

# TIB

TECHNISCHE  
INFORMATIONSBIBLIOTHEK

## LinSearch – Effiziente Indizierung an der Technischen Informationsbibliothek, Hannover

D. Daberkow, P. Mensing, I. Sens, C. Todt  
LIS'2012, 02. August 2012



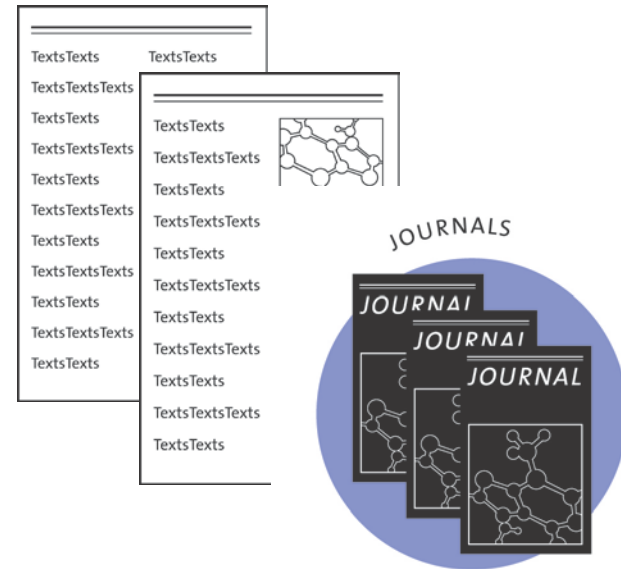
# Die Technische Informationsbibliothek



- **Deutsche Zentrale Fachbibliothek** für Technik sowie Architektur, Chemie, Informatik, Mathematik und Physik
- Bund-Länder-finanziert
- Mitglied der Leibniz-Gemeinschaft (WGL)
- Alle 7 Jahre Evaluierung durch externe international besetzte Gutachtergruppe (zuletzt 2.2011)

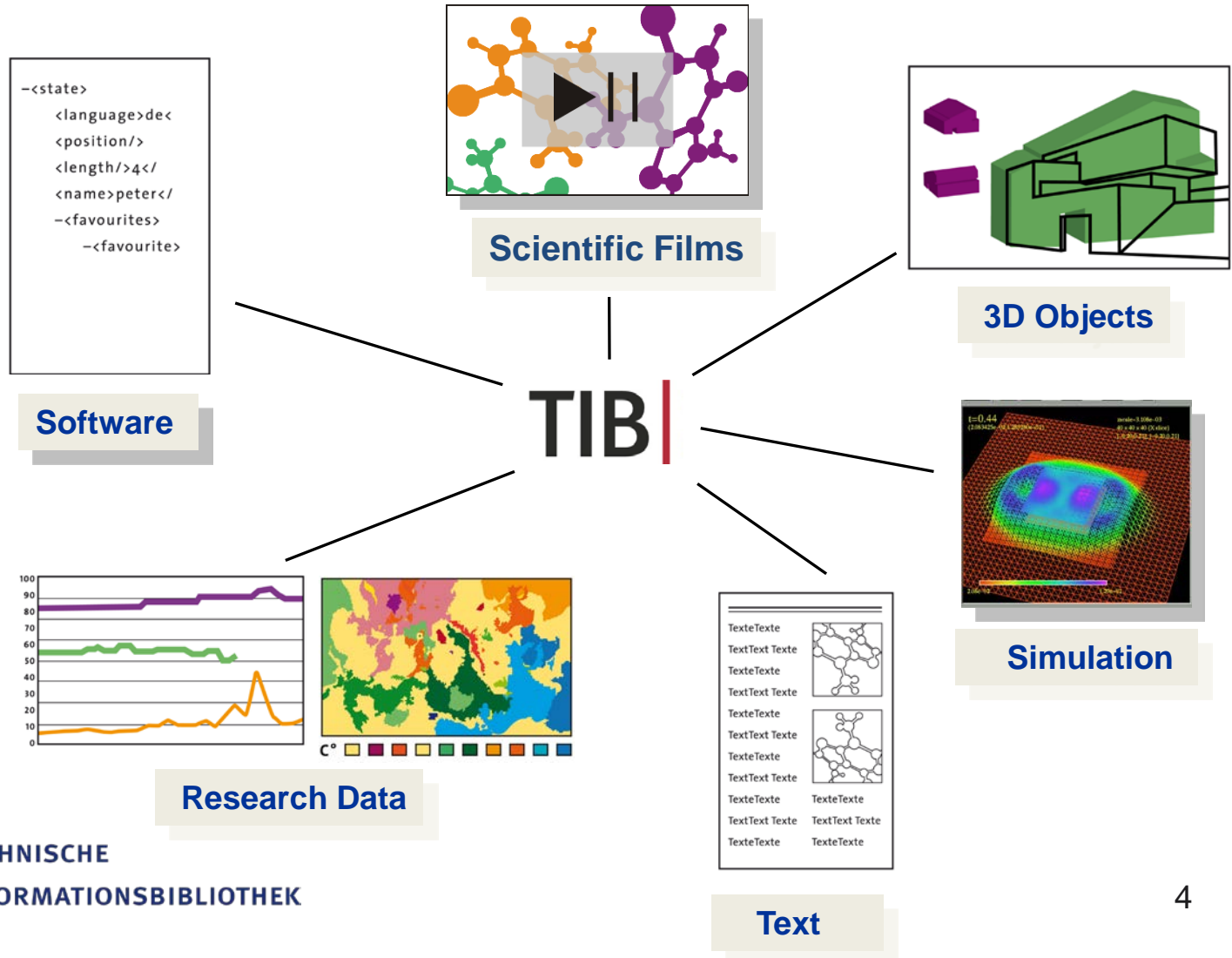
# TIB – heute

## Globaler Literaturversorger für wissenschaftlich-technische Informationen

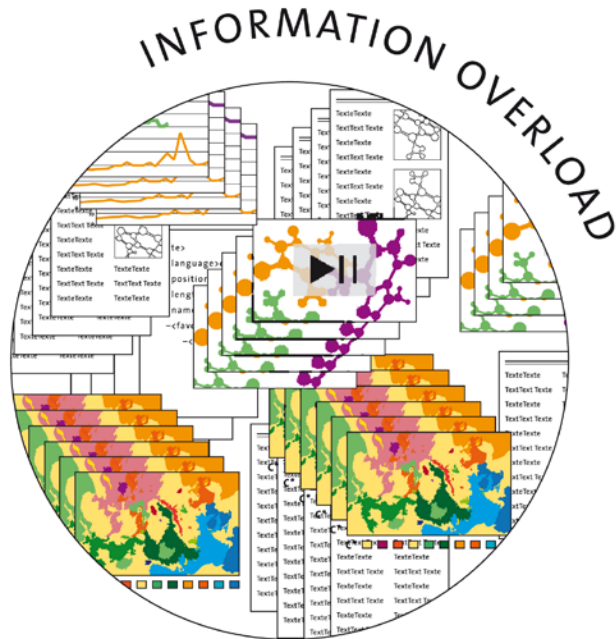


Traditionell textuelle Materialien

# TIB Strategie – Verknüpfung und Suche über alle Typen digitaler Dokumente



# Problembeschreibung



Die ständig wachsende Menge an Daten und Informationen lässt sich nicht mehr nur durch „Manpower“ bewältigen.





# GetInfo – Portal für wissenschaftlich publizierte Fachinformation

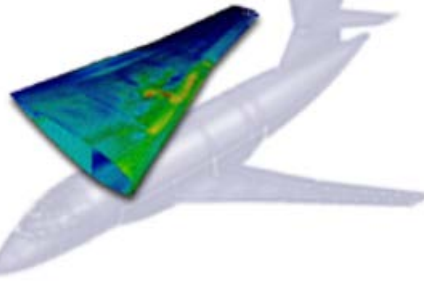
## Lehrbuch

Konstruieren mit Faser-Kunststoff-Verbunden

1. Ausbildung als zentraler Auftrag

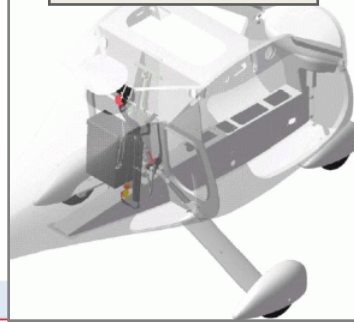


## Simulation

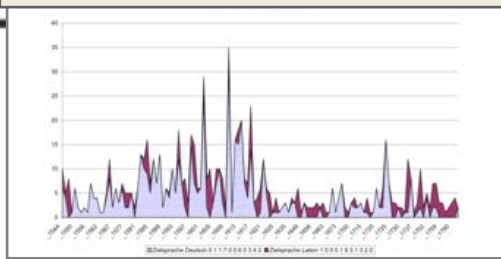


Suche

## CAD Modell



## Meßdaten Experiment



## Film / Experiment



Suchen Sie fachübergreifend in den Beständen der Technischen Informationsbibliothek der Deutschen Zentralbibliotheken für Medizin und Wirtschaftswissenschaften sowie

Flugzeug\* AND Verbundwerk\* AND Festigkeit\*

Suche starten

Suchmaske leeren

## Software / Quellcode

```
if ((list == user.PrimaryMailbox) && (1
return "";
var mboxlist = new String(list).split("
var row_bar = "<TR><TD COLSPAN=2 CLASS=
var row_space = "<TR CLASS=row_normal><
var text = "<TABLE CELLSPACING=0 CELLP
"<TD COLSPAN=2 CLASS=row_top><TABLE C
SPC(8,20) + "</TD><TD WIDTH=100% CLAS
caption + "</TD></TR></TABLE></TD></T
for (var i = 0; i < mboxlist.length; i+
var slot = Config.GetMailslot(mboxlis
```

## Forschungsberichte

Kooperation von CTIC-Kooperationspartnern unter Einbeziehung des Concurrent Engineering

DLR

**SCHLUSSBERICHT**

**Titel:**

HGF-Verbundprojekt: 99/19:

**„Schwarzer Rumpf“ - Realisierung von CFK-Rumpfkomponenten unter Einbeziehung des Concurrent Engineering**

Förderkennzeichen: 01SF9927/6

Laufzeit: 01.07.1999 – 30.06.2002

Gefördert durch Mittel des Bundesministeriums für Bildung und Forschung (BMBF)

**Projektleitung:**

Dipl.-Ing. Holger Wilmes (DLR, Braunschweig)

Prof. Dr.-Ing. Axel Herrmann (jetzt CTC Stade)

## Patente

re schränken Sie Ihren Suchbereich mit der Datenbankauswahl

Portal für Millionen

**Bibliographie**

**COMPOSITE MATERIAL FOR STRUCTURAL APPLICATIONS**

Page bookmark: W02011133351 (A1) - COMPOSITE MATERIAL FOR STRUCTURAL APPLICATIONS

Inventor(s): BOYLE MAUREEN [J], BLAIR DANA [GB], WU YE-JUN [J], HANO YEN-BEIE [J], FLORIANICO BRUCE [J], MACKENZIE PAUL [J]

Applicant(s): HECCEL CORP [J], HECCEL COMPOSITES LTD [GB], BOYLE MAUREEN [J], BLAIR DANA [J], WU YE-JUN [J], HANO YEN-BEIE [J], FLORIANICO BRUCE [J], MACKENZIE PAUL [J]

Classification: - International: B28B7/36, C08J5/26, C08L4/00 - European: C08J5/26

Application number: W02011133351/200208 20110412

Priority number(s): US20010764836 20130431

Also published as: US20011259514 (A1)

**Abstract of W02011133351 (A1)**

Composite materials that contain epoxy resin which is toughened and **reinforced** with thermoplastic materials and a blend of soluble particles. The uncured matrix resins include an epoxy resin component, a soluble thermoplastic component, a curing agent and an insoluble particulate component composed of elastic particles and rigid particles. The uncured resin matrix is combined with a fibrous reinforcement and cured to form composite materials that may be used for structural applications, such as primary structures in **aircraft**.

## Aufsätze aus Zeitschriften / Konferenzen

EXPERIMENTAL AND NUMERICAL STUDY OF THE OPEN-HOLE TENSILE STRENGTH OF CARBONEPOXY COMPOSITES

R. M. O'Higgins\*, G. S. Pathak\*, M. A. McCarthy\* and C. E. McCarthy\*\*

*Keywords: high strength composite, notched strength, failure, prediction, experiments*













Open hole tension tests are a part of the qualification process for composite parts that need to be joined to other parts in aircraft structures [1]. With each new material, a new set of tests is required. To reduce cost, it is desirable to develop analysis tools for the prediction of damage and failure in such tests, so that the amount of testing can be reduced and predictions can be made about material behaviour early in the design process. In this paper, an experimental and numerical study is presented on the notched open hole strength of high strength carbon/epoxy composite (HTAS78). Open hole tension tests have been performed on specimens with three different lay-ups—quasi-isotropic, zero-dominated, and cross-ply—in accordance with procedures as available standards. The data observed are being used to develop several methods for predicting the notched strength and results from one such method using a progressive damage analysis are presented with comparison with experiments. The predictions of specimen stiffness and failure load were found to agree well with experiments. To gain insight into the failure process, damage progression maps are shown.

1. Introduction

Mechanisms of fastener remains the primary means of joining composite components in modern aircraft structures, even though bolted joints are relatively inefficient. The stress concentration due to the hole causes a considerable reduction in both the tensile and compressive strength of a composite laminate. Experience with bolted joints in composite structures for aerospace applications has indicated a high level of complexity in joint design, due to the almost unlimited combinations of composite materials and fiber patterns and the fact that bolted joints in composites fail at loads that are not predicted by either

# GetInfo: Der Index

## Interne Datenbanken

- TIBKat 
- Katalog Medizin 
- Katalog Agrar 
- Katalog Wirtschaftswissenschaften 
- TIBscholar 
- Konferenzbeiträge 
- Zeitschriftenaufsätze 
- TEMA® 
- DKF 
- Forschungsdaten 
- IUD Bahn 
- AV-Medien 

## Über 46 Millionen Datensätze

- ca. 7,2 Mio Katalogdaten
- ca. 30,4 Mio Aufsatzdaten aus Fachdatenbanken
- ca. 8,5 Mio Verlagsdaten

Davon inhaltlich erschlossen: xx %



Formal erschlossen: **100%**

- Heterogene Metadaten
  - Format – Umfang - Inhalt
- Standardisierung in GetInfo-eigenes xml
- Indexierung
  - Suchmaschinen-Index – Drilldown
- Massen-Import- und –Updateroutinen

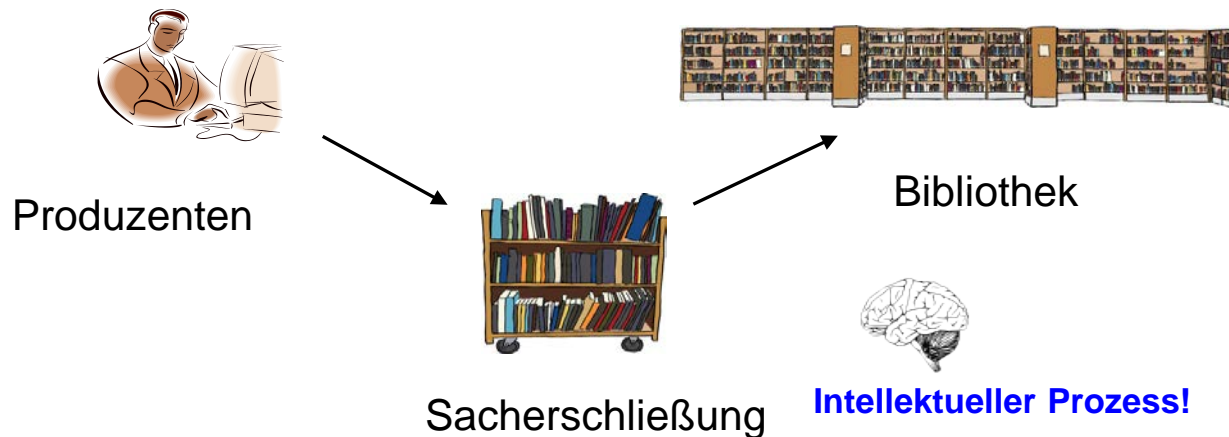
# GetInfo – die Ausgangs(daten)lage

- Zwischen **1,0 Mio. und > x Mio.** Metadatensätze pro Quelle
- Aktualisierungsrate: ca. 100.000 Metadatensätze pro Monat
  - mit minimalem Inhalt (Autor, Titel, Institution pro Dokument)
- **Beispiel:**
  - **title** = Untersuchung von Bodenbewegungen über Speicherkavernen: Schlussbericht zum Forschungsvorhaben
  - **creator** = Schauermann, Volker
  - **issued** = 1978
  - **publisher** = unknown
  - **description** = Umfang: 24 Bl
- Ca. 9 Mio. Datensätze sachlich klassifiziert = 20 %

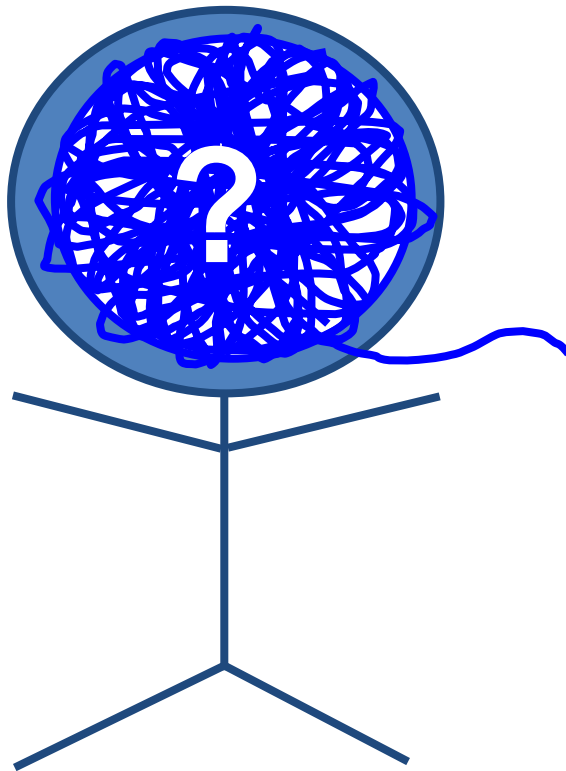


# Status quo

- Inhaltlicher Erschließung: intellektueller Prozess
- → hohe Qualität, hoher Ressourcenverbrauch



# Ich, der Suchende ...



benötige fachspezifisches Wissen

zu ...

für ...

damit ich ...

jetzt, morgen, übermorgen ...

Forschen, entscheiden,

mich weiter- und ausbilden kann.

# Was tun wir ...

vollständige  
standardisierte  
inhaltliche Erschließung  
aller Wissensobjekte

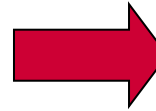


Fachspezifische Suche

The screenshot shows the TIB search interface. At the top, there are tabs for different subjects: **Alle Fächer**, **Architektur**, **Chemie**, **Informatik**, **Mathematik**, **Physik**, and **Technik**. Below the tabs, there is a search bar with the text "Suchen in > 135 Millionen Datensätzen" and a search input field containing "Suchbegriff eingeben (z.B. Autor, Titel, Thema, ...)". To the right of the input field is a search button labeled "Suchen". Below the search bar, there are several navigation links: "erweiterte Suche", "Datenbankauswahl", "Suchtipps", "Suchhistorie", and "Merkliste [0]".

# GetInfo: Fachsuche

- sachliche Erschließung der Metadatenätze ohne SE-Elemente
- Anreicherung der Metadaten
- Zuordnung **aller Metadatenätze** zu den TIB-Fächern



1. Schritt:  
Facheinstieg über  
Drilldown

Kurztitelanzeige	Treffer erschließen
Ihre Suche nach <b>hybrid</b> ergab <b>132.112</b> Treffer <span style="float: right;">Suche in Externe Quellen abgeschlossen</span>	
◀ ▶ 1 2 3 4 5 6 7 8 9 10 11 ▶	
1 Polymeric Composites for Hybrid Microelectronics (Score: 7.2778482) Krivka, I.; Kuzel, R.; Prokes, J. et al.   <b>BL CP</b>   1993 ▶ <a href="#">Zur Detailanzeige</a> ▶ <a href="#">Zur Merkliste hinzufügen</a> ▶ <a href="#">Zur Bestellung</a> ▶ <a href="#">Zeige Datensatz</a>	<b>Autor</b> Li, Y. (296) Wang, J. (294) Wang, Y. (293) ▶ <a href="#">mehr anzeigen</a>
2 Advanced Processing and Novel Applications of Hybrid Multilayer Green Tapes (Score: 7.2778482) Bansky, J.; Engemann, J.; Polzer, E. K. et al.   <b>BL CP</b>   1993 ▶ <a href="#">Zur Detailanzeige</a> ▶ <a href="#">Zur Merkliste hinzufügen</a> ▶ <a href="#">Zur Bestellung</a> ▶ <a href="#">Zeige Datensatz</a>	<b>Fach</b> Technik (92143) Physik (50760) Chemie (24835) Informatik (21815) Mathematik (12094) drilldown.classification.rest (11741) Architektur (9894) ▶ <a href="#">weniger anzeigen</a>
3 Study of the Thermomechanical Problems in Hybrid Assemblies satisfying the Automotive Electronics Conditions (Score: 7.2778482) Aucouturier, J. L.; Woirgard, E.; Zardini, C. et al.   <b>BL CP</b>   1993 ▶ <a href="#">Zur Detailanzeige</a> ▶ <a href="#">Zur Merkliste hinzufügen</a> ▶ <a href="#">Zur Bestellung</a> ▶ <a href="#">Zeige Datensatz</a>	<b>Classification Procedure</b> mapping (81711) lexicon (24814) automatic (21595) ▶ <a href="#">mehr anzeigen</a>
4 Hybrid Chemical and Physical Sensor Arrays (Score: 7.2778482) Atkinson, J. K.; International Society for Hybrid Microelectronics   <b>BL CP</b>   1993 ▶ <a href="#">Zur Detailanzeige</a> ▶ <a href="#">Zur Merkliste hinzufügen</a> ▶ <a href="#">Zur Bestellung</a> ▶ <a href="#">Zeige Datensatz</a>	<b>Dokumentformat</b> Print (121991) Elektronische Ressource (8636) CD-ROM / DVD (1081) ▶ <a href="#">mehr anzeigen</a>
5 Microtechnologies: International Realities and Outlooks. Research and Development in France (Score: 5.532518) Gognepain, J. J.; International Society for Hybrid Microelectronics   <b>BL CP</b>   1993 ▶ <a href="#">Zur Detailanzeige</a> ▶ <a href="#">Zur Merkliste hinzufügen</a> ▶ <a href="#">Zur Bestellung</a> ▶ <a href="#">Zeige Datensatz</a>	<b>Dokumenttyp</b> Aufsatz (Zeitschrift) (71061) Aufsatz (Konferenz) (54662) Ohne Angabe (2413) ▶ <a href="#">mehr anzeigen</a>
6 Laser Soldering of Today's Electronic Components (Score: 5.532518) Tullemans, A. M. J.; International Society for Hybrid Microelectronics   <b>BL</b>	<b>Erscheinungsjahr</b>

# LinSearch - Fachcluster



Architektur

Chemie



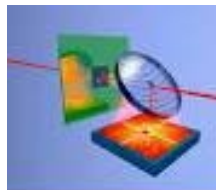
Informatik



Mathematik

$$\left( \begin{array}{l} \lim_{x \rightarrow \infty} \frac{\alpha \left(1 + \frac{1}{x}\right)}{x} \\ \int \frac{e}{x^2} dx \\ \cot\left(\frac{\pi x}{4}\right) \\ \frac{d}{dx} \left( \frac{\ln x}{\sqrt{\frac{x+1}{x-1}}} \right) \end{array} \right)$$

Physik



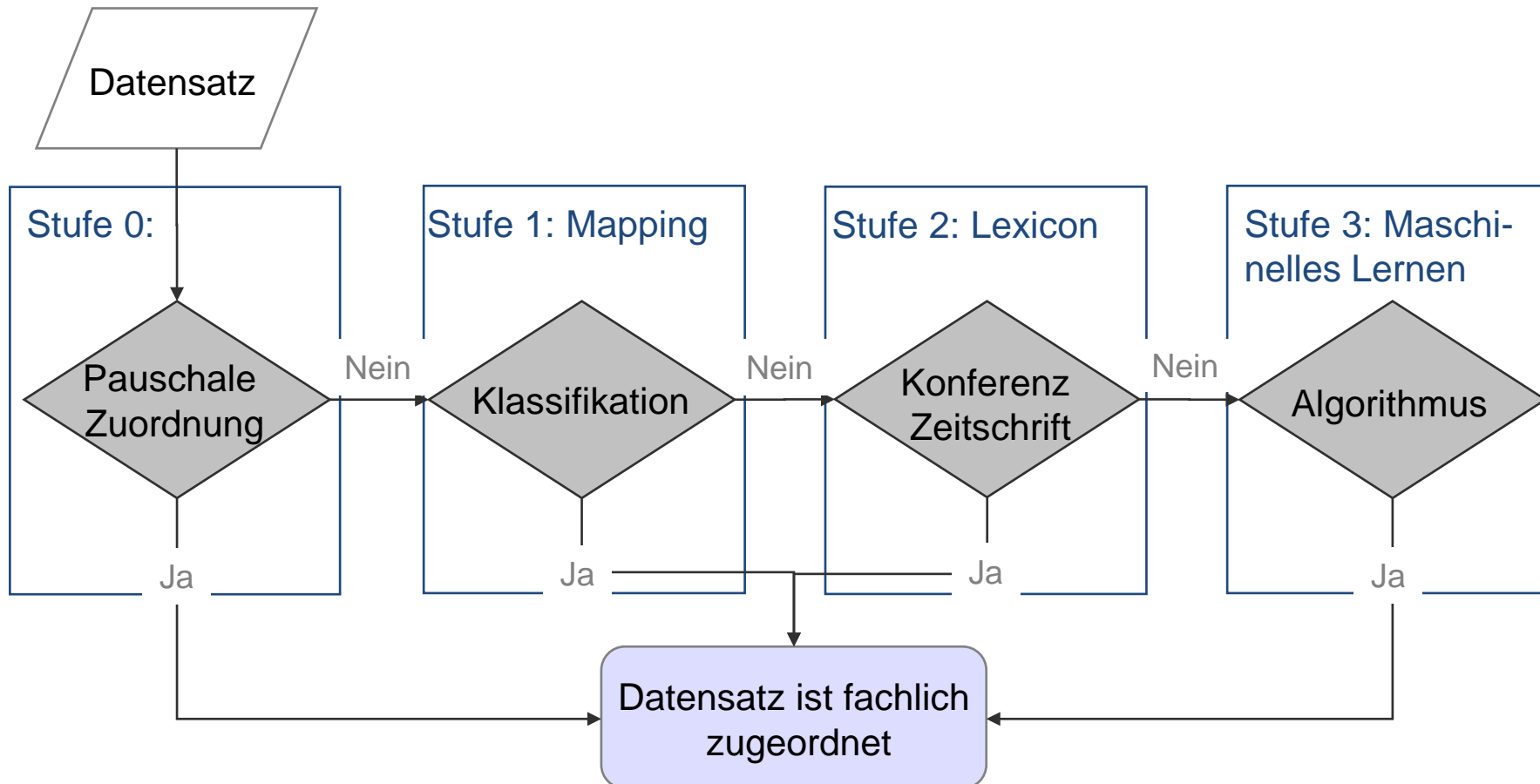
Technik



Weitere Fächer

# LINSearch – Das 4-Stufen-Verfahren

Kombination aus Mapping und automatischer Klassifizierung





## **Quellen werden pauschal zugeordnet**

### **Technik**

Springer-VDI, ASCE, VDE, IUD Bahn, DKF

### **Physik**

APS

### **Chemie**

Thieme (Römpp)

## LINSearch Stufe 1 – Mapping

- BK
- DDC
- Lokale Systematik
- SIGLE
- MSC
- WTI Fachordnung
- PACS
- RVK
- LCC

Vorhandene  
Klassifikations-  
elemente

→ Für alle sechs TIB Fächer + Weitere Fächer

### Titellisten

- Zuordnung von Zeitschriften- und Kongresstiteln zu den Fachgebieten der TIB
- Erstellung von Titellisten mit Formalparametern (ISSN / ISBN) für die technische Umsetzung
- Bereinigung von unsauberer Sachgruppenzuordnungen der ZDB

# LINSearch Stufe 3 – Algorithmus (averbis)

averbis

Verschlagwortungsprojekte



- SWD, GND



- MeSH



- Agrovoc, Umthes

BMW Group



- Labortechnik
  - Laborterminologie, Konzernterminologie
- Innovationsmanagement
  - Vokabular noch offen, Terminologierstellung



- Caritas-Terminologie
- SWD



- TEMA



- Die wesentlichen Inhalte sollen erkannt und strukturiert werden
  - Aus Freitexten werden **strukturierte Daten**
- Die Inhalte sollen schnell und bequem wiederauffindbar sein
  - Effektive und **semantische Suche**
- Relevante Inhalte sollen miteinander vernetzt sein
  - Inhalte benötigen ein gemeinsame **terminologische Basis**
- Die Inhalte sollen auswertbar sein
  - Statistische Auswertung und **Data-Mining**

## LINSearch Stufe 3 – Algorithmus (averbis)

- Alle Datensätze, die in den Stufen 0 bis 2 keiner Klasse zugeordnet werden konnten
- Deutsche und englische Titel
- Linguistische Vorverarbeitung
- Training mithilfe der Mappingtabelle (nur BK)
- Support Vector Machine
- Multilabel Klassifikation



# Algorithmus-Prozess

- Einlesen der Daten
- Linguistische Vorverarbeitung
- Merkmalsgewinnung
- Training
- Vorhersage

# Linguistische Vorverarbeitung

- Tokenisierung
  - z.B. → z.B.
  - polystyrol-related → polystyrol, related
- Stoppworterkennung
- “Das Leben ist schön!” → leben, schön
- Stammformreduktion
  - singer, singing, sing → sing
  - Modelle, Modellierung, Modell → modell
- (Segmentierung)
  - Farbbildanalyse → farb, bild, analyse

# Evaluation

- Zweck: Beurteilung der Güte des Klassifizierers im Anwendungsfall, also auf “neuen” Daten
- “neu”: nicht im Training enthalten
- Bestimmung Test/Trainingsdaten
- Zufällige Aufteilung  
Repräsentativität, Gleichverteilung, Unabhängigkeit
- 90% Training, 10% Test

# Fehlerarten → Wahrheitsmatrix

- Falsch-Negative (FN)
- Falsch-Positive (FP)
- Bei der Multi-Label-Klassifikation kann Modell viele Fehler machen
- Aus Fehlerzahlen (FN, FP) und korrekten Klassifikationen werden dann Performanzwerte bestimmt:
  - Accuracy
  - Precision, Recall, F-Score

# Evaluation

- Schritt 1: Automatische Evaluation
  - Ist-Soll Abgleich: Klassen in Testdaten sind zwar vorhanden, werden aber während der Klassifikation “versteckt”
- Schritt 2: Intellektuelle Prüfung
  - Nur Falsch-Positive überprüft
  - Falsche-Negative nicht geprüft, da Annahme, das BK Zuweisungen in Daten immer korrekt sind
- Hochrechnung der sich daraus ergebenden Korrekturen auf gesamte Dokumentenmenge

→ Typischerweise verbessertes Ergebnis

# Typische Fehler

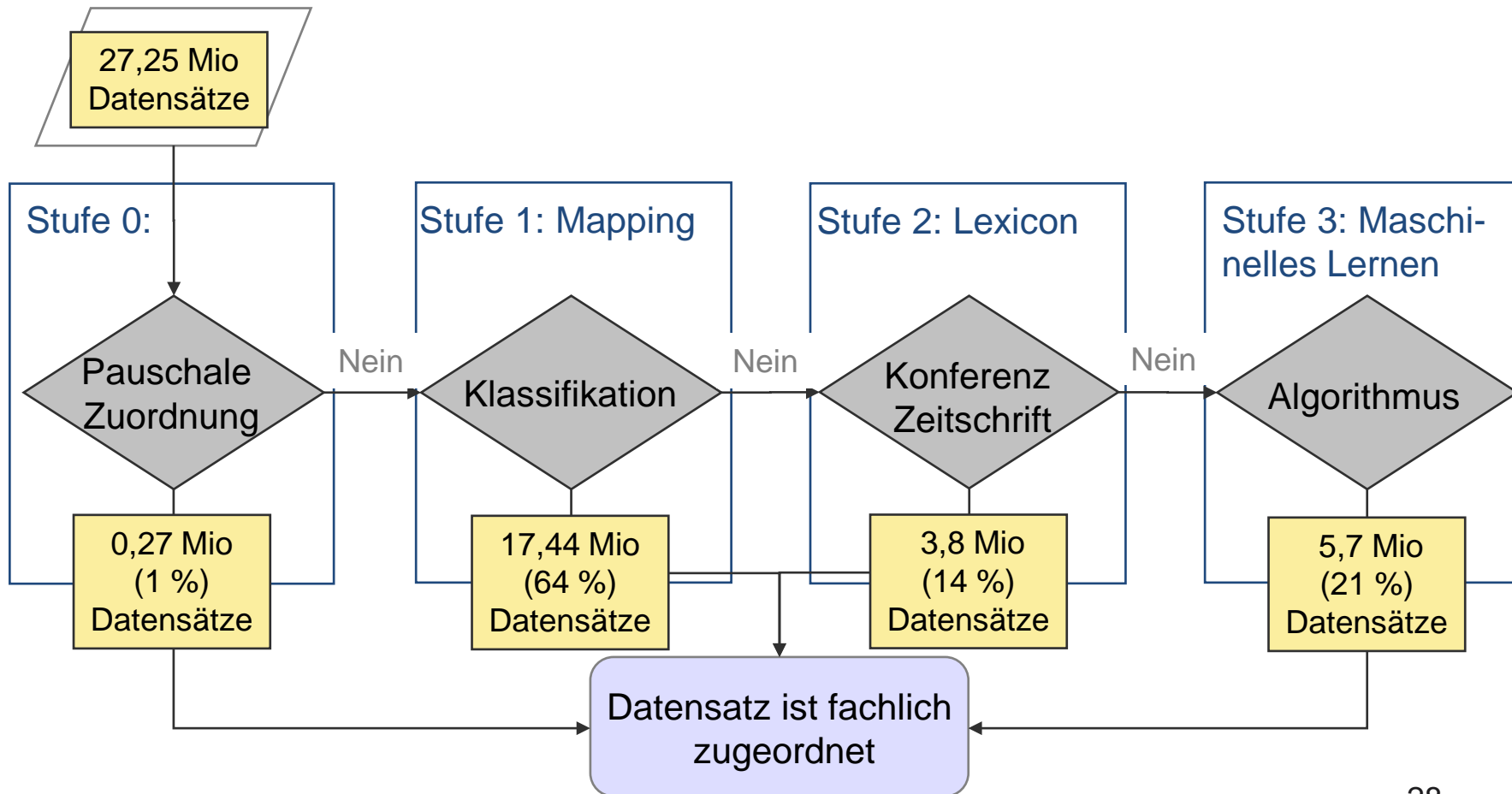
- Zu wenig Text vorhanden: v.a. bei Büchern meist nur sehr kurzer Titel
  - `<dc:title>Qualitative Analyse</dc:title>`
- Begriffliche Überlappungen (“Algebraische Spezifikation”)
- Originalklassifikation anhand Anwendung und nicht technischer Grundlage
  - `<dc:title>Diskrete Strukturen ; Bd. 1: Kombinatorik, Graphentheorie, Algebra</dc:title>`
- Falsche Dokumentensprache im Metadatensatz



# Ergebnis

- Gute Ergebnisse: ca. 87% F-score  
(Technik: 94 %) trotz geringem Informationsgehalt pro Datensatz
- Schnelle Verarbeitung, inkl. Spracherkennung  
 $\leq 5\text{ms}$
- Methodik: viel Statistik, letztendlich wenig Linguistik

# LINSearch – Das 4-Stufen-Verfahren: Ergebnis



# Chancen und Risiken

## Chancen

- Automatische Klassifizierungen können die intellektuelle Sacherschließung (noch) nicht ersetzen
- Unterstützung bei der Sacherschließung
- Gute Eignung für Bestände, die sachlich nicht erschlossen werden

## Risiken

- Mut zur Lücke - nicht alle Inhalte können verarbeitet werden.
- Fehlertoleranz - Inhalte werden falsch zugeordnet.
- Kontrollverlust - Bei der Datenmenge ist eine Analyse nur stichprobenartig möglich.

