsetzung $g \cup f$ und diese ist die totale Abbildung id_X mit Initialtopologie \mathcal{T}_Y .

Mit den totalen Abbildungen war es nun möglich, die Initialtopologien der einzelnen Abbildungen einer Quelle jeweils gesondert zu berechnen und dann deren gemeinsame Verfeinerung zu bilden. Wenn wir nun aber die gemeinsamen Verfeinerungen unserer oben aufgelisteten Topologien berechnen, dann erhalten wir:

$$\begin{aligned} \mathcal{T}_{1,1} := \sup \{\mathcal{T}_{f,1}, \mathcal{T}_{g,1}\} &= \mathcal{T} \left(\begin{array}{ccc} a & & c \\ \downarrow & \nearrow & \downarrow \\ b & & d \end{array} \right) = \mathcal{T}_Z & \text{(OK)} \\ \mathcal{T}_{1,2} := \sup \{\mathcal{T}_{f,1}, \mathcal{T}_{g,2}\} &= \mathcal{T} \left(\begin{array}{ccc} a & & c \\ \downarrow & \nearrow & \downarrow \\ b & & d \end{array} \right) = \mathcal{T}_Z & \text{(OK)} \\ \mathcal{T}_{2,1} := \sup \{\mathcal{T}_{f,2}, \mathcal{T}_{g,1}\} &= \mathcal{T} \left(\begin{array}{ccc} a & \longrightarrow & c \\ \downarrow & & \downarrow \\ b & & d \end{array} \right) \supsetneq \mathcal{T}_Z & \text{(zu fein)} \\ \mathcal{T}_{2,2} := \sup \{\mathcal{T}_{f,2}, \mathcal{T}_{g,2}\} &= \mathcal{T} \left(\begin{array}{ccc} a & \longleftarrow & c \\ \downarrow & & \downarrow \\ b & & d \end{array} \right) & \text{(ungeeignet)} \end{aligned}$$

Dabei gibt die viereckige Anordnung der a, b, c, d mit den Pfeilen eine die jeweilige Topologie erzeugende Relation an.

Die Topologien $\mathcal{T}_{1,1}$ und $\mathcal{T}_{1,2}$ ergeben die Initialtopologie \mathcal{T}_Z bezüglich $g \cup f$. Hingegen ist $\mathcal{T}_{2,1}$ echt feiner als \mathcal{T}_Z , und $\mathcal{T}_{2,2}$ ist mit \mathcal{T}_Z gar nicht vergleichbar,

also weder gröber noch feiner als diese. Damit sind die partiellen Abbildungen g und f auch mit diesen Topologien stetig. Mit der Topologie $\mathcal{T}_{2,2}$ hingegen ist $g \cup f$ gar nicht stetig, denn $(g \cup f)^{-1}[\{a, b\}] = \{a, b\}$ ist das Urbild einer offenen Menge in (Y, \mathcal{T}_Y) und nicht Element von $\mathcal{T}_{2,2}$.

In jedem Fall aber war die Initialtopologie bezüglich $f \cup g$ eine gemeinsame Verfeinerung von initialen Topologien bezüglich f und g . Wenn wir uns dabei aber nur auf die größten initialen Topologien beschränken, dann ist dies im Allgemeinen nicht möglich, wie man leicht sieht, wenn man die wenigen partiellen Abbildungen der Form $i : \subseteq \{a, b\} \hookrightarrow \{a \rightarrow b\}$ daraufhin untersucht. Aus diesem Grunde verzichten wir auf eine derartige Minimalitätsbedingung, denn wir haben ja bereits gesehen, dass auch damit die initialen Topologien für partielle Abbildungen noch nicht eindeutig bestimmt sind.

Wir stellen also fest, dass uns bei den partiellen Abbildungen und somit auch beim äußeren Verbund die Initialtopologie, zumindest nach unserer obigen Definition, Probleme bereiten. Es wäre zwar noch denkbar, es mit einer anderen Definition zu versuchen, aber wir werden hier einen anderen Weg gehen und einfach gar keine Topologie für den äußeren Verbund definieren: Wir *verbieten* einfach diese Abfrage in den topologischen Datenbanken.

Alternativen für den äußeren Verbund Nun sind die äußeren Verbünde zu nützlich, um sie einfach ersatzlos zu streichen. Wir benötigen immer noch eine Lösung für den Zeitraum von der Entwurfsskizze bis zum fertigen Werkplan. Also brauchen wir eine Alternative. Die Verwendung von partiellen Abbildungen $f : \subseteq X \rightarrow Y$ lässt sich nun vermeiden, wenn wir mit unserer bisherigen Konvention brechen und erlauben, „ \uparrow “ als einen sonderbaren Wert von f zu betrachten. Wir setzen also eine partielle Abbildung $f : \subseteq X \rightarrow Y$ einfach zu $f^\uparrow : X \rightarrow Y \cup \{\uparrow\}$ fort. Ist \mathcal{T}_Y eine Topologie für Y , dann können wir diese ebenfalls zu einer Topologie $\mathcal{T}_{Y \cup \{\uparrow\}}$ fortsetzen. Durch diese Fortsetzung haben wir es in der Hand, die Initialtopologie für eine partielle Abbildung f eindeutig zu bestimmen. Durch $\mathcal{T}_{Y \cup \{\uparrow\}}$ wird eine Initialtopologie für X bezüglich $f : \subseteq X \rightarrow (Y, \mathcal{T}_Y)$ eindeutig als die Initialtopologie für X bezüglich der totalen Abbildung

$$f^\uparrow : X \rightarrow (Y \cup \{\uparrow\}, \mathcal{T}_{Y \cup \{\uparrow\}})$$

festgelegt. Diese Überlegung lässt sich auch weiter ausbauen, indem wir sogar mehrere verschiedene „ \uparrow “ zulassen. Diese Vorgabewerte für Elemente von X bilden dann einen Teilraum \underline{Y}_\uparrow von (Y, \mathcal{T}_Y) .

Beispiel 4.10 (Detaildatenbank ohne äußeren Verbund). *Angenommen, es sei $p : (Y, S) \rightarrow I$ eine (topologische) Detaildatenbank. Das heißt (Y, S) ist die Menge der Elemente, aus denen sich Details zusammensetzen, I eine Indexmenge der Details und p eine surjektive Abbildung, die für jedes Element $y \in Y$ angibt, zu welchem Detail $p(y)$ es gehört. „Abbildung“ meint dann in der Regel ein Fremdschlüsselattribut. Es sei \mathcal{T}_p die Finaltopologie für I bezüglich p . Sei zudem (E, T) der topologische Datentyp einer Entwurfszeichnung und $d : \subseteq (E, T) \rightarrow (I, \mathcal{T}_p)$ eine partielle stetige Abbildung, welche die Verwendung eines Details $d(e)$ für jedes Entwurfselement $e \in E$ festlegt, bzw. $d(e) = \uparrow$ falls eine solche Festlegung für e nicht getroffen wurde. Nehmen wir an, dass sich (E, T) aus Volumen-, Flächen-, Kanten- und Knotenelementen zusammensetzt und dass diese durch ein Attribut $\dim : \{0, 1, 2, 3\}$ (also eine Abbildung $\dim :$*

$E \rightarrow \{0, 1, 2, 3\}$ für jedes $e \in E$ gekennzeichnet ist. So ist $\dim(f) = 2$, falls f ein Flächenelement ist.

Die Fortsetzung \underline{Y}_{\uparrow} könnte etwa mit vier verschiedenen „ \uparrow “ in der Form

$$\left\{ \uparrow_{\text{Volumen}} \xrightarrow{R} \uparrow_{\text{Fläche}} \xrightarrow{R} \uparrow_{\text{Kante}} \xrightarrow{R} \uparrow_{\text{Knoten}} \right\}$$

in einer Detailbibliothek \underline{Y} liegen. Die Fortsetzung d^{\uparrow} ist dann definiert für ein $e \in E$ mit $d(e) = \uparrow$ durch

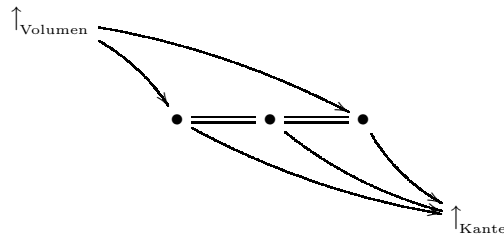
$$d^{\uparrow}(e) := \begin{cases} \uparrow_{\text{Volumen}} & : \dim(e) = 3 \\ \uparrow_{\text{Fläche}} & : \dim(e) = 2 \\ \uparrow_{\text{Kante}} & : \dim(e) = 1 \\ \uparrow_{\text{Knoten}} & : \dim(e) = 0. \end{cases}$$

Die Topologie $\mathcal{T}_{Y_{\uparrow}}$ legt dann auch fest, welches Element eines Details $\sigma_{p=i}(Y, S)$ für ein $i \in I$ das begrenzte Entwurfs-element berührt.

Betrachten wir etwa einen Wandaufbau mit zwei äußeren Ebenen und zwei durch eine innere Ebene verbundenen Schichten. Dann hat das Detail die Form



Die Grenzebenen sind durch „•“ und die Wandschichten durch „====“ gekennzeichnet. Dieses Detail würde dann als



in der erweiterten Detaildatenbank vorliegen. Eine Wand, die diesen Wandaufbau verwendet, darf also mit ihren beiden Außenflächen nur Volumen berühren und sie darf selbst nur von Kanten und Knoten berührt werden.

Damit haben wir einen vollwertigen Ersatz für partielle Fremdschlüssel, indem wir einen Default-Wert für Fremdschlüsselattribute festlegen, der dann ein (oder mehrere) Default-Objekt(e) eines anderen topologischen Datentyps referenziert. Da null als Wert von Schlüsselattributen in SQL-Datenbanken nicht zulässig ist, müssen wir dazu natürlich einen beliebigen anderen festen Wert (z.B. 0) zur Identifizierung von derartigen Vorgabelementen reservieren. Diese Vorgehensweise ersetzt den rechts äußeren Verbund bei den $n : 1$ -Beziehungstypen, wobei der Entitätstyp mit Vielfachheit 1 rechts sei.

Soll bei $n : m$ -Beziehungstypen, etwa A und B , ein (links-) äußerer Verbund $A \bowtie_{\Theta} B$ berechnet werden, dann ist dazu für B mindestens ein Default-Datensatz $\uparrow_B \in B$ vorzusehen. Jedes $a \in A$ muss dann zumindest diesen Datensatz \uparrow_B referenzieren. Dabei heißt „ a referenziert b “ nichts anderes als dass $a \Theta b$ gilt. Entsprechendes gilt für den symmetrischen äußeren Verbund, für den zusätzlich noch mindestens ein $\uparrow_A \in A$ vorzusehen ist.

Stetigkeit und Monotonie in Detaildatenbanken

Wir hatten in unserem Beispiel zu partiellen Fremdschlüsseln gefordert, dass die Abbildung $d : (E, T) \rightarrow (I, \mathcal{T}_p)$ von den Entwurfselementen E zum Index I der konstruktiven Details (Y, S) stetig sein soll. Dies stellt jedoch noch nicht sicher, dass die Projektion

$$\underline{\pi}_E : (E, T) \underline{\bowtie}_{d=p} (Y, S) \rightarrow (E, T)$$

monoton ist, was wir ja als Konsistenzregel für detaillierte Pläne gefordert hatten. Wir wollen nun ein hinreichendes Kriterium für Monotonie in derartigen Detaildatenbanken liefern. Insbesondere kann dieses Kriterium effizient überprüft werden. Bisher kennen wir ja zum Test auf Monotonie nur einen relativ schlechten, weil langsamen Algorithmus. Dabei nehmen wir in Kauf, dass es auch Fälle von Monotonie geben kann, die gegen dieses Kriterium verstoßen. Dieses ist also keine notwendige Bedingung von Monotonie.

Formulieren wir also unsere Behauptung:

Satz 4.28 (Monotonie in Detaildatenbanken). *Sei (Y, S) ein topologischer Datentyp und $p : (Y, S) \rightarrow I$ eine surjektive Abbildung mit finaler Relation F_p . Sei zudem (E, T) ein topologischer Datentyp und $d : (E, T) \rightarrow (I, F_p)$ eine stetige Abbildung. Dann ist $(E, T) \underline{\bowtie}_{d=p} (Y, S)$ die detaillierte Version von (E, T) , in der jedes $e \in E$ durch das Detail $\underline{\sigma}_{p=d(e)}(Y, S) = (Y, S)|_{p^{-1}(d(e))}$ ersetzt ist. Dann ist die Projektionsabbildung*

$$\underline{\pi}_E : (E, T) \underline{\bowtie}_{d=p} (Y, S) \rightarrow (E, T)$$

monoton, falls $(p \times p)[S^+]$ transitiv und $p^{-1}[i]$ für jedes $i \in I$ zusammenhängend ist.

Um die Intention dieses Satzes zu illustrieren, wollen wir Bezeichnungen aus dem Bauwesen auf die obigen Konstruktionen anwenden. Bezeichnen wir also (E, T) als „Entwurf“, und dessen Elemente $e \in E$ entsprechend als „Entwurfselemente“. Die Abbildung $p : (Y, S) \rightarrow I$ heiße „Detailbibliothek“, wobei $i \in I$ die „Bezeichnung“ eines Details sei, und dessen Urbild $Y_i := p^{-1}(i) = \sigma_{p=i}(Y)$ als Teilraum $(Y_i, S_i) := (Y, S)|_{Y_i}$ von (Y, S) sei das „Detail“ i . Ein Detail muss also zusammenhängend sein. In der Praxis kommen Teilräume meist als Plan-ausschnitte oder als Schnittzeichnungen vor. Wir könnten hier also anstelle von „Teilraum“ das Wort „Ausschnitt“ verwenden. Der Raum $(E, T) \underline{\bowtie}_{d=p} (Y, S)$ ist dann die „Detaillierung“ von (E, T) . Die Elemente in der Detaillierung schreiben wir als Paare (e, y) mit Entwurfselement e und Detailelement $y \in Y$. Die Detaillierung selbst besteht also aus der Menge aller derartigen Paare (e, y) mit $d(e) = p(y)$ zusammen mit der Initialtopologie der Projektionen $\pi_E(e, y) := e$ und $\pi_Y(e, y) = y$.

Beweis. Zunächst beobachten wir, dass $\underline{\pi}_E$ surjektiv ist. Ist nämlich ein Entwurfselement $e \in (E, T)$ beliebig gewählt, dann verwendet e ein Detail mit Bezeichnung $i = d(e) \in I$. Da p surjektiv ist, existiert zur Detailbezeichnung i das nichtleere Detail (Y_i, S_i) als Ausschnitt (Teilraum) von (Y, S) . Damit gibt es ein $y \in Y$ mit $d(e) = p(y)$ und damit ist das Paar (e, y) Element von $E \bowtie_{d=p} Y$ und ist somit auch Punkt in der Detaillierung $(E, T) \underline{\bowtie}_{d=p} (Y, S)$. Für dieses (e, y) ist $\underline{\pi}_E(e, y) = e$ und damit ist e Bildpunkt von mindestens einem Element in

$E \times_{d=p} Y$ und damit ist π_E surjektiv. Wir können also das Monotoniekriterium für surjektive Abbildungen anwenden.

Nun zeigen wir noch, dass p unter der oben genannten Voraussetzung monoton ist: Die erste Monotoniebedingung, dass das Urbild eines Punktes (Elements) zusammenhängend sein muss, haben wir ja ausdrücklich vorausgesetzt. Sei nun $(i, j) \in (p \times p)[S]^+$. Zu zeigen ist, dass dann $p^{-1}[\{i, j\}]$ zusammenhängend ist. Zunächst ist $(i, j) \in (p \times p)[S^+]^+$, da wegen $S \subseteq S^+$ auch $(p \times p)[S] \subseteq (p \times p)[S^+]$ und somit auch $(p \times p)[S]^+ \subseteq (p \times p)[S^+]^+$ gilt. Da jedoch $(p \times p)[S^+]$ als transitiv vorausgesetzt wurde, gilt $(p \times p)[S^+] = (p \times p)[S^+]^+$ und folglich $(i, j) \in (p \times p)[S^+]$. Dann existiert jedoch ein $(y_i, y_j) \in S^+$ mit $p(y_i) = i$ und $p(y_j) = j$. Diese Elemente y_i, y_j sind damit auch in $p^{-1}[\{i, j\}]$ enthalten und es ist (y_i, y_j) ein Pfad von y_i nach y_j in $p^{-1}[\{i, j\}]$. Seien nun beliebige Punkte y_1 und y_2 Elemente von $p^{-1}[\{i, j\}]$. Ist nun $p(y_1) = i$, dann existiert ein Weg von y_1 nach y_i in $p^{-1}(i)$, ansonsten existiert ein Weg von y_1 nach y_j in $p^{-1}(j)$, da jedes Urbild eines Punktes durch Wege zusammenhängend ist. Die Vereinigung dieser beiden Urbilder $p^{-1}(i) \cup p^{-1}(j)$ ist das Urbild der beiden Punkte und damit existiert ein Weg in $p^{-1}[\{i, j\}]$ von y_1 nach y_i oder nach y_j . In jedem Fall existiert auch ein Weg von y_i nach y_j und somit ist y_1 durch einen Weg mit y_j in $p^{-1}[\{i, j\}]$ verbunden. Das selbe gilt für y_2 und damit existiert auch ein Weg von y_1 nach y_2 in $p^{-1}[\{i, j\}]$. Da y_1 und y_2 beliebig gewählt wurden, sind alle Punkte in $p^{-1}[\{i, j\}]$ durch Wege verbunden und damit ist diese Menge zusammenhängend und folglich ist p monoton.

Es ist für jedes $e \in (E, T)$ dessen Urbild $\pi_E^{-1}(e)$ als Teilraum (Ausschnitt) der Detaillierung homöomorph zum Detail (Y_i, S_i) mit $i = d(e)$, was man sich leicht überlegt. Jedes derartige Detail ist aber laut der Erkenntnis, dass p monoton ist, zusammenhängend, und damit ist dies auch $\pi_E^{-1}(e)$.

Sei nun $a T^+ b$ ein Paar der Relation T^+ . Dann ist zu zeigen, dass der detaillierte Ausschnitt $\pi_E^{-1}[\{a, b\}]$ zusammenhängend in $(E, T) \times_{p=e} (Y, S)$ ist. Wenn wir nun aus diesem Ausschnitt ein (e_1, y_1) und ein (e_2, y_2) beliebig auswählen, dann ist jedes Entwurfselemente e_1 und e_2 gleich a oder gleich b , denn es gilt $\pi_E(e_1, y_1) \in \{a, b\}$ und somit $e_1 \in \{a, b\}$. Selbiges gilt für e_2 , also insgesamt $\{e_1, e_2\} \subseteq \{a, b\}$. Insbesondere gilt dann $e_1 T^* e_2$ oder $e_2 T^* e_1$ und damit ist auch für jedes $y \in Y$ die Menge $\{(e_1, y), (e_2, y)\}$ zusammenhängend in $(E, T) \times (Y, S)$, denn es ist $(e_1, y) T^* \otimes \Delta_Y (e_2, y)$ oder $(e_2, y) T^* \otimes \Delta_Y (e_1, y)$. Wir wollen nun annehmen, es gelte $e_1 T^* e_2$, denn ansonsten vertauschen wir in der nun folgenden Argumentation e_1 und e_2 . Wir wollen nun zeigen, dass stets ein Weg von (e_1, y_1) nach (e_2, y_2) existiert, womit dann der Zusammenhang von $\pi_E^{-1}[\{a, b\}]$ und damit die Monotonie von π_E gezeigt wäre.

Ist nun $e_1 = e_2$, dann ist $\pi_E(e_1, y_1) = \pi_E(e_2, y_2)$ und somit sind (e_1, y_1) und (e_2, y_2) Elemente von $\pi_E^{-1}(e_1)$, das ja zusammenhängend ist, wie wir oben gesehen haben. Damit existiert ein Weg von (e_1, y_1) nach $(e_1, y_2) = (e_2, y_2)$, der die erste Komponente e_1 konstant hält. Wegen $\pi^{-1}(e_1) \subseteq \pi^{-1}[\{a, b\}]$ liegt dieser Weg auch in $\pi^{-1}[\{a, b\}]$.

Ist hingegen $e_1 \neq e_2$, dann gilt ja $(d \times d)[T] \subseteq (p \times p)[S]^*$, denn die Relation $(p \times p)[S]$ erzeugt die Finaltopologie bezüglich p und wir hatten die Stetigkeit von $d : (E, T) \rightarrow (I, (p \times p)[S])$ vorausgesetzt. Damit ist $(d(e_1), d(e_2)) \in (p \times p)[S]^*$.

Ist hier $d(e_1) = d(e_2)$, verwendet also e_1 das selbe Detail wie e_2 , dann ist (e_1, y_2) ein Element des detaillierten Entwurfsobjekts e_1 und somit existiert ein Pfad in $\pi_E^{-1}(e_1)$ von (e_1, y_1) nach (e_1, y_2) , denn das detaillierte Objekt ist ja

zusammenhängend. Zudem ist $((e_1, y_2), (e_2, y_2))$ ein Pfad, denn es gilt $e_1 T^* e_2$ und somit gilt auch $(e_1, y_2) (T^* \otimes \Delta_Y) (e_2, y_2)$. Damit existiert ebenfalls ein Pfad von (e_1, y_1) nach (e_2, y_2) in $\{(e_1, y_1), (e_2, y_2)\}$.

Ist aber $d(e_1) \neq d(e_2)$, dann verwenden die beiden Entwurfselemente verschiedene Details. Sei (Y_1, S_1) das Detail, welches von e_1 verwendet wird, und (Y_2, S_2) entsprechend das Detail von e_2 . Dann ist

$$\pi_E^{-1}[\{e_1, e_2\}] = \{e_1\} \times Y_1 + \{e_2\} \times Y_2.$$

Da d stetig ist, gilt $d(e_1) F_p^* d(e_2)$, denn es gilt ja $e_1 T^* e_2$. Wegen $d(e_1) \neq d(e_2)$ gilt $d(e_1) F_p^+ d(e_2)$. Damit existiert ein z_1 im Detail Y_1 und ein z_2 in Y_2 mit $p(z_1) = d(e_1)$ und $p(z_2) = d(e_2)$, und damit ist $p(z_1) F_p^* p(z_2)$. Da $(p \times p)[S]$ die Finaltopologie der Abbildung $p : (Y, S) \rightarrow I$ erzeugt und S die selbe Topologie wie S^+ für Y erzeugt, gilt stets $(p \times p)[S]^* = (p \times p)[S^+]^*$. Damit ist $z_1 S^* z_2$ und wegen $z_1 \neq z_2$ ist $z_1 S^+ z_2$. Nun ist also $(e_1, z_1) (T \otimes \Delta_Y) (e_2, z_1) (\Delta_E \otimes S) (e_2, z_2)$ und somit ist das Paar $((e_1, z_1), (e_2, z_2))$ in der transitiven Hülle der Relation des Produktraums und somit ein Pfad von (e_1, z_1) nach (e_2, z_2) in $\pi_E^{-1}[\{a, b\}]$. Damit haben wir auch hier einen Weg von (e_1, y_1) über (e_1, z_1) und (e_2, z_2) nach (e_2, y_2) und somit sind (e_1, y_1) und (e_2, y_2) durch einen Weg in $\pi_E^{-1}[\{a, b\}]$ verbunden.

In jedem Fall haben wir also einen Weg von einem beliebigen Punkt (e_1, y_1) zu einem beliebigen anderen Punkt (e_2, y_2) innerhalb des Urbildes $\pi_E^{-1}[\{a, b\}]$ zweier miteinander verbundener Entwurfselemente $a T^* b$.

Damit ist π_E eine monotone Abbildung. □

Die Eigenschaft, dass für die Abbildung $p : (Y, S) \rightarrow I$ die Relation $(p \times p)[S^+]$ transitiv ist, könnte z.B. dadurch sichergestellt werden, dass für jeden gerichteten Pfad von Detailbezeichnern (i_n, \dots, i_0) in $(I, (p \times p)[S])$ auch ein entsprechender Pfad (y_n, \dots, y_0) mit $p(y_j) = i_j$ von Detailelementen in (Y, S) gefordert wird. Angenommen es gibt z.B. einen Pfad

Außenraum \rightarrow Fassade \rightarrow Ecklösung \rightarrow Traufanschluss

in den Detailbezeichnern. Dann kann festgelegt werden, dass ein Punkt p_A des Außenraum-„details“, ein Punkt p_F des Fassadendetails, ein Punkt p_E der Ecklösung und ein Punkt p_T in der Detaillösung des Traufanschlusses durch einen entsprechenden gerichteten Pfad

$$p_A \xrightarrow{S} p_F \xrightarrow{S} p_E \xrightarrow{S} p_T$$

in (Y, S) verbunden sind. In Abbildung 4.8 wird dies veranschaulicht.

Diese Bedingung ist relativ einfach zu überprüfen. Wir benötigen ohnehin eine relationale Algebra mit transitiver Hülle und mit dieser können wir dann auch S^+ berechnen. Dann haben wir sofort auch $(p \times p)[S^+]$ und können die Transitivität dieser Relation testen durch $(p \times p)[S^+]^2 \subseteq (p \times p)[S^+]$.

Zusammenfassung

Wir haben gesehen, dass Monotonie eine sinnvolle topologische Konsistenzregel für den Entwurf von Planungsinformationssystemen ist. Diese soll aber nur ein Beispiel für sinnvolle topologische Konsistenzregeln sein und ist sicher nicht

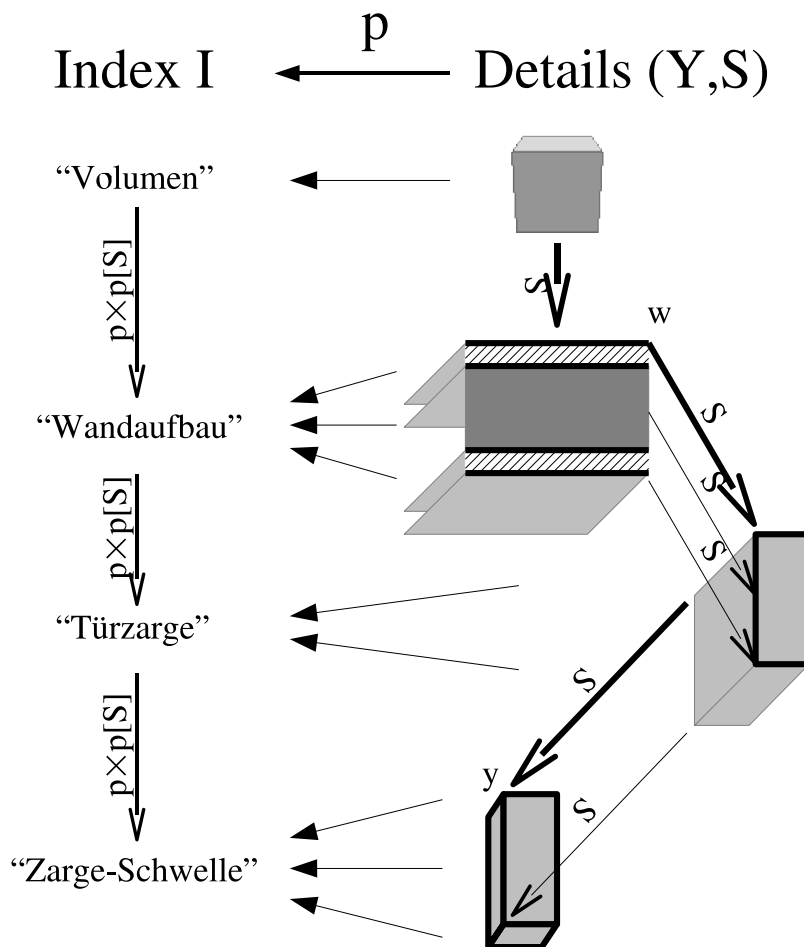


Abbildung 4.8: Ein hinreichendes Kriterium für die Monotonie von Detaillierungen. Die Detailbezeichnung „Zarge-Schwelle“ $\in I$ berührt die Detailbezeichnung „Wandaufbau“ $\in I$. Also muss mindestens ein Punkt im entsprechenden Detail auch in Berührung mit irgend einem Punkt des Wandaufbaus sein. In diesem Fall berührt die mit y gekennzeichnete linke obere Ecke des Zarge-Schwelle-Details entlang des fett gezeichneten gerichteten Pfades der Länge 2 die mit w gekennzeichnete äußere Wandebene des Wandaufbaus. Es gilt also $w \in S^2 y$.

die einzige. Es wird wahrscheinlich weitere praktisch relevante topologische Regeln geben, die nicht auf Monotonie zurückgeführt werden können, sondern eine andere topologische Eigenschaft betreffen. Monotonie ist also möglicherweise keine hinreichende, jedoch offensichtlich eine notwendige Konsistenzregel für den praktischen Datenbankentwurf. Insbesondere ist sie vermutlich noch nicht einmal hinreichend für einen praxistauglichen Entwurf von Detaildatenbanken.

Es müssen also noch weitere Konsistenzregeln gefunden werden. Wir haben hier nun gesehen, wie derart topologische Entwurfsregeln entwickelt werden können:

- Identifiziere ein praktisches Problem beim Entwurf einer Datenbank.
- Versuche, dieses Problem als topologische Eigenschaft zu formulieren.
- Stelle eventuell sicher, dass diese Eigenschaft nicht NP-schwer ist. Ansonsten besteht nämlich keine realistische Chance, dass für diese Eigenschaft ein effizienter Test überhaupt möglich ist.
- Entwickle dann einen Test, der diese Eigenschaft überprüft.

Zukünftige Arbeiten könnten so zu weiteren derartigen Regeln führen.

Kapitel 5

Geometrie

Topologie ist das was noch übrig bleibt, wenn man die geometrische Form beseitigt. Dieses Kapitel diskutiert, wie wir die Geometrie wieder zurückbekommen. Unseren topologischen Räumen soll nun also eine Form gegeben werden und ein Mausloch sich wieder von einer Türöffnung unterscheiden. Wir wollen hier einige geometrische Grundbegriffe wiederholen, mit denen in der Computergraphik gearbeitet wird. Die Beziehung der topologischen Datenbanken zur geometrischen Modellierung wird hier jedoch nur angedeutet und in dieser Arbeit nicht weiter vertieft.

5.1 Vektorräume und affine Räume

Erinnern wir uns an die Definition eines \mathbb{Z} -Moduls. In dieser Definition hatten wir die ganzen Zahlen \mathbb{Z} als Skalar verwendet. Wenn wir in dieser Definition nun die ganzen Zahlen \mathbb{Z} durch die reellen Zahlen \mathbb{R} ersetzen, dann erhalten wir den reellen Vektorraum. Die Begriffe Basis, Dimension und lineare Abbildung sind bei Vektorräumen genauso definiert wie bei den Moduln. Vektorräume unterscheiden sich von Moduln durch die zusätzliche Möglichkeit der Division eines Vektors durch einen (von Null verschiedenen) Skalar.

Ein Vektorraum hat einen besonders hervorgehobenen Bezugspunkt, den Nullvektor, der dem euklidischen Raum fehlt. Einen Raum, der einem Vektorraum ähnlich ist, dem aber sein Nullpunkt abhanden gekommen ist, bezeichnen wir als *affinen Raum*:

Definition 5.1 (affiner Raum). *Sei A eine Menge und V ein Vektorraum. Eine Abbildung $+ : V \times A \rightarrow A$, $(v, a) \mapsto v + a$, bezeichnet man als Aktion von V auf A , wenn $0 + a = a$ und $v + (w + a) = (v + w) + a$ für alle $a \in A$ und alle $v, w \in V$ gilt. Gibt es für je zwei Punkte $a, b \in A$ immer auch einen eindeutig bestimmten Vektor $v \in V$ mit $v + a = b$, dann heißt das Paar $(A, +)$ affiner Raum. Statt $v + a = b$ schreiben wir $\vec{ab} := v$ und vereinfachend statt $(A, +)$ einfach A . Der Vektorraum V heißt zu A assoziierter Vektorraum.*

Der Vektorraum definiert also die Verschiebung der Punkte und umgekehrt haben wir für zwei Punkte a, b den eindeutigen Verschiebungsvektor \vec{ab} .

Jeder Vektorraum V bildet mit $(V, +)$ einen affinen Raum und ist zu sich selbst assoziiert, denn mit der Vektoraddition $+ : V \times V \rightarrow V$ operiert V auf sich

selbst. Umgekehrt können wir aus jedem affinen Raum A einen Vektorraum A' erzeugen. Dazu legen wir einen Punkt $a \in A$ fest und definieren Vektoraddition und Skalarmultiplikation für A wie folgt: Addition ist $x + y := \vec{ax} + \vec{ay} + a$ und Skalarmultiplikation $r \cdot x := (r \cdot \vec{ax}) + a$. Der Punkt a ist dann der Nullpunkt dieses Vektorraums und bestimmt auch das Ergebnis der Addition $x + y$ und der Skalarmultiplikation $r \cdot x$. Damit ist A' zu A assoziiert.

Definition 5.2 (Dimension). *Ist A ein affiner Raum mit assoziiertem Vektorraum V , dann ist $\dim A$, die Dimension von A , gleich $\dim V$, der Dimension von V .*

Der dreidimensionale reelle Vektorraum ist unser Modell des uns umgebenden Raums und der vierdimensionale reelle Vektorraum eignet sich als Modell für die Raumzeit.

5.2 Darstellung geometrischer Körper

Die Verwendung von Topologie haben wir durch die Annahme gerechtfertigt, dass Architektur eine Zerlegung des euklidischen Raums ist. Damit muss die geometrische Beschreibung eines Elements mit der topologischen Beschreibung übereinstimmen. Dazu muss die Vereinigung der geometrischen Realisierung von Elementen einer offenen Menge in unserer Datenbank auch offen im affinen Raum sein. Des Weiteren sollen die Geometrien verschiedener Elemente der Datenbank disjunkt sein, sich also grundsätzlich nicht schneiden.

Da eine Zelle e aus einer Zerlegung des euklidischen Raums E^n hervorgegangen ist, ist ihre geometrische Realisierung $|e|$ (umgangssprachlich: ihre „Geometrie“) einfach die Menge der Punkte, die zu e gehören. Eine Beschreibung der geometrischen Realisierung $|e|$ ist ein Prädikat $p_e(x)$, so dass $|e| = \{x \in E^n \mid p_e(x)\}$. Wir müssen zudem fordern, dass $p_e(x)$ als endliche Formel darstellbar ist, die von einem Rechner verstanden werden kann. Allerdings sollen hier noch weitere Einschränkungen vereinbart werden, die sich nicht selbstverständlich aus den Definitionen der topologischen Datenbanken ergeben.

5.2.1 Punkte

0-Zellen eines topologischen Datentyps sind Eckpunkte, also Elemente P , die von keinem anderen Element berührt werden, für die also gilt: $\text{cl}\{P\} = \{P\}$. Ist der Datentyp ein relationaler Komplex, dann sind es zudem die Elemente mit Dimensionsattribut 0. Wir wollen zunächst vereinbaren, dass P einem einzelnen Punkt q im affinen Raum entspricht, beschrieben durch seine Koordinaten $q = (q_1, q_2, q_3)$ (statt $\{1, 2, 3\}$ geht natürlich auch $\{X, Y, Z\}$ als Indexmenge). Die Forderung, dass verschiedene Elemente sich nicht schneiden, heißt im Fall von Punkten, dass bei verschiedenen 0-Zellen $q \neq t$, die jeweils einen Punkt im affinen Raum darstellen, sich mindestens eine der Koordinaten (q_1, q_2, q_3) von der entsprechenden (t_1, t_2, t_3) unterscheidet. Die Koordinatenattribute bilden also einen Schlüssel für die Punkte.

5.2.2 Linien

Die geometrische Realisierung einer 1-Zelle L ist durch einen Weg $w : [0, 1] \rightarrow \mathbb{R}^3$ beschrieben, wobei nur das Innere des Weges, also $w|_{]0,1[}$, zur 1-Zelle gehört und

die Randpunkte 0 und 1 des Intervalls $[0, 1]$ auf die geometrische Realisierung der Randzellen von L abgebildet werden. Ist also $\{a, b\} = \text{fr}\{L\}$, dann ist $|a| \cup |b| = w(\{0, 1\})$.

5.2.3 Flächen

Eine 2-Zelle F lässt sich ganz analog durch eine stetige Abbildung $f : [0, 1] \times [0, 1] \rightarrow \mathbb{R}^n$ geometrisch darstellen. Dabei dient ebenfalls nur die Einschränkung auf das Innere der Zelle $f|_{]0,1[\times]0,1[}$ der geometrischen Beschreibung der Fläche und die geometrische Realisierung $|r|$ jeder Randzelle $r \in \text{fr}\{v\}$ muss im Rand $f(\{0, 1\} \times [0, 1] \cup [0, 1] \times \{0, 1\})$ liegen.

5.2.4 Volumen

Die 3-Zellen im \mathbb{R}^3 benötigen an sich keine eigene geometrische Beschreibung. Es ist zwar möglich, auch für ein Volumenobjekt V eine stetige Abbildung $v : [0, 1]^3 \rightarrow \mathbb{R}^3$ heranzuziehen, so dass $v(]0, 1[^3)$ die geometrische Realisierung $|V|$ ergibt, und $v \circ \text{fr}[0, 1]^3$ die des Randes. Umgekehrt ist jedoch $|V|$ bereits aus der geometrischen Beschreibung des Randes von V im wesentlichen festgelegt. Wir müssen nur noch angeben an welcher Seite des Randes $|V|$ liegen soll. Volumenmodellierer, die die sogenannte Randdarstellung (Boundary Representation BREP) verwenden, funktionieren auf diese Weise. Ein Beispiel ist das `topology_schema` der ISO 10303, eine genormte Randdarstellung.

5.3 Zusammenfassung

Haben wir also eine n -Zelle c im k -dimensionalen reellen Vektorraum \mathbb{R}^k , dann ist deren geometrische Realisierung $|c| = f_c(]0, 1[^n)$ durch eine stetige Abbildung $f_c : [0, 1]^n \rightarrow \mathbb{R}^k$ gegeben, deren Einschränkung $f_c|_{]0, 1[^n} :]0, 1[^n \rightarrow |c|$ ein Homöomorphismus ist. Für den Rand muss zudem gelten $|\text{fr}\{c\}| = \text{fr}|c| = f_c[\text{fr}[0, 1]^n]$. Der Rand der geometrischen Realisierung muss also gleich der geometrischen Realisierung des Randes sein. Zudem muss für die geometrische Realisierung des gesamten Raums stets die Dimension k des umgebenden Raums festgelegt werden. Des Weiteren kann die Zeit des Bauablaufs und die Abfolge der Planungsversionen ebenfalls als je eine weitere Dimension aufgefasst werden, so dass sich insgesamt ein fünfdimensionaler Raum ergibt. Diese weiteren Dimensionen erfordern aber keine höherdimensional verallgemeinerte Computergraphik. Eine 3D-Graphikmodellierung unter Vernachlässigung der zwei zusätzlichen Zeitattribute müsste ausreichend sein.

Kapitel 6

Anwendung

Die Beispiele im Theorieteil dieser Arbeit sollten bereits veranschaulichen, wie topologische Datenbanken angewendet werden können. Wir wollen nun beschreiben, wie diese Theorie eingesetzt werden kann, um ein Planungssystem aufzubauen. Zunächst soll dazu der Begriff der Planung bestimmt werden. Dann werden wir uns am idealisierten Planungsablauf entsprechend der Leistungsphasen der Honorarordnung für Architekten und Ingenieure (HOAI) orientieren und für einige dieser Phasen die Anforderungen für ein Planungssystem sammeln.

Danach erstellen wir ein entsprechendes Modell, in dem nur diese praktischen Anforderungen erfasst werden. Aspekte der technischen Realisierung als Softwaresystem werden dabei zunächst ignoriert. Ein solches Modell wird als *konzeptuelles Schema* oder als *Domänenmodell* bezeichnet.

Wir werden dieses Modell als sogenannte *Klassendiagramme* in der graphischen Notation der Unified Modeling Language (UML) darstellen. Diese Notation ist standardisiert und kann als Analogie zur „Planzeichenverordnung“ für die Softwareentwicklung angesehen werden. Jedes neu eingeführte Notationselement wird dabei jeweils kurz erläutert. Einen Überblick über diese Sprache bietet [Bal01]. Unter den zahlreichen ausführlichen Darstellungen gibt es etwa [FS98] und natürlich die im Internet frei verfügbaren Originaldokumente auf <http://www.uml.org/>.

6.1 Planungsbegriff

Eine Aufgabe ist allgemein die Definition eines Ziels, das durch geeignetes Handeln erreichbar ist. Bei einer Planungsaufgabe besteht dieses Handeln im Planen. Wir wollen hier den Planungsbegriff aus der handlungsorientierten Arbeitspsychologie zugrunde legen.

Definition 6.1 (Handlung). *Eine Handlung ist eine elementare, zielgerichtete, rationale Folge von Aktionen einer Person, also eine Aktivität unter folgenden Voraussetzungen:*

- *Der Handelnde kennt das zu erreichende Ziel vor Beginn der Aktivität.*
- *Der Handelnde ist in der Lage, seine Aktivität so zu steuern, dass das Ziel tatsächlich erreicht wird, und steuert sie auch entsprechend.*

- *Der Handelnde kann erkennen, wenn das Ziel erreicht wurde, und stellt dann die entsprechende Aktivität ein.*

Zur Steuerung seiner Aktivität muss der Handelnde eine Vorstellung davon haben, welche Aktionen tatsächlich zum Ziel führen. Diese Vorstellung umfasst die ihn umgebende Realität, die möglichen Aktionen und deren zu erwartende Wirkungen und wird in der Arbeitspsychologie als *operatives Abbildsystem* oder OAS bezeichnet [Hac86, S.120][Hac73, S.92][Wie94, S.75].

Zudem ist es zur Steuerung der Aktivität nötig, die Aktionen gedanklich vorwegzunehmen. Diese Vorwegnahme kann verschiedene Formen haben. HACKER bezeichnet sie dann als *Plan*, wenn sie nicht die Aktivität als solche¹, sondern abzuarbeitende konkrete Teilziele festlegt [Hac86, S.159]. Diese können dann durch entsprechende Teilhandlungen oder durch Delegation an Auftragnehmer erreicht werden. Daran angelehnt können wir *Planung* wie folgt definieren:

Definition 6.2 (Plan, Planung). *Ein Plan für ein Ziel (oder eine Aufgabe) ist eine Festlegung von konkreten Teilzielen, nach deren Erfüllung auch das Ziel (der Aufgabe) erfüllt ist. Eine Handlung heißt Planung, wenn ihre Aktivität im Erstellen eines solchen Plans und im Sicherstellen des Erreichens der im Plan festgelegten Teilziele besteht.*

Man beachte, dass dann zu einem Plan stets auch ein Ziel gehört, das mit diesem Plan erreicht werden soll. Die Betonung der Überwachung („Sicherstellen“) der Ausführung des Plans kann auch aus der obigen Begriffsdefinition von „Handlung“ abgeleitet werden, da man sonst nicht von einer zielgerichteten Aktivitätssteuerung sprechen kann.

Für die Gebäude- und Stadtplanung gibt es Beschreibungen von idealisierten Abläufen (vgl. [Joe76, S.13],[Die69], [KS77, S.33ff]). Die am weitesten verbreitete derartige Beschreibung ist die Folge der Leistungsphasen der HOAI².

1. Grundlagenermittlung
Ermitteln der Voraussetzungen zur Lösung der Bauaufgabe durch die Planung
2. Vorplanung (Projekt- und Planungsvorbereitung)
Erarbeiten der wesentlichen Teile einer Lösung der Planungsaufgabe
3. Entwurfsplanung (System- und Integrationsplanung)
Erarbeiten der endgültigen Lösung der Planungsaufgabe
4. Genehmigungsplanung
Erarbeiten und Einreichen der Vorlagen für die erforderlichen Genehmigungen oder Zustimmungen
5. Ausführungsplanung
Erarbeiten und Darstellen der ausführungsfähigen Planungslösung
6. Vorbereitung der Vergabe
Ermitteln der Mengen und Aufstellen von Leistungsverzeichnissen
7. Mitwirkung bei der Vergabe
Ermitteln der Kosten und Mitwirkung bei der Auftragsvergabe

¹Dies bezeichnet HACKER dann als Handlungsschema

²Honorarordnung für Architekten und Ingenieure

8. Objektüberwachung (Bauüberwachung)
Überwachung der Ausführung des Objekts
9. Objektbetreuung und Dokumentation
Überwachung der Beseitigung von Mängeln und Dokumentation des Gesamtergebnisses

Wobei diese Phasen eigentlich gar nicht der Strukturierung des Planungsablaufs dienen, sondern als Grundlage für Honorarabrechnungen dienen. Wegen der allgemeinen Beliebtheit dieses Schemas wollen wir uns grob daran orientieren.

6.2 Planungsaufgabe

In der Architektur ist das Planungsziel ein Gebäude für eine vorgegebene Nutzung. Dies kann eine Umnutzungsplanung von bestehenden Gebäuden, oder eine Neuplanung sein. Das traditionelle Berufsbild des Architekten ist meist an der Neuplanung orientiert, diese verliert jedoch in der Praxis zugunsten anderer Tätigkeiten an Bedeutung. Gelegentlich wirkt sich dieser Wandel sogar auf die Architekturausbildung aus.

Für die Planungsaufgabe definieren wir den Entitätstyp **Projekt** mit Attributen, die das Projekt als Ganzes betreffen, wie z.B. Projektbeginn, Projektbezeichnung, Laufzeit oder Auftraggeber:

Projekt(id:Ganzzahl, bezeichnung:Text, beginn:Datum, ende:Datum, auftraggeber:Personen.ID)

Da zwei verschiedene Projekte nicht die selbe Bezeichnung haben sollen, ist {**bezeichnung**} Kandidatenschlüssel. Ein Entitätstyp **Person** mit Primärschlüssel {**id**} wird hierbei vorausgesetzt. Auf Abbildung 6.1 wird dieser Sachverhalt in UML notiert.

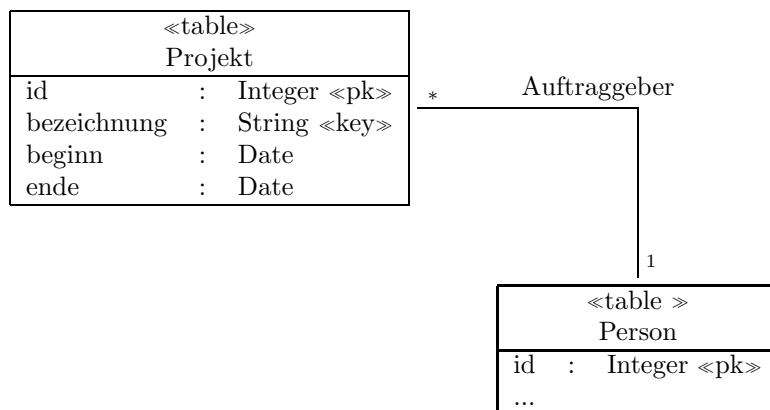


Abbildung 6.1: UML-Notation der Entitätstypen **Projekt** und **Person**. Im oberen Feld steht der Name der Klasse bzw. der Tabelle. Darunter folgt die Attributliste. Das Attribut **auftraggeber** realisiert eine Assoziation.

Wir wollen die verwendeten UML-Notationselemente nun kurz erklären. Ein Entitätstyp entspricht einer *Klasse* der objektorientierten Programmierung und wird in UML als Rechteck mit Klassennamen notiert. Da es sich um eine Tabelle einer relationalen Datenbank handelt, verwenden wir einen sogenannten *Stereotypen* «table». Stereotype dienen dazu, Notationselemente genauer zu charakterisieren. Zur Notation von Attributen und Operationen kann dieses Rechteck in ein Namensfeld, ein Attributfeld und ein Operationenfeld unterteilt sein. Wir werden nur Attribute angeben und lassen daher das Operationenfeld weg.

Beziehungen notiert man als gegebenenfalls beschriftete Verbindungslinien. Die Vielfachheit wird am jeweiligen Ende der Verbindungslinie angegeben: Jedes Projekt hat genau einen Auftraggeber („1“) und eine Person kann mehrere Projekte beauftragen („*“). Die Assoziation „Auftraggeber“ ist also eine Abbildung von den Projekten zu den Personen.

Primärschlüsselattribute kennzeichnen wir durch einen Stereotyp «pk» und Schlüsselkandidaten durch «key». Man beachte auch die Groß- und Kleinschreibung von Klassen-, Attribut- und Beziehungsnamen entsprechend der UML-Konvention.

6.2.1 Das Raum- und Funktionsprogramm

Ein wichtiger Teil der Planungsaufgabe ist das Raum- und Funktionsprogramm. Legen wir hier (frei nach KUHNE [KS77, S.37ff]) fest, was darunter zu verstehen ist:

Definition 6.3 (Funktionsprogramm, Raumprogramm). *Ein Funktionsprogramm ist die Festlegung von Nutzungsbereichen und die Bestimmung der Aufgaben, Arbeitsabläufe, Betriebssysteme³ und Organisationsformen innerhalb des Gesamtgebäudes und der einzelnen Nutzungsbereiche. Ein Raumprogramm ist die Zusammenstellung der einzelnen Räume bzw. Raumgruppen unter Angabe von Funktion, Flächenbedarf, betriebstechnischen Einrichtungen, besonderen bautechnischen Anforderungen, Sondereinrichtungen und erforderlichem Ausbaustandard.*

Diese Programme sind insbesondere deshalb sehr interessant, da sie ein Planungsziel sehr präzise beschreiben und damit die automatisierte Prüfung der Zielerfüllung erleichtern. Daher sollte ein Planungssystem in jedem Fall ein Raumprogramm, idealerweise auch ein Funktionsprogramm vorsehen.

Für Raum- und Funktionsprogramme existiert kein allgemeiner verbindlicher Standard, häufig aber für die darin zu verwendenden Raumnutzungen. Die DIN 13080 – Gliederung des Krankenhauses in Funktionsbereiche und Funktionsstellen – ist z.B. ein genormtes Funktionsprogramm für Krankenhäuser mit Beiblatt 3 als Musterformular dazu.

Mit Teil 2 der DIN 277 – Grundflächen und Rauminhalte von Hochbauten – gibt es ein allgemein anerkanntes Verzeichnis von Flächennutzungen in Gebäuden.

Typisch für diese Normen ist eine hierarchische Gliederung dieser Funktionen, wie etwa die Gliederung nach Funktionsbereich, Funktionsstelle und Teilstelle in der Krankenhausgliederung nach DIN 13080:

³Im Sprachgebrauch des Jahres 1977 ist dies natürlich nicht das Betriebssystem eines Rechners, sondern ein technisches System, das zum Betrieb des Nutzungsbereichs notwendig ist (z.B. ein besonderes Lüftungssystem). n.p.

Funktionsbereich ein Aufgabengebiet (z.B. Untersuchung und Behandlung), gegliedert in Funktionsstellen.

Funktionsstelle jeweils eine Aufgabe (z.B. 1.09 Operation), bei Bedarf gegliedert in Teilstellen.

Teilstelle jeweils eine fachspezifische Aufgabe (z.B. 1.02.03 Augenheilkunde)

Da es offenbar mehrere derartige Funktionsverzeichnisse gibt, wollen wir hier einen Entitätstyp **Funktionsverzeichnis** definieren. Wir nehmen an, dass eine Funktion zunächst unabhängig von einem Funktionsverzeichnis allgemein definiert ist, und haben daher den weiteren Entitätstyp **AllgemeineFunktion**. Jedem Funktionsverzeichnis ist eine Menge von Funktionen, also Entitäten des Typs **AllgemeineFunktion**, zugeordnet. Damit ergibt sich die **Funktion** eines Funktionsverzeichnisses als n:m-Beziehungstyp.

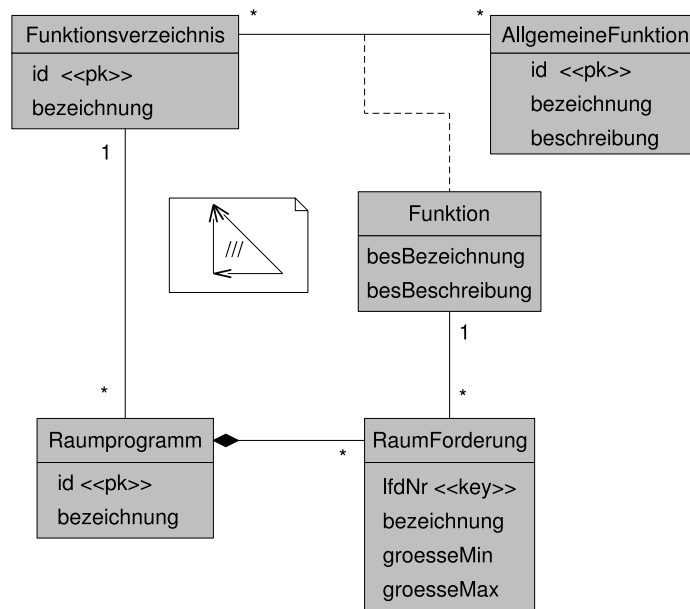


Abbildung 6.2: UML-Diagramm mit Funktionsverzeichnis, Funktion und Raumprogramm. Ein Rechteck mit einem Eselsohr ist ein Kommentarfeld. Das Kommentarfeld mit dem kleinen kommutierenden Dreieck („///“) soll anzeigen, dass das Funktionsverzeichnis des Raumprogramms auch das Funktionsverzeichnis jedes Raums des Raumprogramms sein muss. Kommutierende Diagramme tauchen in UML-Modellen so häufig auf, dass dafür die Einführung eines eigenen Notationssymbols gerechtfertigt wäre.

Dabei bleibt offen, ob zwischen dem Funktionsverzeichnis und der Funktion weitere (statische) Hierarchieebenen liegen. Die Beziehung zwischen **AllgemeineFunktion** und **Funktionsverzeichnis** hat selbst Eigenschaften und ist deshalb als sogenannte Beziehungsklasse dargestellt. Dass mit der Klasse **Funktion** diese Beziehung gemeint ist, wird in UML durch eine gestrichelte

Linie vom Klassensymbol zur Verbindungslinie angezeigt. Ein weiteres neues Element in dieser Abbildung ist die kleine schwarze Raute am Raumprogramm-Ende der Beziehung **Raumprogramm** – **Raumforderung**. Diese zeigt an, dass es sich hier um eine sogenannte *Komposition* handelt. Ein Raumprogramm „besteht aus“ Raumforderungen und wir verbieten, dass diese für sich alleine ohne zugehöriges Raumprogramm existieren. In ER-Modellen bezeichnet man dies als *schwachen* Entitätstyp.

Ein **Raumprogramm** ist (zunächst) eine Liste von Funktionen eines Funktionsverzeichnisses mit Mindestgröße und eventuell besonderen Anforderungen. Daher muss das Raumprogramm festlegen, welches Funktionsverzeichnis es verwendet. Wir haben es also mit einer Abbildung

Raumprogramm → **Funktionsverzeichnis**

zu tun, oder – im Datenbankjargon – mit einem 1:n-Beziehungstyp.

Topologische Datenbanken in UML Bevor wir nun mit der Modellierung fortfahren, wollen wir hier eine Erweiterung der UML-Notation für topologische Datentypen einführen: Wir notieren einen topologischen Raum als Entitätstyp mit dem Stereotyp `<<raum>>`. Dieser Raum umfasst eine Menge von Punkten und definiert auf diesen eine Relation. Anstelle der vollständigen UML-Notation dieses Sachverhaltes wollen wir die Aggregations- oder Kompositionsbeziehung ohne Name und Vielfachheiten notieren und die Beziehung zu den Punkten mit dem Stereotyp `<<punkte>>` oder Stereotypsymbol „...“ notieren, wie auf Abbildung 6.3 angegeben.

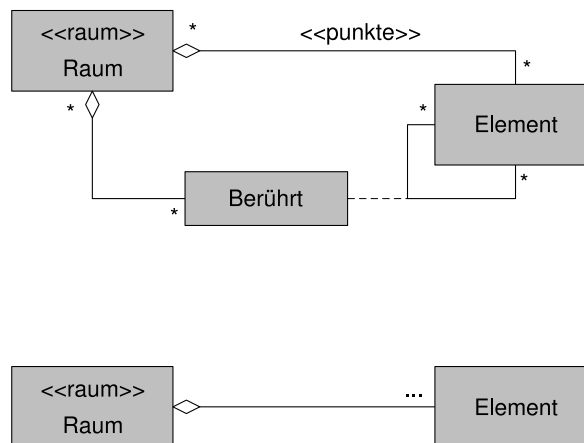


Abbildung 6.3: Oben die ausführliche und unten unsere vereinfachte UML-Notation von topologischen Räumen. Ein Raum setzt sich aus Elementen zusammen, wobei hier ein Punkt mehreren Räumen angehören kann. Die Beziehung **Berührt** ist durch den Raum definiert.

Im Gegensatz zur obigen Beziehung **Raumforderung** – **Raumprogramm**, die in UML als Komposition bezeichnet wird, haben wir die Beziehung der Klasse **Raum** zu den Punkten des Raums hier als eine sogenannte *Aggregation* mit einer *leeren*

Raute notiert. Dies bedeutet, dass Elemente (Punkte) eines Raums auch ohne zugehörigen Raum existieren oder auch mehreren Räumen angehören können – z.B. einem Teilraum. Ein Raum kann alternativ auch als Komposition seiner Punkte notiert werden.

6.2.2 Planungskontext

Bevor ein Planungsziel angestrebt wird, muss der Ausgangspunkt der Planungshandlung bekannt sein. Daher muss unser System ermöglichen, zunächst den Ist-Zustand zu dokumentieren, wie das Umfeld der Planung, Lage und Form des Grundstücks, die benachbarte Bebauung, die städtebauliche Situation und planungsrechtliche Einschränkungen. Auch bestehende und schützenswerte Pflanzen und — insbesondere bei Umnutzungsplanungen — auch noch der aktuelle Baubestand gehören zum Kontext. Umnutzung stellt dabei besonders hohe Anforderungen an die Dokumentation des Ist-Zustandes und an die Dokumentation der Veränderungen, die durch Planungshandlungen vorgenommen werden.

Dabei ist darauf zu achten, dass der Planungskontext als solcher prinzipiell unveränderlich bleibt. Die Dokumentation des Ist-Zustandes vor Beginn der Planung darf also nicht verloren gehen — eine Änderung daran ist also nur dann erlaubt, wenn neue Erkenntnisse über die Ausgangslage für die Planung vorliegen, nicht als Folge der beabsichtigten Planungsmaßnahmen.

In der Umnutzungsplanung kann dann jeder Plan als „Patch“ für den Kontext verstanden werden. Das heißt, dass z.B. ein abzubrechendes Gebäude X nicht aus dem Kontext gelöscht werden darf, sondern es gibt dazu eine Beschreibung der Änderung des Kontextes, den Patch P_X = „Abbruch von X “. Die Anwendung von P_X auf den Kontext erzeugt erst die neue Situation: das Baugrundstück ohne X . Wir haben in Abschnitt 4.4.10 angedeutet, wie diese Operatoren aussehen können.

Mit einer Versionskontrolle ausgestattet können wir nun eine bestimmte Version „PLANUNGSKONTEXT“ festhalten und benötigen somit keinen Entitätstyp für den Planungskontext. Dieser ist vielmehr ein 1 : 1-Beziehungstyp vom Projekt zu den Planungsversionen. Insbesondere besteht aus Sicht der Modellierung kein Unterschied zwischen Neuplanung und Umplanung.

6.2.3 Planungsobjekt

Das Planungsobjekt ist ein Gebäude, der „umbaute Raum“. Die Besonderheiten dieses Objekts sind das Motiv für diese Arbeit und natürlich wird für dieses ein topologischer Datentyp definiert. Dieses Gebäude ist nun in seinen Kontext „eingebettet“ und damit haben wir zwei topologische Räume: Der Planungskontext zu Projektbeginn und die angestrebte Situation nach Fertigstellung des Projekts. Das „Objekt“ ist dabei einfach ein Teilraum dieser angestrebten Situation.

Wir identifizieren also gewisse Elemente dieses Teilraums, bezeichnen diese als „Objekt“ und der Rest ist „Kontext“. Dazu könnten wir einen Entitätstyp **Planungsobjekt** definieren, der in 1 : n -Beziehung zu den Elementen des Raums steht. Es besteht so auch die Möglichkeit, dass mehrere Planungsobjekte in einem einzigen Projekt existieren.

Allerdings gibt „1 : n “ die Vielfachheit dieser Beziehung nicht ganz korrekt wieder: Die Elemente des Planungskontextes gehören nämlich nicht zum Pla-

nungsobjekt, stehen also nicht mit einem, sondern mit keinem Planungsobjekt in Beziehung. Damit ist es richtiger, die „1“ in „1 : n“ durch das Intervall $\{0, 1\}$ zu ersetzen und von einer $\{0, 1\} : n$ -Beziehung zu sprechen. Genau dies sieht UML auch vor: Intervalle $\{a, \dots, b\}$ bei der Angabe der Vielfachheiten einer Beziehung werden dort in der Form $a..b$ angegeben. Statt $a..a$ schreibt man kurz a und Vielfachheit „1“ kann ganz weggelassen werden. Beziehungsenden ohne explizit angegebene Vielfachheit haben also Vielfachheit 1. Wir erinnern, dass Vielfachheit n („beliebig viele“) durch ein Sternchen $*$ angegeben wird. Dies ist ebenfalls eine Kurzschreibweise, und zwar für $0..*$: Beliebig viele aber mindestens 0, denn mit $a..*$ kann auch eine andere untere Grenze a der Vielfachheit festgelegt werden, häufig z.B. $1..*$.

Auf Abbildung 6.4 wird nun die Gliederung des Projekts in mehrere Abschnitte modelliert. Hier wird auch ein neues Notationselement eingefügt: die Einschränkung. Die Assoziation einer Gliederungsebene zu ihren Planungsabschnitten ist als „/final“ gekennzeichnet. Das bedeutet hier, dass die Topologie der Planungsabschnitte die Finaltopologie bezüglich der Zuordnung von Elementen zu Abschnitten ist. Der Schrägstrich vor dem „final“ gibt an, dass diese Finaltopologie aus der Topologie des Planungsstandes zu berechnen ist. Dies ist nicht ganz selbstverständlich, denn es ist durchaus denkbar, umgekehrt eine Finaltopologie im Wertebereich vorzugeben und damit die im Definitionsbereich zulässigen Topologien entsprechend einzuschränken.

Einschränkungen von Elementen werden in UML in geschweifte Klammern $\{ \dots \}$ geschrieben.

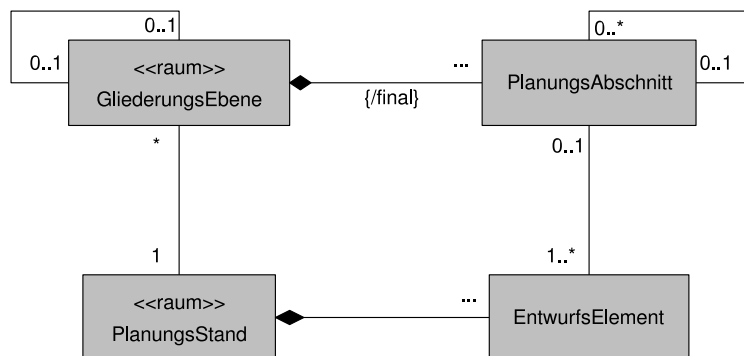


Abbildung 6.4: Anstelle von **Planungsobjekt** modellieren wir hier eine dynamische Hierarchie von Gliederungsebenen wie Gebäude – Geschoss – Zone – Raum, die bei **Entwurfselement** beginnt und bei einem „Planungsobjekt“ enden kann. Für jede Hierarchieebene wird ein **Gliederungsebene**-Objekt erstellt. Ein **Planungsabschnitt** muss mindestens ein Entwurfselement umfassen, hat also auf der **Entwurfselement**-Seite Vielfachheit $1..*$.

Mit dieser Modellierung haben wir eine Kette von Gliederungsebenen, wobei jedes Element einer höheren Ebenen eine Äquivalenzklasse von Elementen der darunter liegenden Ebene ist. Damit ist das Planungsobjekt nur ein besonderer Planungsabschnitt. Es ist derjenige Planungsabschnitt auf der obersten Hierarchieebene im Projekt.

Alternativ ließen sich auch statische Gliederungsebenen verwenden. Allerdings gibt es dann Probleme bei Projekten unterschiedlicher Komplexität (Krankenhaus vs. Einzelgarage), für die unterschiedlich tiefe Gliederungen sinnvoll sind. Bei einer statischen Gliederung müsste aber auch ein einfaches Projekt, wie z.B. die erwähnte Einzelgarage, nach Abschnitt-Geschoss-Station-Raum, gegliedert werden, was nicht sehr sinnvoll ist.

Da die Gliederungsebene dem Projekt zugeordnet ist, gibt es noch eine Kompositionsbeziehung von **Projekt** zu **Gliederungsebene**, die in Abbildung 6.4 nicht angegeben ist.

Auch in die andere Richtung ist eine verfeinernde Gliederung möglich. Dies ist zum einen die Detaillierung, und zum anderen ist eine Zerlegung in Finite Elemente nicht wesentlich von der hier beschriebenen hierarchischen Gliederung verschieden.

6.3 Beteiligte

Zur Berücksichtigung der einzelnen Planungsbeteiligten wie Planer, Fachplaner, ausführende Firmen, Bauherr und Nutzer ist ein Konzept der Rechtevergabe und Bearbeitungssperren nötig. Wir könnten z.B. für ein Projekt Arbeitsgruppen festlegen, wobei jeder Beteiligte einer Gruppe angehören muss, um Planelemente bearbeiten zu können. Jedes Planelement kann dann nur von genau einer Gruppe bearbeitet werden. Damit sind im Prinzip alle Differenzierungen von Rechtevergaben möglich, da so auch für jede Teilmenge von Beteiligten eine eigene Gruppe definiert werden kann. Ein differenzierteres Modell für Rechtevergabe wollen wir hier nicht entwickeln.

Aus Sicht der topologischen Modellierung wäre jedoch interessant zu untersuchen, wie weit eine Bearbeitungssperre räumlich gehen muss. Wird ein Element e von einem Bearbeiter reserviert, dann kann es erforderlich sein, die Hülle $cl e$ oder die Umgebung $St(e)$ von e ebenfalls zu sperren.

6.4 Dokumente

Man unterscheidet zwischen dem internen Aufbau eines Softwaresystems, und demjenigen Teil dieses Systems, der für externe Benutzer sichtbar ist, die *Schnittstelle* dieses Systems. Diese Schnittstelle beinhaltet den Benutzerdialog zum Erstellen, Abfragen, Modifizieren und Löschen der Planelemente und die Ausgabe von Planzeichnungen an Druckern bzw. ganz allgemein die Ausgabe von graphischen Daten, und es beinhaltet ebenfalls den Datenaustausch mit anderen Softwaresystemen, wie z.B. die Erzeugung und das Importieren von IFC-Dateien.

Werden derartige Dokumente an andere Planungsbeteiligte wie etwa ausführende Firmen ausgegeben, dann macht dies natürlich nur dann Sinn, wenn diese Firmen selbst keinen Zugriff zu den im System gespeicherten Daten haben. Dann ist aber genau zu dokumentieren, auf welchem Informationsstand diese Planungsbeteiligten stehen, und deshalb ist die Erzeugung und Verteilung von derartigen Dokumenten ebenfalls genau zu dokumentieren. Dazu definieren wir einen Entitätstyp **Plan**, dem alle Parameter zur Dokumenterzeugung zugeordnet werden, welche notwendig sind, um das entsprechende Dokument

reproduzieren zu können. Bei Grundrissen und Schnitten wäre das etwa die Festlegung der Schnittebene, die Blickrichtung und der Maßstab, bzw. der verwendete Detaillierungsgrad.

Zur korrekten Reproduktion eines Dokuments ist es nötig, auf die entsprechende Version des zugrunde liegenden Modells zurückzugreifen, und wir benötigen daher eine Versionskontrolle. Bei der Verwendung von Detailbibliotheken ist ebenfalls jeweils das Detail in der richtigen Version zu verwenden — für Detailbibliotheken ist somit ebenfalls eine Versionskontrolle nötig.

Weiterhin müssen die Empfänger dieses Dokuments dem Dokument zugeordnet werden können. Wir haben also noch eine $n : m$ -Beziehung `PlanVerteiler` zwischen `Dokument` und `Person`. Da sinnvollerweise der Zeitpunkt der Zusendung im Verteiler mitgespeichert wird, diese Beziehung selbst also Attribute besitzt, führen wir `PlanVerteiler` als Beziehungsklasse aus.

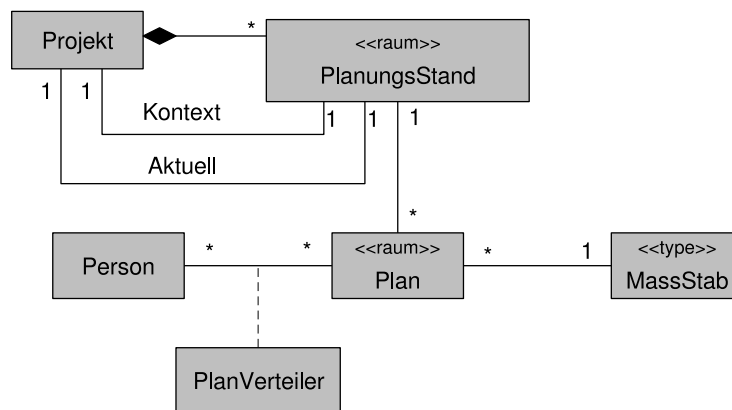


Abbildung 6.5: Die Klasse `Plan` enthält alle Informationen, die zur Plangenerierung notwendig sind. Anhand ihrer Assoziation zu `MassStab` kann z.B. ein entsprechender Detaillierungsgrad bei der Plangenerierung bestimmt werden. Die Klasse `PlanVerteiler` enthält die Informationen, ob und wann ein Projektbeteiligter ein Plandokument auch erhalten hat. Der Typ `PlanungsStand` enthält alle archivierten Versionen des aktuellen Planungsverlaufs. Zwei Planungsstände sind dabei gesondert ausgezeichnet: `Kontext` markiert den Ausgangszustand der Planung und `Aktuell` den aktuellen Zustand und entspricht der Version `HEAD` im `CVS`.

Dieses Diagramm setzt eine Versionskontrolle voraus, auf die wir nun genauer eingehen wollen.

6.5 Versionen

Zur Vereinfachung gehen wir davon aus, dass der Planungskontext irgendwann während der Grundlagenermittlung erstellt und dann eingefroren wird. Es gibt also einen klaren Schnitt zwischen Grundlagenermittlung und Entwurfsplanung. Nachträgliche Änderungen am Planungskontext werden damit ausgeschlossen.

Wir wollen hier außerdem auch ausschließlich topologische Änderungen betrachten. So ist das Verschieben einer Wandöffnung erst dann eine „Änderung“ in diesem Sinne, wenn sie nicht durch einen Homöomorphismus bewerkstelligt werden kann.

Dann erinnern wir uns an Definition 4.28, dass wir zu zwei Räumen X und Y immer ein Paar $(A, N) \in \text{diff}(X, Y)$ von Räumen finden können, so dass $Y = \text{patch}(X, (A, N))$ ist. Die Größe eines minimalen (A, N) ist ein Maß dafür, wie stark sich Y von X unterscheidet. Je geringfügiger also die Änderung ist, desto geringer ist der prinzipielle Platzbedarf zum Abspeichern dieser Änderung. Zudem ist es ausreichend, nur die Punktmenge von A zu speichern, da dessen Topologie $\mathcal{T}_X|_A$ zum Patchen irrelevant ist. Jedes diff ist also ein Paar von Teilräumen, jeweils des alten und des neuen Raums, das wir zudem noch mit einer fortlaufenden Revisionsnummer und dem Änderungsdatum versehen. Derart durchnummerierte Einzeländerungen bezeichnen wir als Revisionen und die Markierung eines Bearbeitungsstandes insgesamt zu einem gewissen Zeitpunkt heißt Ausgabe (release).

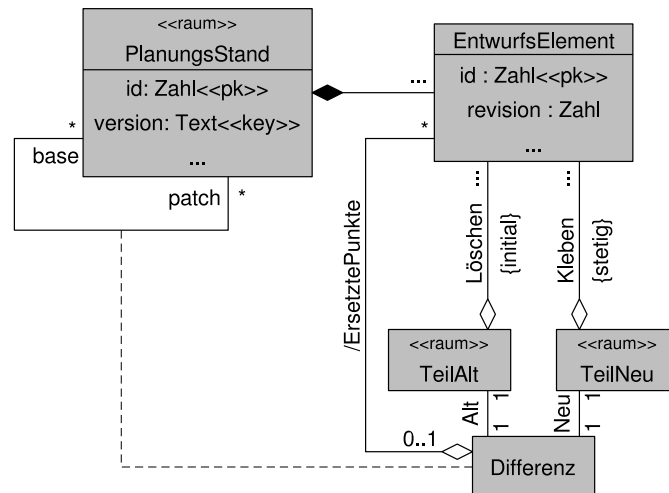


Abbildung 6.6: Jeder `PlanungsStand` steht für eine bestimmte Version der Planung. Die Klasse `Differenz` steht für eine Transaktion zwischen zwei Revisionen. Anstelle der Klasse `TeilAlt` könnte auch die Aggregation `/ErsetztePunkte` verwendet werden, da nur die zu löschende Punktmenge von Interesse ist, nicht aber deren Topologie. Diese ist ja lediglich die Spurtopologie von `TeilAlt` im `PlanungsStand`.

Aus Abbildung 6.6 wird ein Schema für eine Versionskontrolle vorgeschlagen. Hier wird wieder ein neues UML-Element eingeführt: Die abgeleitete Assoziation `/ErsetztePunkte`. Abgeleitete Elemente sind diejenigen Elemente, die sich aus anderen Elementen berechnen lassen, die also redundant sind. Diese werden durch einen dem Namen vorangestellten Schrägstrich gekennzeichnet. `/ErsetztePunkte` lässt sich aus den Assoziationen `Alt` und `Löschen` durch Relationenprodukt ermitteln:

$$\text{/ErsetztePunkte} = \text{Alt} \circ \text{Löschen}.$$

Zum Bearbeiten einer Revision gehen wir analog zu CVS vor: Es existiert ein Aufbewahrungsort (Repository) für alle Revisionen unserer Planung. Dieser befindet sich in der Regel auf einem Server. Durch eine Abfrage (update) erhalten wir die Revision, die wir bearbeiten möchten — in der Regel die aktuelle. Diese wird nun vom Aufbewahrungsort in einen persönlichen Arbeitsbereich (Sandbox) übertragen und kann dort bearbeitet werden. Der Arbeitsbereich liegt in der Regel auf einem Arbeitsplatzrechner bzw. auf auf einem personalisierten Bereich in einem Netzwerk. Gegebenenfalls werden die so in Arbeit befindlichen Elemente für den momentanen Bearbeiter reserviert und für alle anderen Bearbeiter gesperrt (watch & edit). Ist der Bearbeiter fertig, dann wird die nun entstandene Version wieder an den Aufbewahrungsort zurückübertragen (commit), dort mit der abgefragten Version verglichen, ein entsprechendes `diff` erstellt und unter einer neuen Revisionsnummer abgespeichert. Ein derartiger Ablauf wird auch als *lange Transaktion* bezeichnet. Diese sind dadurch gekennzeichnet, dass im Zeitraum zwischen update und commit andere Bearbeiter auf die Datenbasis zugreifen und eventuell ebenfalls Änderungen vornehmen können. Man sagt auch, die Transaktion ist nicht isoliert. Dann kann es aber passieren, dass die Änderungen der beiden Bearbeiter nicht miteinander verträglich sind. Wenn etwa Bearbeiter *A* eine Tür in eine Wand einfügt und Bearbeiter *B* eine andere Tür derart in die selbe Wand einfügt, dass beide Türen kollidieren, dann kann nur einer von ihnen sein commit durchführen. Dazu muss er sich mit seiner Änderung gegenüber dem anderen Bearbeiter durchsetzen. Das Versionssystem muss nun derartige Konflikte erkennen und dann entweder das commit zurückweisen oder die Bearbeiter über derartige Konflikte unterrichten, damit diese geeignet darauf reagieren können. Die Reservierung von Elementen ermöglicht, derartigen Konflikten vorzubeugen.

Bei gegebener Zeit kann nun ein bestimmter Planungsstand als Ausgabe markiert werden, etwa als „EINGEREICHT“ wenn ein Bauantrag eingereicht wurde oder „BESPRECHUNG-42“, die Ausgabe des Planungsstandes zur 42-ten Besprechung.

6.6 Semantik

Als Vorteil der topologischen Modellierung wurde eingangs versprochen, dass sie ermögliche, die räumliche Struktur und die Bedeutung der Elemente dieses Raums jeweils eigenständig zu modellieren, was die Modellierung vereinfacht. Hier soll nun die andere — nicht-räumliche Seite der Modellierung betrachtet werden. Als „Bedeutung“ eines Elements bezeichnen wir also einfach die nicht-räumlichen Eigenschaften wie Nutzung, Materialwahl, Farbgebung, Temperatur oder Belastung. Aus manchen dieser Eigenschaften lassen sich wiederum weitere Eigenschaften ableiten wie etwa die Wärmeleitfähigkeit aus gewähltem Material und der Geometrie oder die mechanische Spannung, die sich aus Belastung, Materialwahl und Geometrie ergibt.

6.6.1 Detailbezeichner

Einen Teil der Semantik haben wir bereits durch die Detailbibliotheken gegeben: Bereits wenn wir ein Detail geeignet bezeichnen, etwa als „Fenster XYZ“, dann wissen wir allein durch die Detailbezeichnung, dass es sich um ein Fen-

ster handelt. Damit ist diese Eigenschaft „Fenster“-sein jedoch nicht in erster Normalform modelliert.

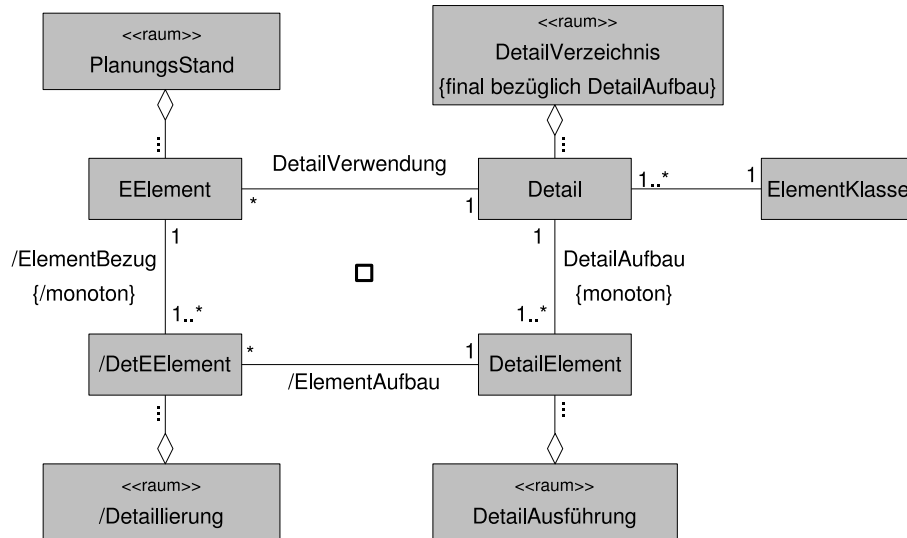


Abbildung 6.7: Klassendiagramm einer Detailbibliothek. Im rechten Teil ist die Detailbibliothek, die sich zusammensetzt aus dem topologischen Raum der Detailausführungen, und deren erlaubter Verbindungen untereinander und dem Verzeichnis der Details. Dass /Detailierung ein Faserprodukt (bzw. Gleichheitsverbund) ist, stellen wir hier durch das in der Mathematik dafür übliche Symbol, das kleine Quadrat in der Mitte, dar. Dieses Symbol ist kein UML-Symbol.

6.6.2 Detailklassen

Wenn wir die Detailbezeichner unserer Detailbibliothek geeignet gliedern, dann ergibt sich eine Klassifizierung der Details wie etwa in „Fenster“, „Türen“ oder „Wände“. Insbesondere benötigen wir nicht jeweils einen eigenen Entitätstyp für jede Elementklasse wie z.B. unsere `IfcDoor` oder auch `IfcWindow` und `IfcWall`. Es genügt der Entitätstyp `ElementKlasse` in dem all diese Klassen zusammengefasst sind. Auf Abbildung 6.7 wird die Detailbibliothek und eine mögliche Anordnung des Typs `ElementKlasse` im Klassendiagramm angegeben. Wir haben also — im Gegensatz zu den IFC — *dynamische* Elementklassen und damit ein einfacheres Datenbankschema, das sich auch dann nicht ändert, wenn wir neue Elementklassen hinzufügen. Allerdings haben wir für die so gewonnene Einfachheit auch einen Preis zu zahlen: Attribute, die für eine Elementklasse spezifisch sind, wie etwa die Schlagrichtung einer Türe, können auf diese Weise nicht sinnvoll dargestellt werden. In diesem Fall sind wir dann doch gezwungen, jeweils einen eigenen Entitätstyp zu definieren oder mit Konzepten ähnlich wie den IFC-Property Sets zu arbeiten. Dies ist jedoch eine Entwurfsentscheidung, die unabhängig davon ist, ob wir nun Topologie einsetzen oder nicht, und die wir deshalb hier nicht mehr weiter diskutieren werden.

6.6.3 Dimensionsklassen

Neben der Definition von Elementklassen bietet sich auch die Klassifikation nach der Dimension der Elemente an: Volumen, Flächen, Kanten und Knoten. Wir erinnern uns, dass wir es mit „Dimension“ dabei nicht ganz genau nehmen müssen und z.B. Wände als zweidimensionale und Türrahmen als eindimensionale Objekte betrachten können, auch wenn diese durch ihre Ausdehnung in der Realität stets ein Volumen einnehmen. Dieses Volumen, wie etwa ein Wandvolumen, taucht jedoch erst bei einem bestimmten Detaillierungsgrad auf.

Volumen

Hier sind die Nutzräume eines Gebäudes im umgangssprachlichen Sinn gemeint. Da diese natürlich eine zentrale Rolle in der Planung spielen, wollen wir einen Entitätstyp `NutzRaum` definieren, der die Entwurfselemente referenziert. Jedes dreidimensionale Entwurfselement kann einen Raum darstellen, und dadurch eine im Raumprogramm geforderte Raumnutzung realisieren.

Damit haben wir eine $1 : 0..1$ -Beziehung von dem Typ `EntwurfsElement` zu `NutzRaum` und von dort eine $* : 1$ von `NutzRaum` zum Typ `RaumForderung` aus dem Raumprogramm. Bei einem Entwurf, der das Raumprogramm erfüllt, hat diese Assoziation Vielfachheit von $1..* : 1$. Wir erwähnen noch, dass die Erfüllung des Raumprogramms eine abfragbare Eigenschaft, aber keine Konsistenzregel ist. Ein Entwurf darf also – aus Sicht der Datenhaltung – sein Raumprogramm durchaus verfehlen.

Wände, Türen, Fenster

Die flächigen Bauelemente begrenzen und verbinden Räume. Zum Begrenzen können wir jeder Raumnutzung die Emission und eine maximale Immission einer bestimmten Wirkung, wie etwa Schall oder Wärme, zuordnen und jedem Detail für ein flächiges Bauelement einen Widerstand und gegebenenfalls eine Pufferung (z.B. Wärmekapazität) dieser Wirkung. Damit ließe sich grob abschätzen, inwieweit die gewünschte Trennwirkung in der Tat erreicht wird. Für eine genauere Modellierung der Ausbreitung derartiger Wirkungen ist dieser Ansatz aber noch zu verfeinern, um Phänomene wie z.B. Körperschall oder Wärmetransport über Bauteilverbindungen korrekt zu berücksichtigen.

Die Bezeichnung „Wirkung“ wurde bewusst abstrakt gewählt, weil derartige Wirkungen auch andere Phänomene wie Schutz der Privatheit (Durchsichtigkeit) oder Sicherheit (als „Immission“ von Unbefugten) umfassen können.

Umgekehrt haben aber flächige Bauelemente auch verbindende Funktionen, wie etwa Fenster als Sichtverbindung von innen nach außen oder Türen zur Sicherstellung der Erschließung. Auch hier könnte das Konzept Emission/Immission ein geeignetes formales Modell für gewünschte Verbindungseigenschaften sein. In diesem Fall ist die Immission erwünscht, was durch eine untere Grenze bzw. durch Angabe eines Minimums modelliert werden kann.

Damit ist es möglich, sowohl einige der Trennungs- als auch der Verbindungsfunktionen von flächigen Baukörpern einheitlich durch folgende Parameter zu beschreiben:

- eine obere und untere Schranke von zulässigen erduldeten oder erwünschten Einwirkungen (Immissionen oder Service),

- eine Wirkung (Emission oder Dienstleistung), die von dem Raum ausgeht,
- die Durchlässigkeit (bzw. Widerstand) des Bauelements für diese Wirkung.
- die Speicherkapazität des Bauelements für diese Wirkung.

Bei der Angabe der Durchlässigkeit ist darauf zu achten, dass nur einige Wirkungen, wie etwa Wärme, sich ganz in Analogie zum Ohmschen Gesetz ausbreiten. Dann reicht es, für die Durchlässigkeit einen einfachen numerischen Widerstandswert anzugeben. Bei anderen Wirkungen besteht diese Analogie jedoch im Allgemeinen nicht, und dort gibt ein einfacher Widerstandswert diese Durchlässigkeit nicht korrekt wieder. Auf Abbildung 6.8 wird ein Modell nur für diese einfache Form der Verbindungs- und Trennungsfunktionen dargestellt.

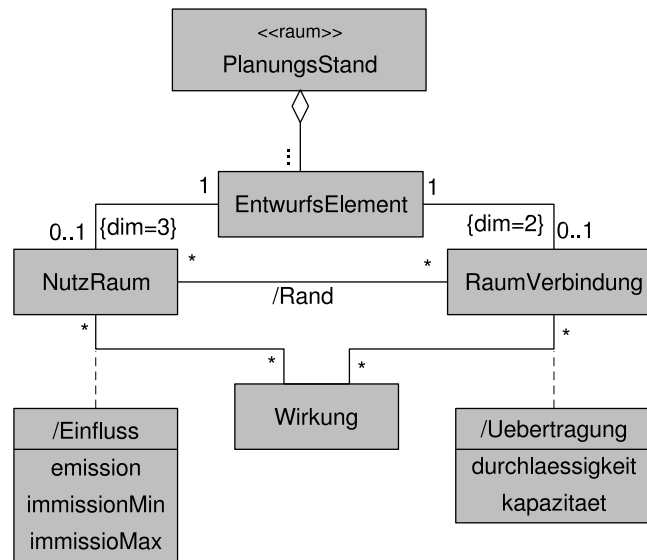


Abbildung 6.8: Eine **Wirkung** repräsentiert einen **Einfluss**, den ein **Raum** (im Sinne von „Zimmer“) auf einen anderen **Raum** haben kann. Die **Attributwerte** werden anhand der **Raumgeometrie** und der zugeordneten **Nutzung** ermittelt, weshalb die **Beziehung Einfluss** als **abgeleitet** gekennzeichnet ist. In der **Detailbibliothek** können dann die **Werte** für die **Wirkungsübertragung** (**durchlässigkeit**) und gegebenenfalls deren **Speicherung** (z.B. **Wärmekapazität**) stehen. Es bleibt offen, auf welche **Weise** diese **Werte** bei der **Berechnung** der **Wirkungsübertragung** verwendet werden.

Ein präziseres Modell dafür wollen wir hier nicht erarbeiten. In jedem Fall aber enthält das hier vorgestellte Modell alle *räumlichen* Informationen, um diese Phänomene mit ausreichender Genauigkeit zu modellieren.

Kanten und Knoten

Was über die Trenn- und Verbindungsfunktionen bei den flächigen Bauelementen gesagt wurde, gilt ebenfalls für die **Kanten** und **Knoten**, da der **Transport**

von Wirkungen wie etwas des Körperschalls auch innerhalb der flächigen Elemente über die Verbindungskanten und Knoten hinweg erfolgen kann. Außerdem können diese Verbindungen von flächigen Elementen untereinander, etwa Wand und Fenster durch den Fensterrahmen, die gewünschte Verbindungswirkung eben dieser Flächen zwischen angrenzenden Volumen, wie etwa Außen und Innen, behindern oder fördern. Der Fensterrahmen schränkt die Beleuchtungswirkung der Fensteröffnung ein und beeinflusst zudem noch den Wärmeverlust.

Es sei übrigens noch einmal daran erinnert, dass in diesem Modell jede Verbindungskante in begrenzende Knotenpunkte mündet, und für diese ist dann ebenfalls eine Detaillierung vorzusehen. Dies hat zur Folge, dass der Bearbeiter gezwungen ist, für jeden Knotenpunkt explizit anzugeben, ob und wie dieser konstruktiv gelöst wurde. Dies ist insbesondere deshalb bedeutsam, da derartige Knotenpunkte in der Praxis gelegentlich übersehen werden. Es ist zwar möglich, auf eine Lösung von Knotenpunkten in voller Absicht zu verzichten und sich von den Verhältnissen auf der Baustelle überraschen zu lassen, es ist jedoch kaum möglich, dies zu übersehen, denn ein Knotenpunkt muss zu jeder Zeit mindestens eine konstruktive Lösung haben und sei es die Markierung „ \uparrow_{Knoten} “ für „nicht gelöst“, wie wir in Abschnitt 4.5.5 ab Seite 179 ausführlich diskutiert hatten. Eine einfache Abfrage, wo denn dieses „Detail“ verwendet wird, gibt alle Stellen im Projekt an, die eine entsprechende Überraschung auf der Baustelle in Aussicht stellen.

6.7 Zusammenstellung

Damit haben wir einiges Material für ein Planungsinformationssystem gesammelt und wollen nun die einzelnen Teildiagramme zu einem gesamten Modell zusammenfassen. Dies ergibt das Klassenmodell auf Abbildung 6.9. Dieses Modell ist noch weit von einer praktischen Realisierung entfernt. Es modelliert nur den Anwendungsbereich, diesen sehr abstrakt und basiert zudem auf angenommenen Lösungen von Problemen, wie etwa der geometrischen Realisierung, der Planungsstände und der Detailbibliotheken.

Das Modell beansprucht daher auch nicht, eine fertige Vorlage für Planungssysteme zu sein. So basiert es auf einem sehr einfachen Modell der gegenseitigen Beeinflussung von Nutzräumen über deren Raumgrenzen hinweg. Dieser Bereich der Wirkungen und deren Ausbreitung im Raum muss für eine Implementierung in jedem Fall überarbeitet werden.

Hier wird aber nur bezweckt, mit einem einfachen ersten Entwurf zu illustrieren, an welchen Stellen eines Planungssystems topologische Datenbanken sinnvoll eingesetzt werden können. Insbesondere ist dieses Modell verhältnismäßig einfach und zudem sehr allgemein gehalten — der größte Teil ist für jede Planung mit Raumbezug geeignet. Nur der Bereich mit dem Funktionsverzeichnis/Raumprogramm ist speziell auf das Bauwesen zugeschnitten. Zwar sind auch die Klassen der Nutzräume und Raumgrenzen bauwesenspezifisch, aber diese sind ebenfalls nur exemplarisch und deren oben geforderte Überarbeitung dieses Bereichs wird wohl zu einer allgemeineren Lösung führen.

Wir haben hier also ein Konzept, wie die Theorie der topologischen Datenbanken praktisch eingesetzt werden kann. Dies ist naturgemäß zunächst auf den *räumlichen* Aspekt der Planung beschränkt und weitere Aspekte wie etwa Projektmanagement, Qualitätssicherung oder Kooperation bleiben dabei ver-

Kapitel 7

Zusammenfassung und Ausblick

In dieser Arbeit wurde die Beantwortung folgender Frage in Aussicht gestellt:

Wie lässt sich die Modellierung der topologischen Eigenschaften von Gebäuden verallgemeinern, und kann diese Verallgemeinerung als Grundlage eines Modells für Gebäude, deren Elemente und Verbindungen dienen?

Der erste Teil der Frage war schnell beantwortet, da bereits 1937 von ALEXANDROFF eine Klasse von topologischen Räumen vorgestellt wurde, die alle endlichen topologischen Räume umfasst und deren Realisierung mit relationalen Datenbanksystemen sich geradezu aufdrängt.

Die topologischen Eigenschaften von Gebäuden erhalten wir, indem wir einfach voraussetzen, dass durch architektonische Räume stets Zerlegungen des Euklidischen Raums in endlich viele Teile definiert sind. Damit ergibt sich durch jede dieser Zerlegungen eine Topologie für diese Teile. Wenn wir also von „topologischen Eigenschaften“ reden, dann betrachten wir einfach diese Topologien. Es gibt nun für verschiedene Maßstabebenen auch verschiedene Zerlegungen, insbesondere erhalten wir je eine Topologie pro Maßstabebene. Da diese Zerlegungen endlich sind, erhalten wir alexandroffsche Topologien, die sich wiederum mit relationalen Datenbanken modellieren lassen.

Dazu wurde hier die entsprechende Erweiterung des relationalen Datenmodells um Topologie vorgeschlagen, die wir als *topologische Datenbank* bezeichnet haben. Dabei zeigte sich, dass eine entsprechend erweiterte Abfragesprache im Allgemeinen eine relationale Algebra mit transitiver Hülle ist. Manche der gegenwärtig auf dem Markt angebotenen Datenbankmanagementsysteme bieten eine solche Abfragesprache als proprietäre Erweiterung von SQL bereits heute an, obwohl diese Erweiterung von CODD erst vor weniger als dreißig Jahren in [Cod79] vorgeschlagen wurde (☺).

Zudem konnte gezeigt werden, dass die topologischen Datenbanken bezüglich des asymptotischen Speicherbedarfs bereits die bestmögliche Lösung zum Speichern von allgemeinen endlichen topologischen Räumen sind. Es hat sich zudem herausgestellt, dass Topologien im Allgemeinen zwar einen bezüglich der Punktezah quadratisch wachsenden Speicherplatzbedarf haben, dass aber im be-

sonderen Anwendungsbereich der Gebäudeproduktmodellierung für Planungsinformationssysteme dieses Effizienzproblem nicht auftreten wird und praktisch immer lineares Wachstum des Speicherplatzbedarfs vorliegt.

Dann wurde auch am Beispiel der monotonen Abbildungen gezeigt, wie topologische Eigenschaften für den Entwurf von derartigen Systemen verwendet werden können: Detaildatenbanken müssen mit monotonen Abbildungen definiert werden.

Auf dieser Theorie aufbauend wurde ein erster Entwurf für den raumbezogenen Teil eines Planungssystems vorgestellt. Dieser Entwurf zeigt bereits, in welche Richtung die weitere Forschung in diesem Bereich gehen kann. Zunächst sind die offen gelassenen Fragen – insbesondere die geometrische Realisierung – zu klären. Dann kann dieser Entwurf zwar als Grundlage für eine Implementierung genommen werden, allerdings ist bis dahin noch viel Arbeit offen. Der Entwurf ist zunächst nur konzeptionell und von der Anwendungsseite gesehen. Er ist noch auf innere Widersprüche (Konsistenz) und auf eventuelle Widersprüche zu praktischen Anforderungen zu analysieren. Weiterhin gibt es Anwendungsbeispiele der Planung, die zwar nicht so offensichtlich räumlicher Natur sind, aber dennoch topologisch modelliert werden können. Insbesondere im sogenannten *Axiomatic Design* [Suh01] ist die „Abhängigkeitsmatrix“ eine Relation und definiert eine Topologie auf den Paaren (fr, ts) je einer funktionellen Anforderung fr und deren entsprechenden technischen Lösung ts . Die von SUH geforderte Eigenschaft „Unabhängigkeit“ entspricht in etwa der T_0 -Trennbarkeit des entsprechenden topologischen Raums. Das hierarchische Vorgehen bei der Lösung eines Entwurfsproblems nach dem Axiomatic Design wiederum entspricht den verschiedenen räumlichen Detaillierungsgraden.

Anhang A

Endliche Wege

Dies ist der in 4.5.3 angekündigte Beweis, dass jeder Weg in einem alexandroffschen Raum eine homotope endliche Abkürzung hat. Dazu werden zunächst einige Grundlagen bereitgestellt, die in der Arbeit bisher nicht erwähnt wurden. Dies sind insbesondere die Begriffe *Homotopie* und der Begriff des *kompakten* topologischen Raums, denn Kompaktheit wird in diesem Beweis verwendet. Wer dies bereits kennt, kann die Einführung überspringen und direkt zum Beweis in Abschnitt A.3 gehen.

A.1 Der Begriff des endlichen Wegs

Wollen wir zunächst unseren Begriff des endlichen Wegs wiederholen und uns dann überlegen, ob es unendliche Wege in alexandroffschen Räumen überhaupt gibt. Ansonsten würde der Satz ja keinen Sinn machen. Hier zunächst eine alternative Definition des endlichen Wegs:

Lemma A.1 (endlicher Weg). *Seien $\underline{X} = (X, \mathcal{T}_X)$ ein topologischer Raum und $w : [0, 1] \rightarrow \underline{X}$ ein Weg in \underline{X} . Dann ist w genau dann endlich, wenn es ein $n + 1$ -Tupel (a_0, a_1, \dots, a_n) von Punkten aus $[0, 1]$ mit $a_0 = 0$ und $a_n = 1$ gibt, so dass für jedes abgeschlossene Teilintervall $[a_{i-1}, a_i]$ für $i = 1, \dots, n$, die Abbildung*

$$w_i : [0, 1] \mapsto \underline{X}, t \mapsto w(a_{i-1} + t \cdot (a_i - a_{i-1}))$$

ein elementarer Weg ist.

Dies ist leicht einzusehen und wird hier nicht bewiesen. Wir erinnern uns, dass wir dann das Tupel (a_0, a_1, \dots, a_n) als einen *Pfad* zu diesem Weg bezeichnen.

Ein elementarer Weg verläuft also höchstens über drei Punkte: Er startet bei $w(0)$, verweilt dann ständig bei $w(0.5)$ und endet bei $w(1)$. Ein konstanter Weg ist ein besonderer elementarer Weg mit $w(0) = w(0.5) = w(1)$. Endliche Wege haben zudem eine endliche Spur und wechseln nur endlich oft zwischen konstanten Wegabschnitten und einzelnen Wegpunkten hin und her. Umgekehrt ist aber ein Weg mit endlicher Spur nicht notwendigerweise endlich.

Lemma A.2 (Unendliche Wege in alexandroffschen Räumen). *Das Einheitsintervall $[0, 1]$ bildet mit der trivialen Topologie $\mathcal{S} := \{\emptyset, [0, 1]\}$ einen*

alexandroffschen Raum $([0, 1], \mathcal{S})$. Dann ist die Abbildung

$$i : [0, 1] \rightarrow ([0, 1], \mathcal{S}), t \mapsto t$$

ein Weg von 0 nach 1 in $([0, 1], \mathcal{S})$ und nicht endlich, und es ist die Abbildung

$$d : [0, 1] \rightarrow ([0, 1], \mathcal{S}), t \mapsto \begin{cases} 1 & : t \in \mathbb{Q} \\ 0 & : t \notin \mathbb{Q} \end{cases}$$

zwar stetig, und somit ein Weg von 0 nach 1 mit endlichem Pfad $\{0, 1\}$, aber ebenfalls nicht endlich.

Die Abbildung i ist nicht endlich, da ihr Bild die unendliche Menge $[0, 1]$ ist. i ist in der Tat ein Weg, denn jede Abbildung in einen trivialen Raum hinein ist stetig. Es gibt aber den zu i homotopen elementaren Weg

$$w_{0,1} : [0, 1] \rightarrow ([0, 1], \{\emptyset, [0, 1]\}), t \mapsto \begin{cases} 0 & : t = 0 \\ 1 & : t > 0 \end{cases}$$

wie wir noch sehen werden.

Dass die zweite Abbildung eine stetige Abbildung sein soll, mag manchen zu Widerspruch veranlassen. Sie ist nämlich die auf $[0, 1]$ eingeschränkte sogenannte Dirichletfunktion und diese ist als reelle Abbildung bekanntermaßen nicht stetig. Dies gilt jedoch für die reellen Zahlen mit der *natürlichen* Topologie, wir verwenden hingegen als Bildbereich die reellen Zahlen mit der trivialen Topologie.

Zwei Wege sind nun zueinander homotop, wenn ein Weg stetig in den anderen Weg überführt werden kann. Ein Kennzeichen von Homotopie ist, dass zwei homotope Wege in einem Raum in der gleichen Weise an eventuell vorhandenen Hindernissen (bzw Löchern) vorbeilaufen. So sind bei einem Schirennen die Wege zweier Slalomläufer dann homotop, wenn sie in der selben Weise um die gesetzten Torstangen herumfahren. Wer ein Tor verpasst, darf zurücklaufen und es noch einmal versuchen, nicht jedoch einfach weiterfahren. Hier die Definition von Homotopie:

Definition A.1 ((relative) Homotopie). Seien \underline{X} und \underline{Y} topologische Räume und $f, g : \underline{X} \rightarrow \underline{Y}$ zwei stetige Abbildungen. Eine Homotopie von f nach g ist eine stetige Abbildung

$$h : [0, 1] \times \underline{X} \rightarrow \underline{Y}$$

mit $h(0, x) = f(x)$ und $h(1, x) = g(x)$ für alle $x \in \underline{X}$. Gilt für eine Teilmenge $A \subseteq \underline{X}$ zudem $h(s, a) = h(t, a)$ für alle $s, t \in [0, 1]$ und für alle $a \in A$, dann heißt h auch Homotopie relativ zu A . Die Abbildungen f und g heißen homotop (relativ zu A), wenn eine solche Homotopie (relativ zu A) von f nach g existiert. Sind zwei Abbildungen f und g zueinander homotop (relativ zu A), schreiben wir $f \simeq g$ (bzw. $f \simeq_A g$).

Dabei trägt $[0, 1] \times \underline{X}$ natürlich die Produkttopologie.

Eine Homotopie ist also eine stetige „Verschiebung“ einer Abbildung in eine andere. Wenn nun bestimmte Teile der Abbildung dabei nicht verschoben werden, dann nennt man dies „relativ“ zu nicht verschobenen Teilen. So ist die Bewegung eines Zeigers auf dem Ziffernblatt einer Uhr eine Homotopie relativ

zu dessen Mittelpunkt, wenn wir die Zeigerstellung als Abbildung der Punkte des Uhrzeigers auf die jeweils darunterliegenden Punkte des Ziffernblattes betrachten.

Eine Homotopie $h : [0, 1] \times \underline{X} \rightarrow \underline{Y}$ kann man wahlweise auch als eine Familie $h_i : \underline{X} \rightarrow \underline{Y}, x \mapsto h_i(x) := h(i, x)$ von stetigen Abbildungen mit Index $i \in [0, 1]$ auffassen, oder sich als Familie von Wegen $w_x : [0, 1] \rightarrow \underline{Y}, t \mapsto w_x(t) := h(t, x)$ mit Index in \underline{X} vorstellen.

Zwei Wege $w, v : [0, 1] \rightarrow \underline{X}$ bezeichnet man als homotop, wenn eine Homotopie $h : [0, 1] \times [0, 1] \rightarrow \underline{X}$ relativ zum Anfangs- und Endpunkt von w existiert. Bei Wegen fordert man also zusätzlich, dass die Homotopie den Anfangs- und den Endpunkt festhält und höchstens die Wegstrecke dazwischen verändert. Ansonsten wären nämlich alle Wege innerhalb einer Wegzusammenhangskomponente zueinander homotop, und daraus würde ein wenig interessanter Begriff resultieren. Wenn wir hier also sagen „Zwei Wege w und v sind zueinander homotop.“, dann meinen wir immer „homotop relativ $\{w(0), w(1)\}$ “. Um Weghomotopie von allgemeiner Homotopie zu unterscheiden, schreiben wir $w \cong v$ anstelle von $w \simeq_{\{w(0), w(1)\}} v$.

Auch bei einem Schirennen wird ja von jedem Teilnehmer verlangt, dass seine Fahrt am Startpunkt beginnt und im Zielpunkt endet. Alle fehlerfreien Abfahrten sind also homotop relativ zu Start- und Zielpunkt, sowohl die Idealinie als auch ein gemütliches Umrunden der Torstangen. Beim sog. „Einfädeln“ hingegen ist der Weg des linken Schis nicht homotop zum Weg des rechten Schis.

Homotopie bzw. relative Homotopie ist eine Äquivalenzrelation. Dies überlegt man sich mit Argumenten analog zur Wegzusammenhangsrelation, indem wir die Homotopie als Familie von Wegen betrachten:

Jede stetige Abbildung $f : \underline{X} \rightarrow \underline{Y}$ ist zu sich selbst homotop, denn die Projektion $\pi : [0, 1] \times \underline{X} \rightarrow \underline{X}, (t, x) \mapsto x$ ist bekanntermaßen stetig und somit ist auch die Abbildung $f \circ \pi : [0, 1] \times \underline{X} \rightarrow \underline{Y}$ stetig. Da $f \circ \pi(i, x) = f(x)$ für alle $i \in [0, 1]$ gilt, ist natürlich auch $f \circ \pi(0, x) = f(x)$ und $f \circ \pi(1, x) = f(x)$ und damit ist $f \circ \pi$ eine Homotopie von f nach f , analog zu einem konstanten Weg c_x von x nach x .

Für jede Homotopie $h : [0, 1] \times \underline{X} \rightarrow \underline{Y}$ von f nach g gibt es die Homotopie zurück in der Form

$$h^- : [0, 1] \times \underline{X} \rightarrow \underline{Y}, (t, x) \mapsto h(1 - t, x)$$

und wegen $h^-(0, x) = h(1, x)$ und $h^-(1, x) = h(0, x)$ ist dies eine Homotopie von g nach f .

Ist h_{df} Homotopie von d nach f und h_{fg} Homotopie von f nach g , dann ist

$$h_{df} * h_{fg} : [0, 1] \times \underline{X} \rightarrow \underline{Y}, (t, x) \mapsto \begin{cases} h_{df}(2t, x) & : x < 0.5 \\ h_{fg}(2t - 1, x) & : x \geq 0.5 \end{cases}$$

eine Homotopie von d über f nach g , wie wir noch sehen werden.

Damit formulieren wir nun unseren Satz:

Satz A.3 (endliche homotope Wege). Sei $\underline{X} = (X, \mathcal{T}_X)$ ein alexandroffscher Raum. Dann gibt es zu jedem Weg $w : [0, 1] \rightarrow \underline{X}$ von a nach b einen dazu homotopen endlichen Weg $s : [0, 1] \rightarrow \underline{X}$.

Bevor wir diesen Satz beweisen, werden wir einige für diesen Beweis benötigte Grundlagen bereitstellen.

A.2 Grundlagen

Zunächst erinnern wir noch einmal an die Definition der reellen Zahlen und beschäftigen uns mit den Grenzwerten von Folgen. Dann werden wir den Begriff des kompakten topologischen Raums einführen, eine Endlichkeitseigenschaft der Topologie. Die Grenzwertbetrachtungen benötigen wir, um zu zeigen, dass jedes abgeschlossene Intervall in den reellen Zahlen kompakt ist. Da dann auch der Definitionsbereich eines Weges stets kompakt ist, hat jeder Weg Endlichkeitseigenschaften, die wir zur Konstruktion des homotopen endlichen Weges ausnutzen werden.

A.2.1 Reelle Zahlen

Wir erinnern, dass die reellen Zahlen diejenigen Zahlen sind, in denen jede Cauchyfolge von Rationalzahlen auch einen Grenzwert hat. Die Berechnung eines *Limes* ist also erst in \mathbb{R} immer möglich.

Da man z.B. in \mathbb{Q} den Exponenten a^b definieren kann, mit $a^0 := 1$ und $a^{n+1} := a \cdot a^n$, wünscht man sich auch das Wurzelziehen als inversen Operator, was jedoch in \mathbb{Q} nicht immer funktioniert. Statt $r^n = q$ schreiben wir auch $r = \sqrt[n]{q}$ und es ist bekannt, dass nicht für jedes $q \in \mathbb{Q}$ und $n \in \mathbb{N}$ ein solches r existiert. Dann gibt es jedoch stets eine sogenannte Cauchyfolge von Rationalzahlen $(r_0, r_1, \dots, r_i, \dots)$, für deren einzelne Folgenglieder r_i sich jeweils die Ausdrücke r_i^n beliebig nahe an q annähern, obwohl keine Rationalzahl r mit $r^n = q$ existiert. Dabei konzentrieren sich die Folgenglieder auf einen immer kleiner werdenden Bereich, dessen „Durchmesser“ sich beliebig nahe an 0 annähert. Dennoch gibt für manche Cauchyfolgen es keine Rationalzahl, die sich durch alle diese ineinandergeschachtelten Bereiche festhalten lässt, jede geht irgendwann verloren — die Folge konvergiert nicht.

Die Menge \mathbb{R} der reellen Zahlen ist nun die Erweiterung der Rationalzahlen derart, dass jede Cauchyfolge einen Grenzwert besitzt und folglich konvergiert. Eine Folge der Form $\Pi = (3, 3.1, 3.14, 3.141, 3.1415, 3.14159, \dots)$, heißt z.B. Dezimalentwicklung und diese hier konvergiert nach π , wenn sie im Sinne des Autors fortgesetzt wird. Für jede Rationalzahl $q \in \mathbb{Q}$ konvergiert außerdem die konstante Folge (q, q, q, q, \dots) gegen q in \mathbb{Q} .

A.2.2 Folgen

Beginnen wir wieder mit den Begriffsdefinitionen und erinnern uns an unsere Definition einer Folge:

Definition A.2 (Folge). *Sei X eine nichtleere Menge. Eine Abbildung $a : \mathbb{N} \rightarrow X$ heißt auch Folge in X . Statt $a : \mathbb{N} \rightarrow X$ schreiben wir einfach $(a_i)_{i \in \mathbb{N}}$ oder kurz (a_i) und statt $a(n)$ schreiben wir a_n und nennen dies ein Folgenglied.*

Alle Verknüpfungen in X lassen sich gliedweise als Verknüpfungen in $X^{\mathbb{N}}$, der Menge aller Folgen in X fortsetzen, wie etwa Addition $(a_i) + (b_i) := (a_i + b_i)$ oder „Wrlmbrmft“ $(a_i) \heartsuit (b_i) := (a_i \heartsuit b_i)$, egal wie „Wrlmbrmft“ \heartsuit definiert ist.

Zur Definition der Cauchyfolgen benötigen wir einen Abstandsbegriff, sind also wieder bei der Topologie. Die Menge \mathbb{Q} ist mit der Definition eines Abstands d in \mathbb{Q} , definiert als $d : \mathbb{Q} \times \mathbb{Q} \rightarrow \mathbb{Q}, (a, b) \mapsto |a - b|$, ein metrischer Raum. Da

der Abstand zweier Rationalzahlen wiederum eine Rationalzahl ist, reicht es, sich auf \mathbb{Q} als Wertebereich zu beschränken. Dies verletzt zwar geringfügig die Definition von Metrik, aber wir wollen ja zunächst so tun, als ob wir die reellen Zahlen noch nicht kennen.

Eine Folge heißt nun Cauchyfolge, wenn es zu jedem Abstand $\varepsilon > 0$ in \mathbb{Q} ein Element $q \in \mathbb{Q}$ gibt, so dass die Folgeglieder irgendwann nur noch höchstens diesen Abstand zu q haben:

Definition A.3 (Cauchyfolge). Sei (a_i) eine Folge in X und (X, d) ein metrischer Raum. (a_i) heißt Cauchyfolge in (X, d) , wenn es zu jedem $\varepsilon > 0$ eine natürliche Zahl n_0 gibt, so dass für alle natürlichen Zahlen $i, j \in \mathbb{N}$ gilt: $d(a_{n_0+i}, a_{n_0+j}) \leq \varepsilon$.

Die Menge aller Cauchyfolgen in (X, d) wollen wir mit $C(X, d)$ oder einfach mit $C(X)$, wenn die Metrik für X klar ist, bezeichnen.

Diese Definition unterscheidet sich etwas von der einleitenden Charakterisierung mit der Existenz eines $q \in \mathbb{Q}$, in dessen Nähe die Folgeglieder irgendwann bleiben, ist aber dazu äquivalent. Zum Einen ist nämlich das a_{n_0} aus der obigen Definition eben dieses q und damit hat jede Cauchyfolge diese Eigenschaft. Umgekehrt ist jede Folge mit dieser Eigenschaft auch eine Cauchyfolge, denn wenn wir ein solches $\varepsilon > 0$ vorgeben, dann ist ja auch $\varepsilon/2 > 0$ ebenfalls ein solcher Abstand. Dann haben alle Folgeglieder „irgendwann“, also ab einem $n_0 \in \mathbb{N}$, höchstens einen Abstand von $\varepsilon/2$ zu einem solchen q . Wenn aber alle Folgeglieder a_{n_0+i} und a_{n_0+j} , die auf n_0 folgen, jeweils den Abstand $d(a_{n_0+i}, q) \leq \varepsilon/2$ und $d(q, a_{n_0+j}) \leq \varepsilon/2$ zu q haben, dann gilt wegen der Dreiecksungleichung auch

$$d(a_{n_0+i}, a_{n_0+j}) \leq d(a_{n_0+i}, q) + d(q, a_{n_0+j}) \leq \varepsilon.$$

Falls nun dieser Punkt q für alle $\varepsilon > 0$ fest gewählt werden kann, sagt man (a_i) konvergiert gegen $p \in X$ und schreibt $\lim(a_i) = p$. Wenn aber ein solcher Punkt nicht existiert, dann „konzentriert“ sich die Folge zwar auf einen immer kleiner werdenden Bereich, dessen „Durchmesser“ gegen Null geht, aber jedes Element darin geht uns irgendwann verloren, wenn der Bereich zu klein wird.

Es ist der Zweck der reellen Zahlen, für jede Cauchyfolge einen Limes zur Verfügung zu stellen und sich ansonsten wie die rationalen Zahlen zu verhalten. Der Trick ist nun, einfach diese Folge selbst zu ihrem Limes zu erklären, ein Trick, der deshalb funktioniert, weil für Cauchyfolgen eine Äquivalenzrelation existiert:

Definition A.4 (Äquivalenz von Cauchyfolgen). Zwei Cauchyfolgen (a_i) und (b_i) heißen äquivalent, wenn die Folge der gliedweisen Differenzen gegen Null konvergiert, wenn also gilt

$$\lim(a_i - b_i) = 0,$$

bzw. wenn $(a_i - b_i)$ eine Nullfolge ist.

Es ist leicht einzusehen, dass diese Relation reflexiv, symmetrisch und transitiv ist. Insbesondere ist die konstante Folge $(1)_{i \in \mathbb{N}}$ äquivalent zu Dezimalentwicklung $(0, 0.9, 0.99, 0.999, 0.9999, \dots)$, denn die gliedweisen Differenzen $(1, 0.1, 0.01, 0.001, 0.0001, \dots)$ konvergieren gegen 0.

Dann ist $\mathbb{R} := C(\mathbb{Q})/\sim$. Eine reelle Zahl $r \in \mathbb{R}$ ist also einfach die Menge der paarweise äquivalenten rationalen Cauchyfolgen. Man kann zeigen, dass es

reicht, sich bei den Cauchyfolgen auf die Dezimalentwicklungen zu beschränken, denn für jede reelle Zahl gibt es eine Dezimalentwicklung als Repräsentant und jede Dezimalentwicklung ist eine konvergente Folge, wie etwa bei $0.999\dots = 1.000\dots$. Daher gibt es auch die alternative Definition der reellen Zahlen als diejenigen Zahlen mit beliebiger Dezimalentwicklung $3.14159\dots$. Zudem kann man zeigen, dass reellen Cauchyfolgen immer einen Limes in \mathbb{R} haben. Damit ist eine entsprechende Erweiterung von \mathbb{R} nicht mehr nötig, bzw. man sagt \mathbb{R} ist bezüglich Grenzwertbildung abgeschlossen.

Wir erinnern uns an das auf Seite 26 gesagte, wonach Addition $a + b$ und Multiplikation $a \cdot b$ zweier reeller Zahlen $a = [(a_i)_{i \in \mathbb{N}}]$ und $b = [(b_i)_{i \in \mathbb{N}}]$ definiert sind, indem wir jeweils einen Repräsentanten $(a_i)_{i \in \mathbb{N}}$ und $(b_i)_{i \in \mathbb{N}}$ beliebig auswählen und diese dann gliedweise verknüpfen. Damit erhalten wir wiederum jeweils eine Cauchyfolge, deren Äquivalenzklasse die gesuchte Summe bzw. das Produkt ist. Es gilt also $a + b := [(a_i + b_i)_{i \in \mathbb{N}}]$ und $a \cdot b := [(a_i \cdot b_i)_{i \in \mathbb{N}}]$. Dabei erhalten wir in der Tat ein eindeutiges Ergebnis, die Verknüpfungen sind also wohldefiniert. Die Wohldefiniertheit der Addition hatten wir bereits anfangs gezeigt, die der Multiplikation zeigt man analog, muss dabei aber ein wenig mehr Aufwand treiben:

Zunächst zeigen wir, dass für alle $b \in \mathbb{R}$ die Gleichung $0 \cdot b = 0$ gilt, dass also das Produkt einer Nullfolge mit einer anderen Cauchyfolge wiederum eine Nullfolge ergibt.

Sei also (z_i) (als Kurzschreibweise für $(z_i)_{i \in \mathbb{N}}$) eine Nullfolge und (b_i) eine Cauchyfolge in \mathbb{Q} , und sein $\varepsilon > 0$ ein gegebener Abstand. Dann ist auch $1 > 0$ ein Abstand und somit gibt es eine Zahl n_0 und ein $b \in \mathbb{Q}$, so dass alle auf n_0 folgenden Folgeglieder b_{n_0+i} einen Abstand $|b - b_{n_0+i}| < 1$ haben. Damit liegen alle diese Folgeglieder im Intervall $[-|b| - 1, |b| + 1]$ und folglich ist der Betrag, also der Abstand zu 0, fast aller Folgeglieder durch $|b| + 1$ nach oben beschränkt: es gilt also $|b_{n_0+i}| \leq |b| + 1$. Wenn wir nun den Abstand ε durch diese Schranke dividieren erhalten wir mit

$$\delta := \frac{\varepsilon}{|a| + 1} > 0$$

einen weiteren Abstand, der größer als 0 ist, und damit gibt es ebenfalls ein m_0 , so dass alle Folgeglieder z_{m_0+i} mit $i \in \mathbb{N}$ in der δ -Umgebung von 0 liegen. Sei k_0 der größeren der beiden Indizes n_0 und m_0 . Dann gilt für die Beträge der Produkte

$$|z_{k_0+i} \cdot b_{k_0+i}| = |z_{k_0+i}| \cdot |b_{k_0+i}| \leq \delta \cdot (|b| + 1) = \varepsilon.$$

Damit ist für alle $i \in \mathbb{N}$ der Betrag $|z_{k_0+i} \cdot b_{k_0+i}|$ der auf k_0 folgenden Folgeglieder stets kleiner als ε , die Folge der Produkte ist also eine Nullfolge, denn der Betrag ist ja der Abstand zu 0.

Damit zeigen wir nun die Wohldefiniertheit der Multiplikation. Wenn wir also zwei Cauchyfolgen gliedweise Multiplizieren und dann eine der Folgen durch eine beliebige andere äquivalente Cauchyfolge ersetzen, dann erhalten wir auch ein äquivalentes Ergebnis.

Gelte also $(a_i) \sim (c_i)$ gilt, dann ist $(a_i - c_i)$ eine Nullfolge. Ersetzen wir nun a_i in $(a_i \cdot b_i)$ durch c_i , dann erhalten wir die Folge $(c_i \cdot b_i)$. Zu zeigen ist, dass deren Differenz $((a_i \cdot b_i) - (c_i \cdot b_i))$ ebenfalls gegen Null konvergiert. Es ist $((a_i \cdot b_i) - (c_i \cdot b_i)) = ((a_i - c_i) \cdot b_i)$. Dies ist das Produkt der Nullfolge $(a_i - c_i)$ mit einer weiteren Cauchyfolge (b_i) und somit ebenfalls eine Nullfolge. Also gilt $[(a_i \cdot b_i)] = [(c_i \cdot b_i)]$, die Multiplikation von reellen Zahlen ist wohldefiniert.

Nun noch ein Satz, den wir später im Beweis benötigen:

Satz A.4 (Konvergenz beschränkter monotoner Folgen). *Eine monoton steigende reelle Folge ist eine Folge (a_i) mit $a_i \leq a_{i+1}$ für alle i . Ist eine solche Folge nach oben beschränkt, also $a_i \leq S$ für ein festes $S \in \mathbb{R}$ und alle $i \in \mathbb{N}$, dann hat (a_i) einen Grenzwert $\lim(a_i)$ in \mathbb{R} . Dieser ist ebenfalls eine obere Schranke. Analoges gilt für monoton fallende Folgen mit unterer Schranke.*

Als Beweis ist zu zeigen, dass eine solche Folge konvergiert. Ist dies der Fall, dann ist diese Folge in $C(\mathbb{R})$ und hat somit einen Limes. Wir behaupten zunächst das Gegenteil und nehmen an, eine Folge (a_i) sei monoton steigend, habe obere Schranke S und sei dennoch nicht konvergent. Dann existiert ein $\varepsilon > 0$, so dass für jeden Punkt $p \in \mathbb{R}$ gilt, dass die Folgeglieder irgendwann das Intervall $]p - \varepsilon, p + \varepsilon[$ verlassen. Dies gilt insbesondere wenn wir $p = a_{n_0} = a_0$ wählen. Da (a_i) monoton steigt, sind alle Glieder $a_i \geq a_{n_0}$ (Beweis per Induktion). Nun gibt es ein a_{n_1} , das das Intervall $]a_{n_0} - \varepsilon, a_{n_0} + \varepsilon[$ verlässt und dann gilt also $a_{n_1} - p > \varepsilon$. Dann wählen wir uns dieses a_{n_1} als neues p und unsere Folge steigt ebenfalls so lange an, bis sie $]a_{n_1} - \varepsilon, a_{n_1} + \varepsilon[$ verlässt und wir ein a_{n_2} erhalten mit $a_{n_2} \geq a_{n_1} + \varepsilon$ und $a_{n_1} \geq a_{n_0} + \varepsilon$, also $a_{n_2} \geq a_{n_0} + \varepsilon + \varepsilon$. Da uns unsere Folgeglieder nie ausgehen, können wir für jedes $k \in \mathbb{N}$ ein $a_{n_k} \geq a_0 + k \cdot \varepsilon$ erreichen. Es gibt nun wegen $\varepsilon > 0$ eine natürliche Zahl s mit $s > (S - a_{n_0})/\varepsilon$ und damit $s \cdot \varepsilon > S - a_{n_0}$ und damit $a_{n_0} + s \cdot \varepsilon > S$. Das Folgeglied a_{n_s} ist aber wegen $a_{n_s} \geq a_0 + s \cdot \varepsilon$ größer als S , das damit keine obere Schranke der Folge ist. Also muss die Behauptung, es gebe eine monoton steigende Folge mit oberer Schranke, die nicht konvergiert, falsch sein. Damit gilt das Gegenteil. Dies ist aber die erste Aussage des Satzes der somit bewiesen ist. Es ist noch zu zeigen, dass der Limes einer monoton steigenden Folge ebenfalls eine obere Schranke ist. Angenommen nun $\lim(a_i)$ sei keine obere Schranke. Dann gibt es ein a_n mit $a_n > \lim(a_i)$. Wegen der Monotonie der Folge müssen dann alle weiteren Folgeglieder a_{n+i} um $\varepsilon := (a_n - \lim(a_i))$ größer als a_n sein. Damit konvergiert die Folge nicht gegen ihren eigenen Limes, was aber nicht sein kann, denn dies widerspricht der Definition des Limes. Also ist der Limes einer monotonen Folge eine Schranke.

A.2.3 Kompaktheit

Kompaktheit ist eine Endlichkeitseigenschaft von topologischen Räumen, die gewisse Beweise (wie diesen hier) ermöglicht. Es erscheint plausibel, dass Endlichkeitseigenschaften zum Beweis von Endlichkeitsaussagen nützlich sind:

Definition A.5 (Kompaktheit). *Sei $\underline{X} = (X, \mathcal{T}_X)$ ein topologischer Raum. Eine Teilmenge $\mathcal{C} \subseteq \mathcal{T}_X$ mit $\bigcup \mathcal{C} = X$ heißt offene Überdeckung von \underline{X} . Der Raum \underline{X} heißt kompakt, wenn zu jeder solchen offenen Überdeckung \mathcal{C} eine endliche Teilmenge $\{A_0, \dots, A_n\} \subseteq \mathcal{C}$ existiert, die ebenfalls Überdeckung von \underline{X} ist. Eine Teilmenge A eines topologischen Raums \underline{X} ist kompakt, wenn der Teilraum $\underline{X}|_A$ kompakt ist.*

Wir sind nun in der glücklichen Lage, dass die Spur eines Weges stets kompakt ist.

Satz A.5 (Kompaktheit von $[0, 1]$ und der Spur eines Weges). *Jedes abgeschlossene Intervall $[a, b]$ in \mathbb{R} (mit seiner natürlichen Topologie) ist kompakt.*

Das Bild einer kompakten Menge unter einer stetigen Abbildung ist ebenfalls kompakt.

Als Beweis nehmen wir eine beliebige offene Überdeckung \mathcal{C} von $[a, b]$. Für jeden Punkt $x \in [a, b]$ existiert dann eine offene Menge $A_x \in \mathcal{C}$ die x enthält. Dann existiert ein $\varepsilon > 0$, so dass $]x - \varepsilon, x + \varepsilon[\subseteq A_x$. Sei nun

$$\mathcal{B}_x := \{]x - \varepsilon, x + \varepsilon[\mid \varepsilon > 0, \text{ Es gibt ein } A \in \mathcal{C} \text{ mit }]x - \varepsilon, x + \varepsilon[\subseteq A\}$$

die Menge aller offenen Kugeln um x , die in irgend einem der $A_x \in \mathcal{C}$ liegen. Ist nun eines der $\varepsilon > b - x$, dann gibt es eine „Kugel“, die das gesamte restliche Intervall $[x, b]$ rechts von x überdeckt. Ansonsten gibt es ein eindeutig bestimmtes maximales ε_x mit $]x - \varepsilon_x, x + \varepsilon_x[\in \mathcal{B}_x$ und zu diesem gehört eine offene Menge $A_x \in \mathcal{C}$.

Nun definieren wir eine Folge, indem wir mit $a_0 := a$ starten. Damit haben wir ein ε_{a_0} und ein zugehöriges $A_{a_0} \in \mathcal{C}$. Zunächst vereinfachen wir unsere Schreibweise: $\varepsilon_i := \varepsilon_{a_i}$ und $A_i := A_{a_i}$.

Dann setzen wir $a_{i+1} := \min(b, a_i + \varepsilon_i)$. Und behaupten, dass diese Folge nach b konvergiert. Man überzeugt sich leicht, dass diese Folge (vorausgesetzt, es ist $a \leq b$) monoton steigend und durch b nach oben beschränkt ist. Dann hat diese Folge einen Grenzwert $\lim(a_i) \leq b$. Wir zeigen nun, dass dieser Grenzwert nicht kleiner b sein kann. Denn sei x dieser Grenzwert mit $x < b$, dann ist x obere Schranke der Folge (a_i) . Wegen $x \in [a, b]$ haben wir ein (maximales) $\varepsilon_x > 0$, so dass gilt $]x - \varepsilon_x, x + \varepsilon_x[\subseteq A_x \in \mathcal{C}$. Wegen der Konvergenz von (a_i) wissen wir, dass dann für ein $n \in \mathbb{N}$ gilt $a_n \in]x - \varepsilon_x/2, x + \varepsilon_x/2[$. Damit ist aber $a_{n+1} > x$, denn $x - \varepsilon_x/2 < a_n \leq x < a + \varepsilon_x/2$ also $x < a_n + \varepsilon_x/2$ und somit ist $a_n - \varepsilon_x/2 < x < a_n + \varepsilon_x/2$ und damit gilt $\varepsilon_x/2 \leq \varepsilon_n$ und folglich $a_{n+1} = a_n + \varepsilon_n \geq a_n + \varepsilon_x/2 > x$. Wir haben also $a_{n+1} > x$ und somit ist x keine obere Schranke der Folge. Dies ist jedoch ein Widerspruch, denn als Grenzwert einer monoton steigenden Folge muss x obere Schranke sein.

Wenn jedoch die Folge (a_i) den Grenzwert b hat, dann gibt es für b ebenfalls eine Umgebung $A_b \in \mathcal{C}$, und somit auch ein $\varepsilon_b > 0$ mit $]b - \varepsilon_b, b + \varepsilon_b[\subseteq A_b$. Damit gibt es ein n mit $a_n \in]b - \varepsilon_b/n, b + \varepsilon_b/n[$ und somit $a_{n+1} = b$. Für die endlich vielen Folgeglieder $a_0, \dots, a_n, a_{n+1} = b$ haben wir endlich viele zugeordnete Umgebungen A_0, \dots, A_n, A_b und jeder Punkt $x \in [a, b]$ liegt in einer dieser Umgebung, denn entweder ist $x = a_i$ für ein $i = 0, \dots, n + 1$, dann ist $x \in A_i$ oder x liegt zwischen zwei Folgegliedern a_i, a_{i+1} . Doch dann ist $a_i < x < a_{i+1} = a_i + \varepsilon_i$ und damit gilt $x \in]a_i - \varepsilon_i, a_i + \varepsilon_i[\subseteq A_i$, also kurz $x \in A_i$. Damit sind die endlich vielen $\{A_0, \dots, A_{n+1}\}$ eine Überdeckung von $[a, b]$, also ist $[a, b]$ kompakt und insbesondere ist dies auch $[0, 1]$.

Die Spur eines Weges ist nun ebenfalls kompakt, denn das Bild einer kompakten Menge unter einer stetigen Abbildung ist wiederum kompakt.

Zum Beweis nehmen wir an, \underline{X} sei ein kompakter Raum und $f : \underline{X} \rightarrow \underline{Y}$ eine stetige Abbildung. Dann ist $f[\underline{X}] \subseteq \underline{Y}$, das Bild dieses Raumes unter f . Wir nehmen zunächst an, dass $f[\underline{X}]$ gleich \underline{Y} ist. Ist nun \mathcal{D} eine offene Überdeckung von \underline{Y} , dann ist $\mathcal{C} := \{f^{-1}[B] \mid B \in \mathcal{D}\}$ eine offene Überdeckung von \underline{X} . Es überdeckt \underline{X} , denn ist $x \in \underline{X}$, dann ist $f(x) \in \underline{Y}$, also ist $f(x) \in B_{f(x)}$ für ein $B_{f(x)} \in \mathcal{D}$ und somit $x \in f^{-1}[B_{f(x)}] \in \mathcal{C}$. Zudem ist wegen der Stetigkeit von f jedes $f^{-1}[B] \in \mathcal{C}$ offen. Da \underline{X} kompakt ist, gibt es eine endliche Teilüberdeckung $\{f^{-1}[B_0], \dots, f^{-1}[B_n]\} \subseteq \mathcal{C}$. Da $f[\underline{X}] = \underline{Y}$ angenommen wurde, gilt für jede Teilmenge $B \subseteq \underline{Y}$, dass $f[f^{-1}[B]] = B$ gilt. Denn ist $p \in f[f^{-1}[B]]$, dann gibt es

ein $x \in f^{-1}[B]$ so dass $f(x) = p$ gilt. Dann gilt wegen $x \in f^{-1}[B]$ auch $f(x) \in B$ also $p \in B$ also gilt $f[f^{-1}[B]] \subseteq B$. Ist umgekehrt $p \in B$, dann ist $p \in \underline{Y}$ und dann gilt wegen $f[\underline{X}] = \underline{Y}$ auch $p \in f[\underline{X}]$. Dann gibt es ein $x \in \underline{X}$ mit $f(x) = p$ und wegen $p \in B$ gilt $x \in f^{-1}[B]$. Damit ist $f(x) \in f[f^{-1}[B]]$ und damit $B \subseteq f[f^{-1}[B]]$. Also ist $\{f[f^{-1}[B_0]], \dots, f[f^{-1}[B_n]]\} = \{B_0, \dots, B_n\} \subseteq \mathcal{D}$. Diese Menge $\{B_0, \dots, B_n\}$ ist eine offene Überdeckung von \underline{Y} . Es ist klar, dass sie nur offene Mengen enthält, denn sie ist ja Teilmenge einer offenen Überdeckung. Zudem ist sie eine Überdeckung, denn ist $y \in \underline{Y}$, dann ist $y \in f[\underline{X}]$ und somit gibt es ein $x \in \underline{X}$ mit $f(x) = y$. Dieses x liegt in einer der endlich vielen Mengen $f^{-1}[B_i]$ und damit liegt y in B_i .

Es bleibt noch zu zeigen, dass dies auch gilt, wenn $f[\underline{X}] \neq \underline{Y}$, wenn f also nicht surjektiv ist. In diesem Fall betrachten wir die surjektive Abbildung $s : \underline{X} \rightarrow f[\underline{X}], x \mapsto f(x)$ und die injektive Abbildung $i : f[\underline{X}] \rightarrow \underline{Y}, y \mapsto y$, wobei $f[\underline{X}]$ Teilraum von \underline{Y} ist: $f[\underline{X}] := \underline{Y}|_{f[\underline{X}]}$. Dann sind i und s stetig und es ist $f = i \circ s$. Nun ist \mathcal{C} genau dann eine offene Überdeckung von $f[\underline{X}]$, wenn es auch eine offene Überdeckung von $s[\underline{X}]$ ist. Als Überdeckung ist dies klar, denn als Menge ist ja $f[\underline{X}] = s[\underline{X}]$. Aber $s[\underline{X}]$ trägt auch die selbe Teilraumtopologie wie $f[\underline{X}]$ und damit ist eine Menge genau dann offen in $s[\underline{X}]$, wenn sie offen in $f[\underline{X}]$ ist.

A.3 Der Beweis

Weitere Hilfssätze des Beweises werden im weiteren Verlauf gegeben. Zunächst soll die Beweisidee angegeben werden:

Zuerst wird beobachtet, dass das Bild (die Spur) eines Weges in einem alexandroffschen Raum ebenfalls alexandroffsch ist. Diese Spur kann also mit offenen Sternen überdeckt werden. Da die Spur eines Weges kompakt ist, reichen dazu endlich viele Sterne. Die Urbilder dieser endlich vielen Sterne sind offene Teilmengen von $[0, 1]$, jede davon eine Vereinigung von offenen Kugeln (bzw. Intervallen) $]x - \varepsilon, x + \varepsilon[$. Das Bild jedes dieser Intervalle liegt also vollständig in einem der endlich vielen Sterne. Diese offenen Intervalle wiederum sind eine offene Überdeckung von $[0, 1]$ und damit gibt es eine endliche Teilüberdeckung davon. Dann gibt es auch eine minimale endliche Teilüberdeckung in der keines der Intervalle fehlen darf. Dann kann angenommen werden, dass jeder Punkt in $[0, 1]$ in höchstens zwei Intervallen liegt. Damit wechselt der Weg endlich oft zwischen jeweils genau einem Intervall und einem Bereich, an dem sich je zwei benachbarte Intervalle überlappen, hin und her. In jeder Überlappung wählen wir nun einen Punkt aus, an dem wir zunächst für den Stern, den wir verlassen werden, und dann den Stern, den wir betreten haben, je einen elementaren Schritt vom Wegpunkt direkt zum Zentrum des Sternes und sofort zurückgehen. Der so modifizierte Weg ist homotop zum ursprünglichen Weg. Damit wechselt der Weg endlich oft zwischen endlichen Abschnitten und Rundwegen innerhalb eines Sternes ab. Ein Rundweg innerhalb eines Sternes kann jedoch entfernt werden und damit bleiben nur noch endlich viele elementare Wegabschnitte übrig.

Damit ist die Aussage bereits bewiesen und der Rest dieses Anhangs führt die obige Argumentation nur noch einmal ausführlich Schritt für Schritt aus und führt dabei weitere dazu nötigen Grundlagen ein:

Lemma A.6. *Ist \underline{X} ein alexandroffscher Raum, und $w : [0, 1] \rightarrow \underline{X}$ ein Weg,*

dann ist die Spur von w ein kompakter alexandroffscher Raum.

Die Spur eines Weges in einen alexandroffschen Raum ist ein Teilraum davon. Wir haben ja bei dem Abfragen von topologischen Datenbanken die Konstruktion der Relation des Teilraums angegeben und so gezeigt, dass dieser dann ebenfalls alexandroffsch ist. Beim Beweis der Korrektheit der Konstruktion haben wir an keiner Stelle die Endlichkeit von \underline{X} vorausgesetzt. Dass die Spur eines Weges zudem kompakt ist, haben wir ebenfalls gezeigt.

Lemma A.7. *Ein kompakter alexandroffscher Raum $\underline{X} = (X, \mathcal{T}_R)$ mit topologischem Datentyp (X, R) hat eine endliche Teilmenge $\{x_0, \dots, x_n\} \subseteq X$, so dass $\{\text{St}_R^*(x_0), \dots, \text{St}_R^*(x_n)\}$ eine offene Überdeckung von \underline{X} ist.*

Der Beweis ist einfach:

Beweis. Wegen $x \in \text{St}_R^*(x)$ gilt, dass $\{\text{St}_R^*(x) \mid x \in \underline{X}\}$ eine offene Überdeckung von \underline{X} ist, denn Sterne sind ja bekanntermaßen offen. Damit gibt es eine endliche offene Teilüberdeckung $\{\text{St}_R^*(x_0), \dots, \text{St}_R^*(x_n)\}$. \square

Die Urbilder $\mathcal{C} = \{w^{-1}[\text{St}_R^*(x_0)], \dots, w^{-1}[\text{St}_R^*(x_n)]\}$ dieser Sterne unseres obigen Lemmas sind dann wiederum eine offenen Überdeckung von $[0, 1]$. Damit ist auch die Menge

$$\mathcal{B} := \{]a - \varepsilon, a + \varepsilon[\subseteq \text{St} \mid a \in [0, 1], \varepsilon > 0, \text{St} \in \mathcal{C}\}$$

eine offene Überdeckung von \mathcal{B} . Das sind alle offenen Intervalle, deren Bild vollständig in einem der Sterne $\text{St}_R^*(x) \in \mathcal{C}$ liegt. Diese wiederum haben wegen der Kompaktheit von $[0, 1]$ eine endliche offene Teilüberdeckung

$$\{]a_1 - \varepsilon_1, a_1 + \varepsilon_1[, \dots,]a_n - \varepsilon_n, a_n + \varepsilon_n[\}.$$

Wir können nun voraussetzen, dass jeder Punkt in $[0, 1]$ in höchstens zwei dieser Intervalle vorkommt¹. Denn gilt dies nicht, dann gibt es einen Punkt p , der in drei (oder mehr) Intervallen liegt. Es gilt also $p \in]a_1, b_1[$ und $p \in]a_2, b_2[$ und zudem $p \in]a_3, b_3[$. Dann gibt es ein Minimum der drei unteren Intervallgrenzen a_1, a_2, a_3 und ein Maximum der oberen Grenzen b_1, b_2, b_3 . Sei a_m dieses Minimum und b_n dieses Maximum. Dann ist jeder Punkt q aus einem der Intervalle, der nicht in $]a_m, b_m[$ liegt, Element des Intervalls $]a_n, b_n[$. Denn q ist dann größer als b_m , da a_m ja minimal ist. Damit ist q größer als p , da p ja in allen drei Intervallen liegt und somit gilt $a_n < p < b_m < q < b_n$, also $a_n < q < b_n$ und damit $q \in]a_n, b_n[$. Dann ist jedoch

$$]a_1, b_1[\cup]a_2, b_2[\cup]a_3, b_3[=]a_m, b_m[\cup]a_n, b_n[,$$

also reicht es anstelle der drei Intervalle nur die zwei „äußeren“ Intervalle auf der rechten Seite der Gleichung zu vereinigen.

Wir können daher schrittweise von unserer endlichen offenen Überdeckung jeweils ein überflüssiges Intervall entfernen, bis wir eine minimale endliche Überdeckung mit offenen Intervallen erhalten. Dabei ist leicht einzusehen, dass die Grenzen 0 und 1 in genau in einem Intervall liegen, denn ist $0 \in]a, b[$, dann ist

¹Dies nennt man den *Grad* einer Überdeckung und wird auch zur Definition der sogenannten Überdeckungsdimension herangezogen

$a < 0 < b$ und somit $]a, b[= [0, b[$. Ist nun 0 in einem weiteren Intervall $]c, d[$ dann gilt ebenfalls $]c, d[= [0, d[$. Damit ist eines der beiden eine Teilmenge des anderen Intervalls und sind nun $]0, b[$ und $]0, d[$ verschieden, dann entfernen wir einfach das kleinere der beiden.

Um unsere Konstruktion des endlichen homotopen Wegs vornehmen zu können, benötigen wir einige wohlbekanntete Eigenschaften homotoper Wege. Wir erinnern uns, dass „homotoper Weg“ immer bedeutet homotop relativ zu Start- und Zielpunkt.

Lemma A.8 (Eigenschaften homotoper Wege). *Sei $a, b \in [0, 1]$. Ein Weg $w_{ab} : [0, 1] \rightarrow [0, 1]$ von a nach b ist immer homotop zur linearen Interpolation von a nach b , also zu $i_{ab} : [0, 1] \rightarrow [0, 1], t \mapsto a + t \cdot (b - a)$. Insbesondere gilt dann*

- *Ein Weg $w : [0, 1] \rightarrow [0, 1]$ in $[0, 1]$ von 0 nach 1 ist homotop zum identischen Weg $\text{id} : [0, 1] \rightarrow [0, 1], t \mapsto t$.*
- *Ein Rundweg $r : [0, 1] \rightarrow [0, 1]$ von x nach x in $[0, 1]$ ist homotop zum konstanten Weg c_x .*

Ist $f : \underline{X} \rightarrow \underline{Y}$ eine stetige Abbildung und sind w_{ab} und v_{ab} homotope Wege in \underline{X} , dann sind $f \circ w_{ab}$ und $f \circ v_{ab}$ homotope Wege in \underline{Y} .

*Ist w_{ab} Weg von a nach b und w_{bc} ein Weg von b nach c , und sind v_{ab} und v_{bc} jeweils homotop zu w_{ab} und w_{bc} dann ist auch $w_{ab} * w_{bc}$ homotop zu $v_{ab} * v_{bc}$.*

*Die Komposition $w_{ab} * w_{ab}^-$ eines Weges w_{ab} von a nach b mit seinem Rückweg w_{ab}^- von b nach a ist homotop zum konstanten Weg $c_a : [0, 1] \rightarrow \underline{X}, t \mapsto a$.*

Ein Rundweg $w_{aa} : [0, 1] \rightarrow \underline{X}$ in einem alexandroffschen Raum \underline{X} , der innerhalb eines Sterns $\text{St}(x)$ für ein $x \in \underline{X}$ verläuft, ist homotop zum konstanten Weg c_a .

Die ersten vier Aussagen besagten, dass es bezüglich Homotopie egal ist, in welcher Geschwindigkeit ein Weg gegangen wird, man darf auch stehenbleiben und sogar beliebige Abschnitte vor- und zurücklaufen. Das heißt auch, dass es bis auf Homotopie beliebig viele Möglichkeiten gibt, die Komposition von Wegen zu definieren. Bisher haben wir immer gesagt, dass der erste Weg mit doppelter Geschwindigkeit durchlaufen wird, und von der Mitte der Zeit an der zweite Weg ebenfalls mit doppelter Geschwindigkeit fortgesetzt wird. Man kann aber auch z.B. den ersten Weg in 42-facher Geschwindigkeit innerhalb von $[0, \frac{1}{42}]$, dann den zweiten Weg in 84-facher Geschwindigkeit innerhalb von $[\frac{1}{42}, \frac{3}{84}]$ durchheilen und dann in der restlichen Zeit $[\frac{3}{84}, 1]$ einfach am Zielpunkt stehend abwarten. Betrachten wir homotope Wege als „gleich“, dann ist also der konstante Weg neutral bezüglich der Komposition, also gilt, $w_{ab} * c_b$ ist homotop zu w_{ab} und zu $c_a * w_{ab}$. Außerdem ist dann die Komposition von Wegen bis auf Homotopie assoziativ: es ist also $(w_{ab} * w_{bc}) * w_{cd}$ homotop zu $w_{ab} * (w_{bc} * w_{cd})$, denn diese beiden Wege unterscheiden sich nur in der Geschwindigkeit, in der sie durchlaufen werden.

Wenn wir einen Weg mit seinem Rückweg verketteten, dann hätten wir bis auf Homotopie auch einfach stehenbleiben können, erhalten also den konstanten Weg. Da dieser zudem (wieder bis auf Homotopie) „neutral“ ist, haben wir mit dem Rückweg einen bis auf Homotopie inversen Weg bezüglich der Komposition.

Weiterhin können endlich viele Wegabschnitte durch dazu homotope Wegabschnitte ersetzt werden und ergeben so wieder einen homotopen Weg.

Die letzte Aussage ist wichtig, da wir mit ihr die Endlichkeit konstruieren. Sobald nämlich unsere potenziell unendlichen Wegabschnitte nur noch Rundwege in Sternen sind, können wir sie durch konstante und somit endliche Wegabschnitte ersetzen. Statt also innerhalb eines Sternes herumzulaufen, können wir auch einfach darin stehenbleiben.

Beweis. Zunächst halten wir fest, dass für zwei reelle stetige Abbildungen $f : \mathbb{R} \rightarrow \mathbb{R}$ und $g : \mathbb{R} \rightarrow \mathbb{R}$ auch die Abbildungen Summe $f + g$, Differenz $f - g$ und (reelles) Produkt $f \cdot g$ stetig sind. Dies ist allgemein bekannt und wird hier nicht bewiesen. Dann ist die Interpolation von a nach b als eine solche Verknüpfung von reellen stetigen Abbildungen darstellbar: die konstanten Abbildungen c_a und c_b und die identische Abbildung $\text{id} : \mathbb{R} \rightarrow \mathbb{R}$. Wegen der Abbildungsvorschrift $a + t \cdot (b - a)$ von i gilt: $i = c_a + \text{id} \cdot (c_b - c_a)|_{[0,1]}$ ist stetig. Zudem gilt $i(0) = a + 0 \cdot (b - a) = a$ und $i(1) = a + 1 \cdot (b - a) = a + b - a = b$, also ist i ein Weg von a nach b .

Für die ersten drei Aussagen beweisen wir einfach, dass je zwei Wege w und v in $[0, 1]$ mit je gleichem Start- und gleichem Zielpunkt zueinander homotop sind. Dann ist jeder Weg von 0 nach 1 homotop zur Identität und jeder Rundweg mit Basispunkt x homotop zum konstanten Weg c_x .

Sind $w, v : [0, 1] \rightarrow [0, 1]$ zwei Wege von x nach y , dann sind w und v stetige Abbildungen mit $w(0) = x = v(0)$ und $w(1) = y = v(1)$. Wir interpolieren einfach für jedes $i, t \in [0, 1]$ die Wegpunkte $w(t)$ linear zu $v(t)$ mit Parameter i . Es ist $h_i(t) := w(t) + i \cdot (v(t) - w(t))$ diese Interpolation. Wegen $w(t) \in [0, 1]$ und $v(t) \in [0, 1]$ ist für jedes $i \in [0, 1]$ auch die Interpolation $h_i(t)$ zwischen $w(t)$ und $v(t)$ und somit im Intervall $[0, 1]$. Damit haben wir die Abbildung

$$h : [0, 1] \times [0, 1] \rightarrow [0, 1], (i, t) \mapsto h_i(t) := w(t) + i \cdot (v(t) - w(t)).$$

Dann gilt wegen $i = \pi_1(i, t)$ und $t = \pi_2(i, t)$ für die Abbildung

$$h(i, t) = w(\pi_2(i, t)) + \pi_1(i, t) \cdot (v(\pi_2(i, t)) - w(\pi_2(i, t)))$$

und somit ist h als eine algebraische Verknüpfung

$$h = w \circ \pi_2 + \pi_1 \cdot (v \circ \pi_2 - w \circ \pi_2)$$

von stetigen reellen Abbildungen

$$\begin{aligned} v \circ \pi_2 &: [0, 1] \times [0, 1] \rightarrow [0, 1] \\ w \circ \pi_2 &: [0, 1] \times [0, 1] \rightarrow [0, 1] \\ \pi_1 &: [0, 1] \times [0, 1] \rightarrow [0, 1] \end{aligned}$$

darstellbar und damit selbst stetig. Zudem gelten

$$\begin{aligned} h_0(t) &= w(t) + 0 \cdot (v(t) - w(t)) = w(t) \\ h_1(t) &= w(t) + 1 \cdot (v(t) - w(t)) = w(t) + v(t) - w(t) = v(t) \end{aligned}$$

und damit ist h Homotopie von w nach v . Gleichzeitig ist für beliebiges i

$$\begin{aligned} h_i(0) &= w(0) + i \cdot (v(0) - w(0)) = x + i \cdot (x - x) = x + 0 = x \\ h_i(1) &= w(1) + i \cdot (v(1) - w(1)) = y + i \cdot (y - y) = y + 0 = y. \end{aligned}$$

Also hält h den Anfangs- und Endpunkt fest, ist also relativ zum Anfangs- und Endpunkt von w bzw. v und damit eine Homotopie von Wegen.

Zeigen wir nun, dass die Komposition einer Abbildung mit je zwei homotopen Wegen wiederum zwei homotope Wege ergibt. Sind also w und v homotope Wege in \underline{X} und ist $f : \underline{X} \rightarrow \underline{Y}$ stetig, dann existiert eine Homotopie $h : [0, 1] \times [0, 1] \rightarrow \underline{X}$ von w nach v , relativ zu Start- und Endpunkt von w . Dann ist $f \circ h$ stetig, denn f und h sind ja stetig. $f \circ h(0, t) = f(h(0, t)) = f(w(t)) = f \circ w(t)$, also ist $(f \circ h)_0 = f \circ w$. $f \circ h(1, t) = f(h(1, t)) = f(v(t)) = f \circ v(t)$, also ist $(f \circ h)_1 = f \circ v$. Damit ist $f \circ h$ eine Homotopie von $f \circ w$ nach $f \circ v$. Sie ist auch relativ bezüglich Anfangs- und Endpunkt, denn $f \circ h(i, 0) = f(h(i, 0)) = f(w(0)) = f \circ w(0) = f \circ v(0)$ unabhängig von der Wahl von i . Analog zeigt man dies für den Endpunkt $w(1)$.

Nun zeigen wir noch, dass homotope Wegabschnitte jeweils ausgetauscht werden können: Sei w_{ab} ein Weg von a nach b , w_{bc} ein Weg von b nach c und v_{ab} und v_{bc} zu diesen jeweils homotop.

Dann existieren zwei Homotopien

$$\begin{array}{ll} h_{ab} : [0, 1] \times [0, 1] \rightarrow \underline{X} & \text{von } w_{ab} \text{ nach } v_{ab} \\ h_{bc} : [0, 1] \times [0, 1] \rightarrow \underline{X} & \text{von } w_{bc} \text{ nach } v_{bc} \end{array}$$

Dann definieren wir die Abbildung

$$H : [0, 1] \times [0, 1] \rightarrow \underline{X}, (i, t) \mapsto \begin{cases} h_{ab}(i, 2t) & : t \leq 0.5 \\ h_{bc}(i, 2t - 1) & : t \geq 0.5, \end{cases}$$

die „Komposition von Homotopien“ ganz analog zur Komposition von Wegen, und behaupten, dies sei eine Homotopie von $w_{ab} * w_{bc}$ nach $v_{ab} * v_{bc}$. Die durch H erzeugte Familie von Abbildungen $(H_i)_{i \in [0, 1]}$, definiert durch $H_i(t) := H(i, t)$ „beginnt“ in der Tat bei $i = 0$ mit dem Weg $H_0 = w_{ab} * w_{bc}$ und „endet“ mit $H_1 = v_{ab} * v_{bc}$, wie man leicht durch Einsetzen der entsprechenden Werte für i sieht. Es ist ja z.B. $h_{ab}(0, t) = w_{ab}(t)$ da h_{ab} ja eine Homotopie von w_{ab} nach v_{ab} ist.

Zu zeigen ist noch, dass H zudem eine stetige Abbildung ist: Betrachten wir zunächst zwei Einschränkungen von H :

$$\begin{aligned} H_{ab}(i, t) &:= H|_{[0, 1] \times [0, 0.5]}(i, t) := h_{ab}(i, 2t) \\ H_{bc}(i, t) &:= H|_{[0, 1] \times [0.5, 1]}(i, t) := h_{bc}(i, 2t - 1) \end{aligned}$$

Dann haben wir nur in h das zweite Argument t jeweils durch einen algebraischen Ausdruck $2t$ bzw. $2t - 1$ ersetzt, und das erste Argument wurde unverändert gelassen. Diese Ausdrücke definieren die stetigen Abbildungen

$$\begin{aligned} \lambda_{2t} &: [0, 0.5] \rightarrow [0, 1], t \mapsto 2t \\ \lambda_{2t-1} &: [0.5, 1] \rightarrow [0, 1], t \mapsto 2t - 1 \\ \text{id}_{[0, 1]} &: [0, 1] \rightarrow [0, 1]. \end{aligned}$$

Nun ist das direkte Produkt $f \times g : X \times Y \rightarrow A \times B$ zweier stetiger Abbildungen $f : X \rightarrow A$ und $g : Y \rightarrow B$ ebenfalls eine stetige Abbildung zwischen den Produkträumen. Ist nämlich eine Menge U offen in $A \times B$, dann ist sie laut Definition der Produkttopologie die Vereinigung einer Menge

$$\mathcal{U} = \{U_A \times U_B \subseteq U \mid U_A \in \mathcal{T}_A, U_B \in \mathcal{T}_B\}$$

von kartesischen Produkten von offenen Mengen in A und B , gelegentlich auch als „offene Kästchen“ bezeichnet. Das Urbild eines solchen Kästchens unter dem direkten Produkt zweier Abbildungen ist aber gleich dem kartesischen Produkt der Urbilder der Faktoren, also

$$(f \times g)^{-1}[U_A \times U_B] = f^{-1}[U_A] \times g^{-1}[U_B],$$

und da f und g stetig sind ist dies das kartesische Produkt zweier offener Mengen und damit ebenfalls ein „offenes Kästchen“ und folglich offen im Produktraum. Das Urbild der Vereinigung von Mengen ist zudem die Vereinigung der Urbilder der Mengen. Also ist das Urbild von U als Vereinigung von offenen Kästchen gleich der Vereinigung der Urbilder all dieser Kästchen in U . Da diese offen sind, ist auch $(f \times g)^{-1}[U]$ als deren Vereinigung offen und damit ist $(f \times g)$ stetig.

Es sind also die direkten Produkte

$$\begin{aligned} \text{id}_{[0,1]} \times \lambda_{2t} : [0, 1] \times [0, 0.5] &\rightarrow [0, 1] \times [0, 1], (i, t) \mapsto (i, 2t) \\ \text{id}_{[0,1]} \times \lambda_{2t-1} : [0, 1] \times [0.5, 1] &\rightarrow [0, 1] \times [0, 1], (i, t) \mapsto (i, 2t - 1) \end{aligned}$$

ebenfalls stetige Abbildungen. Nun gilt aber

$$\begin{aligned} H_{ab}(i, t) &= h_{ab}(i, 2t) \\ &= h_{ab}(\text{id}_{[0,1]}(i), \lambda_{2t}(t)) \\ &= h_{ab}(\text{id}_{[0,1]} \times \lambda_{2t}(i, t)) \\ &= h_{ab} \circ (\text{id}_{[0,1]} \times \lambda_{2t})(i, t) \\ H_{bc}(i, t) &= h_{bc}(i, 2t - 1) \\ &= h_{bc}(\text{id}_{[0,1]}(i), \lambda_{2t-1}(t)) \\ &= h_{bc}(\text{id}_{[0,1]} \times \lambda_{2t-1}(i, t)) \\ &= h_{bc} \circ (\text{id}_{[0,1]} \times \lambda_{2t-1})(i, t). \end{aligned}$$

Damit ist H_{ab} die Komposition $H_{ab} = h_{ab} \circ (\text{id}_{[0,1]} \times \lambda_{2t})$ zweier stetiger Abbildungen und damit selbst stetig. Das selbe gilt für $H_{bc} = h_{bc} \circ (\text{id}_{[0,1]} \times \lambda_{2t-1})$.

Wir haben also zwei stetige Abbildungen H_{ab} und H_{bc} mit gemeinsamem Definitionsbereich

$$[0, 1] \times [0, 0.5] \cap [0, 1] \times [0.5, 1] = [0, 1] \times \{0.5\}.$$

und entsprechend gemeinsamer Einschränkung

$$H_{ab}|_{[0,1] \times \{0.5\}} = H_{bc}|_{[0,1] \times \{0.5\}}.$$

Damit ist

$$H = H_{ab} \cup H_{bc}$$

die gemeinsame Fortsetzung zweier stetiger Abbildungen. Nun ist der Definitionsbereich jeder dieser beiden Abbildungen abgeschlossen in deren Vereinigung. In anderen Worten, es ist sowohl $[0, 1] \times [0, 0.5]$ als auch $[0, 1] \times [0.5, 1]$ eine abgeschlossene Teilmenge von $[0, 1] \times [0, 1]$. Damit ist H als deren gemeinsame Fortsetzung ebenfalls stetig. Dies gilt ganz allgemein, dann sind A und B jeweils abgeschlossene Teilmengen eines Raums \underline{X} , sind zudem $f_A : \underline{A} \rightarrow \underline{Y}$ und $f_B : \underline{B} \rightarrow \underline{Y}$ stetige Einschränkungen einer Abbildung $f : X \rightarrow Y$, dann ist deren gemeinsame Fortsetzung $f_A \cup f_B$ eine stetige Abbildung von $A \cup B$ nach \underline{Y} .

Wir setzen hier $X = A \cup B$, denn dies ist keine echte Einschränkung, und nur dieser Fall interessiert uns hier. Jede offene Menge $U \subseteq Y$ hat ein offenes Urbild $f_A^{-1}[U]$ in \underline{A} und ein offenes Urbild $f_B^{-1}[U]$ in \underline{B} . Wobei \underline{A} und \underline{B} jeweils Teilräume von \underline{X} sind. Dann existieren jeweils in \underline{X} offene Mengen U_A und U_B , so dass deren Schnitt mit dem jeweiligen Definitionsbereich dieses offene Urbild ergibt, also $U_A \cap A = f_A^{-1}[U]$ und $U_B \cap B = f_B^{-1}[U]$. Zudem ist die Restmenge $A \setminus B$ offen, denn es ist $A \setminus B = (A \cup B) \setminus B = X \setminus B$ Komplementmenge der abgeschlossenen Menge B . Aus dem selben Grund ist $B \setminus A$ offen in $A \cup B$. Dann sind auch $U_B \cup (A \setminus B)$ und $U_A \cup (B \setminus A)$ offene Mengen und damit ist deren Schnitt

$$M = (U_B \cup (A \setminus B)) \cap (U_A \cup (B \setminus A))$$

ebenfalls eine offene Menge. Diese Schnittmenge ist aber gleichzeitig das Urbild $(f_A \cup f_B)^{-1}[U]$ der offenen Menge U bezüglich der gemeinsamen Fortsetzung unserer Abbildungen. Liegt nämlich ein Punkt $p \in M$, dann liegt dieser Punkt sowohl in $U_B \cup (A \setminus B)$ als auch in $U_A \cup (B \setminus A)$. Wenn dieser Punkt p in A liegt, dann liegt er nicht in $B \setminus A$ und muss daher in U_A liegen. Folglich gilt $p \in U_A \cap A = f_A^{-1}[U]$ und somit gilt $f_A(p) \in U$. Wenn p nicht in A liegt, dann liegt er auch nicht in $A \setminus B$ und muss folglich in U_B liegen. Dann gilt aber $f_B(p) \in U$. Damit gilt $M \subseteq (f_A \cup f_B)^{-1}[U]$.

Sei nun umgekehrt ein Punkt p Element von $(f_A \cup f_B)^{-1}[U]$, dann ist $p \in A$ oder $p \in B$. Aus Symmetriegründen betrachten wir nur den Fall $p \in A$, denn ansonsten muss $p \in B$ gelten und dann vertauschen wir einfach A mit B . Dann ist $f_A(p) \in U$ bzw. $p \in f_A^{-1}[U]$. Dann ist $p \in U_A$ und somit $p \in U_A \cup (B \setminus A)$. Ist p gleichzeitig in B , dann ist auch $p \in U_B$ und somit $p \in U_B \cup (A \setminus B)$. Ist p jedoch nicht in B , dann gilt $p \in A \setminus B$ und somit ebenfalls $p \in U_B \cup (A \setminus B)$. Also ist in jedem Fall $p \in (U_A \cup (B \setminus A)) \cap (U_B \cup (A \setminus B)) = M$. Es gilt also $(f_A \cup f_B)^{-1}[U] \subseteq M$ und mit der oberen Teilmengenbeziehung ergibt sich

$$(f_A \cup f_B)^{-1}[U] = M.$$

Damit wissen wir, dass das Urbild von jeder beliebigen offenen Menge U bezüglich der gemeinsamen Fortsetzung $f_A \cup f_B$ ebenfalls eine offene Menge ist und damit ist auch diese Abbildung stetig.

Da nun H eine solche gemeinsame Fortsetzung von stetigen Abbildungen ist, haben wir mit H ebenfalls eine stetige Abbildung und damit ist H eine Homotopie der Wege $w_{ab} * w_{bc}$ und $v_{ab} * v_{bc}$ (relativ zum Start- und Endpunkt).

Damit haben wir übrigens auch den Beweis nachgeholt, dass die Komposition zweier Wege ebenfalls ein Weg ist, da diese ganz analog als eine gemeinsame Fortsetzung zweier stetiger Abbildungen mit jeweils abgeschlossenem Definitionsbereichen definiert ist. Diesen sind wir bisher schuldig geblieben.

Zeigen wir nun noch, dass der Rückweg w_{ab}^- eines Weges w_{ab} in \underline{X} von a nach b bis auf Homotopie invers zu diesem Weg ist:

Dazu modifizieren wir die Komposition $w_{ab} * w_{ab}^-$ eines Weges w_{ab} und dessen Rückweg w_{ab}^- anhand eines Parameters $i \in [0, 1]$, indem wir den ersten Wegabschnitt des Weges w_{ab} um den Faktor i langsamer gehen und somit nicht immer das Ende dieses Weges erreichen sondern nur bis zur Stelle $w_{ab}(i)$ gelangen. Dann kehren wir an dieser Stelle um und schlendern mit gleichem „Gemütlichkeitsfaktor“ i wieder zurück nach a . Dabei ist bei $i = 0$ (Igel) die Geschwindigkeit ebenfalls 0 und der Weg ist der konstante Weg von a nach a , denn wir bleiben ja die Hälfte der Zeit einfach stehen und kommen so nur bis a . Dann kehren

wir um, indem wir die restliche Zeit ebenfalls einfach bei a beginnend stehen bleiben, bis wir endlich wieder bei a angelangt sind. Bei $i = 1$ (Hase) hingegen laufen wir pflichtbewusst den gesamten Weg w_{ab} komplett hin und auch wieder zurück.

Durch Nachrechnen zeigt sich, dass die nun folgende Argumentation genau zu der oberen anschaulich beschriebenen Konstruktion führt:

Die Wegkomposition eines Weges und seines Rückweges entspricht der Abbildungskomposition dieses Weges mit der Wegkomposition der identischen Abbildung als Weg in $[0, 1]$ mit ihrem Rückweg $\text{id}^-(t) := 1 - t$. Es gilt also

$$w_{ab} * w_{ab}^- = w_{ab} \circ (\text{id}_{[0,1]} * \text{id}_{[0,1]}^-).$$

Nun ist diese Wegkomposition $\text{id}_{[0,1]} * \text{id}_{[0,1]}^-$ ein Rundweg in $[0, 1]$ mit Basispunkt 0 , von dem wir wissen, dass er homotop zum konstanten Weg c_0 in $[0, 1]$ ist. Dann ist aber auch $w_{ab} \circ c_0$ homotop zu $w_{ab} \circ (\text{id}_{[0,1]} * \text{id}_{[0,1]}^-)$ wie wir ebenfalls gesehen haben. Wegen $w_{ab} \circ c_0 = c_a$ ist also der konstante Weg c_a homotop zu $w_{ab} * w_{ab}^-$.

Wir haben noch zu zeigen, dass ein Rundweg w_{xx} innerhalb eines Sterns $\text{St}(y)$ homotop zum konstanten Weg c_x ist:

Sei \underline{X} alexandroffscher Raum mit topologischem Datentyp (X, R) und $w_{xx} : [0, 1] \rightarrow \underline{X}$ ein Rundweg in \underline{X} , dessen Spur $\text{Bild}(w)$ in einem Stern $\text{St}_R^*(y)$ für ein y in \underline{X} liegt, also $w_{xx}([0, 1]) \subseteq \text{St}_R^*(y)$. Zunächst nehmen wir an $x = y$.

Dann definieren wir die Abbildung

$$h : [0, 1] \times [0, 1], (i, t) \mapsto h(i, t) := \begin{cases} w_{xx}(t) & : i > 0 \\ x & : i = 0. \end{cases}$$

Dies ist eine Homotopie von w_{xx} zum konstanten Weg c_x . Diese Abbildung ist stetig, denn ist A eine offene Menge in \underline{X} , dann gilt entweder $x \in A$ oder $x \notin A$. Ist $x \in A$, dann ist $\text{St}_R^*(x) \subseteq A$, denn A ist ja offen, also ist zu jedem Punkt in A auch dessen Stern in A . Dann ist aber das gesamte Bild von h in A , denn entweder ist $i > 0$ und dann ist $h(i, t) = w_{xx}(t) \in \text{St}_R^*(x)$ oder $i = 0$ und dann ist $h(i, t) = x \in \text{St}_R^*(x)$. Damit haben wir im Falle $x \in A$ das Urbild $h^{-1}[A] = [0, 1] \times [0, 1]$, eine offene Menge.

Ist hingegen $x \notin A$, dann kann auch für $h(i, t) \in A$ nicht $i = 0$ gelten. Gilt hier also $h(i, t) \in A$, dann gilt $i > 0$ und damit auch $h(i, t) = w_{xx}(t)$. Also ist $h^{-1}[A] =]0, 1] \times w^{-1}[A]$. Das Intervall $]0, 1]$ ist offen in $[0, 1]$. $w^{-1}[A]$ ist ebenfalls offen in $[0, 1]$, da w_{ab} stetig ist. Damit ist auch das Produkt $]0, 1] \times w^{-1}[A]$ offen in $[0, 1] \times [0, 1]$. Damit ist h stetig.

Diese Abbildung ist zudem eine Homotopie von c_x nach w_{xx} , denn es ist $h(0, t) = x = c_x(t)$ und $h(1, t) = w_{xx}(t)$. Zudem ist sie relativ zu Start- und Endpunkt, denn $h(i, 0) = x = w_{xx}(0) = w_{xx}(1) = x = h(i, 1)$ unabhängig von der Wahl von $i \in [0, 1]$.

Ist nun das Zentrum y des Sterns vom Start- und Endpunkt x des Rundwegs w_{xx} verschieden, dann gilt $x R^* y$ und damit existiert der elementare Weg w_{xy} von x nach y und sein Rückweg w_{xy}^- von y nach x , deren Spur ebenfalls in $\text{St}_R^*(x)$ liegt. Es ist $w_{xy} * w_{xy}^-$ homotop zum konstanten Weg c_x , also ist auch $(w_{xy} * w_{xy}^-) * w_{xx} * (w_{xy} * w_{xy}^-)$ homotop zu w_{xx} . Dann ist auch der innere Teilweg $w_{xy}^- * w_{xx} * w_{xy}$ ein Rundweg von Zentrum y des Sterns aus, bleibt ebenfalls innerhalb dieses Sterns und ist somit homotop zu c_y . Damit ist $w_{xy} * (w_{xy}^- * w_{xy}^-) * w_{xx} * (w_{xy} * w_{xy}^-)$ homotop zu c_y .

$w_{xx} * w_{xy}) * w_{xy}^-$ homotop zu $w_{xy} * c_y * w_{xy}^-$, dies wiederum zu $w_{xy} * w_{xy}^-$ und als Komposition von Hin- und Rückweg homotop zu c_x .

Wenn wir „ist homotop zu“ als Zeichen „ \cong “ schreiben entspricht dies:

$$\begin{aligned} w_{xx} &\cong c_x * w_{xx} * c_x \\ &\cong (w_{xy} * w_{xy}^-) * w_{xx} * (w_{xy} * w_{xy}^-) \\ &\cong w_{xy} * (w_{xy}^- * w_{xx} * w_{xy}) * w_{xy}^- \\ &\cong w_{xy} * (c_y) * w_{xy}^- \\ &\cong w_{xy} * w_{xy}^- \\ &\cong c_x. \end{aligned}$$

Und deutet so an, wie algebraische Methoden in der Topologie eingesetzt werden. \square

Mit diesen Instrumenten ausgestattet, erzeugen wir nun aus einem beliebigen Weg $w : [0, 1] \rightarrow \underline{X}$ in einem alexandroffschen Raum einen dazu homotopen endlichen Weg $s : [0, 1] \rightarrow \underline{X}$.

Wir wissen bereits, dass es zu $w : [0, 1] \rightarrow \underline{X}$ eine minimale Menge endlich vieler offener Intervalle $\{]a_0, b_0[, \dots ,]a_n, b_n[\}$ gibt, die das Einheitsintervall offen überdecken, so dass zu jedem derartigen Intervall $]a_i, b_i[$ dessen Bild $w(]a_i, b_i[)$ vollständig in einem Stern $\text{St}_r^*(x_i)$ liegt. Wir unterstellen, dass diese Intervalle nach unteren Grenzen aufsteigend sortiert vorliegen.

Durch vollständige Induktion über n zeigen wir, dass dann ein zu w äquivalenter endlicher Weg w_n von $w(0)$ nach $w(1)$ existiert.

Sei $n = 0$, dann ist diese Überdeckung $\{]a_0, b_0[\}$ bzw. $\{ [0, 1] \}$, da die Intervallgrenzen a_0 und b_0 dann etwas über die Raumgrenzen 0 und 1 „hinausragen“. Dann ist die Spur von w vollständig in einem Stern $\text{St}_R^*(x)$ enthalten. Sei a Start- und b Zielpunkt des Weges. Dann gilt $a R^* x$ und $b R^* x$ und damit existiert ein elementarer Weg w_{ax} von a nach x mit Rückweg $w_{xa} = w_{ax}^-$ und ein elementarer Weg w_{bx} von b nach x mit Rückweg $w_{xb} = w_{bx}^-$ und damit ist w homotop zu $(w_{ax} * w_{xa}) * w * (w_{bx} * w_{xb})$. Dann aber ist der innere Teil $w_{xa} * w * w_{bx}$ ein Rundweg in einem Stern von x nach x und somit homotop zum konstanten Weg c_x und damit ist

$$\begin{aligned} w &\cong c_a * w * c_b && \text{konstante Wege anhängen} \\ &\cong (w_{ax} * w_{xa}) * w * (w_{bx} * w_{xb}) && \text{durch Hin- und Rückweg ersetzen} \\ &\cong w_{ax} * (w_{xa} * w * w_{bx}) * w_{xb} && \text{Umgruppieren} \\ &\cong w_{ax} * c_x * w_{xb} && \text{Rundweg im Stern kürzen} \\ &\cong w_{ax} * w_{xb} = w_0. \end{aligned}$$

Wobei „ \cong “ wieder die Relation „ist homotop relativ zu Anfangs- und Endpunkt“ anzeigt. Diese ist als Äquivalenzrelation transitiv, und somit ist w homotop zum endlichen Weg $w_{ax} * w_{xb}$, den wir hier mit w_0 bezeichnet hatten.

Gelte nun diese Behauptung, dass für eine feste Zahl n , ein endlicher homotoper Weg existiere (also für $n + 1$ Intervalle), sei aber die Anzahl der überdeckenden Intervalle $n + 2$, also $\{]a_0, b_0[, \dots ,]a_n, b_n[,]a_{n+1}, b_{n+1}[\}$. Dann gibt es einen Parameter $t \in [0, 1]$, der in den beiden letzten Teilintervallen liegt, also $t \in]a_n, b_n[\cap]a_{n+1}, b_{n+1}[$. Dann gibt es den Teilweg $w|_{[0,t]}$ von $w(0)$ nach $w(t)$,

der durch die endlich vielen Intervalle $\{]a_0, b_0[, \dots ,]a_n, b_n[\}$ überdeckt ist und somit laut Induktionsannahme einen dazu homotopen endlichen Weg w_n . Der restliche Weg $w|_{[t,1]}$ verläuft nun innerhalb eines Sterns $\text{St}_R^*(x)$ von $w(t)$ nach $w(1)$. Dann liegt $w(t) \in \text{St}_R^*(x)$, also gilt $w(t) R^* x$ und damit existiert der elementare Weg $s_{w(t)x}$ von $w(t)$ nach x und sein Rückweg $s_{xw(t)}$, und damit ist $s_{w(t)x} * s_{xw(t)} * w|_{[t,1]}$ homotop zu $w|_{[t,1]}$. Auch der Zielpunkt $w(1)$ liegt im Stern um x und damit haben wir ebenfalls den elementaren Weg $s_{w(1)x}$ mit Rückweg $s_{xw(1)}$ und damit ist $w|_{[t,1]}$ homotop zu $s_{w(t)x} * s_{xw(t)} * w|_{[t,1]} * s_{w(1)x} * s_{xw(1)}$. Dann ist wiederum der innere Teil des Wegs $s_{xw(t)} * w|_{[t,1]} * s_{w(1)x}$ ein Rundweg innerhalb eines Sterns, kann also herausgekürzt werden und es bleibt $w|_{[t,1]}$ ist homotop zu $s_{w(t)x} * s_{xw(1)}$, den wir mit s_{t1} bezeichnen wollen. Damit ist also der erste Wegabschnitt $w|_{[0,t]}$ homotop zum endlichen Weg w_n , der zweite Abschnitt ist homotop zu s_{t1} und damit ist die Komposition $w|_{[0,t]} * w|_{[t,1]}$ homotop zur Komposition der endlichen Wege $w_n * s_{t1}$, die wiederum einen endlichen Weg ergibt, und dieser ist homotop zu w . Insbesondere liegt die Spur von s in der Spur von w , denn die Zentren der Sterne wurden ja ursprünglich aus der Spur von w ausgewählt.

Damit ist jeder Weg in einem alexandroffschen Raum bis auf Homotopie ein endlicher Weg, was zu beweisen uns wir hier vorgenommen hatten.

Anhang B

Monotone Mengen

B.1 Algorithmus zur Entscheidung von Monotonie.

Wir definieren die Funktion `monotonep` als ein LISP-Prädikat. In LISP ist ein Prädikat eine Abbildung, deren Namen mit einem „p“ endet oder auch nicht und dessen Ergebnis in der Regel als Wahrheitswert interpretiert wird. Dabei ist in LISP der Wert `nil` auch gleichzeitig der Wahrheitswert `false` und jeder andere Wert (etwa "Willi") hat Wahrheitswert `true` und somit gibt es dort keinen prinzipiellen Unterschied zwischen Funktionen und Prädikaten. Insbesondere ist das Symbol `t` ein Repräsentant von `true`.

Die Funktion `monotonep` hat als Argumente eine Menge A , eine weitere Menge X , die A umfasst, und eine Relation R auf X . Es gilt stets

$$\text{monotonep}(A, X, R) \neq \text{nil}$$

bzw. in LISP-Syntax

$$(\text{monotonep } A \ X \ R)$$

genau dann, wenn A eine monotone Menge in (X, \mathcal{T}_R) ist. Eine detaillierte Beschreibung der Programmiersprache LISP findet sich in [Ste90], die auch im Internet an vielen Stellen frei verfügbar ist (Stichwort: CLTL2).

Wir behaupten nun, dass folgender Algorithmus in LISP Monotonie einer endlichen Teilmenge A in einem endlichen topologischen Datentyp (X, R) entscheidet und dass dessen Laufzeit polynomiell bezüglich der Größe der angegebenen Parameter A , X und R ist:

```

1:(defun monotonep (A X R)
2:
3:  (do* ( ( coA      (set-difference X A) )
4:        ( coAlter  coA
5:            (set-difference coAlter extComp) )
6:        ( extComp  (list (first coAlter))
7:                  (list (first coAlter)) ) )
8:
9:    ( (null coAlter) t ) ; externe Komponenten erledigt
10:
11:   (do* ( ( xC extComp ) )
12:         ( (null xC) ) ; extComp berechnet
13:         (let* ( ( StxC  (set-difference
14:                       (star xC X R)
15:                       A )
16:                 ( BdStxC (set-difference
17:                          (boundary StxC X R)
18:                          A) ) )
19:
20:           (setf xC      (set-difference BdStxC
21:                                       extComp )
22:                 extComp (append xC extComp) ) )
23:
24:   (do* ( ( StBdCA
25:         (intersection
26:           (union (star extComp X R)
27:                 (boundary extComp X R) )
28:           A )
29:         (remove (first endpoint) StBdCA) )
30:         ( endpointp #'(lambda (p)
31:                       (declare (special StBdCA))
32:                       (subsetp StBdCA
33:                               (star (list p) X R) ) ) ) )
34:         ( endpoint (member-if endpointp StBdCA)
35:                   (member-if endpointp StBdCA) ) )
36:         ( (null endpoint)
37:           (if (not (null StBdCA))
38:               (return-from monotonep nil) ) )
39:         (declare (special StBdCA)) )
40:)

```

Dabei gehören die Zeilennummern nicht zum Programm und sollen nur die Argumentation erleichtern.

Um zu zeigen, dass ein Algorithmus korrekt ist, geht man üblicherweise in zwei Schritten vor:

1. Zunächst untersucht man, ob der Algorithmus *partiell* korrekt ist, das heißt, ob er ein korrektes Ergebnis liefert, wenn er anhält.
2. Erst in einem weiteren Schritt überlegt man dann, dass der Algorithmus bei jeder gültigen Eingabe in der Tat auch anhält.

Ein partiell korrekter Algorithmus, der (bei gültiger Eingabe) stets anhält, heißt total korrekt oder einfach nur korrekt. Diese Bezeichnungen „partiell“ und „total“ stehen übrigens im engen Zusammenhang mit den entsprechenden Bezeichnungen für Abbildungen.

Danach wollen wir kurz darauf eingehen, welche Komplexität dieser Algorithmus hat, wie lange wir also (in Abhängigkeit von der Größe der Eingabe, aber unabhängig von der Geschwindigkeit eines Computers, auf dem dieser Algorithmus läuft) auf das Ergebnis warten müssen. Allerdings lohnt sich eine ausführliche Komplexitätsbetrachtung bei diesem Algorithmus nicht, da er nicht sonderlich effizient ist. Wir kennen ja bereits ein effizient zu berechnendes (aber auch echt strengeres) Monotoniekriterium. Die Effizienz des hier vorliegenden Algorithmus ist auch beeinflusst durch die hier verwendete Darstellung von Mengen $\{a_1, \dots, a_n\}$ als Listen der Form $(a_1 \dots a_n)$, wie sie in LISP üblich ist. So benötigt z.B. das Suchen, Hinzufügen und Entfernen eines Elements lineare Zeit. Allerdings gibt es (auch in LISP) diesbezüglich effizientere Algorithmen¹. In einer Hinsicht ist dieser Algorithmus dennoch effizient – seine Korrektheit ist relativ leicht zu zeigen.

Die Parameter der Funktion `monotonep` sind in der Funktionssignatur (Zeile 1) angegeben: Dies sind eine Menge A eine Menge X und eine Relation $R \subseteq X \times X$, wobei über Implementierungsdetails von R nichts gesagt ist. A und X seien Mengen, dargestellt als Listen mit $A \subseteq X$ (in LISP: `(subsetp A X)`). Es ist zwar üblich, bei den Mengenoperatoren wie etwa `union`, `set-difference` und `intersection` optionale Schlüsselwortparameter `test` und `test-not` als Alternativen für die Gleichheitsrelation `#'eql` mit anzugeben. Würden wir dieser Konvention folgen, dann hätte `monotonep` die Form

```
(defun monotonep (A X R &key (test #'eql)) . Body )
```

Die häufige Verwendung von Formen wie `(union A B :test test)` anstelle von `(union A B)` (für $A \cup B$) erschwert jedoch die Lesbarkeit, weshalb wir hier darauf verzichten.

Da der Algorithmus `do*`-Schleifen verwendet, wollen wir zunächst deren Bedeutung erläutern. Eine `do*`-Schleife ist ein LISP-Ausdruck mit folgendem Aufbau:

```
(do* ( Var1 ... Varn ) ( Test . Ergebnis ) . Rumpf ),
```

beginnend mit dem Symbol `do*`, gefolgt von der so genannten `do`-Liste, einer Test und Ergebnisliste und einer Liste von Ausdrücken im Rumpf. Man beachte die zwei Punkte in den Listen. Die Schleife wiederholt den Rumpfausdruck so lange, bis der Testausdruck erfüllt ist, und gibt dann das angegebene Ergebnis zurück.

Die `do`-Liste ist die Liste der Deklaration der Schleifenvariablen, die u.A. die Form `(symbol init step)` oder die Form `(symbol init)` haben können. Dabei ist `symbol` jeweils ein LISP-Symbol und `init` und `step` ist jeweils ein Ausdruck. `Test` ist ein Ausdruck, dessen Ergebnis als Wahrheitswert interpretiert wird, und dieser wird gefolgt von `Ergebnis`, einer (gegebenenfalls leeren) Liste von Ausdrücken. Der `Rumpf` ist ebenfalls eine derartige Liste von Ausdrücken. Wir

¹Die ersten Implementierungen von LISP waren wenig effizient. Die Ineffizienz von LISP lebt heute noch als nicht mehr gerechtfertigter Mythos weiter.

merken noch an, dass in LISP $(a\ b\ .\ (c\ d))$ eine alternative Schreibweise für $(a\ b\ c\ d)$ ist.

Eine `do*`-Schleife wiederholt den Schleifenrumpf so lange, bis der Testausdruck erfüllt ist und berechnet dann ein Ergebnis dieser Iteration. Die Auswertung einer `do*`-Schleife beginnt damit, dass zunächst für alle deklarierten Variablensymbole eventuell bestehende Variablenbindungen gesichert werden, so dass eine Bindung an Variablenwerte nur lokal innerhalb dieser Schleife gelten. Danach werden die Variablen initialisiert. Dazu wird in der Reihenfolge der Deklaration der Variablen in der `do`-Liste jeweils das zweite Element der Liste ausgewertet und dessen Ergebnis an das zugehörige Variablensymbol gebunden. Erst nach dieser Bindung wird das jeweils darauf folgende Symbol initialisiert, so dass der Initialisierungsausdruck auf die Werte bereits vorher initialisierter Symbole zugreifen kann.

Danach wird der Test-Ausdruck ausgewertet. Wenn dieser Test fehlschlägt (also `nil` zurückgibt), werden die weiteren Ausdrücke zur Berechnung des Ergebnisses ignoriert und die Ausdrücke des Schleifenrumpfs ausgewertet. Danach folgt die Aktualisierung aller Variablen, für die ein Aktualisierungsausdruck *step* angegeben ist. Diese erfolgt ebenfalls sequenziell. Danach wird erneut der Test ausgeführt. Dies wird nun so lange wiederholt, bis der Test „wahr“ (also einen von `nil` verschiedenen Wert) ergibt.

Initialisierung und Aktualisierung der Schleifenvariablen geschehen sequenziell. Es gibt auch die Variante `do` (ohne „*“), in der dies parallel geschieht. Für unsere Zwecke ist die sequenzielle Variante besser geeignet.

Dann wird der Wert des `do*`-Ausdrucks berechnet, indem alle auf den Test folgenden Ergebnisausdrücke der Liste ausgewertet werden. Das Ergebnis des letzten Ausdrucks ergibt das Ergebnis der Iteration. Fehlen diese Ergebnisausdrücke, dann ist `nil` dieses Ergebnis.

Die äußere `do*`-Schleife unseres Algorithmus (Zeilen 3 bis 39) iteriert nun über alle Zusammenhangskomponenten des Raums $(X, \mathcal{T}_R)|_{X \setminus A}$, also der Komplementmenge von A in X . Bei jeder Iteration ist diese Zusammenhangskomponente ab Zeile 24 an die Variable `extComp` gebunden.

Bei der Initialisierung von `do*` (Zeile 3 bis 7), wird die Variable `coA`, das Komplement von A , mit $X \setminus A$ initialisiert (Zeile 3). Es folgt die Initialisierung von `coAIter` mit `coA` (Zeile 4), also ebenfalls $X \setminus A$, da die Initialisierung sequenziell verläuft. Die Variable `extComp` wird (zunächst) an eine Liste mit genau einem Element gebunden, und zwar entweder ein beliebig ausgewähltes Element aus `coAIter`, also $X \setminus A$, oder `nil` falls `coAIter` die leere Menge ist (Zeile 6). Das Symbol `nil` steht in LISP gleichzeitig auch für die leere Liste.

Ist `coAIter` nach dieser Initialisierung die leere Menge, dann ist $X \subseteq A$ und damit ist wegen der Vorbedingung $A \subseteq X$ auch $X = A$ und folglich ist A trivialerweise monoton in (X, R) . In Zeile 9 ist dies die Abbruchbedingung der `do*`-Schleife, die in diesem Fall korrekt mit Ergebnis `t`, also `true` anhält.

Ansonsten ist `coAIter` nicht die leere Menge. Dann gibt es auch einen Punkt p in X , der nicht in A liegt. Die „Menge“ $\{p\}$ ist dann als Liste (p) mit genau einem solchen Element p an `extComp` gebunden, und der Algorithmus erreicht Zeile 11. Die erste innere `do*`-Schleife (Zeilen 11 bis 22) berechnet nun die Zusammenhangskomponente $[p]$ von p im „Restraum“

$$(X, \mathcal{T}_R) \setminus A = (\text{coA}, \mathcal{T}_R|_{\text{coA}}).$$

Um dies zu beweisen, werden wir Folgendes zeigen:

1. An Zeile 12 gilt stets $p \in \text{extComp}$ und für jeden Punkt q in extComp und in \mathbf{xC} gibt es einen Weg von q nach p in $(X, \mathcal{T}_R) \setminus A$. Es gilt also immer $p \in \text{extComp} \subseteq [p]$.
2. Die Schleife terminiert nach endlich vielen Schritten.
3. Wenn die Schleife terminiert, ist $[p] \subseteq \text{extComp}$.

Lemma B.1. *In Zeile 12 sind alle Punkte in \mathbf{xC} und alle Punkte in extComp durch einen Weg in $\underline{\text{coA}} := (X \setminus A, \mathcal{T}_R|_{X \setminus A})$ mit p verbunden und es gilt stets $p \in \text{extComp}$.*

Beweis. Bei Betreten der do^* -Schleife in Zeile 11 wird die Schleifenvariable \mathbf{xC} mit extComp initialisiert. Damit haben wir bereits 0 Schleifendurchläufe und es gilt $\{p\} = \text{extComp} = \mathbf{xC}$. Jeder Punkt dieser Menge ist ganz offensichtlich durch einen Weg in $\underline{\text{coA}}$ mit p verbunden, es ist $p \in \text{extComp}$ und wir haben bereits den Anfang eines Induktionsbeweises. Durch vollständige Induktion über die Anzahl n von Schleifenwiederholungen zeigen wir nun die Behauptung:

Für 0 Schleifendurchläufe haben wir die Behauptung des obigen Lemmas eben bereits gezeigt.

Gelte die Behauptung stets für n Schleifendurchläufe und habe es aber $n + 1$ Schleifendurchläufe bis Zeile 12 gegeben. Dann gilt bis zum n -ten Schleifendurchlauf die Induktionsbehauptung, dass die Mengen \mathbf{xC} und in extComp Teilmengen von $[p]$ sind. Da es noch einen weiteren Schleifendurchlauf gibt, kann \mathbf{xC} nicht die leere Menge sein, denn sonst würde $(\text{null } \mathbf{xC})$ gelten und die Iteration an Zeile 12 abbrechen².

Dann betritt der Algorithmus den let^* -Ausdruck, eine syntaktische Struktur, in der ähnlich wie bei do^* lokale Variablen deklariert und initialisiert werden können und danach eine Folge von Ausdrücken ausgeführt wird, bei let^* aber ohne Wiederholungen. Folgende Ausdrücke sind also äquivalent:

$$(\text{do}^* \quad ((\text{var-1 init-1}) (\text{var-2 init-2}) \dots (\text{var-n init-n})) \\ \quad (\text{t expr-1} \dots \text{expr-n}))$$

$$(\text{let}^* \quad ((\text{var-1 init-1}) (\text{var-2 init-2}) \dots (\text{var-n init-n})) \\ \quad \text{expr-1} \dots \text{expr-n})$$

Zeile 14 berechnet $\text{St}_R^*(\mathbf{xC})$. Wir setzen voraus, dass eine entsprechende LISP-Funktion star definiert ist, die dieses Ergebnis liefert. Jeder Punkt q , der sowohl in $\text{St}_R^*(\mathbf{xC})$ als auch in $\underline{\text{coA}}$ liegt, ist nun mit p durch einen Weg in $\underline{\text{coA}}$ verbunden, denn es gilt qR^*x für ein $x \in \mathbf{xC}$ und damit existiert ein Weg von q nach x in (X, R) und laut Induktionsvoraussetzung existiert ein Weg von x nach p in $\underline{\text{coA}}$. Damit existiert ein Weg von q nach p in $\underline{\text{coA}}$. In Zeile 13 wird die Variable StxC an $\text{St}_R^*(\mathbf{xC}) \setminus A$ gebunden, also besteht diese Menge genau aus denjenigen Punkten $q \in \text{St}_R^*(\mathbf{xC})$, die auch in $\underline{\text{coA}}$ liegen, und somit existiert für jeden Punkt in StxC ein Weg in $\underline{\text{coA}}$ nach p .

Ganz analog berechnet Zeile 17 die abgeschlossene Hülle $\text{Bd}_R^*(\text{StxC})$ des zuvor berechneten StxC unter Verwendung einer als gegeben angenommenen LISP-Funktion boundary , und bei Zeile 16 wird BdStxC an $\text{Bd}_R^*(\text{StxC}) \setminus A$ gebunden. Auch hier gibt es zu jedem Punkt in BdStxC einen Weg in $\underline{\text{coA}}$ nach p . Wir erinnern uns, dass die abgeschlossene Hülle dem Stern im dualen Raum

² null ist also ein Prädikat, das nicht mit „p“ endet.

entspricht und dieser durch die Transponierte der Relation definiert ist. Es gilt also $\text{Bd}_R^* = \text{St}_{R^T}^*$.

In Zeile 20 wird nun der Variablen \mathbf{xC} der Wert von $\text{BdStxC} \setminus \text{extComp}$ zugewiesen. Danach gilt $\mathbf{xC} \subseteq \text{BdStxC}$ und damit gilt die Behauptung des Lemmas bezüglich \mathbf{xC} . Insbesondere sind nun \mathbf{xC} und extComp disjunkt. In Zeile 22 wird der unter extComp gespeicherten Menge noch die Menge \mathbf{xC} hinzugefügt und die so erweiterte Menge wiederum extComp zugeordnet. Nun bestand laut Induktionsvoraussetzung extComp nur aus Punkten, die durch einen Weg in coA mit p verbunden sind und \mathbf{xC} besteht ebenfalls nur aus derartigen Punkten und somit auch deren Vereinigung. Zudem galt $p \in \text{extComp}$ vor der Erweiterung laut Voraussetzung und dies gilt damit auch danach. Die Iteration wird nun bei Zeile 12 fortgesetzt, da in Zeile 11 kein Aktualisierungsausdruck steht. Damit gilt nach der $n + 1$ -ten Iteration die Behauptung von Lemma B.1 und somit ist dieses Lemma durch vollständige Induktion über die Zahl der Wiederholungen von Zeile 12 bewiesen. \square

Lemma B.2. *Die do*-Schleife ab Zeile 11 terminiert nach endlich vielen Schritten.*

Beweis. Die Menge extComp wächst mit jedem Iterationszyklus in Zeile 22 monoton an. Wir haben ebenfalls gesehen, dass extComp stets Teilmenge der Zusammenhangskomponente $[p]$ in coA bleibt. Damit ist das Wachstum von extComp nach oben durch eine endliche Menge beschränkt. Die Endlichkeit von X wurde ja vorausgesetzt und es gilt $[p] \subseteq X$ und folglich ist $[p]$ endlich. Dann muss es nach endlich vielen Iterationen eine Iteration geben, in der extComp nicht weiter anwächst. Es ist jedoch \mathbf{xC} genau diejenige Menge, um die extComp angewachsen ist, also muss in diesem Zyklus \mathbf{xC} die leere Menge sein. Dann gilt aber $(\text{null } \mathbf{xC})$ und die Iteration bricht an Zeile 12 ab. \square

Lemma B.3. *Wenn die do*-Schleife ab Zeile 11 terminiert, ist $[p] \subseteq \text{extComp}$.*

Beweis. Sei S eine Relation in $X \setminus A$, die die Spurtopologie von \mathcal{T}_R in $X \setminus A$ erzeugt, z.B. $S = R^+|_{\text{coA}}$.

Sei $q \in [p]$ beliebig, dann ist q äquivalent zu p bezüglich der von S erzeugten Äquivalenzrelation $(S \cup S^T)^*$. Dann gibt es eine natürliche Zahl n mit $(q, p) \in (S \cup S^T)^n$. Ist $n = 0$, dann gilt $q = p$ und damit wegen $p \in \text{extComp}$ auch $q \in \text{extComp}$. Ansonsten gibt es einen endlichen Weg $q = a_1, \dots, a_n = p$ in coA mit $a_{i-1} (S^+ \cup S^{+T}) a_i$ für alle $i = 2, \dots, n$.

Durch vollständige Induktion über die Anzahl z der Schleifendurchläufe wollen wir zeigen, dass die Iteration nicht aufhört, solange $q \neq \text{extComp}$ gilt. Dazu behaupten wir, dass an Zeile 12 für jeden Punkt $q \in [p]$ aus $q \notin \text{extComp}$ folgt, dass jeder Weg w_{qp} von q nach p in coA über einen Punkt $x \in \mathbf{xC}$ führt. Dann ist $\mathbf{xC} \neq \emptyset$ und es gilt nicht $(\text{null } \mathbf{xC})$ und die Iteration wird fortgesetzt.

Für $z = 0$ gilt $\mathbf{xC} = (p)$, und da jeder Weg w_{qp} von q nach p in p endet, gilt die Behauptung.

Gelte nun die Behauptung nach z Iterationen und sei q auch nach $z + 1$ Iterationen noch nicht in extComp enthalten. Dann ist zu zeigen, dass nach der $z + 1$ -ten Iteration an Zeile 12 die an die Variable \mathbf{xC} gebundene Menge nicht leer ist. Wegen Zeile 22 gilt stets $\mathbf{xC} \subseteq \text{extComp}$ und damit ist q auch nicht in \mathbf{xC} enthalten. Dann führte bisher jeder Weg w_{qp} von q nach p laut Induktionsvoraussetzung über einen Punkt $x \in \mathbf{xC}$. Ohne Einschränkung ist

dieser Weg endlich, denn da X als endlich vorausgesetzt wurde, hat w_{qp} eine endliche Spur in $\underline{\text{coA}}$ und damit gibt es einen Pfad $(q, a_1, \dots, a_{n-1}, p)$ mit $n+1$ Wegpunkten von w_{qp} so dass für alle a_{i-1}, a_i mit $i = 1, \dots, n$ gilt $(a_{i-1}, a_i) \in S^+$ oder $(a_i, a_{i-1}) \in S^+$. Dabei setzen wir $a_0 = q$ und $a_n = p$. Dies gilt, da die Spur von w_{ab} ebenfalls ein alexandroffscher Raum ist und w_{ab} auch ein Weg in seiner eigenen Spur als Teilraum von (X, R) ist. Damit gibt es in der Spur von w_{ab} , also zu jedem Weg v_{xy} von x nach y eine (laut Anhang A sogar homotope) endliche Abkürzung $(x = x_0, x_1, \dots, x_m = y)$ mit $x_{j-1}R^*x_j$ oder $x_jR^*x_{j-1}$ für $j = 1, \dots, m$. Diese Abkürzung ist wiederum ein Weg in $[p]$. Damit liegt laut Induktionsvoraussetzung mindestens eines dieser a_i in \mathbf{xC} und dieses a_i ist ein Wegpunkt von w_{xy} . Sei nun i minimal mit $a_i \in \mathbf{xC}$. Dann ist $i > 0$, denn ansonsten wäre ja wegen $a_i = a_0 = q$ der Punkt q Element von \mathbf{xC} , was wir laut Voraussetzung ausgeschlossen hatten. Dann ist $a_{i-1} \notin \mathbf{xC}$ und es gilt $a_{i-1}R^*a_i$ oder $a_iR^*a_{i-1}$.

Nun berechnet der Algorithmus in Zeile 16 $\text{BdStxC} = \text{Bd}_S^*[\text{St}_S^*[\mathbf{xC}]]$, also alle Punkte $a \in \text{coA}$, für die ein $x \in \mathbf{xC}$ existiert, so dass $a(S^* \cup S^{T^*})x$ gilt, wie man leicht überprüft. Damit ist $a_{i-1} \in \text{BdStxC}$.

In Zeile 20 wird nun die Variable \mathbf{xC} an $\text{BdStxC} \setminus \text{extComp}$ gebunden, wodurch $a_{i-1} \in \mathbf{xC}$ gilt, denn a_{i-1} ist ja Element von BdStxC und nicht Element von extComp . Damit ist für die nächste Iteration $\mathbf{xC} \neq \emptyset$, bzw. der Ausdruck $(\text{null } \mathbf{xC})$ gleich nil und wird somit nicht an der Zeile 12 abgebrochen. \square

Wir haben also gesehen, dass der Algorithmus, wenn er Zeile 24 erreicht, eine Zusammenhangskomponente $[p]$ von $\underline{\text{coA}}$ an die Variable extComp gebunden hat. Wir sagen nun, der do^* -Ausdruck an Zeile 24 *akzeptiert* eine Zusammenhangskomponente extComp , wenn er nicht an Zeile 38 gelangt. An dieser Zeile hält der Algorithmus mit Ergebnis false an. Akzeptiert der Ausdruck hingegen die Zusammenhangskomponente, dann terminiert er normal und die Iteration der äußeren do^* -Schleife wird wiederholt.

Lemma B.4 (Akzeptieren bei Monotonie). *Ist A monoton in (X, R) , dann erreicht der Algorithmus nicht die Zeile 38 und akzeptiert also jede Zusammenhangskomponente extComp von $\underline{\text{coA}}$.*

Beweis. Sei also A monoton in (X, R) und $[p]$ eine Zusammenhangskomponente von $X \setminus A$, dann ist $[p]$ auch zusammenhängend in (X, R) . Dann sind $\text{St}_R^*([p])$ und $\text{Bd}_R^*([p])$ ebenfalls zusammenhängend in (X, R) . Dann ist auch deren Vereinigung $\text{St}_R^*([p]) \cup \text{Bd}_R^*([p])$ zusammenhängend in (X, R) , da diese nicht disjunkt sind und die Vereinigung nichtdisjunkter zusammenhängender Mengen ebenfalls eine zusammenhängende Menge ergibt. Sei nun

$$\text{StBdCA} := (\text{St}_R^*([p]) \cup \text{Bd}_R^*([p])) \cap A$$

der Schnitt dieser Menge mit A . Dann ist jede Teilmenge $U \subseteq \text{StBdCA}$ ebenfalls zusammenhängend in A , denn sei $x \in U$ ein beliebiger Punkt, dann gilt xR^*c oder cR^*x für einen Punkt c in $[p]$ und damit existiert ein Weg von x nach c in $U \cup [p]$ und da $[p]$ zusammenhängend ist existiert ebenfalls ein Weg von c nach p in $U \cup [p]$. Also existiert ein Weg von x nach p in $U \cup [p]$. Dann ist auch $(U \cup [p]) \cap A$ zusammenhängend, da A monoton in (X, R) ist. Es ist jedoch U genau derjenige Teil von $U \cup [p]$, der in A liegt und somit ist $U = (U \cup [p]) \cap A$ zusammenhängend.

Es ist nun genau dann jede Teilmenge von StBdCA zusammenhängend, wenn es einen *gerichteten* Pfad $a_0 R^+ a_1 \cdots a_{n-1} R^+ a_n$ in X gibt, der jeden Punkt in StBdCA mindestens einmal besucht. Genau dann gibt es auch einen sogenannten *gerichteten Hamilton-Pfad* $u_0 R^+ u_1 \cdots u_{m-1} R^+ u_m$ in $(U, R^+|_U)$, der jeden Punkt in U genau einmal besucht. Wir wollen zunächst Missverständnissen vorbeugen und klarstellen, dass wir uns hier *nicht* mit dem berühmten NP-vollständigen Problem beschäftigen. Es wird hier nur der Sonderfall von Hamilton-Pfaden in *transitiven* Graphen betrachtet, und wir wollen diesen im Folgenden als *surjektiven gerichteten Pfad* bezeichnen. Wir erhalten diesen Pfad, indem wir einfach aus $a_0 R^+ a_1 \cdots a_{n-1} R^+ a_n$ alle Duplikate und alle Elemente streichen, die nicht in U sind. Gilt nämlich $a_0 \notin U$ können wir den Pfad mit a_1 beginnen lassen und entsprechend den Pfad auch bei $a_n \notin U$ vom Ende her wegekürzen. Haben wir dann noch Elemente $a \notin U$ in der Mitte unseres Pfades übrig, dann liegen diese ja in der Form $u_i R^+ a R^+ u_{i+1}$ vor. Diesen Teilpfad können wir wegen der Transitivität von R^+ durch $u_i R^+ u_{i+1}$ ersetzen, also das a aus unserem Pfad streichen. Das selbe gilt, wenn ein Element $u \in U$ mehrmals im Pfad vorkommt. Dieses können wir ebenfalls wegen der Transitivität von R^+ herausstreichen.

Wir wollen zunächst zeigen, dass der Algorithmus die Existenz eines solchen Pfades korrekt erkennt und erst später untersuchen, ob dies eine charakteristische Eigenschaft monotoner Teilmengen ist. Dabei sucht die zweite do^* Schleife jeweils den letzten Punkt a_n eines solchen Pfades und entfernt diesen. Sie tut dies so lange, bis in StBdCA kein weiterer derartiger Punkt existiert. Es ist genau dann jede Teilmenge von $A \cap \text{St}_R^*(\text{extComp})$ zusammenhängend, wenn am Ende der Iteration die noch verbleibende Menge StBdCA die leere Menge ist.

Angenommen $a_0 R^+ a_1 \cdots a_{n-1} R^+ a_n$ sei ein gerichteter Pfad in X , der jeden Punkt in StBdCA einmal besucht. Dann hat a_n die Eigenschaft $\text{StBdCA} \subseteq \text{St}_R^*(\{a_n\})$. Sei nun jedoch a_j ein weiterer Punkt mit dieser Eigenschaft, dann gilt $a_0 R^+ a_1 \cdots a_{n-1} R^+ a_n R^+ a_j$. Der Punkt a_j kann also an das Ende des Pfades angehängt werden. Wegen der Transitivität von R^+ kann zudem dieses a_j aus der Mitte des Pfades $a_0 R^+ a_1 \cdots a_{n-1} R^+ a_n$ entfernt werden, denn dann kann $\cdots a_{j-1} R^+ a_j R^+ a_{j+1} \cdots$ in unserem Pfad durch $\cdots a_{j-1} R^+ a_{j+1} \cdots$ ersetzt werden. Wenn solche weiteren Punkte a_j existieren sagt man, der topologische Raum ist nicht *trennbar* (oder genauer: nicht T_0 -trennbar). Trennbarkeit wollen wir in dieser Arbeit jedoch nicht behandeln.

Bei der Initialisierung der do^* -Schleifenvariablen wird StBdCA in Zeile 24 an die Menge $(\text{St}_R^*(\text{extComp}) \cup \text{Bd}_R^*(\text{extComp})) \cap A$ gebunden. Das Symbol centerp wird mit dem Prädikat $\text{St}_R^*(x) \supseteq \text{StBdCA}$ initialisiert. centerp ist für einen Punkt x genau dann wahr, wenn jede Umgebung um x die an die Variable StBdCA gebundene Menge umfasst. Durch die special -Deklarationen in Zeile 39 und Zeile 31 wird sichergestellt, dass bei Verwendung von centerp jeweils der aktuelle Wert von StBdCA verwendet wird. Danach wird die Variable center mit der maximalen Teilliste (x, \dots) von StBdCA initialisiert, deren erste Komponente x die Eigenschaft $\text{centerp}(x)$ hat oder, wenn ein solches x nicht existiert, mit der leeren Liste nil , die gleichzeitig den Wahrheitswert false darstellt.

Wenn nun in StBdCA ein solcher Pfad existiert, dann ist center nicht nil und die erste Komponente von center ist der Endpunkt eines solchen Pfades. Bei der nächsten Iteration wird nun dieser Endpunkt von StBdCA entfernt und es wird der jeweils vorhergehende Punkt gesucht. Man zeigt leicht durch Induktion über die Anzahl der Schleifendurchläufe bzw. Induktion über die Anzahl

der Elemente in **StBdCA**, dass, falls ein solcher Pfad existiert, der Algorithmus in Zeile 29 jeweils einen weiteren Punkt vom Ende dieses Pfades aus **StBdCA** entfernt. Wurden nun alle Punkte entfernt, dann bilden diese in der Reihenfolge ihres Entfernens den gesuchten surjektiven gerichteten Pfad. Kann irgendwann kein solcher Endpunkt gefunden werden, obwohl die Menge nicht leer ist, dann existiert auch kein surjektiver gerichteter Pfad, denn dazu ist die Existenz eines solchen Endpunktes notwendig. Oben hatten wir gezeigt, dass es aufgrund der Transitivität von R^+ nicht darauf ankommt, welcher der Kandidaten für einen Endpunkt vorher gewählt wurde, und dass nicht etwa falsche Punkte entfernt wurden, die nun hier benötigt werden. (In trennbaren Räumen ist dieser Punkt zudem stets eindeutig bestimmt.)

Die **do***-Schleife bricht ab, wenn ein solcher Endpunkt nicht existiert. Ist dann **StBdCA** nicht die leere Menge, dann terminiert der gesamte Algorithmus in Zeile 38 mit **nil**, also **false**. Ist hingegen **StBdCA** die leere Menge, dann terminiert die **do***-Schleife normal, hat korrekterweise erkannt, dass in $(U, R^+|_U)$ ein surjektiver gerichteter Pfad existiert, und akzeptiert die Menge U . \square

Wir haben nun behauptet, zu einem topologischen Datentyp (X, R) gibt es genau dann einen surjektiven gerichteten Pfad in (X, R^+) , wenn jede Teilmenge A von X zusammenhängend in (X, R) ist. Die eine Richtung der Inklusion ist klar, denn wenn ein solcher Pfad $(u_0 R^+ \cdots R^+ u_n)$ in (X, R^+) existiert, dann haben wir ja gesehen, dass wir wegen der Transitivität von R^+ alle Elemente aus diesem Pfad streichen können, die nicht in A liegen und erhalten so den Pfad $(a_0 R^+|_A \cdots R^+|_A a_m) =: (u_0 R^+ \cdots R^+ u_n)|_A$ als surjektiven gerichteten Pfad in A . Damit sind beliebige Elemente a_i und a_j in A durch den Teilpfad $(a_i R^+ \cdots R^+ a_j)$ verbunden, also ist A monoton in (X, R) . Sei umgekehrt jede Teilmenge $A \subseteq X$ zusammenhängend in (X, R) , dann zeigen wir durch vollständige Induktion über die Anzahl der Elemente in X , dass ein surjektiver gerichteter Pfad in (X, R^+) (damit aber i.A. nicht in (X, R)) existiert.

Ist die Anzahl von Elementen gleich Null, dann könnten wir den leeren Pfad $()$ als surjektiven gerichteten Pfad für den leeren Raum betrachten. Allerdings gibt es im leeren Raum keine Wege, die wir diesem Pfad zuordnen können. Der leere Raum ist natürlich dennoch wegzusammenhängend.

Wenn es genau ein Element $x \in \{x\} = X$ gibt, dann ist das 1-Tupel (x) der gesuchte Pfad.

Gelte nun die Behauptung für jeden topologischen Datentyp (X, R) der Größe n , habe X aber $n + 1$ Elemente. Sei also jede Teilmenge A von (X, R) zusammenhängend, dann ist auch jede Teilmenge der Form $X_a := X \setminus \{a\}$ für alle $a \in X$ zusammenhängend und jede Teilmenge von X_a als Teilmenge von X laut Annahme zusammenhängend. Dann ist $|X_a| = n$ und deshalb existiert laut Induktionsbehauptung für jedes der X_a ein surjektiver gerichteter Pfad $a_1 R^+ \cdots R^+ a_n$. Dann existiert für X_{a_n} ebenfalls ein solcher Pfad $b_1 R^+ \cdots R^+ b_n$. Es ist also $X \setminus \{a_n\} \subseteq \text{St}_R^*(b_n)$. Es ist aber auch die Teilmenge $\{a_n, b_n\} \subseteq X$ zusammenhängend und es gilt $a_n \neq b_n$. Damit ist $a_n R^+ b_n$ oder es ist $b_n R^+ a_n$, denn es gibt ja einen Weg von a_n nach b_n in $\{a_n, b_n\}$. Ist $b_n R^+ a_n$, dann können wir a_n einfach an das Ende des Weges in X_{a_n} anhängen und erhalten mit $b_1 R^+ \cdots R^+ b_n R^+ a_n$ einen surjektiven gerichteten Pfad in (X, R^+) . Gilt hingegen $a_n R^+ b_n$, dann ist $a_n \in \text{St}_R^*(b_n)$. Also gilt $X = \text{St}_R^*(b_n)$. Dann existiert jedoch auch ein solcher Pfad $c_1 R^+ \cdots R^+ c_n$ in X_{b_n} . Dann ist $c_n R^+ b_n$ und wir erhalten mit $c_1 R^+ \cdots R^+ c_n R^+ b_n$ ebenfalls einen surjektiven

gerichteten Pfad in (X, R^+) . Es sei noch einmal darauf hingewiesen, dass das allgemeine, NP-schwere Problem eines Hamilton-Pfades in einem simplen Graphen (X, R) von diesem viel einfacheren Spezialfall eines surjektiven gerichteten Pfades in (X, R^+) mit der transitiven Hülle von R genau zu unterscheiden ist.

Es verbleibt noch zu zeigen, dass eine Menge A genau dann monoton in (X, R) liegt, wenn die innere do^* -Schleife an Zeile 24 jede Zusammenhangskomponente $[p]$ in $X \setminus A$ akzeptiert, und dass der Algorithmus korrekt ist, dass also die äußere do^* -Schleife ab Zeile 3 jede Zusammenhangskomponente untersucht und der Algorithmus dann terminiert.

Wir haben oben schon gezeigt, dass der Algorithmus bei monotonen Mengen A in $\underline{X} = (X, R)$ jede Zusammenhangskomponente $[p]$ in $\underline{X \setminus A} := (X \setminus A, R^+|_{X \setminus A})$ akzeptiert.

Es bleibt nun noch zu zeigen, dass die Umkehrung ebenfalls gilt und A in \underline{X} monoton ist, wenn jede derartige Zusammenhangskomponente akzeptiert wird. Sei also $C \subseteq \underline{X}$ eine zusammenhängende Menge. Ist $C \subset A$, dann ist $C \cap A = C$ trivialerweise ebenfalls zusammenhängend. Ist $C \cap A = \emptyset$, dann erhalten wir mit $C \cap A$ ebenfalls trivialerweise eine zusammenhängende Menge. Die Monotonie-Eigenschaft ist also nur in den Fällen interessant, bei denen C Punkte in A hat und es zudem Punkte in C gibt, die nicht in A liegen. Seien nun x, y Punkte in $C \cap A$. Dann existiert ein Pfad $(x, p_0, p_1, \dots, p_{n-1}, p_n, y)$ in C von x nach y . Wir wollen zeigen, dass es dann immer auch einen Pfad $(x, q_0, q_1, \dots, q_{m-1}, q_m, y)$ innerhalb von $C \cap A$ gibt. Sei i der Index des ersten Punktes p_i , der nicht in A liegt, dann können wir den Pfad aufteilen in zwei Teilpfade (x, p_0, \dots, p_i) und $(p_i, \dots, p_{n-1}, p_n, y)$. Sei nun j der Index des ersten Punktes p_j von (p_i, \dots, y) , der in A liegt, dann haben wir nun drei Teilpfade (x, \dots, p_{i-1}, p_i) und $(p_i, \dots, p_{j-1}, p_j)$ und (p_j, \dots, y) . Dann liegt (x, \dots, p_{i-1}) vollständig in A , (p_i, \dots, p_{j-1}) vollständig in $X \setminus A$ und (p_j, \dots, y) ist der Rest des Pfades, der echt kürzer als (x, p_0, \dots, p_n, y) ist. Nun liegt der Pfad (p_i, \dots, p_{j-1}) in $X \setminus A$ und somit liegen alle Punkte dieses Pfades in der Zusammenhangskomponente $[p_i]$. Da das Paar (p_{i-1}, p_i) ebenfalls ein Pfad ist, gilt $p_{i-1} R^* p_i$ oder $p_i R^* p_{i-1}$ und damit ist p_{i-1} Element der im Algorithmus mit StBdCA bezeichneten Menge. Genauso liegt p_j in dieser Menge, da ja $p_{j-1} \in [p_i]$ gilt und der Pfad (p_{j-1}, p_j) existiert. Dann ist aber (p_{i-1}, p_j) ein Pfad, da der Algorithmus laut Voraussetzung StBdCA akzeptiert und die Menge $\{p_{i-1}, p_j\}$ somit zusammenhängend ist.

Damit kann der Pfad $(x, p_0, \dots, p_{i-1}, p_i, \dots, p_{j-1}, p_j, \dots, p_n, y)$ zu dem Pfad $(x, p_0, \dots, p_{i-1}, p_j, \dots, p_n, y)$ gekürzt werden. Nach spätestens n derartigen Kürzungen verläuft der Pfad vollständig in A und damit auch vollständig in $C \cap A$. Damit sind die zwei beliebig gewählten Punkte x, y in $C \cap A$ durch einen Weg in $C \cap A$ verbunden und damit ist diese Menge zusammenhängend. Damit akzeptiert der Algorithmus nur monotone Mengen.

Es bleibt nur noch zu zeigen, dass der Algorithmus jede Zusammenhangskomponente $[p]$ in $X \setminus A$ untersucht. Es wird bei jeder Schleifenwiederholung $n + 1$ in Zeile 5 die zuletzt untersuchte Zusammenhangskomponente $[p_n]$ von coAIter entfernt und danach ein weiterer Punkt p_{n+1} aus $\text{coAIter} \setminus [p_n]$ ausgewählt, um $[p_{n+1}]$ zu berechnen. Damit schrumpft coAIter bei jeder Iteration n mindestens um p_n , ist also irgendwann leer und der Algorithmus terminiert spätestens dann in Zeile 9 mit \mathbf{t} , wenn A monoton ist, ansonsten bereits vorher mit \mathbf{nil} .

Damit ist der Algorithmus korrekt.

Die Komplexität (also Rechenzeit) des Algorithmus soll nur grob nach oben abgeschätzt werden, da er nicht besonders effizient ist. Ohne Beweis behaupten wir, die folgenden Algorithmen implementieren **star** und **boundary**. Dabei wird über die eingegebenen Mengen B , X und die Relation R , implementiert als Liste von zweielementigen Listen, in mehreren verschachtelten Schleifen iteriert.

```
(defun star (B X R)
  (do* ( ( result B )
        ( todo B done )
        ( done () () ) )

    ( (null todo) (intersection X result) )

    (dolist (a-b R)
      (if (and (member (second a-b) todo)
                (not (member (first a-b) result)))
          (progn (pushnew (first a-b) done)
                  (push (first a-b) result) ) ) ) ) )

(defun boundary (B X R)
  (star B X
    (mapcar #'(lambda (a-b)
                (list (second a-b) (first a-b)))
            R ) ) )
```

Der von der Funktion **star** verwendete **progn**-Ausdruck bedeutet nichts anderes als die Zusammenfassung aufeinander folgender Ausdrücke zu einem Gesamtausdruck. Die Ausdrücke (**progn** a b) und (**let*** $()$ a b) sind äquivalent. Die Standardfunktionen **member** und **pushnew** für die Mengenoperationen suchen die Argumentlisten sequenziell ab und haben damit lineare Laufzeit bezüglich der Größe der eingegebenen Mengen $n = |B| + |X| + |R|$. Damit haben auch **star** und **boundary** polynomielle Laufzeit, es gibt also einen konstanten Exponenten $k \in \mathbb{N}$ und eine Konstante $c \in \mathbb{N}$, so dass spätestens nach $n^k + c$ Rechenschritten das Ergebnis feststeht. Dies ist zwar keine besonders gute obere Schranke aber da der Algorithmus ohnehin nicht sonderlich effizient ist, erscheint es nicht lohnend, seine Komplexität genauer zu berechnen. Natürlich gibt es auch effizientere Algorithmen als die hier vorgestellten, etwa durch den Einsatz von Binärbäumen, Hashtables etc. Wenn effizientere Implementierungen z.B. objektorientiert programmiert sind, dann könnten **star** und **boundary** nach folgenden Methoden berechnet werden. Es wird vorausgesetzt, dass für **unify**, **subtract** etc. entsprechende Methoden einer Klasse **Math-Set** oder **Relation** existieren. Zudem wird die Existenz einer Methode **relate-all**((**M Math-Set**) (**R Relation**)) angenommen. Die Komplexität der folgenden Algorithmen hängt dann stark von den Implementierungen dieser Methoden ab. Die äußere **do***-Schleife hingegen iteriert $n + 1$ mal, wenn n die Dimension von (X, R) ist.

```

(defmethod star ((B Math-Set) (X Math-Set) (R Relation))

  (do* ( ( result B (unify B done) )
        ( todo  B (subtract done result) )
        ( done  (relate-all B R) (relate-all todo R) )

        ( (empty todo) (intersect X result) )
  )

  (defmethod boundary ((B math-set) (X math-set) (R math-set))
    (star B X (transpose R) )
  )

```

Auch das Ergebnis von `monotonep` wird durch verschachtelte Iterationen über die Eingabemengen gewonnen und verwendet dabei nur Operationen mit polynomieller Rechenzeit, ist also selbst ebenfalls polynomiell und hat zudem erhebliche Reserven zur Optimierung, die wir hier aber nicht mehr nutzen können, denn es ist in der Tat die auf Seite 178 befürchtete Situation eingetreten, und diese Arbeit ist hier zu Ende.

Literaturverzeichnis

- [AF90] ARKHANGEL'SKIĬ, A.V. und V. V. FEDORCHUK: *The Basic Concepts and Constructions of General Topology*. In: ARKHANGEL'SKIĬ, A. V. und L. S. PONTRYAGIN (Herausgeber): *General Topology I*, Band 17 der Reihe *Encyclopaedia of Mathematical Sciences*, Kapitel 1. Springer-Verlag, 1990.
- [Ale37] ALEXANDROFF, PAVEL: *Diskrete Räume*. *Matematiečeskij Sbornik*, 44(2):501–519, 1937.
- [Aug95] AUGENBROE, GODFRIED: *Combine 2 Final Report*. Technischer Bericht, CEC-JOULE Report, Brussels, 1995.
- [Bal01] BALZERT, HEIDE: *UML kompakt*. Spektrum, Akad. Verl., 2001.
- [Bau75] BAUMGART, BRUCE G.: *Winged-Edge Polyhedron Representation for Computer Vision*. In: *National Computer Conference*, 1975.
- [BF] BILL, RALF und DIETER FRITSCH: *Grundlagen der Geo-Informationssysteme*. Wichmann.
- [Bot06] BOTH, PETRA VON: *Ein systemisches Projektmodell für eine kooperative Planung komplexer Unikate*. Doktorarbeit, 2006.
- [Bre01] BREUNIG, MARTIN: *On the Way to Component-Based 3D/4D Geoinformation Systems*, Band 94 der Reihe *Lecture Notes in Earth Sciences*. Springer, 2001.
- [Bro88] BROWN, KENNETH S.: *Buildings*. Springer-Verlag, Berlin, Heidelberg, New York, 1988.
- [Can95] CANTOR, GEORG: *Beiträge Zur Begründung der Transfiniten Mengenlehre*. *Mathematische Annalen*, 46(4):481–512, 1895.
- [Can62] CANTOR, GEORG: *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*. Olms, Reprograf. Nachdr. d. Ausg. Berlin, Springer, 1932 Auflage, 1962.
- [Cod79] CODD, EDGAR F.: *Extending the Database Relational Model to Capture More Meaning*. *ACM Trans. Database Syst.*, 4(4):397–434, 1979. <http://doi.acm.org/10.1145/320107.320109>, db/journals/tods/Codd79.html.

- [Cod83] CODD, EDGAR F.: *A relational model of data for large shared data banks*. Commun. ACM, 26(1):64–69, 1983. <http://doi.acm.org/10.1145/357980.358007>.
- [Cod90] CODD, EDGAR F.: *The Relational Model for Database Management*. Addison-Wesley, 1990.
- [Dam00] DAMRATH, RUDOLF: *Relationale Prozessmodellierung in kooperativer Gebäudeplanung (DFG-SPP1103) [online]*. <http://www.bauinf.uni-hannover.de/dfg-spp1103/>, 2000.
- [Die69] DIETZE, PETER: *Der Bewertungsprozess und Verfahren zur Bewertung in der Stadtplanung*. In: JOEDICKE, JÜRGEN (Herausgeber): *Bewertungsprobleme in der Bauplanung*. Krämer, 1969.
- [DK02a] DAMRATH, PROF. DR.-ING. RUDOLF und M. KÖNIG: *Relationale Prozessmodellierung in kooperativer Gebäudeplanung*. In: *Zwischenbericht über die 1. Förderphase des DFG Schwerpunktprogramms 1103*, www.dfg-spp1103.de. <http://www.dfg-spp1103.de/de/zwischenberichte/Phase1/Hannover.htm>, 2002.
- [DK02b] DAMRATH, RUDOLF und MARKUS KÖNIG: *Relationale Prozessmodellierung in kooperativer Gebäudeplanung, Folien zum 2. Berichtskolloquium Darmstadt 2002 [online]*. http://www.bauinf.uni-hannover.de/dfg-spp1103/bericht_darmstadt_2002-04-19.pdf, 2002. zuletzt besucht: 11. Oktober 2007.
- [Eas99] EASTMAN, CHARLES M.: *Building Product Models: computer environments supporting design and construction*. CRC Press LLC, 1999.
- [Eng78] ENGELKING, RYSZARD: *Dimension theory*. North-Holland, 1978.
- [Ern74] ERNÉ, MARCEL: *Struktur- und Anzahlformeln für Topologien auf endlichen Mengen*. manuscripta math., (11):221 – 259, 1974.
- [FML02] FRADIN, DAVID, DANIEL MENEVEAUX und PASCAL LIENHARDT: *Partition de l'espace et hiérarchie de cartes généralisées : application aux complexes architecturaux [online]*. <http://afig2002.univ-lyon1.fr/ACTES/fradin.pdf>, 2002.
- [FS98] FOWLER, MARTIN und KENDALL SCOTT: *UML distilled*. Addison-Wesley, Nachdr. Auflage, 1998.
- [Gie05] GIELSDORF, FRANK: *Ausgleichsrechnung und raumbezogene Informationssysteme*. Habilitationsschrift. Technische Universität Berlin, 2005.
- [Hac73] HACKER, WINFRIED: *Allgemeine Arbeits- und Ingenieurpsychologie*. Dt. Verl. d. Wiss., 1973.
- [Hac86] HACKER, WINFRIED: *Arbeitspsychologie*. Huber, Neufassung von „Allgemeine Arbeits- und Ingenieurpsychologie“ Auflage, 1986.
- [Har74] HARARY, FRANK: *Graphentheorie*. Oldenbourg, 1974.

- [Hat02] HATCHER, ALLEN: *Algebraic Topology*. Cambridge University Press, 2002. Verfügbar auch unter <http://www.math.cornell.edu/~hatcher/>.
- [Hau78] HAUSDORFF, FELIX: *Grundzüge der Mengenlehre*. Chelsea, Repr. Auflage, 1978.
- [Her86a] HERRLICH, HORST: *Einführung in die Topologie*. Heldermann, 1986.
- [Her86b] HERRLICH, HORST: *Topologie 1: Topologische Räume*. Heldermann, 1986.
- [Hid95] HIDDERS, J.: *An Isotopic Invariant for Planar Drawings of Connected Planar Graphs*. Technischer Bericht, Eindhoven University of Technology, 1995.
- [Hof82] HOFFMANN, CHRISTOPH M.: *Group-theoretic algorithms and graph isomorphism*. Springer, 1982.
- [ISO03] ISO - INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (Herausgeber): *Integrated generic resource: Geometric and topological representation*, Band ISO 10303-42 der Reihe *Industrial automation systems and integration – Product data representation and exchange*. ISO, 2003.
- [Jän01] JÄNICH, KLAUS: *Topologie*. Springer, 7. Aufl. Auflage, 2001.
- [Joe76] JOEDICKE, JÜRGEN: *Angewandte Entwurfsmethodik für Architekten*. Krämer, 1976.
- [Kan87] KANT, IMMANUEL: *Kritik der reinen Vernunft, Zweite hin und wieder verbesserte Auflage*. Riga, 1787. zitiert nach [KH98].
- [KH98] KANT, IMMANUEL und INGEBORG [HRSG.] HEIDEMANN: *Kritik der reinen Vernunft*. Reclam, Stuttgart, 1998.
- [KS77] KUHNE, VOLKER und HANS SOMMER: *Projektsteuerung im Hochbau*. Bauverlag, 1977.
- [Kur68] KURATOWSKI, KAZIMIERZ: *Topology*, Band II. Acad.Pr., New ed., rev. and augm. Auflage, 1968.
- [LAF⁺] LIEBICH, THOMAS, YOSHINOBU ADACHI, JAMES FORESTER, JUHA HYVARINEN, KARI KARSTILA und JEFFREY WIX: *Industry Foundation Classes, IFC2x Edition 3, [online]*. http://www.iai-international.org/Model/R2x3_final/index.htm. zuletzt besucht: 11. Oktober 2007.
- [LG03] LORENTZ, DIANA und JOAN GREGOIRE: *Oracle Database SQL Reference 10g Release 1 (10.1)*. http://download-uk.oracle.com/docs/cd/B14117_01/server.101/b10759/toc.htm, 2003. Part No. B10759-01.
- [Lie94] LIENHARDT, PASCAL: *N-dimensional generalized combinatorial maps and cellular quasi-manifolds*. International Journal of Computational Geometry and Applications, 4(3):275–324, 1994.

- [Mai83] MAIER, DAVID: *The theory of relational databases*. Pitman, 1983.
- [Män88] MÄNTYLÄ, MARTTI: *An introduction to solid modeling*. Computer Science Pr., 1988.
- [MS74] MARCH, LIONEL und PHILIP STEADMAN: *The geometry of environment*. MIT Pr., 1. US ed. Auflage, 1974.
- [OS01] OTT, THOMAS und FRANK SWIACZNY: *Time-Integrative Information Systems*. Springer, Berlin, Heidelberg, New York, 2001.
- [Oss92] OSSA, ERICH: *Topologie*. Nummer 42 in *Vieweg-Studium*. Vieweg, 1992.
- [Oze00] OZEL, FILIZ: *Spatial Databases and the Analysis of Dynamic Processes in Buildings*. In: *CAADRIA 2000 Proceedings of the Fifth Conference on Computer Aided Architectural Design Research in Asia*, Seiten 97–106, Singapur, 2000. <http://www.arch.usyd.edu.au/kcdc/journal/IJDC2/files/caadria/caadria2000.html>.
- [PB03] PAUL, NORBERT und PATRICK ERIK BRADLEY: *Topological Houses*. In: *Internationales Kolloquium über Anwendungen der Informatik und Mathematik in Architektur und Bauwesen, IKM*, 2003. (online) <http://e-pub.uni-weimar.de/volltexte/2005/353/>.
- [Rös98] RÖSCH, NORBERT: *Topologische Beziehungen in Geo-Informationssystemen*. Doktorarbeit, Universität Karlsruhe (Geodät. Inst.), <http://digbib.ubka.uni-karlsruhe.de/volltexte/33698>, 1998.
- [Sch75] SCHUBERT, HORST: *Topologie Eine Einführung*. Teubner, 4. Auflage Auflage, 1975.
- [Sed91] SEDGEWICK, ROBERT: *Algorithmen*. Addison-Wesley, 1. Aufl. Auflage, c1992 erschienen 1991.
- [Ste90] STEELE, GUY L.: *Common LISP*. Digital Pr., 2. ed. Auflage, 1990.
- [Sti80] STILLWELL, JOHN: *Classical topology and combinatorial group theory*. Springer, 1980.
- [Suh01] SUH, NAM P.: *Axiomatic design*. Oxford University Press, 2001.
- [Sut63] SUTHERLAND, IVAN E.: *Sketchpad : a man-machine graphical communication system*. Doktorarbeit, Massachusetts Institute of Technology, 1963. [online]<http://www.cl.cam.ac.uk/TechReports/UCAM-CL-TR-574.pdf>.
- [Sut80] SUTHERLAND, IVAN E.: *Sketchpad : a man-machine graphical communication system*. Outstanding dissertations in the computer sciences. New York : Garland Publ., 1980.
- [The97] THE GIS HISTORY PROJECT: *The GBF/DIME Case Study [online]*. <http://www.geog.buffalo.edu/ncgia/gishist/DIME.html>, 1997.

- [TK78] TSICHRITZIS, DENNIS und ANTHONY C. KLUG: *The ANSI/X3/SPARC DBMS Framework Report of the Study Group on Database Management Systems*. Information Systems, 3(3):173–191, 1978.
- [Ves03] VESPERMAN, JENNIFER: *Essential CVS*. O'Reilly&Associates, Sebastopol, CA, 2003.
- [Wei85] WEILER, KEVIN: *Edge-Based Data Structures for Solid Modeling in Curved-Surface Environments*. j-IEEE-CGA, 5(1):21–40, Jan 1985.
- [Wie94] WIENDIECK, GERD: *Arbeits- und Organisationspsychologie*. Quintessenz, 1994.
- [Zuk81] ZUKAV, GARY: *Die tanzenden Wu Li Meister*. Rowohlt, 1981.

Abbildungsverzeichnis

1.1	IfcDoor in EXPRESS-G	6
1.2	COMBINE IDM Kernmodell	8
1.3	Gebäudetopologie des SPP 1103	10
1.4	DIME	11
1.5	Winged-Edge-Dach	15
1.6	G-Map-Dach	20
2.1	Zimmerrelation	32
2.2	Zimmer- und Bauteilerrelation	34
2.3	Relationenprodukt	36
2.4	offene Kugel	47
2.5	Innerer-, Rand- und Äußerer Punkt	52
2.6	Architektonische Zerlegung in Flatland	59
2.7	Fundamentalgruppe einer Säulenhalle	61
2.8	Unterm modul und Faktormodul	64
2.9	Türkoordinaten	73
2.10	CWKomplex	77
4.1	Gemeinsame Verfeinerung	130
4.2	Disjunkte Kopien eines Raums	136
4.3	topologischer Verbund	153
4.4	Stapel und Diff	162
4.5	Trivialität eines Wandabschnitts	164
4.6	Naiver Verbund mit Bauteilbibliothek	166
4.7	Monotoner Verbund mit Bauteilbibliothek	179
4.8	Monotone Bauteilbibliothek	191
6.1	UML-Notation eines Entitätstyps	198
6.2	Funktionsverzeichnis	200
6.3	UML-Notation für topologische Räume	201
6.4	Planungsobjekt	203
6.5	Plandokumente	205
6.6	Versionsverwaltung von Räumen	206
6.7	Detailbibliothek	208
6.8	Wirkung und Wirkungsübertragung	210
6.9	Das gesamte System	212

Index

- Abbildung, **40**
 - , Gleichheit, 41
 - , direktes Produkt von –en, 42
 - , einer Relation, 43
 - , identische, 41
 - , konstante, 42
 - , lineare, **65**
 - Bild, **65**
 - Kern, **65**
 - , monotone, *siehe* Monotonie
 - , natürliche, 43
 - , partielle, **41**
 - , stetig und partiell, **181**
 - Anheftungsabbildung, 182
 - , stetige, **53**
 - in Datenbanken, 118, **119**
 - und Maßstab, 54
 - , zelluläre, 76
 - skomposition, 42
 - Bild, 43
 - Einschränkung einer –, 43
 - Einsetzungs–, 42
 - in LISP, 43
 - Faktorisierung, 175
 - Fortsetzung einer –, 43
 - gemeinsame Fortsetzung, 44
 - Stetigkeit, 228–229
 - Inklusion, 41
 - Involution, 19
 - Produkt–, 42
 - Projektion, 42
 - Tupel als –, 42
 - Urbild, 43
 - Verknüpfung als –, 44
 - Verknüpfung von –en, 44
- Abstand, *siehe* Metrik
- Antisymmetrie, *siehe* Eigenschaften einer Relation
- Äpfel
 - , mit Birnen addieren, 65
- Äquivalenz
 - klasse, **33**
 - relation, **33**
 - erzeugte –relation, **33**
 - und Wegzusammenhang, 171
 - Homöomorphismus als –, 55
 - Homotopie, 217
 - Quotientenmenge, 33
 - von Cauchyfolgen, **219**
 - von Kategorien, **80**
 - Wegverbundenheit, 170
 - Zerlegung einer Menge, **34**
- Architekturausbildung, 198
- Architektonischer Raum, *siehe* Quotientenraum
- Ästhetik
 - Vermeiden der Beurteilung, 163
- Attributsymbole, *siehe* Menge
- Axiomatic Design, 214
- Bahn, 19
- Basis, **63**
- Bedeutung
 - , eines Plans, 54
 - Trennen und Verbinden, 209
- Berufsbild, 198
- Berührungspunkt, *siehe* Hülle, abgeschlossene
- Beschreibungssprache, 5
 - EXPRESS
 - NULL-Werte, 6
 - topology_schema, 6
 - EXPRESS und EXPRESS-G, 5
- Beziehungstyp, 163
- Birnen
 - , mit Äpfeln addieren, 65
- Datenbank, **96**
 - , topologische, **118**

- und alexandroffsche Räume, 122
 - Normalisierung, 163
- Datenbankschema, **96**
- Datentyp
 - , topologische
 - Kategorie der –, **121**
- Detailbibliothek
 - Notwendigkeit der Monotonie, 164–165
- Dimension
 - dynamisch und statisch, 107
 - einer Relation, **126**, 160
 - eines affinen Raums, **194**
 - eines Moduls, **63**
 - Überdeckungsdimension, 224
- Dirichletfunktion, 216
- do*-Schleife, 235
 - Abbruchbedingung, 235
 - Iterationsergebnis, 235
 - Schleifenrumpf, 235
 - Schleifenvariablen, 235
 - Semantik, 236
- Domänenmodell, 196
- Dualität
 - alexandroffscher Räume, 127
 - DIME, 12
 - G-Maps, 21
 - topologischer Datenbanken, 127
 - Winged-Edge, 16
- Entitätstyp, 163
- Euler-Poincaré-Formel, 16
- Euleroperatoren, 161
- Folge, **218**
 - , beschränkt monoton, 221
 - Äquivalenz, **219**
 - Cauchyfolge, 24, 26, **219**
 - Konvergenz einer –, 26
- Fundamentalgruppe, 60
- funktionale Abhängigkeit, **94**
- Funktionsprogramm
 - Definition, 199
 - Nutzen, 199
 - Standard, 199
 - Krankenhäuser, 199
- Funktor, **79**
 - , identischer, 79
 - zum Unterscheiden von Objekten, 79
- Geoinformationssystem, **11**
 - als Vorbild, 2
- Geschossstapel, 161
- GIS, *siehe* Geoinformationssystem
- Graph
 - , ungerichteter
 - als CW-Komplex, 76
 - , ungerichteter
 - Kategorie der –en, 78
 - als Planungswerkzeug, 66
 - gerichteter, **68**
 - Homologiegruppe, 72
 - Hypergraph, 126
 - ungerichteter, **66**
 - als topologischer Raum, 67
 - und Randoperator, 68
 - Zyklus und Zykel, 69
- Hamilton-Pfad, 239
 - NP-Vollständigkeit, 240
- Handlung, **196**
- HOAI, 197
- Homöomorphismus, **55**, 55–56
 - und Graphenisomorphismus, 56
- Homotopie, **216**
- Hülle
 - , abgeschlossene, 51, **51**
 - in topologischen Datenbanken, **127**
 - , transitive, **36**
 - in Abfragesprachen, 39, 91
 - in gegenwärtigen RDBMS, 213
 - und stetige Datenbankabbildungen, 120
 - und topologische Verfeinerung, 135
 - , transitive und reflexive, **36**
- IDM — integriertes Datenmodell, 6
- Innenpunkt, *siehe* Inneres
- Inneres, **51**
- Irreflexivität, *siehe* Eigenschaften einer Relation
- Isomorphismus
 - , allgemeiner, 79

- , von Moduln, **66**
- Homöomorphismus, 55
- Kategorie, **77**
 - , opposite, 78
 - , triviale, 78
 - der kleinen $-en$, 78
- Kern
 - , einer linearen Abbildung, **65**
 - , offener, 51, *siehe* Inneres
- Kette, 65
- Klassendiagramm, 196
- Kompaktheit, **221**
 - bei alexandroffschen Räumen, 224
 - des Einheitsintervalls, 221
- Komplex
 - , algebraischer
 - Kategorie der $-e$, 78
 - , architektonischer, 99
 - , relationaler, **106, 112**
 - Informationsverluste, 112–114
 - Morphismus eines $-$, **106, 112**
 - algebraischer, **72**
 - CW-Komplex, **74**
- Komplexität, 235
 - , polynomielle, 243
 - von COMBINE, 7
 - von $\mathcal{D}Top$, 124
- konzeptuelles Schema, *siehe* Domänenmodell
- Korrektheit von Algorithmen, 234
- Kugel
 - , abgeschlossene, 51
 - , offene, **46**
- let*-Ausdruck, 237
- Linearkombination, 65
- lokale Variablen, *siehe* let*-Ausdruck
 - special, 240
- Mannigfaltigkeit, **74**
- Matrix, **101**
 - , partielle, **108**
 - Einheit, 108
 - Multiplikation, **108**
 - Rechenregeln, 109
 - $-$ produkt, **101**
 - als Relation, 103
 - Kategorie der Matrizen, 102
- Menge, **22**
 - , abgeschlossene, **49**
 - , offene, 49
 - Beispiele, 23
 - Gleichheit, 27
 - Teilmenge, 27
 - Beispiele, 27–28
 - Vereinigung, Schnitt, Rest, **28**
 - Zerlegung einer $-$, **34**
- Metrik, **45**
 - , euklidische, 45
 - , triviale, 45
 - Manhattan-, 46
- metrischer Raum, *siehe* Metrik
- Modellierungsschichten, 4
 - Referenzarchitektur, 5, 9
- Modul, **62**
 - , freier, **63**
 - , trivialer, 63
 - Faktor-, 64
 - freier $-$ über einer Menge, 64
 - Quotienten-, *siehe* Faktor-
 - Unter-, 63
- Monotonie, **168**
 - injektiver Abbildungen, 177, *siehe* Monotonie von Mengen
 - Kriterium für Detaildatenbanken, 188
 - Nutzen für Detailbibliotheken, 164–165
 - surjektiver Abbildungen, 174–175
 - und Morphismen, 169
 - von Abbildungskompositionen, 176
 - von Mengen, **177**
- Morphismus
 - Abbildung, 44
 - Funktor in Kat , 78
 - lineare Abbildung, 65
 - monotone Abbildung, 169
 - Relation, 35
 - Selektion, Projektion und Umbenennung, 87
 - stetige Abbildung, 55

- stetige Datenbankabbildung, 119
- stetige und partielle Abbildung, 182
- von Komplexen, **73**
- von ungerichteten Graphen, 67
- Normalformen, 93–95
 - , erste, **93**
- offene Überdeckung, **221**
- Operation, 44
- operatives Abbildsystem, 197
- Operator, 44
- Paar, *siehe* Tupel
- Pfad, **171**
 - und Weg, 172
- Plan, **197**
- Planung, *siehe* Plan, Handlung
- Planungssystem
 - Plan, 204
 - Rechtevergabe, 204
- Prädikat
 - für Relationen, **83**
 - in LISP, **233**
- Rand, **51**
 - , algebraischer, 68
- Randpunkt, *siehe* Rand
- Raum, *siehe* metrischer –, topologischer –
 - , affiner, **193**
 - , alexandroffscher, **60**
 - Kategorie der –, **121**
 - und dualer –, 60
 - und topologische Datenbanken, 122
 - , architektonischer, 99
 - , diskreter, 50
 - , indiskreter, 50
 - , kompakter, *siehe* Kompaktheit
 - , leerer, 50
 - , metrisierbarer, 50
 - , natürlicher, 50
 - bei GIS, 11
 - Quotienten–, 58
 - architektonischer –, 57
 - Teil–, 50
 - Umgebung einer Türe, 50
 - und Fundamentalgruppe, 60
- Raumprogramm
 - Definition, 199
 - Nutzen, 199
 - Standard, 199
- Realität und Mathematik, 98
- Referentielle Integrität, **96**, 95–96
- Reflexivität, *siehe* Eigenschaften einer Relation
- Relation, **30**
 - , topologische Verfeinerung und Vergrößerung, 131
 - , finale, 141, **142**
 - , initiale, 141, **142**
 - in Oracle-SQL, 142
 - , topologisch feinere und gröbere, **128**
 - , vereinigungsverträglich, 30, 82
 - enprodukt, **35**
 - Abfragen, 89, *siehe* Relationenalgebra
 - Abfrageschema, **89**
 - Äquivalenz–, *siehe* Äquivalenz
 - Beispiele, 30–31
 - Dimension einer –, **126**
 - direktes Produkt von –en, **137**
 - Eigenschaften, **31**
 - Einschränkung, **174**
 - Grad einer –, **126**
 - Kategorie der –en, 78
 - Potenzen einer –, **36**
- Relationenalgebra
 - Abfrageschema, **89**
 - äußerer Verbund, **180**, 179–181
 - Ersatz, 186
 - Verbot, 186
 - Gleichheitsverbund, 90
 - und Relationenprodukt, 90
 - natürlicher Verbund, **90**
 - kartesisches Produkt, **90**
 - Potenzen einer –, 92
 - Prädikat, **83**
 - Projektion, **85**
 - auf keine Attribute, 86
 - Relationenprodukt, 90
 - Selektion, **84**
 - Θ -Verbund, 90
 - topologisch, 144–155

- Umbenennung, **86**
- Vereinigung, Schnitt und Rest, **83**
- Vereinigungsverträglichkeit, 82
- Relationenschema, **80**
- Prädikat, **83**
- Relation zu $-$, **80**
- Tupel zu $-$, **80**
- Schlüssel, **94**
 - kandidaten, 95
 - Primärschlüssel, 95
- Stachel, 19
- Stern, **114**
- Superschlüssel, **94**
- Symmetrie, *siehe* Eigenschaften einer Relation
- Tabelle, *siehe* Relation
- Tabellendefinition, *siehe* Relationenschema
- Tomaten
 - , zu Äpfeln und Birnen addieren, 65
- Topologie, **49**
 - Verfeinerung, Vergrößerung
 - Assoziativität, 131
 - , feinere und gröbere, **128**
 - , finale, **140**
 - , initiale, **140**
 - bei partiellen Abbildungen, **184**
 - Mehrdeutigkeit bei partiellen Abbildungen, 184
 - , von Mengen erzeugte, *siehe* Subbasis
 - , von Relation erzeugte, 116
 - im DFG SPP 1103, 9
 - in COMBINE, 6
 - in den Industry Foundation Classes, 9
 - in DIME, 11–12
 - in G-Maps, 21
 - in RATAS, 7
 - in STEP, 6
 - in Winged-Edge, 14–16
 - Speicherplatzkomplexität, 124
 - Spezialisierungspräordnung, 122
 - Subbasis, **130**
 - Verfeinerung, Vergrößerung, **129**
 - topologischer Raum, *siehe* Topologie
 - Kategorie der $-e$, 78
 - Transitivität, *siehe* Eigenschaften einer Relation
 - Trennbarkeit, 124, **240**
 - Trivialität, **164**
 - Tupel, **30**
 - Umgebung, **48**
 - , metrische, **46**
 - Eigenschaften, 46
 - UML, *siehe* Unified Modeling Language
 - Unified Modeling Language, 196
 - Aggregation, 201
 - Attribute, 199
 - berechnete Werte, 203
 - Beziehungen, 199
 - Vielfachheit, 199
 - Beziehungsklassen, 200
 - Einschränkung, 203
 - Entitätstyp, 198
 - Klasse, 198
 - Komposition, 201
 - Stereotyp, 199
 - Versionskontrolle
 - Aufbewahrungsort, 207
 - Ausgabe, 206
 - diff und patch, **159**
 - Repository, *siehe* Aufbewahrungsort
 - Revision, 206
 - Transaktion
 - commit, 207
 - update, 207
 - Weg, *siehe* stetige Abbildung, **54**
 - , elementarer, **170**
 - , endlicher, **173, 215**
 - , unendlicher
 - Existenz, 215
 - endliche Abkürzung, 173
 - innerhalb eines Sterns, 230
 - Rückweg, **170**
 - und homotoper endlicher $-$, 173, 217

- und Pfad, 172
- Wohldefiniiertheit, 25
- \mathbb{Z} -Modul
 - Kategorie der $-e$, 78
- Zahlen
 - , reelle, 218
- Zelle, **74**
- Zusammenhang, **165**
 - Kriterium in \mathcal{DTop} , 174
 - Komplexität, 174
 - und Wegzusammenhang, 167
 - von Bildern, 168
 - Weg-, **165**