



Hochwertige Bildsuche mittels empirisch fundierter semantischer Verfahren

Zur Erlangung des akademischen Grades eines
Doktors der Ingenieurwissenschaften

der Fakultät für Informatik
am Karlsruher Institut für Technologie (KIT)

genehmigte

Dissertation

von

Andreas Walter

aus Baden-Baden

Tag der mündlichen Prüfung: 28.10.2010

Erster Gutachter: Prof. em. Dr. Dr. h. c. Peter C. Lockemann

Zweiter Gutachter: Prof. Dr. Rudi Studer

Für Anulka

Prolog

Schöne Jahre des selbständigen, projektorientierten wie auch wissenschaftlichen Arbeitens liegen hinter mir. Umfangreiche Erfahrungen über den Aufbau, die Stärken und Schwächen von Bildsuchmaschinen konnte ich in fotomarktplatz.de sammeln. Weiterhin eine Menge Erfahrung, wie man Systeme praxistauglich erstellen muss. Zusätzlich die Erfahrung, aktuelle Forschungsergebnisse praxistauglich zu integrieren im EU Projekt IMAGINATION. Und schließlich konnte ich lernen, angeleitet zuerst durch Prof. Klemens Böhm am IPD am KIT Karlsruhe und anschließend durch meinen Mentor und über die Jahre auch Freund Gabor Nagypal am FZI, wie man wissenschaftlich arbeitet und publiziert. Alle drei Faktoren führten zur Idee dieses Buches. Somit möchte ich mich an dieser Stelle speziell bei Gabor Nagypal für seine Unterstützung bedanken. Danken möchte ich auch allen Mitarbeitern am FZI, speziell Raphael Volz, für das schöne und angenehme Arbeitsumfeld. Besonderer Dank an meinen Doktorvater Professor Peter Lockemann, dass ich bei ihm trotz seiner Emeritierung noch promovieren durfte. Allen an diesem Buch beteiligten Personen sei hiermit herzlich gedankt.

Das Schönste der letzten Jahre war, dass ich meine Frau Anna kennen gelernt habe, wir uns ineinander verliebt haben und wir schließlich im August 2008 geheiratet haben. *Kochanie Anulka, ich danke Dir für Deine Geduld während der Erstellung dieses Buches und widme es Dir. Ich liebe Dich.*

Inhaltsverzeichnis

1	Einleitung	21
1.1	Motivation	21
1.2	Herausforderung und resultierende Fragestellungen	23
1.3	Ziele und Lösungsschritte	25
1.4	Überblick über die Arbeit	27
2	Suchen von Bildmotiven	29
2.1	Bildsuchmaschinen bei Bildagenturen	29
2.2	Erstellung von Suchergebnissen durch Bildarchivare	30
2.2.1	Ablauf der Bildsuche	30
2.2.2	Definition: qualitativ hochwertige Suchergebnisse	32
2.3	Bildsuche mittels Bildsuchmaschinen	32
2.3.1	Ablauf	32
2.3.2	Bemessung des Gesamtaufwands	34
2.4	Szenario: Bildsuche für ein Magazin	34
2.4.1	Ablauf der Bildsuche	34
2.4.2	Diskussion	35
2.4.3	Begriffsdefinition: Arten von Suchanfragen	36
2.4.4	Ermittlung der Schwachpunkte von Bildsuchmaschinen	36
2.5	Semantische Ansätze zur Verbesserung der Bildsuche	37
2.5.1	Aspekte von Semantik im Kontext der Bildsuche	37
2.5.2	Mögliches Suchergebnis einer „intelligenten Bildsuchmaschine“	38

2.6	Semantische Bildsuche in einem Filmarchiv	39
2.7	Zusammenfassung	40
3	Aufbau der Bildsuchmaschine ImageNotion	41
3.1	Textbasierte Bildsuchmaschinen	42
3.1.1	Die Notwendigkeit zur Erstellung von Metadaten	42
3.1.2	Architektur und Komponenten	42
3.1.3	Lexikalische Probleme in der Bildsuche	44
3.2	Ontologien innerhalb der Bildsuche	45
3.2.1	Definition Ontologie	45
3.2.2	Arten von Ontologien	46
3.3	Mögliche Orte zur Integration semantischer Verfahren	49
3.3.1	Anfrageerweiterung	50
3.3.2	Semantische Metadaten	53
3.3.3	Beschriftung von Bildteilen	56
3.3.4	Analyse 1: Integration der semantischen Verfahren bei Bild- agenturen	56
3.3.5	Entscheidung zum Einsatzort semantischer Verfahren	58
3.4	Architektur und Komponenten von ImageNotion	59
3.4.1	Benutzerschnittstelle	59
3.4.2	Anwendungsschicht	64
3.4.3	Datenhaltung	67
3.5	Verwandte Arbeiten zu ImageNotion	67
3.5.1	System von Vallet (2005)	67
3.5.2	MIA Demonstrator (2006)	68
3.5.3	SemSpace (2007)	68
3.5.4	ALIPR (2007)	68
3.5.5	Gegenüberstellung der Arbeiten mit ImageNotion (Konzeption)	69
3.6	Zusammenfassung	69
4	Kategorisierung von Suchanfragen	73

<i>INHALTSVERZEICHNIS</i>	9
4.1 Motivation	73
4.2 Fragestellungen an ein geeignetes Modell	74
4.3 Kategorisierung von Abfolgen einer Suchanfrage in Sitzungen	75
4.4 Bestehende Modelle zur Beschreibung von Bildinhalten	76
4.4.1 Das Modell von Shatford	77
4.4.2 Das 10-Ebenenmodell von Jaimes und Chang	77
4.5 Kategorisierung der Art und dem Bezug von Suchanfragen	78
4.5.1 Überblick über die durchgeführten Anpassungen	78
4.5.2 Kategorien des Modells	79
4.6 Zusammenfassung	81
5 Analyse von Suchanfragen an fotomarktplatz.de	83
5.1 Ergebnisse bisheriger Untersuchungen	83
5.1.1 Anfragen an die Bildsuchmaschine einer Bildagentur	84
5.1.2 Anfragen durch Bildredakteure in einem Zeitungsverlag	84
5.1.3 Gründe für die eigene Untersuchung	85
5.2 Die Metasuchmaschine fotomarktplatz.de	86
5.2.1 Bildsuche mit fotomarktplatz.de	87
5.2.2 Benutzergruppen in fotomarktplatz.de	87
5.3 Analyse 2: Bildsuche in fotomarktplatz.de	87
5.4 Auswertung der Ergebnisse	88
5.4.1 Umfang und genereller Ablauf einzelner Sitzungen	88
5.4.2 Sind lexikalische Anfrageerweiterungen notwendig?	90
5.4.3 Sind Suchanfragen miteinander verwandt?	91
5.4.4 Sind semantische Anfrageerweiterungen notwendig?	91
5.4.5 Themen der Suchanfragen	93
5.4.6 Schlussfolgerung aus den Ergebnissen	94
5.5 Zusammenfassung	95
6 Onlineevaluation zur semantischen Bildsuche	97

6.1	Untersuchungen zum Verständnis von Benutzern für semantische Ansätze	98
6.2	Verfahren für die Evaluation der semantischen Bildsuche	99
6.2.1	Eingabe einer semantischen Suchanfrage	99
6.2.2	Browsen der Ontologie	101
6.2.3	Verfahren zur Neuformulierung einer gegebenen Suchanfrage .	101
6.2.4	Weitere Verfahren	103
6.2.5	Navigation durch Bildteile	103
6.3	Empirische Studie 1: Onlineevaluation der semantischen Bildsuche . .	104
6.4	Auswertung der Ergebnisse	105
6.4.1	Erfahrung der Benutzer mit semantischen Anwendungen . . .	106
6.4.2	Verständnis für die semantische Bildsuche	106
6.4.3	Neuformulierung von Suchanfragen	110
6.4.4	Verständnis für Relationen	111
6.4.5	Navigation durch Bildteile	113
6.5	Resultierender Aufbau der Komponente Bildsuche in ImageNotion . .	114
6.6	Zusammenfassung	116
7	Erstellung semantischer Metadaten	119
7.1	Methodologien zur Ontologierstellung	119
7.1.1	Begriff: Methodologie	120
7.1.2	Oft verwendete Methodologien	120
7.1.3	Standardisierte Ontologien	121
7.1.4	Fazit	121
7.2	Empirische Studie 2: Erstellbarkeit semantischer Metadaten mit vorgegeben Ontologien	121
7.3	Aufbau und Durchführung des Workshops	122
7.4	Fragen und Beobachtungen während des Workshops	124
7.5	Diskussion mit den Teilnehmern	125
7.6	Resultierende Anforderungen	128

<i>INHALTSVERZEICHNIS</i>	11
7.7 Verwandte Untersuchungen	129
7.7.1 SOBOLEO: Taxonomien für Bookmarks	129
7.7.2 Konzeptionelle Betrachtung von Hepp	129
7.7.3 Schlussfolgerungen	130
7.8 Zusammenfassung	130
8 Ablauf der Bildbeschriftung bei Bildagenturen	133
8.1 Erstellung einer textbasierten Bildbeschriftung	133
8.1.1 Betrachtung des entstehenden Aufwands	133
8.1.2 Beispiel: Beschriftung eines Bilds in einem Filmarchiv	134
8.1.3 Iterativer Ablauf zur Erstellung von Bildbeschriftungen	135
8.2 Empirische Studie 3: Zeitaufwand zur Erstellung textbasierter Bildbeschriftungen	136
8.3 Einsparpotential durch semantische Verfahren	137
8.4 Arbeitsablauf zur Erstellung semantischer Metadaten	138
8.5 Zusammenfassung	139
9 Ontologieentwicklung als Reifungsprozess	141
9.1 Vorüberlegung zur benutzergetriebenen Erstellung von Ontologien	141
9.2 Verwandte Arbeiten	142
9.2.1 Kollaborative Methoden und Anwendungen zur Ontologieentwicklung	143
9.2.2 Semantic Media Wiki	144
9.3 Das Ontology Maturing Process Model	145
9.3.1 Reifung von Wissen	145
9.3.2 Resultierende Phasen zur Ontologieerstellung	146
9.3.3 Anwendung des Modells	147
9.4 Zusammenfassung	148
10 Semantische Bildbeschriftung mit ImageNotion	149
10.1 Die ImageNotion Methodologie	149

10.1.1	Phasen des Modells	150
10.1.2	Resultierende arbeitsintegrierte Ontologieerstellung	151
10.2	Aufbau der Komponente „Verwaltung Wissensbasis“	152
10.2.1	Der ImageNotion Browser	152
10.2.2	Der ImageNotion Editor	153
10.2.3	Beispiel aus der Implementierung	154
10.3	Empirische Studie 4: Anwendung der ImageNotion Methode	159
10.3.1	Aufbau und Durchführung des Workshops	159
10.3.2	Beobachtungen während des Workshops	159
10.3.3	Betrachtung der erstellten Ontologien und Bildbeschriftungen	160
10.3.4	Diskussion mit den Teilnehmern	161
10.4	Fazit	162
10.5	Zusammenfassung	163
11	Zeitaufwand für die semantische Bildbeschriftung	165
11.1	Mögliche Erstellung von semantischen Metadaten	165
11.1.1	Beschriftung mit einer Taxonomie	166
11.1.2	Beschriftung mit einem Instanzenmodell	166
11.1.3	Beschriftung mit Text	167
11.2	Empirische Studie 5: Zeitdauer zur Beschriftung eines Filmarchivs	168
11.2.1	Dauer einzelner Arbeitsschritte	169
11.2.2	Zeitdauer der Nutzung eines Instanzenmodells	169
11.2.3	Zeitdauer zur Nutzung einer Taxonomie	171
11.2.4	Zeitdauer zur Erstellung textbasierter Beschriftungen	172
11.2.5	Gegenüberstellung der Messungen	173
11.2.6	Fazit	174
11.3	Zusammenfassung	174
12	Automatische Erkennung von Bildinhalten	177
12.1	Erkennung von Objekten auf Bildteilen	178

12.2 Erstellung von Trainingsdaten 179

 12.2.1 Universelle Objekterkenner 179

 12.2.2 Spezialisierte Objekterkenner 180

 12.2.3 Fazit 180

12.3 Identifikation von Objekten 180

 12.3.1 Zuordnung von Objekten 180

 12.3.2 Erkennung von Merkmalen 181

12.4 Kategorisierung der auftretenden Fehler 182

12.5 Laboruntersuchung 1: Objekterkenner im praktischen Einsatz 184

 12.5.1 Motivation 184

 12.5.2 Vergleichsbasis: manuell erstellte Bildbeschriftungen 185

 12.5.3 Auswahl geeigneter Objekterkenner 185

12.6 Auswahl geeigneter Textinterpretierer 186

 12.6.1 Evaluation 187

 12.6.2 Bewertung und Überlegungen zu den Ergebnissen 191

 12.6.3 Zusammenfassung 192

13 Kombination automatischer Verfahren 193

13.1 Grundidee im EU Projekt IMAGINATION 193

13.2 Ziele und Anforderungen 194

 13.2.1 Ziele der Kombination automatischer Verfahren 195

 13.2.2 Anforderungen 195

13.3 Verwandte Arbeiten 196

 13.3.1 Semantische Bibliotheken 196

 13.3.2 Kombination von Textinterpretierer und Objekterkenner 196

 13.3.3 Kombination von Gesicht- und Personenerkenner 197

13.4 Hierarchisierung von Objekterkennern 197

 13.4.1 Aufbau eines geeigneten Instanzenmodells 198

 13.4.2 Beispiel eines entstehenden Instanzenmodells 200

13.5 Kombination der Ergebnisse 200

13.5.1	Genereller Ablauf der Kombination von Ergebnissen	200
13.5.2	Kombination gleichartiger Ergebnisse	201
13.5.3	Kombination hierarchischer Ergebnisse	203
13.5.4	Kombination mit einem Textinterpretierer	205
13.6	Aufbau der Komponente „Kombination automatischer Prozesse“ . . .	207
13.6.1	Wrapper	207
13.6.2	Aufbau der Steuerung	208
13.7	Laboruntersuchung 2: Messung der kombinierten Ergebnisse von Ob- jekterkennern	208
13.7.1	Durchführung der Evaluation	208
13.7.2	Ergebnisse der Evaluation	209
13.7.3	Fazit	212
13.8	Zusammenfassung	214
14	Hochwertige Suchergebnisse mit ImageNotion	215
14.1	Bildsuche in einem Filmarchiv	215
14.2	Textbasierte Bildsuche „Thurman“	216
14.3	Semantische Suche „Uma Thurman“	217
14.4	Semantische Suche „Quentin Tarantino“	217
14.5	Interpretation der Suchanfrage „Weibliche Schauspieler“	218
14.6	Strukturierte Verfeinerung der Suchanfrage „Schauspieler“	220
14.7	Navigation durch das Bildarchiv	221
14.8	Zusammenfassung	221
15	Zusammenfassung und Ausblick	225
15.1	Zusammenfassung	225
15.2	Ausblick	230
15.2.1	Einsatz von ImageNotion	230
15.2.2	Offene wissenschaftliche Fragestellungen	231
A	Datenbasen der Untersuchungen	235

Abbildungsverzeichnis

2.1	Ablauf einer Sitzung zur Bildsuche	33
2.2	Korrekt interpretiertes Ergebnis der Anfrage <i>Lachende Frau neben einem Oldtimer von Mercedes Benz stehend</i>	38
3.1	Architektur und Komponenten einer textbasierten Bildsuchmaschine .	43
3.2	Klassifizierung von Ontologien	46
3.3	Ein Thesaurus für Synonyme und unterschiedliche Sprachen einzelner Wörter	47
3.4	Ausschnitt aus einem Instanzenmodell für ein Filmarchiv	48
3.5	Semantische Beschriftung einer Szene aus dem Film <i>Cleopatra</i>	54
3.6	Architektur und Komponenten der Bildsuche ImageNotion	60
4.1	Angepasstes Modell zur Kategorisierung von Suchanfragen	78
5.1	Die Metasuchmaschine fotomarktplatz.de	85
5.2	Verteilung der Benutzer in fotomarktplatz.de nach Bildnutzung	86
5.3	Kategorisierung der Abfolgen von Suchanfragen einer Sitzung	90
5.4	Kategorisierung der Suchanfragen nach Art und dem Bezug der Anfrage	91
5.5	Hauptthemen innerhalb von 1000 Suchanfragen	93
6.1	Vorschlag semantischer Elemente für die Texteingabe „Petain“	100
6.2	Zusammenfassung von semantischen Elemente für die Anfrage <i>Philipppe Petain</i>	102
6.3	Navigation durch Bildteile, Anfrage <i>Angela Merkel</i>	103
6.4	Persönliche Erfahrung der Benutzer mit semantischen Anwendungen .	106

6.5	Top 10 für zu verwendende semantische Elemente	112
6.6	Top 10 für benannte Relationen	112
6.7	Gewünschte Formulierung einer Suchanfrage auf einem Bildteil	114
7.1	Semantische Bildbeschriftung mit PhotoStuff	122
8.1	Beschriftung eines Bilds aus einer Filmszene	134
8.2	Iterativer Ablauf zur Erstellung von Bildbeschriftungen	135
8.3	Arbeitsablauf zur Erstellung semantischer Bildbeschriftungen	138
9.1	Phasen der Reifung von Wissen	145
9.2	Vier Phasen des Ontologie Maturing Process Model	146
10.1	Phasen der ImageNotion Methodologie	150
10.2	Beschriftung des Bilds „Uma Thurman neben einem Jaguar“	155
10.3	Suche nach semantischen Elementen mit dem ImageNotion Browser	156
10.4	Erstellung des semantischen Elements „Auto“	156
10.5	Einordnung von „Auto“ in eine Taxonomie	157
10.6	Erstellung der Beschriftung „Jaguar E-Type“ für einen Bildteil	158
10.7	Resultierende semantische Metadaten für das Beispiel	158
11.1	Bildbeschriftung in einem Filmarchiv mittels einer Taxonomie	166
11.2	Bildbeschriftung in Filmarchiv mit einem Instanzenmodell	167
11.3	Bildbeschriftung in Filmarchiv mittels gleichwertigem Text	167
11.4	Vergleich der Zeitdauer zur Bildbeschriftung mittels Text, Taxonomie und Instanzenmodell	173
12.1	Ablauf der Extraktion und Erkennung von Objekten	178
12.2	Erkennung der Emotionen <i>glücklich, überrascht, traurig</i> und <i>wütend</i>	181
12.3	Zuordnung des Geschlechts in Bildteilen, die Personen zeigen	182
12.4	Fehlerarten <i>Bildteil nicht erkannt, Fantomobjekte, falsch erkanntes Objekt, falsch identifiziertes Objekt</i>	183
13.1	Grundidee der Kombination automatischer Verfahren	194

13.2 Ein Instanzenmodell für Objekterkenner	199
13.3 Kombination gleichartiger Ergebnisse	202
13.4 Kombination hierarchischer Ergebnisse	204
13.5 Kombination von Textinterpretierer mit Gesichterkenner und Gesicht- tidentifikation	206
13.6 Aufbau der Komponente „Kombination automatischer Prozess“ in ImageNotion	207
13.7 Vergleich richtig und falsch erkannter Bildteile und der resultierenden Ersparnis an Arbeitsschritten	213
14.1 Ergebnis der textbasierten Bildsuche „Thurman“	216
14.2 Ergebnis der semantischen Bildsuche <i>Uma Thurman</i>	217
14.3 Ergebnis der semantischen Bildsuche <i>Quentin Tarrantino</i>	218
14.4 Ergebnis der semantischen Bildsuche <i>Weibliche Schauspieler</i>	219
14.5 Hintergrundwissen des semantischen Element <i>Uma Thurman</i>	219
14.6 Vorschläge zur strukturierten Verfeinerung der Suchanfrage <i>Schau- spieler</i>	220
14.7 Navigation durch das Bildarchiv mit Bildteil <i>John Travolta</i>	221
14.8 Suchergebnis nach Klick auf Bildteil <i>John Travolta</i>	222
15.1 Innenansicht des ImageNotion Broschüre 2010	231

Tabellenverzeichnis

3.1	Vergleich verwandter Arbeiten mit ImageNotion	70
3.2	Übersicht der resultierenden Fragestellungen in ImageNotion	71
4.1	Kategorisierung der Abfolgen von Suchanfragen in Sitzungen	76
4.2	Kategorien für die Art und dem Bezug von Suchanfragen	80
5.1	Generelle Auswertung der Suchanfragen und Sitzungen in fotomarkt- platz.de	89
6.1	Ergebnisse für den Aufgabenbereich „Verständnis für die semantische Bildsuche“	107
6.2	Ergebnisse für den Aufgabenbereich „Neuformulierung von Suchan- fragen“	111
6.3	Ergebnisse für den Aufgabenbereich „Neuformulierung von Suchan- fragen“	113
6.4	Umsetzung der Komponente Bildsuche in ImageNotion	115
7.1	Erfahrungen der Benutzer mit Verfahren zur Bildbeschriftung	123
8.1	Ergebnis der Messung des Zeitaufwands zur Beschriftung angezeigter Personen	136
10.1	Schnittstelle des ImageNotion Browsers	152
10.2	Schnittstelle des ImageNotion Editors	153
11.1	Messung der Zeitdauer einzelner Arbeitsschritte zur Bildbeschriftung	169
11.2	Zeitdauer zur Beschriftung eines Bildarchivs mit einem Instanzenmodell	170

11.3 Abgeleitete Zeitdauer zur Beschriftung eines Bildarchivs mit einer Taxonomie	171
11.4 Abgeleitete Zeitdauer zur Beschriftung eines Bildarchivs mit textbasierten Metadaten	172
12.1 Übersicht über die Inhalte im Filmarchiv	185
12.2 Arbeitsschritte für Erstellung und Korrektur von Bildteilen	188
12.3 Übersicht über die Ergebnisse der Objekterkennung	189
12.4 Erkannte semantische Elemente durch Textinterpretierer	190
13.1 Übersicht über die Ergebnisse der Kombinationsverfahren (Basis: Tabelle 12.2)	210

Kapitel 1

Einleitung

1.1 Motivation

Der Bildagentur¹ ActionPress aus Hamburg wurden alleine am Abend der Oscar Verleihung 2010 von Fotografen mehr als 50.000 digitale Bilder der Preisträger und Teilnehmer geschickt. Somit kommen täglich bei Bildagenturen, sowie in Bildarchiven von Internetanbietern, eine Vielzahl an neuen Bildern hinzu und müssen dort verwaltet werden². Dabei stellt sich die Frage: *Wie findet man in einem Bildarchiv möglichst schnell und präzise die gewünschten Bilder zu einem bestimmten Thema?*

Schon im privaten Umfeld werden oft Bilder von Freunden oder Familienmitgliedern gesucht, beispielsweise zur Erstellung eines Kalenders mit Bildern der Enkelkinder als Geschenk für die Oma. Meist wird hierbei der gesamte Bildbestand auf dem Computer gesichtet. Dieses *Durchstöbern* des ganzen Bildarchivs ist in einem kommerziellen Umfeld unmöglich, da keine präzisen Suchergebnisse [FTN⁺00] erstellt werden! In dieser Arbeit soll daher der Schwerpunkt auf der Suche nach kommerziell verwertbaren Bildern bei Bildagenturen liegen. Dort haben die Akteure präzise Anforderungen. Sie haben eine genaue Vorstellung, was für Bilder sie in einem Bildarchiv finden möchten [MS00]. Ein häufiges Szenario für Suchanfragen durch kommerzielle Nutzer sind Angestellte in Verlagen. Diese benötigen Bildmotive, um damit Artikel für eine Tageszeitung oder ein Wochenmagazin zu illustrieren. Dabei möchten diese Nutzer *schnell die gewünschten Bildern finden*, denn lange Sitzungen zur wiederholten Formulierung von verschiedenen Suchanfragen zu einem

¹Eine Bildagentur verwaltet die Lizenzen kommerzieller Bilder. Bildnutzer wie Verlage oder Firmen erwerben diese Lizenzen beispielsweise zum Abdruck der Bilder in Magazinen oder Broschüren.

²So beträgt der Bildbestand von großen Bildagenturen wie ActionPress, Bildmaschine, DPA oder Superbild mehrere Millionen Bilder pro Bildarchiv.

gewünschten Bildinhalt sind zeitaufwändig und somit für diese Verlage teuer.

Zur Bildsuche werden Suchmaschinen für Bilder verwendet³. Diese Bildsuchmaschinen sollen das Suchen nach Bildern in gleicher Art und Weise wie die Suche nach Webseiten⁴ zu einem Thema ermöglichen. In einer Suchmaske werden passend zum gewünschten Bildmaterial Stichwörter eingegeben, anschließend wird ein Suchergebnis erzeugt [FTN⁺00]. Auf den ersten Blick sollte hierbei die Erstellung einer Bildsuche kein größeres Problem als die Suche nach textbasierten Webseiten darstellen. Doch dem ist für Bilder leider nicht so, denn *Computer erkennen von selbst nicht die Inhalte und Zusammenhänge auf Bildern* [GW01]!

Somit geht dem schnellen und präzisen Finden von Bildern eine weitere Frage voraus, nämlich wie man dann überhaupt Bilder finden kann. Hierzu müssen erst *Metadaten* für die Bilder erstellt werden um zu beschreiben, was auf dem Bild zu sehen ist [FTN⁺00]. Diese Metadaten müssen anschließend indiziert werden, so dass eine Bildsuche überhaupt möglich wird. Zwei generelle Ansätze zur Erstellung von Metadaten für Bilder sind möglich. Dies sind die *manuelle* Bildbeschriftung durch Bildredakteure [MS00] sowie die *automatische* Erstellung von Beschriftungen mittels speziellen und trainierbaren Objekterkennungsalgorithmen [AHAQI⁺07].

Die Erstellung von manuellen Bildbeschriftungen ist *zeitaufwändig* und *teuer*, denn ein Bildredakteur⁵ muss raten, welche Begriffe in Suchanfragen verwendet werden könnten. Dadurch fehlen bei der manuellen Beschriftung oft relevante Suchbegriffe, da diese nicht in die Metadaten hinzugefügt wurden [MS00]. Automatische Verfahren erkennen entweder nicht alle wichtigen Objekte wie Gesichter, Personen oder Gegenständen auf einem Bild oder markieren Stellen, die keine Objekte anzeigen. Die Erstellung von Metadaten für Bilder ist somit nicht nur zeitaufwändig und teuer sondern weiterhin sind die Metadaten oft *unvollständig* und *fehlerhaft*.

Unvollständige und fehlerhafte Bildbeschriftungen haben daher starke Auswirkungen auf die Qualität von Suchergebnissen. Ein deutlicher Beleg hierfür ist, dass bei der Bildsuche deutlich mehr Seiten eines Suchergebnisses gesichtet werden, als zum Beispiel bei der Suche nach Webseiten [FTN⁺00]. Ein Bild, welches den „schiefen Turm von Pisa“ zeigt und mit den Stichwörtern „Pisa, Turm, schief“ als Metadaten beschriftet wurde, wird bei der Suche „Turm Pisa“ korrekt angezeigt. Dieses Bild wird allerdings in den bei Bildagenturen überwiegend rein textbasierten Bildsuchmaschinen bei Suchanfragen wie „historische Bauwerke Toskana“ nicht gezeigt. Dies hat zur Konsequenz, dass Bildsucher allgemeine Anfragen (im vorigen Beispiel „Toskana“) formulieren und sich manuell eine Vielzahl von Bildern ansehen

³z.B. die Bildsuchmaschine von ActionPress ist erreichbar unter www.actionpress.com

⁴z.B. bei Google, www.google.com

⁵In dieser Arbeit meint „Bildredakteur“ ein Angestellter in einer Bildagentur, der auf die Erstellung von Bildbeschriftungen spezialisiert ist und diese erstellt

müssen, um selbst die geeigneten Bilder in einem Bildarchiv zu finden [MS00].

Die Suche nach geeigneten Bildern gleicht daher oft einem „Stochern im Nebel“ und ist so für einen Bildsuchenden nicht nur sehr zeitaufwändig und somit teuer sondern oft auch sehr frustrierend [MS00].

1.2 Herausforderung und resultierende Fragestellungen

Die Verbesserung der Suchergebnisse in Bildsuchmaschinen ist daher die Herausforderung und Mission dieser Arbeit. Es sollen Verfahren und Lösungen zur *Erzielung hochwertiger Suchergebnisse* im Umfeld von Bildarchiven bei Bildagenturen gefunden werden. Ein *hochwertiges Suchergebnis* soll in Anlehnung an das gängige Verfahren zur Messung der Qualität eines Suchergebnisses (precision und recall [BYRN99]) ein möglichst *präzises* und *vollständiges* Suchergebnis sein. Ein *präzises Suchergebnis* soll die geeignetsten Bilder zu einer Anfrage enthalten und die darin wiederum geeigneten Bilder sollten möglichst an den ersten Stellen des Suchergebnisses stehen. Ein *vollständiges Suchergebnis* soll möglichst alle Bilder enthalten, die zur Suchanfrage passen. Die weitere Herausforderung ist dabei, dass entsprechende *Verfahren praktisch einsetzbar sind* innerhalb eines Bildarchivs bei einer Bildagentur. Daher müssen die zu schaffenden Lösungen verständlich, anwendbar und umsetzbar sein.

Der Hauptansatz zur Schaffung von präzisen Suchergebnissen ist der *Einsatz semantischer Verfahren*. Allgemein meint Semantik die „Bedeutung von Dingen einer bestimmten Art“ [Fre92]. Im Kontext der Bildsuche gibt es dadurch drei unterschiedliche Aspekte von Semantik [JJfC00], diese sind die *Bedeutung einer gestellten Suchanfrage*, die *Bedeutung von Bildbeschriftungen* sowie die *Bedeutung von abgebildeten Inhalten* auf den Bildern selbst.

Bei der *Bedeutung der Suchanfrage* soll eine Bildsuchmaschine die Suchanfrage korrekt verstehen und umsetzen. Mit Hilfe von vorhandenem Hintergrundwissen über die eingegebenen Begriffe kann so zum Beispiel die allgemeine Eingabe „historische Bauwerke Toskana“ umgesetzt werden in detailliertere Eingabeteile wie „Schiefer Turm von Pisa“ oder „Kathedrale Florenz“. Nagypal [Nag07] hat in seiner Dissertation unter anderem gezeigt, dass durch diese *Anfrageerweiterung* von Suchanfragen mit semantischen Elementen die Präzision der entstehenden Suchergebnisse deutlich verbessert werden kann.

Mit der *Bedeutung von Bildbeschriftungen* werden die Metadaten selbst und deren Bedeutung betrachtet. Statt der textbasierten Erstellung von Metadaten werden hierbei einzelne *semantische Elemente* erstellt, die umfangreich mit Hintergrundwis-

sen versehen werden können und zur Bildbeschriftung verwendet werden. Für den „Schiefen Turm von Pisa“ kann man so hinterlegen, dass dieser sich unter anderem in „Pisa“ (einer „Stadt“ in „Italien“) befindet sowie ein „Bauwerk“ ist. Hollink hat in ihrer Dissertation gezeigt, dass die so entstehenden Bildbeschriftungen dadurch eine deutlich höhere Qualität erhalten und im Vergleich zu textbasierten Metadaten deutlich vollständigere und präzisere Suchergebnisse generiert werden können [Hol06] [Nag07].

Diese beiden generellen Ansätze zur Verbesserung der Bildsuche durch *Anfrageerweiterung* und *Erstellung semantischer Metadaten* sollen in dieser Arbeit weiter verfolgt und erweitert werden. Im Gegensatz zu den Arbeiten von Nagypal und Hollink, die nur im Umfeld von Laborumgebungen betrachtet wurden, soll in dieser Arbeit auch die praktische Umsetzbarkeit in Bildarchiven von Bildagenturen mit betrachtet werden. Es ergeben sich daraus die folgenden Fragestellungen:

- **Frage 1: Wie sollen semantische Verfahren eingesetzt werden?** Es soll mit dieser Frage zum Einen ermittelt werden, (a) *an welcher Stelle in einer Bildsuchmaschine semantische Verfahren eingesetzt werden sollen*. Ist also eine Anfrageerweiterung ausreichend? Oder ist die Erstellung von semantischen Bildbeschriftungen der geeignetere Ansatz? Zum Anderen soll ermittelt werden, (b) *ob semantische Verfahren überhaupt von Bildsuchern benötigt werden* und (c), *ob die entstehenden Möglichkeiten zur Bildsuche für Nutzer in der Praxis verständlich sind und wie diese eingesetzt werden sollten*.
- **Frage 2: Wie und von wem werden semantische Hintergrundinformationen erstellt?** Der Erfolg semantischer Anwendungen hängt maßgeblich von der *Umsetzbarkeit* und der *Akzeptanz* durch die Benutzer ab [Hep07]. Semantisches Hintergrundwissen wird in solchen Anwendungen überwiegend getrennt von der Nutzung zur Beschriftung von Ressourcen erstellt und in Ontologien [Gru93] abgebildet. Es ergeben sich dadurch die folgenden Teilfragen: a) *Sind getrennt erstellte, vorgefertigte Ontologien ausreichend und praktisch anwendbar?*, b) *Wie können Bildredakteure in diesen Prozess eingebunden werden und selbst Hintergrundwissen erstellen?* und c) *Reduzieren semantische Verfahren den Zeitaufwand zur Erstellung von Metadaten?* Da die Erstellung von semantischem Hintergrundwissen selbst zeitaufwändig ist, muss ermittelt werden, unter welchen Bedingungen die semantische Bildbeschriftung schneller ist als die textbasierte Beschriftung von Bildern und somit in einer Bildagentur überhaupt wirtschaftlich umsetzbar ist.
- **Frage 3: Verbessert die Kombination der Ergebnisse automatischer Verfahren zur Beschriftung von Bildteilen das Gesamtergebnis?** Hierbei soll nicht ein neuer Algorithmus zur Objekterkennung entwickelt werden.

Die Frage ist, ob und wie eine geeignete Kombination der Einzelergebnisse verschiedener Objekterkenner das Gesamtergebnis verbessern kann.

Frage 3 geht somit dem dritten Aspekt von Semantik auf Bildern nach. Die *Bedeutung von abgebildeten Inhalten* soll ebenfalls in dieser Arbeit betrachtet werden. Hierbei ist die Beschriftung von Bildteilen nötig [JJfC00]. Die Beschriftung von Bildteilen erfordert allerdings für einen Bildredakteur einen zusätzlichen Zeitaufwand. Der Einsatz von Objekterkennern zur automatischen Erstellung von Bildteilen bietet sich also an. Wie bereits erwähnt, sind deren Ergebnisse aber oft sehr fehlerhaft und unvollständig, so dass entsprechende Maßnahmen zur Verbesserung des Gesamtergebnisses einer automatischen Beschriftung von Bildteilen nötig sind.

1.3 Ziele und Lösungsschritte

Das Ziel dieser Arbeit ist, die zu ermittelnden Antworten auf die drei formulierten Fragestellungen in einem entsprechenden Gesamtsystem mit dem Namen *ImageNotion* [WN07a] umzusetzen. Dieses soll hochwertige Suchergebnisse mittels semantischer und automatischer Verfahren zur Beschriftung von Bildteilen ermöglichen.

Da die praktische Umsetzbarkeit der Lösungen und Verfahren erreicht werden soll, ist ein weiteres Ziel die Durchführung einer *Reihe von empirischen Untersuchungen* mit Hilfe von Bildsuchern und Bildredakteuren im Umfeld von Bildagenturen. Daher werden die einzelnen Lösungsschritte mit Hilfe von *Analysen von Suchanfragen, Benutzerworkshops mit Fokusgruppen* und einer *Onlineevaluation des ImageNotion Prototyps* begleitet.

Die folgenden Schritte sollen zur Lösung der Fragestellungen auf Basis von geeigneten empirischen Untersuchungen und Evaluationen führen:

- **Art der Integration semantischer Verfahren** Zur Beantwortung von *Frage 1 a* (Ort zur Integration semantischer Verfahren) werden fünf Geschäftsführer von Bildagenturen in einem Workshop befragt und aus den Antworten ihre entstehenden Anforderungen abgeleitet.
- **Analyse von Suchanfragen an Bildagenturen** Zur Beantwortung von *Frage 1 b* (ob und in welcher Art semantische Verfahren überhaupt von Bildsuchern benötigt werden?), wird analysiert, wie Suchanfragen an Bildarchive von Bildagenturen gestellt werden. Ermittelt wird hierbei, ob überhaupt Hintergrundwissen bei der Bildsuche ausgenutzt werden muss. Hierzu werden

mehr als 100.000 Suchanfragen, die 2006 an die Metasuchmaschine fotomarktplatz.de⁶ gestellt wurden ausgewertet.

- **Gewünschte Art des Einsatzes semantischer Verfahren durch Bildsucher.** Mehr als 130 Teilnehmern haben an einer durchgeführten Onlineevaluation von ImageNotion zur Evaluation von semantischen Verfahren und der Nutzung von beschrifteten Bildteilen innerhalb der Bildsuche teilgenommen. Hierdurch soll *Frage 1 c* (ob die entstehenden Möglichkeiten zur Bildsuche für Nutzer in der Praxis verständlich sind und wie diese eingesetzt werden sollten) beantwortet werden.
- **Einsetzbarkeit bestehender semantischer Verfahren zur Bildbeschriftung** Zur Beantwortung der *Frage 2 a* (reichen herkömmliche Verfahren zur Erstellung von semantischem Hintergrundwissen aus), werden Benutzerworkshops mit sechs fachlich geschulten Bildredakteuren, die innerhalb des IMAGINATION Projekts zur Verfügung standen, durchgeführt und entsprechend ausgewertet.
- **Entwicklung eines Verfahrens zur Erstellung von Hintergrundwissen durch Benutzer** Ebenfalls mit diesen Benutzern und zur Beantwortung von *Frage 2 b* (wie können Bildredakteure in den Prozess der Erstellung von Hintergrundwissen mit eingebunden werden und dieses selbst erstellen) wird ermittelt, wie ein Verfahren zur Erstellung von Hintergrundwissen durch die Benutzer selbst aufgebaut werden muss und dieses evaluiert.
- **Zeitaufwand zur Bildbeschriftung mittels semantischer Verfahren durch Bildredakteure** Etwa 750 manuell beschriftete Bilder und Bildteile durch Bildredakteure der Bildagentur photo12 dienten zum Vergleich verschiedener semantischer Verfahren und der Ermittlung des dabei möglichen Zeitgewinns. Hierdurch kann die *Frage 2 c* (können semantische Verfahren die Erstellung von semantischen Metadaten beschleunigen) beantwortet werden.
- **Evaluation von Objekterkennung in einem Filmarchiv** Zur Beantwortung von *Frage 3* (kann die Kombination von verschiedenen automatischen Verfahren das Gesamtergebnis verbessern) werden diese 750 manuell beschrifteten Bilder und deren Bildteile als Vergleichsbasis in einer Laborevaluation verwendet, um die Qualität der entstehenden Ergebnisse von Objekterkennung zu ermitteln. Hieraus sind Rückschlüsse möglich, wie entsprechende Verbesserungen durchgeführt werden können.

⁶Fotomarktplatz.de ist eine vom Autor entwickelte Metasuchmaschine zur parallelen Abfrage der Bildsuchmaschinen von Bildagenturen. Stand 2006 waren mehr als 100 Bildsuchmaschinen an fotomarktplatz.de angebunden. Der Betrieb von fotomarktplatz.de wurde 2009 vom Betreiber PIAG AG, Baden-Baden eingestellt

1.4 Überblick über die Arbeit

Der Aufbau dieser Arbeit nach ihren einzelnen Kapiteln ist wie folgt:

Kapitel 2 geht näher auf den Begriff *qualitativ hochwertige Suchergebnisse ein* und zeigt welche Probleme diesem Ziel aktuell in Bildsuchmaschinen bei Bildagenturen gegenüberstehen.

Das *Kapitel 3* dient zum Aufbau einer geeigneten Systemarchitektur für die geplante ImageNotion Anwendung. Hierbei wird ermittelt, welche semantischen Verfahren geeignet sind und wie diese sich sinnvoll integrieren lassen.

Die *Kapitel 4 und 5* gehen der Frage nach, ob semantische Verfahren zur Bildsuche überhaupt in Bildsuchmaschinen von Bildagenturen benötigt werden und falls ja, wie diese geeignet eingesetzt werden müssen.

Innerhalb von *Kapitel 6* wird die Frage zur sinnvollen Einsetzbarkeit semantischer Verfahren aus der Perspektive des Benutzers betrachtet. Können diese in einer Bildsuchmaschine Verfahren zur semantischen Bildsuche verstehen und anwenden und sind die Verfahren von den Benutzern gewollt?

Die *Kapitel 7 bis 10* befassen sich mit der praktisch nutzbaren Erstellung von Hintergrundwissen, dass für die Nutzung von semantischen Verfahren geeignet erstellt werden muss, damit es anschließend die Schaffung hochwertiger Suchergebnisse ermöglichen kann. *Kapitel 11* untersucht durch die Auswertung der Gesamtzeit zur Erstellung semantischer Bildbeschriftung, ob diese Form der Bildbeschriftung von Bildern einen Zeitgewinn bringt. Dauert dieses Verfahren länger als die Erstellung textbasierter Beschriftungen, ist ein sinnvoller Einsatz semantischer Verfahren in der Praxis zu teuer und somit ihr Einsatz fragwürdig.

Die *Kapitel 12 und 13* betrachten die Einsetzbarkeit von Objekterkennung zur automatischen Erstellung von Beschriftungen für Bildteile und wie deren Gesamtergebnisse durch eine geeignete Kombination der Einzelergebnisse verbessert werden können.

Mit einem entsprechenden Durchlauf durch das ImageNotion System wird in *Kapitel 14* gezeigt, wie die Umsetzung semantischer Suchanfragen dort erfolgt und somit die gewünschten hochwertigen Suchergebnisse entstehen.

Eine Zusammenfassung dieser Arbeit und einen Ausblick geben das abschließende *Kapitel 15*.

Kapitel 2

Suchen von Bildmotiven

Hochwertige Suchergebnisse bei der Bildsuche haben im Umfeld von Bildagenturen eine große Bedeutung. Die Betreiber möchten möglichst *viele Bilder an ihre Kunden verkaufen*. Die Käufer der Bilder möchten möglichst *schnell ihre gewünschten Bilder finden*.

Die Betrachtung der Bildsuche in Bildagenturen ist daher der Fokus in dieser Arbeit. Weiterhin sind sowohl Angestellte in Bildagenturen als auch Bildsucher erfahren im Umgang mit der Bildsuche, so dass deren Aussagen innerhalb empirischer Untersuchungen von einem großen Wert sind und verlässliche Rückschlüsse auf zu treffende Maßnahmen zur Verbesserung der Bildsuche zulassen.

Die genauere Betrachtung des Umfelds der Bildsuche in Bildagenturen, die nötigen Abläufe zur Bildsuche und die damit mögliche Definition des Begriffs *qualitativ hochwertiges Suchergebnis* ist daher das Ziel dieses Kapitels. Weiterhin sollen generelle Ansätze gezeigt werden, die das Schaffen solcher Suchergebnisse in dieser Arbeit ermöglichen sollen.

2.1 Bildsuchmaschinen bei Bildagenturen

Bildagenturen bieten Bilder zur kommerziellen Nutzung an, damit Bildkäufer diese kaufen und für eine vordefinierte Nutzung wie Abdruck in einem Magazin oder einer Webseite verwenden können. Die Bildsuche und der Erwerb von Lizenzen findet heutzutage überwiegend über das Internet mittels Bildsuchmaschinen statt. Für hohe Verkaufsumsätze haben Bildagenturen ein hohes Interesse daran, dass Bildkäufer die gewünschten Bilder in ihren Bildarchiven finden. So besteht auf beiden Seiten ein großes Interesse an qualitativ hochwertigen Suchergebnissen.

Doch die Nutzung von Bildsuchmaschinen bei Bildagenturen führt oft zu un-

korrekten oder unvollständigen Ergebnissen [FTN⁺00] und ist dadurch zeitaufwändig¹. Dies steht im klaren Gegensatz zum Interesse von Bildagenturen und Bildkäufer, hochwertige Ergebnisse erhalten zu können.

In den folgenden Sektionen wird daher gezeigt, wie Angestellte in Bildagenturen selbst für ihre Kunden Ergebnisse zu Suchanfragen erstellen. So kann der Begriff *qualitativ hochwertiges Suchergebnis* geeignet definiert werden. Anschließend wird gezeigt, wie ein Bildkäufer selbst vorgehen muss, um seine gewünschten Bildmotive mittels einer Bildsuchmaschine zu ermitteln. Beide Abläufe werden verglichen, um die bestehenden Schwächen von Bildsuchmaschinen aufzuzeigen und so generelle Maßnahmen zur Verbesserung der Bildsuche ermitteln zu können.

2.2 Erstellung von Suchergebnissen durch Bildarchivare

Bildarchivare sind Fachkräfte zur Bildverwaltung und zur Durchführung von Suchanfragen in Bibliotheken und bei Bildagenturen. Innerhalb von Bildagenturen beantworten Bildarchivare Suchanfragen, die durch Kunden an sie gestellt werden. Basierend auf eigenen Beobachtungen der Arbeit von Bildarchivaren innerhalb des IMAGINATION Projekts² wird nun beschrieben, wie Bildarchivare diese Suchanfragen durchführen und das gewünschte Bildmaterial ermitteln. Die dabei entstehenden Suchergebnisse können als *qualitativ hochwertig* betrachtet werden, denn die Angestellten sind geschult und sehr erfahren in der Bildrecherche.

2.2.1 Ablauf der Bildsuche

Anfragen an einen Bildarchivar werden üblicherweise per Telefon gestellt. Ein Kunde teilt dem Bildarchivar mit, an was für Bildinhalten er interessiert ist. Im folgenden Beispiel sucht ein Mitarbeiter eines Verlags Bilder zu *Wichtige Politiker, die zum Fall der Mauer beigetragen haben*, um ein Lehrbuch durch Verwendung geeigneter Bilder zu illustrieren. Die Erfüllung der Suchanfrage durch einen Bildarchivar erfolgt dann in den folgenden Schritten:

1. **Suchanfrage verstehen.** Der Bildarchivar muss *verstehen*, was mit der gestellten Anfrage gemeint ist. Im aktuellen Beispiel erkennt der Bildarchivar,

¹Gleiches gilt auch für Bildsuchmaschinen zur privaten Nutzung wie soziale Netzwerkseiten bei Facebook oder Flickr. Die Suchergebnisse dort sind allerdings oft noch deutlich schlechter

²Die Beschreibung des IMAGINATION Projekts befindet sich in Sektion 13.1 dieser Arbeit

2.2. ERSTELLUNG VON SUCHERGEBNISSEN DURCH BILDARCHIVARE 31

dass das Thema der Anfrage die *Deutsche Wiedervereinigung*³ ist und Bilder gesucht werden, die *Personen* mit dem Beruf *Politiker* abbilden.

2. **Suchanfrage einschränken.** Falls die Suchanfrage dem Bildarchivar noch zu allgemein erscheint, stellt er dem Kunden weitere Fragen zur sinnvollen Einschränkung der gegebenen Anfrage.

Im aktuellen Beispiel wird der Kunde gefragt, ob die Anfrage einen Fokus auf einen oder mehrere bestimmte Politiker haben soll. Als möglicher Fokus schlägt der Bildarchivar dem Kunden Bilder des damaligen *Bundeskanzlers der BRD Helmut Kohl* vor.

Ausnutzung von Hintergrundwissen: Um die Suche gezielt einzuschränken, benötigt der Bildarchivar *Hintergrundwissen* über eine Anfrage, um diese umsetzen zu können. In diesem Fall das Wissen, dass Politiker wie unter anderem *Helmut Kohl, Hans-Dietrich Genscher, Michail Gorbatschow* und *George Bush* zu Schritten zur Erlangung der deutschen Wiedervereinigung beigetragen haben.

3. **Suchen geeigneter Bilder** Zur Erfüllung der Anfrage durchsucht der Bildarchivar nun die relevanten Teile des Bildarchivs.
4. **Erstellen einer Auswahl.** Innerhalb der gefundenen Bilder wählt der Bildarchivar die geeignetsten aus. Zur gestellten Anfrage könnten dies Nahaufnahmen einzelner relevanter Politiker sein und Bilder, die mehrere Politiker zusammen zeigen.

Ordnung der Auswahl nach Relevanz: Für Bildarchivare haben diejenigen Bilder eine hohe Relevanz, auf denen die gesuchten Bildinhalte am Deutlichsten abgebildet sind. So hat im Beispiel ein Bild, auf dem Helmut Kohl nur im Hintergrund zu sehen ist, eine niedrige Relevanz. Hingegen hat ein Bild, das Helmut Kohl deutlich erkennbar und als wichtigen Teil des Bilds zeigt, eine hohe Relevanz. Danach werden die Bilder geordnet.

5. **Hinzufügen thematisch verwandter Bilder.** Üblicherweise fügen Bildarchivare zusätzlich Bilder zur Auswahl hinzu, die thematisch verwandt sind und damit für den Anfragenden auch in Frage kommen könnten. Zur gestellten Anfrage *Deutsche Wiedervereinigung* könnten dies Bilder der *Montagsdemonstrationen in Leipzig* oder der *deutschen Botschaft in Prag* sein, an denen wichtige Ereignisse, welche zur Wiedervereinigung führten, stattgefunden haben.

Ausnutzung von zusätzlichem Hintergrundwissen Um thematisch verwandte Bilder hinzufügen oder vorschlagen zu können, muss der Bildarchivar über

³<http://de.wikipedia.org/wiki/DeutscheWiedervereinigung>

zusätzliches Hintergrundwissen verfügen. Hierdurch kann er Themen und Bilder ermitteln, die zur gestellten Anfrage in Verbindung stehen.

6. **Senden der Auswahl.** Der Bildarchivar übersendet diese an den Anfragenden als Suchergebnis.

2.2.2 Definition: qualitativ hochwertige Suchergebnisse

Zusammengefasst hat der Bildarchivar im gezeigten Ablauf die Suchanfrage des Kunden **(a)** richtig verstanden, erweitert und umgesetzt. Hierbei hat er **(b)** sein Hintergrundwissen über die genannte Anfrage ausgenutzt, um die vage Anfrage „Politiker“ richtig auf konkrete Personen wie „Helmut Kohl“ umsetzen zu können. Weiterhin hat er **(c)** das Suchergebnis geordnet nach der Wichtigkeit der gezeigten Motive innerhalb der Bilder. Insgesamt entstand eine *präzise* und zur Anfrage *vollständige* Auswahl, die als Suchergebnis an den Kunden zurück gesendet wurde.

Basierend auf dem gängigen Messverfahren durch Präzision und Recall [BYRN99] hat ein Suchergebnis dann eine hohe Qualität, wenn es sich um ein möglichst *präzises* und *vollständiges* Suchergebnis handelt.

Qualitativ hochwertiges Suchergebnis: Ein Suchergebnis, das nach den Schritten **(a)**, **(b)** und **(c)** entsteht und zur gestellten Anfrage eine hohe Präzision und Vollständigkeit hat sowie nach Bildinhalt geordnet ist, soll im Folgenden dieser Arbeit als *qualitativ hochwertiges Suchergebnis* definiert sein.

2.3 Bildsuche mittels Bildsuchmaschinen

Wie soeben gezeigt, nimmt ein Bildarchivar einem Kunden die Recherchearbeit bei der Bildsuche ab. Im Gegensatz dazu muss ein Kunde bei der Nutzung von Bildsuchmaschinen die Bildsuche und Auswahl von geeigneten Bildern selbst durchführen.

2.3.1 Ablauf

Der genaue Ablauf der Bildsuche durch Benutzer wird im Folgenden dargestellt. Das generelle Ziel der Benutzung einer Suchmaschine ist das Erfüllen eines bestehenden *Informationsbedürfnisses* [Tay68]. Das Informationsbedürfnis im Bezug auf die Bildsuche ist somit, zu einem bestimmten Thema geeignete Bilder zu finden. Zur Durchführung einer Suchanfrage wird hierzu eine Bildsuchmaschine verwendet [MS00]. Eine Folge von hintereinander gestellten Suchanfragen zu einem gewünsch-

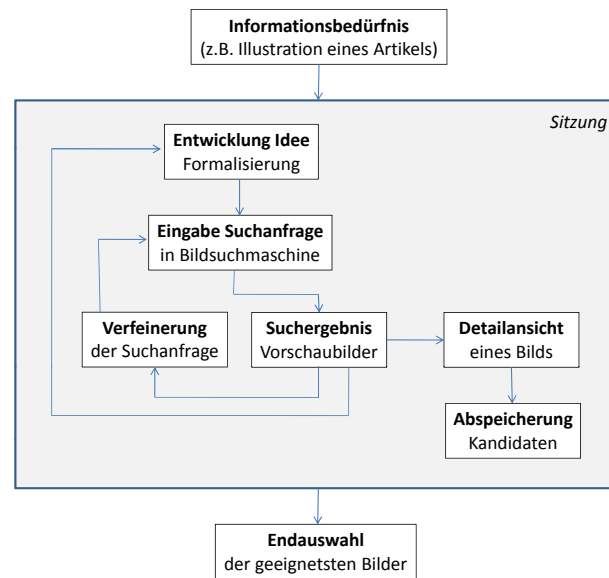


Abbildung 2.1: Ablauf einer Sitzung zur Bildsuche

ten Thema wird hierbei als *Sitzung* bezeichnet [MS00]. Abbildung 2.1 zeigt einen Überblick über den gesamten Ablauf einer Sitzung.

1. **Formalisierung eines Informationsbedürfnisses** Das Informationsbedürfnis wird zuerst formalisiert, indem dies als *Idee* ausgedrückt wird.
2. **Eingabe Suchanfrage** Anschließend wird diese Idee angepasst auf die Eingabemöglichkeiten einer Suchmaschine. Der übliche Weg ist hierzu die Eingabe von passenden *Stichwörtern* [VF05].
3. **Betrachtung des Suchergebnisses** Der Benutzer betrachtet das von der Suchmaschine zurückgegebene Suchergebnis. Für gewöhnlich wird das Suchergebnis in Form kleiner Vorschaubilder dargestellt.
4. **Detailansicht geeigneter Bilder** Für Bilder, die der Benutzer für geeignet hält, lässt er sich die Detailansicht anzeigen. Die dabei geeignetsten Bilder speichert der Benutzer als mögliche Kandidaten für sein gewünschtes Bildmotiv ab.
5. **Verfeinerung einer Suchanfrage** Liefert eine Suchanfrage zu viele oder zu unpräzise Ergebnisse zurück, wird die Suchanfrage eingeschränkt. Hierfür werden geeignete Stichwörter *eingeschränkt* hinzugefügt.

6. **Neuformulierung von Anfragen** Hat eine Suchanfrage nicht oder nur unzureichend zu gewünschtem Bildmaterial geführt, wird eine weitere Idee durch den Benutzer generiert und der Ablauf beginnt von neuem.

2.3.2 Bemessung des Gesamtaufwands

Zur Erfüllung des Bedürfnisses nach gewünschtem Bildmaterial muss ein Benutzer für gewöhnlich eine Reihe von Suchanfragen durchführen [MS00]. *Je mehr Suchanfragen innerhalb einer Sitzung durchgeführt müssen, umso höher ist der entstehende Gesamtaufwand an Zeit, um das gewünschte Bildmaterial zu erhalten.*

Idealerweise besteht demnach eine Sitzung aus genau einer Suchanfrage, die dem Benutzer direkt die gewünschten Bilder zurückliefert. Dieses Ergebnis wäre dann aus der Sicht eines Bildsuchers das qualitativ hochwertigste Suchergebnis, das eine Bildsuchmaschine erstellen kann.

2.4 Szenario: Bildsuche für ein Magazin

Mit dem folgenden Anwendungsszenario soll gezeigt werden, wie aufwändig es ist, speziell Ideen zu umfangreichen Suchanfragen auf eine Bildsuchmaschine⁴ abzubilden. Als Szenario dient der in einem Verlag häufig vorkommende Fall, dass ein Artikel für ein Magazin durch Bilder geeignet illustriert werden soll.

2.4.1 Ablauf der Bildsuche

Informationsbedürfnis Ein Bildredakteur erhält den Auftrag, Bilder für einen Artikel über das alljährlich statt findende *Oldtimer Meeting in Baden-Baden* zu finden. Von der Redaktion erhält er hierbei die Vorgaben, dass die *Ehrengastmarke des nächsten Meetings Mercedes Benz*⁵ ist und die Bilder ausdrücken sollen, *wie schön ein Besuch dieses Meetings ist.*

Ablauf der Sitzung Mit den Vorgaben macht sich der Bildredakteur an die Arbeit und entwickelt die Idee, Bilder zu suchen, die eine *Lachende Frau neben*

⁴Es wird im Szenario davon ausgegangen, dass eine textbasierte Bildsuchmaschine eingesetzt wird. Diese nimmt Stichwörter entgegen und bildet diese direkt auf die textbasierten Bildbeschriftungen ab. Die Bildbeschriftungen sind wie üblich als Metadaten vorhanden. Beschriftungen für Bildteile sind nicht vorhanden. Auf dieser Basis arbeiten die bei Bildagenturen heutzutage üblicherweise eingesetzten Bildsuchmaschinen

⁵Dies war im Jahr 2008 beim 32. Oldtimer Meeting der Fall.

einem Oldtimer von Mercedes Benz stehend zeigen. Er verwendet hierzu die Bildsuchmaschine einer Bildagentur und beginnt die Bildsuche.

1. *Suchanfrage: „Mercedes Benz“*: Die Suchanfrage „Mercedes Benz“ liefert dem Benutzer eine Vielzahl unterschiedlichster Bildmotive rund um den Konzern *Daimler AG*, Bilder von *Motorsport*, dem *Mercedes Benz Museum*, aktuellen Modellreihen und darunter auch Modelle, deren Baujahre älter als 20 Jahre⁶ sind. Dem Benutzer ist das Ergebnis *zu unpräzise*.

2. *Suchanfrage: „Mercedes Benz Oldtimer“*: Da die erste Suchanfrage zu viel Suchergebnisse generiert hat, wird die Anfrage weiter eingeschränkt zu „Mercedes Benz Oldtimer“. Viele der relevanten Bilder, die der Benutzer noch im ersten Suchergebnis als Oldtimer wahrgenommen hat, werden in diesem Ergebnis nicht mehr angezeigt. Der Benutzer nimmt als Grund an, dass demnach vielen Bilder die Beschriftung „Oldtimer“ fehlt. Das Ergebnis ist demnach *sehr unvollständig*.

3. *Suchanfrage: „Mercedes Benz Auto Frau stehen“*: Um sich dem Informationsbedürfnis neu anzunähern, formuliert der Benutzer die Anfrage neu. Das Suchergebnis soll nun Bilder zeigen, die eine Frau stehend neben einem beliebigen Auto von Mercedes Benz zeigen. Wiederum hält der Benutzer das Ergebnis für unvollständig, da wenige Bilder gezeigt werden, und sehr unpräzise. Innerhalb der gezeigten Bilder sitzt die Frau im Auto und steht nicht daneben.

4. *Suchanfrage: „Auto Frau stehen“*: Verwundert über das schlechte Suchergebnis formuliert der Benutzer die Anfrage allgemeiner durch „Auto Frau stehen“, um nur Bilder zu erhalten, die eine stehende Frau neben einem Auto zeigen. Das Ergebnis erscheint dem Benutzer nun deutlich vollständiger, es ist aber *sehr unpräzise* hinsichtlich seiner Anfrage. So muss er eine Vielzahl der Bilder im Suchergebnis ansehen, um die geeigneten Bilder zu finden.

Weitere Suchanfragen Der Benutzer stellt in der Sitzung weitere Variationen von Anfragen, unter anderem „Mercedes 170s“ oder „Frau Mercedes 300“. Dies sind *spezifische* Namen von Automodellen von Mercedes Benz.

2.4.2 Diskussion

Im Szenario hatte der Benutzer das Problem, dass er eine Vielzahl unterschiedlicher Suchanfragen stellen musste, um geeignete Bilder für sein gewünschtes Informationsbedürfnis zu erhalten. Die Sitzung war für den Benutzer somit *sehr zeitaufwändig*. Die Hauptgründe hierfür waren, dass entweder zur Anfrage passende Bilder nicht angezeigt wurden, das Ergebnis somit unvollständig war, oder die Anfragen von

⁶Der Begriff Oldtimer wird unterschiedlich definiert, ab 20 Jahre alter spricht man von Youngtimern, Veteranen sind Autos aus den 60er und 70er Jahren

der Bildsuchmaschine nicht richtig umgesetzt worden sind und somit kein oder ein unpräzises Ergebnis entstand.

Im Fall des Suchworts *Oldtimer* wäre es hilfreich gewesen, wenn die Suchmaschine alle in Frage kommenden Bilder, die einen Oldtimer anzeigen, ermittelt und angezeigt hätte. Da dies aber nicht der Fall war, musste der Benutzer selbst eine Reihe von spezifischen Suchanfragen mit Modellnamen von Autos, die als Oldtimer gelten, generieren.

2.4.3 Begriffsdefinition: Arten von Suchanfragen

Im Weiteren dieser Arbeit werden Suchanfragen basierend auf [Gud95] nach drei Abstraktionsebenen unterschieden, diese sind:

- **Generelle Anfragen** Eine generelle Suchanfrage bezieht sich auf eine beliebige Abbildung einer Sache auf einem Bild. Die Suchanfrage *Frau* soll beispielsweise Bilder zeigen, die eine beliebige Frau abbilden.
- **Spezifische Anfragen** Eine spezifische Suchanfrage bezieht sich auf eine bestimmte Sache und unter Verwendung eines exakten Bezeichners. Die Suchanfrage *Mercedes Benz* im Kontext der Suche nach Autos meint somit Autos vom spezifischen Hersteller Mercedes Benz.
- **Imaginäre Anfragen** Im Englischen wird diese Ebene als *abstract* bezeichnet. Die Übersetzung nach *abstrakt* würde bedeuten „unkonkret, unscharf“. Da mit dieser Ebene aber die gedankliche Interpretation einer Sache gemeint ist, wird der Begriff *imaginär* als Übersetzung verwendet. *Oldtimer* ist eine imaginäre Anfrage, denn damit ist die Interpretation für „alte Autos“ gemeint.

2.4.4 Ermittlung der Schwachpunkte von Bildsuchmaschinen

Durch den Vergleich der Abläufe bei der Bildsuche eines Bildarchivars mit denen bei der Nutzung einer Bildsuchmaschine werden nun die vorhandenen Schwachpunkte von Bildsuchmaschinen gezeigt

- **Umsetzung der Anfrage: Eine Suchmaschine muss die Suchanfrage interpretieren können.** Die Anfrage im Szenario ist komplex und muss daher zuerst interpretiert werden. Sie enthält alle drei mögliche Abstraktionsebenen für Suchanfragen, die entsprechend von der Suchanfrage umgesetzt werden müssen. „Frau“ als generelle Anfrage muss alle Bilder ermitteln, die

eine beliebige Frau zeigen, „Mercedes Benz“ alle Bilder, die etwas von diesem Automobilhersteller zeigen. „Oldtimer“ als imaginärer Anfrageteil muss interpretiert werden auf alle alten Autos.

- **Ausnutzung von Hintergrundwissen: Eine Suchmaschine muss die Zusammenhänge der vorhandenen Bilder kennen.** Für *generelle* und *imaginäre* Anfrageteile muss Hintergrundwissen ausgenutzt werden. Die Suchmaschine muss also zum Beispiel ermitteln können, welche Baureihen von Mercedes Benz als Oldtimer eingestuft sind.
- **Ordnung der Ergebnisse: Bilder müssen nach Bildinhalt angeordnet werden.** Bilder, die eine Frau und einen Oldtimer in Detailaufnahme zeigen, hätten im Szenario eine höhere Präzision zur gestellten Anfrage gehabt als Bilder, die nur eine Frau oder ein Auto anzeigen. Die Möglichkeit zur Erstellung einer solchen Ordnung nach Bildinhalt ist daher wünschenswert.

2.5 Semantische Ansätze zur Verbesserung der Bildsuche

Die Integration von semantischen Ansätzen in eine Bildsuchmaschine stellt den Hauptansatz dar, um die genannten Schwächen von Bildsuchmaschinen beseitigen zu können [JJfC00]. Allgemein meint Semantik die *Bedeutung von Dingen einer bestimmten Art* [Fre92].

2.5.1 Aspekte von Semantik im Kontext der Bildsuche

Im Kontext der Bildsuche gibt es drei unterschiedliche Aspekte von Semantik [JJfC00], diese sind die *Bedeutung einer gestellten Suchanfrage*, die *Bedeutung von Bildbeschriftungen* sowie die *Bedeutung von abgebildeten Inhalten* innerhalb der Bilder selbst.

1. **Semantik der Suchanfrage.** Es sind Verfahren nötig, damit Suchanfragen korrekt von einer Bildsuchmaschine verstanden und umgesetzt werden können.
2. **Semantik der Bilder.** Die Bildsuchmaschine benötigt Hintergrundwissen, um die Bilder zu verstehen und in Verbindung miteinander setzen zu können. So können die zu einer Anfrage gehörenden Anfrageteile sowie thematisch verwandte Anfragen mit berücksichtigt werden.

3. **Semantik von Bildteilen.** Basierend auf beschrifteten Bildteilen können Suchergebnisse nach der Relevanz von Bildteilen geordnet und gruppiert werden.

Durch die Beseitigung der genannten Schwachpunkte mittels semantischen Ansätzen kann also eine „intelligente Bildsuchmaschine“ entstehen. Sie hat Wissen über die Bedeutung von Suchanfragen, den Inhalten auf Bildteilen und Hintergrundwissen über die abgebildeten Motive.

2.5.2 Mögliches Suchergebnis einer „intelligenten Bildsuchmaschine“

Lachende Frau neben Oldtimer von Mercedes Benz stehend

Suchergebnis Anfrage erzeugte 231 Treffer Seite 1 / 6 [▶](#)







 DX-02-023	 43-432-123	 PK0212-B
 DX-02-030	 MN-546-31	 GTXAGD1

Abbildung 2.2: Korrekt interpretiertes Ergebnis der Anfrage *Lachende Frau neben einem Oldtimer von Mercedes Benz stehend*

Eine solche „intelligente Bildsuchmaschine“ hätte die Suchanfrage des Szenarios wie folgt umsetzen können:

1. Finde alle Bilder, die einen Bildteil mit einer Frau enthalten. Ordne die Bilder so, dass zuerst Bilder angezeigt werden, die eine lachende und stehende Frau anzeigen.

2. Finde in diesem Zwischenergebnis alle Bilder, die zusätzlich neben der Frau ein Auto abbilden. Ordne diese Bilder an den Anfang des Suchergebnisses
3. Ermittle alle Automodelle von Mercedes Benz, die Oldtimer sind. Finde diese Bilder und bevorzuge diese in der Ordnung
4. Gebe das so ermittelte und geordnete Suchergebnis zurück.

Abbildung 2.2 zeigt ein Beispiel des so entstandenen Suchergebnisses. Das erste Bild in Abbildung 2.2 zeigt daher das Bild mit höchster Präzision, es zeigt wie gewünscht eine lachende Frau die sehr nahe neben einem Oldtimer von Mercedes stehend. Im zweiten Bild steht eine Frau weiter neben einem solchen Oldtimer. Das dritte Bild ist schon unpräziser, da zusätzlich zur Frau ein Mann auf dem Bild ist und beide vor dem Auto stehen. Die weiteren Bilder zeigen entweder eine Frau neben einem Oldtimer einer anderen Marke oder nur Oldtimer.

Mit den gezeigten Schritten hätte eine Suchanfrage genügt, statt wie im Szenario gezeigt mehr als fünf Anfragen, um das gewünschte Suchergebnis zu erhalten. Es wäre also wie in dieser Arbeit gewünscht *ein hochwertiges Suchergebnis bei der Verwendung einer Bildsuchmaschine* entstanden.

2.6 Semantische Bildsuche in einem Filmarchiv

Als Beispielszenario zur Anwendung von semantischen Suchanfragen soll im Weiteren dieser Arbeit die Suche nach Bildern in einem Filmarchiv dienen. Das Filmarchiv selbst wurde im IMAGINATION Projekt vom Projektpartner photo12⁷ zur Verfügung gestellt. Es enthält 750 Bilder, die Schauspieler wie unter anderem *Uma Thurman*, *Clint Eastwood* und *Elizabeth Taylor* zeigen. Die Bilder zeigen Szenen aus Filmen wie *Pulp Fiction*, *Dirty Harry* oder *Die Frau aus dem Nichts*.

Elizabeth Taylor ist eine in *London* geborene *Schauspielerin* (englisch „actor“). Ihr oft verwendeter Spitzname ist „Liz“. Sie hatte unter anderem die Hauptrolle in Filmen wie *Butterfield 8* und *Die Frau aus dem Nichts* (englischer Originaltitel „Secret Ceremony“).

In der Bildsuche erfahrene Teilnehmer haben in einem IMAGINATION Workshop folgende Suchanfragen formuliert, die sie an ein solches Bildarchiv stellen würden:

- (1) „Liz Taylor“: bei der Suche wird der Spitzname von Elizabeth Taylor verwendet.

⁷photo12 (photos12) Paris, www.photo12.com

- (2) „Die Frau aus dem Nichts“: Alle Bilder zum Film „Die Frau aus dem Nichts“ sollen gezeigt werden.
- (3) „Schauspieler England“: Es sollen alle Bilder von Schauspielern aus England angezeigt werden
- (4) „Film Elizabeth Taylor“ Es sollen alle Bilder von Filmszenen angezeigt werden, die Elizabeth Taylor zeigen
- „(5) „Portrait Elizabeth Taylor““: Es sollen alle Bilder gezeigt werden, die Elizabeth Taylor in Detailaufnahme zeigen.

Bei den Anfragen (1) und (2) sind Probleme wie Synonyme und verschiedene Sprachen zu berücksichtigen. Die Anfragen (3) und (4) sind semantische Anfragen und erfordern somit Verfahren zur korrekten Umsetzung der Suchanfrage. Anfrage (5) erfordert die Auswertung von beschrifteten Bildteilen.

Diese Anfragen sollen im Weiteren dieser Arbeit durch geeignete semantische Verfahren ermöglicht werden, damit die Generierung von hochwertigen Suchergebnissen mittels einer Bildsuchmaschine möglich wird.

2.7 Zusammenfassung

In diesem Kapitel wurde gezeigt, dass die Nutzung von Bildsuchmaschinen innerhalb einzelner Sitzungen aufwändig ist. Ein Benutzer ist häufig gezwungen, eine Vielzahl an Suchanfragen zu einem gewünschten Bildmotiv zu stellen, bis er geeignete Suchergebnisse erhält. Dieses wiederholte Stellen von Suchanfragen wird schnell sehr zeitaufwändig.

Der Vergleich der Bildsuche durch einen Bildarchivar im Vergleich zur Abarbeitung einer Anfrage durch eine Bildsuchmaschine hat gezeigt, dass die Integration von semantischen Verfahren der Hauptansatz ist, um hochwertige Suchergebnisse schaffen zu können. Darauf basierend ist es möglich, dass eine Bildsuchmaschine die Bedeutung der Suchanfrage sowie die Bedeutung von Bildbeschriftungen kennt. Weiterhin ist das Wissen um die Bedeutung von Bildteilen wichtig, damit Suchergebnisse basierend auf den Bildinhalten geordnet und gruppiert werden können. Aus der Sicht der Bildsucher soll mit diesen Maßnahmen somit der Gesamtaufwand zum Finden von gewünschten Bildmotiven verringert werden.

Die genaue Art und Weise der Integration semantischer Ansätze in eine Bildsuchmaschine wird im nächsten Kapitel betrachtet.

Kapitel 3

Aufbau der Bildsuchmaschine ImageNotion

Das erklärte Hauptziel in dieser Arbeit ist die Erstellung einer Bildsuchmaschine, die mit Hilfe von empirischen Untersuchungen um geeignete semantische Verfahren erweitert werden soll, um die Schaffung hochwertiger Suchergebnisse zu ermöglichen. Geeignete Verfahren und Maßnahmen hierfür werden in diesem Kapitel ermittelt. Die Suchmaschine wird mit dem Namen *ImageNotion* bezeichnet.

Zum Aufbau der Bildsuchmaschine werden zuerst die Architektur und die Komponenten einer herkömmlichen, textbasierten Bildsuchmaschine vorgestellt. Darauf basierend können die möglichen Orte zur Integration semantischer Verfahren und deren Beitrag zur Schaffung hochwertiger Suchergebnisse gezeigt werden. Dadurch kann die Frage 1 (a), *an welcher Stelle in einer Bildsuchmaschine semantische Verfahren eingesetzt werden sollen*, der in dieser Arbeit zu beantwortenden Fragestellungen beantwortet werden.

Anschließend können Architektur und Komponenten von ImageNotion vorgestellt werden. Hierbei wird gezeigt, wie die einzelnen Komponenten durch in der Literatur bekannten bestehenden Lösungen umgesetzt werden können. Ebenso wird dargestellt, welche Komponenten durch eigene Lösungen in dieser Arbeit erstellt oder geeignet erweitert werden müssen und welche konkreten Fragestellungen sich daraus für den weiteren Verlauf dieser Arbeit ergeben. Zum Abschluss des Kapitels wird ein Vergleich der Konzeption der ImageNotion Systemarchitektur mit verwandten Arbeiten durchgeführt.

3.1 Textbasierte Bildsuchmaschinen

Textbasierte Bildsuchmaschinen sind bei Bildagenturen überwiegend im Einsatz¹.

Im vorigen Kapitel wurde beschrieben, welche Schwachstellen bei der Nutzung dieser Suchmaschinen bestehen. Architektur und Komponenten einer textbasierten Bildsuchmaschine werden daher in dieser Sektion vorgestellt, um diese anschließend durch geeignete semantische Verfahren erweitern zu können.

3.1.1 Die Notwendigkeit zur Erstellung von Metadaten

Computer können die einzelnen Wörter textbasierter Dokumente selbst lesen und indizieren, so dass anschließend eine textbasierte Suche nach Dokumenten möglich wird. Auf dieser Basis arbeiten zum Beispiel webbasierte Suchmaschinen wie Google [BP98]. Die Erstellung einer textbasierten Suchmaschine kann also *ohne zusätzlichen manuellen Beschriftungsaufwand* der Dokumente erfolgen.

Im Gegensatz hierzu stellt für einen Computer ein digitales Bild lediglich eine Folge von Koordinaten im zweidimensionalen Raum dar, die jeweils den Farbwert für einen Bildpunkt beinhalten [GW01]. Zur Anzeige des Bilds werden diese Bildpunkte geladen und zur Darstellung an die Grafikkarte des Computers gesendet. Somit kann ein Computer digitale Bilder anzeigen, laden und speichern. *Den Inhalt eines Bilds selbst versteht ein Computer aber nicht.*

Um die Suche nach Bildern zu ermöglichen, ist daher zuerst manueller Aufwand nötig. Zu jedem Bild müssen dazu *Metadaten* erstellt werden, die beschreiben, welche Inhalte auf dem Bild abgebildet sind [FTN⁺00]. Diese Metadaten können anschließend indiziert werden, so dass die Bildsuche in einer Bildsuchmaschine möglich wird. Häufig verwendet werden hierbei die Standards IPTC [Int05] und Dublin Core [PNNJ05]. Innerhalb solchen Standards werden Felder vorgegeben, die für bestimmte Inhalte vorgesehen sind. In IPTC gibt es zum Beispiel die Felder „Aufnahmeort“ zur Angabe des Aufnahmeorts, „Copyright“ zur Angabe des Rechteinhabers und „Keywords“ zur Angabe von Stichwörtern.

3.1.2 Architektur und Komponenten

Zusätzlich zu Komponenten, die eine Bildsuche ermöglichen, benötigt daher eine Bildsuchmaschine auch Komponenten zur Erstellung und Verwaltung von Bildbe-

¹Aus eigener Erfahrung: innerhalb der Metasuchmaschine fotomarktplatz.de waren 2006 über 100 Bildsuchmaschinen von Bildagenturen angebunden. Alle basierten auf dem IPTC Standard und waren vollständig textbasiert.

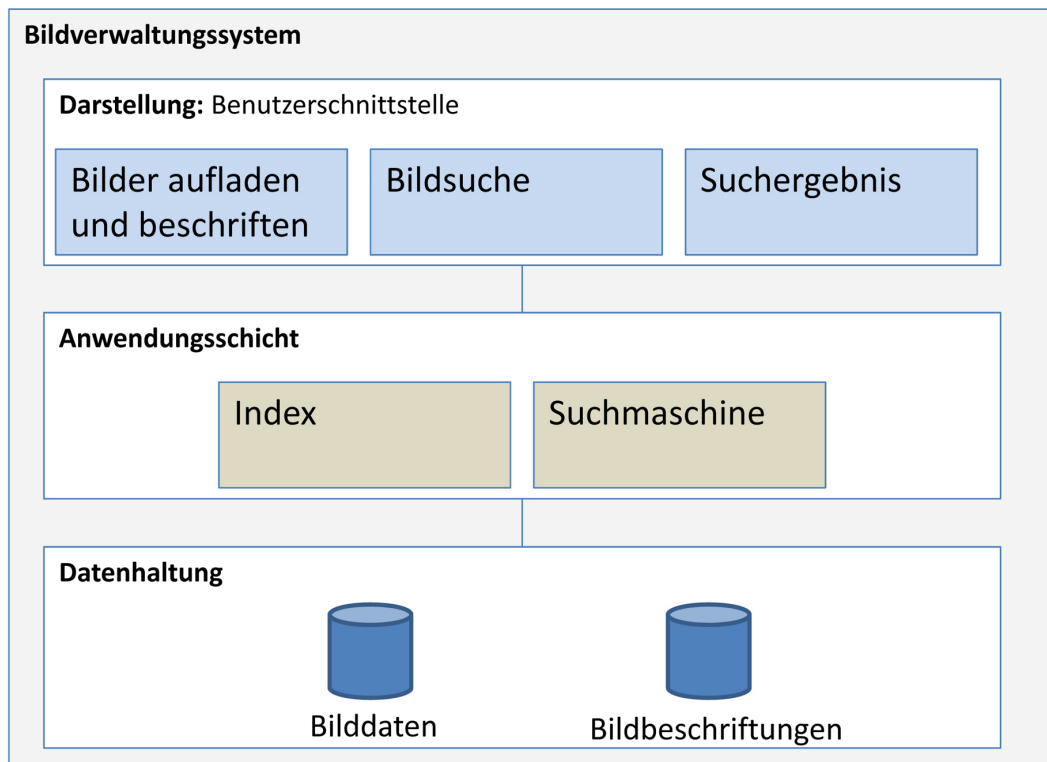


Abbildung 3.1: Architektur und Komponenten einer textbasierten Bildsuchmaschine

schriftungen. Im Folgenden wird, basierend auf [AHAQI⁺07], die typische Architektur und die vorhandenen Komponenten einer textbasierten Bildsuchmaschine vorgestellt.

Abbildung 3.1 zeigt den generellen Aufbau einer Bildsuchmaschine. Die Architekturen setzen üblicherweise die *Drei-Schichten-Architektur* zur Trennung von Darstellung, Anwendung und Datenhaltung [Ram01] um. Darauf basierend besteht eine Bildsuchmaschine generell aus den folgenden drei Schichten und Komponenten:

- **Datenhaltung:** Innerhalb der Datenhaltung werden die *Bilddaten* selbst gespeichert sowie die Metadaten, also die *Bildbeschriftungen*.
- **Anwendungsschicht:** Die Anwendungsschicht enthält die Komponenten *Index* und *Suchmaschine*. Der Index verwendet die textbasierten Bildbeschriftungen und bereitet diese auf. So können anschließend auch in großen Datenbeständen Suchanfragen in kurzer Zeit beantwortet werden. Die Suchmaschine hat die Aufgabe, die Beantwortung von Suchanfragen zu koordinieren.
- **Darstellungsschicht:** Die Darstellungsschicht ermöglicht die *Bildsuche* durch

die Eingabe von Suchanfragen sowie der Anzeige von *Suchergebnissen*. Hierzu werden Suchanfragen an die Suchmaschine gesendet. Diese fragt mit der Eingabe den Index ab und generiert basierend auf dem zurückkommenden Ergebnis die Suchergebnisse. Das Ergebnis wird in Form von Suchergebnisseiten aufbereitet und dargestellt. Die Darstellung kann beispielsweise in Form von Webseiten erfolgen. Jede Suchergebnisseite besteht aus einer Menge von kleinen Vorschaubildern. Diese sind verlinkt mit einer jeweiligen Detailansicht, welche das Bild in vergrößerter Form sowie die Metadaten anzeigt. Die Komponente *Bilder aufladen und beschriften* ermöglicht das *Aufladen* neuer Bilder sowie die *Beschriftung* bestehender Bilder. Alle Eingaben werden an die Suchmaschine weitergegeben. Diese gibt die Bilddaten und Bildbeschriftung an die Datenhaltung zur Speicherung weiter. Anschließend wird der Index benachrichtigt, damit dieser seine Daten entsprechend aktualisieren kann.

3.1.3 Lexikalische Probleme in der Bildsuche

Im vorigen Kapitel wurden bereits die Schwachstellen textbasierter Bildsuchmaschinen aus der Sicht eines Bildsuchers ermittelt. Hierbei lag der Fokus auf dem korrekten Interpretieren von Suchanfragen durch Ausnutzung von Hintergrundwissen, sowie dem fehlenden Verständnis für die Inhalte auf Bildteilen.

Ein weiteres Problem ist die *Unstrukturiertheit textbasierter Metadaten* [Nag07]. Metadaten in textbasierten Bildsuchmaschinen bestehen häufig aus einer Abfolge von Stichwörtern. Dies stellt zwar ein sehr einfach anwendbares Verfahren zur Erstellung von Bildbeschriftungen dar. Allerdings sind die so entstehenden Metadaten unstrukturiert [BSW⁺07], die Bildsuchmaschine kennt also nicht die Zusammenhänge der einzelnen Stichwörter. Dadurch treten die folgenden *lexikalischen Probleme* auf [BSW⁺07], die häufig zu unvollständigen und unkorrekten Suchergebnissen führen:

- **Berücksichtigung alternativer und synonyme Bedeutungen:** Alternative und synonyme Bedeutungen zu einer Eingabe, z.B. *Orange* ist ein Synonym zu *Apfelsine*, werden bei der Suche nicht berücksichtigt.
- **Trennung Homonyme:** Homonyme Wörter werden in Suchanfragen wie Schloss (kann bedeuten *Bauwerk* oder *Schließanlage*) oder Kohl (*Helmut Kohl* vs. *Gemüsekohl*) nicht getrennt und führen zur Anzeige vieler unkorrekter Bilder im Ergebnis.
- **Unterstützung anderer Sprachen:** Oft sind in Bildarchiven von Bildagenturen Bilder in unterschiedlichen Sprachen beschriftet, zum Beispiel durch

Aufkauf von Bildbeständen aus dem Ausland. Die unterschiedlichen Sprachen werden nicht bei der Bildsuche berücksichtigt.

- **Schreibfehler:** Bildbeschriftungen enthalten oft Schreibfehler. Dadurch werden die entsprechenden Bilder bei Suchanfragen nicht angezeigt.

Bei der Verbesserung der Bildsuche hin zu hochwertigen Suchergebnissen müssen also die genannten lexikalischen Probleme zusätzlich berücksichtigt werden.

3.2 Ontologien innerhalb der Bildsuche

Ontologien [Gru93] ermöglichen die Abbildung von Hintergrundwissen, so dass dieses einem Computer zugänglich und von diesem ausgenutzt werden kann. *Die Nutzung semantischer Ansätze in der Bildsuche erfordert somit zusätzlich die Erstellung entsprechender Ontologien, damit das dort vorhandene Hintergrundwissen ausgenutzt werden kann.*

Daher werden im Folgenden zuerst der Begriff Ontologie und mögliche Arten von Ontologien für die Bildsuche genannt. Im nächsten Abschnitt wird dann gezeigt, welche semantischen Verfahren sich konkret für die gewünschte Erzielung hochwertiger Suchergebnisse eignen und welche Arten von Ontologien hierbei verwendet werden müssen.

3.2.1 Definition Ontologie

Die Definition von Gruber für den Begriff Ontologie stellt die anerkannteste und gleichzeitig allgemeinste Definition dar. Nach Gruber ist eine Ontologie eine *specification of a conceptualization* [Gru93], also der Festlegung einer Konzeptualisierung. Eine Konzeptualisierung ist hierbei das Modellieren von Dingen der realen Welt. Einzelne Elemente des Modells müssen dabei explizit sein, so dass alle Benutzer des Modells das gleiche Verständnis darüber haben. Die Art der Modellierung gibt hierbei vor, wie die Dinge der Welt beschrieben und miteinander in Beziehung gesetzt werden können. Eine Ontologie ist demnach die *formale, explizite Spezifikation einer gemeinsamen Konzeptualisierung* [Smi08]. Ontologien dienen somit zur Beschreibung von *Dingen und Sachverhalten* und somit auch zur *Abbildung von Hintergrundwissen*.

Ontologien werden im Weiteren mit dem Fokus auf die Integrierbarkeit in eine Bildsuchmaschine und mittels Beispielen aus der Praxis beschrieben. Für die entsprechenden formalen Darstellungen von Ontologien wird auf die Literatur verwiesen (z.B. [HKS06] [Gua98]).

3.2.2 Arten von Ontologien

Eine Ontologie kann sein ...

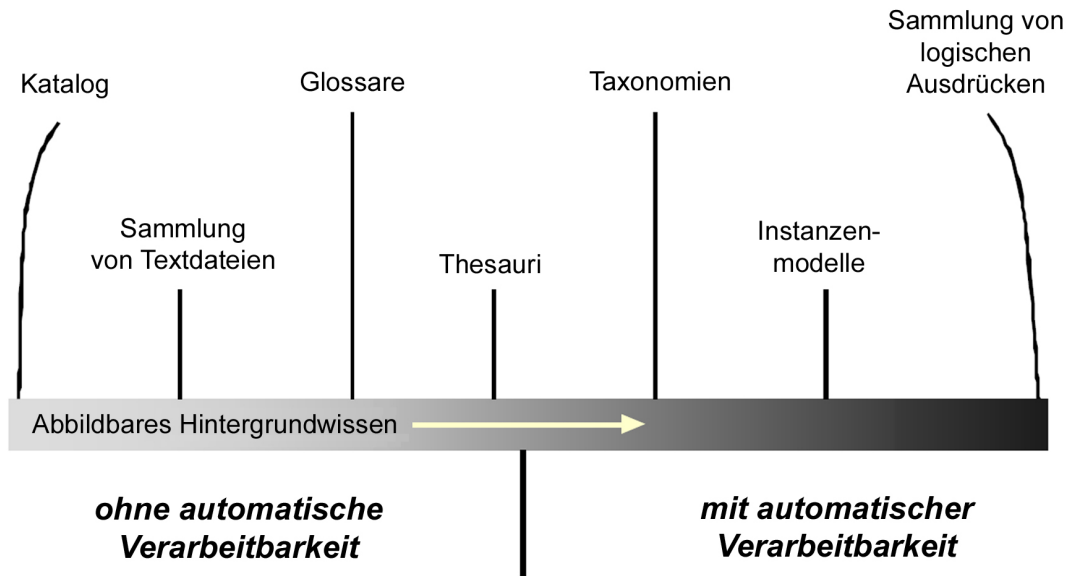


Abbildung 3.2: Klassifizierung von Ontologien

Basierend auf Welty und Lehmann [Wel99] lassen sich Ontologien wie in Abbildung 3.2 gezeigt kategorisieren. Die einfachsten Formen von Ontologien sind demnach bereits in einer textbasierten Bildsuchmaschine enthalten. Ein *Katalog* ist das Bildarchiv selbst mit eindeutigen Bildnamen. Die *Sammlung von Textdateien* sind die Bildbeschriftungen selbst auf Basis von textbasierten Metadaten. Unter einem *Glossar* wird eine Liste von Wörtern mit genauen Erklärungen verstanden. In einer Bildsuchmaschine könnte dies die Festlegung von einzelnen Begriffen und deren Beschreibungen sein, die anschließend einheitlich zur Bildbeschriftung verwendet werden.

Zur Verbesserung der Bildsuche mittels geeigneten semantischen Verfahren sind die in Abbildung 3.2 weiterhin gezeigten Arten von Ontologien *Thesauri*, *Taxonomien* und *Instanzenmodelle* interessant. Diese erlauben die umfangreiche Abbildung von Hintergrundwissen, welches in der Bildsuche ausgenutzt werden kann. Taxonomien und Instanzenmodelle sind zusätzlich automatisch verarbeitbar, ein Computer kann somit das darin vorhandene Wissen ausnutzen, interpretieren und verarbeiten [Gru93].

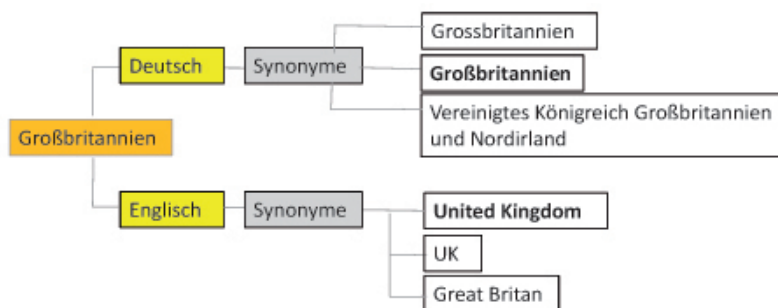


Abbildung 3.3: Ein Thesaurus für Synonyme und unterschiedliche Sprachen einzelner Wörter

Thesaurus (kontrolliertes Vokabular)

Unter einem Thesaurus wird heutzutage und basierend auf der Definition von Gruber auch in dieser Arbeit ein *kontrolliertes Vokabular* verstanden, dessen Elemente zueinander in Beziehung gesetzt werden können². Ein Thesaurus kann damit die beschriebenen lexikalischen Probleme in der Bildsuche lösen. So können beispielsweise durch den in Abbildung 3.3 gezeigten Thesaurus für Synonyme und unterschiedliche Sprachen alle synonymen und englischen Bedeutungen zu einer deutschsprachigen Suchanfrage ermitteln und bei der Bildsuche berücksichtigen („Großbritannien“ wird hierzu erweitert zu „Großbritannien, UK, Great Britain, ...“).

Taxonomien

Innerhalb einer Taxonomie können *semantische Elemente* definiert werden. Jedes semantische Element beschreibt ein *Ding* oder einen *Sachverhalt*, zum Beispiel „Elizabeth Taylor“. Dieses konkrete Element wird formal als *Instanz* bezeichnet [Gru93]. Der Sachverhalt „Schauspieler(in)“ beschreibt entsprechend die Menge aller Schauspieler. Die Zusammenfassung einer gleichartigen Menge von Instanzen als ein semantisches Element wird als *Konzept* bezeichnet [Gru93].

Eine Relation ist eine Beziehung zwischen zwei semantischen Elementen [Wel99]. Innerhalb einer Taxonomie sind *hierarchische Relationen zwischen zwei Konzepten* und *zwischen Instanzen mit Konzepten* erlaubt. „Frau aus dem Nichts“ kann so zum Beispiel die Relation zu „Psychothriller“ erhalten. Die Vererbungseigenschaft (Transitivität) der hierarchischen Relationen sichert die maschinelle Verarbeitung von Taxomien zu. Hat das semantische Element „Psychothriller“ die hierarchische Rela-

²In der Literatur wird manchmal auch eine Hierarchie als Thesaurus bezeichnet. Hierarchische Strukturen sollen aber innerhalb dieser Arbeit von Computern auswertbar sein bezüglich ihrem vorhandenen Hintergrundwissen und sind daher den Taxonomien zugeordnet.

tion zu „Film / Genre“, kann somit abgeleitet werden, dass „Frau aus dem Nichts“ ein „Film“ ist.

Taxonomien eignen sich in der Bildsuche daher zur Auflösung von *generellen* und *imaginären* Suchanfragen hin zu *spezifischeren* Anfragen wie zum Beispiel der Anfrage „Film“ hin zu spezifischen Teilanfrageteilen wie „Butterfield 8“ (ist ein „Film“) oder „Die Frau aus dem Nichts“.

Instanzenmodelle

Zusätzlich zu Taxonomien erlauben Instanzenmodelle das Erstellen von Relationen zwischen Instanzen. Instanzenmodelle erlauben somit die umfangreiche Abbildung der Zusammenhänge vorhandener Bildinhalte in einem Bildarchiv.

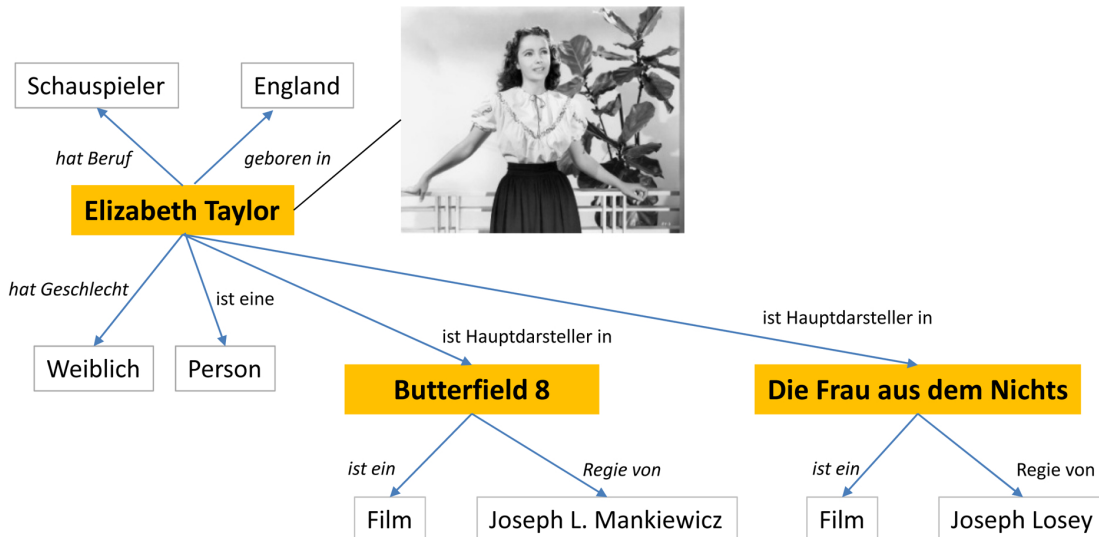


Abbildung 3.4: Ausschnitt aus einem Instanzenmodell für ein Filmarchiv

Abbildung 3.4 zeigt einen möglichen Teil eines Instanzenmodells zur Beschreibung der nötigen semantischen Elemente für die in Sektion 2.6 genannten Suchanfragen. Daher beinhaltet das Instanzenmodell unter anderem die Elemente „Elizabeth Taylor“, „Butterfield 8“ und „Die Frau aus dem Nichts“. Die Relationen zwischen den semantischen Elementen erlauben unter anderem das Hintergrundwissen auszudrücken, dass „Elizabeth Taylor“ eine „Schauspielerin“ ist, eine „weibliche“ „Person“ ist, die in „England“ geboren ist und die Hauptrolle in den Filmen „Butterfield 8“ und „Die Frau aus dem Nichts“ spielt.

Instanzenmodelle erlauben somit ebenso wie Taxonomien die Auflösung von generellen und imaginären Suchanfragen hin zu spezifischen Suchanfragen. Aufgrund

3.3. MÖGLICHE ORTE ZUR INTEGRATION SEMANTISCHER VERFAHREN⁴⁹

von umfangreicherem Hintergrundwissen durch Relationen können auch semantische Suchanfragen wie „Filmszenen Elizabeth Taylor“ direkt aufgelöst werden.

Sammlung von logischen Ausdrücken

Die weitere Form einer Ontologie laut Abbildung 3.2 sind eine Sammlung von logischen Ausdrücken. Diese ermöglichen zum Beispiel auszudrücken, *dass „Oldtimer“ alle Autos sind, die in Deutschland älter als 25 Jahre sind.* Es können also zu den erstellten Taxonomien und Instanzenmodellen Regeln formuliert werden.

3.3 Mögliche Orte zur Integration semantischer Verfahren

Generell gibt es, basierend auf den in Sektion 2.5 genannten Aspekten von Semantik bei Bildern zwei mögliche Orte in einer textbasierten Bildsuchmaschine zur Integration von semantischen Verfahren.

Zum Einen können diese an der Stelle der Umsetzung von Suchanfragen eingesetzt werden und dienen hierbei der Erweiterung von Suchanfragen. Die textbasierten Metadaten der Bilder selbst werden mit diesen Verfahren nicht verändert [CCC10]. Innerhalb dieser Arbeit werden diese Verfahren zusammengefasst unter dem Begriff *Anfrageerweiterung*.

Zum Anderen ist es möglich, die textbasierten Metadaten selbst zu verändern. Bei der *semantischen Bildbeschriftung* werden zur Erstellung von Metadaten für die Bilder direkt semantische Elemente aus einer Ontologie verwendet [VFaC05].

Zusätzlich zu diesen beiden Verfahren ermöglicht die Beschriftung von Bildteilen, dass einer Bildsuchmaschine die Bedeutung von Bildteilen zugänglich gemacht werden kann.

Aufgrund der beiden möglichen Integrationsorte soll daher die folgende Fragestellung innerhalb der folgenden Sektionen beantwortet werden:

Frage 1 (a): An welcher Stelle in einer Bildsuchmaschine sollen semantische Verfahren eingesetzt werden?

3.3.1 Anfrageerweiterung

Verfahren zur Anfrageerweiterung werden als eigene Komponente zwischen die Komponenten *Suchanfrage* und *Suchmaschine* gestellt. Die Suchanfrage wird entgegengenommen, erweitert und erst dann an die Suchmaschine zur Erstellung des Suchergebnisses weitergegeben [CCC10]. Für jedes der folgenden Verfahren zur Erweiterung von Anfragen wird beschrieben, welchen Beitrag es zur Schaffung von hochwertiger Suchergebnisse leisten kann.

Lexikalische Anfrageerweiterung

Verfahren zur lexikalischen Anfrageerweiterung sollen die beschriebenen lexikalischen Problem in der Bildsuche lösen [CCC10]. Zur Erweiterung von Anfragen ist zuerst *die Erstellung geeigneter Thesauri nötig*. So müssen zur Erkennung von Synonymen und Homonymen zuerst *Synonymlisten* erstellt werden. Entsprechend müssen *Übersetzungslisten* erstellt werden, um die einzelnen Wörter einer Sucheingabe richtig übersetzen zu können. Um zu einer Suche ähnliche Begriffe vorschlagen zu können, muss eine Taxonomie erstellt werden, die Begriffe hierarchisch ordnen kann. Entsprechend kann ein Thesaurus erstellt werden, der häufige Schreibfehler zu einzelnen Wörtern auflistet.

Auf der Basis des erstellten Thesaurus (oder Thesauri) läuft die Anfrageerweiterung dann in den folgenden drei Schritten ab [CCC10]:

1. Nehme die Suchanfrage entgegen und finde diese Wörter im Thesaurus.
2. Ermittle das kontrollierte Vokabular an Wörtern im Thesaurus
3. Nehme die Suchanfrage zusammen mit dem kontrollierten Vokabular aus dem Thesaurus und sende diese an die Suchmaschine.

Die Suchanfrage aus dem Beispielszenario „(1) Liz Taylor“ (siehe Abschnitt 2.6) kann mit lexikalischer Anfrageerweiterung also mit Synonymen erweitert werden zu „Liz | Elizabeth Taylor“. Entsprechend kann die Anfrage „(2) Die Frau aus dem Nichts“ erweitert werden mit der englischsprachigen Übersetzung zu „Die Frau aus dem Nichts | Secret Ceremony“.

Durch die lexikalische Anfrageerweiterung soll ein *vollständigeres Suchergebnis* entstehen. Natürlich muss mindestens eines der Wörter der erweiterten Anfrage in den Metadaten geeigneter Bilder im Archiv vorhanden sein, damit es im Suchergebnis erscheinen kann.

3.3. MÖGLICHE ORTE ZUR INTEGRATION SEMANTISCHER VERFAHREN⁵¹

In einem textbasierten Bildarchiv ist daher die Nutzung von Verfahren zur lexikalischen Anfrageerweiterung ein erster Schritt in Richtung von hochwertigeren Suchergebnissen, da diese durch die Lösung der lexikalischen Probleme *vollständiger* werden.

Semantische Anfrageerweiterung

Die lexikalische Anfrageerweiterung hat eine Anfrage nur textbasiert erweitert, aber die Anfrage selbst nicht interpretiert. Somit ist mit lexikalischen Verfahren noch nicht die gewünschte semantische Bildsuche möglich.

Die semantische Anfrageerweiterung basiert daher auf Taxonomien oder Instanzenmodellen, welche die *Zusammenhänge der in einem Bildarchiv vorhandenen Bilder beschreiben*. Darauf basierend läuft die semantische Anfrageerweiterung wie folgt ab [CCC10]:

1. Nehme die Suchanfrage und finde das passende semantische Element in der Taxonomie (bzw. im Instanzenmodell).
2. Ermittle die zur Anfrage in Relation stehenden semantischen Elemente (z.B. die Hierarchie unterhalb des semantischen Elements).
3. Ermittle die lexikalischen Wörter, die mit den gefundenen semantischen Elementen verknüpft sind.
4. Sende die so interpretierte Suchanfrage an die Suchmaschine.

Unter Verwendung des in Abbildung 3.4 gezeigten Instanzenmodells kann so die Suchanfrage aus dem Beispielszenario „(3) Schauspieler England“ erweitert werden zu „Schauspieler | Actor | Liz | Elizabeth Taylor | England | London“. Entsprechend kann die Anfrage „(4) Film Elizabeth Taylor“ erweitert werden zu „Film | Butterfield 8 | Die Frau aus dem Nichts | Elizabeth Taylor“.

Durch die semantische Anfrageerweiterung können Suchergebnisse gerade auf den ersten Rängen deutlich *präziser* und *vollständiger* werden [Nag07]. Das Ergebnis wird vollständiger, da auch Bilder gefunden werden, die statt „Film“ die Titel „Butterfield 8“ oder „Die Frau aus dem Nichts“ enthalten. Je mehr der Filmtitel im Instanzenmodell enthalten sind (z.B. „Cleopatra“ ist ein weiterer Film mit Elizabeth Taylor), desto vollständiger wird das entsprechende Ergebnis. Das Suchergebnis wird präziser, da die Suchanfrage selbst korrekt von der Suchmaschine verstanden wurde und so nur die Bilder anzeigt, die eine Verbindung mit Filmtiteln haben.

Die Nutzung von Verfahren zur semantischen Anfrageerweiterung ermöglicht die gewünschte Interpretation und korrekte Auflösung von Suchanfragen und bietet so *einen deutlichen Schritt in Richtung hochwertiger Suchergebnisse*.

Problem bei der Abbildung von semantischen Elementen auf Text:

Da die semantischen Elemente auf Text abgebildet werden, geht das dabei enthaltene Hintergrundwissen verloren [Nag07]. Dadurch besteht bei sehr umfangreichen Erweiterungen die Gefahr, dass die Suchergebnisse wieder unpräziser werden. Ein Beispiel ist die Erweiterung „London“ (in (3)). Diese kann auch zu Suchergebnissen führen, die Aufnahmen von der Stadt London zeigen, da der textbasierten Suchmaschine nicht bekannt ist, dass der Kontext der Suchanfrage Schauspieler sind.

Indizierung erweiterter Bildbeschriftungen

Sowohl bei der lexikalischen wie auch der semantischen Anfrageerweiterung können die resultierenden Suchanfragen recht lange dauern, wenn diese zur Laufzeit der Suche durchgeführt werden müssen. Zusätzlich kann die Ermittlung geeigneter semantischer Elemente bei der Anfrageerweiterung recht lange dauern [Nag07]. Da die Benutzer von Suchmaschinen schnelle Suchergebnisse gewohnt sind, müssen diese innerhalb weniger Sekunden erzeugt werden. Beobachtungen haben gezeigt, dass Suchanfragen von Benutzern abgebrochen werden, wenn diese länger als zehn Sekunden dauern [VFaC05]. Lang andauernde Suchanfragen, auch wenn sie hochwertige Ergebnisse liefern, sind somit nicht sinnvoll, da sie nicht beim Benutzer ankommen.

Es bietet sich daher an, die Anfrageerweiterungen ebenfalls zu indizieren [CCC10]. Der in Metadaten fehlende Inhalt der Beziehungen von Thesauri beziehungsweise die textuellen Beschriftungen von semantischen Elementen einer Taxonomie oder Instanzenmodells wird dann zusätzlich mit den vorhandenen Metadaten in den Index geschrieben. Fehlt zum Beispiel das Wort „Film“ in den Metadaten eines Bildes zu „Butterfield 8“, wird es mit indiziert. Anschließend können Suchanfragen direkt an den Index weiter gegeben werden und die Bildsuche funktioniert wieder wie gewohnt schnell.

Problem der dynamischen Aktualisierung des Indexes Allerdings besteht das große Problem dieses Verfahrens darin, dass der Index die Änderungen in einer Ontologie nicht mit bekommt, der Index also nicht dynamisch reagiert [KPJ09]. Werden Änderungen in einer Ontologie durchgeführt, müssen daher anschließend alle Bilder neu indiziert werden, um den Index zu aktualisieren. Gerade in großen Bildbeständen können diese Aktualisierungen sehr lange dauern.

Da sich Bildarchive kontinuierlich erweitern und oft neue Sachverhalte hinzukommen, ist dies ein sehr störendes Problem bei Bildagenturen, da diese ihre Änderungen gerne möglichst direkt ihren Bildkunden zugänglich machen möchten.

Automatische Vorschläge von Suchanfragen (Autocomplete)

Schließlich ist die Generierung von automatischen Vorschlägen von Suchanfragen ein weiteres mögliches Verfahren zur Erweiterung von Suchanfragen (sogenanntes Autocomplete). Die Erweiterung der Anfrage findet hierbei während der Eingabe einer Suchanfrage durch den Benutzer statt [HM06].

Sobald ein Benutzer Zeichen in das Suchfeld eingibt, werden diese an die Suchmaschine übersandt. Die Suchmaschine ermittelt Wörter oder semantische Elemente, welche die Eingabe zu ganzen Wörtern oder zusammengehörenden Wortfolgen (z.B. Namen von Personen) vervollständigen und bei einer Bildsuche zu Ergebnissen führen.

Der Beitrag dieses Verfahrens hin zu hochwertigen Suchergebnissen ist, dass hiermit die Menge an Suchanfragen in einer Sitzung minimiert werden kann. Der Benutzer muss nicht mehr ausschliesslich „raten“, welche Worte zu Suchergebnissen führen, sondern bekommt passende Wörter direkt angezeigt und kann bereits vor der Durchführung einer Suche entscheiden, welches vorgeschlagene Suchwort er verwenden möchte.

3.3.2 Semantische Metadaten

Im Gegensatz zu den Verfahren zur Anfrageerweiterung muss bei der Nutzung von *semantischen Metadaten* die Beschriftung selbst geändert werden. Im Folgenden werden zuerst die Vorteile dieses Verfahrens für hochwertige Suchergebnisse genannt. Anschließend wird beschrieben, wie *semantische Metadaten erstellt werden* und darauf basierend *semantische Suchanfragen gestellt werden* können.

Bei der Erstellung *semantischer Metadaten* wird die Bildbeschreibung nicht mit Text, sondern mittels semantischer Elemente einer Ontologie durchgeführt. Hierbei werden entweder Taxonomien oder Instanzenmodelle verwendet, da diese das umfangreiche Abbilden von Hintergrundwissen und weiterhin eine maschinelle Verarbeitung ermöglichen [Hol06].

Die Hauptvorteile semantischer Beschriftungen verglichen mit textbasierten Bildbeschriftungen sind schnell ersichtlich: *semantische Beschriftungen haben eine höhere Qualität im Vergleich zu textbasierten Beschriftungen* [Hol06] [VFAC05] [vOTSP07]. Es entstehen einheitliche Bildbeschriftungen aufgrund der Verwendung einheitlicher semantischer Elemente, es müssen keine Transformationen zu Text durchgeführt werden (wodurch kein Wissen verloren geht) und die Indizierung sowie Gewichtung von semantischen Bildbeschriftungen ist hochwertiger.

In Labormessungen wurde gezeigt, dass mit semantischen Metadaten im Ver-

gleich zu textbasierten Metadaten deutlich vollständigere und präzisere Suchergebnisse generiert werden können [Nag07] [Hol06]. *Die Erstellung semantischer Metadaten erscheint daher als die geeignetste Form zur Erstellung hochwertiger Suchergebnisse in einer Bildsuchmaschine.*

Die Abläufe zur Erstellung von semantischen Metadaten und bei der semantischen Bildsuche werden nun gezeigt.

Erstellung von semantischen Beschriftungen

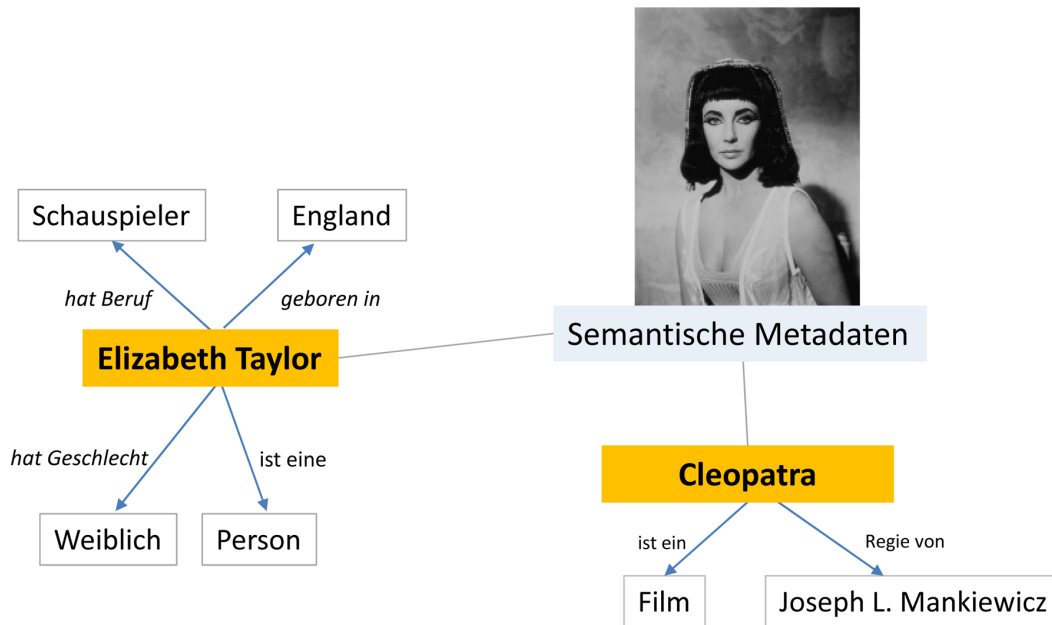


Abbildung 3.5: Semantische Beschriftung einer Szene aus dem Film Cleopatra

Bei der semantischen Bildbeschriftung werden einzelne semantische Elemente einer Ontologie ausgewählt und diese den Metadaten hinzugefügt. So entstehen *semantische Metadaten*. Abbildung 3.5 zeigt ein Beispiel zur Beschriftung einer Szene aus dem Film *Cleopatra*. Auf dem Bild wird die Schauspielerin Elizabeth Taylor gezeigt. Entsprechend wurden aus einem Instanzenmodell die semantischen Elemente *Cleopatra* und *Elizabeth Taylor* ausgewählt und zusammen mit den Vorhandenen Relationen den semantischen Metadaten hinzugefügt.

Semantische Bildsuche

Bei der semantischen Bildsuche wählt der Benutzer direkt ein semantisches Element aus einer Ontologie aus beziehungsweise eingegebener Text wird auf das am Besten passende semantische Element abgebildet. Die Suchmaschine bildet diese Anfrage dann direkt auf die semantischen Metadaten ab.

Da *Elizabeth Taylor* unter anderem mit den semantischen Elementen *Schauspieler* und *weiblich* verknüpft ist, kann dieses Hintergrundwissen ausgenutzt werden, um die Anfrage „(3) Schauspieler England“ des Beispielszenarios ohne weitere Transformation umzusetzen. Die Suchmaschine ermittelt hierzu im ersten Schritt die semantischen Elemente aus einer Ontologie³. Anschließend werden alle Bilder ermittelt, die als semantische Metadaten bzw. durch deren Relationen mit dem Element „Schauspieler“ beschriftet wurden. Zurück kommen zum Beispiel Bilder, die *Uma Thurman* oder *Elizabeth Taylor* zeigen. Dieses Ergebnis wird gefiltert nach allen Bildern, bei denen die semantischen Metadaten mit *England* beschriftet wurden beziehungsweise über die Relationen der ermittelten Schauspieler eine Verbindung zu *England* haben. Da für *Elizabeth Taylor* die Relation *geboren in England* besteht (siehe Abbildung 3.5), bleiben als Suchergebnis alle Bilder übrig, die Elizabeth Taylor zeigen.

Somit werden keine unkorrekten Ergebnisse generiert, wie beispielsweise Luftaufnahmen von England, da die Bedeutung der Suchanfrage der Suchmaschine bekannt ist. Dieses Beispiel zeigt also anschaulich, warum die semantische Suche vollständigere und präzisere Suchergebnisse erzielt.

Strukturierte Neuformulierung von Suchanfragen

Neben der bereits mit der semantischen Anfrageverfeinerung gezeigten Möglichkeit zur Erweiterung genereller Suchanfragen zu konkreteren Anfragen („Schauspieler“ zu „Elizabeth Taylor“) erlaubt die semantische Suche auch die *strukturierte Neuformulierung von Suchanfragen*.

Dadurch kann innerhalb einer Bildsuche nach „Elizabeth Taylor“ die Bildsuchmaschine durch Nutzung der Ontologie feststellen, dass nach einer *Schauspielerin* in *Filmszenen* gesucht wurde. Die Bildsuchmaschine kann dann beispielsweise ermitteln, mit welchen anderen Schauspielern Elizabeth Taylor innerhalb von weiteren Bildern abgebildet ist und die Namen dieser Schauspieler zur strukturierten Neuformulierung einer gegebenen Suchanfrage vorschlagen.

³Es wird hierbei davon ausgegangen, dass diese Elemente in der Ontologie vorhanden sind und zur Erstellung semantischer Metadaten verwendet wurden

3.3.3 Beschriftung von Bildteilen

Mit Hilfe der Beschriftung von Bildteilen wird einer Bildsuchmaschine die Bedeutung einzelner Bildteile zugänglich gemacht. Bei der manuellen Beschriftung von Bildteilen markiert ein Benutzer einen Bildteil und verknüpft diesen anschließend mit einem semantischen Element einer Ontologie (oder erstellt eine Textbeschriftung). Die manuelle Erstellung von Beschriftungen für Bildteile ist somit sehr zeitaufwändig.

Die Beschriftung von Bildteilen ermöglicht anschließend die Ordnung eines Suchergebnisses nach der Größe von Bildteilen. Auch können, wie durch Benutzer in IMAGINATION gefordert, in der Anzeige eines Suchergebnisses Bilder mit einer Beschriftung auf einem Bildteil bevorzugt werden.

Falls passende Bildteile auf Bildern beschriftete wurden, kann die Suchmaschine anschließend die Suchanfrage „(5) Portrait Elizabeth Taylor“ aus dem Beispielszenario auswerten. Hierzu werden alle Bilder ermittelt, deren Bildteile mit „Elizabeth Taylor“ versehen sind. Da ein Portrait eine Person in Nahaufnahme zeigt, werden alle Bilder herausgefiltert, die mehr als eine Person als beschrifteten Bildteil zeigen (dies erfordert semantische Bildbeschriftungen). Anschließend werden die Bilder herausgefiltert, bei denen die Größe des Bildteils kleiner als einen bestimmten Prozentsatz (z.B. 50 Prozent) auf dem Bild ausmacht. Übrig bleibt so das gewünschte Suchergebnis mit Portaitaufnahmen von Elizabeth Taylor .

Beschriftete Bildteile tragen somit zu einer hohen Präzision gerade auf den ersten Rängen eines Suchergebnisses bei, denn es können dort Bilder gezeigt werden, welche die Suchanfrage direkt auf einen Bildteil abbilden.

3.3.4 Analyse 1: Integration der semantischen Verfahren bei Bildagenturen

Bei der Erstellung von Bildsuchmaschinen mit semantischen Verfahren wird unter Laborbedingungen davon ausgegangen, dass man aufgrund der höheren Qualität direkt auf semantische Metadaten aufsetzt (z.B. in [Hol06] und [VFAC05]). Ist dieser Ansatz praktisch umsetzbar? Während des IMAGINATION Projekts wurden fünf Geschäftsführer von Bildagenturen hinsichtlich ihren Anforderungen und Meinungen zur Integration von semantischen Techniken befragt.

In einem Workshop wurden die die Möglichkeiten semantischer Verfahren vorgestellt. Anschließend fand eine Diskussion mit der Fragestellung „Anforderungen und Meinungen zum Einsatz semantischer Verfahren“ statt. Im Folgenden werden die Antworten der Geschäftsführer von Bildagenturen wiedergegeben:

- (A1) „*Semantische Bildsuche vielversprechend, da eine hohe Qualität der Such-*

3.3. MÖGLICHE ORTE ZUR INTEGRATION SEMANTISCHER VERFAHREN⁵⁷

ergebnisse möglich sein zu scheint“

- (A2) *„Brauchen unsere Kunden wirklich bessere Suchverfahren?“*
- (A3) *„Können unsere Kunden mit einer semantischen Bildsuchmaschine umgehen?“⁴*
- (A4) *„Wir haben bereits mehrere hunderttausend Bilder im Archiv. Müssen diese nun komplett mit semantischen Beschriftungen neu versehen werden?“*
- (A5) *„Die Neubeschriftung aller Bilder ist für uns zu teuer und unwirtschaftlich. Dann können wir diese Verfahren nicht einsetzen.“*
- (A6) *„Semantische Anfrageerweiterung vielversprechend, da auch für textbasierte Metadaten eine hohe Qualität der Suchergebnisse möglich wird.“*
- (A7) *„Wir würden gerne die Erstellung semantischer Metadaten nur für neue und wichtige Bilder in unserem Archiv durchführen.“*
- (A8) *„Kann ein Bildarchiv kontinuierlich von einer textbasierten zu einer semantischen Bildsuchmaschine umgestellt werden?“*
- (A9) *„Semantische Bildbeschriftung verspricht Zeitersparnis.“*
- (A10) *„Können textbasierte Bildbeschriftungen automatisch in semantische Bildbeschriftungen umgewandelt werden?“*
- (A11) *„Können Bildteile automatisch beschriftet werden?“*
- (A12) *„Wie zuverlässig arbeiten automatische Verfahren?“*
- (A13) *„Bildteile sollten direkt als semantischen Metadaten erstellt werden“*
- (A14) *„Woher kommen eigentlich die Ontologien?“*

Als ein zusammenfassendes Fazit lässt sich sagen, dass semantische Verfahren generell in Bildsuchmaschinen als vielversprechend angesehen werden, diese also *nicht komplett abgelehnt werden* (siehe (A1) und (A6)). Gleichzeitig bestehen Zweifel, ob diese wirklich *von ihren Kunden benötigt* werden und wenn ja, ob sie damit umgehen können (siehe (A1) und (A2)). Die hauptsächlichen Sorgen sind die *Integrierbarkeit der semantischen Verfahren* in ihr Archiv, speziell aufgrund von Kosten und Aufwand. Eine berechnete Frage, die oft bei der Erstellung von semantischen

⁴Interessanterweise hat keiner der Teilnehmer gefragt, ob ihre Angestellten mit der Erstellung von semantischen Metadaten umgehen können. Doch auch diese Gruppe sollte die Verfahren verstehen. Daher werden sie in dieser Arbeit nicht vergessen und in einem eigenen Workshop befragt.

Anwendungen vernachlässigt wird, ist die Frage nach der *Herkunft von Ontologien* (A14).

In den weiteren Abschnitten werden die einzelnen Antworten aufgegriffen und davon Entscheidungen für den Aufbau von ImageNotion abgeleitet.

3.3.5 Entscheidung zum Einsatzort semantischer Verfahren

Speziell die Aussagen (A4) und (A5) zeigen, dass im Unterschied zu Laborumgebungen eine Bildsuchmaschine bei Bildagenturen nicht direkt in ein auf semantischen Metadaten basierendes System umgestellt werden kann. Obwohl die zu erwartende Qualität der Suchergebnisse dann am Höchsten wäre, sind der Aufwand und somit die Kosten zu hoch, um alle Bilder in einem Archiv komplett neu mit semantischen Bildbeschriftungen zu versehen.

Gleichzeitig wird mit der Aussage (A7) der Wunsch geäußert, neue oder sehr wichtige Bilder im Archiv mit semantischen Metadaten zu versehen. Die Aussage (A8) formuliert den Wunsch entsprechend deutlicher, nämlich mit der Frage, „ob ein Bildarchiv kontinuierlich umgestellt werden kann hin zu einer auf semantischen Metadaten basierenden Bildsuchmaschine“.

Grundlegende Entwurfsentscheidung und Antwort zu Frage 1 (a) Die Frage, an welcher Stelle in der Bildsuche semantische Verfahren eingesetzt werden sollen, *lässt sich somit nicht eindeutig beantworten*. Die direkte Nutzung semantischer Metadaten ist nicht möglich, da der Aufwand zu hoch ist. Gleichzeitig erscheint den Betreibern von Bildagenturen die Erstellung von semantischen Metadaten der bessere Ort für semantische Verfahren zu sein, da hierdurch die höchste Qualität für Suchergebnisse zu erwarten ist.

Die grundlegende Entwurfsentscheidung und gleichzeitig Antwort zu Frage 1 (a) lautet daher:

Frage 1 (a): An welcher Stelle in einer Bildsuchmaschine sollen semantische Verfahren eingesetzt werden?

Grundlegende Entwurfsentscheidung: *ImageNotion muss als eine Hybridlösung aus textbasierten und semantischen Metadaten betrieben werden. Der Einsatz semantischer Verfahren muss sowohl an der Stelle der Anfrageerweiterung wie im Bereich der Erstellung von semantischen Metadaten erfolgen.*

Mit dieser Entwurfsentscheidung wird ermöglicht, dass einzelne Teile des Archivs mit semantischen Metadaten versehen werden können. Gleichzeitig muss die textbasierte Bildsuche für den Rest des Archivs weiter betrieben werden. Zur Verbesserung der textbasierten Bildsuche sollen die Verfahren zur lexikalischen und semantischen Anfrageerweiterung eingesetzt werden. Dies ermöglicht, dass ein kon-

tinuierlicher Übergang von einem auf textbasierten Metadaten hin zu einer auf semantischen Metadaten basierender Bildsuchmaschine möglich ist.

Die Beschriftung von Bildteilen kann nur (oder überwiegend) durch automatische Verfahren durchgeführt werden, da eine manuelle Beschriftung zu zeitaufwändig ist (A11). Zur Erhöhung der Zuverlässigkeit dieser Verfahren (A12) ist die Idee in dieser Arbeit, die Ergebnisse einer Reihe von verschiedenen automatischen Verfahren geeignet zu kombinieren.

3.4 Architektur und Komponenten von ImageNotion

Aufgrund der getroffenen Entwurfsentscheidung wird *ImageNotion* als eine *Hybridlösung* erstellt, welche sowohl mit textbasierten als auch mit semantischen Metadaten umgehen kann. Semantische Verfahren werden daher sowohl an der Stelle der Anfrageerweiterung wie auch im Bereich der Erstellung von Metadaten eingesetzt. Die geplante Beschriftung von Bildteilen durch automatische Verfahren erfordert entsprechende Komponenten. Ebenso muss es für Benutzer möglich sein, deren Ergebnisse zu validieren und zu korrigieren.

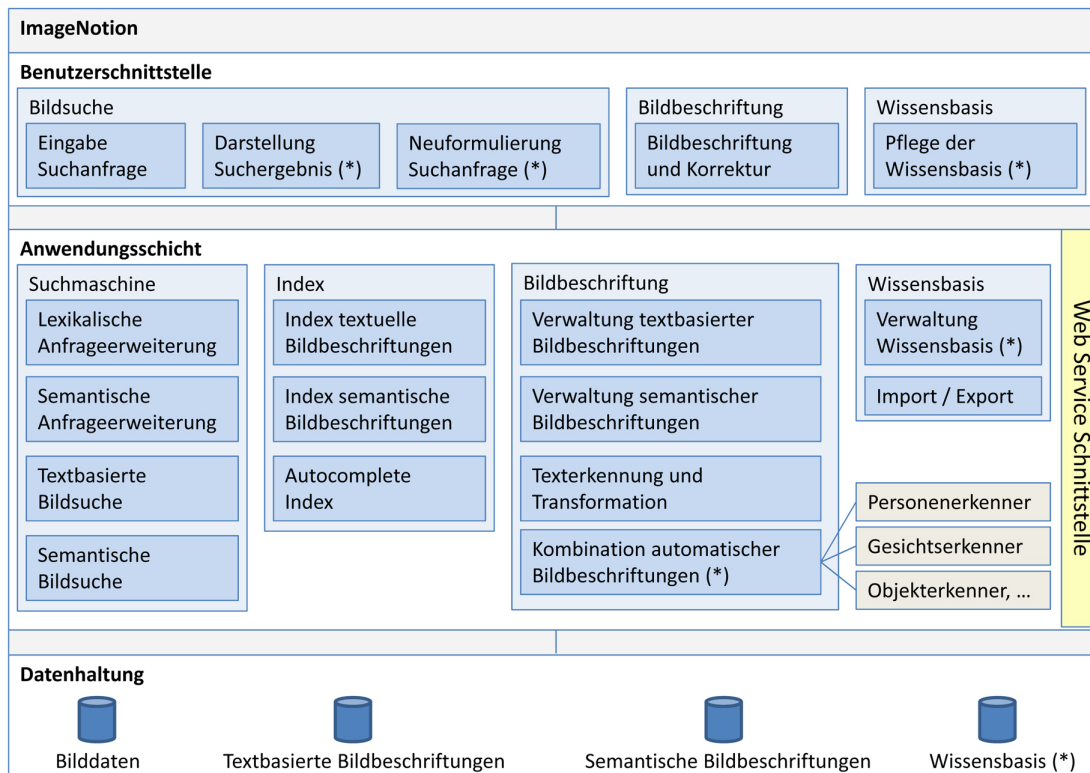
Abbildung 3.6 gibt einen Überblick über die daraus entstehende Gesamtarchitektur und Komponenten von ImageNotion. Die einzelnen Komponenten werden jeweils beschrieben. Falls kein weiterer Handlungsbedarf innerhalb dieser Arbeit für die jeweilige Komponente besteht, wird ein mögliches Verfahren aus der Literatur zur Umsetzung der Komponente genannt. Falls Handlungsbedarf besteht, werden daraus Fragestellungen abgeleitet, die innerhalb der weiteren Arbeit gelöst werden sollen.

3.4.1 Benutzerschnittstelle

Die Benutzerschnittstelle besteht aus den drei Bereichen *Bildsuche*, *Bildbeschriftung* und *Wissensbasis*. Im Vergleich zur in Abbildung 3.1 gezeigten Architektur einer textbasierten Bildsuchmaschine ist somit die Komponente *Wissensbasis* neu. Sie soll die Pflege der für die semantischen Verfahren verwendeten Ontologien ermöglichen.

Bildsuche

Die Komponente *Bildsuche* in ImageNotion besteht aus den drei Teilen *Eingabe Suchanfrage*, *Darstellung Suchergebnis* und *Neuformulierung Suchanfrage*:



(*) : Aufbau / Abläufe müssen im weiteren Verlauf dieser Arbeit noch ermittelt werden

Abbildung 3.6: Architektur und Komponenten der Bildsuche ImageNotion

- **Eingabe Suchanfrage:** Ein Textfeld zur Eingabe einer Suchanfrage ist der für Bildsucher gewohnte Startpunkt zur Formulierung einer Suchanfrage [VFAC05]⁵. Falls semantische Bildbeschriftungen vorhanden sind, werden geeignete semantische Elemente während der Eingabe mittels dem *Index über semantische Elemente* ermittelt und angezeigt (Indizierung von Anfrageerweiterungen). Zusätzlich werden, solange in einem Archiv nicht alle Bilder mit semantischen Bildbeschriftungen versehen sind, passende textbasierte Suchanfragen angezeigt. Der Benutzer kann dann seine gewünschte Suchanfrage auswählen. Abhängig von seiner Auswahl startet er dann direkt eine semantische Suchanfrage oder beginnt mit einer textbasierten Suchanfrage. Die resultierende Anfrage wird an die *Suchmaschine* weiter gegeben.
- **Darstellung Suchergebnis:** Falls Beschriftungen für Bildteile oder seman-

⁵Eine möglich Alternative wäre, die Ontologie anzuzeigen und den Benutzer ein gewünschtes Element zum Starten einer Suchanfrage auswählen zu lassen

tische Bildbeschriftungen vorhanden sind, werden diese bei der Darstellung der Bilder im Suchergebnis zusätzlich zu textbasierten Bildbeschriftungen mit angezeigt.

- **Neuformulierung Suchanfrage:** Hintergrundwissen über die Bilder und deren Zusammenhänge aus vorhandenen Ontologien soll ausgenutzt werden, um dem Benutzer Vorschläge zur Neuformulierung seiner Suchanfrage zu geben.

Fragestellungen Bildsuche Bezüglich der Bildsuche sind zwei wichtige Antworten zu klären, die innerhalb der Diskussion mit Geschäftsführern von Bildagenturen aufgetreten sind:

- (A2) Brauchen unsere Kunden wirklich bessere Suchverfahren?
- (A3) Können unsere Kunden mit einer semantischen Bildsuchmaschine umgehen?

Zum Einsatz semantischer Verfahren in einer Bildsuche ist also eine entsprechende *Motivation der Betreiber von Bildagenturen* nötig, damit diese an einer entsprechenden Verwendung interessiert sind. Die resultierenden Fragestellungen innerhalb dieser Arbeit lauten daher:

Frage 1 (b)

Werden semantische Verfahren zur Verbesserung der Bildsuche benötigt?

Zur Beantwortung dieser Frage werden in Kapitel 5 Suchanfragen an fotomarktplatz.de geeignet analysiert.

Frage 1 (c)

Sind die entstehenden Möglichkeiten zur Bildsuche für Nutzer in der Praxis verständlich und wie sollen diese eingesetzt werden?

Zur Beantwortung dieser Frage wird in Kapitel 6 eine Onlineevaluation durchgeführt. Damit kann anschließend beantwortet werden, wie die Komponenten „Eingabe Suchanfrage“ und „Neuformulierung Suchanfrage“ aufgebaut werden sollen.

Die genaue *Darstellung des Suchergebnisses*, speziell der Anzeige semantischer Metadaten ist ebenfalls eine zu beantwortende Frage. Diese Frage wird aber innerhalb der Onlineevaluation nur aus der inhaltlichen Sicht gesehen, also „Welche Informationen sollen angezeigt werden“. Die Frage, „wie diese Informationen dargestellt werden sollen“ müssten mit empirischen Untersuchungen zur Benutzerschnittstelle selbst gelöst werden und verbleiben innerhalb dieser Arbeit als Ausblick.

Bildbeschriftung und Korrektur

Die Komponente *Bildbeschriftung* einer textbasierten Bildsuchmaschine wird erweitert um die Möglichkeit zur manuellen Erstellung von Bildteilen und zur Validierung der Ergebnisse von automatisch erstellten Bildbeschriftungen. Entsprechend muss es möglich sein, Bildteile zu markieren und mit Beschriftungen zu versehen. Basierend auf der Aussage (A13) werden die Beschriftungen für Bildteile direkt mit semantischen Beschriftungen erstellt.

Da ImageNotion eine Hybridlösung ist, wird weiterhin die Erstellung textbasierter Metadaten für die Bilder ermöglicht. Zusätzlich werden Möglichkeiten zur Erstellung semantischer Metadaten in ImageNotion angeboten.

In dieser Arbeit ist nur die Integration der genannten Funktionalitäten selbst in ImageNotion von Interesse. Es werden daher keine empirischen Untersuchungen zum geeignetsten Aufbau der Benutzerschnittstelle selbst durchgeführt.

Pflege der Wissensbasis

Bei der Komponente *Pflege der Wissensbasis* handelt es sich um eine neue Komponente. Sie soll die Bearbeitung bestehender semantischer Elemente einer Ontologie und die Erstellung neuer semantischer Elemente ermöglichen.

Fragestellungen Der Grund für diese Komponente resultiert aus der gestellten Frage (A14): Woher kommen eigentlich die Ontologien?

Ontologien werden für gewöhnlich getrennt erstellt von der eigentlichen Nutzung in einer semantischen Anwendung [FLGP02]. Dieser Ansatz funktioniert in Bereichen, bei denen sich der Inhalt der Ontologie selten ändert, zum Beispiel in medizinischen Anwendungen zur Abfrage von Krankheiten.

Sollen aber Ressourcen mit semantischen Verfahren beschrieben werden (z.B. Webseiten mit SOBOLEO [ZBS09] oder wie in dieser Arbeit Bilder), besteht das Problem, dass kontinuierlich neue Inhalte hinzukommen. Fehlen in den Ontologien semantische Elemente für diese Inhalte, können diese nicht wie gewünscht beschrieben werden. Benutzer müssen dann warten, bis die Ontologien entsprechend aktualisiert worden sind [Hep07].

Der Erfolg einer Bildsuchmaschine, die auf semantischen Verfahren basieren soll, hängt daher maßgeblich davon ab, *wie gut die Ontologien die zu beschreibenden Bildinhalte abdecken*. Innerhalb dieser Arbeit wird daher die folgende These aufgestellt.

These zur Ontologierstellung

Die Erstellung von semantischen Metadaten erfordert Möglichkeiten, damit die Benutzer selbst Ontologien erweitern und erstellen können, die zu ihren Bildinhalten passen.

Zur Überprüfung der These werden die folgenden Fragen in dieser Arbeit aufgestellt:

Frage 2 (a)

Sind getrennt erstellte, vorgefertigte Ontologien ausreichend und praktisch anwendbar?

Zur Beantwortung dieser Frage wird in Kapitel 7 ein Workshop mit Bildredakteuren durchgeführt und geprüft, inwieweit der getrennte Ansatz der Ontologierstellung für Bildarchivare verständlich und umsetzbar ist.

Frage 2 (b)

Wie können Bildredakteure in diesen Prozess eingebunden werden und selbst Hintergrundwissen erstellen?

Wie muss hierzu ein geeignetes Verfahren aussehen und wie muss es in den Ablauf der Bildbeschriftung integriert werden?

Zur Beantwortung dieser Frage wird in den Kapiteln 9 und Kapitel 10 ein geeignetes Verfahren ermittelt und mit Hilfe von Bildredakteuren evaluiert.

In Abschnitt 3.3.4 wurde der Kommentar (A9) „*Semantische Bildbeschriftung verspricht Zeitersparnis.*“ gegeben. Es stellt sich die Frage, ob dem wirklich so ist? Ist also die Erstellung von semantischen Bildbeschriftungen in der Praxis wirklich schneller oder gar zeitaufwändiger? Wäre die Beschriftung zeitaufwändiger, würde diese bedeuten, dass dieses Verfahren aufgrund von zu hohen Kosten durch Betreiber von Bildagenturen abgelehnt werden würde. Umgekehrt wäre es eine starke Motivation zur Nutzung der semantischen Bildbeschriftung, wenn dieses Verfahren einen deutlichen Zeitgewinn mit sich bringen würde, da dann Kosten eingespart werden können. Eine weitere Fragestellung in dieser Arbeit lautet daher

Frage 2 (c)

Reduzieren semantische Verfahren den Zeitaufwand zur Erstellung von Metadaten?

Zur Beantwortung dieser Frage werden im Kapitel 11 entsprechende Zeitmessung bei der Nutzung der semantischen Bildbeschriftung mit dem Zeitaufwand für die Erstellung von textbasierten Metadaten verglichen.

3.4.2 Anwendungsschicht

Die Anwendungsschicht wird um semantische Verfahren zur Anfrageerweiterung und zur semantischen Bildsuche und Erstellung von Metadaten erweitert. Es ergeben sich hieraus die Komponenten *Suchmaschine*, *Index*, *Bildbeschriftung* und *Wissensbasis*. Zusätzlich soll die Anwendungsschicht eine *Web Service Schnittstelle* haben.

Suchmaschine

In die Suchmaschine wurden die Komponenten lexikalische Anfrageerweiterung sowie semantische Anfrageerweiterung integriert. Weiterhin beherrscht die Suchmaschine die *textbasierte Suche* sowie die *semantische Suche*. Innerhalb der Suchmaschine entstehen keine Fragestellungen, die detaillierter in dieser Arbeit untersucht werden müssen.

Lexikalische und semantische Anfrageerweiterung: Diese Komponenten arbeiten wie in Abschnitt 3.3.1 beschrieben. Für schnelle Suchergebnisse werden die möglichen Anfrageerweiterungen indiziert. Die Geschäftsführer der Bildagenturen gaben an, dass es ein geeigneter Kompromiss sei, wenn der Index regelmäßig nachts (in Zeiten geringer Last) aktualisiert wird. Die dabei entstehenden Probleme der dynamischen Indexaktualisierung (siehe Abschnitt 3.3.1) werden also in Kauf genommen, man akzeptiert also im Eintausch gegen Geschwindigkeit einen nur tagesaktuellen Index.

Textbasierte und semantische Suche: Die textbasierte Suche arbeitet wie bisher. Die semantische Suche erhält entweder direkt ein semantisches Element als Suchanfrage oder bildet eine Texteingabe auf das am besten passende Element ab. Anschließend wird der Index aufgerufen zur Durchführung der Suchanfrage.

Index

Der Index wurde unterteilt in die drei Teilkomponenten *Index textuelle Beschriftungen* (verwaltet wie bisher den Index textbasierter Metadaten), *Index semantische Beschriftungen* und *Autocomplete Index*.

Index semantische Beschriftungen: Die Komponente *Index semantische Beschriftungen* soll die semantischen Metadaten indizieren und gewichten, um daraus in schneller Zeit geeignete Suchanfragen und Ordnungen der Ergebnisse zu erstellen.

Das Verfahren von Nagypal [Nag07] ist hierbei das aus der Literatur am besten geeignetste Verfahren zur Indizierung semantischer Metadaten. Die semantischen Beschriftungen werden hierbei in einen Volltextindex abgebildet und gewichtet. Die

Gewichtung hängt dabei von der Art der semantischen Elemente ab. Elemente, die direkt für die Bildbeschriftung verwendet werden (z.B. „Elizabeth Taylor“) erhalten eine hohe Gewichtung. Damit verbundene semantische Elemente (z.B. „Schauspieler“) erhalten eine entsprechend niedrigere Gewichtung, entsprechend niedriger gewichtet werden weitere Oberkonzepte einer Taxonomie. So entstehen präzise Suchergebnisse speziell auf den ersten Rängen eines Suchergebnisses in wie gewünscht kurzer Suchzeit (was durch Nagypal umfangreich durch entsprechende Evaluationen belegt wurde [Nag07]).

Dieses Verfahren wird innerhalb von ImageNotion um eine weitere Form der Gewichtung erweitert. Semantische Beschriftungen von Bildteilen werden bevorzugt gegenüber von Beschriftungen, die das ganze Bild betreffen. Innerhalb der Bildteile werden große Bildteile vor kleinen Bildteilen bevorzugt. So wird erreicht, dass Bilder, die ein Motiv auf dem Bild dominant anzeigen, im Suchergebnis bevorzugt werden. Dies führt zur gewünschten Ordnung eines Suchergebnisses nach Bildinhalt.

Ein Index über semantische Metadaten kann weiterhin dynamisch aktualisiert werden. Werden Änderungen an semantischen Elementen durchgeführt, so können die betroffenen Bilder abgefragt werden. So muss die Aktualisierung des Indexes nur für diese Bilder und dadurch nur einem Teil des Indexes durchgeführt werden. Der Index kann somit immer aktuell gehalten werden, ohne dass dieser jeweils komplett neu erstellt werden muss.

Solange es in einem Bildarchiv sowohl textbasierte wie auch semantische Metadaten gibt, werden der textbasierte und der semantische Index parallel abgefragt. Bilder mit semantischen Beschriftungen werden aufgrund der erwarteten höheren Qualität der damit generierbaren Suchergebnisse bevorzugt und erst danach Bilder gezeigt, welche nur textbasierte Beschriftungen haben.

Index Autocomplete: Die Komponente *Index Autocomplete* erlaubt das automatische Vorschlagen von Suchanfragen (wie in Sektion 3.3.1 beschrieben) bei Benutzereingaben.

Bildbeschriftung

Die Komponente Bildbeschriftung wurde gegenüber Abbildung 3.1 umfangreich erweitert. Statt bisher nur manuell *textbasierte Metadaten* erstellen zu können, ist mit ImageNotion parallel dazu die semantische Bildbeschriftung möglich. Weiterhin werden automatische Prozesse integriert, sowohl durch die Komponente *Texterkennung und Transformation* als auch der Komponente *Kombination automatischer Beschriftungen*.

Textbasierte und semantische Bildbeschriftung: Über diese Komponenten ist die Eingabe von Metadaten möglich. Zur Erstellung textbasierter Bildbe-

schriftungen gibt es weiterhin Felder zur Eingabe von Text. Für die Erstellung semantischer Bildbeschriftung müssen semantische Elemente aus den Ontologien ausgewählt werden können, damit diese für die Erstellung von semantischen Metadaten verwendet werden können.

Texterkennung und Transformation: Mit der Frage (A10) wurde von den Geschäftsführern von Bildagenturen (siehe Abschnitt 3.3.4) die Frage gestellt, ob textbasierte Metadaten automatisch in semantische Metadaten umgewandelt werden können.

Hierzu eignet sich der Einsatz von Texterkennern und Texttransformierern. Diese werden trainiert auf die Erkennung von Dingen einer bestimmten Domäne [BF06]. So können beispielsweise aus Wikipedia die Namen aller Filme oder Schauspieler ermittelt werden und diese im Texterkenner als Datenbasis hinterlegt werden. Findet dieser entsprechende Stellen in den textbasierten Metadaten, können diese zu semantischen Elemente transformiert werden. Wird zum Beispiel der Text „Elizabeth Taylor“ gefunden, kann dieses in das entsprechend (vorhandene) Element einer Ontologie transformiert werden. Zusätzlich ist mit diesem Verfahren das Vorschlagen von möglichen, neuen semantischen Elementen für eine Ontologie möglich [BF06].

Eine mit diesen Verfahren arbeitende Komponente *Texterkennung und Transformation* wird daher zu diesem Zweck in ImageNotion vorgesehen.

Kombination automatischer Prozesse: Da einzelne automatische Prozesse wie ein Gesichtserkennung oder Objekterkennung häufig wichtige Stellen nicht erkennen, ist der neuartige Ansatz in ImageNotion, eine Vielzahl dieser Algorithmen zu verwenden und die Ergebnisse sinnvoll zu kombinieren, um ein präziseres und damit besseres Gesamtergebnis ermöglichen zu können.

Fragestellungen:

Hinsichtlich der Idee der Kombination der Ergebnisse automatischer Verfahren zur Beschriftung von Bildteilen wird die folgende Fragestellung aufgestellt:

Frage 3

Verbessert die Kombination der Ergebnisse automatischer Verfahren zur Beschriftung von Bildteilen das Gesamtergebnis?

Wissensbasis

Die Komponente Wissensbasis ermöglicht zum Einen die *Verwaltung der Wissensbasis* und zum Anderen den *Import / Export* von Ontologien.

Verwaltung Wissensbasis Entsprechend einer Antwort auf die formulierte

Frage 2 (b) muss diese Komponente geeignete Methoden anbieten, um neue semantische Elemente erstellen zu können. Weiterhin muss die Anzeige und Bearbeitung bestehender semantischer Elemente möglich sein.

Weiterhin soll diese Komponente auf Anfragen geeignetes Hintergrundwissen den Komponenten Bildsuchmaschine und Index zur Verfügung stellen. Hierzu werden sogenannte *Reasoner* eingesetzt [Gua98]. Ein Beispiel hierzu ist die Eingabe „Schauspieler“. Zu dieser Eingabe sollen dann die Namen aller passenden Schauspieler beziehungsweise die dazu gehörenden semantischen Elemente zurückgegeben werden.

Import / Export Diese Komponente dient zum Import und Export von Ontologien in ImageNotion. Mit dem Import können geeignete, bereits bestehende Ontologien in ImageNotion integriert und genutzt werden. Der Export von Ontologien soll es erlauben, diese mit anderen Anwendungen auszutauschen. So können diese durch spezielle Anwendungen zum Beispiel auf Konsistenz geprüft werden.

Web Service Schnittstelle

Die Web Service Schnittstelle soll es externen Anwendungen ermöglichen, die ImageNotion Anwendung als Dienst zu verwenden und so höherwertigere Anwendungen erstellen zu können.

3.4.3 Datenhaltung

Die Komponenten *Bilddaten* und *textbasierte Bildbeschriftungen* bleiben wie auch in einer textbasierten Bildsuchmaschine erhalten. Hinzu kommt die Speicherung der *semantischen Bildbeschriftungen* sowie die Speicherung der *Wissensbasis*.

3.5 Verwandte Arbeiten zu ImageNotion

Das vorgestellte ImageNotion System soll nun zum Abschluss dieses Kapitels mit bestehenden, verwandten Arbeiten verglichen werden. Damit lässt sich zeigen, wo in ImageNotion neue Ansätze erforderlich sind.

3.5.1 System von Vallet (2005)

Vallet [VFAC05] hat 2005 eine Architektur für ein semantisches Bildverwaltungssystem vorgestellt. Dieses basiert komplett auf semantischen Metadaten, die mittels

einer Taxonomie gebildet werden. Über einen Index werden die Metadaten indiziert und semantischen Elemente der Beschriftung gewichtet. Eine Beschriftung von Bildteilen ist nicht möglich. Die Wissensbasis besteht aus einer importierten Ontologie. Automatische Verfahren zur Erkennung von Motiven auf Bildteilen werden nicht eingesetzt.

3.5.2 MIA Demonstrator (2006)

Der MIA Demonstrator [Hol06] ermöglicht die Erstellung von semantischen Metadaten für Gemälde (deren digitale Abbilder im System). Mit dem MIA Demonstrator können auch semantische Beschriftung für Bildteile erstellt werden. Weiterhin werden Texterkenner eingesetzt, um gleichartige semantische Elemente aus verschiedenen Ontologien zusammenfügen zu können. Beide Prozesse wurden als Aufbereitungsprozesse durchgeführt, also außerhalb des Systems. Das System wurde also nur auf Basis semantischer Metadaten betrieben. Die Anfragen selbst wurden zur Laufzeit ausgewertet und eine Ordnung erstellt, so dass kein Index eingesetzt wurde. Innerhalb des MIA Demonstrators wurden automatische Verfahren zur Erkennung wichtiger struktureller Teile in einem Gemälde, wie zum Beispiel dem Hintergrund des Bildes oder wichtiger zusammengehörender Teile, eingesetzt und diese Bereiche markiert.

3.5.3 SemSpace (2007)

Das System SemSpace [vOTSP07] ist eine weitere semantische Bildsuchmaschine, die beim W3C als Demonstrator zum Aufbau einer semantischen Bildsuchmaschine gezeigt wird. SemSpace erlaubt den Import von Ontologien, mit denen anschließend semantische Metadaten für Bildteile und das ganze Bild erstellt werden können. Bei generellen Suchanfragen (nach Konzepten) werden im Suchergebnis die möglichen semantischen Elemente zur Konkretisierung einer Anfrage vorgeschlagen. Automatische Verfahren wurden keine integriert.

3.5.4 ALIPR (2007)

Das ALIPR System [LW06] ist ein textbasiertes Bildverwaltungssystem. In diesem System wird ein Objekterkenner eingesetzt. Der Objekterkenner ist trainierbar und soll so langfristig die Erkennung einer Vielzahl verschiedener Objekte lernen. ALIPR ermöglicht die Erweiterung von Suchanfragen um Synonyme. Die Suchanfragen werden mit Hilfe eines textbasierten Indexes ermittelt. Semantische Verfahren zur Anfrageverfeinerung und Erstellung von Metadaten werden nicht eingesetzt.

3.5.5 Gegenüberstellung der Arbeiten mit ImageNotion (Konzeption)

Tabelle 3.1 vergleicht die Komponenten bestehender Arbeiten mit ImageNotion. Die ersten drei Arbeiten setzen komplett auf semantische Metadaten auf. Hingegen setzt ALIPR komplett auf textbasierten Metadaten auf. Man sieht, dass ImageNotion aufgrund der Ergebnisse der Befragung von Geschäftsführern von Bildagenturen als neuen Ansatz die hybride Verwaltung von textbasierten und semantischen Metadaten vorsieht. Entsprechend sind nur bei ImageNotion auch beide Verfahren zur Anfrageerweiterung integriert sowie beide Möglichkeiten zur Bildbeschriftung.

Die deutlichsten Neuerungen sind weiterhin in ImageNotion die Ziele, geeignete Verfahren zur Erweiterbarkeit von Ontologien durch Benutzer zu ermöglichen und eine hohe Qualität von Algorithmen zur Erkennung von Bildteilen zu erhalten durch die geeignete Kombination der von diesen erzeugten Ergebnissen.

Ebenso generiert nur ImageNotion automatische Vorschläge von Suchanfragen mittels Autocomplete. Allerdings soll dies an dieser Stelle nur als Randnotiz erwähnt werden, da deren Einfluss auf eine bessere Qualität von Suchergebnissen nicht weiter in dieser Arbeit untersucht wird.

Die Gründe für die mit Stern markierten Komponenten in der geplanten ImageNotion Systemarchitektur in Abbildung 3.6 sind also jetzt ersichtlich aufgrund der erstellten Fragestellungen und dem Vergleich der ImageNotion Systemarchitektur mit bestehenden Arbeiten in Tabelle 3.1. Die genaue Umsetzung dieser Komponenten in ImageNotion soll mit Hilfe von empirischen Untersuchungen ermittelt werden, um eine hohe Einsatzbarkeit der gefundenen Lösungen gewährleisten zu können.

3.6 Zusammenfassung

Innerhalb dieses Kapitels wurde die ImageNotion Anwendung aufgebaut durch die Konzeption einer geeigneten Architektur und die dort nötigen Komponenten.

Ausgehend von der Architektur einer textbasierten Bildsuchmaschine wurde in diesem Kapitel untersucht, an welchen Stellen semantische Verfahren zur Verbesserung der Bildsuche integriert werden können. Von den möglichen Stellen Anfrageerweiterung und Erstellung von semantischen Metadaten ist zwar unter Laborbedingung die komplette Nutzung des zweiten Verfahrens die bessere Lösung, da diese zu den in dieser Arbeit gewünschten, hochwertigen Suchergebnissen führen.

Die Befragung und Diskussion von Geschäftsführern von Bildagenturen hat

System	Vallet	MIA	Sem Space	ALIPR	ImageNotion (Konzeption)
Bildsuche					
Lexikalische Anfrageerweiterung	-	-	-	x	x
Semantische Anfrageerweiterung	x	x	x	-	x
Textbasierte Bildsuche	-	-	-	x	x
Semantische Bildsuche	x	x	x	-	x
Index					
Index textuelle Beschriftungen	-	-	-	x	x
Index semantische Beschriftungen	x	-	-	-	x
Index Autocomplete	-	-	-	-	x
Bildbeschriftung					
Textbasierte Bildbeschriftung	-	-	-	x	x
Semantische Bildbeschriftung	x	x	x	-	x
Beschriftung Bildteile	-	x	x	x	x
Verwendung Texterkenner	-	x	-	-	x
Automatische Erkennung Bildteile	-	x	-	x	x
Kombination automatischer Verfahren	-	-	-	-	x
Wissensbasis					
Verwaltung Wissensbasis	x	x	x	-	x
Erweiterbarkeit Wissensbasis durch Benutzer	-	-	-	-	x
Import /Export	x	x	x	-	x

Tabelle 3.1: Vergleich verwandter Arbeiten mit ImageNotion

<p>Frage 1: Wie sollen semantische Verfahren eingesetzt werden?</p> <p>Frage 1 (a): An welcher Stelle in einer Bildsuchmaschine sollen semantische Verfahren eingesetzt werden.</p> <p><i>Antwort:</i> Sowohl in Anfrageerweiterung wie der Erstellung semantischer Metadaten. ImageNotion wird als hybrides System erstellt, das den kontinuierlichen Übergang von textuellen zu semantischen Metadaten ermöglicht.</p> <p>Frage 1 (b): Werden semantische Verfahren zur Verbesserung der Bildsuche benötigt?</p> <p>Frage 1 (c): Sind die entstehenden Möglichkeiten zur Bildsuche für Nutzer in der Praxis verständlich und wie sollen diese eingesetzt werden?</p>
<p>Frage 2: Wie und von wem werden semantische Hintergrundinformationen erstellt?</p> <p>Frage 2 (a): Sind getrennt erstellte, vorgefertigte Ontologien ausreichend und praktisch anwendbar?</p> <p>Frage 2 (b): Wie können Bildredakteure in diesen Prozess eingebunden werden und selbst Hintergrundwissen erstellen?</p> <p>Frage 2 (c): Reduzieren semantische Verfahren den Zeitaufwand zur Erstellung von Metadaten?</p>
<p>Frage 3: Verbessert die Kombination der Ergebnisse automatischer Verfahren zur Beschriftung von Bildteilen das Gesamtergebnis?</p>

Tabelle 3.2: Übersicht der resultierenden Fragestellungen in ImageNotion

aber gezeigt, dass dieses Vorgehen in der Praxis nicht möglich ist. Der Hauptgrund hierbei ist, dass große Bildbestände komplett neu mit semantischen Metadaten versehen werden müssten. Dies ist zu teuer und somit nicht umsetzbar. Speziell diese Tatsache führte zur grundlegenden Entwurfsentscheidung für den Aufbau von ImageNotion als hybrides System. Es soll den kontinuierlichen Übergang in einem Bildarchiv von textbasierten hin zu semantischen Metadaten erfolgen. Die Bildagenturen können so entweder zuerst nur Verfahren zur Anfrageerweiterung verwenden. Weiterhin können diese beginnen, Teile des Bildarchivs oder neue Bilder mit semantischen Metadaten zu versehen. So kann eine kontinuierliche Verbesserung der Suchergebnisse hin zu hochwertigen Ergebnissen erfolgen.

Tabelle 3.2 gibt noch einmal einen Überblick über die Fragestellungen, die sich daraus ergeben haben und die in den weiteren Kapiteln dieser Arbeit entsprechend beantwortet werden.

Kapitel 4

Kategorisierung von Suchanfragen

Die folgenden zwei Kapitel dienen zur Beantwortung der zu klärenden Frage 1 (b) in dieser Arbeit *Werden semantische Verfahren zur Verbesserung der Bildsuche benötigt?* Zur Beantwortung dieser Frage werden Suchanfragen an Bildagenturen ausgewertet und geeignet kategorisiert.

Durch die Kategorisierung von Suchanfragen soll ermittelt werden, welche *Abfolgen von Suchanfragen* durchgeführt werden und welche *Art sowie Bezug von Suchanfragen* diese haben. Mit den *Abfolgen von Suchanfragen* wird ersichtlich, inwieweit tatsächlich Anfrageerweiterungen sinnvoll sind (siehe Sektion 3.3.1), um die Abfolgen an Suchanfragen verringern zu können. Durch die *Art sowie Bezug von Suchanfragen* kann ermittelt werden, ob die Suchanfragen überwiegend spezifisch oder eher generell sind und wie oft sich die Anfragen auf Bildteile beziehen. Werden viele generelle Suchanfragen gestellt, bedeutet dies, dass Verfahren zur semantischen Anfrageerweiterung benötigt werden, um die Suchanfragen interpretieren zu können (siehe Sektion 3.3.1). Aus den so gewonnenen Ergebnissen kann mit einer geeigneten Auswertung von Suchanfragen die Frage 1 (b) entsprechend beantwortet werden.

Ziel dieses Kapitels ist, für eine solche Analyse zuerst die geeigneten Möglichkeiten zur Kategorisierung von Suchanfragen nach *Abfolgen* und *Art sowie Bezug* von Suchanfragen zu finden.

4.1 Motivation

Falls an eine Bildsuchmaschine überwiegend spezifischen Anfragen gestellt¹ werden (z.B. Personennamen, Produktnamen, Städtenamen), so ist die Bearbeitung der An-

¹Bei Bildarchiven wie ddp oder dpa ist dies der Fall, dort suchen die Benutzer häufig nach Namen der Personen des Tagesgeschehens oder nach Datum des Tages.

frage durch textbasierte Bildsuchmaschinen einfach und direkt möglich. Die Eingabe wird auf den Index abgebildet und führt zu präzisen Suchergebnissen². Die Länge einzelner Sitzungen ist dann recht kurz, denn einzelne Suchergebnisse sind bereits sehr hochwertig³. Es wären dann keine weiteren gravierenden Maßnahmen nötig, um die Bildsuche zu verbessern.

Müssen häufig lexikalische Phänomene wie Synonyme oder unterschiedliche Sprachen verwendet werden oder werden häufig auch generelle Anfragen gestellt⁴ (z.B. „Deutsche Politiker“), so ist das Suchergebnis dann von einer hohen Qualität, wenn alle Personen im Ergebnis sind, die im Bildarchiv gespeichert sind und Politiker aus Deutschland sind oder waren. Wie im vorigen Kapitel gezeigt, erfordern diese Suchanfragen den Einsatz semantischer Verfahren, um die *Suchergebnisse einzelner Suchanfragen verbessern* und so die *Länge einer Sitzung verkürzen* zu können.

Das Ziel der folgenden Analyse ist daher, *Abfolgen von Sitzungen* und die *Art sowie den Bezug von Suchanfragen* zu untersuchen.

4.2 Fragestellungen an ein geeignetes Modell

Durch eine geeignete Kategorisierung sollen Suchanfragen ausgewertet werden, um die *Frage 1 (b) Werden semantische Verfahren zur Verbesserung der Bildsuche benötigt?* zu beantworten.

Laut den im vorigen Kapitel genannten möglichen semantischen Verfahren (siehe Sektion 3.3) entstehend die folgenden Teilfragen:

- *(F1) Sind lexikalische Anfrageerweiterungen notwendig?* Lexikalische Anfrageerweiterungen sind dann notwendig, wenn in Sitzungen häufig unterschiedliche Synonyme oder Sprachen verwendet werden. Lexikalische Verfahren können diese Abfolgen von Suchanfragen verkürzen und so die benötigten Anfragen in einer Sitzung verringern. Dadurch werden gewünschte Bilder schneller gefunden.
- *(F2) Sind semantische Anfrageerweiterungen notwendig?* Semantische Anfrageerweiterungen sind dann notwendig, wenn häufig generelle oder imaginäre Suchanfragen gestellt werden. Dann hilft die semantische Anfrageerweiterung dabei, diese Anfragen richtig aufzulösen und umzusetzen. Diese Verfahren reduzieren demnach ebenfalls für einen Bildsucher die Anzahl an Suchanfragen

²Vollständig nur, falls alle Metadaten geeigneter Bilder den Text der Suchanfrage enthalten

³(zumindest inhaltlich, die Qualität der Bilder selbst ist hier nicht gemeint)

⁴Entsprechende Anfragen stellen zum Beispiel Bildredakteure in Verlagen oder Werbeagenturen.

und die damit nötige Zeit, da direkter die gewünschten Ergebnisse angezeigt werden.

- (F3) *Sind Suchanfragen miteinander verwandt?* Falls die Suchanfragen miteinander verwandt sind, z.B. zuerst „Frosch“, dann „Kröte“, sind der Einsatz von Verfahren zur Neuformulierung von Suchanfragen hilfreich. Im eben genannten Beispiel könnten dann weitere Arten von Froschlurchen vorgeschlagen werden.
- (F4) *Sind Beschriftungen für Bildteile notwendig?* Werden häufig Suchanfragen gestellt, die sich direkt auf Bildmotive beziehen, verbessern beschriftete Bildteile die Präzision der Ergebnisse speziell auf den ersten Rängen eines Suchergebnisses. Diese zeigen dann direkt Bilder mit dem gewünschten Motiv.

Innerhalb der Analyse soll zusätzlich ermittelt werden, nach welchen Bildmotiven häufig gesucht wird. Dies zeigt, welche Verfahren zur automatischen Erkennung von Bildteilen benötigt werden. Falls zum Beispiel häufig nach Personen gesucht wird, die innerhalb der Bilder abgebildet sein sollen, bietet sich der Einsatz von Personen- und Gesichtserkennern an.

Frage 1 (b) kann dann mit den folgenden Extremen beantwortet werden: wenn alle vier Fragen mit nein beantwortet werden müssen, so sind semantische Verfahren in keiner Weise notwendig. Wenn alle vier Fragen mit ja beantwortet werden können, sind alle möglichen Verfahren zwingend notwendig, um hochwertige Suchergebnisse schaffen zu können.

Zur Kategorisierung von F1 und F3 können die *Abfolgen einer Suchanfragen* verwendet und nach *Synonym, unterschiedlichen Sprachen, konkretere Anfrage* und *verwandte Anfrage* kategorisiert werden.

Im Folgenden wird zusätzlich nach einem Modell gesucht, welches Suchanfragen bezüglich der Fragen F2 und F4 beantworten kann. Hierzu müssen die *Art der Suchanfrage* zur Beantwortung der Frage F2 und der *Bezug der Anfrage* auf ein Bildmotiv oder das ganze Bild (für Frage F4) kategorisiert werden können.

4.3 Kategorisierung von Abfolgen einer Suchanfrage in Sitzungen

Tabelle 4.1 gibt einen Überblick über die Kategorisierung von Abfolgen in einer Suchanfrage, die in der Analyse verwendet werden soll. Die gezeigten Kategorien ergaben sich bei der Analyse der Suchanfragen im nächsten Kapitel.

Art der Suchanfrage	Beschreibung
Sprache	Stellen der Suchanfrage in einer anderen Sprache
Synonym	Stellen der Suchanfrage mit einem Synonym
Konkretere Anfrage	Anfrage wird konkreter gestellt
Verwandte Anfrage	Stellen von ähnlichen Anfragen

Tabelle 4.1: Kategorisierung der Abfolgen von Suchanfragen in Sitzungen

Die erste Möglichkeit zur Neuformulierung einer Suchanfrage ist dabei das Verwenden einer anderen Landessprache für die Suchanfrage. So kann zum Beispiel die Anfrage⁵ „Buchkreuzer“ zu „Bookcrosser“ umformuliert werden.

Die Verwendung von *Synonymen* bilden die zweite Möglichkeit zur Neuformulierung einer gestellten Suchanfrage. Die Anfrage „Bus von Innen“ wurde neu formuliert mit „Innenansicht Bus“.

Bei generellen Suchanfragen wurden *konkretere Anfragen* gestellt. Die Anfrage „Herz“⁶ kann zum Beispiel konkretisiert werden zu „Herz Organ“.

Schließlich ist die vierte Möglichkeit zur Neuformulierung einer Suchanfrage, eine *verwandte Anfrage* zu verwenden. So passt zur Anfrage „Aitutaki“ der Vorschlag zur Suche nach „Bora Bora“. Beide Anfragen sind verwandt zueinander aber keine Synonyme, denn es handelt sich um verschiedene Atolle.

4.4 Bestehende Modelle zur Beschreibung von Bildinhalten

Innerhalb der folgenden Modelle meint „beschreiben“ die mündliche Wiedergabe durch Benutzer, was auf einem Bild zu sehen ist. So ist beschreibbar, wie Benutzer Bilder sehen. Entsprechend können diese Modelle zur Kategorisierung von Bildbeschriftungen verwendet werden (z.B. in [Hol06]). Aus der Literatur die gängigen Modelle sind hierbei das *Modell von Shatford* und das *10-Ebenenmodell von Jaimés und Chang*.

⁵Alle folgenden Beispiele wurden in fotomarktplatz.de gestellt innerhalb der analysierten Suchanfragen im nächsten Kapitel.

⁶„Herz“ kann als Ergebnis unter anderem Symbole, Bilder der Aktion „Ein Herz für Kinder“ und das Organ Herz liefern.

4.4.1 Das Modell von Shatford

Das Modell von Shatford [Sha86] dient zur Kategorisierung der Bestandteile eines Bilds mit den Fragen „Wer ist auf dem Bild“ und „Was ist auf dem Bild“. Weiterhin soll mit den Fragen „Wo entstand das Bild“ und „Wann entstand das Bild“ beschrieben werden, an welchem Ort und zu welcher Zeit das Bild entstand. Die Fragestellungen werden im Modell drei Abstraktionsebenen unterteilt. Die Ebenen sollen beschreiben, was einzelne Bildteile (bzw. das ganze Bild) *allgemein* (z.B. „ein Bildteil zeigt ein Auto“), *spezifisch* (z.B. „ein Bildteil zeigt einen Mercedes SL 300, Baujahr 1968) beziehungsweise *imaginär* (z.B. „ein Bildteil zeigt einen Oldtimer“) aussagt.

4.4.2 Das 10-Ebenenmodell von Jaimes und Chang

Das 10-Ebenenmodell von Jaimes und Chang [JJfC00] ist eine Verfeinerung des Modells von Shatford. Es dient der Ermittlung von Kriterien zur Indizierung von Bildbeschriftungen. Jaimes und Chang trennen hierbei nach Beschreibungen, die sich auf Bildteile beziehungsweise auf das ganze Bild beziehen.

Die Beschreibungen für das Bild werden unterteilt in den *perzeptiven* und den *visuellen* Bereich. Der *perzeptive Bereich* betrifft die Gesamtwahrnehmung des Bildes, also ob es sich um ein Farbbild oder Schwarz / Weiß- Bild handelt, die dominanten Farben und die dominanten Bereiche in einem Bild.

Der *visuelle Bereich* betrifft wichtige Bildteile und den Gesamtzusammenhang des Bildes. Entsprechend dem Modell von Shatford werden ebenfalls die Abstraktionsebenen *allgemein*, *spezifisch* und *imaginär* verwendet, um den visuellen Bereich zu beschreiben. Die allgemeine Ebene für ein ganzes Bild könnte zum Beispiel „Touristenattraktion in Pisa“ lauten, wenn das Bild spezifisch den „Schiefen Turm von Pisa“ zeigt. Eine entsprechende Interpretation des Bildteils innerhalb der imaginären Ebene könnte dann „Urlaub in der Toskana“ sein.

Das 10-Ebenenmodell von Jaimes und Chang soll in der folgenden Analyse von Suchanfragen als Modell verwendet werden, denn es ermöglicht die gewünschte Einteilung von Suchanfragen nach unterschiedlichen Abstraktionsebenen und weiterhin die Trennung in Suchanfragen, die Bildteile oder das Bild als Ganzes betreffen.

4.5 Kategorisierung der Art und dem Bezug von Suchanfragen

Jaimes und Chang entwickelten ihr Modell mit der Frage nach der geeigneten Indizierung von Beschriftungen mit der Fragestellung „Welche Ebene des Modells soll wie indiziert werden?“ Da in der vorliegenden Arbeit mit der geplanten Analyse aber Art und Bezug einer Suchanfrage von Interesse sind mit der Fragestellung „Wie können Suchanfragen nach dem Modell kategorisiert werden?“, muss es entsprechend angepasst werden.

4.5.1 Überblick über die durchgeführten Anpassungen

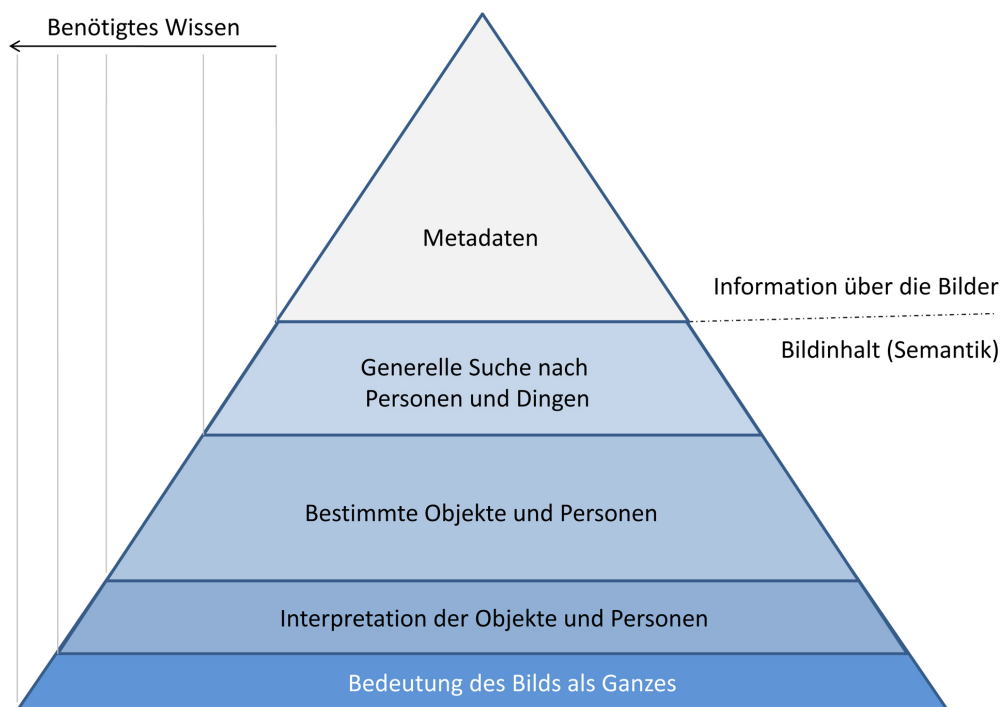


Abbildung 4.1: Angepasstes Modell zur Kategorisierung von Suchanfragen

Abbildung 4.1 zeigt das angepasste Modell zur Kategorisierung der Art und dem Bezug von Suchanfragen. Die folgenden Änderungen im Modell zur geeigneten Kategorisierung von Suchanfragen wurden durchgeführt:

Verzicht auf den perceptiven Bereich Eine Anpassungen ist der Verzicht auf den perceptiven Bereich. Bei Jaimes und Chang dient dieser zur Indizierung

von Bildern mit ähnlicher Form und Farbe. Diese Anfragen können nicht durch textbasierte Suchanfragen ausgedrückt werden. So entfallen vier Ebenen aus dem ursprünglichen Modell.

Unterteilung nach Personen und Dingen Zusätzlich ist die Unterteilung für Anfragen nach Bildteilen nach Personen und Dingen als Anpassung vorgenommen worden. Dies ermöglicht die spätere Auswahl geeigneter Verfahren zur automatischen Erkennung von Bildteilen in dieser Arbeit.

Metadaten über die Bilder Standards zur Erstellung von Metadaten (z.B. IPTC) bieten häufig Felder an, um Metadaten über die Bilder wie Name des Fotografen oder der Bildagentur, die Bildnummer oder das Aufnahmedatum einzugeben. Nach diesen Metadaten über die Bilder kann ebenfalls gesucht werden. Daher wird die Kategorie „Metadaten“ dem Modell hinzugefügt.

Abbildung 4.1 deutet weiterhin mit der jeweils gezeigten Höhe der senkrechten Linie an, dass in Anlehnung an die Jaimes und Jang ein Bildsucher vom Umfang her ansteigendem Wissen über Sachverhalt benötigt, um spezifische und imaginäre Suchanfragen formulieren zu können. Gerade hier können semantische Verfahren hilfreich sein, z.B. beim Start mit der Suchanfrage „Schauspieler“, um anschließend die spezifischen Namen von Schauspielern vorgeschlagen zu bekommen.

4.5.2 Kategorien des Modells

Tabelle 4.2 zeigt eine Übersicht über die entstandenen Kategorien des Modells und die dort einzuordnenden Suchanfragen. Im Folgenden werden die einzelnen Kategorien genauer beschrieben.

Kategorie 1: Metadaten über die Bilder

Die Kategorie der *Metadaten über die Bilder* (im Folgenden kurz als „Metadaten“ bezeichnet) deckt Anfragen ab, die sich auf die Suche nach Metadaten zu einem Bild beziehen. Die wichtigsten Felder des Standards IPTC⁷ werden hierbei verwendet. Dies betrifft den Aufnahmeort, ein bestimmtes Datum oder Zeitintervall sowie die Bildnummer. Weiterhin sind Suchanfragen nach Bildern einer bestimmten Bildagentur möglich.

⁷Dieser wird fast überwiegend von Bildagenturen eingesetzt

Metadaten		
Suche in Metadaten	Aufnahmeort	Ort der Aufnahme, z.B. „Baden-Baden“
	Zeit, Zeitintervall	Bilder von einem bestimmten Datum oder Intervall
	Bildnummer	Suche eines Bildes unter Verwendung der Bildnummer, z.B. „P1243de“
	Bildagentur	Suche nach Bilder einer bestimmten Bildagentur, z.B. „ActionPress“
Suche nach Bildinhalt: Was ist auf dem Bild abgebildet?		
Generelle Anfragen		
Generelle Suche nach Personen und Dingen	Person (generell)	Suche nach einer Art von Person, z.B. „Frau“, „Mädchen“, „Schauspieler“
	Dinge (generell)	Suche nach Dingen einer bestimmten Art, z.B. „Auto“
Spezifische Anfragen		
Suche nach bestimmten Personen oder Dingen	Person (spezifisch)	Suche nach einer bestimmten Person, z.B. „Elizabeth Taylor“
	Dinge (spezifisch)	Suche nach bestimmten Dingen, z.B. „Mercedes SL 300“
Imaginäre Anfragen		
Was repräsentiert ein Objekt oder eine Person	Person (imaginär)	Suche nach der Bedeutung einer Person, z.B. „Bundeskanzler“
	Dinge (imaginär)	Suche nach der Bedeutung von Dingen, z.B. „Oldtimer“
Anfragen das ganze Bild betreffend		
Was ist die Bedeutung des ganzen Bilds	Thema (Gesamtszene)	Welche Aussage hat das ganze Bild, z.B. „Weihnachten“

Tabelle 4.2: Kategorien für die Art und dem Bezug von Suchanfragen

Suche nach Bildinhalt (Kategorien 2 -4)

Mit Suchanfragen zum Bildinhalt wird nach auf dem Bild dargestellten und somit sichtbaren Dingen gesucht.

Kategorie 2: generelle Anfragen: Die generellsten Suchanfragen betreffen hierbei die den Bereich der *Generelle Anfragen*. Für Personen sind dies Suchanfragen zur Art der Person, zum Beispiel „Gruppe von Jugendlichen“. Neben den Personen können dies im Bereich der Objekte Suchanfragen nach Dingen einer bestimmten Art sein, zum Beispiel „Blumen“.

Kategorie 3: spezifische Anfragen: Der Bereich der *Spezifische Anfragen* deckt spezifische Anfragen zu Personen und Dingen ab. Dies kann eine bestimmte Person sein, also zum Beispiel „Helmut Kohl“, oder ein bestimmter Gegenstand, zum Beispiel bestimmte Blumen wie „rote Rosen“. Es wird hierbei also Wissen um Personennamen oder bestimmte Bezeichner von Dingen benötigt.

Kategorie 4: imaginäre Anfragen: Der Bereich der *Imaginäre Anfragen* sind schließlich Anfragen zur Interpretation von abgebildeten Objekten. Im Bereich von Personen kann die Bedeutung zum Beispiel der Beruf einer Person sein, also bei Helmut Kohl „Altkanzler“ bzw. im Jahr 1989 „Bundeskanzler“, und bei Dingen deren mögliche Bedeutung als Symbol, also bei „roten Rosen“ wäre dies „Symbole zum Ausdruck der Liebe“. In diesem Bereich wird also zusätzliches Hintergrundwissen benötigt, um die gezeigten Personen und Dinge interpretieren zu können.

Kategorie 5: Suchanfragen das ganze Bild betreffend

Suchanfragen, die das ganze Bild betreffen, zielen auf eine gewünschte Gesamtszene oder das Thema eines Bilds ab. Mögliche Gesamtszenen sind auch die Interpretation oder Aussage des Bilds. Ein „verliebttes Paar“ ist eine mögliche Interpretation eines Bilds, auf dem sich ein Mann und eine Frau umarmen. Suchanfragen zu Themen bilden Szenen ab, die ein Thema abbilden, zum Beispiel Bilder eines Kinofilms, eines politischen Ereignisses oder Themen des Zeitgeschehens wie „Ostern“ oder „Weihnachten“.

Suchanfragen das ganze Bild betreffend könnten somit ebenfalls noch einmal nach konkret, imaginär und spezifisch unterteilt werden.

4.6 Zusammenfassung

Die in diesem Kapitel erstellten Kategorisierungen von Suchanfragen nach *Abfolgen* und *Art und Bezug* von Suchanfragen ermöglichen wie gewünscht zur Beantwortung

der in Sektion 4.2 formulierten Fragestellungen die entsprechende Kategorisierung von Suchanfragen.

Mit Hilfe dieser Kategorisierungen werden daher im nächsten Kapitel Suchanfragen an die Suchmaschine fotomarktplatz.de ausgewertet.

Kapitel 5

Analyse von Suchanfragen an fotomarktplatz.de

Mit den erstellten Möglichkeiten zur *Kategorisierung von Abfolgen einer Suchanfrage* in Sitzungen und *Kategorisierung der Art und dem Bezug von Suchanfragen* aus dem letzten Kapitel sollen nun die zu beantwortenden Fragestellungen aus Sektion 4.2 beantwortet werden.

Hierzu dient eine Analyse von Suchanfragen aus fotomarktplatz.de. Mehr als 100.000 Suchanfragen bilden die Datenbasis für die Analyse. Die Suchanfragen wurden von mehr als 6.000 registrierten Benutzern der Bildsuchmaschine gestellt. Dies lässt eine umfangreiche Analyse der Verwendung von Bildsuchmaschinen und ihren Suchanfragen zu.

Durch die Auswertung der ermittelten Daten und Bewertung der Ergebnisse kann somit durch Analyse von Suchanfragen die Frage beantwortet werden, *ob zur Verbesserung der Bildsuche semantische Verfahren benötigt werden.*

5.1 Ergebnisse bisheriger Untersuchungen

Mit dem Fokus dieser Arbeit auf die Bildsuche im Umfeld von Bildagenturen wurden entsprechende bisherige Untersuchungen in der Literatur ermittelt. Die folgenden zwei Arbeiten liefern hierbei die relevantesten Auswertungen.

Anzumerken ist, dass der Zugriff auf die Suchanfragen von professionellen Bildanbietern häufig schwierig ist aufgrund von wirtschaftlichen Interessen der Betreiber („Die Konkurrenz soll das nicht sehen“). Die folgenden Untersuchungen konnten nur aufgrund direkter Kontakte zu den Betreibern der Bildsuchmaschinen her-

gestellt werden. Das Gleiche gilt im Fall der eigenen Untersuchung in fotomarktplatz.de. Die große Menge an Suchanfragen konnte nur erhalten werden, da ich als Technischer Leiter von fotomarktplatz.de Zugriff auf die Log-Daten hatte und vor der Auswertung die Ergebnisse entsprechend der Anforderungen der Betreiber anonymisiert habe.

5.1.1 Anfragen an die Bildsuchmaschine einer Bildagentur

Ziel der Untersuchung von Jörgensen 2005 [JJ05] war die Auswertung von Sitzungen von Benutzern und die dabei gestellten Suchanfragen an eine Bildagentur. Der Fokus lag hierbei auf der Art und Weise, wie Sitzungen verlaufen und wie lange Sitzungen insgesamt dauern. Hierzu wurden 400 Sitzungen beobachtet und dabei insgesamt 1.556 Suchanfragen ausgewertet.

Als Ergebnis wurde festgestellt, dass pro Sitzung durchschnittlich etwa vier Anfragen gestellt wurden. Bei 56 Prozent der Suchanfragen wurde nur ein Suchwort verwendet und bei etwa 24 Prozent der Suchanfragen zwei Suchwörter. Die meisten Suchanfragen werden somit mit weniger als drei Begriffen in der Anfrage gestartet. Innerhalb einer Sitzung wurden 26 Prozent der Anfragen durch Hinzufügen von Wörtern eingeschränkt, bei 10 Prozent der Sitzungen wurden Suchwörter entfernt. Der überwiegende Teil von 64 Prozent entfiel auf das Stellen einer komplett neuen Anfrage.

5.1.2 Anfragen durch Bildredakteure in einem Zeitungsverlag

Die Untersuchung von Markkula and Sormunen aus dem Jahr 2002 [MS00] hat sich mit dem Suchverhalten von Bildredakteuren in einem finnischen Zeitungsverlag beschäftigt.

Bei der Studie wurden 27 Sitzungen von Bildredakteuren und deren 108 Suchanfragen kategorisiert. Die Kategorisierung hat sich am Modell von Shatford orientiert und wurde auf die Suchanfragen in einem Zeitungsverlag angepasst. Die Anfragen wurden unterschieden nach *allgemein*, *spezifisch* und *imaginär*. Die Untersuchung hat die Anfragen nicht nach dem Bezug auf Bildteile und dem ganzen Bild unterschieden.

Das Ergebnis der Kategorisierung war, dass 54 Prozent der insgesamt 108 Anfragen spezifische Anfragen waren und weiterhin sich 29 Prozent auf generelle Objekte wie zum Beispiel „Fahrzeug“ bezogen oder imaginäre Anfragen waren. Es wurden somit sowohl *spezifische*, *generelle* und *imaginäre* Suchanfragen verwendet.



Abbildung 5.1: Die Metasuchmaschine fotomarktplatz.de

Bei der Untersuchung wurde ebenso wie bei Jörgensen weiterhin beobachtet, dass die Bildredakteure oft Anfragen mit wenigen Stichwörtern gestellt haben und bewusst ein großes Suchergebnis in Kauf nahmen. Im Durchschnitt wurden so pro Suchanfrage 8.9 Suchergebnisseiten aufgerufen. Danach wurden die einzelnen Suchergebnisseiten gesichtet und die relevanten Bilder daraus ermittelt. Die Begründung war, dass „ihrer Erfahrung nach oft relevante Suchwörter in den Beschriftungen fehlen. Dadurch konnte nur durch das Browsen vieler Anfragen die gewünschten Bilder gefunden werden“.

Pro Sitzung wurden im Durchschnitt vier Suchanfragen gestellt. 57 Prozent dieser Anfragen wurden weiter eingeschränkt durch das Hinzufügen von Stichwörtern.

5.1.3 Gründe für die eigene Untersuchung

Mit den eben genannten Untersuchungen lassen sich die Fragestellungen aus Sektion 4.2 nicht hinreichend beantworten. Zum Einen, weil die Anzahl an Suchanfragen mit 100 recht gering ist und so die statistische Relevanz zu gering ist für allgemeinere Folgerungen aus der Analyse. Weiterhin, weil nicht untersucht wurde, ob sich die Suchanfragen auf Bildteile oder das ganze Bild beziehen. Auch ist die genaue Art der Abfolge von Suchanfragen nicht ersichtlich.

Ein weiter Grund ist, dass die Untersuchung von Markkula und Sormunen in einem Zeitungsverlag durchgeführt wurde. Aus eigenen Erfahrungen aus fotomarkt-

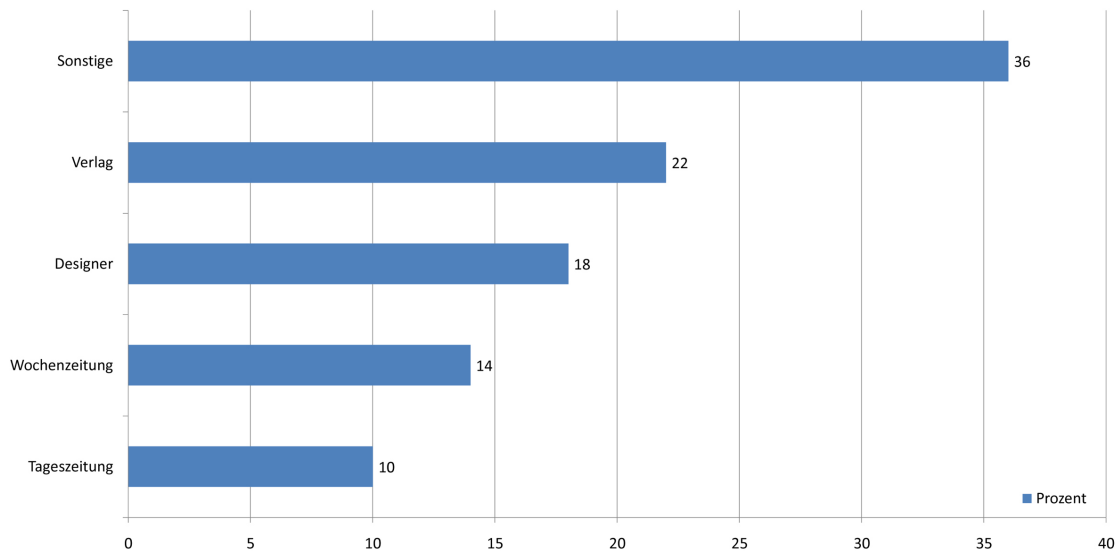


Abbildung 5.2: Verteilung der Benutzer in fotomarktplatz.de nach Bildnutzung

platz.de suchen die Redakteure dort tatsächlich (wie die Untersuchung zeigte) sehr spezifisch. Es wird meist nur nach Datum (Bilder von heute) gesucht oder meist sehr spezifisch (z.B. Namen von Politikern). Eine Einteilung nach der Art der Bildsuche ist also in Zeitungsverlagen wenig zielführend, mit anderen Worten, dort besteht tatsächlich ein sehr geringer Bedarf, die Bildsuche zu verbessern.

Um aus einer Analyse daher die Fragestellung geeignet beantworten zu können, ist der Ansatz der eigenen Untersuchung, eine große Datenbasis von mehr als 100.000 Suchanfragen zu verwenden und diese auszuwerten. Weiterhin sollten Suchanfragen von verschiedenen Benutzergruppen mit unterschiedlichen Aspekten betrachtet werden. Dies ermöglicht die allgemeinere Untersuchung des Suchverhaltens, also nicht nur wie bei Markkula fokussiert auf die Belange der Recherche in einem Zeitungsverlag. Aus diesen Gründen scheint die Verwendung von fotomarktplatz.de gut geeignet.

5.2 Die Metasuchmaschine fotomarktplatz.de

Ziel von fotomarktplatz.de (siehe Abbildung 5.1) war die parallele Abfrage der Bildarchive von professionellen Bildarchiven und Fotografen aus dem deutschsprachigen Raum¹.

Hierzu können diese Bildanbieter entweder die Bilddaten in die zentrale Da-

¹Der Betrieb der Suchmaschine selbst wurde im April 2009 eingestellt; ein direkt vergleichbares Projekt kann unter www.fotofinder.net gefunden werden

tenbank von fotomarktplatz.de aufladen. Als weitere Möglichkeit wird die Abfrage der von den Bildanbietern betriebenen Bildsuchmaschinen auf deren Webseiten über eine Serviceschnittstelle angeboten.

5.2.1 Bildsuche mit fotomarktplatz.de

Durch die Abfrage der Bildbestände von mehr als hundert Bildanbietern (Stand aus dem Jahr 2006) ist für einen Benutzer durch die Nutzung eines zentralen Portals die parallele Bildsuche in einem Bildbestand von über zehn Millionen Bilder möglich. Hierzu hat der Benutzer eine Suchanfrage in fotomarktplatz.de eingegeben und erhielt anschließend ein über alle angebotenen Bildagenturen zusammengefasstes und aufbereitetes Suchergebnis.

5.2.2 Benutzergruppen in fotomarktplatz.de

Die Suchmaschine fotomarktplatz.de wird überwiegend durch kommerzielle und geübte Benutzer verwendet. Innerhalb von fotomarktplatz.de ist die Suche als registrierter Nutzer sowie als nicht registrierter Nutzer möglich. In der Analyse soll das Hauptaugenmerk auf Suchanfragen von registrierten Benutzern liegen. Zum Zeitpunkt der Messung gab es 6.317 registrierte Benutzer in fotomarktplatz.de. Wie Abbildung 5.2 zeigt, sind 22 Prozent der Nutzer Verlagsmitarbeiter, die beispielsweise Bücher oder Magazine illustrieren. 18 Prozent der Nutzer sind Gestalter, die Bilder überwiegend für kreative Zwecke wie Werbematerialien oder Anzeigen nutzen. Weiterhin waren 14 Prozent der Nutzer Redakteure einer Wochenzeitung und 10 Prozent der Benutzer Redakteure in einer Tageszeitung. Es gibt somit in fotomarktplatz.de wie gewünscht verschiedene Benutzergruppen, welche jeweils unterschiedliche Verwendungen mit gefundenen Bildern planen.

5.3 Analyse 2: Bildsuche in fotomarktplatz.de

Als Datenbasis der Studie dienten 134.943 Suchanfragen, die im Zeitraum vom 1. Februar 2006 - 31. März 2006 in fotomarktplatz.de mit protokolliert worden sind. Über eine Kennung für jeden eindeutigen Benutzer konnten in den Protokolldateien einzelne Suchanfragen zu Sitzungen zusammengefasst werden.

Bei der Protokollierung der Suchanfragen wurde in fotomarktplatz.de jeweils gespeichert, welche Stichwörter zu einer Suchanfrage vom Benutzer eingegeben worden sind. Über zusätzliche Log-Dateien der Bildsuchmaschine weiterhin, wie viele einzelne Suchergebnisseiten, Vorschaubilder und Detailansichten gezeigter Bilder

angeklickt worden sind. Einzelne Mengen an Suchergebnissen wurden dabei (leider) nicht mit protokolliert. Allerdings kann von hohen Ergebnismengen von 500 und mehr Bildern bei Anfragen ausgegangen werden aufgrund des hohen Bildbestands in fotomarktplatz.de von mehr als 10 Millionen Bildern.

Da die manuelle Kategorisierung von über 100.000 Suchanfragen deutlich zu viel Zeit in Anspruch nehmen würde, wurde zur Kategorisierung von Suchanfragen eine zufällige Folge von 1.000 Suchanfragen ausgewählt und entsprechend kategorisiert.

5.4 Auswertung der Ergebnisse

Die Auswertung der Ergebnisse wird unterteilt nach den vier gestellten Fragestellungen in Sektion 4.2.

Zusätzlich wird entsprechend dem Vorgehen von Jörgensen in seiner Untersuchung [JJ05] mit dem generellen Ablauf der Frage nachgegangen, wie einzelne Sitzungen ablaufen und wie umfangreich diese sind. So kann der damit verbundene Aufwand gezeigt werden.

5.4.1 Umfang und genereller Ablauf einzelner Sitzungen

Die Auswertung des generellen Ablaufs einer Sitzung konnte automatisiert durch ein geeignetes Aggregieren über alle 134.943 protokollierten Suchanfragen vom 1. Februar 2006 - 31. März 2006 an fotomarktplatz.de durchgeführt werden.

Tabelle 5.1 gibt einen Überblick der Ergebnisse zum generellen Ablauf von Sitzungen und Suchanfragen.

Durchschnittlich wurden 1,77 Stichwörter zum Stellen einer Suchanfrage verwendet. Dies bestätigt sowohl die Feststellungen von Jörgensen und Sormunen, dass zumeist ein beziehungsweise maximal zwei Suchwörter verwendet werden.

Im Suchergebnis wurden in fotomarktplatz.de durchschnittlich 6,1 Seiten mit durchschnittlich etwa 150 angezeigten Vorschaubildern pro Suchergebnis (jede Seite zeigt 25 Bilder) führt. Dies bestätigt somit die Feststellung von Markkula and Sormunen [MS00], dass bei der Bildsuche ein großes Suchergebnis in Kauf genommen wird und durch Sichten der einzelnen Vorschaubilder die relevanten Bilder herausgepickt werden. Diese Feststellung kann zusätzlich an der hohen Menge an abgerufenen Detailansichten pro Suchanfrage gesehen werden, bei dem mit insgesamt 309.870 Detailansichten durchschnittlich 2,3 Detailansichten pro Suchanfrage abgerufen worden sind.

Suchanfragen und angezeigte Suchergebnisse	
Zeitraum der Messung	1. Februar 2006 - 31. März 2006
Anzahl Suchanfragen	134.943 Anfra- gen
Durchschnittliche Anzahl an verwendeten Stichwörtern	1,8 Stichwörter
Durchschnittliche Anzahl angezeigter Ergebnisseiten	6,1 Seiten
Anzeige von Detailansichten einzelner Bilder	
Anzahl angezeigter Detailansichten	309.870 Bilder
Durchschnittliche Anzahl angezeigter Detailansichten pro Suchanfrage	2,3 Bilder
Sitzungen	
Anzahl Sitzungen insgesamt	42.894 Sitzun- gen
Durchschnitt Anzahl an Suchanfragen registrierter Benutzer pro Sitzung	10,91 Anfragen

Tabelle 5.1: Generelle Auswertung der Suchanfragen und Sitzungen in fotomarkt-platz.de

Die Annahme von Jörgensen [JJ05], dass pro Idee, zu der Bilder gesucht werden sollen, eine Sitzung verwendet wird, wird auch in dieser Auswertung unterstellt. Innerhalb der insgesamt 42.894 Sitzungen wurden durch die registrierten Benutzer durchschnittlich 10,91 Anfragen pro einzelner Sitzung durchgeführt. Hieraus ist ersichtlich, dass die Bildsuche sehr aufwändig ist, da oft mehr zehn unterschiedliche Anfragen gestellt werden müssen, um die geeigneten Bilder für ein gewünschtes Bildmotiv zu finden.

Der Wert von mehr als zehn Anfragen ist deutlich höher als die gemessenen Werte pro Sitzung durch Jörgensen und Sormunen (vier Suchanfragen). Man kann annehmen, dass aufgrund des Zugriffs auf mehr als hundert Bildarchive die Benutzer davon ausgingen, dass ihr gewünschtes Bildmaterial dort vorhanden sein muss. Entsprechend haben sie über mehrere verschiedene Suchanfragen in einer Sitzung zur Erlangung des gewünschten Bildmaterials iteriert.

Bewertung der generellen Abläufe Aus dem generellen Ablauf der Sitzungen kann somit die Aussage von Markkula bestätigt werden. Der Aufwand im Vergleich zu einem Zeitungsarchiv (durchschnittlich vier bis fünf Suchanfragen pro Sitzung) ist dabei in einem Bildarchiv mit einer Vielzahl an Themen und einem Bestand von mehreren Millionen Bilder mit mehr als zehn Suchanfragen pro Sitzung noch aufwändiger. Zusätzlich wird viel Zeit in Anspruch genommen, um große Mengen an Vorschaubilder zu sichten, um daraus die geeignetsten Bilder ermitteln

zu können.

Es besteht somit auf Basis der *langen Sitzungen* und dem damit verbundenen Zeitaufwand tatsächlich die Notwendigkeit, den *Aufwand zur Bildsuche für Benutzer zu minimieren*.

5.4.2 Sind lexikalische Anfrageerweiterungen notwendig?

Entsprechend dem erstellten Modell zur *Kategorisierung von Abfolgen einer Suchanfrage* aus Sektion 4.3 wurden die 1.000 zufällig ausgewählten Suchanfragen kategorisiert

539 Suchanfragen, also mit 53,9 Prozent mehr als die Hälfte der 1.000 manuell ausgewerteten Suchanfragen, wurden neu formuliert. Dies bestätigt somit die Messung von Sormunen (57 Prozent).

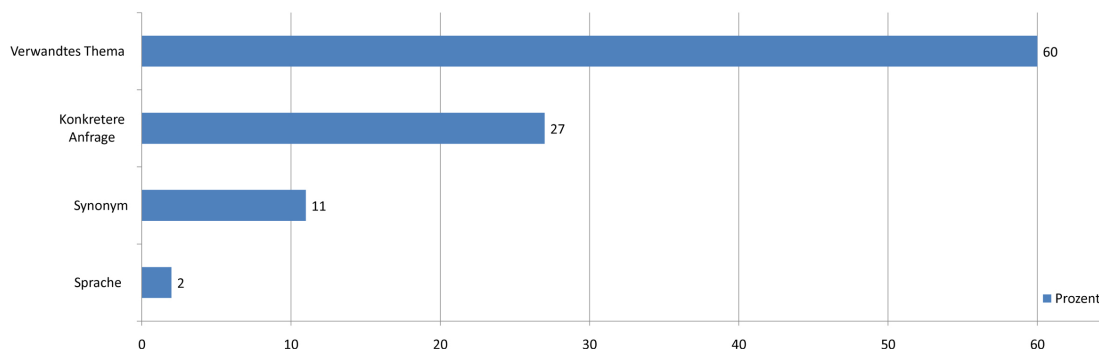


Abbildung 5.3: Kategorisierung der Abfolgen von Suchanfragen einer Sitzung

Abbildung 4.1 zeigt die Verteilung der Suchanfragen nach den vier möglichen Kategorien. 60 Prozent dieser Anfragen waren *verwandte Anfragen*. 27 Prozent der Anfragen wurden durch konkretere Anfragen neu formuliert. Weitere 11 Prozent der Anfragen wurden neu formuliert mit Hilfe von Synonymen und zwei Prozent durch das Verwenden einer anderen Landessprache, obwohl die überwiegende Menge an Beschriftungen in fotomarktplatz.de nur in deutscher Sprache vorhanden waren.

Bewertung der Ergebnisse Lexikalische Anfrageerweiterungen sind somit notwendig. Sie können die Menge an notwendigen Suchanfragen einer Sitzung einschränken. 13 Prozent der Anfragen können verringert werden durch den Einsatz von Thesauri für Synonyme und unterschiedliche Landessprachen.

5.4.3 Sind Suchanfragen miteinander verwandt?

Wie Abbildung 4.1 zeigt, handelt es sich bei 60 Prozent der Anfragen um verwandte Anfragen. Oft wurde zum Beispiel nach Namen von Schauspielern gesucht, die bei einer gleichen Veranstaltung teilnahmen oder in der gleichen Serie mitspielen. *Der Einsatz von Verfahren zum Vorschlag geeigneter Neuformulierung zum Finden verwandter Bilder zu einer gegebene Anfrage ist daher als notwendig zu bewerten.* Für 27 Prozent der Suchanfragen hin zu konkreteren Anfragen würde der Vorschlag von geeigneten Suchanfragen das „Raten“ von geeigneten Suchwörtern reduzieren.

5.4.4 Sind semantische Anfrageerweiterungen notwendig?

Die 1.000 zufällig ausgewählten Suchanfragen vom 06. Februar 2006 wurden ausgewertet und manuell entsprechend dem *Modell zur Kategorisierung der Art und dem Bezug von Suchanfragen* in Sektion 4.3 kategorisiert

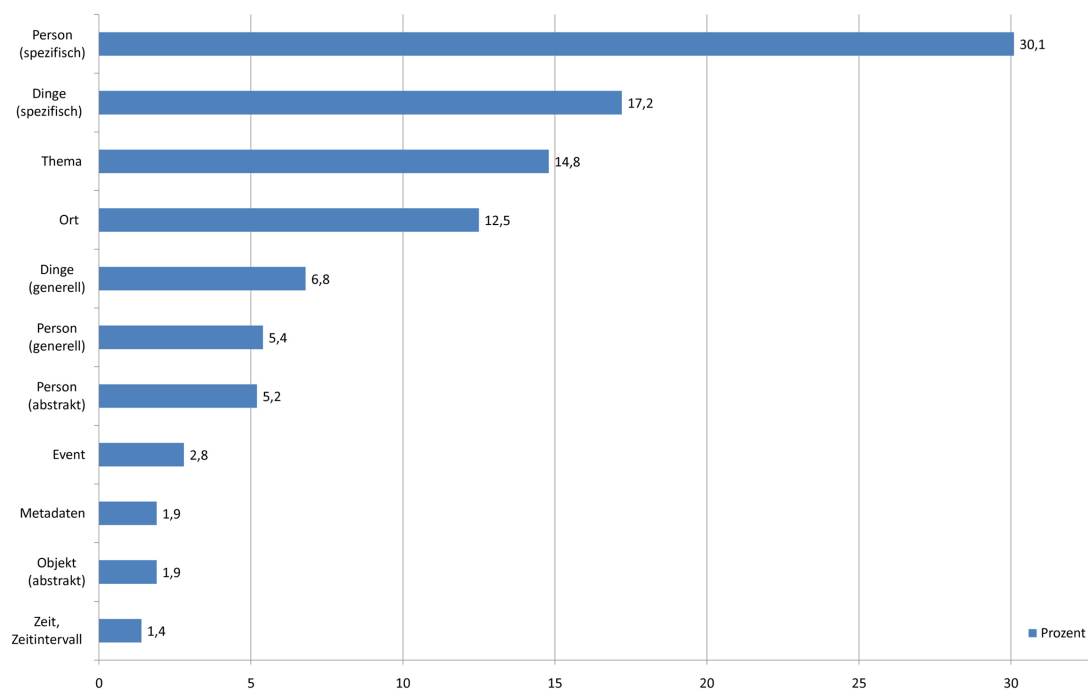


Abbildung 5.4: Kategorisierung der Suchanfragen nach Art und dem Bezug der Anfrage

Da bei der Kategorisierung auffiel, dass Anfragen zu bestimmten Veranstaltungen auftraten, also das ganze Bild betreffende Suchanfragen, wurde hierfür eine

eigene Kategorie „Event“ gebildet. Abbildung 5.4 zeigt die entstandene prozentuale Verteilung der dabei entstandenen Kategorien.

Mehr als 30 Prozent der Anfragen (30,1 Prozent) betrafen hierbei die Kategorie *Person (spezifisch)*, es wurden also gezielt Bilder gesucht, die bestimmte Personen als Bildteil abbilden. Beispiele hierzu sind die Anfragen „Michael Schumacher“ (Formel 1 Fahrer 2006) oder „Jessica Schwarz“ (deutsche Schauspielerin).

Den zweiten Platz in der Auswertung nimmt mit 17,2 Prozent der Bereich *Dinge (spezifisch)* ein, beispielsweise die Suchanfragen „Liegestuhl“ oder „Weißspitzenriffhai“ ein.

Die Kategorien *Person (spezifisch)* und *Dinge (spezifisch)* zusammen haben somit einen Anteil von 47,3 Prozent an den insgesamt gestellten Suchanfragen. Der gemessene Wert für spezifische Anfragen von Markkula von 54 Prozent und somit ungefähr der Hälfte der Anfragen kann somit bestätigt werden.

Auf dem dritten Platz mit ca. 15 Prozent innerhalb der Auswertung liegen Anfragen, die das *ganze Bild* betreffen, wie zum Beispiel die Suchanfragen „Fasching“ (aktuelles Zeitgeschehen), „Versöhnung“ (Interpretation des Bildes) oder „Tatort“ (Deutsche Krimiserie). 12,5 Prozent der Anfragen beziehen sich auf den Ort der Aufnahme, also auf Metadaten eines Bilds.

Auf den Plätzen fünf und sechs liegen die Kategorien *Dinge (generell)* mit 6,8 Prozent und *Personen (generell)* mit 5,2 Prozent. Somit betreffen insgesamt etwa 12 Prozent den Bereich der generellen Anfragen zu Objekten.

Mit dem Platz vier in der Auswertung fielen 12,5 Prozent der Suchanfragen auf die Kategorie *Ort* und somit der Fragestellung „Wo entstand das Bild“. Die restlichen Kategorien nehmen einen eher geringen Anteil an, also dem *Event* mit 2,8 Prozent, *Objekt (imaginär)* mit 1,9 Prozent der Anfragen sowie nach der Zeit mit 1,4 Prozent und sonstigen Anfragen zu Metadaten mit 1,9 Prozent.

Bewertung der Ergebnisse: 47,3 Prozent der Anfragen sind spezifische Anfragen nach Personen oder Dingen, 12,2 Prozent der Anfragen sind generelle Suchanfragen und weitere 7,1 Prozent der Anfragen sind imaginärer Art. Somit sind für 20 Prozent der Suchanfragen entsprechende Interpretationen nötig (z.B. Anfrage „Frau“ abbilden auf „Namen aller Bilder die weibliche Personen zeigen“). Entsprechend diesen Zahlen und dem erwähnten häufigen Suchen nach verwandten Begriffen *ist der Einsatz semantischer Anfrageerweiterungen notwendig* und diese Frage mit ja zu beantworten.

Sind Beschriftungen für Bildteile notwendig?

Abbildung 5.4 zeigt das Ergebnis der Unterscheidung der Suchanfragen nach ihrem Bezug. Zwei Drittel der Suchanfragen beziehen sich direkt auf Bildmotive

(Personen und Dinge). Somit beziehen sich nur etwa ein Drittel der Suchanfragen auf das Thema des ganzen Bilds.

Zur Bevorzugung von Ergebnissen, die ein Bildmotiv direkt abbilden sind somit Beschriftungen für Bildteile erforderlich. Die weitere Unterscheidung des Bezugs der Anfrage auf Dinge und Personen zeigt, dass sich etwa vierzig Prozent der Anfragen auf Personen beziehen. *Somit ist der umfangreiche Einsatz von automatischen Verfahren empfehlenswert, um Personen und Gesichter automatisch innerhalb eines Bilds erkennen können.*

5.4.5 Themen der Suchanfragen

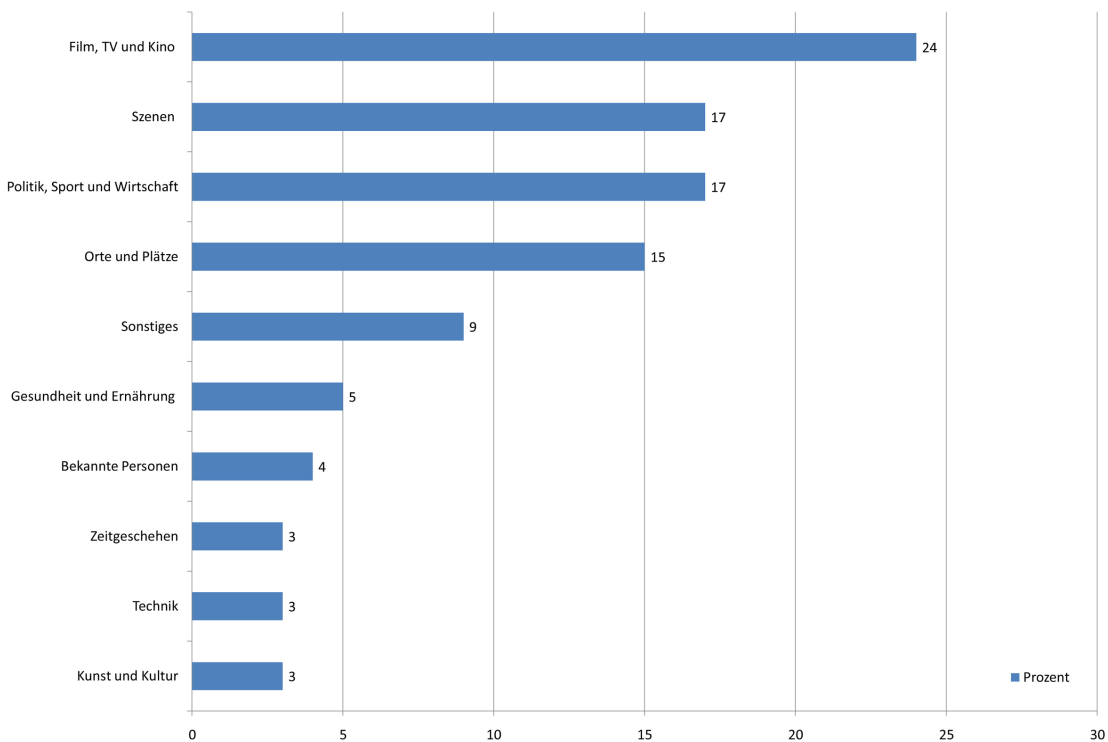


Abbildung 5.5: Hauptthemen innerhalb von 1000 Suchanfragen

Von weiterem Interesse in der Auswertung war, auf welche Themen sich die Suchanfragen beziehen. Das Wissen um die verwendeten Themen ermöglicht den Betreibern von Bildsuchmaschinen, die von den Benutzern am meisten benötigten Bildmotive zu kennen. So bietet es sich an, sich bei der Erstellung von Bildbeschriftungen schwerpunktmässig auf diese Motive zu konzentrieren.

Abbildung 5.5 zeigt die entstandene Verteilung der Suchanfragen auf zehn

identifizierte Hauptthemenbereiche. Mit 24 Prozent der Anfragen haben sich hierbei die meisten Anfragen auf das Thema *Film, TV und Kino* bezogen, also Anfragen zu Schauspielern, Filmtiteln und TV Serien. Auf den zweiten Platz fielen mit 17 Prozent Anfragen zum Thema *Szenen*, also der Darstellung bestimmter Szenen wie zum Beispiel „Leute im Bus“ oder „Mann vor Bildschirm“. Den dritten Platz nahmen mit 17 Prozent Anfragen aus dem Themenbereich *Politik, Sport und Wirtschaft*, also den Hauptbereichen einer Tageszeitung, ein. Beispiele für Anfragen hierzu sind „Staatsbesuch Rau“ oder „Pressekonferenz Deutsche Bank“. Auf den vierten Platz fielen Anfragen zum Hauptthema *Orte und Plätze* wie z.B. „Skyline New York“ oder „Frauenkirche Dresden“. Die weiteren ermittelten Hauptthemen sind *Gesundheit und Ernährung* mit 5 Prozent, *Bekannte Personen* mit 4 Prozent, *Zeitgeschehen* mit 3 Prozent, *Technik* mit 3 Prozent und *Kunst und Kultur* mit 3 Prozent. Weitere 9 Prozent fielen auf den Bereich *Sonstiges*.

Bilder zu den Themen *Film, TV und Kino, Szenen* sowie *Politik, Sport und Wissenschaft* nehmen dabei die ersten drei Plätze ein. Auf Basis dieser Untersuchung wurde daher die Entscheidung getroffen, als Bildmotive für weitere empirische Untersuchungen mit Bildsuchern und Bildredakteuren in dieser Arbeit zum Einen Bilder zur Domäne *Film* sowie zum Anderen zur Domäne *Politiker der EU* zu verwenden.

5.4.6 Schlussfolgerung aus den Ergebnissen

Auf Basis der Untersuchung konnten alle vier Fragen, die in der Untersuchung geklärt werden sollten, positiv beantwortet werden. Der *Einsatz lexikalischer Verfahren zur Anfrageerweiterung* ist genauso sinnvoll und notwendig zur Verbesserung der Bildsuche wie der *Einsatz semantischer Verfahren zur Anfrageerweiterung*. Weiterhin ist die Beschriftung von Bildteilen hilfreich für die Erstellung hochwertiger Suchergebnisse, da Bilder bevorzugt werden können, die Bildteile tatsächlich abbilden. Da die Suchanfragen oft zueinander in Verwandtschaft stehen, ist der Einsatz von *Verfahren zur Neuformulierung von Suchanfragen* zusätzlich sinnvoll.

Die Antwort auf Frage 1 (b) lautet daher wie folgt:

Frage 1 (b)

Werden semantische Verfahren zur Verbesserung der Bildsuche benötigt?

Antwort

Ja, semantische Verfahren zur Verbesserung der Bildsuche werden benötigt. Sie sollen die hohe Anzahl an Suchanfragen pro Sitzung und damit den Zeitaufwand für die Benutzer reduzieren. Hierzu eignen sich die Verfahren *lexikalische Anfrageerweiterung, semantische Anfrageerweiterung, Beschriftung von Bildteilen* sowie die *Neuformulierung von Suchanfragen*.

5.5 Zusammenfassung

Zur Beantwortung der Frage 1 (b) wurde eine Analyse von mehr als 100.000 Suchanfragen an fotomarktplatz.de durchgeführt. Das Ergebnis ist, dass tatsächlich ein Handlungsbedarf innerhalb von Bildsuchmaschinen vorhanden ist, um die Bildsuche zu verbessern.

Die lexikalische und semantische Anfrageerweiterung soll die notwendige Menge an Suchanfragen einer Sitzung verkürzen. Die Beschriftung von Bildteilen kann die Präzision der Suchergebnisse erhöhen, da dann tatsächlich Bilder gezeigt werden können, die das passend zur Suchanfrage gewollte Motiv abbilden. Schließlich kann die Neuformulierung von Suchanfragen ebenfalls die nötige Menge an Suchanfragen reduzieren, da den Benutzern gezielt geeignete Suchanfragen vorgeschlagen werden können.

Zur Unterstützung der Bildsuche hin zu hochwertigen Suchergebnissen ist daher der Einsatz semantischer Verfahren sinnvoll und notwendig.

Kapitel 6

Onlineevaluation zur semantischen Bildsuche

Bei der Untersuchung des Einsatzes semantischer Verfahren in Suchmaschinen liegt der Fokus häufig auf der Systemseite, also inwieweit die Verfahren die Präzision und die Vollständigkeit eines Suchergebnisses erhöhen können (z.B. in [Nag07] und [Yoo09]). Die Benutzerseite wird hierbei meist vernachlässigt. In diesem Kapitel soll daher der Fokus auf der Seite des Benutzers liegen. *Versteht dieser überhaupt die Möglichkeiten der mittels Semantik möglichen Bildsuche?* Können diese Verfahren somit dazu beitragen, dass Benutzer präzisere Suchanfragen formulieren und somit hochwertigere Ergebnisse erhalten können? Und falls ja, wie sollen diese Verfahren eingesetzt werden?

Innerhalb des IMAGINATION Projekts wurde früh ein erster Prototyp der ImageNotion Anwendung entwickelt. Mit diesem Prototyp wurde die in diesem Kapitel gezeigte Onlineevaluation durchgeführt. Ziel der Befragung war es zu ermitteln, inwieweit Benutzer *die semantische Bildsuche verstehen* und ob die *Verfahren zur Formulierung von Suchanfragen basierend auf semantischen Verfahren* von den Benutzern verstanden werden. Die Untersuchung sollte hierbei eine große Menge an Benutzern mit einbeziehen, was letztlich zur Entscheidung führte, eine Onlineevaluation durchzuführen (Teile der Ergebnisse wurden auch publiziert in [WNN08] und [WN07c]).

Die Ergebnisse der Onlineevaluation ermöglichen somit die Beantwortung der Frage 1 (c) dieser Arbeit *Sind die entstehenden Möglichkeiten zur Bildsuche für Nutzer in der Praxis verständlich und wie sollen diese eingesetzt werden?*

6.1 Untersuchungen zum Verständnis von Benutzern für semantische Ansätze

Basierend auf verwandten Arbeiten sollen bestehende Ergebnisse aus der Literatur gezeigt werden, die sich ebenfalls mit der Frage nach dem *Verständnis von Benutzern für semantische Ansätze* beschäftigt haben. Dies soll motivieren, dass auch innerhalb der Bildsuche semantische Ansätze verstanden werden können.

Mit der Frage nach dem *Verständnis von Benutzern für semantische Ansätze* ist gemeint, ob diese die Möglichkeiten der semantischen Suche verstehen und nachvollziehen können. Semantische Suche heißt hierbei, dass Benutzer nicht nur *spezifisch* suchen, sondern auch *generelle* oder *imaginäre* Suchanfragen formulieren und die resultierenden Ergebnisse verstehen können [Hol06].

Jaimes und Chang haben mit ihrem Modell konzeptionell gezeigt, wie sich generelle, spezifische und imaginäre Beschriftungen für Bildteile und dem ganzen Bild unterteilen und indizieren lassen [JJfC00]. Inwieweit Benutzer dieses konzeptionelle semantische Modell innerhalb der Bildsuche verstehen, wurde dort nicht untersucht.

Daher ging Hollink in ihrer Untersuchung der Frage nach, inwieweit Benutzer semantische Ansätze in der Bildsuche verstehen, um damit Bilder nach diesen Ebenen zu beschreiben [Hol06]. Sie gab hierzu dreißig Teilnehmern in einem Workshop drei Texte, die den Inhalt eines Bildes als Text beschreiben. Die Bilder selbst zu den Beschreibungstexten wurden den Benutzern nicht gezeigt.

Die Beschreibungstexte wurden bewusst sehr generell gehalten um zu sehen, inwieweit die Benutzer diese Beschreibungen präzisieren. Die Benutzer sollten daher die Texte in Bildbeschriftungen überführen. Im Ergebnis waren 74 Prozent der erstellten Beschriftungen generell, weitere 16 Prozent spezifisch und 10 Prozent der Beschriftungen imaginär. Aus der Tatsache, dass in allen drei Ebenen Beschriftungen genannt und dabei häufig generelle Beschreibungen zu spezifischeren Anfragen präzisiert wurden, leitete Hollink ab, dass die Benutzer dann auch diesen Ablauf innerhalb einer Bildsuche verstehen und nachvollziehen können. Eine entsprechende Untersuchung im tatsächlichen Einsatz der semantischen Suche hat sie aber nicht durchgeführt.

Im letzten Kapitel wurde bereits gezeigt, dass Benutzer in fotomarktplatz.de tatsächlich generelle, spezifische und imaginäre Suchanfragen stellen. Dies deutet auf eine eher unbewusste Nutzung von semantischem Hintergrundwissen bei der Bildsuche hin, zeigt aber noch nicht, dass die Benutzer auch bewusst damit umgehen können. Im Gegensatz zu Hollink soll daher die Onlineevaluation zeigen, inwieweit Benutzer tatsächlich semantische Suchanfragen formulieren können. Dies ermöglicht, die gestellte Frage nicht nur durch Rückschlüsse, sondern *direkt durch die ersichtli-*

chen Vorgehen der Benutzer in der Bildsuche beantworten zu können.

Die Benutzer sollen daher in der eigenen Untersuchung Aufgaben erhalten, um semantische Suchanfragen mit ImageNotion durchzuführen. Ermittelt wird dann, ob sie diese Aufgaben erfolgreich umsetzen können. Falls dies überwiegend der Fall ist, so sind die entstehenden Möglichkeiten der semantischen Suche für Benutzer verständlich und somit praktisch umsetzbar.

6.2 Verfahren für die Evaluation der semantischen Bildsuche

In diesem Abschnitt werden bestehende *Verfahren zum Stellen von Suchanfragen* vorgestellt. Diese sollen zur Evaluation der semantischen Bildsuche mittels dem Stellen und Neuformulieren von Suchanfragen verwendet werden. Falls nötig, werden diese Verfahren erweitert oder angepasst, um eine semantische Bildsuche zu ermöglichen. Ziel in der Onlineevaluation ist dann, diese Verfahren hinsichtlich ihrer Verständlichkeit für Benutzer zum Stellen von semantischen Suchanfragen nach Bildern zu untersuchen.

Die aus der Literatur bekannten Verfahren zum *Stellen von Suchanfragen* lassen sich unterteilen nach Verfahren zur *Eingabe einer Suchanfrage*, *Browsen von Ontologien*, *Neuformulierung einer gegebenen Suchanfrage* und falls beschriftete Bildteile vorhanden sind Verfahren zur *Navigation durch Bildteile*.

6.2.1 Eingabe einer semantischen Suchanfrage

Bei der *textbasierten Eingabe einer Suchanfrage* [AHAQI⁺07] werden Stichwörter eingegeben und über den Index ein Suchergebnis generiert (wie in Sektion 3.1 beschrieben). Diese Form der Suche sind die Benutzer von Bildsuchmaschinen gewöhnt und wird von diesen aufgrund der Einfachheit des Verfahrens verstanden¹.

Verwandte Arbeiten

Die folgenden Verfahren zur *Eingabe einer semantischen Suchanfrage* sind aus der Literatur bekannt. Entsprechende umfangreiche Evaluationen mit Benutzern konnten nicht ermittelt werden. Daher ist in der Onlineevaluation zu ermitteln, ob diese Verfahren von Benutzern verwendet werden können. Werden diese Verfahren rich-

¹Die Suche nach Webseiten bei Google funktioniert in der gleichen Form.

tig angewendet und verstanden, können diese aus der Sicht der Benutzer zu den in dieser Arbeit gewünschten hochwertigeren Suchergebnissen führen.

Vorschlag semantischer Elemente für die Suche Bei der ersten Möglichkeit wird Text eingegeben und passende semantische Elemente zur Eingabe werden angezeigt [vOTSP07]. Basierend auf Sektion 3.3.1 führend diese Suchanfragen zu präzisieren Suchanfragen und somit präziseren Ergebnissen. Dies ist also der erste Aspekt in die Richtung zu hochwertigen Suchergebnissen.

Auflösung einer semantischen Suchanfrage Die zweite Möglichkeit ist die Auflösung einer semantischen Anfrage in spezifische Anfrageteile [Nag07]. Die Suchanfrage *EU Präsident* kann dann zum Beispiel durch das System in die Anfrage *Manuel Jose Barroso* und *Romano Prodi* unterteilt werden. Basierend auf Sektion 3.3.1 führen so interpretierte Suchanfragen zu vollständigeren und präziseren Suchergebnissen. Dies ist also der zweite Aspekt in die Richtung zu hochwertigen Suchergebnissen.

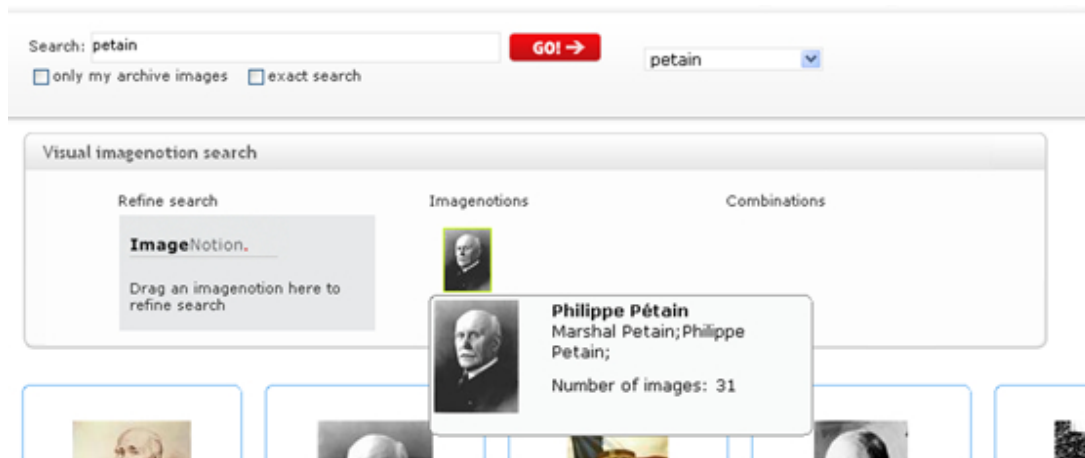


Abbildung 6.1: Vorschlag semantischer Elemente für die Texteingabe „Petain“

Umsetzung im ImageNotion Prototyp

Beide Verfahren sind im ImageNotion Prototyp integriert. Abbildung 6.1 zeigt die Umsetzung zum Vorschlag semantischer Elemente. Eine Benutzer hat „Petain“ als Suchanfrage eingegeben und erhält als Vorschlag das semantische Element *Philippe Petain* zum Starten einer semantischen Suchanfrage. In ImageNotion werden einzelne semantische Elemente visuell durch entsprechende Bilder dargestellt. Durch klicken auf das Element *Philippe Petain* kann die entsprechende semantische Suchanfrage für Bilder dieses französischen Generals aus dem ersten Weltkrieg gestartet werden.

6.2.2 Browsen der Ontologie

Das Browsen der Ontologie ermöglicht die Anzeige der Inhalte einer Ontologie.

Verwandte Arbeiten

Der Benutzer klickt einzelne Elemente in der Ontologie an und bekommt die damit verbundenen textuellen Beschreibungen und semantischen Elemente (z.B. in der Hierarchie einer Taxonomie) angezeigt (z.B. in [vOTSP07] [VFaC05]). Er kann so sein gewünschtes Element auswählen und eine semantische Suche starten.

Umsetzung im ImageNotion Prototyp

Das Browsen der Ontologie wurde in der eben beschriebenen Art in ImageNotion durch Integration eines *Ontologiebrowsers* umgesetzt. In der Onlineevaluation soll von Interesse sein, ob die Benutzer durch eine Ontologie navigieren können und wollen, um eine geeignete semantische Suchanfrage zu starten.

6.2.3 Verfahren zur Neuformulierung einer gegebenen Suchanfrage

Die Neuformulierung einer gegebenen Suchanfrage meint das Vorschlagen geeigneter Einschränkungen einer Suchanfrage. Die Suchanfrage soll so präziser werden.

Verwandte Arbeiten

Aus der Literatur sind drei Verfahren zur Neuformulierung einer gegebenen Suchanfrage bekannt, das *Vorschlagen ähnlicher Suchanfragen*, das *Vorschlagen verwandter Suchanfragen* und das *Thematische Zusammenfassen von Suchergebnissen*.

Vorschlag ähnlicher Suchanfragen. Basierend auf textbasierten Metadaten ist zum Einen das Vorschlagen *ähnlicher Suchanfragen* [LW06] möglich. Zu einer Suchanfrage werden so über einen Thesaurus ähnliche Wörter zu einer Eingabe ermittelt. Weiterhin können häufig zusammen mit der Anfrage ermittelte Wörter in den Beschriftungen ermittelt werden und zur Präzision der Suchanfrage gezeigt werden (z.B. „Petain“ nach „Philippe Petain“).

Vorschlag verwandter Suchanfragen Basierend auf semantischen Verfahren lassen sich weiterhin auch verwandte Suchanfragen vorschlagen. Wie zum Beispiel in der semantischen Bildsuchmaschine *PhotoStuff* [vOTSP07] ist dies die An-

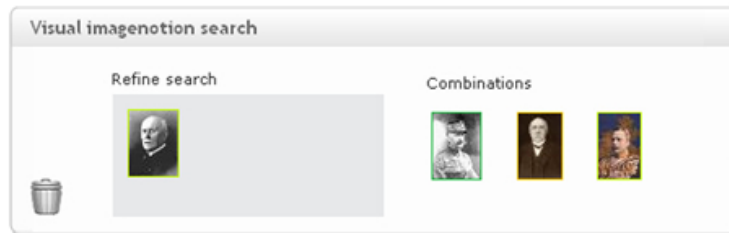


Abbildung 6.2: Zusammenfassung von semantischen Elementen für die Anfrage *Philippe Pétain*

zeige von weiteren Astronauten bei der Anfrage nach dem Namen eines bestimmten Astronauten.

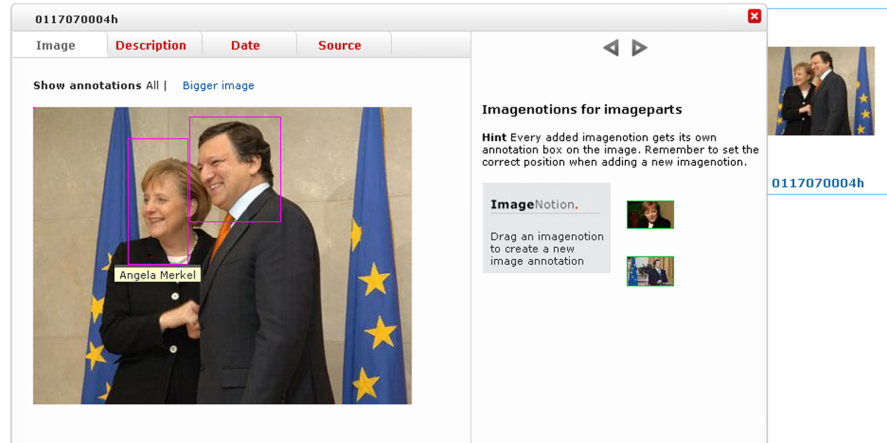
Thematische Zusammenfassung von Suchergebnissen. Ein weiteres Verfahren ist die *Thematische Zusammenfassung von Suchergebnissen* [Yoo09]. Hierbei werden gleichartige Beschriftungen der Bilder im Suchergebnis ermittelt und dem Benutzer als thematische Zusammenfassung angezeigt. Der Benutzer kann dadurch ein gewünschte Cluster auswählen und damit eine präzisere Suchanfrage starten. Ein Beispiel ist die Anfrage „Paris“, bei der anschließend Bilder in Bilder gruppiert werden können zu Clustern wie „Paris Hilton“, „Paris Eiffelturm“ oder „Paris Seine“.

Umsetzung in ImageNotion Prototyp:

In ImageNotion werden *ähnliche Suchanfragen* direkt in die Suche mittels Anfragerweiterung (siehe Sektion 3.3.1) einbezogen. Dieses Verfahren wird somit nicht explizit dem Benutzer angezeigt und daher auch nicht evaluiert.

Das Verfahren *Vorschlägen verwandter Suchanfragen* wird wie beschrieben in ImageNotion umgesetzt. In der Evaluation wird hierbei die Frage gestellt, ob die Benutzer die zum Vorschlägen verwandter Suchanfragen nötigen Relationen verstehen können und nutzen möchten.

Ebenso wird im ImageNotion Prototyp die *thematische Zusammenfassung von Suchergebnissen* umgesetzt und als *Zusammenfassung semantischer Elemente* bezeichnet (siehe Abbildung 6.2), da zur Erstellung von Clustern semantische Elemente der Metadaten ermittelt werden. Bei der semantischen Suche *Philippe Pétain* werden daher unter anderem *Joseph Joffre* oder *Georges Clemenceau* vorgeschlagen, da es im Archiv viele Bilder gibt, die jeweils beide Personen zusammen auf Bildern zeigen.

Abbildung 6.3: Navigation durch Bildteile, Anfrage *Angela Merkel*

6.2.4 Weitere Verfahren

Aus der Literatur sind weitere Verfahren zum Stellen von Suchanfragen bekannt. Da diese aber nicht die semantische Bildsuche unterstützen oder ermöglichen, werden diese nur kurz vorgestellt.

Verfahren auf Ähnlichkeitssuche ermöglichen nach den perceptiven Bereiche [JJfC00] eines Bilds wie Farbe, Strukturen und Formen zu suchen und finden hierzu ähnliche Bilder [LW06]. Da die Suchanfragen hierbei nicht semantisch sind, werden diese für die Evaluation nicht weiter betrachtet.

Die Gruppierung eines Suchergebnisses auf einer Landkarte ermöglicht das schnelle Auffinden der Bilder eines bestimmten Ortes. So könnten die Bilder bei der Suche „Paris“ auf einer Karte nach ihren Sehenswürdigkeiten geordnet werden. Das Verfahren nutzt somit nicht semantische Verfahren, sondern in die Bilder gespeicherte GPS Informationen. Daher wird das Verfahren für die Evaluation nicht weiter betrachtet.

6.2.5 Navigation durch Bildteile

Die Navigation durch Bildteile [vOTSP07] ermöglicht das Klicken auf einen beschrifteten Bildteils. Anschließend wird eine neue Suche gestartet.

Umsetzung in ImageNotion Prototyp: Dieses Verfahren wurde im ImageNotion Prototyp wie eben beschrieben umgesetzt, allerdings werden neue Suchen dann als semantische Suchanfragen basierend auf den semantischen Beschriftungen der Bildteile gestartet. Abbildung 6.3 zeigt ein Beispiel für die Navigation durch

Bildteile in ImageNotion. Das Bild enthält eine Beschriftung für den Bildteil, der *Angela Merkel* zeigt. Durch einen Klick auf den Bildteil wird eine entsprechende semantische Suche nach *Angela Merkel* gestartet. Für eine möglich Erweiterung der Navigation durch Bildteile sollen in der Evaluation die Benutzer gefragt werden, ob sie tatsächlich immer eine neue Suche starten möchten bei dem Klick auf einen Bildteil oder ob die aktuell bestehende Suche mit berücksichtigt werden soll.

6.3 Empirische Studie 1: Onlineevaluation der semantischen Bildsuche

Zur Beantwortung der Frage 1 (c) dieser Arbeit *Sind die entstehenden Möglichkeiten zur Bildsuche für Nutzer in der Praxis verständlich und wie sollen diese eingesetzt werden?* soll im Folgenden eine Onlineevaluation durchgeführt werden.

Die folgenden Ziele wurden an die Onlineevaluation gestellt, die durch entsprechende Fragestellungen beantwortet werden sollen:

- **(Ziel 1) Können die Benutzer mit der semantischen Bildsuche umgehen?** Um diese Frage zu beantworten, bekommen die Benutzer Suchaufgaben zur Formulierung semantischer Suchanfragen gestellt. Wenn diese von einer großen Prozentzahl der Benutzer durchgeführt werden können, kann man daraus schließen, dass die semantische Bildsuche generell für Benutzer verständlich ist.
- **(Ziel 2) Können die Benutzer mit semantischen Verfahren zur Neuformulierung von Suchanfragen umgehen?** Um diese Frage zu beantworten, bekommen die Benutzer ebenfalls entsprechende Suchaufgaben gestellt. Können diese Anfragen von einer großen Prozentzahl der Benutzer durchgeführt werden, lohnt sich die Erstellung entsprechender Möglichkeiten in einer Bildsuchmaschine, da diese von den Benutzern anwendbar sind.
- **(Ziel 3) Verstehen die Benutzer die Bedeutung von Relationen?** Es soll weiterhin ermittelt werden, ob Benutzer die Bedeutung von Relationen verstehen. Hierzu sollen sie entsprechende Aufgaben und Fragen gestellt bekommen. Werden Relationen verstanden, lohnt es sich, diese umfangreicher bei der Erstellung von Instanzenmodellen zu berücksichtigen, damit diese anschließend in entsprechenden Suchanfragen ausgenutzt werden können².

²Ein Beispiel wäre „Finde alle Bilder von portugiesischen Politikern, die bei der EU Kommission arbeiten.“

- **(Ziel 4) Wie soll die Navigation durch Bildteile funktionieren?** Es soll ermittelt werden, ob die Benutzer eine Bedeutung in beschrifteten Bildteilen sehen und die Navigation durch Bildteile durchführen können. Zusätzlich soll die Frage geklärt werden, wie eine Bildsuche beim Klicken auf einen Bildteil umgesetzt werden soll.

Der ImageNotion Prototyp wurde auf einem öffentlich zugänglichen Server im Forschungszentrum Informatik Karlsruhe installiert. 200 Bilder aus der Domäne *EU Politiker* wurden von Bildagenturen innerhalb des IMAGINATION Projekts mit semantischen Metadaten versehen und beschriftet. Die semantischen Metadaten nebst den ursprünglich bereits vorhandenen textbasierten Metadaten wurden in den Prototyp installiert.

Zur Durchführung der Onlineevaluation wurde ein Fragebogen erstellt und als Webseite formatiert zugänglich gemacht. Die Benutzer können im Fragebogen die durchzuführenden Aufgaben lesen und erhalten Eingabefelder zur Beantwortung entsprechender Fragen.

Der Fragebogen wurde als E-Mail über die Verteiler von Partnern des IMAGINATION Projekts, sowie weiterhin mit der Unterstützung von EU Projekten wie Theseus und Mature-IP über deren Verteiler versandt. Insgesamt konnten so mehr als 1.000 Einladungen versandt werden. Zur Motivation der Onlineevaluation wurde als Preis ein Apple iPod unter allen Teilnehmern verlost, welche den Fragebogen komplett ausgefüllt haben. Mehr als 250 Benutzer begannen mit dem Ausfüllen des Fragebogens, wovon 137 Teilnehmer den Fragebogen komplett ausgefüllt haben. Die kalkulierte Zeitdauer zur Bearbeitung aller Fragen lag bei 20 Minuten. Die Ergebnisse der 137 komplett ausgefüllten Fragebögen wurden ausgewertet.

Zur Ermittlung, ob die einzelnen Aufgaben zur Bildsuche von den Benutzern erfolgreich umgesetzt werden konnten, wurden die Log-Dateien der ImageNotion Anwendung ausgewertet.

6.4 Auswertung der Ergebnisse

Die Auswertung der Ergebnisse wird unterteilt nach den vier formulierten Teilzielen der Onlineevaluation.

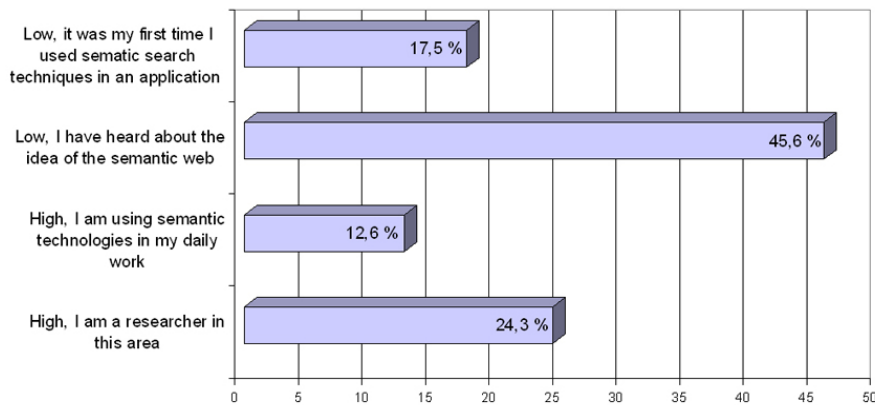


Abbildung 6.4: Persönliche Erfahrung der Benutzer mit semantischen Anwendungen

6.4.1 Erfahrung der Benutzer mit semantischen Anwendungen

Die erste Fragestellung betraf die *persönlichen Erfahrung der Benutzer mit semantischen Anwendungen*. Abbildung 6.4 zeigt die Verteilung der dabei genannten Antworten. 17,5 Prozent der Benutzer hatten keinerlei Erfahrungen mit semantischen Anwendungen, 45,6 Prozent der Benutzer haben schon einmal vom *Semantic Web* [SBLH06] und dessen Ideen gehört. Mehr als 70 Prozent der Anwender haben somit wenig Erfahrung mit semantischen Anwendungen, so dass die in der Evaluation durchgeführten Suchmöglichkeiten *für diese neu sind*. Die weiteren ersichtlichen Benutzergruppen nutzen semantische Anwendung regelmäßig (12,6 Prozent) oder forschen in diesem Gebiet (24,3 Prozent).

6.4.2 Verständnis für die semantische Bildsuche

Der erste Bereich der Aufgaben und Fragen betrifft (Ziel 1), also *Können die Benutzer mit semantischer Bildsuche umgehen?*

Ergebnisse

Tabelle 6.1 gibt einen Überblick der Ergebnisse auf die gestellten Aufgaben und Fragen.

Innerhalb der ersten Frage wurde den Benutzern in einem einleitenden Text die Idee der semantischen Bildsuche vorgestellt und dazu mögliche Beispiele genannt.

Frage	Ja (%)	Nein (%)
Habe Sie die Idee der semantischen Bildsuche verstanden?	95,6	4,4
Benutzer, die textbasierte Suche für „José Manuel Durão Barroso“ durchführen konnten	94,1	5,9
Benutzer, die semantische Suche „Vorschlag semantischer Elemente“ für Eingabe „Barroso“ durchführen konnten	88,3	11,7
Empfinden Sie das Ergebnis der semantischen Suche als besseres Ergebnis im Vergleich zur textbasierten Suche?	78,4	3,1
Benutzer, die semantische Suche „EU Präsidenten“ durchführen konnten	72,3	27,7
Empfinden sie das gezeigten Suchergebnis für „EU Präsidenten“ als korrekt?	69,3	30,7
Sollten die gezeigten semantischen Suchverfahren in Bildsuchmaschinen eingeführt werden?	71,5	28,5
Benutzer, die Bilder von „José Manuel Durão Barroso“ über den Ontology-Browser gefunden haben	64,2	35,8
Empfinden sie den Ontology-Browser als hilfreich für die Bildsuche	37,2	72,8

Tabelle 6.1: Ergebnisse für den Aufgabenbereich „Verständnis für die semantische Bildsuche“

Anschließend wurden Sie gefragt, ob „die Idee der semantischen Bildsuche verstanden wurde“. 95,6 Prozent der Benutzer haben die Frage mit Ja beantwortet. *Der generelle Zweck der Evaluation war somit den meisten Benutzern ersichtlich.*

Umsetzung textbasierter Suchanfragen: Die erste Suchaufgabe bestand darin, eine textbasierte Bildsuche durchzuführen. Die Aufgabe war daher „Starten Sie eine Suche nach dem aktuellen Präsident der EU Kommission José *Manuel Durão Barroso*“. Aufgrund der Analyse der Log Dateien war ersichtlich, dass 94,1 Prozent der Benutzer die Suchanfrage erfolgreich durchführen konnten. Sie gaben zu meist „Barroso“ oder „Manuel Barroso“ als Suchanfrage ein. Technische Probleme im Prototyp (z.B. mit unterschiedlichen Internetbrowsern) werden als Ursache angenommen, warum die anderen 5,9 Prozent der Benutzer die Anfrage nicht umsetzen konnten, denn bei diesen Benutzern konnte keine Suchanfrage in den Log Dateien gefunden werden. *Die textbasierte Suche konnte also von der überwiegenden Menge an Benutzern erfolgreich durchgeführt werden.*

Vorschlag semantischer Elemente: Die zweite Suchaufgabe bestand darin, mit Hilfe der ImageNotion Anwendung ihre soeben durchgeführte Bildsuche basierend auf einer semantischen Suchanfrage nach *Manuel Barosso* zu stellen. Die ImageNotion Anwendung listet hierbei zu einer Suchanfrage passende semantische Elemente auf. Die Aufgabe des Benutzers lag daher darin, das geeignete semantische Element in dieser Anzeige auszuwählen und dieses anzuklicken, um eine entsprechende semantische Suche zu starten. In den Log Files konnten bei 88,3 Prozent der Benutzer festgestellt werden, dass sie die Aufgabe entsprechend korrekt umgesetzt haben. Die anderen Benutzer hatten entweder technische Probleme oder haben die Aufgabe nicht verstanden und daher nicht richtig umgesetzt. *Das Verfahren „Vorschlag semantischer Elemente“ konnte somit ebenfalls von den meisten Benutzern durchgeführt werden.*

Zur dieser Suchaufgabe wurden die Benutzer zusätzlich gefragt, ob sie „das Ergebnis der semantischen Suche als besseres Ergebnis im Vergleich zur textbasierten Suche empfinden“. ImageNotion sortiert die Bilder der semantischen Suche neu. Bilder mit großen beschrifteten Bildteilen, die Barroso zeigen wurden vor Bilder gezeigt, die ihn nur in Gruppenaufnahmen zeigen. 78,4 Prozent der Benutzer empfanden dieses so erstellte Ergebnis als das bessere Ergebnis. *Man kann somit schließen, dass die Benutzer eine Sortierung der Bilder nach Bildinhalt bevorzugen.* Ein geringer Anteil von 3,3 Prozent empfand dieses Ergebnis als schlechter. Die restlichen 18,3 Prozent hielten beide Ergebnisse gleich gut.

Auflösung einer semantischen Suchanfrage Die dritte Suchaufgabe in diesem Bereich lag darin, die „Bilder aller EU Präsidenten im Bildarchiv“ zu finden. Diese Aufgabe konnte in mehreren Wegen richtig gelöst werden, was auch die Untersuchung der Log Dateien ersichtlich machte. Der kürzeste Weg und aus der

Sicht der semantischen Bildsuche *richtige* Weg ist die Eingabe von z.B. „EU“, „Präsident“ oder „EU Präsident“. ImageNotion hat dann das semantische Element „EU Präsident“ angezeigt. Durch Klick auf dieses Element kann dann die Suche aller Bilder, die EU Präsidenten zeigen gestartet werden. Das Suchergebnis zeigt Bilder an, welche die beiden im Archiv vorhandenen Präsidenten der EU Kommission *Manuel Barroso* und *Romano Prodi* zeigen. Ein alternativer Weg um diese Bilder zu finden waren textbasierte Suchanfragen wie zum Beispiel „Prodi Barroso“. Da hier aber die semantische Suche aufgrund dem Wissen des Benutzers nachgebildet wurde, wurde diese Art der Anfrage als falsch gewertet. Ebenso Eingaben wie „president“ und keine weiteren Aktionen. 72,3 Prozent der Benutzer konnten mit dem als richtig gewerteten Ablauf die Bildsuche richtig durchführen. Entsprechend haben 27,7 Prozent der Benutzer die Aufgabe nicht lösen können.

Zu dieser Suchaufgabe wurden die Benutzer gefragt, ob sie „das gezeigte Suchergebnis für „EU Präsidenten“ als korrekt empfinden?“. 69,3 Prozent der Benutzer haben die Frage mit ja beantwortet. *Sie konnten somit das gezeigte Ergebnis nachvollziehen und als richtig einstufen.* Gründe für die Beantwortung mit nein können falsche Suchanfragen sein. Die Eingabe „president“ zeigte auch Bilder von *George W. Bush* (ehemaliger Präsident der USA). Ein Grund für ein als falsch empfundenes Suchergebnis kann auch sein, dass die Benutzer zwar die Suche richtig durchgeführt haben, aber weitere bisherige Präsidenten der EU Kommission wie Manuel Marín erwartet haben. Diese waren aber im Archiv nicht vorhanden.

Die Frage, ob „die gezeigten semantischen Suchverfahren in Bildsuchmaschinen eingeführt werden sollen“ wurde mit einer deutlichen Mehrheit von 71,5 Prozent der Teilnehmer mit ja beantwortet.

Nutzung des Ontologiebrowsers. Die abschließend letzte Aufgabe dieses Teilbereichs bestand darin, Bilder von Manuel Barroso über den Ontologiebrowser zu finden. Im ImageNotion Prototyp war dafür der integrierte Ontologiebrowser aufzurufen und zum semantischen Element „Manuel Barroso“ zu navigieren (durch Absteigen in der Hierarchie). Anschließend kann das semantische Element angeklickt und damit direkt eine semantische Suche gestartet werden. 64,2 Prozent der Benutzer konnten diese Form der semantischen Suche erfolgreich durchführen. Allerdings scheint diese Form der Suche nicht gewollt zu sein, denn nur 37,2 Prozent der Benutzer fanden diese Form der Suche für hilfreich zur Bildsuche. Dies deckt sich mit der Aussage von Vallet [VF05], dass die Benutzer das Starten einer Suchanfrage mit einem Textfeld bevorzugen. *Entsprechend wird es als nicht nötig betrachtet, sich weiter mit dem Ontologiebrowser als Startpunkt einer semantischen Suche zu befassen.*

Fazit

Hollink hat in ihrer Untersuchung zur Erstellung semantischer Bildbeschriftungen *gefolgert*, dass die semantische Bildsuche von Benutzern verstanden wird. Eine empirische Untersuchung zur Anwendung der Bildsuche selbst hat sie aber nicht durchgeführt. Der erste Frageteil dieser Evaluation hat sich die Beantwortung dieser Frage zum Ziel gesetzt. Die Ergebnisse der gestellten Suchanfragen und die überwiegend hohen Prozentzahlen zeigen, dass die Frage, ob Verfahren zur semantischen Bildsuche verstanden und gewollt werden, klar mit ja beantwortet werden kann. Mit der Zusammenfassung der Ergebnisse kann die Frage 1 (c) dieser Arbeit somit erfolgreich mit ja beantwortet werden:

Frage 1 (c)

Sind die entstehenden Möglichkeiten zur Bildsuche für Nutzer in der Praxis verständlich und wie sollen diese eingesetzt werden?

Antwort

Benutzer verstehen die Idee der semantischen Bildsuche, können semantische Suchanfragen formulieren und nachvollziehen und empfanden weiterhin die entstehenden Ergebnisse als besser im Vergleich zum textbasierten Suchergebnis. Als Verfahren zur Formulierung von Suchanfragen sind *Vorschlag semantischer Elemente* und *Auflösung einer semantischen Suchanfrage* zu verwenden.

6.4.3 Neuformulierung von Suchanfragen

Der zweite Bereich der Fragen betraf (Ziel 2), also „Können die Benutzer mit semantischen Verfahren zur Neuformulierung von Suchanfragen umgehen?“.

Tabelle 6.2 gibt einen Überblick über die gestellten Fragen und Aufgaben in Bereich 2.

Zusammenfassung von semantischen Elementen: Die erste Aufgabe in diesem Bereich lautete „Suchen Sie nach Barroso. Schränken Sie anschließend mit einem geeigneten semantischen Element das Suchergebnis so ein, dass Barroso und Merkel zusammen angezeigt werden“. In dieser Aufgabe wurde es als nicht richtig gewertet, wenn aus den Log Dateien ersichtlich war, dass Benutzer eine textbasierte Suchanfrage „Merkel Barroso“ gestartet haben. Als richtig wurde gewertet, wenn im Bereich „Mögliche Neuformulierungen der Suchanfrage“ eines Suchergebnisses das dort angezeigte semantische Element „Angela Merkel“ verwendet und angeklickt wurde. Anschließend wurde das gewünschte Suchergebnis erstellt. 63.6 Prozent der Benutzer konnten diese Suchaufgabe erfolgreich umsetzen. 54,9 Prozent der Benutzer

Frage	Ja (%)	Nein (%)
Benutzer, die Suchanfrage „Barroso“ einschränken konnten mit den semantischen Elementen <i>Manuel Barroso</i> und <i>Angela Merkel</i> ?	63,5	36,5
Ist die Zusammenfassung semantischer Elemente hilfreich zur Neuformulierung von Suchanfragen?	54,7	45,3
Benutzer, die Suchanfrage „Merkel“ neuformulieren konnten hin zu verwandten Bilder des „EU Councils Kopenhagen“	59,1	41,9
Soll bei der Neuformulierung zu verwandten Themen die bestehende Suchanfrage mit berücksichtigt werden	88,7	11,3
Ist der Vorschlag semantisch verwandter Themen hilfreich zur Neuformulierung von Suchanfragen	72,1	28,9

Tabelle 6.2: Ergebnisse für den Aufgabenbereich „Neuformulierung von Suchanfragen“

hielten diese Form der Neuformulierung von Suchanfragen für hilfreich.

Vorschlag verwandter Suchanfragen: Die zweite Aufgabe in diesem Bereich war die Suche nach „Angela Merkel“ und der anschließenden Suche nach Bildern des „EU Council Kopenhagen“ durch Verwendung eines geeigneten semantischen Elements. Im Bereich „Verwandte Suchanfragen“ wird hierbei im Suchergebnis bei der Suche nach „Angela Merkel“ das semantische Element „EU Council Kopenhagen“ in ImageNotion angezeigt. Der Klick auf dieses semantische Element wurde als richtig für diese Aufgabe bezeichnet. 59,2 Prozent der Benutzer konnten diese Aufgabe erfolgreich durchführen. Die Frage, ob beim Starten der Suchanfrage dann die bestehende Suchanfrage mit berücksichtigt werden soll, beantworteten von diesen Benutzern 88,7 Prozent mit ja. Die gewünschte Anfrage lautet dann also „Angela Merkel auf dem EU Council Kopenhagen“. 72,1 Prozent dieser Benutzer halten dieses Verfahren für hilfreich zur Neuformulierung von Suchanfragen.

6.4.4 Verständnis für Relationen

Motiviert vom Vorgehen Hollinks, aus einer Aufgabenstellung Rückschlüsse zu ziehen, war ich interessiert an der Frage, ob Benutzer ein Verständnis für Relationen haben. Die Annahme hierbei ist, dass Relationen auch sinnvoll in der Bildsuche mit ausgenutzt werden können, wenn die Benutzer Relationen zu einem semantischen Element formulieren können.

Semantisches Element	Prozent
Portugal	16
EU commission	14
Politician	14
President	9
EU	6
European commission	6
man	4
person	4
European Union	3
Italy	3

Abbildung 6.5: Top 10 für zu verwendende semantische Elemente

Relation	Prozent
is president of	24
works for	8
has nationality	6
is born in	6
is member of	6
has position	4
is head of	4
lives in	4
has meeting with	2
has profession	2

Abbildung 6.6: Top 10 für benannte Relationen

Entsprechend bekamen die Benutzer im dritten Teil der Befragung die folgende Aufgabe gestellt: „Lesen Sie den Wikipedia Artikel zu Manuel Barroso³. Listen Sie anschließend auf, mit welchen semantischen Elementen Sie diesen Politiker beschreiben würden und welche *benannten Relationen*⁴ Sie hierbei erstellen würden.“ Die Relationen sollten von den Benutzern selbst formuliert werden, eine Auswahlliste an vorgegebenen Relationen lag hierbei nicht vor.

Insgesamt gaben die Benutzer 98 Nennungen zu semantischen Elementen, die sie verwenden würden. Weiterhin wurden 76 benannte Relationen genannt, welche die Benutzer verwenden würden. Tabelle 6.6 listet die hierbei jeweils am meisten genannten Nennungen auf. Die semantischen Elemente betrafen oft den Beruf von Manuel Barroso (z.B. „Politician“, „President“, „EU commission“) sowie seine Nationalität („Portugal“, „Italien“ (wobei Italien falsch ist, Barroso ist Portugiese). Entsprechend wurden auch Relationen benannt, um die Rolle von Barroso in der EU Kommission genauer zu beschreiben („is head of“, „is member of“, „has profession“) sowie der Nationalität („lives in“, „has nationality“, „is born in“).

Aus diesem Ergebnis kann man schließen, dass die Benutzer in der Lage sind, semantische Elemente und Relationen zu formulieren, die geeignet wären, um dieses Hintergrundwissen in Ontologien abzubilden. Allerdings muss erwähnt werden, dass nur von 56 Prozent der Teilnehmer die Frage beantwortet wurde. Die anderen Teilnehmer haben die Frage übersprungen. So kann es auch sein, dass die meisten Nennungen von den Benutzern erstellt wurden, die bereits Erfahrungen mit semantischen Anwendungen haben (siehe Abbildung 6.4).

³<http://de.wikipedia.org/wiki/JoseManuelDuraoBarroso>

⁴Eine entsprechende Einleitung zu Relationen wurde im Text gegeben

Die Frage, ob Benutzer ein Verständnis für Relationen haben, wird daher unter Vorbehalt mit ja beantwortet, denn die Aussage konnte in der Evaluation nicht auf alle Teilnehmer verallgemeinert werden.

6.4.5 Navigation durch Bildteile

Der vierte und letzte Teil der Befragung betraf (Ziel 4), „Wie soll die Navigation durch Bildteile funktionieren?“. Tabelle 6.3 gibt einen Überblick über die gestellten Fragen dieses Bereichs.

Frage	Ja (%)	Nein (%)
Benutzer, die Suchanfrage „Angela Merkel“ durch Klick auf einen Bildteil, der sie abbildet, ausführen konnten	76,6	23,4
Halten Sie beschriftete Bildteile für sinnvoll?	69,3	30,7
Fänden Sie eine große Menge an beschrifteten Bildteilen auf einem Bild als störend?	53,3	46,7

Tabelle 6.3: Ergebnisse für den Aufgabenbereich „Neuformulierung von Suchanfragen“

Als Suchaufgabe zur Navigation durch Bildteile sollten die Benutzer wiederum nach „Barroso“ suchen. Anschließend sollten sie ein beliebiges Bild im Suchergebnis, welches Barroso zusammen mit Angela Merkel zeigt öffnen, um auf dem Bild selbst eine Suchanfrage nach „Angela Merkel“ durchzuführen. 76,6 Prozent der Benutzer und damit mehr als zwei Drittel der Teilnehmer konnten diese Suche erfolgreich durchführen. Die Frage, „ob die Beschriftung von Bildteilen für sinnvoll gehalten wird“, wurde von fast 70 Prozent mit ja beantwortet.

Anschließend bekamen die Teilnehmer die Frage gestellt, was ihrer Meinung nach beim Klick auf einen Bildteil passieren soll. Zur Auswahl standen (1) eine neue Suche starten, (2) eine boolesche Suche unter Berücksichtigung der aktuellen Suchanfrage und dem geklickten Bildteil durchführen, (3) eine gewichtete Suche durchführen, welche den Bildteil bevorzugt und (4) eine gewichtete Suche durchführen, welche die bestehende Anfrage bevorzugt.

Abbildung 6.7 zeigt die Auswertung zu dieser Frage. In der Tat ist für etwa die Hälfte der Teilnehmer das aus der Literatur bekannte und bestehende Verfahren (1), also eine komplett neue Suche zu starten, das richtige Verfahren (46,1 Prozent). 4,4 Prozent der Benutzer gaben an, statt einer Suche lieber Details zum gezeigten Bildteil zu erhalten. Die andere Hälfte der Teilnehmer mit 50,1 Prozent bevorzugten das Starten einer kombinierten Suche mit den Varianten (2), (3) und (4).

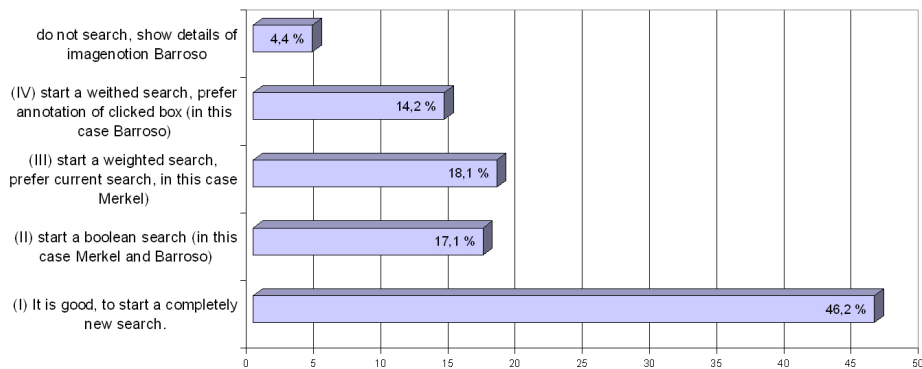


Abbildung 6.7: Gewünschte Formulierung einer Suchanfrage auf einem Bildteil

Die abschließende Frage war, „ob eine große Menge an beschrifteten Bildteilen als störend empfunden wird“. 53,3 Prozent der Teilnehmer haben diese Frage mit ja beantwortet. Entsprechend wäre es sinnvoll, Mechanismen zu haben, welche die Beschriftungen beispielsweise nach dem Thema der aktuell gestellten Suchanfrage des Benutzers filtern könnten.

6.5 Resultierender Aufbau der Komponente Bildsuche in ImageNotion

Auf Basis der gewonnenen Ergebnisse der Onlineevaluation kann nun der genaue Aufbau der Komponente *Bildsuche* in Systemarchitektur von ImageNotion (siehe Sektion 3.4) beschrieben werden. Tabelle 6.4 gibt einen Überblick der resultierenden Komponente

Eingabe Suchanfrage: Da von jeweils mehr als 70 Prozent der Bildsucher die Verfahren zur Eingabe und Umsetzung von semantischen Suchanfragen verstanden wurden, können Verfahren zum *Vorschlag semantischer Elemente* und *Auflösung einer semantischen Suchanfrage* umgesetzt werden. Der Ontologie Browser wurde nur von etwa 35 Prozent der Benutzer als hilfreich für die Bildsuche eingestuft und wird daher nicht weiter bezüglich der Bildsuche in ImageNotion ausgebaut.

Neuformulierung von Suchanfragen: Aus den Ergebnissen zur Neuformulierung einer Suchanfrage kann somit auch der genaue Aufbau der Komponente „Neuformulierung Suchanfrage“ der gezeigten Systemarchitektur in Sektion 3.4 genannt werden. Die Komponente soll die beiden in der Evaluation verwendeten Verfahren *Zusammenfassung semantischer Elementen* und *Vorschlag verwandter Suchanfrage*

Verfahren	Ergebnis
Eingabe semantischer Suchanfragen	
Vorschlag semantischer Elemente	Überwiegend von Benutzern verstanden, somit umsetzbar
Auflösung einer semantischen Suchanfrage	Von 72 Prozent der Benutzern verstanden, somit umsetzbar
Suche mit dem Ontologiebrowser	Nur von 37 Prozent als hilfreich bewertet, somit Umsetzung nicht hilfreich für Benutzer bei der Bildsuche
Neuformulierung von Suchanfragen	
Zusammenfassung von semantischen Elementen	Von 63 Prozent der Benutzer verstanden, somit umsetzbar
Vorschlag verwandter Suchanfragen	Von 59 Prozent der Benutzer verstanden, somit umsetzbar
Verständnis für Relationen	
Verständnis für Benutzer für Relationen	Ja, für etwa 50 Prozent der Benutzer
Umfangreicher Einsatz von Relationen bei der Bildsuche	Ausblick
Navigation durch Bildteile	
Starten einer Suchanfrage mit Klick auf Bildteil	Von 76 Prozent der Benutzer verstanden, somit umsetzbar
Art der zu verwendenden Suche bei Klick auf Bildteil	Sowohl das Starten einer neuen Suchanfrage wie auch die Berücksichtigung der aktuellen Suchanfrage wurde von Benutzern gewünscht. Daher werden in ImageNotion beide Verfahren umgesetzt
Kontext der Suchanfrage selbst und Anfragen einer Sitzung berücksichtigen	
Ordnung der Suchergebnisse basierend auf dem Kontext der bisherigen Sitzung	Ausblick
Filterung der semantischen Metadaten basierend auf Suchanfrage	Ausblick

Tabelle 6.4: Umsetzung der Komponente Bildsuche in ImageNotion

gen enthalten, da sie von der Mehrheit der Benutzer angewandt und als hilfreich eingestuft wurden.

Da die *Navigation durch Bildteile* ebenfalls überwiegend durch die Benutzer verstanden wurde, kann auch diese Form der Neuformulierung von Suchanfragen in ImageNotion übernommen werden. Die Umsetzung der Suchanfrage kann, wie durch die Benutzer gewünscht, entweder durch Starten einer neuen Suchanfrage mit den semantischen Metadaten des Bildteils erfolgen oder auch zusätzlich die semantischen Elemente der aktuellen Suchanfrage hinzunehmen.

Verfahren, die weitere empirische Untersuchungen benötigen

Wie die Onlineevaluation gezeigt hat, besteht für etwa 50 Prozent der Benutzer ein generelles Verständnis für Relationen. Wie diese Relationen bei der Bildsuche genau eingesetzt werden können und durch Benutzer verstanden werden, muss aber durch zusätzliche Konzeptionen und entsprechende Evaluationen ermittelt werden. Diese Fragen stellen somit eine neue Aufgabenstellung und verbleiben als Ausblick.

Etwa 50 Prozent der Benutzer gaben an, dass sie eine große Menge an Bildteilen auf Bildern stören würde. Dies erfordert Mechanismen, die beispielsweise aufgrund der gegebenen Suchanfrage die Bildteile filtern, z.B. nur Beschriftungen für Bildteile von Personen anzuzeigen, wenn nach einer Person gesucht wird. Hierfür muss also der Kontext der Suchanfrage ausgenutzt werden. Eine entsprechende Nutzung des Kontext der Suchanfrage und dem Kontext der bestehenden Suchanfrage der aktuellen Sitzung zur Ordnung der Suchergebnisse sind weitere, kontextabhängige Fragen. Die Konzeption derartiger Kontexte und entsprechende Evaluationen bilden ebenfalls eine eigene Aufgabenstellung und verbleiben daher als Ausblick.

6.6 Zusammenfassung

Mit Hilfe der Onlineevaluation konnte die Frage 1 (c) in dieser Arbeit mit ja beantwortet werden. Benutzer verstehen die Idee der semantischen Suche, können semantische Suchanfragen formulieren und nachvollziehen und empfanden weiterhin die entstehenden Ergebnisse als besser im Vergleich zum textbasierten Suchergebnis.

Die Komponente zur Neuformulierung von Suchanfragen soll erstellt werden, indem die Verfahren *Zusammenfassung semantischer Elemente* und *Vorschlag verwandter Suchanfragen* dort integriert werden. Ein generell bestehendes Verständnis für Relationen besteht bei den Benutzern, so dass diese bei der Erstellung von Hintergrundwissen mit berücksichtigt werden sollten. Die Benutzer haben weiterhin die Beschriftung von Bildteilen als wichtig angesehen und konnten die Navigation durch Bildteile anwenden.

Die letzten beiden Kapitel haben sich damit beschäftigt, ob semantische Verfahren zur Bildsuche überhaupt benötigt werden und entsprechende Verfahren zur semantischen Bildsuche überhaupt verstanden werden. Beide Fragen konnten mit ja beantwortet werden. Daher geht es innerhalb der nächsten Kapitel darum, wer überhaupt in welcher Art und Weise das nötige Hintergrundwissen in Ontologien erstellt, um semantische Metadaten erstellen zu können. Denn *nur mit Hilfe von Ontologien und semantischen Metadaten kann eine semantische Bildsuche überhaupt stattfinden und somit zu hochwertigen Suchergebnissen in der Bildsuche führen.*

Kapitel 7

Erstellung semantischer Metadaten

Das Ziel der folgenden Kapitel ist die Beantwortung der gestellten *Frage 2* in dieser Arbeit (siehe Sektion 3.6). Die generelle Frage dabei lautet *Wie und von wem werden semantische Hintergrundinformationen erstellt?*, also Ontologien, die zur Erstellung semantischer Metadaten benötigt werden.

Die Motivation dieser Frage liegt in der von Hepp getroffenen Feststellung „Der Erfolg einer semantischen Anwendung hängt maßgeblich von der *Umsetzbarkeit* und der *Akzeptanz* durch die Benutzer ab [Hep07].“. Die *Akzeptanz* semantischer Verfahren durch Bildsucher wurde bereits im vorigen Kapitel erfolgreich gezeigt. Eine weitere Gruppe muss diese Verfahren akzeptieren, dies sind die *Bildarchivare*. Diese benötigen zur Erstellung von semantischen Metadaten für hochwertige Suchergebnisse entsprechende Ontologien und müssen diese benutzen können.

In diesem Kapitel wird daher zuerst ermittelt, welches gängige Verfahren zur Erstellung einer semantischen Anwendung sich anbietet. In einem Workshop mit Fachanwendern soll ermittelt werden, ob dieses Verfahren durch Bildarchivare umsetzbar ist und es akzeptiert wird oder Erweiterungen und Änderungen in diesem Verfahren nötig sind.

7.1 Methodologien zur Ontologieerstellung

Hintergrundwissen über Dinge einer Domäne wird in eine Ontologie abgebildet (siehe Sektion 3.2). Entsprechend müssen Ontologien erstellt werden. *Methodologien* sollen die dabei nötigen Abläufe vorgeben.

7.1.1 Begriff: Methodologie

Eine Methodologie definiert hinsichtlich einer Ontologie einen Satz von Methoden zur Erstellung und Pflege [SSSS01]. Die gängigen Methodologien definieren Schritte zur *Erstellung einer Ontologie*, *Nutzung der Ontologie* und *Pflege der Ontologie* als iterative Prozesse. Ontologien sollen damit über die Zeit verbessert werden.

7.1.2 Oft verwendete Methodologien

Die Methodologien *On-to-Knowledge* und *METHONTOLOGY* sind aus der Literatur bekannte und oft verwendete Methodologien zur Ontologieerstellung [FLGP02]. Deren wichtigste Schritte und Ziele werden im Folgenden dargestellt. Die Details der Methodologien finden sich unter [FLGP02].

On-to-Knowledge

Ziel von On-to-Knowledge [FLGP02] ist die Erstellung von Ontologien für Anwendungen, die auf semantischen Verfahren basieren. Die wichtigen Hauptschritte von On-to-Knowledge sind *Refinement* und *Application*.

Während der Phase *Refinement* werden entweder initial oder in weiteren iterativen Schritten semantische Elemente ermittelt, die für die spätere Anwendung benötigt werden. Im Kontext von Bildsuchmaschinen sind dies alle Dinge und Sachverhalte, die benötigt werden, um anschließend geeignete semantische Metadaten für die Bilder im Archiv erstellen zu können. Mit der Phase *Application* wird die Ontologie in die Anwendung integriert und dort genutzt.

Iterative Schritte aus *Refinement* und *Application* sollen die Ontologie auf sich ändernde Sachverhalte in der Anwendung anpassen. Es werden also fehlende semantische Elemente ermittelt und der Ontologie hinzugefügt.

METHONTOLOGY

METHONTOLOGY [FLGP02] besteht aus denselben Grundsritten wie On-to-Knowledge. Zusätzliches Ziel von METHONTOLOGY ist der Fokus auf der *gemeinsamen Begriffsbildung* (conceptualization). Damit ist gemeint, dass Begriffe gebildet werden sollen, die von den Nutzern der Ontologien einheitlich verstanden werden. Entsprechend müssen diese zusätzlich dokumentiert werden. METHONTOLOGY eignet sich daher zur Erstellung von Ontologien, die auch über eine größere Benutzergruppe einheitlich verstanden werden sollen.

7.1.3 Standardisierte Ontologien

Ein einheitliches Verständnis zu einer Ontologie über eine große Zahl von Benutzern ermöglicht die *Standardisierung von Ontologien*. Die Ontologie CIDOC-CRM [CDG⁺05] beispielsweise erhält mehr als 500 Konzepte zur Beschreibung von Bild- und Museumsinhalten. Der Vorteil standardisierter Ontologien ist, dass dann Informationen einheitlich über eine Vielzahl von Systemen eingesetzt werden können. Werden semantische Metadaten mit einer standardisierten Ontologie erstellt, können die damit verbundenen semantischen Metadaten über die Grenzen einer Anwendung hinweg ausgetauscht werden. So entsteht *semantische Interoperabilität* über mehrere Systeme [WN07b]. Da Bildagenturen oft mit weiteren Agenturen (z.B. im Ausland) zusammenarbeiten und sich gegenseitig im jeweiligen Land vertreten, ist diese Möglichkeit des einheitlichen Austauschs von Daten wünschenswert für deren Betreiber.

Bekannte und oft verwendete standardisierte Ontologien, die sich für die Erstellung von semantischen Metadaten eignen [WN07b], sind das eben erwähnte CIDOC-CRM sowie MPEG-7 [Mar05].

7.1.4 Fazit

Gängige Methodologien zur Ontologierstellung sehen somit *eine getrennte Ontologierstellung* bzw. Pflegeprozess vor, der *von der eigentlichen Nutzung in der Anwendung getrennt* ist.

Experten zur Erstellung von Ontologien kümmern sich somit um die *Inhalte der Ontologien*, während in der Anwendung zum Beispiel Bildarchivare diese vorgefertigten Ontologien *zur Erstellung semantischer Metadaten benutzen*.

Durch die Einigung auf die Inhalte über eine Gruppe hinweg können standardisierte Ontologien entstehen. Der Vorteil von standardisierten Ontologien ist, dass Anwendungen ihr Hintergrundwissen untereinander einheitlich austauschen können.

7.2 Empirische Studie 2: Erstellbarkeit semantischer Metadaten mit vorgegeben Ontologien

Aufgrund der gezeigten Methodologien ist der gängige Aufbau einer semantischen Anwendung, eine vorab erstellte Ontologie in die Anwendung zu integrieren. Für eine Bildsuchmaschine bietet es sich dabei an, eine standardisierte Ontologie zu verwenden.

Mit Hilfe von Bildarchivaren soll innerhalb eines Workshops ermittelt werden,

ob der Ansatz zur getrennten Ontologieerstellung und Nutzung von vorgefertigten Ontologien praktisch umsetzbar ist. Es soll daher die gestellte *Frage 2 (a)* dieser Arbeit beantwortet werden, also *Sind getrennt erstellte, vorgefertigte Ontologien ausreichend und praktisch anwendbar?* Falls dieser Ansatz nicht praktisch umsetzbar ist, müssen alternativ Verfahren ermittelt werden, bei denen die Benutzer selbst die notwendigen Ontologien erstellen können.

7.3 Aufbau und Durchführung des Workshops

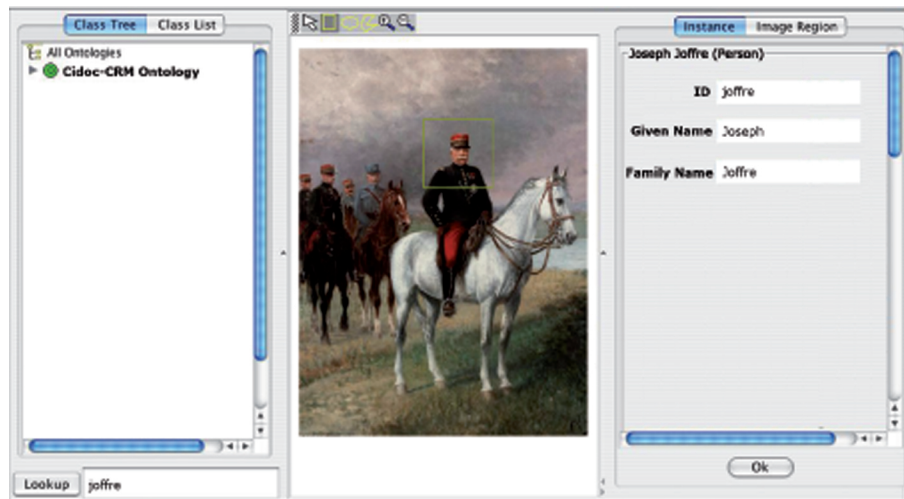


Abbildung 7.1: Semantische Bildbeschriftung mit PhotoStuff

Für den Workshop wurde eine Bildsuchmaschine aufgebaut (im Folgenden als Testsystem“ bezeichnet), welche die Erstellung von semantischen Metadaten ermöglicht. Hierzu wurde die Software PhotoStuff [vOTSP07] ausgewählt. PhotoStuff erlaubt die Erstellung semantischer Bildbeschriftungen für Bildteile und dem ganzen Bild. In das Testsystem wurden 622 Bilder aus der *Domäne Erster Weltkrieg* integriert, die von Bildagenturen aus dem IMAGINATION Projekt zur Verfügung gestellt wurden. Weiterhin wurde die Ontologie CIDOC-CRM importiert. Abbildung 7.1 zeigt ein Beispiel, hierbei wird ein Bildteil, welcher Joseph Joffre¹ zeigt, mit der entsprechenden Instanz *Joseph Joffre* des Konzepts *Person* erstellt.

6 Teilnehmer wurden zum Workshop eingeladen. Bei den Teilnehmern handelt es sich um *Fachbenutzer*, sie sind Angestellte in Bildagenturen und ihre Aufgabe dort ist die Erstellung von Bildbeschriftungen.

¹ein französischer General aus dem ersten Weltkrieg

Die Aufgabe der Teilnehmer war die Erstellung semantischer Metadaten für die vorhandenen Bilder im Testsystem. Hierzu wurde in einem Vortrag zuerst erklärt, wie die Anwendung PhotoStuff funktioniert und es wurden Beispiele gegeben, wie semantische Metadaten mittels der Verwendung von CIDOC-CRM erstellt werden können.

Benutzer	1	2	3	4	5	6
Beruf	Archivar	Archivar	Geschäftsführer Bild-agentur	Bibliothekar	Bibliothekar	Bibliothekar
Erfahrung im Umgang mit Computern?	Hoch	Normal	Normal	Hoch	Normal	Normal
Erfahrung mit Verfahren zur Bildbeschriftung?						
Textbasierte Verfahren	IPTC	IPTC	IPTC	IPTC, Dublin Core	Dublin Core	Dublin Core
Thesauri	Ja	Nein	Nein	Ja	Ja	Ja
Ontologien (Taxonomien und Instanzenmodelle)	Nein	Nein	Nein	Nein	Nein	Nein

Tabelle 7.1: Erfahrungen der Benutzer mit Verfahren zur Bildbeschriftung

Zu Beginn des Workshops wurden die Teilnehmer hinsichtlich ihres Berufs und ihren bisherigen Erfahrungen mit der Bildbeschriftung unter der Verwendung von textbasierten und semantischen Verfahren befragt. Tabelle 7.1 belegt zum Einen, dass es sich bei allen Teilnehmern um erfahrene Benutzer der Bildbeschriftung handelt. Die ersten beiden Teilnehmer sind Bildarchivare, Teilnehmer 3 ist Geschäftsführer der Bildagentur photo12 (Projektpartner IMAGINATION). Die Teilnehmer 4 bis 6 sind Bibliothekare des IMAGINATION Projektpartners BSMC². Alle Teilnehmer geben normale bis hohe Erfahrungen im Umgang mit Computern an, womit generelle Bedienungsprobleme im Workshop nicht zu erwarten sind. Alle Teilnehmer sind erfahren in der Nutzung textbasierter Verfahren zur Erstellung von Metadaten. In Bibliotheken scheint der Einsatz von Thesauri üblich, da alle drei Bibliothekare Erfahrungen im Umgang mit Thesauri haben. Mit Taxonomien und Instanzenmodellen hatte keiner der Teilnehmer Erfahrungen.

²biblioteca storia moderna, Rom, www.bsmc.it

Der Workshop bestand aus einer zweistündigen *Beobachtung der Teilnehmer* mit der Arbeit von PhotoStuff und einer anschließenden *Diskussion* mit den Teilnehmern.

7.4 Fragen und Beobachtungen während des Workshops

Die folgenden Fragen und Probleme traten während des Workshops mit dem Testsystem auf:

- **(F1) Wozu braucht man die ganzen Konzepte in CIDOC-CRM?** Ein Teilnehmer fragte, wozu man das Konzept „E81 Information Carrier“ benötigt. Ihm wurde erklärt, dass dieses Konzept zur Beschreibung, auf welchem Medium etwas gespeichert ist (z.B. CD-ROM) dient. Der Teilnehmer gab an, dass „ein solches Konzept bei Bildagenturen nicht gebraucht werde, da die Bilder immer im Archiv liegen würden“. Viel wichtiger sei angeben zu können, „welcher Agentur das Bild gehöre“.
- **(F2) Wie können die Bildteile wie gewünscht beschrieben werden?** Ein Teilnehmer wollte das Bild in Abbildung 7.1 beschreiben. Hierbei gab er an, dass er ausdrücken möchte „Joseph Joffre ist ein französischer General“ sowie „der Bildteil, welcher das Pferd zeigt, solle ebenfalls beschrieben werden“. Geeignete Konzepte für „General“ und „Pferd“ waren aber in der Ontologie nicht enthalten.
- **(F3) Wann ist ein Element als Konzept und wann als Instanz sinnvoller?** Die meisten Teilnehmer stellten die Frage, was der Unterschied zwischen einem Konzept und einer Instanz sei. Konkret machten sich zwei Teilnehmer Gedanken, wie sie denn das semantische Element „Pferd“ erstellen könnten. Ist es eine „Instanz von *Säugetier* oder selbst ein Konzept, da es unterschiedliche Pferderassen gibt?“.

Die folgenden Beobachtungen wurden während des Workshops gemacht:

- **(B1) Schwierigkeiten mit der Anwendung PhotoStuff:** PhotoStuff ist ein Demonstrator für die semantische Bildbeschriftung des W3C. Entsprechend fehlerbehaftet ist die Anwendung. Die Anwender störten sich über häufige Fehlermeldungen bzw. es traten auch Abstürze auf. Weiterhin empfanden alle Anwender die Benutzung der Anwendung zum Beschriften von Bildteilen als sehr kompliziert.

- **(B2) Langes Suchen passender Konzepte:** Die Benutzer verbrachten die überwiegende Zeit mit dem Suchen von passenden Konzepten. Sie waren entsprechend über fehlende Konzepte verärgert oder wunderten sich über deren Einordnung in der Ontologie, z.B. „Warum ist Person ein Unterkonzept von „E 20: Biological Object“ und nicht von „Säugetier“.
- **(B3) Es wurden wenig Bilder beschrieben:** Entsprechend dem hohen Zeitaufwand für die Suche und das Nachdenken über Konzepte wurden in der Zeit des Workshops sehr wenig Bilder beschrieben, gerade einmal 5 Stück (und diese unter Anleitung durch Betreuer des Workshops).
- **(B4) Workshop-Teilnehmer resignierten:** Drei der sechs Teilnehmer des Workshops gaben nach einer Stunde ihre Versuche auf, die Bilder zu beschriften und surfen stattdessen im Internet. Auf entsprechende Anfragen gaben Sie an „das ist mir zu kompliziert“, „ich verstehe das nicht“, „ich kann die Bilder nicht beschriften wie ich will“.

7.5 Diskussion mit den Teilnehmern

Der zweite Teil des Workshops bestand aus einer Diskussionsrunde mit den Betreuern und allen Teilnehmern. Die Diskussion wurde moderiert. Die erste Frage hierbei lautete: „Wie ist ihre Meinung zu dem vorgestellten Aufbau des Testsystem zur Erstellung semantischer Metadaten“. Die folgenden Antworten wurden gegeben:

- **Anwendung sehr kompliziert.** Benutzer 4 gab an, dass die „Bedienung der Anwendung sehr kompliziert sei“. Dies entspricht der Beobachtung *B1* während des Workshops.
- **Ontologie oft unvollständig.** Benutzer 2 gab an, dass „die Ontologie oft unvollständig war“. Dies belegt die Beobachtung *B2*. Es wurde sehr lange nach passenden Konzepten gesucht ohne diese zu finden. Entsprechend gab Benutzer 2 an, (und griff damit Frage *F2* weiter auf), dass „die Ontologie benötigte Elemente zur Erstellung von Bildbeschriftungen nicht enthielt“. Die Bildbeschriftung konnte so nicht wie gewünscht durchgeführt werden.
- **Ontologie enthält viele nicht benötigte Konzepte** Benutzer 1 gab an, dass die Ontologie für ihn „nicht nachvollziehbare Konzepte“ enthielt. Er griff damit die während des Workshops gestellte Frage *F1* weiter auf.
- **Bildbeschriftung so nicht möglich:** Benutzer 3 gab an, dass „mit dem gezeigten Verfahren die Erstellung von Bildbeschriftungen praktisch nicht möglich ist“.

- **Benutzung wurde aufgegeben:** Benutzer 4 gab noch einmal an (entsprechend der Beobachtung B_4), dass er „die Benutzung des Testsystems aufgegeben hat, da es ihn zu anstrengend und zu kompliziert gewesen sei, damit Bildbeschriftungen zu erstellen“.
- **Komplizierter Formalismus:** Benutzer 1 griff noch einmal seine während des Workshops gestellte Frage F_3 auf und gab an, dass er „die Trennung und Bedeutung von Konzept und Instanz nicht verstanden habe“.

Die zweite Frage vom Moderator wurde gestellt als „Wie ist ihre Meinung zur vorgeschlagenen Nutzung vorgefertigter Ontologien zur Erstellung semantischer Metadaten?“. Die folgenden Antworten wurden gegeben:

- **Verfahren ist teuer.** Benutzer 3 gab an, dass ihm die getrennte Ontologieerstellung als sehr teuer erscheint, denn da „müsste man ja weiteres Personal beschäftigen oder beauftragen für die Erstellung von Ontologien“.
- **Umgang mit fehlenden Elementen?** Benutzer 6 fragte, „wie dann mit fehlenden Elementen umgegangen werden soll, also wie dann die Bilder beschriftet werden sollen“. Als Antwort wurde gegeben: „In der Tat müssten die Benutzer bei fehlenden Elementen immer warten, bis die Ontologie aktualisiert wird!“
- **Bildbestände erweitern sich kontinuierlich.** Auf die gegebene Antwort kam die weitere Frage von Benutzer 3 „unsere Bildbestände ändern sich kontinuierlich, oft kommen neue Themen hinzu“. Also Antwort wurde gegeben: „Ja, dies würde sehr oft die Erweiterung der Ontologie erfordern“.

Als dritte Frage wurde den Benutzern die Frage gestellt, „Halten Sie den getrennten Ansatz der Ontologieerstellung und Nutzung der Ontologie durch Bildarchive zur Bildbeschriftung für ausreichend und in der Praxis durchführbar?“. Die ziemlich klar von den Benutzern gegebenen Antworten lauteten

- **Getrennte Ontologieerstellung nicht ausreichend.** Die getrennte Ontologieerstellung ist „nicht ausreichend, da die Bilder nicht wie gewünscht beschriftet werden können“ (Benutzer 2).
- **Getrennte Ontologieentwicklung unpraktisch.** Getrennte Ontologieentwicklung erfordert „häufiges Warten auf Änderungen in der Ontologie, damit diese anschließend für die Bilder verwendet werden können“ (Benutzer 4).
- **Getrennte Ontologieerstellung zu umständlich und teuer.** Die getrennte Ontologieentwicklung ist zu „umständlich und teuer“ (Benutzer 3).

Aus diesen klar gegebenen Antworten kann man somit recht klar Frage 2 (a) dieser Arbeit beantworten wie folgt:

Frage 2 (a)

Sind getrennt erstellte, vorgefertigte Ontologien ausreichend und praktisch anwendbar?

Antwort

Nein, der getrennte Ansatz der Ontologieerstellung ist für Bildarchivare *weder ausreichend noch praktisch durchführbar*, denn wie die Antworten zeigen ist dieser Ansatz zu umständlich und unpraktisch und weiterhin zu teuer. Hinzu kommt, dass die Elemente einer getrennt erstellten Ontologie oft nicht verstanden werden oder Elemente vorhanden sind, die nicht benötigt werden.

Innerhalb der Diskussion stellte der Moderator somit eine weitere Frage, nämlich „Welche Vorschläge die Teilnehmer haben, um eine praktisch durchführbare Erstellung von semantischen Metadaten zu ermöglichen?“

Die Teilnehmer gaben die folgenden Antworten und Vorschläge:

- **(Idee 1) Initiale semantische Elemente erstellen.** Benutzer 2 gab an, dass es hilfreich gewesen wäre, wenn fehlende semantische Elemente initial hätten hinzugefügt werden können, zum Beispiel das Konzept „Pferd“ zur Beschriftung des in Abbildung 7.1 gezeigten Bilds. Er hätte dann nämlich „anschließend seine gewünschte semantische Beschriftung des Bildteils durchführen können“.
- **(Idee 2) Dokumentation der semantischen Elemente.** Benutzer 4 gab an, dass es hilfreich gewesen wäre zum Verständnis der Elemente in CIDOC-CRM, wenn eine Dokumentation vorgelegen wäre. Als Vorschlag wurde hierbei genannt, Links zu Wikipedia legen zu können
- **(Idee 3) Mehrsprachigkeit unterstützen.** Benutzer 3 gab an, dass es schön wäre, wenn die semantischen Elemente in mehreren Sprachen beschrieben werden könnten, um die mehrsprachige Bildsuche zu ermöglichen.
- **(Idee 4) Synonyme unterstützen.** Benutzer 3 gab weiterhin an, dass es schön wäre, wenn man zu den einzelnen semantischen Elementen synonyme Bedeutungen hinterlegen könnte
- **(Idee 5) Formalismus vereinfachen.** Benutzer 1 fragte, ob die „Unterscheidung von Instanzen und Konzepten entfallen könne“. Dies wurde als Antwort bejaht.
- **(Idee 6) Erweiterung bestehender Elemente ermöglichen.** Benutzer 2 gab an, dass auch „eine Erweiterung bestehender Elemente schön wäre“. So

hätte er es schaffen können auszudrücken, dass *Joseph Joffre* ein *General* im ersten Weltkrieg aus Frankreich war.

- **(Idee 7) Ontologieerstellung in die Anwendung zur Bildbeschriftung integrieren** Auf die Frage des Moderators, ob zur Ontologieerstellung eine weitere Anwendung erstellt oder genutzt werden soll, oder besser in die Anwendung zur Bildbeschriftung mit eingebunden sein soll, gab Benutzer 4 die Antwort „besser in der Anwendung selbst, so können die neu erstellten Elemente direkt zur Bildbeschriftung verwendet werden“. Benutzer 3 merkte an, dass „der Wechsel zwischen zwei Anwendungen und nötige Abgleich der Systeme zu zeitaufwändig sei“.

7.6 Resultierende Anforderungen

Aus den genannten Beobachtungen des Workshops und speziell aus den gegebenen Antworten und Ideen der Teilnehmer in der Diskussion lassen sich somit die folgenden Anforderungen ableiten, um eine praktisch nutzbare Erstellung semantischer Metadaten zu ermöglichen:

- **(OA1) Benutzer in Ontologieerstellung einbeziehen:** Aus (Idee 1) lässt sich fordern: Die Benutzer selbst müssen in die Ontologieerstellung mit integriert werden und eine zentralere Rolle in der Ontologieentwicklung einnehmen. Fehlende semantische Elemente sollten sie erstellen sowie bestehende Elemente überarbeiten können.
- **(OA2) Formalismus vereinfachen:** Aus (Idee 5) kann man fordern, dass die Unterscheidung von Instanzen und Konzepten auf der Benutzerschnittstelle verschwinden sollte³.
- **(OA3) Dokumentation der Elemente:** Neu soll hier sein basierend auf (Idee 2), dass zur Dokumentation Verknüpfungen zu Wikipedia gelegt werden, anstatt die Benutzer eigene Texte erstellen zu lassen.
- **(OA4) Schaffung eines arbeitsintegrierten Arbeitsablaufs:** Die Nutzung mehrerer Anwendungen, also einem getrennten Editor für die Ontologie als eigene Anwendung, sowie einer Anwendung zur Bildbeschriftung ist zu kompliziert und zeitaufwändig. Basierend auf (Idee 7) muss die Ontologieerstellung daher in den Ablauf zur Erstellung semantischer Metadaten selbst mit integriert werden.

³Dies ist formal möglich, weiterhin kann bei Exporten der Ontologie die Unterscheidung wieder hergestellt werden, indem man Blätter als Instanzen und Knoten als Konzepte ausgibt

- **(OA5) Flexible Formalisierung:** Entsprechend der (Idee 6) sollte es möglich sein, einen flexiblen Übergang von Thesauri hin zu Taxonomien und Instanzenmodellen zu haben, indem eine Hierarchisierung und das Hinzufügen von Relationen möglich ist. Weiterhin soll die einfachste Form sein, kontrollierte Vokabulare für Synonyme und Mehrsprachigkeit erstellen zu können, zur Umsetzung von (Idee 3) und (Idee 4). Mit anderen Worten, die einfachste Form der Ontologien, die erstellt werden können, sollen Thesauri sein.

7.7 Verwandte Untersuchungen

Verwandte Untersuchungen unter der Berücksichtigung der Belange von Benutzern zur Frage, *wie semantische Anwendungen erfolgreich für Benutzer umgesetzt werden können*, sind SOBOLEO und die konzeptionelle Betrachtung dieser Problematik von Hepp.

7.7.1 SOBOLEO: Taxonomien für Bookmarks

Ziel von SOBOLEO ist die Verwendung einer Taxonomie zur hierarchischen Kategorisierung von Bookmarks für Webseiten [ZBS09]. Als Hauptproblem bei der Erstellung der Anwendung werden genannt, „dass die Benutzer in Evaluationen nicht in der Lage waren, fehlende semantische Elemente zur Eingliederung neuer Bookmarks erstellen zu können“. Die Benutzer „brachen dann die Arbeit mit dem System ab und verwendeten textbasierte Bookmarksysteme“. Die semantische Anwendung SOBOLEO wurde somit von den Benutzern „nicht akzeptiert und die Anwendung war somit nicht geeignet umsetzbar.“

Entsprechend sollte es möglich sein, dass „die Benutzer selbst die fehlenden semantischen Elemente erstellen sollten und zwar direkt in der Anwendung, wenn diese zur Erstellung von Bookmarks benötigt werden“.

7.7.2 Konzeptionelle Betrachtung von Hepp

Hepp hat eine konzeptionelle Betrachtung der Probleme bei der Erstellung semantischer Anwendung durchgeführt [Hep07]. Als die gravierendsten Probleme gab er hierbei an: (1) Inhalte in der Domäne ändern sich schneller als die Inhalte in den Ontologien, (2) Ontologieerstellung ist teuer, (3) Hoher Kommunikationsaufwand von Benutzern mit Erstellern der Ontologien, (4) Verständnis für Ontologien und (5) Wem gehören die Ontologien?

Eine entsprechende Untersuchung mit Benutzern wurde hierbei nicht durchgeführt. Der durchgeführte Workshop in diesem Kapitel belegt aber speziell die Punkte (1), (2) und (4). Es haben oft Inhalte gefehlt und eine getrennte Ontologieerstellung durch Experten wurde von den Bildarchivaren als zu zeitaufwändig und teuer angesehen. Weiterhin wurden die Inhalte der verwendeten Ontologie CIDOC-CRM oft nicht verstanden.

Als einen möglichen Lösungsansatz gibt Hepp in seiner Arbeit an, dass die Benutzer in die Ontologieerstellung mit einbezogen werden sollten, um eine bessere Abbildung von benötigtem Hintergrundwissen in Ontologien zu ermöglichen. Dies führt zu einer Akzeptanz der semantischen Anwendung, da fehlende Inhalte in der Ontologie von den Benutzern hinzugefügt werden können, die Ontologieerstellung billiger wird, der Kommunikationsaufwand geringer wird und die Benutzer die Inhalte der Ontologie besser verstehen. Auf dieser Basis kann eine Umsetzbarkeit einer semantischen Anwendung erreicht werden. Eine entsprechende Methodologie zur Umsetzung dieses Lösungsansatzes wurde aber von Hepp nicht genannt.

7.7.3 Schlussfolgerungen

Aus der eigenen Untersuchung und den genannten verwandten Arbeiten lässt sich somit schlussfolgern, dass gerade in Domänen wie Bilder oder Bookmarks mit sich häufig ändernden Inhalten, *die Akzeptanz und der Erfolg semantischer Anwendungen von den vorhandenen Ontologien selbst abhängt.*

Aufgrund der hierbei ermittelnden Hauptprobleme, dass Ontologien zur Nutzung in der Anwendung entweder zu kompliziert oder unvollständig sind, sowie dem getrennten Arbeitsablauf zur Ontologieerstellung und Nutzung ergibt sich somit die Anforderung, dass Benutzer geeignet selbst in die Erstellung der Ontologien mit einbezogen werden.

7.8 Zusammenfassung

Der durchgeführte Workshop mit Bildarchivaren hat gezeigt, dass die bestehenden Methodologien zur Erstellung von Ontologien in der Praxis *nicht akzeptiert werden und nicht anwendbar sind.* Frage 2 (a) in dieser Arbeit musste daher mit nein beantwortet werden. Vorgefertigte Ontologien reichen nicht aus, um eine umfangreiche und gewünschte Erstellung semantischer Metadaten durch Bildarchivare zu ermöglichen!

Entsprechend den eigenen Ergebnissen und den Ergebnissen der verwandten Arbeiten entsteht nun somit die Forderung, dass Benutzer in die Erstellung von On-

tologien mit einbezogen werden. Hinsichtlich der Erstellung semantischer Metadaten konnten Anforderungen durch Ableitung aus den genannten Ideen der Teilnehmer des Workshops formuliert werden.

Ziel in den nächsten Kapiteln ist nun das Finden einer geeigneten Methode, um den genannten Anforderungen gerecht werden zu können.

Kapitel 8

Ablauf der Bildbeschriftung bei Bildagenturen

Die Forderung nach der Einbeziehung von *Benutzern in die Ontologierstellung* und der *Schaffung eines arbeitsintegrierten Ablaufs von Ontologieentwicklung und der Erstellung semantischer Metadaten* macht es nötig, die genauen Arbeitsabläufe zu kennen, wie Bildbeschriftungen durch Bildarchivare erstellt werden. Mit der Kenntnis der nötigen Arbeitsschritte bei der bisher üblichen Erstellung von textbasierten Bildbeschriftungen kann ermittelt werden, wie diese geändert und erweitert werden müssen, damit eine praxistaugliche Erstellung von semantischen Metadaten möglich wird. Daher soll der Ablauf der Bildbeschriftung in diesem Kapitel betrachtet werden.

Ein weiteres Interesse in diesem Kapitel ist, wie umfangreich textbasierte Bildbeschriftungen ausfallen und wie hoch der Zeitaufwand für ihre Erstellung ist. In Kapitel 11 soll dieser Zeitaufwand mit dem Aufwand zur Erstellung semantischer Metadaten verglichen werden zur Beantwortung von *Frage 2 (c)* dieser Arbeit.

8.1 Erstellung einer textbasierten Bildbeschriftung

Im Folgenden wird gezeigt, wie aufwändig die Erstellung von Bildbeschriftungen ist und dies durch ein Beispiel aus der Praxis entsprechend gezeigt.

8.1.1 Betrachtung des entstehenden Aufwands

Das Hauptproblem bei der Bildbeschriftung ist, dass ein Bildarchivar raten muss, welche Begriffe ein Bildsucher später in Suchanfragen verwenden könnte [MS00].



Keywords	Person, Nicole Kidman, Schauspieler, weiblich, Frau , Australien, Hut Person, Jude Law, Schauspieler, männlich, Mann, Großbritannien, Grossbritannien, Seitliches Profil, Nahaufnahme, 2 (zwei) Personen, Filmszene, Unterwegs nach Cold Mountain, Film, Drama, Regie Anthony Minghella
Copyright	Bildagentur photo12

Abbildung 8.1: Beschriftung eines Bilds aus einer Filmszene

Ein Bildarchivar bildet daher semantische Zusammenhänge mit Hilfe von Text in die Metadaten ab. Es werden neben spezifischen Angaben auch generelle Angaben gegeben. Weiterhin werden oft unterschiedliche Sprachen verwendet sowie synonyme Bedeutungen angegeben, um die lexikalischen Probleme bei der Bildsuche einschränken zu können. Ebenso werden Themen angegeben, zu denen ein Bild gefunden werden könnte. Bildredakteure geben an, dass sie pro Bild etwa 30 Beschriftungen erstellen. Dies zeigt, dass die Erstellung von Bildbeschriftungen basierend auf Text sehr aufwändig ist.

8.1.2 Beispiel: Beschriftung eines Bilds in einem Filmarchiv

Somit werden für das in Abbildung 8.1 gezeigte Bild tatsächlich eine Vielzahl möglicher Beschriftungen erstellt, wie zum Beispiel die spezifische Beschriftung ‘Nicole Kidman‘ und die generelle Beschriftung „Frau Hut“, beziehungsweise die imaginäre Beschriftung „australische Schauspielerin“. Das Bild zeigt die deutschen Übersetzungen der in englischer Sprache von Bildredakteuren bei photo12 erstellten Bildbe-

schriftung. Wie man sieht, wurde für eine eigentlich einfache Szene, „das Bild zeigt *Nicole Kidman* und *Jude Law* im Film *Unterwegs nach Cold Mountain*“ bereits mehr als zwanzig verschiedene Stichwörter verwendet werden.

8.1.3 Iterativer Ablauf zur Erstellung von Bildbeschriftungen

Der hier gezeigte iterative Ablauf zur Erstellung von Bildbeschriftungen entstand auf der Basis von Beobachtung der Arbeit von Bildarchivaren bei der Bildagentur photo12 sowie durch Befragung der Bildarchivare.

Innerhalb von Bildagenturen arbeiten für gewöhnlich mehrere Bildarchivare an der Erstellung von Bildbeschriftungen. Sie bilden daher eine kollaborativ arbeitende Gruppe von Fachbenutzern in einer kleinen Gruppe von drei bis zehn Personen.

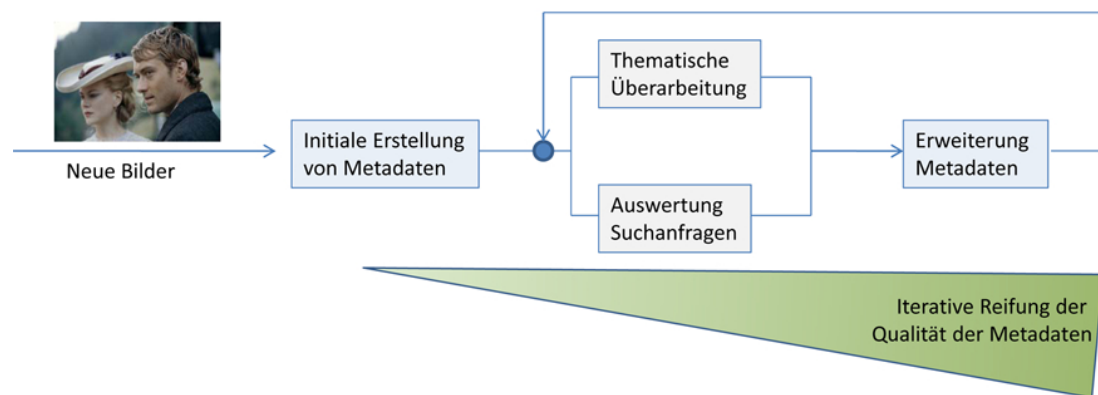


Abbildung 8.2: Iterativer Ablauf zur Erstellung von Bildbeschriftungen

Der Ablauf zur Erstellung von Bildbeschriftungen erfolgt dabei nach den in Abbildung 8.2 gezeigten iterativen Schritten:

- **Neue Bilder im Bildarchiv.** Neue Bilder gelangen in das Bildarchiv. Diese tragen zumeist initiale Beschriftungen, die vom Fotograf selbst durchgeführt wurden, wie zum Beispiel Aufnahmedatum und Aufnahmeort.
- **Initiale Erstellung von Metadaten.** Ein Bildarchivar erstellt initiale Metadaten für das Bild und veröffentlicht das Bild, zum Beispiel in der webbasierten Bildsuchmaschine für Kunden der Agentur.
- **Auswertung von Suchanfragen und thematische Überarbeitung.** Anhand der Auswertung von Suchanfragen, die über die Bildsuchmaschine protokolliert werden, ermitteln die Bildarchivare Stichwörter, die zu den Metadaten

einzelner Bilder hinzugefügt werden müssen, damit diese in entsprechenden Suchanfragen mit auftauchen. So ermöglichen die Bildarchivare vollständige Ergebnisse in zukünftigen Suchanfragen. Entsprechend führen regelmäßige thematische Überarbeitungen der Bilder zu Stichwörtern, die den Metadaten einzelner Bilder hinzugefügt werden sollten. So könnte ein Thema „Australien“ sein, so dass Bilder, die Nicole Kidman zeigen, um das Wort „Australien“ erweitert werden müssten. Erst dann können diese Bilder bei Suchanfragen wie „australische Schauspielern“ erscheinen.

- **Erweiterung der Metadaten.** Die ermittelten Stichwörter werden den Metadaten der betroffenen Bilder hinzugefügt

Die Erstellung von Bildbeschriftungen ist somit ein iterativer Prozess. Somit kann man sagen, dass *Bildbeschriftungen über die Zeit reifen* (siehe Abbildung 8.2).

8.2 Empirische Studie 3: Zeitaufwand zur Erstellung textbasierter Bildbeschriftungen

Um den Zeitaufwand zur Erstellung von textbasierten Bildbeschriftungen zu ermitteln, wurde eine entsprechende Messung mit Bildarchivaren durchgeführt. Aufgrund des entsprechend zu erwarteten hohen Zeitaufwands wird anschließend gezeigt, wie semantische Verfahren helfen können, diesen zu reduzieren.

Benutzer	ph12-1	BSMC-1	BSMC-2	AVG
Vorgegebene Zeit (min)	60	60	60	60
Anzahl Bilder beschriftet	237	172	160	189,7
Summe Beschriftungseinträge	435	347	345	375,7
- pro Bild	1,84	2,02	2,16	2
- pro Minute	7,25	5,78	5,75	6,3

Tabelle 8.1: Ergebnis der Messung des Zeitaufwands zur Beschriftung angezeigter Personen

Um zu ermitteln, wie lange Mitarbeiter von Bildagenturen und Bibliotheken für die Erstellung von Bildbeschriftungen benötigen, wurde hierzu innerhalb des IMAGINATION Projekts eine Messung durchgeführt. Zwei Bibliothekare aus der italienischen Bibliothek für moderne Geschichte (BSMC) sowie ein Angestellter der

Bildagentur photo12 aus Frankreich sollten hierzu eine Stunde lang Bilder, die Personen aus dem Bereich „Politiker EU“ zeigen, beschriften. Die Teilnehmer sollten als Einträge in der Bildbeschriftung angeben, wie viele weibliche und männliche Personen auf dem Bild zu sehen sind. Falls mehr als vier Personen auf dem Bild zu sehen waren, sollte das Bild als Gruppenaufnahme beschriftet werden. Weiterhin sollten bei Aufnahmen, die den ganzen Körper der Personen zeigen, entsprechend beschriftet werden. Es durften Abkürzungen verwendet werden, wodurch zum Beispiel die Einträge „1m 1f bo“ in der Bildbeschriftung dafür stand, dass ein Bild eine Frau und einen Mann in einer Ganzkörperaufnahme zeigen. Die Einträge sollen also generelle Informationen über die abgebildeten Personen abbilden.

Wie Tabelle 8.1 zeigt, konnten für die vorgegebene Art der Beschriftungseinträge pro Teilnehmer im Durchschnitt etwa 190 Bilder pro Stunde beschriftet werden. Insgesamt wurden hierbei im Durchschnitt etwa 380 Einträge erstellt. *Pro Minute sind dies im Durchschnitt sechs Stichwörter, die als Einträge erstellt werden.* Da bewusst Abkürzungen und generelle, also kurze Stichwörter als Einträge verwendet wurden, kann dieses Ergebnis als die Mindestdauer zur Erstellung eines Eintrags in der Bildbeschriftung angesehen werden.

Zur Erstellung von dreißig Einträge in einer Bildbeschriftung braucht man demnach *pro Bild mindestens fünf Minuten*¹. Bildredakteure gaben auf die gestellte Frage, wie hoch sie den Aufwand schätzen, eher zehn Minuten an. Zur manuellen, textbasierten Beschriftung von 100.000 Bildern bräuchte man somit mindestens ein ganzes Kalenderjahr Zeit.

8.3 Einsarpotential durch semantische Verfahren

Die Einsparung von Zeit zur Bildbeschriftung ist daher sehr gewünscht innerhalb von Bildagenturen. Ein Bereich, in dem semantische Verfahren bei der Bildbeschriftung helfen können, sind die lexikalischen Verfahren. Dies reduziert die Notwendigkeit zur Erstellung synonyme Bedeutungen und von Beschriftungen in mehreren Landessprachen.

Wie gezeigt werden in textbasierten Beschriftungen semantische Beziehungen manuell nachgebildet und Beschriftung für einzelne Bilder immer wieder neu geschrieben. Generelles Einsarpotential bietet hierbei die Nutzung von Instanzenmodellen. Wichtig aber ist, dass der Zeitaufwand zur Erstellung von Ontologien und die anschließende Nutzung für semantische Metadaten nicht höher ist als der ursprüngliche Zeitaufwand zur Erstellung textbasierter Metadaten. In Instanzenmodellen kön-

¹berechnet aus 30 Stichwörter / durchschnittlich sechs Stichwörter pro Minute (siehe Tabelle 8.1)

nen die Beziehungen der semantischen Elemente beschrieben werden und das damit verbundene Hintergrundwissen ist somit auch in den Metadaten vorhanden. So genügt die Verknüpfung der spezifischen Elemente und jeweils einmalige Erstellung von Relationen zu weiteren semantischen Elementen.

8.4 Arbeitsablauf zur Erstellung semantischer Metadaten

Zum Abschluss dieses Kapitels soll gezeigt werden, wie die semantische Erstellung von Metadaten durch sinnvolle Erweiterung der bestehenden Abläufe durchgeführt werden kann. Angenommen wird dabei bereits, dass die Benutzer die Erstellung von fehlenden semantischen Elementen selbst durchführen, sowie diese geeignet überarbeiten können. Beim Arbeitsablauf wird von der Erstellung von semantischen Metadaten ausgegangen.

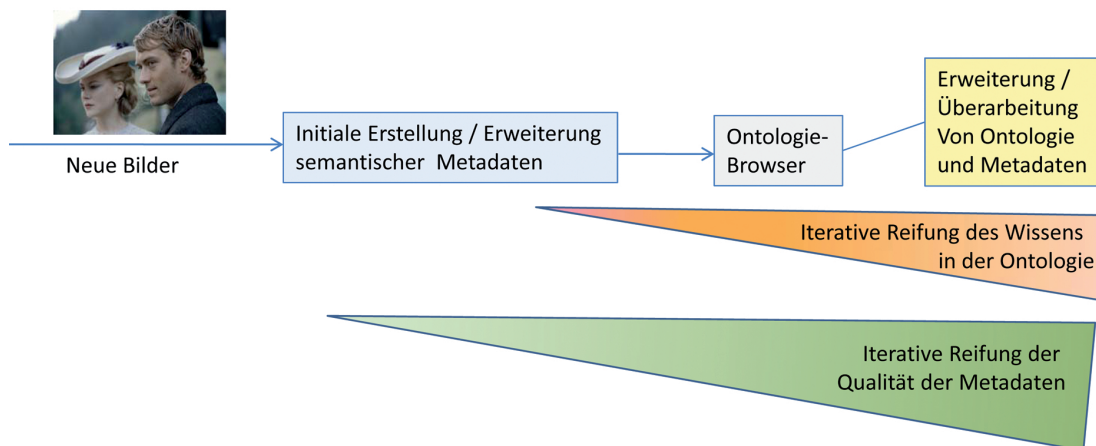


Abbildung 8.3: Arbeitsablauf zur Erstellung semantischer Bildbeschriftungen

Abbildung 8.3 zeigt den resultierend Arbeitsablauf unter Einbeziehung der Erstellung von Ontologien. Es entstehen die folgenden Arbeitsschritte:

- **Hinzufügen einer Beschriftung.** Ein semantisches Element aus einer Ontologie soll zur Erstellung semantischer Metadaten verwendet werden.
- **Aufruf Ontologiebrowser.** Zur Auswahl des semantischen Elements wird der in der Anwendung integrierte *Ontologiebrowser* aufgerufen. Dieser ermöglicht die Suche nach semantischen Elementen in der Ontologie. Falls ein semantisches Element fehlt, wird dieses *erstellt*. Falls ein semantisches Element erweitert werden soll, wird dieses *überarbeitet*.

- **Hinzufügen des semantischen Elements zur Bildbeschriftung.** Das im Ontologiebrowser gewählte semantische Element wird den semantischen Metadaten hinzugefügt.

Die Bildarchivare können durch diesen Ablauf fehlende semantische Elemente direkt in die Ontologie hinzufügen und anschließend direkt verwenden. So entsteht der gewünschte, praxistaugliche Ablauf zur Erstellung semantischer Metadaten.

Auch bei semantischen Metadaten bleibt ein Aspekt der Reifung der Qualität der Bildbeschriftung, dass neue Elemente über die Zeit den Metadaten hinzugefügt werden.

Abbildung 8.3 soll aber einen weiteren Aspekt verdeutlichen, der mit der Erstellung semantischer Metadaten hinzukommt. Die kontinuierliche Erweiterung von semantischen Elementen einer Ontologie führt *ebenfalls zu einer Reifung der Qualität der Bildbeschriftungen*. Wird einem semantischen Element eine neue Relation und somit neues Hintergrundwissen hinzugefügt, wird dieses automatisch auch als Hintergrundwissen über die damit beschrifteten Bilder verwendbar. Somit kann man sagen, dass „*semantische Bildbeschriftungen mit der Erweiterung von Ontologien um wichtiges Hintergrundwissen reifen*“.

8.5 Zusammenfassung

Innerhalb dieses Kapitels wurde gezeigt, wie hoch der Aufwand und die dabei benötigte Zeit zur Erstellung textbasierter Bildbeschriftungen ist. Weiterhin wurde gezeigt, wie die bestehenden Arbeitsabläufe erweitert werden müssen, damit eine praxistaugliche Erstellung semantischer Metadaten möglich wird.

Mit den Überlegungen zur *Reifung von Ontologien* und *Reifung von Bildbeschriftungen* wird nun im nächsten Kapitel ein entsprechendes Modell gezeigt, damit die Bildarchivare wie gewünscht fehlende semantische Elemente selbst einer Ontologie hinzufügen können.

Kapitel 9

Ontologieentwicklung als Reifungsprozess

Die Überlegungen der Kollegen am FZI Andreas Schmidt zur *Reifung von Wissen* [Sch05] sowie die Suche von Braun und Zacharias nach einer geeigneten Methode zur *Erstellung von Taxonomien für das semantische Bookmarksystem SOBOLEO* und die eigenen Überlegungen zur *Reifung von Bildbeschriftungen mit semantischen Metadaten* führten zur gemeinsamen Entwicklung des in diesem Kapitel gezeigten *Ontology Maturing Process Model* [BSWZ07] [BSW⁺07] (zu Deutsch Reifungsprozessmodell für Ontologien).

Speziell die gewünschte Einbeziehung von Benutzern zur *kollaborativen Erstellung und Erweiterung von semantischen Elementen* (OA1) sowie eine *flexible Formalisierung der so entstehenden Ontologien* (OA5) soll damit möglich werden. Hierzu soll das *Ontology Maturing Process Model* zur Beschreibung dienen, wie Ontologien in Anlehnung an die Reifung von Wissen in Gruppen reifen können.

Mit diesem Prozessmodell können anschließend darauf basierende Methoden zur praktisch nutzbaren Erstellung von Ontologien durch Benutzer innerhalb von semantische Anwendungen entwickelt werden.

9.1 Vorüberlegung zur benutzergetriebenen Erstellung von Ontologien

Bei der Suche nach einem geeigneten Prozessmodell für die benutzergetriebene Ontologieentwicklung sollen die in Sektion 7.6 abgeleiteten Anforderungen gelten. Zusätzlich wurden die folgenden Überlegungen angestellt:

- **Erstellung von leichtgewichtigen Ontologien** Experten zur Erstellung von Ontologien versehen diese oft umfangreich mit Regeln. Diese Regeln können anschließend von Computern interpretiert werden. *Die Nutzung der Ontologien durch Computer steht dort im Vordergrund* [Gru93].

Mit dem Fokus auf Benutzer, die Ontologien in einer Anwendung zur Beschriftung von Ressourcen nutzen, *stehen die semantischen Elemente selbst im Vordergrund*. Diese müssen geeignet zur Nutzung für die Beschreibung von Ressourcen vorhanden sein. Benutzer sollen daher fehlende Elemente hinzufügen und bestehende Elemente überarbeiten können. Dies wird als die Erstellung von *leichtgewichtigen Ontologien* bezeichnet.

- **Kleine Gruppen arbeiten kollaborativ zusammen** Bei der Erstellung von Ontologien arbeiten kleine Gruppen zusammen, also beispielsweise Bildarchivare in einer Bildagentur oder eine Gruppe von Kollegen, die in einer Firma strukturiert Bookmarks von Webseiten untereinander teilen möchten.

Es ist daher von Interesse, wie Benutzer in kleinen Gruppen Ontologien erstellen können und wie darin erkannt wird, welche semantischen Elemente von den Benutzern als wichtig empfunden werden. Wichtige Elemente führen anschließend zu einem gemeinsamen Verständnis für die vorhandenen semantischen Elemente und dadurch zu einer einheitlichen Nutzung in der semantischen Anwendung.

- **Verwendung in Anwendung als Motivation** Es wird angenommen, dass *die Ontologieerstellung für die Benutzer ein Mittel zum Zweck ist*. Ontologien werden also nicht einfach so von Benutzern erstellt.
- **Einfacher Formalismus** Für ein einfaches Verständnis der Schritte zur Ontologieerstellung sollte eine geeignete Methode auf einem einfachen Formalismus basieren.

Die hauptsächlich daraus abgeleiteten Fragestellungen sind demnach, *wie neue Elemente entstehen, wie sie in der Gruppe weiter entwickelt werden* und somit von *Benutzern akzeptierte Ontologien entstehen können*, die von allen Teilnehmern in der Gruppe verstanden und verwendet werden in einer semantischen Anwendung.

9.2 Verwandte Arbeiten

Innerhalb der Recherche nach verwandten und geeigneten Arbeiten zur Definition eines Prozessmodells zur benutzergetriebenen Ontologieentwicklung sind zum einen *kollaborative Methoden und Anwendungen zur Ontologieentwicklung* von Interesse.

Weiterhin wird aufgrund ihrer Einfachheit die *Methode in Semantic Media Wiki* vorgestellt und wie dort semantische Informationen abgebildet werden.

9.2.1 Kollaborative Methoden und Anwendungen zur Ontologieentwicklung

Im Folgenden wird gezeigt, welche Erweiterungen von herkömmlichen Methodologien (siehe Sektion 7.1) und welche Anwendungen zur kollaborativen Ontologieerstellung möglich sind.

Erweiterung herkömmlicher Methodologien

Verteilte Ontologieerstellung: *HCOME* [KVA04] ermöglicht die kollaborative Entwicklung von Ontologien. Der Ansatz ist hierbei, dass dezentralisiert Ontologien von Interessensgruppen entwickelt werden. Diese werden anschließend durch entsprechende Verfahren zu einer gesamten Ontologie zusammengeführt. Dieser Ansatz sorgt dann für ein besseres „gemeinsames Verständnis“ der Ontologien über eine größere Benutzergruppe hinweg. *Der Prozess der Ontologieentwicklung bleibt aber weiterhin ein getrennter Prozess.* So können die Ontologien also nicht wie gewünscht durch Benutzer bei Bedarf in der Anwendung erweitert werden.

HCOME kann daher als ein Ansatz genutzt werden, Ontologien die durch kleine Benutzergruppen erstellt werden, zu kombinieren und so über mehrere Systeme hinweg verwendbare Ontologien zu erstellen. Somit fehlt noch der Aspekt, wie Benutzer selbst diese Ontologien erstellen können.

Wiki basierte Ontologieerstellung mit Benutzerintegration: Mit *myOntology* [SH07] sollen die Benutzer selbst angeben können, welche Konzepte in einer Ontologie sie gerne erstellen möchten. Hierzu werden, speziell um die Aspekte *Dokumentation* und *Diskussion* in einer Benutzergruppe zu ermöglichen, für jedes Konzept eine eigene Seite eines Wikis erstellt. Die Benutzer können dann dokumentieren, welche Konzepte sie erstellen möchten und die Gruppe kann dies kommentieren und so über die Konzepte diskutieren. Auch eine prototypische Implementierung existiert. Bei der Evaluation von *myOntology* wurde festgestellt, dass „motivierende Aspekte für Benutzer wichtig sind, damit diese Ontologien erstellen“. Deren Ansatz ist daher, durch Spiele wie *OntoGame* „die Benutzer zu motivieren, an der Ontologieerstellung teilzunehmen“ [SH08].

Wie *myOntology* zeigt, ist die Motivation zur Erstellung von Ontologien durch Benutzer wichtig. *Daher wird in unserer Lösung der Aspekt der arbeitsintegrierten Ontologieentwicklung betont.* Benutzer, die Ontologien in einer Anwendung benöti-

gen, sind auch motiviert, diese zu erweitern. Dieser Aspekt wurde durch die Befragung der Bildarchivare in Sektion 7.5 belegt.

Anwendungen

Collaborative Protege [TN07] erlaubt die kollaborative Entwicklung von Ontologien. Die Anwendung ist allerdings auf Experten in der Ontologierstellung ausgelegt. Entsprechend bietet diese Anwendung Möglichkeiten zur umfangreichen Formalisierung von Ontologien an. Ähnliches gilt für den Ontologieeditor *OntoEdit* [SEA⁺02].

Basierend auf unseren Annahmen sollen Benutzer leichtgewichtige Ontologien erstellen können. Entsprechend sind die Funktionen in *Collaborative Protege* zu umfangreich und nicht einfach genug für Benutzer.

Bewertung

Eine Überarbeitung von Ontologien, die durch Benutzer einer semantischen Anwendung erstellt worden sind, ist mit HCOME möglich. Dadurch können die Ontologien überarbeitet und zusammengefügt werden. Experten in der Ontologierstellung erhalten so die tatsächlich von Benutzern neu benötigten semantischen Elemente und können diese entsprechend formalisieren. Benutzer erstellen leichtgewichtige initiale Ontologien, Experten in der Ontologierstellung können diese dann geeignet weiter formalisieren, aber auch reinigen, strukturell überarbeiten und mit Regeln versehen.

Im weiteren Kapitel geht es daher um die noch fehlende Definition eines Prozessmodells, damit Benutzer diese leichtgewichtigen Ontologien erstellen können.

9.2.2 Semantic Media Wiki

Das Semantic Wiki (kurz SMW) dient zur Aufbereitung von Informationen in Wiki-Seiten [VKV⁺06]. Die mit dem SMW verbundene Methode ist die Erstellung von *typisierten Verweisen* und *Attributen*. Ein typisierter Verweis beschreibt die Art eines Verweises auf einen anderen Artikel. Mit dem typisierten Verweis „Hauptstadt von“ „Italien“ kann man so ausdrücken, dass Rom die Hauptstadt von Italien ist. So erlauben diese typisierten Verweise das Abfragen ähnlicher Seiten. Im Beispiel wären das weitere Seiten, die mit „Hauptstadt von“ typisiert sind wie „Berlin“, „Athen“ oder „Warschau“. Typisierte Attribute erlauben die genauere Beschreibung von Daten. So kann mit dem Attribut „PreisEuro:1,99“ (und der Angabe, dass Euro ein numerischer Typ ist) angegeben werden, dass die angegebene Zahl „1.99“ eine Preisangabe in Euro ist. Die basierend auf dem Wiki Gedanken kollaborativ erstellbaren typisier-

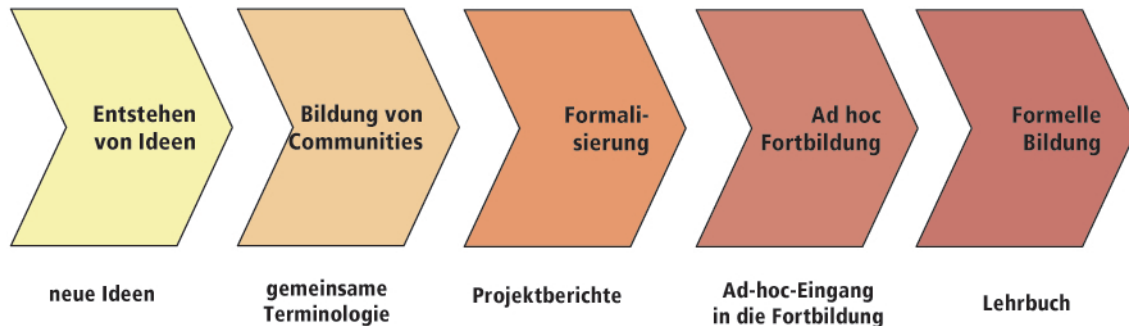


Abbildung 9.1: Phasen der Reifung von Wissen

ten Verweise und Attribute erlauben anschließend *die automatische Verarbeitbarkeit der so hinterlegten semantischen Informationen*.

Einsetzbarkeit für semantische Metadatenerstellung: Semantic Media Wiki erlaubt so basierend auf einem einfachen Formalismus unter der Einbeziehung von Benutzern *in der arbeitsintegrierten Arbeit des Schreibens von Wikipedia Artikeln die Erstellung semantischer Beziehungen*. Weiterhin ist eine *Dokumentation dieser Beziehungen* durch ein entsprechendes Erstellen eines Artikels im SMW möglich. So setzt SMW die in Sektion 7.6 definierten Anforderungen innerhalb Wikis um, die auch für eine Ontologieerstellung zur Erstellung semantischer Metadaten gefordert sind.

Allerdings zielt das SMW nicht auf die Erstellung von benötigten Ontologien wie Taxonomien und Instanzenmodellen ab [SH07]. Das in den weiteren Sektionen dieses Kapitels entwickelte Modell ist allerdings speziell von der Einfachheit der Methode in SMW motiviert.

9.3 Das Ontology Maturing Process Model

9.3.1 Reifung von Wissen

Schmidt hat sich mit der Frage beschäftigt, wie Wissen innerhalb einer Gruppe reift [Sch05]. Als Ergebnis hat er fünf Phasen definiert. Abbildung 9.1 zeigt ein Beispiel der Phasen angewandt auf die Erstellung eines Lehrbuchs. In Schritt 1 entstehen *neue Ideen*, zum Beispiel aufgrund von neuen Forschungsergebnissen. In Schritt 2 entstehen *gemeinsame Terminologien* in einer Fachgruppe, zum Beispiel über zu verwendende Begriffe für die neuen Forschungsergebnisse, die gemeinsam

verwendet werden. Schritt 3 ist die anschließende *Formalisierung dieses Wissens* durch Berichte oder Dokumentationen. Durch den Eingang dieses Wissens in Kurse in Schritt 4 wird das Wissen *in größeren Gruppen verbreitet*. Die Erstellung eines Lehrbuchs zu diesem Inhalt führt schließlich in Schritt 5 zur *formellen Bildung*.

Die Reifung von Wissen als Modell führt somit von Ideen einzelner Benutzer auf ein Wissen, das gemeinsam genutzt und verfeinert wird und anschließend formalisiert wird. *Das Wissen reift somit über die Zeit.*

9.3.2 Resultierende Phasen zur Ontologierstellung

Das *Ontologie Maturing Process Model* ist die *Abbildung des Wissensreifungsprozess auf die Erstellung von Ontologien*, denn in Ontologien wird Hintergrundwissen über Dinge und Sachverhalte abgebildet, wodurch dort das Wissen reift. *Eine Ontologie reift demnach über die kontinuierliche Erweiterung um semantische Elemente.*

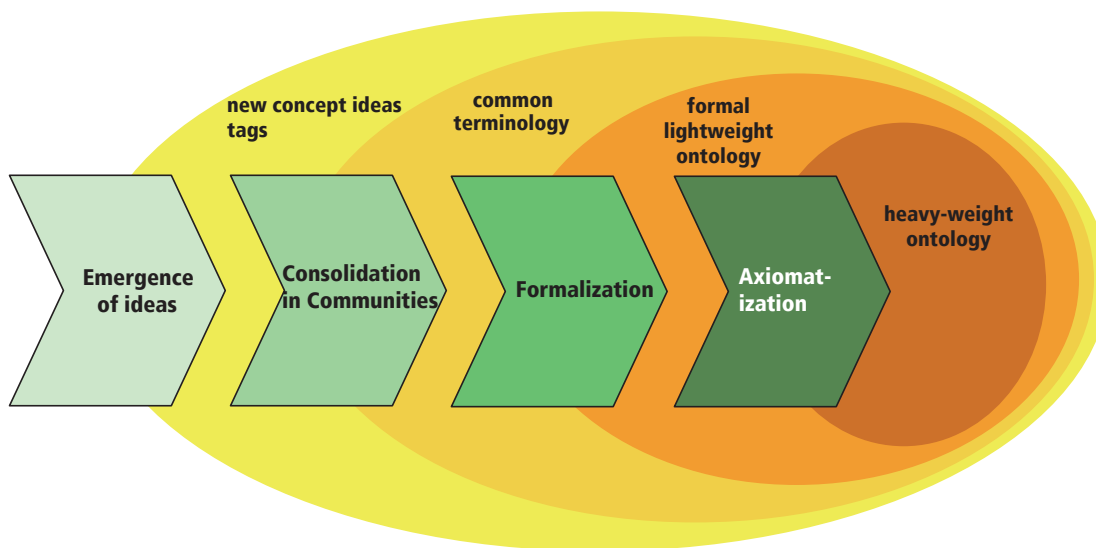


Abbildung 9.2: Vier Phasen des Ontologie Maturing Process Model

Abbildung 9.2 zeigt die Anpassung des Wissensreifungsmodells auf die Ontologierstellung. Demnach reifen Ontologien über die folgenden vier iterativen Schritte:

- **Schritt 1: Entstehung neuer semantischer Elemente.** Ein Benutzer stellt fest, dass ein semantisches Element in der Ontologie fehlt. Daher wird es *initial* erstellt und anschließend in der semantischen Anwendung verwendet. Dies kann im einfachsten Fall die Erstellung eines Bezeichners sein, welches das se-

semantische Element beschreibt. Das semantische Element wird in die Ontologie mit aufgenommen.

- **Schritt 2: Konsolidierung der semantischen Elemente.** Der Benutzer selbst oder die weiteren Benutzer in der Gruppe greifen neu erstellte semantische Elemente auf. Sie werden durch das Hinzufügen von textuellen Informationen oder einer Dokumentation *präzisiert*. Entsprechend werden unpassende semantische Elemente *verworfen*. So setzen sich die semantischen Elemente durch, die von allen Benutzern akzeptiert und verwendet werden. Es entsteht insgesamt durch Schritt 2 ein kontrolliertes Vokabular.
- **Schritt 3: Formalisierung.** Die Benutzer fügen zu den semantischen Elementen Relationen hinzu. Diese können entweder hierarchisch sein zur Erstellung von Taxonomien, oder benannte Relationen zur Erstellung von Instanzenmodellen. Dies führt zu den gewünschten leichtgewichtigen Ontologien.
- **Schritt 4: Axiomatisierung.** Schritt 4 ist die Einbindung von Experten zur Ontologieerstellung. Diese können durch entsprechende Methodologien (wie z.B. der oben erwähnten HCOME Methodologie) die leichtgewichtigen Ontologien mehrerer Gruppen zusammenführen, die Ontologien formalisieren sowie Korrekturen vornehmen.

Alle vier Schritte können iterativ durchgeführt werden. So wird angenommen, dass für die erstellten Relationen innerhalb der Relationen auch eine Konsolidierung stattfindet. So setzen sich oft benötigte Relationen durch und werden zur Beschreibung von gleichartigen semantischen Elementen verwendet, z.B. eine mögliche Festlegung, dass die Relation „ist Hauptdarsteller in“ bei jedem „Schauspieler“ in der Ontologie anzugeben ist.

9.3.3 Anwendung des Modells

Das *Ontology Maturing Process Model* beschreibt den Prozess der benutzergetriebenen Entwicklung von Ontologien auf der Basis der Reifung von Wissen. Es zeigt, dass *initiale Ideen*, also fehlende semantische Elemente, die basierend auf der Arbeit in einer semantischen Anwendung entdeckt werden, von einzelnen Benutzern erstellt werden. Die Gruppe der Benutzer greift diese Ideen auf und *überarbeitet* diese. Die flexible Formalisierung der semantischen Element wird durch das *schrittweise Hinzufügen von Relationen* möglich. Schließlich können anschließend herkömmliche Methoden verwendet werden, um die Ontologien durch Experten weiter zu formalisieren und mit Regeln zu versehen.

Eine geeignete Umsetzung basierend auf dem *Ontology Maturing Process Model* muss somit Schritte definieren, die passend zur semantischen Anwendung die vier Phasen des Reifemodells abbilden können.

9.4 Zusammenfassung

In diesem Kapitel wurde gezeigt, dass eine geeignete Einbeziehung von Benutzern in die Ontologieerstellung darin liegt, die Schritte der *Reifung von Wissen* auf die nötigen Schritte zur *Erstellung von Ontologien zu übertragen*. So können neue semantische Elemente für eine Ontologie initial durch einzelne Benutzer erstellt werden und diese anschließend durch die ganze Gruppe überarbeitet und formalisiert werden. So setzen sich in der Gruppe diejenigen semantischen Elemente durch, die von allen akzeptiert und verstanden werden.

Ziel im nächsten Kapitel ist die Definition einer Methode zur Anwendung des *Ontology Maturing Process Model* für die Anwendung in der *Erstellung semantischer Metadaten*.

Kapitel 10

Semantische Bildbeschriftung mit ImageNotion

In diesem Kapitel wird die *ImageNotion Methodologie* vorgestellt [WN07a]. Diese ist eine Umsetzung des *Ontology Maturing Process Model*. Die ImageNotion Methodologie ermöglicht die gewünschte benutzergetriebene Erstellung der benötigten Ontologien zur Erstellung semantischer Metadaten.

Auf Basis dieser Methodologie wird weiterhin gezeigt, wie die Komponente *Verwaltung Wissensbasis* in ImageNotion entsprechend aufgebaut wird und ein Beispiel der Anwendung der Komponente mit ihrer Implementierung gezeigt. Darauf basierend soll eine empirische Untersuchung mit Bildarchivaren die Umsetzbarkeit der Erstellung von semantischen Metadaten für Benutzer mit der ImageNotion Methodologie prüfen.

10.1 Die ImageNotion Methodologie

Für einen *einfachen Formalismus* (*Anforderung OA2*, siehe Sektion 7.6) wird in der ImageNotion Methodologie *nicht zwischen Instanzen und Konzepten unterschieden*. Jedes semantische Element wird dabei in der Anwendung selbst als *imagenotion*¹ bezeichnet. Jedes semantische Element soll einen *eindeutigen Bezeichner* haben, z.B. den Namen einer Person.

Zur Umsetzung der *Anforderung OA3* nach einer *Dokumentation* der Bedeutung der semantischen Elemente ist es möglich, zu jeder *imagenotion* einen Link zu einem Wikipedia Artikel zu legen. Weiterhin wird für die Anzeige auf der Benutzero-

¹zusammengesetzt aus den englischen Wörtern image und notion, zu Deutsch so viel wie „Bedeutung eines Bilds“

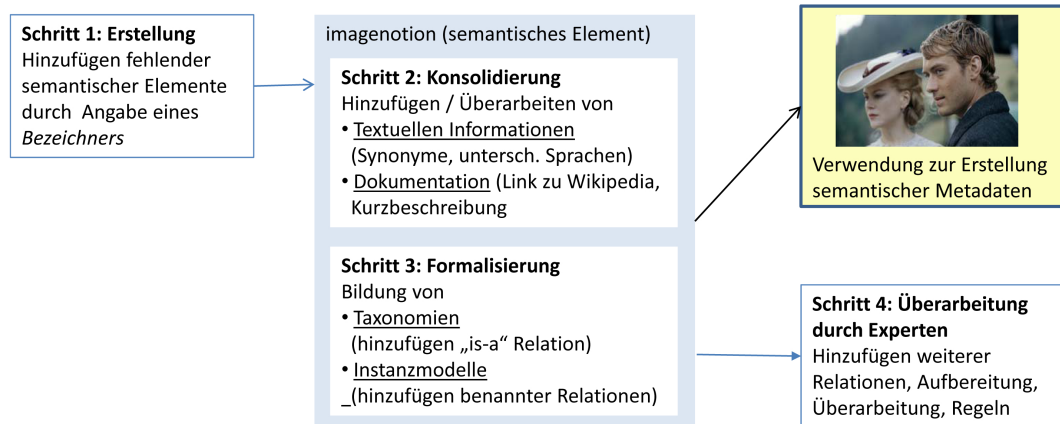


Abbildung 10.1: Phasen der ImageNotion Methodologie

berfläche ein zusätzliches Feld vorgesehen, welches einen kurzen Beschreibungstext der *imagenotion* enthalten kann.

Für ein schnelles Verständnis für einzelne *imagenotions* durch Benutzer sollen diese visualisiert dargestellt werden. Hierzu werden Bilder angezeigt, die mit der jeweiligen *imagenotion* beschrieben wurden. Innerhalb des IMAGINATION Projekts entstand so die Aussage „Bilder werden mit Bildern beschrieben“.

Zur Umsetzung der *Anforderung OA4* wird die Erstellung von *Synonymen und verschiedenen Sprachen* zur Beschriftung einzelner *imagenotions* ermöglicht. Dies erlaubt initial die Erstellung von kontrollierten Vokabularen, also Thesauri.

Zur Umsetzung der *Anforderung OA1* wird dem Benutzer selbst die *Erstellung und Überarbeitung semantischer Elemente* auf der Basis des *Ontology Maturing Prozess Model* ermöglicht.

Resultierende Ontologien und Bildbeschriftung mit ImageNotion: Mit Hilfe der ImageNotion Methodologie werden somit Ontologien erstellt, die aus *imagenotions* und Relationen zwischen diesen bestehen. Die daraus resultierenden semantischen Bildbeschriftungen bestehen darauf basierend jeweils aus einer Menge von *imagenotions*.

10.1.1 Phasen des Modells

Abbildung 10.1 gibt einen Überblick über die folgenden *vier resultierenden Phasen zur Erstellung von Ontologien* zur Erstellung semantischer Metadaten durch Benut-

zer:

- **Phase 1: Erstellung.** Stellt ein Benutzer während der Erstellung semantischer Metadaten fest, dass eine benötigte *imagenotion* zur Bildbeschriftung fehlt, erstellt er ein entsprechendes Element unter Angabe eines eindeutigen Bezeichners
- **Phase 2: Konsolidierung.** Stellen die Benutzer der Gruppe für einzelne *imagenotions* fest, dass textbasierte Informationen fehlen, werden diese in Form von Synonymen hinzugefügt. Es können hierbei unterschiedliche Landessprachen verwendet werden. Entsprechend können zur Dokumentation der *imagenotions* Links zu Wikipedia Seiten gelegt werden und kurze Beschreibungstexte erstellt werden.
- **Phase 3: Formalisierung.** Für *imagenotions*, die mit weiterem Hintergrundwissen versehen werden sollen, werden Relationen erstellt. Dies können hierarchische Relationen sein für Taxonomien oder auch benannte Relationen für Instanzenmodelle.
- **Phase 4: Überarbeitung durch Experten (Axiomatisierung).** Die vierte Phase, die Axiomatisierung, ist die Integration von Experten zur Ontologierstellung. Diese sollen die von Benutzern erstellten Ontologien entweder weiter formalisieren oder die Ontologien überarbeiten.

Alle erstellten *imagenotions* in den Ontologien zur Bildbeschriftung, unabhängig von ihrer formalen Ausprägung oder Vollständigkeit, können zur Erstellung semantischer Metadaten verwendet werden.

Die Anforderung *OA4* wird durch die *Phase 3 Formalisierung* umgesetzt. Ein flexibler Formalismus ist möglich, da das Hinzufügen von Relationen den Übergang von Thesauri zu Taxonomien und Instanzenmodellen zulässt.

Ein vollständiges Beispiel zur Anwendung der ImageNotion Methodologie wird im nächsten Abschnitt nebst der Erstellung der nötigen Komponenten in der ImageNotion Systemarchitektur gezeigt.

10.1.2 Resultierende arbeitsintegrierte Ontologierstellung

Die resultierende Ontologierstellung ist *arbeitsintegriert*. Werden fehlende Elemente entdeckt, können diese direkt über die ImageNotion Methodologie erstellt werden. Anschließend können diese direkt und somit ohne ein Wechseln der Anwendung und ohne ein Warten auf die Aktualisierung der Ontologien durch Experten für die Erstellung semantischer Metadaten verwendet werden.

So entsteht insgesamt der in Sektion 8.4 als notwendig geforderte, *praktisch einsetzbare Arbeitsablauf zur Erstellung semantischer Metadaten*.

10.2 Aufbau der Komponente „Verwaltung Wissensbasis“

Innerhalb der folgenden Sektion wird der Aufbau der Komponente „Verwaltung Wissensbasis“ innerhalb der ImageNotion Systemarchitektur gezeigt (siehe Sektion 3.4).

Zur Anzeige semantischer Elemente dient der *ImageNotion Browser*. Er soll alle nötigen Methoden anbieten, damit die Elemente der Ontologie auf der Benutzerschnittstelle angezeigt werden können. Entsprechend soll der *ImageNotion Editor* Methoden anbieten, um die Erstellung und Erweiterung von Ontologien basierend auf der ImageNotion Methodologie zu ermöglichen.

10.2.1 Der ImageNotion Browser

Suche und Navigation	Anzeige
searchImagenotion showSubImagenotions showSuperImagenotions	loadImageNotionDetails

Tabelle 10.1: Schnittstelle des ImageNotion Browsers

Tabelle 10.1 zeigt die Methoden der Schnittstelle des *ImageNotion Browsers*. Die Schnittstelle ist unterteilt nach Methoden zur *Suche und Navigation* sowie zur *Anzeige*. Zur Suche nach geeigneten Elementen dient die Methode *searchImagenotion*. Sie erhält eine textuelle Eingabe und sucht danach passende semantische Elemente basierend auf dem Bezeichner und den Synonymen. Die Methode *showSubImageNotions* dient zur Navigation. Sie soll alle semantischen Elemente anzeigen, die unterhalb eines aktuell angezeigten Elements liegen. Entsprechend soll die Methode *showSuperImageNotions* die Navigation zu darüber liegenden Elementen ermöglichen. Die Methode *loadImageNotionDetails* soll alle Details eines semantischen Elements zur Anzeige laden.

Erstellung (Phase 1)	Konsolidierung ((Phase 2)
createImagenotion createRelation	add(edit,delete)Synonym add(edit,delete)WikipediaLink add(edit,delete)ShortDescription deleteImageNotion mergeWithImagenotion
Formalisierung (Phase 3)	Axiomatisierung (Phase 4
addRelation editRelation deleteRelation moveImageNotion	über Komponente „Import / Export“

Tabelle 10.2: Schnittstelle des ImageNotion Editors

10.2.2 Der ImageNotion Editor

Tabelle 10.2 zeigt den Überblick der Schnittstelle des *ImageNotion Editors*. Entsprechend der ImageNotion Methodologie wurde die Schnittstelle nach den vier dort auftretenden Phasen unterteilt in die Bereiche *Erstellung*, *Konsolidierung*, *Formalisierung* und *Axiomatisierung*.

Die Methoden im Bereich *Erstellung* erlauben das Hinzufügen neuer semantischer Relationen sowie neuer benannter Relationen. Die Methode *createImageNotion* erhält als Text den für ein neues semantisches Element zu verwendenden Bezeichner. Mit diesem Bezeichner wird das Element abgespeichert. Entsprechend erhält *createRelation* den Namen einer zu erstellenden benannten Relation.

Im Bereich *Konsolidierung* erlauben die Methoden das Bearbeiten der textbasierten Inhalte einer *imagenotion*. Daher ist die Erstellung, Bearbeitung und das Löschen von Synonymen ((add, edit, delete) *Synonym*) möglich. Gleiches gilt für passende Links zu Wikipedia Artikeln (add, edit, delete) *WikipediaLink*) sowie für der Kurzbeschreibung ((add, edit, delete) *ShortDescription*) zu einem semantischen Element. Als eine weitere Möglichkeit zur Konsolidierung erlaubt die Methode *mergeWithImageNotion* zusätzlich das Zusammenfassen von jeweils zwei Elementen der Ontologie, die von unterschiedlichen Benutzern erstellt worden sind aber den gleichen Sachverhalt beschreiben.

Die flexible Formalisierung der semantischen Elemente ist mit den Methoden des Bereichs *Formalisierung* möglich. Es ist das Hinzufügen von Relationen (addRelation), sowie die Bearbeitung (editRelation) und das Löschen (deleteRelation) existierender Relationen möglich.

Die Methode *moveImageNotion* erlaubt das Verschieben eines semantischen Elements in der Hierarchie einer Taxonomie. So ist zum Einen der Übergang von Thesauri zu Taxonomien möglich. Weiterhin können Fehler eines Benutzers bei der Einordnung eines Elements in die Hierarchie korrigiert werden.

Die *Axiomatisierung* selbst benötigt keine Methoden. Stattdessen wird die Komponente *Ontologie Import / Export* in ImageNotion verwendet. Sie erlaubt den Export bestehender Ontologien. So können diese durch Experten der Ontologieerstellung weiter verarbeitet werden. Entsprechend erlaubt der Import, extern überarbeitete Ontologien wieder zurück in das ImageNotion System geschrieben werden können. Über den Import von Ontologien ist weiterhin möglich, eine initiale Ontologie aufzuladen. Diese muss dann nur noch an Stellen überarbeitet werden, die noch nicht vollständig mit benötigten *imagenotions* für die Bildbeschriftung abgedeckt sind. Es ist also nicht erforderlich, dass in der Anwendung selbst immer mit einer leeren Ontologie begonnen werden zu muss.

10.2.3 Beispiel aus der Implementierung

Die Komponente „Verwaltung Wissensbasis“ wurde entsprechend der beschriebenen Methoden für den Ontologie Browser und den Ontologie Editor implementiert.

Bei der Implementierung wurde großen Wert auf eine möglichst fehlerfreie Benutzeroberfläche gelegt. Die Benutzer sollen in Evaluationen und bei der Nutzung nicht von ständigen Fehlermeldungen abgelenkt werden (dies war mit PhotoStuff im Benutzerworkshop der Fall).

Die Erstellung textueller Informationen zu den Elementen einer Ontologie ist in mehreren Sprachen möglich zur Unterstützung der mehrsprachigen Suche. Jeder Bildarchivar kann in seinen Benutzereinstellungen eingeben, welche Landessprache er verwendet, z.B. „Englisch“ oder „Deutsch“.

Ein Beispiel aus der Anwendung soll nun die Erstellung semantischer Metadaten mit der ImageNotion Anwendung anschaulich zeigen. Hierzu soll das in Abbildung gezeigte Bild in Abbildung 10.2 in der ImageNotion Anwendung mit semantischen Metadaten versehen werden.

Das Bild zeigt die Schauspielerin *Uma Thurman* neben einem *Jaguar E-Type*. Hierzu sollen die entsprechenden semantischen Metadaten erstellt werden. Abbildung 10.2 gibt hierzu ein Beispiel eines Ausschnitts eines möglichen Instanzenmodells. So kann zu *Uma Thurman* das Hintergrundwissen hinterlegt werden, dass sie eine *Schauspielerin* ist. Entsprechend kann zu *Jaguar E-Type* angegeben werden, dass es sich um einen *Oldtimer* handelt der *Automarke Jaguar* und es sich um das *Modell E-Type* handelt.

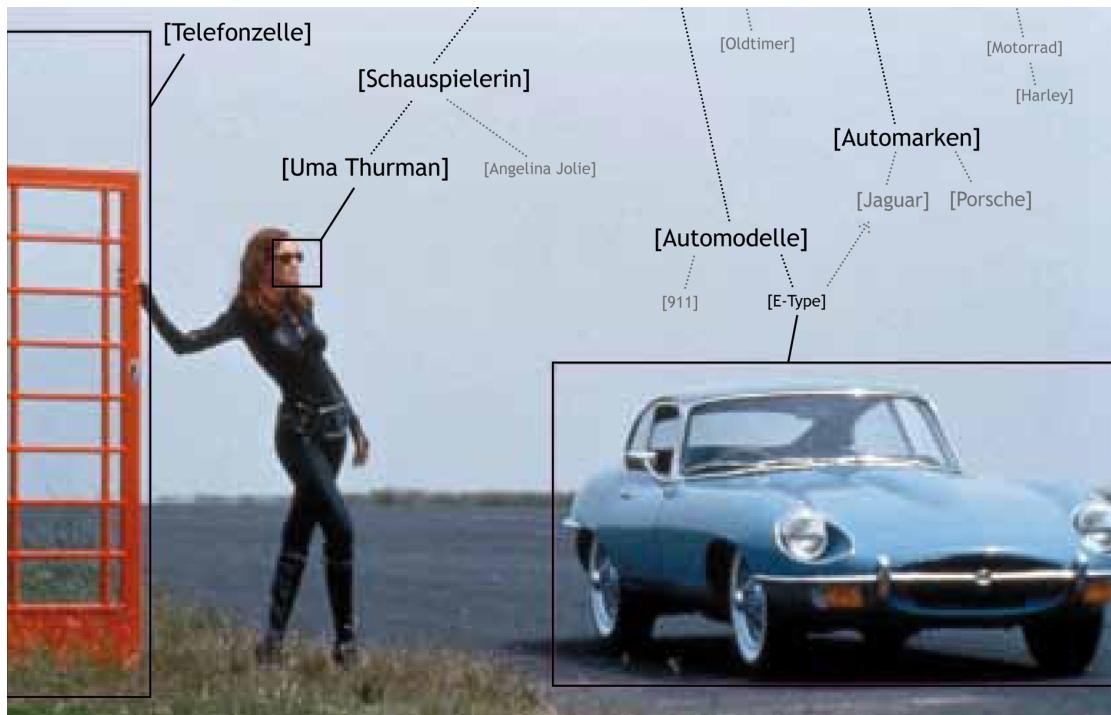


Abbildung 10.2: Beschriftung des Bilds „Uma Thurman neben einem Jaguar“

Die semantischen Elemente *Uma Thurman* und *Jaguar E-Type* sollen zur Erstellung der Metadaten für das Bild verwendet werden. Der Bildarchivar möchte weiterhin ausdrücken, dass *Jaguar E-Type* ein *Auto* ist. Im Folgenden wird gezeigt, wie der Bildarchivar hierzu mit ImageNotion vorgeht. Es soll angenommen werden, dass das die *imagenotion* „Auto“ noch nicht vorhanden ist. Daher muss es von einem Bildarchivar erstellt werden, um den Bildteil in Abbildung 10.2 wie gewünscht beschreiben zu können.

In ImageNotion kann der Benutzer zur Erstellung von *imagenotions* auf zwei verschiedenen Wegen vorgehen. Entweder er nutzt direkt den ImageNotion Browser, um darin fehlende *imagenotions* zu entdecken und erstellen zu können. Oder er entdeckt fehlende *imagenotions* direkt während der Erstellung semantischer Metadaten. Es wird nun zuerst die erste Variante gezeigt.

Wie Abbildung 10.3 zeigt, kann der Benutzer im ImageNotion Browser mit der Eingabe von Stichwörtern die Ontologie durchsuchen. Die Eingabe „Ca“ (Beginn einer Suche nach „Car“) schlägt vorhandene semantische Elemente wie „David Carradine“ oder „Carlos Sorin“ vor. „Car“ ist als semantisches Element in der Ontologie nicht vorhanden, ebenso nicht ein Element „Auto“.

Zur Erstellung einer *imagenotion* klickt der Benutzer auf einen Link „Create

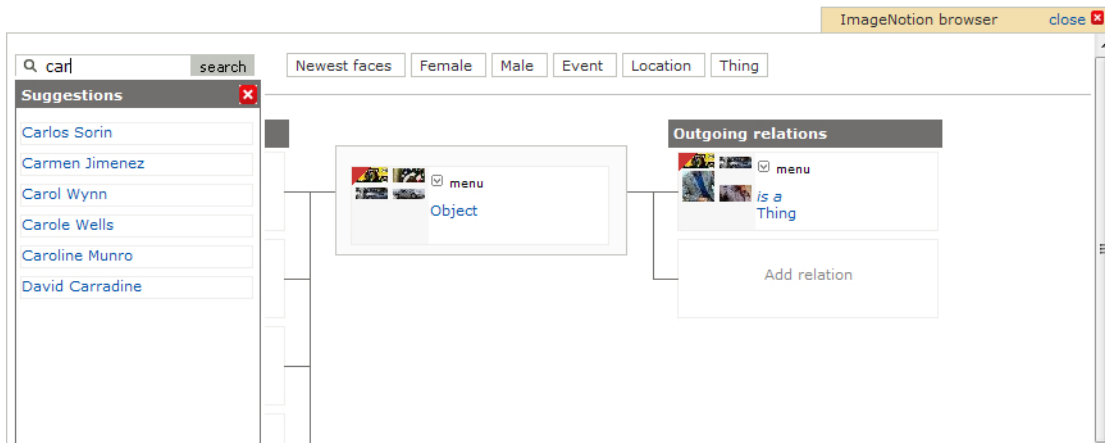


Abbildung 10.3: Suche nach semantischen Elementen mit dem ImageNotion Browser

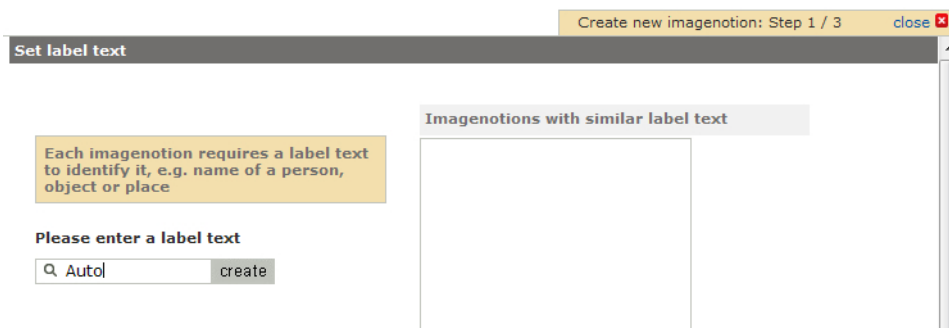


Abbildung 10.4: Erstellung des semantischen Elements „Auto“

The screenshot shows a web interface for creating a new 'imagenotion'. At the top, a yellow bar indicates 'Create new imagenotion: Step 3 / 3' with a 'close' button. Below this is a dark grey header 'Set description text'. The main content area is white and contains the following elements:

- The text 'Auto' is displayed.
- A button labeled 'Change label text' is positioned below the text.
- A section titled 'Add short description text (optional)' contains an empty text input field.
- A yellow bar at the bottom of this section contains the text 'Click on save to create this imagenotion' and a 'save' button.
- A section titled 'Selected relation' contains a dropdown menu with 'is a' selected.
- To the right of the dropdown is a list of relation options: 'menu' (with a checked checkbox) and 'Object' (with an unchecked checkbox).
- Below the relation list is a button labeled 'Change selected relation'.

Abbildung 10.5: Einordnung von „Auto“ in eine Taxonomie

new imagenotion“. Der in Abbildung 10.4 gezeigte Dialog erscheint. Der Benutzer gibt nun als gewünschten Bezeichner den Text „Auto“ ein und klickt auf den Button „Create“

Anschließend kann der Benutzer, wenn er dies möchte, die *imagenotion* direkt in die Hierarchie einer Taxonomie einordnen. Abbildung 10.5 zeigt, dass der Benutzer die neu erstellte *imagenotion* direkt unterhalb von *Objekt* einordnet. Jedes *Auto* ist demnach ein *Objekt*. Findet keine Angabe einer Relation statt, so wird das neue Element unterhalb der höchsten Kategorie „Thing“ eingeordnet, wodurch Thesauri erstellt werden können.

Der zweite Weg zur Erstellung von *imagenotions* ist direkt aus der Arbeit der Bildbeschriftung möglich. Dieser Weg zeigt den wichtigen Aspekt der arbeitsintegrierten Erstellung von Ontologien. In Abbildung 10.6 ist der Dialog zur Erstellung semantischer Metadaten gezeigt. Das zu beschriftende Bild wird angezeigt nebst den Dialogen zur Erstellung von Beschriftungen für einen Bildteil oder dem ganzen Bild. In den jeweiligen Dialogen werden ebenfalls mit der Eingabe von Text passende semantische Elemente vorgeschlagen. Der Benutzer stellt hierbei fest, dass das Element „Jaguar E-Type“ fehlt. ImageNotion bietet nun direkt einen Link an („Create new imagenotion for Jaguar E-Type“) zur Erstellung einer neuen *imagenotion*. Als Bezeichner wird automatisch die Eingabe des Benutzers übernommen, also „Jaguar E-Type“. Die weiteren Schritte sind wie oben beschrieben. Nach der Erstellung der neuen *imagenotion Jaguar E-Type* kann der Benutzer diese auswählen und der semantischen Bildbeschriftung des aktuell geöffneten Bildes hinzufügen.

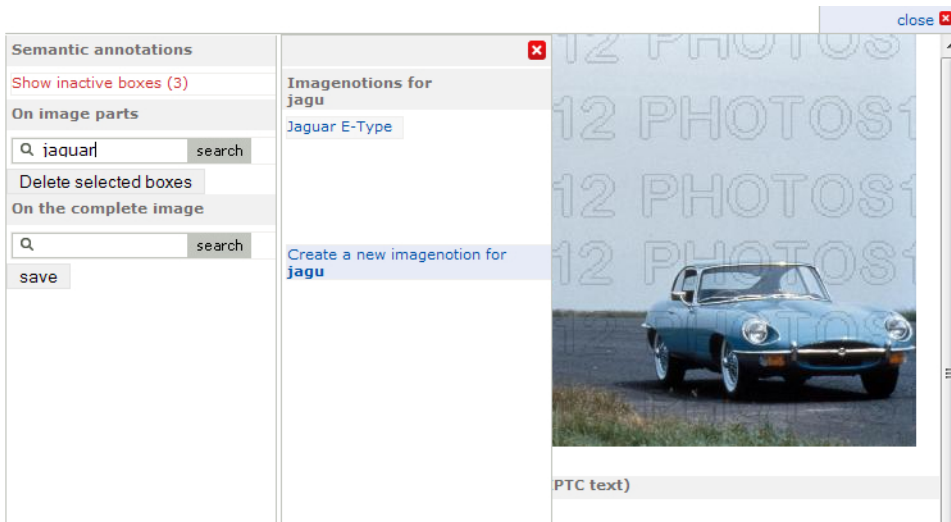


Abbildung 10.6: Erstellung der Beschriftung „Jaguar E-Type“ für einen Bildteil

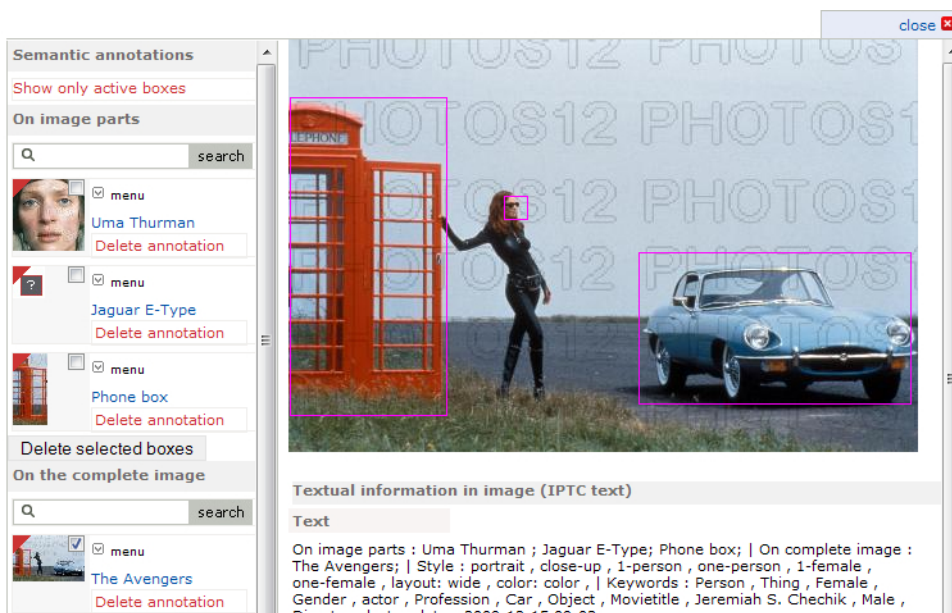


Abbildung 10.7: Resultierende semantische Metadaten für das Beispiel

Abbildung 10.7 zeigt die von einem Benutzer insgesamt erstellten semantischen Metadaten für das aktuell geöffnete Bild. Er hat weitere Bildteile beschriftet, nämlich den Bildteil, der *Uma Thurman* zeigt mit dem entsprechenden semantischen Element, sowie die *Telefonzelle*. Weiterhin hat er das ganze Bild beschrieben mit der *imagenotion The Avengers*, also dem Titel der gezeigten Filmszene.

Wie das Beispiel zeigt, konnte der Benutzer wie gewünscht alle nötigen Schritte zur Erstellung semantischer Bildbeschriftungen selbst durch die Erstellung und Verwendung von entsprechenden *imagenotions* durchführen.

10.3 Empirische Studie 4: Anwendung der ImageNotion Methode

Mit den gleichen Benutzern wie des ersten Workshops (siehe Kapitel 7) wurde ein zweiter Workshop zur Erstellung semantischer Metadaten durchgeführt. Ziel des Workshops war zu ermitteln, ob und inwieweit Benutzer mit der Umsetzung der ImageNotion Methode die Erstellung von semantischen Metadaten durchführen können.

10.3.1 Aufbau und Durchführung des Workshops

Diesmal wurde die Implementierung der ImageNotion Anwendung verwendet. Die Imagenotion Methodologie wurde darin umgesetzt. Die bereitgestellten 622 Bilder wurden in die ImageNotion Anwendung aufgeladen. Als Ontologie wurde eine stark vereinfachte Version von CIDOC-CRM als initiale Startversion zur Verfügung gestellt. Die Ontologie enthält nur eine Taxonomie mit den Elementen *Thing*, *Event*, *Person* und *Objekt*. Wiederum war die Aufgabe, geeignete semantische Bildbeschriftungen für diese Bilder zu erstellen und zwei Stunden Zeit für die Erstellung von Bildbeschriftungen gegeben.

10.3.2 Beobachtungen während des Workshops

Die sechs Teilnehmer des Workshops waren in diesem Workshop sichtlich stärker auf die Erstellung von semantischen Metadaten konzentriert. Entsprechend schienen auch mehr Bilder beschriftet worden zu sein.

Fragen zum Inhalt der Ontologie wurden überwiegend nur noch dann an die Betreuer des Workshops gestellt, wenn es sich um Unklarheiten über die Erstellung

von Relationen zwischen zwei *imagenotions* gehandelt hat. Falls es Fragen zu bestehenden Elementen der Ontologie gab, wurden diese innerhalb der Gruppe der Redakteure besprochen. Sie haben sich dann unter anderem auf zu verwendende Bezeichner und Relationen geeinigt.

Da die Bildredakteure aus unterschiedlichen Ländern waren (Frankreich und Italien), waren sie froh um die Möglichkeiten, kontrollierte Vokabulare erstellen zu können, welche die mehrsprachige Suche ermöglichen. So wurde beispielsweise die *imagenotion* „Erster Weltkrieg“ mit den Bezeichnern „First World War“, „Première Guerre mondiale“ und „Prima guerra mondiale“ versehen. Nach dem Erstellen neuer Bezeichner in bestehenden *imagenotions* wurden häufig direkt Suchanfragen mit diesen Bezeichnern durch die Redakteure gestellt. Die Redakteure waren dabei sichtlich erfreut, dass sich ihre Änderungen direkt auf alle Bilder auswirkten, die als semantische Metadaten für Bilder verwendet wurden. Entsprechend schienen sie dadurch motiviert, weitere *imagenotions* anzulegen.

Auf einzelnen Bildern haben die Redakteure ermittelt, welche semantischen Elemente benötigt werden und diese anschließend in ImageNotion gesucht. Falls diese nicht vorhanden waren, wurden neue *imagenotions* erstellt und direkt als Bildbeschriftung verwendet. Auffällig dabei war, dass die *imagenotions* meist erweitert wurden um Texteinträge und Relationen, wenn diese wieder verwendet wurden für weitere Bildbeschriftungen. So hat ein Teilnehmer die *imagenotion* „Joseph Joffre“ für das erste Bild erstellt. Bei der Verwendung von „Joseph Joffre“ auf dem zweiten Bild fiel dem Benutzer ein, dass es sich bei *Joseph Joffre* um einen *französischen General* handelt und so zusätzlich die *imagenotions* *General* und *Frankreich* erstellt. Beide *imagenotions* wurden als Relationen zu *Joseph Joffre* hinzugefügt.

10.3.3 Betrachtung der erstellten Ontologien und Bildbeschriftungen

Die Auswertung der erstellten semantischen Metadaten zeigt, dass die Benutzer diesmal viel besser mit dem zur Verfügung gestellten Abläufen klar kamen. Statt der nur fünf beschrifteten Bilder wie im ersten Workshop wurden 68 Bilder mit semantischen Metadaten versehen. Insgesamt haben die Benutzer innerhalb des Workshops 46 semantische Elemente erstellt. Fünf Synonyme sowie 18 Links wurden innerhalb der Konsolidierung der Elemente erstellt. Weiterhin wurden 40 Relationen zwischen den Elementen erstellt. Die mögliche Relation im verwendeten ImageNotion Prototyp waren „is-a“ zur Erstellung von Taxonomie und „unnamed Relation“ zur initialen Angabe, dass eine nicht hierarchische Relation zwischen zwei *imagenotions* besteht.

Die reine Betrachtung der erstellten Ontologien und Bildbeschriftungen zeigt somit, dass 12 mal mehr Bilder mit dieser Methode beschriftet werden konnten als

bei der Verwendung von vorgefertigten Ontologien. Auch haben die Benutzer bereits 40 Relationen für 46 semantische Elemente erstellt. Die Methode wurde damit zum Einen recht gut verstanden, zum Anderen haben die Bildredakteure im Workshop gemerkt, dass die Qualität der Bildbeschriftungen besser wird, je mehr Relationen hinzugefügt werden und somit nötiges Hintergrundwissen abgebildet wird, um hochwertige Suchergebnisse bei der Bildsuche zu ermöglichen.

10.3.4 Diskussion mit den Teilnehmern

Nachdem zwei Stunden im Workshop vorbei waren, fand eine Diskussion mit den Teilnehmern statt. Die Teilnehmer sollten die gezeigte ImageNotion Methode kommentieren. Die folgenden Antworten wurden genannt:

- **Hinzufügen von imagenotions:** Benutzer 2 gab an, dass es nun durch das direkte Hinzufügen von imagenotions möglich ist, überhaupt sinnvolle semantische Metadaten zu erstellen.
- **Erweiterung von imagenotions erweitert Bildbeschriftungen:** Benutzer 3 gab an, dass es „sehr praktisch sei, viele Bilder gleichzeitig mit Beschriftungen erweitern zu können, da man nur noch die entsprechende imagenotion erweitern muss. So könne man viel Zeit bei der Überarbeitung von Bildbeschriftungen sparen“.
- **Methode ist gut verständlich:** Benutzer 5 gab an, dass er die Abläufe der Methode für gut verständlich hält: „Fehlen imagenotions, werden diese einfach erstellt. Und oft verwendete imagenotions werden kontinuierlich erweitert, um deren Qualität zu erhöhen“.
- **Vereinfachung der Benutzerschnittstelle wünschenswert:** Benutzer 2 gab an, dass die Benutzerschnittstelle des ImageNotion Prototyps noch zu umständlich ist. Als Vorschlag wurde gegeben, Menüs an einzelnen imagenotions hinzuzufügen, welche direkt Funktionen zum Editieren beinhalten.
- **Protokollierung und Chat wäre hilfreich:** Benutzer 1 gab an, dass „es sehr hilfreich war, mit den anderen Benutzern über einzelne imagenotions und deren Ausgestaltung zu sprechen. Eine entsprechende Chat-Funktion in der Anwendung wäre daher für eine verteilte Arbeit hilfreich.“ Benutzer 3 gab an, dass es „weiterhin hilfreich wäre, wenn zu jeder *imagenotion* protokolliert werden könnte, welche Aktionen bereits durchgeführt wurden, um die Erweiterungen und Änderungen nachvollziehen zu können.“

*Im Vergleich zum ersten Workshop zur Erstellung von semantischen Metadaten mit diesen Bildredakteuren haben die Teilnehmer sich mit der gezeigten ImageNotion Methodologie deutlich besser zurechtgefunden. Aus den Beobachtungen und Kommentare der Teilnehmer lässt sich als Hauptgründe zusammengefasst sagen, dass dies zum Einen an der *besseren Verständlichkeit der Benutzer für die selbst erstellten Ontologien* lag und zum Anderen *die Ontologie nach Bedarf zur Beschriftung von Bildern erweitert werden konnte*.*

Die abschließende Frage, welche Methode zur Erstellung von semantischen Metadaten im IMAGINATION Projekt eingesetzt werden soll, führte zum Beschluss, dass die vorgeschlagene ImageNotion Methodologie hierfür im IMAGINATION Projekt verwendet werden soll.

Innerhalb des IMAGINATION Projekts haben diese Bildredakteure im Anschluss an diesen Workshop zusammen weitere 722 Bilder aus der Domäne *Kinobilder* bereitgestellt und basierend auf der ImageNotion Methode beschriftet. Hierzu haben die Benutzer selbst 506 semantische Elemente erstellt und ein Instanzenmodell erstellt. Das Instanzenmodell besteht aus 1049 Relationen zwischen den semantischen Elementen. 498 Relationen sind hierbei hierarchische Relationen, die restlichen 551 Relationen sind benannte Relationen. 2238 einzelne Verknüpfungen zwischen semantischen Elementen und der Bilder im Archiv zur Erstellung semantischer Metadaten wurden erstellt. Somit wurde jedes Bild mit durchschnittlich 3,1 semantischen Bildbeschriftungen versehen.

10.4 Fazit

Die ermittelten Mengen an erstellten semantischen Metadaten in den gezeigten Untersuchungen zeigen, dass Benutzer mit der ImageNotion Methodologie selbst in der Lage sind, leichtgewichtige Ontologien zur geeigneten Erstellung semantischer Metadaten erstellen können. Die fehlenden *imagenotions* wurden hierbei erstellt, konsolidiert und mit Relationen formalisiert.

Dies bestätigt somit die in Sektion 3.4.2 aufgestellte These *Die Erstellung semantischer Metadaten erfordert Möglichkeiten, damit die Benutzer selbst Ontologien erweitern und erstellen können, die zu ihren Bildinhalten passen*. Die erstellte ImageNotion Methode bietet hierfür die möglichen und nötigen Schritte an.

Die Antwort auf die Frage 2 (b) dieser Arbeit lautet daher:

Frage 2 (b)

Wie können Bildredakteure in diesen Prozess eingebunden werden und selbst Hintergrundwissen erstellen?

Antwort

Die ImageNotion Methode ermöglicht auf Basis des *Ontology Maturing Process Model* die nötigen Schritte, damit Benutzer die zur Erstellung von semantischen Metadaten benötigten und leichtgewichtigen Ontologien erstellen können. Fehlen semantische Elemente für die Bildbeschriftung, werden diese von einem Benutzer initial hinzugefügt. Anschließend werden diese *imagenotions* durch das Hinzufügen von textuellen Informationen zu einem kontrollierten Vokabular konsolidiert. Das Hinzufügen von Relationen ermöglicht schließlich den flexiblen Übergang zu Taxonomien und Instanzenmodellen. Die durch Benutzer erstellten *imagenotions* und Ontologien erlauben so einen praxistauglichen Ablauf zur Erstellung semantischer Metadaten.

10.5 Zusammenfassung

Mit der ImageNotion Methode können die Benutzer selbst die zur Erstellung semantischer Metadaten benötigten Ontologien erstellen, indem fehlende *imagenotions* erstellt werden. *Es entsteht somit ein praktisch nutzbarer, arbeitsintegrierter Ablauf zur Erstellung von Ontologien, die zur Erstellung semantischer Metadaten benötigt werden.*

Dadurch ist nun als Fazit der letzten drei Kapitel sicher gestellt, dass die Bildarchivare das für hochwertige Suchergebnisse benötigte Hintergrundwissen geeignet in die Bildsuchmaschine abbilden können. Dies ermöglicht die Nutzung der Ontologien innerhalb der Bildsuche zur Interpretation von Suchanfragen und somit zu präziseren und vollständigeren Suchergebnissen.

Im nächsten Kapitel soll nun zusätzlich untersucht werden, wie zeitaufwändig die Ontologierstellung ist und ob damit ein Zeitgewinn bei der Erstellung semantischer Metadaten möglich ist. Denn nur wenn dies der Fall ist, können diese Verfahren bei Bildagenturen eingesetzt werden. Ist der Zeitaufwand höher, wären entsprechende Verfahren nämlich zu teuer und würden von den Betreibern abgelehnt werden.

Kapitel 11

Zeitaufwand für die semantische Bildbeschriftung

Dieses Kapitel betrachtet die Frage, wie hoch der entstehende Aufwand zur Erstellung semantischer Metadaten in der Praxis ist. Basierend auf der ImageNotion Methode ist es nun Bildarchivaren selbst möglich, Ontologien zu erstellen. Neben der Zeit zur Erstellung von Bildbeschriftungen ist jedoch weitere Zeit zu berücksichtigen für das Erstellen der benötigten Ontologien. Es ist dabei wichtig, dass der dann resultierende Gesamtaufwand an Zeit nicht höher ist, als die bisherige Nutzung textbasierter Beschriftungen. Denn dann wären die semantischen Verfahren aufgrund des dabei entstehenden hohen Zeitaufwands nicht praktisch einsetzbar.

Zur Beantwortung der Frage 2 (c) dieser Arbeit *Reduzieren semantische Verfahren den Zeitaufwand zur Erstellung von Metadaten?* wird daher eine Auswertung der durch Bildarchivare erstellten Metadaten in einem Filmarchiv durchgeführt

11.1 Mögliche Erstellung von semantischen Metadaten

In dieser Sektion wird durch den Vergleich der möglichen Erstellung semantischer Metadaten gezeigt, wie sich der Einsatz von Taxonomien und Instanzenmodellen auf die Anzahl an zu erstellenden Beschriftungen auswirkt. Darauf basierend wird in der nächsten Sektion eine Auswertung der Zeitaufwände mit diesen Verfahren zur Beschriftung eines Filmarchivs durchgeführt.

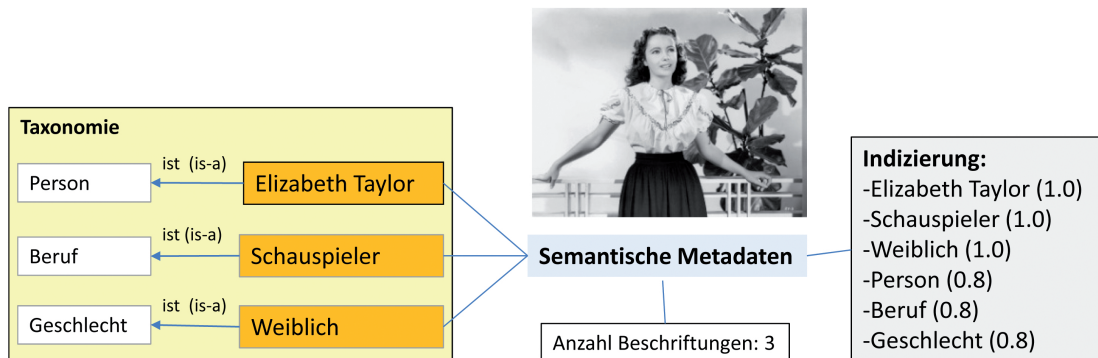


Abbildung 11.1: Bildbeschriftung in einem Filmarchiv mittels einer Taxonomie

11.1.1 Beschriftung mit einer Taxonomie

Abbildung 11.1 zeigt eine mögliche Nutzung einer Taxonomie zur Bildbeschriftung. Das Bild zeigt *Elizabeth Taylor*. Die Bildarchivare haben in der Taxonomie die semantischen Elemente *Elizabeth Taylor*, *Schauspieler* und *weiblich* erstellt. Um auszudrücken, dass Elizabeth Taylor auf dem Bild gezeigt wird, diese weiblich ist und eine Schauspielerin, werden diese Elemente zur Erstellung der semantischen Metadaten verwendet. *Jeder auszudrückende Sachverhalt muss also durch ein entsprechendes semantisches Element ausgedrückt werden und als jeweils einzelne Bildbeschriftung verwendet werden.* In diesem Beispiel mussten daher drei Bildbeschriftungen durchgeführt werden. Diese können anschließend indiziert werden. Generelle Suchen wie „weiblich Schauspieler“ zum Finden von weiblichen Schauspielern im Bildarchiv werden somit anschließend möglich.

11.1.2 Beschriftung mit einem Instanzenmodell

Abbildung 11.2 zeigt die entstehende Beschriftung des gleichen Bildes mit Hilfe eines Instanzenmodells. Die Hintergrundinformationen über *Elizabeth Taylor*, also dass sie *weiblich* ist und von Beruf *Schauspielerin* werden über benannte Relationen ausgedrückt. Als hierarchische Relation bleibt die Relation zu *Person*.

Wie aus der Abbildung 11.2 ersichtlich, sinkt damit die Menge an zu erstellenden semantischen Metadaten mittels eines solchen Instanzenmodells von drei Beschriftungen pro Bild mittels der Taxonomie auf nur noch eine zu erstellende Beschriftung pro Bild ab.

Gleichzeitig entsteht aber ein höherer Aufwand zur Bearbeitung der Ontologie,

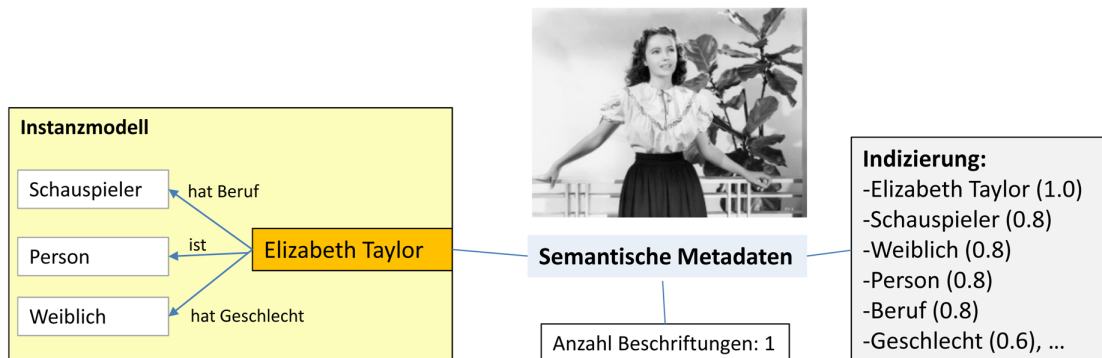


Abbildung 11.2: Bildbeschriftung in Filmarchiv mit einem Instanzenmodell

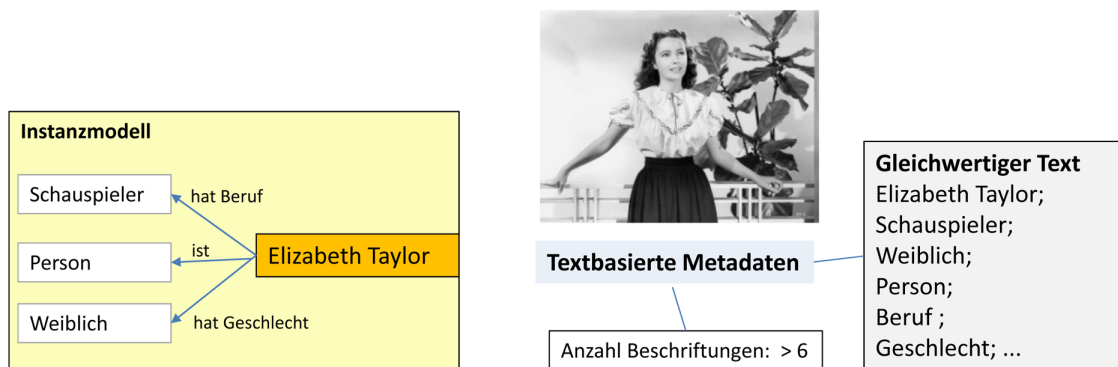


Abbildung 11.3: Bildbeschriftung in Filmarchiv mittels gleichwertigem Text

denn die benannten Relationen einzelner semantischer Element müssen nun erstellt werden. Es stellt sich somit die Frage, *ob die Nutzung eines Instanzenmodells in der Praxis schneller ist als die Nutzung einer Taxonomie?*

11.1.3 Beschriftung mit Text

Es stellt sich die zusätzliche Frage, *ob die Erstellung semantischer Metadaten mit Taxonomien und Instanzenmodellen schneller ist als die Erstellung von textbasierten Bildbeschriftungen?*

Abbildung 11.3 zeigt, dass für textbasierte Beschriftungen keine Ontologien erstellt werden müssen. Allerdings ist der Aufwand zur Erstellung von einzelnen Beschriftungen hoch. Um den mit dem Instanzenmodell beschriebenen Sachverhalt zu

Elizabeth Taylor zu beschreiben, müssen mehr als sechs Stichwörter erstellt werden.

Wie in Sektion 8.1 beschrieben, schreiben die Bildredakteure das Hintergrundwissen zu einzelnen Bildteilen jeweils mit in die textbasierte Bildbeschriftung, z.B. „Elizabeth Taylor, Schauspielerin, ...“. Diese Angaben entsprechen den durch Relationen im eben gezeigten Instanzenmodell abgebildeten Hintergrundwissen. Im Unterschied zur semantischen Bildbeschriftung muss dieses Hintergrundwissen aber für jedes Bild bei der Verwendung der textbasierten Bildbeschriftung neu geschrieben werden. Dadurch kann man davon ausgehen, dass oft einzelne Begriffe vergessen werden. Somit sind textbasierte Bildbeschriftungen nicht nur zeitaufwändiger zu erstellen, sie sind zusätzlich im Vergleich zu semantischen Metadaten oft unvollständiger und uneinheitlicher.

11.2 Empirische Studie 5: Zeitdauer zur Beschriftung eines Filmarchivs

Um die Frage zu beantworten, welche der genannten Verfahren wie viel Beschriftungszeit benötigt, wurde eine Messung des Zeitaufwand zur Beschriftung eines Filmarchivs innerhalb des IMAGINATION Projekts durchgeführt.

Die Bildarchivare der Projektpartner photo12 und BSMC haben ein Filmarchiv bestehend aus 722 Bildern komplett mit semantischen Metadaten versehen. Hierbei haben sie ein Instanzenmodell erstellt und dieses zur Bildbeschriftung verwendet. Die Bildbeschriftungen selbst wurden mit der ImageNotion Anwendung erstellt.

Die dabei protokollierten Log-Daten und Ontologien sowie semantische Metadaten selbst ermöglichen die Auswertung des Filmarchivs nach der *Dauer einzelner Arbeitsschritte* zur Erstellung semantischer Metadaten. Aufgrund dieser Durchschnittswerte können die Verfahren zur Bildbeschriftung mit Instanzenmodellen, Taxonomien und textbasierten Verfahren im Folgenden miteinander verglichen werden.

Da der Zeitaufwand zu hoch war, alle Bilder zusätzlich auch mit Hilfe von Taxonomien beziehungsweise mittels Text zu beschriften, um die tatsächlichen Zeiten und auch Qualitätsunterschiede messen zu können, wird im Folgenden eine Modellrechnung durchgeführt. Die Arbeitsschritte werden ausgehend von den jeweils nötigen Beschriftungen und Schritten zur Ontologieerstellung für Taxonomien und textbasierte Metadaten umgerechnet und so der nötige Zeitaufwand aufgrund der bekannten Zeiten für einzelne Arbeitsschritte berechnet.

Arbeitsschritt	Zeitdauer (Sekunden)
Schreiben eines Keywords (für textbasierte Beschriftung)	10
Erstellung Bezeichner <i>imagenotion</i>	10
Erstellung einer Relation	20
Erstellung einer semantischen Beschriftung	10

Tabelle 11.1: Messung der Zeitdauer einzelner Arbeitsschritte zur Bildbeschriftung

11.2.1 Dauer einzelner Arbeitsschritte

Tabelle 11.2 gibt einen Überblick über die durchschnittliche Dauer einzelner Arbeitsschritte wie dies mit den Log-Daten in ImageNotion ermittelt wurden. Zur Messung der Werte in den Log-Daten wurde in der Anwendung immer die Startzeit beim Beginn eines Arbeitsschritt und die Endzeit beim Ende des Arbeitsschritts gemessen. Hieraus konnte ermittelt werden, dass ein Bildarchivar im Durchschnitt 10 Sekunden zur Erstellung einer *imagenotion* durch entsprechende Angabe eines Bezeichners benötigt hat. 20 Sekunden wurden durchschnittlich benötigt, um zu einem semantischen Element eine Relation zu erstellen. Zur Nutzung und Auswahl eines semantischen Elements für die Bildbeschriftung wurden durchschnittlich zehn Sekunden benötigt. Aus der Messung aus Sektion 8.2 ergibt sich ebenfalls zur Erstellung textbasierten Stichwörtern in einer Textbeschriftung ein Wert von 10 Sekunden.

11.2.2 Zeitdauer der Nutzung eines Instanzenmodells

Das resultierende Instanzenmodell im Filmarchiv wurde hinsichtlich der Anzahl an erstellten semantischen Elementen und Relationen ausgewertet. Weiterhin wurde ermittelt, wie lange die Erstellung der semantischen Metadaten basierend auf diesem Instanzenmodell gedauert hat (siehe Tabelle 11.2).

Für das Instanzenmodell wurden insgesamt 506 *imagenotions* erstellt und mit insgesamt 1049 Relationen versehen. Hierbei sind insgesamt 498 hierarchische Relationen und 551 benannte Relationen erstellt worden.

Tatsächlich genutzt zur Erstellung semantischer Metadaten wurden 431 der erstellten *imagenotions*. Jedes dieser *imagenotions* hatte durchschnittlich 2,15 benannte Relationen und eine hierarchische Relation. Mit diesen semantischen Elementen wurden insgesamt 2238 Bildbeschriftungen erstellt.

Insgesamt hat die Erstellung des Instanzenmodells 443 Minuten gedauert. Die Erstellung der semantischen Metadaten hat 804 Minuten gedauert. Die Gesamt-

Bezeichner	Wert
Anzahl imagenotions erstellt	506
Summe erstellter Relationen	1049
Hierarchische Relationen	498
Benannte Relationen	551
Zeitdauer Erstellung semantischer Elemente (Minuten)	84
Zeitdauer Erstellung Relationen (Minuten)	350
Gesamtzeit Erstellung Instanzenmodell (Minuten)	443
Semantische Elemente in Bildbeschriftung verwendet	431
Durchschnittliche Anzahl benannter Relationen der in Bildbeschriftung verwendeten semantischen Elemente	2,15
Durchschnittliche Anzahl hierarchischer Relationen der in Bildbeschriftung verwendeten semantischen Elemente	1
Anzahl Bilder im Archiv	722
Anzahl Bildbeschriftungen erstellt	2238
Durchschnittliche Anzahl Beschriftungen pro Bild	4,24
Zeitdauer Erstellung semantischer Beschriftungen (Minuten)	373
Gesamtzeit Erstellung semantischer Metadaten mit Instanzenmodell (Minuten)	804
Gesamtzeit in Stunden	13, 4

Tabelle 11.2: Zeitdauer zur Beschriftung eines Bildarchivs mit einem Instanzenmodell

Bezeichner	Wert
Anzahl semantischer Elemente erstellt	506
Summe erstellter Relationen	1049
Hierarchische Relationen	498
Zeitdauer Erstellung semantischer Elemente (Minuten)	84
Zeitdauer Erstellung Relationen (Minuten)	166
Gesamtzeit Erstellung Taxonomie (Minuten)	250
Semantische Elemente in Bildbeschriftung verwendet	431
Durchschnittliche Anzahl Relationen der in Bildbeschriftung verwendeten semantischen Elemente	1
Anzahl Bilder im Archiv	722
Anzahl Bildbeschriftungen erstellt (<i>2238 * 2,15</i>)	4812
Durchschnittliche Anzahl Beschriftungen pro Bild	6,7
Zeitdauer Erstellung semantischer Beschriftungen (Minuten)	802
Gesamtzeit Erstellung semantischer Metadaten mit Taxonomie (Minuten)	1052
Gesamtzeit in Stunden	17, 5

Tabelle 11.3: Abgeleitete Zeitdauer zur Beschriftung eines Bildarchivs mit einer Taxonomie

zeit zur Beschriftung der Bilder mit dem Instanzenmodell hat somit 13,4 Stunden betragen.

11.2.3 Zeitdauer zur Nutzung einer Taxonomie

Basierend auf den gemessenen Werten für das Instanzenmodell kann abgeleitet werden, wie lange die Erstellung der entsprechenden Taxonomie gedauert hätte (siehe Tabelle 11.3). Hierbei wären nur hierarchische Relationen erstellt worden. Die Zeit zur Erstellung der Taxonomie, aus 498 semantischen Elementen bestehend, hätte dadurch nur 250 Minuten benötigt.

Da im Instanzenmodell 2,15 benannte Relationen erstellt wurden pro in der Beschriftung verwendetem semantischen Element, muss man bei der Nutzung der

Bezeichner	Wert
Gesamtzeit Erstellung Ontologie (Minuten)	0
Anzahl Bilder im Archiv	722
Durchschnittliche Anzahl Beschriftungen mit Taxonomie pro Bild	6,7
Resultierende Anzahl gleichwertiger textbasierter Beschriftungen (Verdopplung) pro Bild	13,4
Anzahl textbasierter Beschriftungen gesamt	9674
Zeitdauer Erstellung textbasierter Beschriftungen (Minuten)	1612
Gesamtzeit in Stunden	26,9

Tabelle 11.4: Abgeleitete Zeitdauer zur Beschriftung eines Bildarchivs mit textbasierten Metadaten

Taxonomie wie in Abbildung 11.1 gezeigt die entsprechenden semantischen Elemente nicht durch Relationen, sondern durch direkte Beschriftung mit dem Bild in die semantischen Metadaten mit einbinden. Entsprechend müssen statt 2238 Beschriftungen insgesamt 4812 Beschriftungen erstellt werden.

Die Erstellung der semantischen Metadaten mit einer solchen Taxonomie würde dann 1052 Minuten (berechnet aus $4812 \text{ Beschriftungen} * 10 \text{ Sekunden} / 60 = 802 \text{ Minuten} + 250 \text{ Minuten für die Ontologieerstellung}$) und somit insgesamt 17,5 Stunden benötigen.

11.2.4 Zeitdauer zur Erstellung textbasierter Beschriftungen

Um die Zeitdauer zu berechnen, die zur Erstellung textbasierter Bildbeschriftungen entstanden wäre, wurde die hierzu nötige Menge an nötigen Stichwörtern errechnet. Wie Tabelle 11.4 zeigt, wurden hierbei 13,4 zu verwendende Stichwörter ermittelt. Diese ergeben sich aus der Verdopplung der 6,7 nötigen Bildbeschriftungen in der Taxonomie, da jeweils ein weiteres Stichwort in der Beschriftung mit anzugeben ist, welches die generelle Information des semantischen Element in der Taxonomie ausdrückt, also z.B. „Frau“ für „Elizabeth Taylor“ (siehe Abbildung 11.3).

Entsprechend müssen 9674 Stichwörter für die Erstellung von textbasierten Bildbeschriftungen, damit diese mit dem Instanzenmodell und seinen Hintergrundinformationen gleichwertig sind. Vorhandene Synonyme und unterschiedliche Sprachen eines Instanzenmodells werden hierbei in der Auswertung nicht mit berücksich-

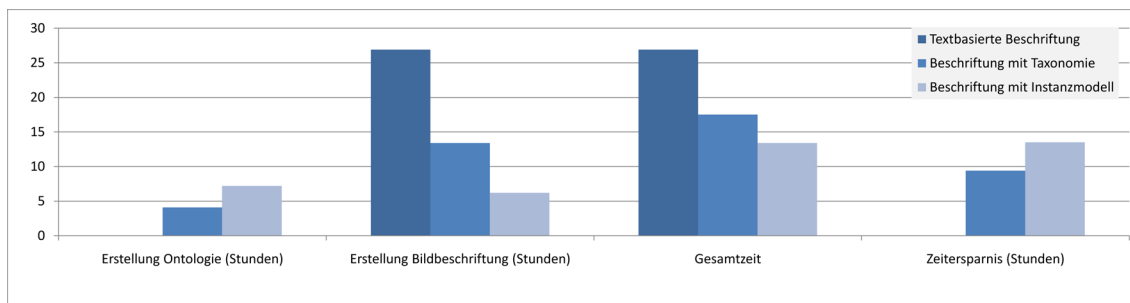


Abbildung 11.4: Vergleich der Zeitdauer zur Bildbeschriftung mittels Text, Taxonomie und Instanzenmodell

tigt, diese hätten die Gesamtzeit zusätzlich erhöht. Die Erstellung der textbasierten Bildbeschriftungen würde somit 26,9 Stunden (basierend auf der ermittelten Mindestzeitdauer von 10 Sekunden zum Eintrag eines Stichworts zur Bildbeschriftung) betragen. Auch qualitative Aspekte können in dieser Modellrechnung nicht berücksichtigt werden, also beispielsweise zu ermitteln, wie häufig wichtige Begriffe in den Metadaten einzelner Bilder vergessen wurden.

11.2.5 Gegenüberstellung der Messungen

Ausgehend von der durchgeführten Modellrechnung wäre im Vergleich zur textbasierten Bildbeschriftung somit die *Nutzung einer Taxonomie um 34 Prozent schneller* (siehe Abbildung 11.4). Die Nutzung des Instanzenmodells ist bei gleicher Qualität der semantischen Metadaten sogar *um 50,1 Prozent schneller im Vergleich zur textbasierten Bildbeschriftung*. Somit sorgt die Nutzung eines Instanzenmodells nicht nur dafür, dass alle Beschriftungen konsistent und aufgrund des abgebildeten Hintergrundwissens von einer hohen Qualität sind. Es sorgt auch dafür, dass die Bildarchivare rein rechnerisch und abgeleitet aus einer Praxismessung bis zu 50 Prozent an Zeit zur Erstellung von Bildbeschriftungen sparen können im Vergleich zu den in Bildarchiven überwiegend verwendeten textbasierten Verfahren.

*Der wichtigste Einfluss auf den Zeitgewinn mit Instanzenmodellen ist die Häufigkeit der Verwendung der erstellten *imagenotions* in der Bildbeschriftung. Jede *imagenotion* wurde durchschnittlich 4,15 Mal in semantischen Metadaten verwendet. Man kann hier annehmen, dass der Zeitgewinn der Wiederverwendung nicht besteht, wenn ein semantisches Element nur einmal oder seltener zur Erstellung semantischer Metadaten verwendet wird. Dann ist der Zeitaufwand zur Erstellung des semantischen Elements zu hoch. Da sich in einem Bildarchiv oft viele Bilder zu einem Thema oder Ding befinden, scheint dieses Problem in Bildagenturen nicht ge-*

geben. Von einer häufigen Wiederverwendung einzelner semantischer Element kann daher ausgegangen werden.

Bei Instanzenmodell dauert die Erstellung der Ontologie am längsten mit 443 Minuten. Bei der Taxonomie werden hierfür nur 250 Minuten benötigt. Anschließend sind aber mit 2238 nötigen Beschriftungen mit dem Instanzenmodell deutlich weniger Beschriftungsschritte als im Vergleich zur Taxonomie mit 4812 gleichwertigen Beschriftungen nötig. Zwar ist bei der textbasierten Bildbeschriftung keine Erstellung einer Ontologie nötig, dafür sind aber mit 9674 textbasierten Beschriftungen das ganze Hintergrundwissen aus der Ontologie immer wieder in die Bildbeschriftungen abzubilden.

In der Zeit zur Erstellung des Instanzenmodells mit 443 Minuten hätten textbasiert 2658 textbasierte Bildbeschriftungen erstellt werden können. Bei einem Bildarchiv mit weniger als 200 Bildern wäre somit die textbasierte Bildbeschriftung schneller gewesen, wenn das gleiche Instanzenmodell erstellt worden wäre. Ein Bildarchiv muss somit eine hinreichende Größe haben, damit die Erstellung semantischer Metadaten schneller ist als die gleichwertige Verwendung textbasierter Metadaten. Bei Bildarchiven von Bildagenturen ist dies der Fall.

11.2.6 Fazit

Auf der Basis der Auswertung der Erstellung semantischer Metadaten für ein Filmarchiv lässt sich daher Frage 2 (c) dieser Arbeit wie folgt beantworten:

Frage 2 (c)

Reduzieren semantische Verfahren den Zeitaufwand zur Erstellung von Metadaten??

Antwort

Die Auswertung des Zeitaufwands zur Beschriftung eines Filmarchivs mit 722 Bilder hat gezeigt, dass semantische Verfahren die Bildbeschriftung beschleunigen. Im Vergleich zur textbasierten Bildbeschriftung führen die Nutzung einer Taxonomie zu einer Zeitersparnis von 34 Prozent und die Nutzung eines Instanzenmodells sogar zu einer Zeitersparnis von 50,1 Prozent.

11.3 Zusammenfassung

Semantische Verfahren ermöglichen somit nicht nur die gewünschten hochwertigen Bildbeschriftungen. Sie führen gleichzeitig zu einer Zeitersparnis bei der Bildbeschrif-

tung um bis zu 50,1 Prozent bei der Nutzung von Instanzenmodellen im Vergleich zur textbasierten Bildbeschriftung!

Eine praxistaugliche Nutzung der semantischen Bildbeschriftung ist somit möglich, da der Zeitaufwand für die Bildbeschriftung deutlich gesenkt werden kann.

Kapitel 12

Automatische Erkennung von Bildinhalten

Der dritte Aspekt hin zu hochwertigen Suchergebnissen (siehe Sektion 2.5) ist das *Wissen einer Bildsuchmaschine um den Inhalt von Bildteilen*. Darauf basierend können die Bilder im Suchergebnis nach dem Inhalt der Bildteile geordnet und gefiltert werden. Die manuelle Beschriftung von Bildteilen ist aber sehr zeitaufwändig. Wünschenswert sind daher Verfahren zur automatischen Erkennung von Bildinhalten. Hierzu werden Objekterkennung eingesetzt [AHAQI⁺07].

Ziel dieses Kapitels ist es daher, zuerst den generellen Ablauf von Objekterkennung vorzustellen und welche Verfahren es gibt. Dies ermöglicht die Auswahl geeigneter Objekterkennung in einem Bildarchiv und die Klassifizierung der Fehlerarten, die beim Einsatz von Objekterkennung auftreten.

Eine Evaluation soll weiterhin zeigen, wie gut diese automatischen Verfahren zur Erkennung von Bildinhalten in der Praxis funktionieren. Daher werden Objekterkennung zur Erkennung von Inhalten innerhalb der Bilder eines Filmarchivs eingesetzt und die Ergebnisse mit zuvor von Bildarchivaren manuell erstellten Bildbeschriftungen verglichen.

Auf Basis der Evaluationsergebnisse soll abgeleitet werden, wie das Gesamtergebnis automatischer Verfahren verbessert werden könnte. Die Idee hierbei soll sein, die Ergebnisse einzelner Objekterkennung geeignet zu kombinieren. Die damit verbundene Frage (3) in dieser Arbeit lautet dabei *Verbessert die Kombination der Ergebnisse automatischer Verfahren zur Beschriftung von Bildteilen das Gesamtergebnis?*

12.1 Erkennung von Objekten auf Bildteilen

Im Folgenden werden der generelle Arbeitsablauf und mögliche Arten von Objekterkennern vorgestellt. Ziel ist hierbei, zu ermitteln, welche Objekterkennner sich für den Einsatz in einem Bildarchiv eignen und wie diese dort integriert werden müssen.

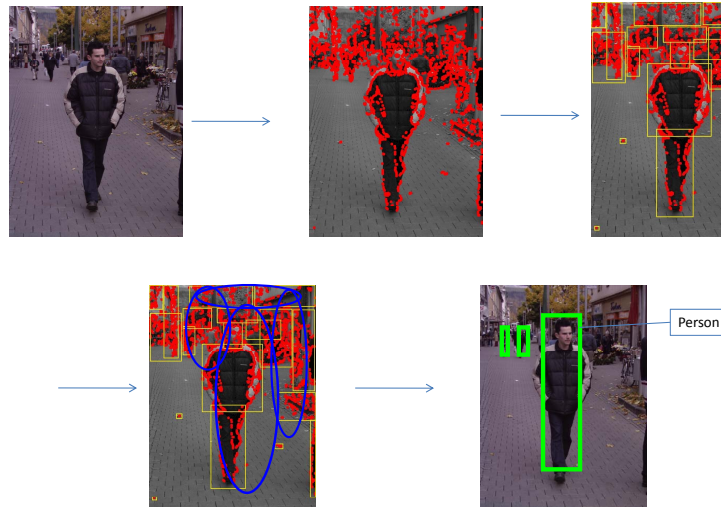


Abbildung 12.1: Ablauf der Extraktion und Erkennung von Objekten

Abbildung 12.1 zeigt basierend auf dem Beispiel der Erkennung von Personen die generell nötigen fünf Schritte zur Objekterkennung [KSH⁺04]. Diese sind das (1) Laden des Bilds, (2) Erkennen von wichtigen Bildpunkten, (3) Zusammenfassung von Bildteilen zu Objekten, (4) Erkennung des Objekts und (5) Beschriftung des Bildes.

Zusammenfassung von Bildteilen zu Objekten. Der schwierigste Teil hierbei ist *Schritt 3*, also das Zusammenfassen von Bildteilen zu Objekten, da die Verfahren eine möglichst hohe Toleranz gegenüber unterschiedlichen Bildbelichtungen, dem Hintergrund der Objekte sowie verdeckter Teile der Objekte zu erkennen den Objekte durch andere Objekte haben müssen. Unterschiedliche Objekterkennner arbeiten somit hauptsächlich verschieden in Schritt 3 und setzen hierbei Algorithmen wie k-means, self organizing maps und neurale Netzwerke, ein (Details zu den Algorithmen siehe [AHAQI⁺07]).

Unterschied der Erkennung von Standbildern und Videobildern. Mit Schritt 3 ist auch der Unterschied in der Erkennung von Objekten in Videos im

Vergleich zu Standbildern ersichtlich. Ein Video ist eine *Sequenz von Bildern*, so dass ein Objekt aus mehreren Einzelbildern extrahiert werden kann. Das Problem von teilweise verdeckten Objektteilen ist dann nicht so gravierend [AHAQI⁺07]. Auch können Menschen, Tiere oder bewegliche Objekte leichter erkannt werden, da sich diese in der Sequenz der Bilder bewegen und so das Objekt leichter vom Hintergrund getrennt werden kann. *Die Erkennung von Objekten innerhalb von Standbildern ist somit ein deutlich schwierigeres Problem als die Erkennung von Objekten in Videos.*

Erkennung von Objekten. Eine weitere Schwierigkeit liegt anschließend in der Erkennung von Objekten durch einen geeigneten Abgleich mit vorhandenen *Trainingsdaten*.

Diese Trainingsdaten werden zur Erkennung von Objekten mit den extrahierten Objekten abgeglichen. Hierbei ermittelt der Objekterkenner jeweils, wie hoch die Wahrscheinlichkeit ist, dass ein Bildteil einem Objekt aus den Trainingsdaten entspricht. Eingesetzte Verfahren hierzu sind unter anderem Support-Vector-Maschinen [LZLM05], statistische Modelle [LZLM05] und Klassifizierer [CGSW03]. Das als am wahrscheinlichsten erkannte Objekt wird anschließend zur Beschriftung des entsprechenden Bildteils verwendet.

Unabhängig vom eingesetzten Verfahren ist somit für alle Objekterkenner *das Training der eingesetzten Algorithmen mittels entsprechender Bilder nötig*, welche die zu erkennenden Objekte abbilden.

12.2 Erstellung von Trainingsdaten

Basierend auf der *Art der Erstellung von Trainingsdaten* kann nach zwei verschiedenen Arten von Objekterkennern unterschieden werden. Dies sind die Erstellung von *universellen* und *spezialisierten* Objekterkennern.

12.2.1 Universelle Objekterkenner

Universelle Objekterkenner *lernen* während der Bilderkennung aufgrund der *Rückmeldung durch Benutzer* kontinuierlich hinzu.

Ein solcher universeller Objekterkenner wird zum Beispiel im System ALIPR [LW06] eingesetzt. Der Benutzer kann hierbei ein Bild aufladen, das System schlägt anschließend Bilder vor, die basierend auf den Trainingsdaten in Frage gleiche Objekte zeigen. Der Benutzer gibt anschließend an, welche Bilder als richtig ermittelt worden sind und welche nicht. Durch die Rückmeldung des Benutzers zu erkannten Objekten lernt ALIPR dann die Trainingsbilder kennen, die zur Erkennung einzelner

Objekte verwendet werden sollen.

Das Hauptproblem der universellen Objekterkenner ist, dass pro zu erkennendem Objekt für präzise Ergebnisse ein *Trainingsbestand von mehr als 100 Bildern pro Objekt* vorhanden sein sollte [LW06]. Weiterhin ist das Training der Objekterkenner somit sehr zeitaufwändig und fehleranfällig, da nämlich nur präzise Rückmeldungen des Benutzers zu qualitativ guten Ergebnissen in der anschließenden Erkennung dieser Objekten innerhalb unbekannter Bilder führen können.

12.2.2 Spezialisierte Objekterkenner

Spezialisierte Objekterkenner benötigen während der Nutzung zur Erkennung von Objekten in neuen Bildern keinen weiteren Trainingsaufwand. Auf der Basis eines umfangreichen und korrekt beschriebenen Bildbestandes erstellen die spezialisierten Objekterkenner einmalig ihr Trainingsmodell über das zu erkennende Objekt.

Die Entwickler geben den spezialisierten Objekterkenner somit Trainingsdaten mit, damit diese bestimmte Objekte wie Personen [STMA09], Gesichter [KE06], Tiere [AHAQI⁺07], Autos oder Flugzeuge [STMA09] mit einer hohen Präzision erkennen können.

12.2.3 Fazit

Aufgrund der meist besseren Trainingsmodelle [AHAQI⁺07] und weiterhin, da während der Nutzung der spezialisierten Objekterkenner zur Bilderkennung keine Erstellung von Trainingsdaten mehr nötig ist, eignen sich diese besser im praktischen Einsatz zur Erkennung von Objekten auf Bildteilen. *Daher konzentriert sich diese Arbeit im Weiteren auf den Einsatz spezialisierter Objekterkenner.*

12.3 Identifikation von Objekten

Mit Hilfe der *Identifikation von Objekten* sollen erkannte Objekte zusätzlich identifiziert werden. Die Objektidentifikation kann unterteilt werden nach der *Zuordnung von Objekten* und nach der *Erkennung von Merkmalen* innerhalb eines Bildteils.

12.3.1 Zuordnung von Objekten

Mit der *Zuordnung von Objekten* wird ein erkanntes Objekt genauer *zugeordnet*. Dies kann zum Beispiel die Identifikation einer abgebildeten Person sein, wenn ein



Abbildung 12.2: Erkennung der Emotionen *glücklich*, *überrascht*, *traurig* und *wütend*

Objekt als Gesicht erkannt wurde. Oder die genaue Zuordnung einer Automarke, wenn ein Auto erkannt wurde.

Hierzu ist der Einsatz von *Objektzuordnern* nötig [TCZZ06]. Objektzuordner sind dabei ebenfalls spezialisierte Objekterkenner, die aber zusätzlich auf die Zuordnung von spezifischen Objekten trainiert werden. So ist zum Beispiel ein Gesichtserkenner ein spezieller Objekterkenner, der zusätzlich auf die zu erkennenden Gesichter im Bildbestand trainiert wird. Entsprechend kann ein Erkenner für Automarken diese zuordnen können, wenn er zuvor speziell auf zu erkennende Automarken trainiert wurde.

Beim Einsatz von Objektzuordnern muss somit bei der Verwendung in einem Bildarchiv die Möglichkeit zur Erstellung spezieller Trainingsdaten pro zuzuordnendem Objekt vorgesehen werden.

12.3.2 Erkennung von Merkmalen

Ein weiterer Aspekt bei der Identifikation von Objekten ist die *Erkennung von Merkmalen* in einem erkannten Bildteil. Innerhalb eines Bildteils sollen so entweder *weitere wichtige Bildteile* gefunden werden, wie zum Beispiel das Autokennzeichen, wenn ein Auto gefunden wurde.

Weiterhin sollen *Eigenschaften des extrahierten Bildteils* erkannt werden. Bei Gesichtern können dies zum Beispiel *Emotionen der Person* sein oder das *Geschlecht* und das *geschätzte Alter* einer Person. Abbildung 12.2 zeigt einen Emotionserkenner zur Erkennung der *Emotion einer Person*, diese sind glücklich, überrascht, traurig

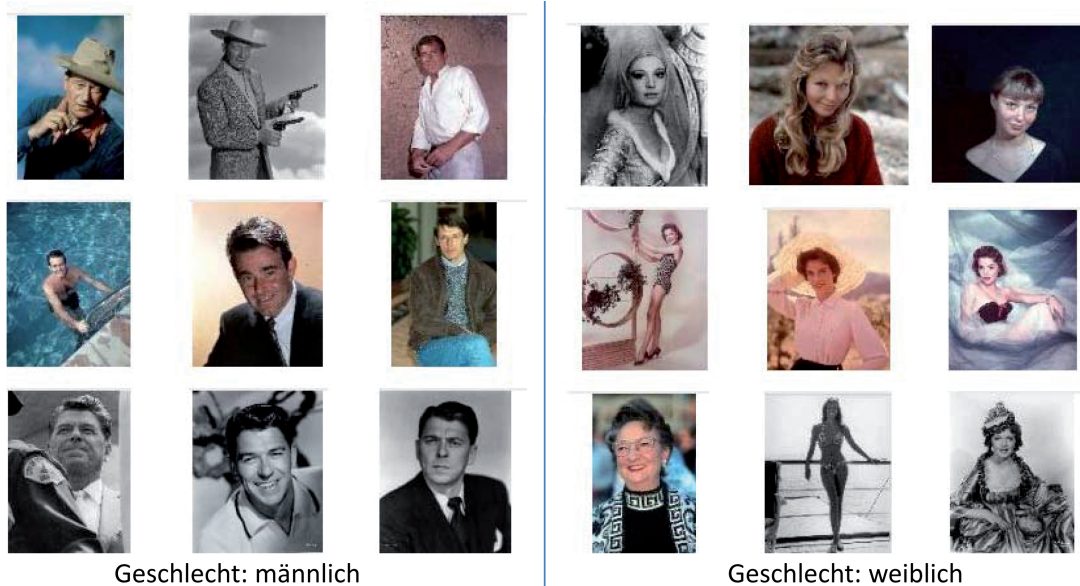


Abbildung 12.3: Zuordnung des Geschlechts in Bildteilen, die Personen zeigen

und wütend. Abbildung 12.3 zeigt einen Geschlechterkennner zur Erkennung der Eigenschaft *Geschlecht einer Person*, also männlich beziehungsweise weiblich.

Die Erkennung von Details erfolgt somit durch den Einsatz von weiteren spezialisierten Objekterkennern. Der gezeigte Emotionskennner in Abbildung 12.2 wurde beispielsweise laut Aussagen des Herstellers Fraunhofer IIS (Projektpartner im IMAGINATION Projekt) mit mehr als 100.000 Bildern trainiert, die Emotionen aufgrund einer Klassifikation von mehr als 10.000 Punkten im Gesicht ermöglichen.

12.4 Kategorisierung der auftretenden Fehler

Aus den in [DJLW08] und [SGCD09] beschriebenen Problemen bei der Objekterkennung und Identifikation lassen sich bei der Erkennung und Identifikation von Objekten die folgenden möglichen Fehler unterscheiden. Abbildung 12.4 gibt hierzu einen Überblick.

- **(1) Objekt wurde nicht erkannt.** Ein Objekt wurde im Bild nicht erkannt, wenn es von einem Objekterkennner nicht beschriftet wurde. Mögliche Ursachen können sein, dass das Objekt zu extreme Positionen hat, zum Beispiel Gesichter von liegenden Personen oder Seitenansichten von Personen. *Der Benutzer muss daher die Beschriftung des Bildteils selbst durchführen.*



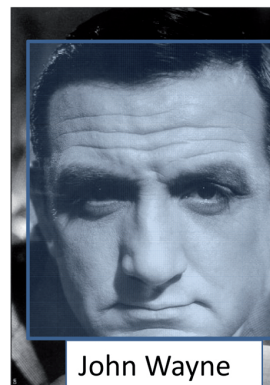
(1) Objekt nicht erkannt (Gesicht)



(2) Fantomobjekte



(3) Objekt falsch erkannt



(4) Objekt falsch zugeordnet

Abbildung 12.4: Fehlerarten *Bildteil nicht erkannt*, *Fantomobjekte*, *falsch erkanntes Objekt*, *falsch identifiziertes Objekt*

- **(2) Fantomobjekte.** Dieser Fehler liegt vor, wenn ein Bildbereich markiert wird, der kein Objekt zeigt, und wird als Fantomobjekt oder Fantomgesicht bezeichnet [SGCD09]. *Ein Benutzer muss derartige Fehlbeschriftungen manuell löschen* und hat somit zusätzlichen Aufwand.
- **(3) Objekte falsch erkannt.** Bei diesem Fehler wird ein Objekt zwar markiert, aber falsch erkannt. In Abbildung 12.4 wurde das abgebildete Flugzeug falsch als Person erkannt. Ein Benutzer hat hier also den Aufwand, diese Beschriftungen von Bildteilen als falsch zu erkennen und anschließend zu korrigieren.
- **(4) Objekte falsch zugeordnet.** Dieser Fehler entsteht bei der Identifikation von Objekten. Das Objekt oder zu erkennende Detail wird falsch identifiziert. Die in Abbildung 12.4 gezeigte Person „Lino Ventura“ wurde falsch als „John

Wayne“ identifiziert. Der Benutzer muss somit die spezifische Information zur erkannten Person korrigieren.

Konsequenz: Ergebnisse müssen validiert werden. Die Ergebnisse bei der Objekterkennung und Identifikation müssen daher aufgrund möglicher Fehler durch Benutzer validiert werden. Auftretende Fehler müssen dann manuell korrigiert werden können.

Zeitersparnis mit automatischen Verfahren Aufgrund der gezeigten Fehlerarten kann nur dann eine Zeitersparnis beim Einsatz von automatischen Verfahren zur Bildbeschriftung ermöglicht werden, *wenn die Anzahl der nötigen Arbeitsschritte zur Erstellung manueller Korrekturen kleiner als die nötigen Arbeitsschritte zur manuellen Beschriftung aller Bildteile ist.*

<p>Ersparnis durch automatische Verfahren</p>
--

<p><i>Arbeitsschritte für manuelle Korrekturen < Arbeitsschritte Beschriftung der Bildteile</i></p>
--

12.5 Laboruntersuchung 1: Objekterkenner im praktischen Einsatz

Genereller Zweck des Einsatzes automatischer Verfahren in einem Bildarchiv soll sein, *dass die zeitaufwändige Beschriftung von Bildteilen minimiert wird.* Die Arbeitsschritte für manuelle Korrekturen müssen also deutlich kleiner sein als die Arbeitsschritte zur manuellen Beschriftung der Bildteile.

12.5.1 Motivation

In der folgenden Evaluation ist von Interesse, wie viele Fehler geeignete Objekterkenner in einem Filmarchiv erzeugen. Dies repräsentiert ein zufällig gewähltes Bildarchiv bei einer Bildagentur. Wichtig ist hierbei zu erwähnen, dass es sich bei deren Bildern um qualitativ hochwertige Bilder handelt. Die Bilder sind gut belichtet, von guter Schärfe und liegen auch in hoher Auflösung vor. Entsprechend werden die Ergebnisse der folgenden Evaluation nicht verfälscht durch unkorrekte Ergebnisse aufgrund einer schlechten Bildqualität.

Mit der Evaluation soll ermittelt werden, wie oft die genannten Fehlerarten bei der Objekterkennung auftreten und ob diese Fehlermengen hinsichtlich der nötigen Korrekturschritte vertretbar sind.

Inhalt	Anzahl
Bilder gesamt	722
Bildteile Personen gesamt	1008
Personen gesamt	326
Weibliche Personen	110
Männliche Personen	216
Verschiedene Filmtitel	73
Verschiedene Schauspieler	171
Uma Thurman	185
Ewan McGregor	121
Clint Eastwood	34
Autos	22
Flugzeuge	9

Tabelle 12.1: Übersicht über die Inhalte im Filmarchiv

12.5.2 Vergleichsbasis: manuell erstellte Bildbeschriftungen

Als Basis für den Vergleich der Ergebnisse der Objekterkenner dienen 722 Bilder, die von Bildarchivaren der Bildagentur photo12 und BSMC innerhalb des IMAGINATION Projekts mit semantischen Metadaten versehen wurden. Auf Bildteilen, die wichtige Personen und Objekte zeigen, wurden manuelle Beschriftungen durchgeführt.

Tabelle 12.1 gibt einen Überblick über die von den Bildredakteuren manuell erstellten Beschriftungen auf wichtigen Bildteilen im Filmarchiv. 326 verschiedene Personen wurden auf Bildteilen beschriftet. Von diesen sind 110 Personen vom Geschlecht weiblich und 216 Personen sind männlich. Die Filmszenen zeigen Szenen von insgesamt 73 Filmen, die 171 verschiedene Schauspieler anzeigen. Andere Personen haben beispielsweise den Beruf Regisseur. Auf 185 der Bilder ist Uma Thurman zu sehen, auf weiteren 121 Bildern Ewan McGregor sowie auf 34 Bildern Clint Eastwood. 22 Bilder zeigen weiterhin Autos als Bildteile und 9 Bilder zeigen Flugzeuge.

12.5.3 Auswahl geeigneter Objekterkenner

Ein Filmarchiv enthält häufig Aufnahmen von Personen als Ganzes oder den Oberkörper sowie den Gesichtern der Schauspieler, welche in den Filmen mitspielen. Da in Kapitel 4 ermittelt wurde, dass sich etwas 40 Prozent der Suchanfragen auf Personen bezogen haben, ist die Nutzung von *Personenerkennern*, *Gesichtserkennern* und weiterhin zur Identifikation von Personen *Gesichtsdetektion* notwendig. Ebenso bie-

tet sich der Einsatz von *Geschlechts-* und *Emotionserkennern* an. Zur Erkennung der Autos und Flugzeuge im Filmarchiv eignen sich *Autoerkenner* und *Flugzeug-erkenner*. Die folgenden Implementierungen von entsprechenden Objekterkennern wurden in der Evaluation verwendet:

- **Personenerkener NTUA** Der Personenerkener des IMAGINATION Projektpartners NTUA [STMA09] wird zur Erkennung von ganzen Personen und Körperteilen wie zum Beispiel dem Oberkörper einer Person verwendet.
- **Gesichterkenner IIS** Der vom Fraunhofer IIS zur Verfügung gestellte Gesichtserkener wird zur Erkennung von Gesichtern verwendet.
- **Gesichterkenner face.com** Weiterhin wird der frei nutzbare Gesichtserkener face.com[WHT09] verwendet zum Vergleich der Ergebnisse mit dem Gesichtserkener von IIS.
- **Geschlechtserkener IIS** Zur Erkennung des Details Geschlechts wird der Geschlechtserkener von IIS verwendet.
- **Gesichtidentifikation IIS** Zur Identifikation von Gesichtern wird die Gesichtsidentifikation von IIS verwendet
- **Gesichtsidentifikation face.com** Weiterhin wird zum Vergleich der Ergebnisse der Gesichtsidentifikation der Gesichtsidentifizierer von face.com eingesetzt.
- **Flugzeugerkennung und Autoerkennung NTUA** Die von NTUA entwickelte spezialisierten Objekterkener für Flüge und Autos wurden weiterhin eingesetzt, um die im Bildarchiv vorhandenen Autos und Flugzeuge erkennen zu können.

Auf den Einsatz des Emotionserkener von Fraunhofer IIS wurde in der Evaluation verzichtet. Der Grund hierfür war die Erkenntnis, dass verschiedene Bildarchivare die Emotionen einer Person oft subjektiv unterschiedlich eingeschätzt haben. Da somit die Beschriftungen der Bildredakteure nicht einheitlich waren, war es schwierig, die richtige Emotion zu ermitteln, welche der Emotionserkener hätte erkennen sollen.

12.6 Auswahl geeigneter Textinterpretierer

Textinterpretierer ermöglichen die Zuordnung beziehungsweise Interpretation von wichtigen Textstellen in vorhandenen, textbasierten Metadaten eines Bildes hin zu

semantischen Metadaten. Hat ein Textinterpretierer in einer textbasierten Beschriftung „Elizabeth Taylor“ als weibliche Person zugeordnet, so kann dieses Hintergrundwissen ausgenutzt werden, um automatisch ein semantisches Element in einer Ontologie zu erstellen. Der Textinterpretierer könnte so ein Element *Elizabeth Taylor* erstellen und automatisch die Relation „hat Geschlecht“ weiblich vorgeschlagen. Dies unterstützt somit die Abbildung von Hintergrundwissen in Ontologien. So kann, falls zusätzlich auch erkannt wird, dass es sich bei der gefundenen Person um eine Schauspielerin handelt, zusätzlich auch die Relation „hat Beruf“ *Schauspieler* vom Textinterpretierer vorgeschlagen werden. Daher sollen auch die möglichen Ergebnisse von Textinterpretierern in der Evaluation von Interesse sein. Als geeignete Textinterpretierer beziehungsweise zu erkennende Domänen sollen in einem Filmarchiv folgende Dinge und Sachverhalte erkannt werden:

- **Erkennung von Namen** Der *Namenserkenner* soll Stellen in den textbasierten Metadaten finden, die sich auf Namen von Personen beziehen. Hierzu wurde als Wissensbasis für den Namenserkenner aus dem XML Export der Liste von Vornamen aus Wikipedia erstellt. Der Namenserkenner übernimmt dann bei gefundenen Vornamen weitere groß geschriebene Wörter neben dem Namen und generiert hieraus den Bezeichner eines gefundenen semantischen Elements für eine Person.
- **Erkennung von Geschlecht.** Der *Geschlechtserkenner* soll das Geschlecht von Namen erkennen. Hierzu wurde aus dem XML Export der Liste weiblicher Vornamen und männlicher Vornamen die Wissensbasis für den Geschlechtserkenner erstellt.
- **Erkennung von Schauspielern.** Ebenfalls aus dem Namen soll ermittelt werden, ob es sich hierbei um einen Schauspieler handelt. Als Wissensbasis hierzu wurde der XML Export der Liste aller Schauspieler aus Wikipedia verwendet.
- **Erkennung von Filmtiteln.** Basierend auf der Wissensbasis der Filmtitel aus dem Internetportal IMDB¹ wurde ein Textinterpretierer für Filmtitel erstellt.

Die Textinterpretierer selbst wurden während des IMAGINATION Projekts vom Partner JSI auf der Basis von [BF06] entwickelt.

12.6.1 Evaluation

Durchführung der Evaluation Die Evaluation wurde durchgeführt, indem jeder der genannten Objekterkenner die Bilder ohne Beschriftungen erhielt und au-

¹The internet movie database, www.imdb.com

Task	Ablauf	Schritte gesamt
Manuelle Beschriftung Bildteil (Hinzufügen fehlender Bildteil) (Beschriftung Bildteil)	Markierung Bildteil Auswahl sem. Element Zuweisung Bildteil	3
Entfernung Phantom	Auswahl Bildteil Löschen Bildteil	2
Korrektur von falsch erkannt	Auswahl Bildteil Auswahl sem. Element Zuweisung Bildteil	3
Korrektur von falsch identifiziert	Auswahl Bildteil Auswahl sem. Element Zuweisung Bildteil	3

Tabelle 12.2: Arbeitsschritte für Erstellung und Korrektur von Bildteilen

tomatisch Beschriftungen durchgeführte. Die beiden Gesichtidentifizierer erhielten zusätzlich jeweils fünf Trainingsbilder von *Uma Thurman*, *Ewan McGregor* und *Clint Eastwood*, um diese anschließend identifizieren zu können. Die Beschriftungen wurden dann mit den manuell erstellten Beschriftungen verglichen, um so die richtig und falsch erkannten Beschriftungen zu zählen und nach Fehlerart zu kategorisieren.

Arbeitsschritte für Beschriftungen Tabelle 12.2 gibt einen Überblick über die einzelnen Arbeitsschritte die nötig sind zur Erstellung von manuellen Bildbeschriftungen (3 Arbeitsschritte), Entfernung von Phantomen (2 Arbeitsschritte), Korrektur falsch erkannter Bildteile (3 Arbeitsschritte) sowie der Korrektur falsch identifizierter Objekte (3 Arbeitsschritte). Die manuelle Erstellung von Bildbeschriftungen dauert demnach genau so lange wie die Korrektur einer falschen Zuweisung durch Objekterkenner. Die genannten Mengen an Arbeitsschritten wurden zur Auswertung der Ergebnisse der Objekterkenner verwendet, um den insgesamt nötigen Korrekturaufwand zu messen.

Ergebnisse der Objekterkenner Tabelle 12.3 gibt einen Gesamtüberblick der Ergebnisse. Deutlich fällt auf, dass nur drei Objekterkenner nennenswert die nötigen Arbeitsschritte reduzieren. Dies sind die beiden Gesichtserkenner von IIS und face.com sowie der Geschlechterkenner von IIS. Diese drei Verfahren reduzieren die nötigen Schritte zur Beschriftung von Gesichtern und dem Geschlecht der Person um mehr als 30 Prozent, der Gesichtserkenner von IIS sogar um 57,5 Prozent. Der Gesichtserkenner von IIS erkennt dabei zwar einige Gesichter weniger als der von

Objekterkenner	Richtig erkannte Bildteile	Nicht erkannte Bildteile	Phantome	Falsch erkannt	Falsch zugeordnet	Arbeitsschritte manuelle Erstellung	Nötige Korrekturschritte	Summe Arbeitsschritte	Ersparnis Prozent
Manuelle Beschriftung Person / Gesicht	1008	0	0	0	0	3024	0	3024	0
Personen-erkenner NTUA	304	704	438	10	0	2112	906	3018	0,002
Gesichts-erkenner IIS	699	317	167	0	0	951	334	1285	57,5
Gesichts-erkenner face.com	746	262	504	0	0	786	1008	1794	40,7
Geschlechts-erkenner IIS	521	487	0	0	172	1461	516	1977	34,6
Manuelle Identifikation Gesicht	340	0	0	0	0	1020	0	1020	0
Gesichts-identifikation IIS	65	275	0	0	64	825	192	1017	0,01
Gesichts-identifikation face.com	73	267	0	0	73	801	219	1020	0
Manuelle Beschriftung Autos	22	0	0	0	0	66	0	66	0
Manuelle Beschriftung Flugzeuge	9	0	0	0	0	18	0	18	0
Autoerkenner NTUA	2	20	135	0	0	60	270	330	–
Flugzeugerkenner NTUA	1	8	95	0	0	24	190	214	–

Tabelle 12.3: Übersicht über die Ergebnisse der Objekterkenner

Bezeichner	Im Archiv vorhanden	Erkannt	in Prozent
Name	326	306	93,9
Geschlecht weiblich	110	102	93,7
Geschlecht männlich	216	204	94,4
Filmtitel	73	53	72,6
Schauspieler	171	138	80,7

Tabelle 12.4: Erkannte semantische Elemente durch Textinterpretierer

face.com (699 vs. 746 Gesichter), allerdings produziert der Gesichtserkennung mit 504 Phantomgesichtern (vs. 167 bei IIS) deutlich mehr falsche Beschriftungen.

Hohe Zahl an Phantomen beim Personenerkennung Sehr störend beim Personenerkennung der NTUA ist, dass eine hohe Anzahl an falschen Beschriftungen erstellt wurde. Dadurch wurden insgesamt mehr Fehler (448) als richtige Beschriftungen (304) erstellt. Entsprechend ist ein Zeitersparnis mit nur sechs Schritten gegenüber der manuellen Beschriftung so gut wie nicht vorhanden. Ein Einsatz in einer Bildsuchmaschine ist daher fraglich.

Autoerkennung und Flugzeugerkennung nicht praktisch einsetzbar Gänzlich unbrauchbar sind der Autoerkennung und der Flugzeugerkennung. Diese haben ein Vielfaches mehr an Beschriftungen erstellt, als eigentlich passende Objekte im Archiv wären. Der Arbeitsaufwand zur Korrektur ist dadurch um ein Vielfaches höher als der Aufwand zur Erstellung von manuellen Beschriftungen. Der Einsatz dieser beiden Objekterkennung ist daher nicht geeignet in einer Bildsuchmaschine einer Bildagentur.

Wenig Gesichter identifiziert Ähnlich wie der Personenerkennung tragen auch die von IIS und face.com eingesetzten Gesichtserkennung zu keiner Reduktion des Arbeitsaufwands bei. Dies kann aber auch daran gelegen haben, dass diesen die falschen Trainingsdaten zur Verfügung gestellt wurden.

Ergebnisse der Textinterpretierer Tabelle 12.4 gibt einen Überblick über den Vergleich der von Textinterpretierer erkannten semantischen Elementen mit den manuell erstellten Elementen. Wie der Vergleich zeigt, haben die Textinterpretierer Namen und das Geschlecht von Personen im Bildarchiv mit je mehr als 90 Prozent sehr gut erkannt. Bei den Filmtiteln kann die etwas niedrige Rate von 72,6 Prozent auch daran liegen, dass einige der Filmtitel in den textbasierten Metadaten mit den französischen Originaltiteln standen, die Wissensbasis enthielt aber nur die englischsprachigen Titel.

Textinterpretierer tragen somit hilfreich zur Erkennung wichtiger semantischer Elemente in bestehenden textbasierten Metadaten bei.

12.6.2 Bewertung und Überlegungen zu den Ergebnissen

Als Bewertung der Ergebnisse kann gesagt werden, dass die Objekterkennung *insgesamt noch zu viele Fehler generieren und damit die Ergebnisse noch nicht wirklich zufriedenstellend sind.*

Um die Ergebnisse zu verbessern, können daher folgende Überlegungen zu den Ergebnissen dienen:

- **Reduzierung der Fehler nötig** Als einzige Objekterkennung in der Evaluation haben die Gesichtserkennung von IIS bzw. face.com zu einer deutlichen Reduktion der nötigen Arbeitsschritte geführt. Gleichzeitig hätten diese noch Potential zu mehr Einsparung, wenn die Menge an falsch erkannten Markierungen reduziert werden könnte.
- **Kombination der Ergebnisse sinnvoll** Bei der Betrachtung der Gesichtserkennung fällt weiterhin auf, dass die Menge an richtig erkannten Gesichtern sich bei beiden Objekterkennern unterscheidet. Daraus kann man schließen, dass man durch Verwendung beider Ergebnisse und einer geeigneten Kombination die Menge an richtig erkannten Objekten hätte erhöhen können. Beim Personenerkennung ist störend, dass dieser eine hohe Menge an Fehlern erstellt hat. Es stellt sich daher die Frage, ob es nicht sinnvoll wäre, wenn ein Gesichtserkennung noch einmal detailliert die vom Personenerkennung erstellten Bereiche überprüft. Für eine markierte Person in einem Filmarchiv sollte auch das Gesicht gut sichtbar sein. Wird also kein Gesicht gefunden, läge es nahe, die markierte Stelle als falsch einzustufen.
- **Unterstützung von Objektidentifikation durch Textinterpretierer erscheint hilfreich** Die Betrachtung der Textinterpretierer hat gezeigt, dass diese sehr zuverlässig Namen erkennen können. Gleichzeitig hat die Gesichtsidentifikation sehr wenig richtige Ergebnisse generiert. Es stellt sich also die Frage, inwieweit die Kombination der Ergebnisse von Textinterpretierer und Gesichtserkennung zum Einen zu automatisch erstellbaren Trainingsdaten für die Gesichtsidentifikation und weiterhin zu einer höheren Menge an richtig erkannten Gesichtern führen kann.

Diese Überlegungen führten zur Idee des nächsten Kapitels, also zu einer umfangreichen Kombination der Ergebnisse automatischer Verfahren zur Text- und Objekterkennung.

12.6.3 Zusammenfassung

Innerhalb dieses Kapitels wurden Objekterkennungsvorgänge vorgestellt und gezeigt, wie diese sich in einem Bildarchiv einsetzen lassen. Weiterhin konnten die dabei auftretenden Fehler der Verfahren nach Fehlerarten unterteilt werden. Die Evaluation hat als gravierendstes Problem ergeben, dass häufig Stellen die kein Objekt zeigen, in einem Bild falsch markiert werden. Dadurch ist für einen Benutzer zusätzlicher Korrekturaufwand nötig.

Damit der Einsatz automatischer Prozesse überhaupt lohnenswert in einem Bildarchiv ist, muss die Menge an richtig erkannten Bildteilen möglichst hoch sein sowie die Fehler durch die Objekterkennungsmethoden möglichst niedrig. *Nur dann können sich die nötigen Arbeitsschritte im Vergleich zur manuellen Bildbeschriftung deutlich reduzieren.*

Die entsprechende Idee der Kombination der Ergebnisse automatischer Verfahren soll hierzu ein Lösungsansatz sein und wird daher im nächsten Kapitel entsprechend konzeptioniert sowie durch eine Evaluation untersucht.

Kapitel 13

Kombination automatischer Verfahren

Die Evaluation im letzten Kapitel hat angeregt, die Menge richtig erkannter Personen auf Bildteilen dadurch zu erhöhen, dass mehrere Gesichterkenner miteinander kombiniert werden. Gleichzeitig besteht die Aussicht auf weniger falsch erkannte Bildteile durch die Kombination verschiedener Objekterkenner wie Gesicht- und Personenerkennung, da die beiden Verfahren ihre Ergebnisse gegenseitig bestätigen können.

Ziel dieses Kapitels ist daher, eine geeignete Kombination automatischer Verfahren zur Erkennung von Bildteilen zu ermöglichen. So sollen vollständigere und gleichzeitig hochwertigere Gesamtergebnisse beim Einsatz von Objekterkennern entstehen. Die vorgeführten Verfahren werden anschließend evaluiert, um mögliche Verbesserungen messen zu können.

13.1 Grundidee im EU Projekt IMAGINATION

Abbildung 13.1 zeigt die Grundidee der Kombination automatischer Verfahren, wie diese im EU Projekt IMAGINATION¹ formuliert wurde. Im IMAGINATION Projekt wurde als Ausgangspunkt ein Bild mit textbasierten Metadaten angenommen². Ein Textinterpretierer soll in diesen textbasierten Metadaten wichtige Stellen ermitteln und daraus semantische Metadaten erstellen. Zur automatischen Erkennung von

¹IMAGINATION Projekt Webseite: www.imagination-project.org

²Anmerkung: In der eigenen Arbeit wird allerdings aus der Sicht eines Bildarchivs einer Bildagentur davon ausgegangen, dass nur für bestehende Bilder textbasierte Beschriftungen vorliegen. Für neue Bilder im Archiv liegen also keine textbasierten Beschriftungen vor.

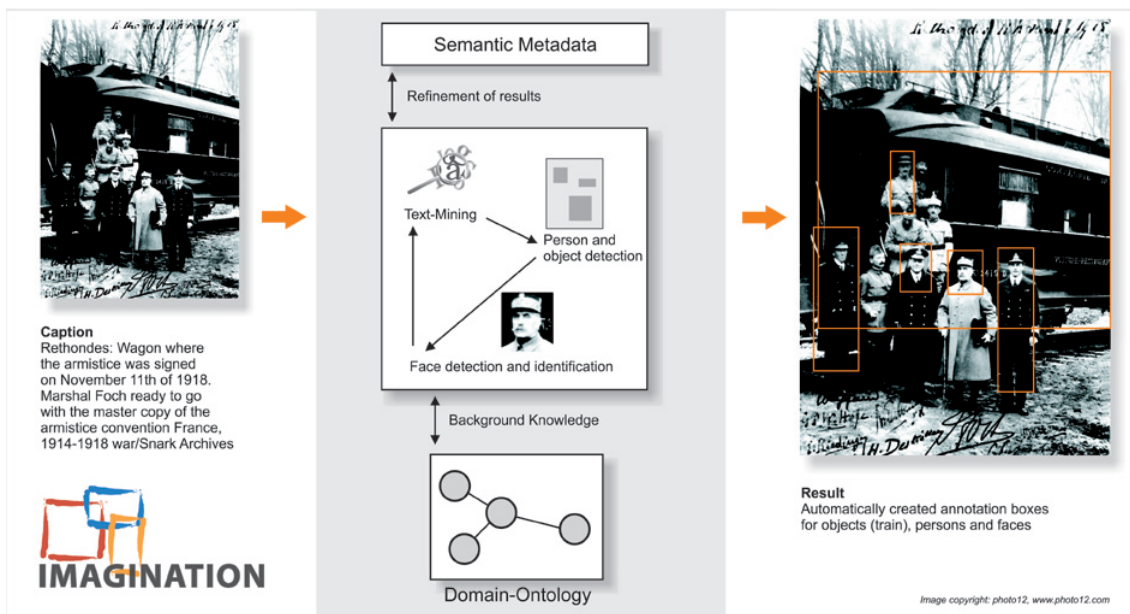


Abbildung 13.1: Grundidee der Kombination automatischer Verfahren

Bildteilen sollen verschiedenartige Objekterkenner eingesetzt werden. Hierbei sollen die vom Textinterpretierer gefundenen Informationen im Text geeignet mit berücksichtigt werden, um die Namen von Personen und Arten abgebildeter Objekte auf den Bildteilen beschriften zu können. Die Ergebnisse sollen geeignet kombiniert werden, um qualitativ hochwertige, automatisch erstellte semantische Metadaten für die Bilder generieren zu können.

Die entsprechende Umsetzung dieser Idee wird in den folgenden Abschnitten gezeigt.

13.2 Ziele und Anforderungen

Basierend auf den Ergebnissen der Evaluation einzelner automatischer Verfahren im letzten Kapitel lassen sich die folgenden Ziele bei der Kombination automatischer Verfahren und die daraus resultierenden Anforderungen formulieren.

13.2.1 Ziele der Kombination automatischer Verfahren

Die Ziele der Kombination automatischer Verfahren sind die Schaffung eines *umfangreicheren* und *hochwertigeren* Gesamtergebnisses.

Umfangreicheres Gesamtergebnis. Der Einsatz mehrerer verschiedener Verfahren führt zu einem umfangreicheren Gesamtergebnis. Die Benutzer müssen dann weniger Bildteile manuell markieren. Ein Beispiel ist, wenn zusätzlich zu einem Personenerkennung ein Objekterkennung eingesetzt wird. Das Ergebnis ist dann umfangreicher, da Personen und Objekte auf den relevanten Bildteilen markiert werden.

Hochwertigeres Gesamtergebnis. Durch geeignete Kombination soll es möglich werden, dass die Anzahl richtig erkannter Bildteile im Gesamtergebnis steigt und die Menge an falsch erkannten Bildteilen im Gesamtergebnis reduziert wird. So entsteht ein hochwertigeres Gesamtergebnis, denn die nötigen Arbeitsschritte zur manuellen Erstellung fehlender Bildteile und der Korrektur von Fehlern werden reduziert.

13.2.2 Anforderungen

Um die genannten Ziele erreichen zu können, lassen sich die folgenden Anforderungen formulieren [WN08]:

- **Verwendung semantischer Metadaten.** Die einzelnen Ergebnisse der automatischen Verfahren müssen als semantische Metadaten erstellt werden. Nur dann lassen sich diese anschließend geeignet kombinieren. Auf der Basis textbasierter Beschriftungen kann zum Beispiel nur schwierig ermittelt werden, dass sich „Person“ und „Elizabeth Taylor“ jeweils auf Bildteile beziehen, die eine Person zeigen.
- **Flexible Integrierbarkeit neuer Objekterkennung.** In der Arbeit von [SGCD09] wurde ein Personenerkennung und Gesichtserkennung statisch miteinander verbunden. Die Implementierung der Steuerung der Kombination der Ergebnisse hätte also erweitert werden müssen, wenn zusätzlich ein weiterer Gesichtserkennung verwendet worden wäre. Es soll daher ein Verfahren entwickelt werden, das neue Objekterkennung flexibel integrieren kann
- **Flexible Kombination der Ergebnisse.** Die Ergebnisse der Objekterkennung sollen flexibel miteinander auf Basis von semantischen Informationen kombiniert werden. *Dies erfordert entsprechende Strategien zur Kombination der Ergebnisse.*

- **Kombination der Ergebnisse mit einem Textinterpretierer.** Falls ein Textinterpretierer eingesetzt wird und textbasierte Metadaten von diesem zu semantischen Metadaten übertragen werden konnten, sollen diese Ergebnisse mit den Ergebnissen des Objekterkenners kombiniert werden können. Wurde zum Beispiel eine Person auf einem Bildteil durch den Objekterkennner markiert und vom Textinterpretierer der Name einer Person, kann dieser Name dem Bildteil der erkannten Person zugewiesen werden.

13.3 Verwandte Arbeiten

Bei der Suche nach verwandten Arbeiten wurden die folgenden Arbeiten ermittelt, welche die Kombination verschiedener Verfahren zur Verbesserung der Gesamtergebnisse von Objekterkennern betrachten. Es wird jeweils ermittelt, wie die jeweiligen Verfahren genutzt werden können, um die eigene Kombination automatischer Verfahren mit den gestellten Anforderungen realisieren zu können.

13.3.1 Semantische Bibliotheken

Hamid und Quadir [AHAQI⁺07] (ähnlich auch in [LZLM05]) nutzen *semantische Bibliotheken* zur Klassifizierung von Teilen und Merkmalen, aus denen ein Objekt besteht. So wurden zum Beispiel die Farben und Formen von Tieren wie der Kuh oder dem Tiger hinterlegt. Zur Erkennung von Tieren wurden diese semantischen Bibliotheken eingesetzt und ihr verwendeter Objekterkennner für Tiere hat diese Informationen ausgenutzt und auf die jeweiligen Tiere *zurückgeschlossen*.

Einsatz bei Kombination automatischer Verfahren Objekte werden bei Hamid und Quadir somit durch Beschreibung wichtiger Teile erkannt. Die semantische Bibliothek wurde hierbei als Hierarchie aus Objektteilen und Merkmalen aufgebaut. Eine solche Hierarchisierung soll bei der Kombination automatischer Verfahren dazu dienen, die einzelnen Objekterkennner zu hierarchisieren nach Objekterkennung und Identifizierung.

13.3.2 Kombination von Textinterpretierer und Objekterkennner

Die Kombination von Textinterpretierer und Objekterkennner wird in der Literatur häufig beschrieben, z.B. bei [DJLW08] und [CGSW03]. Der Textinterpretierer wird hierbei meist verwendet, um automatisiert Trainingsdaten zum Training universeller Objekterkennner zu erstellen [LW06]. Weiterhin kann mit der Kenntnis um zu

erwartende Objekte, die aus dem Text ermittelt wurde, die Menge an zu verwendenden Objekten für den Objekterkenner eingeschränkt werden. Dann steigt die Wahrscheinlichkeit, dass ein Objekt richtig erkannt wurde.

Einsatz bei Kombination automatischer Verfahren Wie im letzten Kapitel schon erwähnt, sollen ebenfalls Textinterpretierer eingesetzt werden. Diese sollen wie beschrieben direkt semantische Informationen liefern. So ist direkt Hintergrundwissen integriert wie zum Beispiel, dass es sich bei dem *Elizabeth Taylor* um eine *Person* handelt, die *weiblich* ist. Da auch die Objekterkenner ihre Beschriftungen mittels semantischer Elemente erstellen sollen, z.B. Personenerkenner nutzen das Element *Person*, lassen sich die Ergebnisse von Textinterpretierer und Objekterkenner kombinieren. Auf einem Portraitbild von *Elizabeth Taylor* könnten die beiden genannten semantischen Beschriftungen kombiniert werden so dass der gefundene Bildteil statt mit *Person* aufgrund der vorhandenen Relation *Elizabeth Taylor ist eine Person* direkt mit dem semantischen Element *Elizabeth Taylor* beschriftet wird.

13.3.3 Kombination von Gesicht- und Personenerkennung

Schwarz [SGCD09] hat in seiner Arbeit die Kombination von Gesicht- und Personenerkennung vorgeschlagen, um dadurch mehr Personen erkennen zu können, von denen wichtige Teile auf dem Bild verdeckt sind, aber das Gesicht noch sichtbar ist. Dadurch konnte er 8,1 Prozent mehr Personen in seinem Trainingsbildbestand erkennen.

Einsatz bei Kombination automatischer Verfahren Eine solche Kombination ist geplant. Allerdings soll auch die umgekehrte Richtung möglich sein, also ob in erkannten Bereichen, welche Personen zeigen, Gesichter erkannt werden oder nicht, um die Menge an falschen Beschriftungen zu verringern.

In der genannten Arbeit von Schwarz [SGCD09] wurden die Ergebnisse zusammengeführt durch einen entsprechenden Kombiniierer der Ergebnisse. Dieser kann aber nur Ergebnisse von Bildteilen, die Personen und Gesichter zeigen, durchführen. Es soll daher ein Verfahren entwickelt werden, das die Ergebnisse beliebiger Objekterkenner flexibel miteinander verbinden kann.

13.4 Hierarchisierung von Objekterkennern

Die folgende *Hierarchisierung von Objekterkennern* soll ermöglichen, dass neue Objekterkenner in ein System zur Kombination automatischer Verfahren *flexibel* integriert werden können und ermittelt werden kann, wie die gelieferten Ergebnisse des Objekterkenners mit weiteren *Ergebnissen anderer Objekterkenner kombiniert* werden

sollen. Die Hierarchisierung ist hierbei wie in der letzten Sektion erwähnt angelehnt an die Hierarchisierung von Objekten in [AHAQI⁺07].

Im vorigen Kapitel wurden die Objekterkenner unterteilt nach universellen und spezialisierten Objekterkennern. Aufgrund der besseren Trainingsdaten erkennen spezialisierte Objekterkenner ein Objekt *präziser*. So kann ein universeller Objekterkenner zwar häufig Formen und Linien zu einem Objekt zusammenfassen, kann aber das Objekt nicht zuordnen. Der kombinierte Einsatz eines universellen Objekterkenners und eines Personenerkenners würde somit dazu führen, dass zwei unabhängige Objekterkenner einen Bildbereich erkennen und sich ihre Ergebnisse somit gegenseitig bestätigen können. Gleichzeitig kann als präziseres Ergebnis angegeben werden, dass der Bildteil eine Person zeigt. Gleiches gilt für den parallelen Einsatz zweier spezialisierter Objekterkenner. Wird ein Fahrzeugerkenner und ein Autoerkenner eingesetzt und finden beide ein Fahrzeug auf einem Bildteil, kann mit einer hohen Wahrscheinlichkeit angenommen werden, dass sich auf dem Bildteil tatsächlich ein Auto befindet.

Für die Kombination von Objekterkennern ist somit wichtig, dass die Zusammenhänge der Objekterkenner bekannt sind, also „Welcher Objekterkenner präzisiert die Ergebnisse eines anderen Objekterkenners?“. *Hierzu bietet sich die Bildung einer Taxonomie an.*

Weiterhin wurden im vorigen Kapitel Verfahren zur Identifikation von Objekten vorgestellt. Diese beschreiben den Inhalt eines gefundenen Bildteils näher. So kann ein Geschlechterkenner das Merkmal Geschlecht für ein gefundenes Gesicht ermitteln. Ebenso kann ein Gesichtserkenner eine Person identifizieren. Zur Beschreibung, wie einzelne Objekterkenner weitere Details und Identifikationen finden können, bietet sich die Bildung eines Instanzenmodells mit den entsprechenden Relationen *findet Merkmal* und *identifiziert an*.

13.4.1 Aufbau eines geeigneten Instanzenmodells

Eine Hierarchisierung und Zuordnung von Objekterkennern mit einem Instanzenmodell wird daher wie folgt aufgebaut:

- **Taxonomie der Objekterkenner** Objekterkenner, die Ergebnisse anderer Objekterkenner spezieller erkennen können, werden als Hierarchie aufgebaut. Die hierarchische Relation hat den Namen *präzisiert*.
- **Details eines Objekts** Objekterkenner, die Merkmale eines Objekts liefern, werden über die benannte Relation *findet Merkmal* mit dem jeweiligen Objekterkenner verbunden.

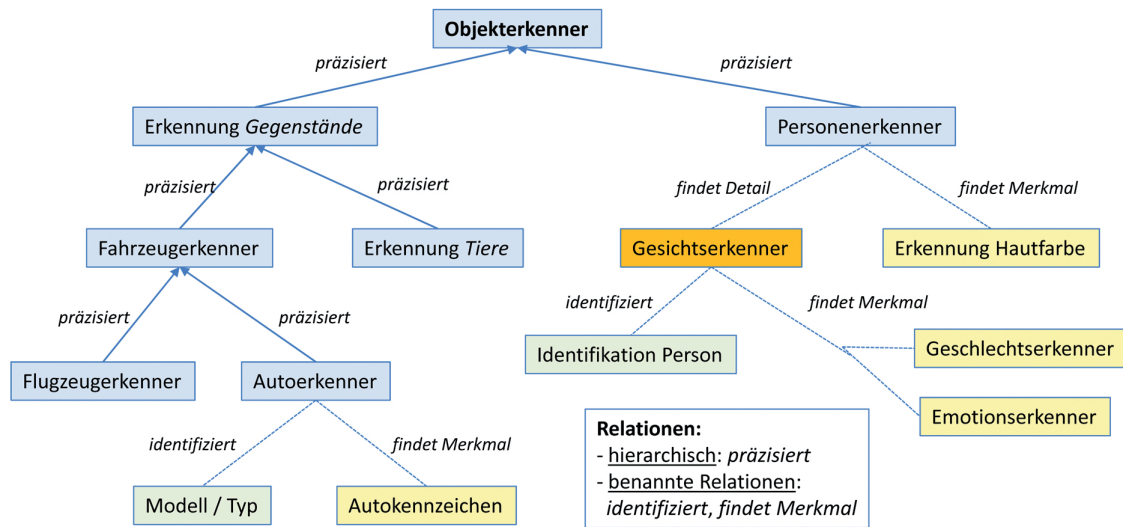


Abbildung 13.2: Ein Instanzenmodell für Objekterkennung

- **Identifikation eines Objekts** Objektidentifizierer, die ein Objekt genauer identifizieren können, werden über die benannte Relation *identifiziert* mit dem jeweiligen Objekterkennung verbunden.

Umsetzung der Anforderung nach flexibler Integrierbarkeit von Objekterkennern Wird ein neuer Objekterkennung in einem System zur Kombination automatischer Verfahren hinzugefügt, kann dieser so wie gewünscht flexibel integriert werden, indem das Instanzenmodell entsprechend erweitert wird und die benötigte Relation verwendet wird.

Umsetzung der Anforderung nach flexibler Kombinierbarkeit der Ergebnisse Aufgrund der gegebenen Relationen kann eine Komponente zur Kombination der Gesamtergebnisse anschließend ermitteln, wie die Ergebnisse eines neuen Objekterkenners mit den Ergebnissen anderer Objekterkennung kombiniert werden können.

Im Folgenden wird zuerst ein Beispiel genannt, wie ein solches Instanzenmodell gebildet werden kann. In den weiteren Sektionen wird dann gezeigt, wie die Ergebnisse basierend auf dem Instanzenmodell geeignet kombiniert werden können.

13.4.2 Beispiel eines entstehenden Instanzenmodells

Auf Basis der im vorigen Kapitel genannten spezialisierten Objekterkennung soll ein Instanzenmodell zur Hierarchisierung von Objekterkennern der Gegenstände *Flugzeuge* und *Auto* sowie für die Erkennung von Personen erfolgen. Als Objekterkennung sollen entsprechend Flugzeugerkennung, Autoerkennung, Personenerkennung, Gesichtsdetektion, Geschlechtererkennung, Emotionserkennung und Gesichtsidentifikation eingesetzt werden.

Abbildung 13.2 zeigt das daraus entstehende Instanzenmodell. Auf der ersten Hierarchie ist die *Erkennung von Gegenständen* und *Erkennung von Personen*. Der Personenerkennung wird somit eingeordnet als „Personenerkennung“ präzisiert Ergebnisse von „Objekterkennung“. Eine mögliche Unterteilung von „Erkennung Gegenstände“ ist 'Erkennung Fahrzeuge' und „Erkennung Tiere“. Innerhalb der „Erkennung Fahrzeuge“ können dann der Autoerkennung und der Flugzeugerkennung in die Hierarchie eingeordnet werden. Sind gleichzeitig ein Fahrzeugerkennung und ein Autoerkennung vorhanden, können die Ergebnisse kombiniert werden. Sind ein Autoerkennung und ein Flugzeugerkennung vorhanden, können ihre Ergebnisse gegenseitig ausgeschlossen werden. Denn finden beide ein Objekt und markieren es je als „Flugzeug“ beziehungsweise „Auto“, so kann abgeleitet werden, dass ein Fehler vorliegt, da ein Bildteil nicht gleichzeitig ein Flugzeug und ein Objekt zeigen kann³. Wird weiterhin ein Erkennung für Autokennzeichen verwendet, kann dieser eingegliedert werden als „Autokennzeichen“ *findet Merkmal* für ein „Auto“

Innerhalb der Hierarchie in der Abbildung 13.2 lassen sich der Gesichtserkennung im Instanzenmodell einordnen als „Gesichtserkennung“ *findet Detail* von „Personenerkennung“. Denn das Gesicht ist ein Detail einer Person. Entsprechend lassen sich der „Emotionserkennung“ und der „Emotionserkennung“ einordnen als „Emotionserkennung“ *findet Merkmal* für „Gesicht“ und „Geschlechtererkennung“ *findet Merkmal* für „Geschlecht“. Schließlich kann die Gesichtsidentifikation eingeordnet werden mit *identifiziert* Person.

13.5 Kombination der Ergebnisse

13.5.1 Genereller Ablauf der Kombination von Ergebnissen

Die Kombination von Ergebnissen der automatischen Prozesse erfordert eine *Steuerkomponente*. Diese kennt die vorhandenen Objekterkennung im System. Nacheinander

³Möglich wäre auch, eine weitere Relation in einem umfangreichen Instanzenmodell „schließt aus“ einzuführen, um anzugeben, welche Ergebnisse einzelner Objekterkennung sich definitiv ausschließen.

ruft die Steuerkomponente diese Objekterkennung auf und sammelt deren Ergebnisse ein. Anschließend können diese Ergebnisse mit Hilfe der gegebenen Relationen des Instanzenmodells kombiniert werden.

Die einfachste Form der Kombination ist, einfach alle Ergebnisse zusammenzufügen. So entsteht wie gewünscht ein *umfangreiches Ergebnis*. Allerdings enthält es, wie durch die Evaluation im letzten Kapitel gezeigt, auch alle Fehler, die durch Objekterkennung erstellt wurden.

Weiterhin ist das so gebildete Ergebnis noch nicht gefiltert. Findet zum Beispiel ein Personenerkennung eine Person und ein Gesichtserkennung ein Gesicht, überschneiden sich die erstellten Markierungen auf den Bildteilen. Eine sinnvolle Filterung dieser Bereiche und somit Zusammenfassung ist daher wünschenswert.

Falls mehrere gleichartige Objekterkennung eingesetzt werden, schreiben diese ihre Ergebnisse in die semantischen Metadaten. Auch diese Ergebnisse müssen gefiltert werden: finden mehrere Objekterkennung den gleichen Bildteil, werden Duplikate von Beschriftungen erstellt, die zusammengefasst werden müssen zu einer verbleibenden Beschriftung für den zu beschreibenden Bildteil.

Entsprechend sollten die Ergebnisse von Textinterpretierern mit den Ergebnissen für beschriftete Bildteile zusammengefasst werden, also zum Beispiel die direkte Zuweisung des Namens einer Person, der im Text gefunden wurde, für den entsprechenden Bildteil.

Die gewünschte Schaffung von hochwertigen Ergebnissen durch Objekterkennung soll im Folgenden durch die geeignete *Kombination gleichartiger Ergebnisse*, der *Kombination hierarchischer Ergebnisse*, sowie durch die *Kombination mit einem Textinterpretierer* ermöglicht werden.

13.5.2 Kombination gleichartiger Ergebnisse

Die Kombination gleichartiger Ergebnisse soll die Ergebnisse von Objekterkennung filtern, *die gleiche Objekte erkennen können und somit im Instanzenmodell an der gleichen Stelle eingeordnet worden sind*.

Am Beispiel von mehreren Gesichtserkennung soll gezeigt werden, welche möglichen Kombinationen entstehen. Abbildung 13.3 gibt einen Überblick zu den möglichen zwei Kombinationen gleichartiger Ergebnisse und deren Filterung:

- **Kombination und Filterung aller Ergebnisse (combineAll)** Hierbei werden alle Ergebnisse zusammengefasst, die sich auf verschiedene Bildteile beziehen. Weiterhin werden Ergebnisse, die sich auf den gleichen Bildteil beziehen, gefiltert zu einer einzigen Beschriftung. Falls die Kombination gleich-



Abbildung 13.3: Kombination gleichartiger Ergebnisse

artiger Ergebnisse deutlich mehr richtige Ergebnisse generiert und sich der Anteil an falschen Stellen nicht zu sehr aufsummiert, sollte diese Kombination hochwertigere Ergebnisse generieren können.

- **Kombination gleicher Ergebnisse (combineSame)** Diese Kombination übernimmt nur diejenigen Beschriftungen, die sich jeweils auf den gleichen Bildteil beziehen. Die Ergebnisse zu nicht übereinstimmenden Bildteilen werden verworfen und somit nicht übernommen. Dieses Verfahren reduziert das Ergebnis auf nur diejenigen Bildteile, die übereinstimmend beschriftet wurden von beiden Gesichtserkennern. Angenommen wird, dass dadurch das Gesamtergebnis richtiger wird, da sich die Menge an falsch erkannten Bildteilen deutlich reduzieren sollte und so der Korrekturaufwand sinkt. Allerdings wird die Menge an richtig erkannten Stellen ebenfalls dann reduziert, wenn nur ein Gesichtserkennung diesen Teil zwar als richtig erkannt hat, der andere Gesichtserkennung aber nicht.

Beide Verfahren sollen evaluiert werden hinsichtlich ihrem Einfluss auf die Reduzierung der nach einer Kombination noch nötigen Arbeitsschritte.

Werden mehr als zwei Objekterkennung eingesetzt, bieten sich *Voting Verfahren* an. Hierbei sammelt die Steuerung alle Vorschläge („Votes“) und fasst diese nach einer Strategie zusammen. Die Strategie könnte sein *Mehrheit der Votes* [Lon02], also es werden die überwiegend richtig genannten Vorschläge übernommen, oder *Gewichtung der Votes* [Lon02], falls bekannt ist, dass einige der Objekterkennung bessere Vorschläge machen als die anderen verwendeten Objekterkennung.

13.5.3 Kombination hierarchischer Ergebnisse

Die Kombination hierarchischer Ergebnisse soll die Ergebnisse von Objekterkennung filtern, *die hierarchisch zusammenhängende Objekte erkennen können und somit im Instanzenmodell an zwei verschiedenen Stellen entlang einer Hierarchie eingeordnet sind*.

Die Steuerung erkennt hierarchisch zusammenhängende Objekte durch Anwendung des Instanzenmodells zur Hierarchisierung von Objekterkennung. Liegen die Objekte im selben Ast der Hierarchie, sind sie hierarchisch zusammenhängend.

Am Beispiel eines Personenerkennung und eines Gesichtserkennung soll gezeigt werden, welche möglichen Kombinationen entstehen. Abbildung 13.4 gibt einen Überblick zu den möglichen zwei Kombinationen hierarchisch zusammenhängender Ergebnisse und deren Filterung:

- **Kombination und Filterung aller Ergebnisse (combineHierarchica-**

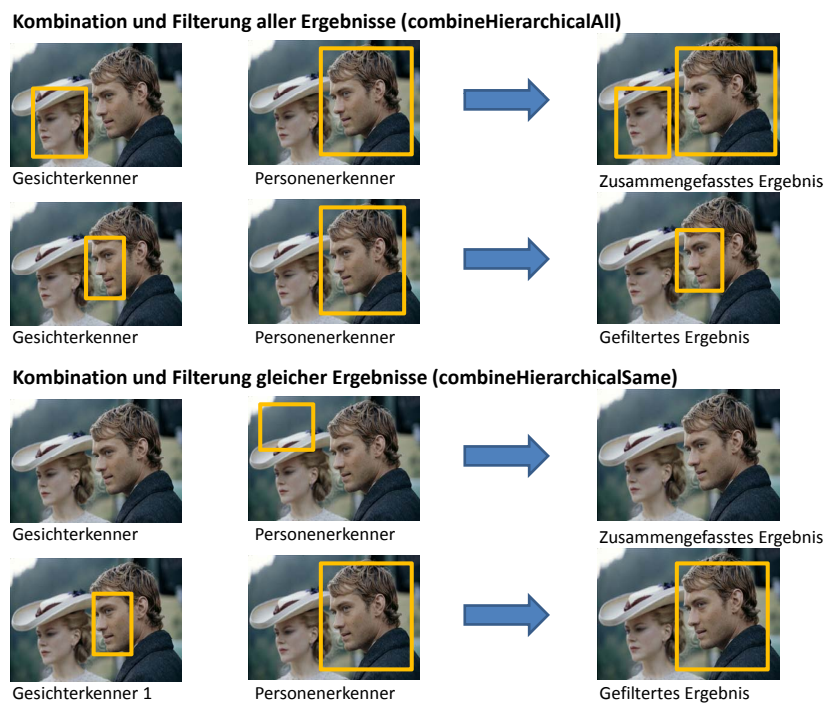


Abbildung 13.4: Kombination hierarchischer Ergebnisse

IAII) Hierbei werden all Ergebnisse zusammengefasst, die sich auf verschiedene Bildteile beziehen. Weiterhin werden Ergebnisse, die sich auf den gleichen Bildteil beziehen, gefiltert zu einer einzigen Beschriftung⁴ Falls die Kombination gleichartiger Ergebnisse deutlich mehr richtige Ergebnisse generiert und sich der Anteil an falschen Stellen nicht zu sehr aufsummiert, sollte diese Kombination hochwertigere Ergebnisse generieren können.

- **Kombination gleicher Ergebnisse (combineHierarchicalSame)** Diese Kombination übernimmt nur diejenigen Beschriftungen, die sich jeweils auf den gleichen Bildteil beziehen. Die Ergebnisse zu nicht übereinstimmenden Bildteilen werden gefiltert. Dieses Verfahren reduziert das Ergebnis nur auf die Bildteile, die übereinstimmend beschriftet wurden von Gesichterkenner und Personenerkenner.

Erkennung von Widersprüchen Durch die Hierarchie im Instanzenmodell können weiterhin Widersprüche erkannt werden. Hierzu prüft die Steuerung, welche der von den Objekterkennern beschrifteten Bildteile sich auf den jeweils gleichen Bildteil beziehen. Stammen diese Ergebnisse von verschiedenen Ästen innerhalb der Hierarchie, so liegt bei den gegebenen Ergebnissen ein Widerspruch vor. Zum Beispiel könnte ein Bildteil gleichzeitig von einem Autoerkenner als „Auto“ und von einem Personenerkenner als „Person“ erkannt worden sein. Die Steuerung kann diesen Bildteil dann entweder als Konflikt markieren und dem Benutzer die Korrektur überlassen oder die betroffenen Bildteile löschen.

13.5.4 Kombination mit einem Textinterpretierer

Falls ein Textinterpretierer eingesetzt wird, können weitere Kombinationen erfolgen. Diese nutzen das Hintergrundwissen der semantischen Metadaten aus. So können gleichartige Beschriftungen ermittelt werden, zum Beispiel dass sich das semantische Element „Nicole Kidman“ auf eine Person bezieht.

Am Beispiel von Textinterpretierer, Gesichterkenner und Gesichtidentifikation sollen mögliche Kombinationen nun gezeigt werden (Abbildung 13.5 gibt einen Überblick):

- **Kombination semantisch gleichartiger Beschriftungen** Die Steuerung ermittelt hierbei zu den gefundenen semantischen Elemente des Textinterpre-

⁴Anmerkung: Während des IMAGINATION Projekts ergaben Befragungen von Bildsuchern in Workshops, dass sie präzisere Beschriftungen für Bildteile bevorzugen. Wurde also ein Bildteil als Person und Gesicht beschriftet, gaben sie an, dass als gefilterter Bildteil nur die präzisere Markierung für das Gesicht bleiben soll. Aufgrund dieser Angaben wird in dieser Arbeit die Filterung des jeweils präziseren Bildteils angewendet.

Kombination von Texterkenner mit Gesichtserkennung und Gesichtidentifikation

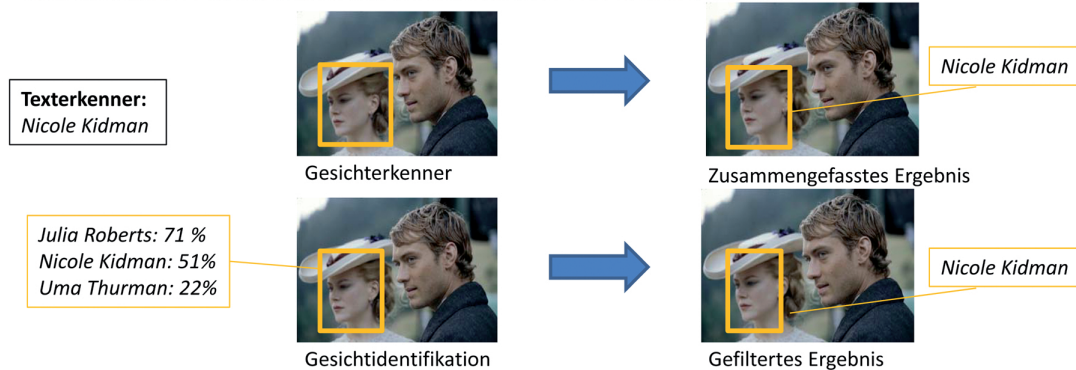


Abbildung 13.5: Kombination von Textinterpretierer mit Gesichtserkennung und Gesichtidentifikation

tierers aus den textbasierten Metadaten, ob gleichartige Beschriftungen durch Objekterkennung erstellt wurden. Wurde zum Beispiel vom Textinterpretierer „Nicole Kidman“ ermittelt und von einem Gesichtserkennung exakt ein Bildteil auf dem Bild, so sind die Beschriftungen gleichartig, da sich beide auf Personen beziehen. Entsprechend kann die Steuerung beide Beschriftungen kombinieren und den Bildteil mit „Nicole Kidman“ beschriften. Falls zusätzlich ein Geschlechterkennung eingesetzt wird, funktioniert dieses Verfahren auch, wenn zwei Personen auf dem Bild abgebildet sind, von denen jeweils eine Person männlich und die andere als weiblich erkannt wurde. Die entsprechenden semantischen Elemente vom Textinterpretierer enthalten nämlich auch die Information zum Geschlecht einer im Text gefundenen Person. Entsprechend können diese Informationen zur Verbindung von Text und Bildteile verwendet werden.

- **Kombination mit Objektidentifikation** Die Steuerung ermittelt, ob ein Verfahren zur Objektidentifikation mehrere Vorschläge zurückgegeben hat zur Beschriftung des Bildes. Bei der Gesichtidentifikation werden häufig mit der Angabe der Wahrscheinlichkeit der Identifikation als Prozentwert mehrere als Identifikation mögliche Namen angegeben. Falls einer der Namen mit dem semantischen Element des vom Textinterpretierer erkannten Namens übereinstimmt, kann diese Beschriftung dem Bildteil zugewiesen werden.

Textinterpretierer helfen somit, die Beschriftungen für die Bildteile zu präzisieren. Entsprechend den Vorschlägen in [DJLW08] und [CGSW03] zum automatisierten Aufbau von Trainingsdaten mit Hilfe von Textinterpretierern kann auch

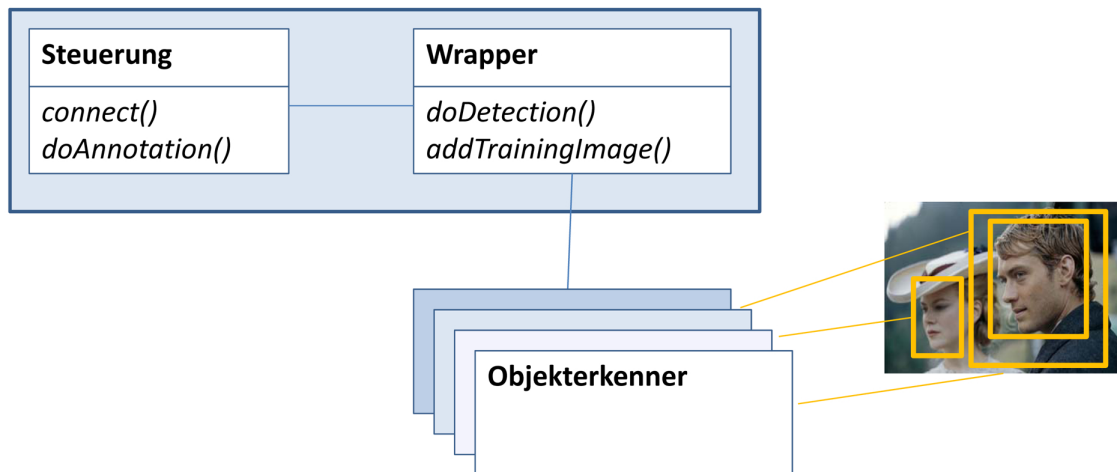


Abbildung 13.6: Aufbau der Komponente „Kombination automatischer Prozess“ in ImageNotion

hier dieses Vorgehen mit eingesetzt werden. Wurden zum Beispiel Kombinationen von Textinterpretierer und Gesichtidentifikation durchgeführt, können die Bildteile als neue Trainingsdaten für das Gesicht verwendet werden und so automatisch hinzugefügt werden.

13.6 Aufbau der Komponente „Kombination automatischer Prozesse“

Abbildung 13.6 gibt einen Überblick über die in ImageNotion erstellte Komponente *Kombination automatischer Prozesse*.

13.6.1 Wrapper

Wrapper dienen zur einheitlichen Einbindung der verschiedenen Objekterkennung und deren Implementierungen in ImageNotion. Die Schnittstelle besteht hierbei aus den Methoden `doDetection` und `addTrainingImage`. Mit `doDetection` wird der Objekterkennung von der Steuerung aufgerufen. Das Bild selbst wird zum Hinzufügen von Beschriftungen mit übergeben. Der Wrapper implementiert diese Methode entsprechend, um seinen aufzurufenden Objekterkennung ansprechen zu können. Das Bild wird an den Objekterkennung übergeben und der Wrapper wartet anschließend auf

die Ergebnisse des Objekterkenners. Die resultierenden Ergebnisse werden in das Objektformat in ImageNotion übertragen und textbasierte Beschriftungen eines Objekterkenners werden übersetzt in semantische Beschriftungen.

Die Methode *addTrainingImage* ermöglicht der *Steuerung*, geeignete Objekterkennern aufzurufen, damit diese Trainingsbilder hinzufügen können für zu erkennende Objekte. So kann zum Beispiel nach Validierung von automatisch erkannten Bildteilen die Steuerung die vorhandenen Gesichtidentifizierer aufrufen, damit diese den für die erkannte und validierte Person gefundenen Bildteil als neues Trainingsbild hinzufügen.

13.6.2 Aufbau der Steuerung

Die Steuerung dient zur Koordination der Objekterkennern und zur Kombination der durch diese erstellten Ergebnisse. Zur Kombination der Ergebnisse wurden in ImageNotion die oben beschriebenen Varianten zur Kombination gleichartiger und hierarchisch verwandter Ergebnisse implementiert.

Mit der Methode *connect* melden sich einzelne Objekterkennern bei der Steuerung an. Hierbei gibt der jeweilige Wrapper mit an, um was für einen Objekterkennern es sich basierend auf der gegebenen Hierarchisierung von Objekterkennern handelt. So können die Objekterkennern wie gewünscht flexibel in ImageNotion integriert werden.

Mit der Methode *doAnnotation* erhält die Steuerung zum Beispiel nach dem Aufladen eines neuen Bildes den Befehl zur Durchführung automatischer Bildbeschriftungen. Die Steuerung ruft anschließend mit *doDetection* alle die Wrapper der angebotenen Objekterkennern auf und wartet auf die Ergebnisse. Sind alle Ergebnisse da, werden zuerst die Ergebnisse der gleichartigen Objekterkennern und anschließend die Ergebnisse hierarchisch verwandter Ergebnisse kombiniert.

13.7 Laboruntersuchung 2: Messung der kombinierten Ergebnisse von Objekterkennern

13.7.1 Durchführung der Evaluation

Es werden dieselben Bilder des Filmarchivs wie für die durchgeführte Evaluation in vorherigen Kapitel verwendet. Die genannten Möglichkeiten zur Kombination der gleichartigen Ergebnisse sowie die Kombination hierarchischer Ergebnisse wurden implementiert. Die folgenden Installationen ergaben sich für die Evaluation:

13.7. LABORUNTERSUCHUNG 2: MESSUNG DER KOMBINIERTEN ERGEBNISSE VON OBJE

- **combineAll** Implementierung der Variante *Kombination und Filterung aller Ergebnisse* gleichartiger Objekterkennung. Zur Evaluation wurden die Ergebnisse der Gesichtserkennung von IIS und face.com verwendet.
- **combineSame** Implementierung der Variante *Kombination und Filterung gleicher Ergebnisse* gleichartiger Objekterkennung. Zur Evaluation wurden die Ergebnisse der Gesichtserkennung von IIS und face.com verwendet.
- **combineHierarchicalSame** Implementierung der Variante *Kombination und Filterung gleicher Ergebnisse* hierarchisch verbundener Objekterkennung. Zur Evaluation wurden die Ergebnisse des Gesichtserkenners von IIS und des Personenerkenners von NTUA verwendet.

Von der Installation combineHierarchicalAll wurde abgesehen, da hiermit die große Menge an falschen Beschriftungen des Personenerkenners von NTUA nicht reduziert werden kann. Daher war in der Evaluation von Interesse, wie stark die falschen Ergebnisse reduziert werden können, wenn nur die Ergebnisse übernommen werden, in denen der Gesichtserkennung ein Gesicht findet.

13.7.2 Ergebnisse der Evaluation

Tabelle 13.1 gibt einen Überblick über die Ergebnisse der Evaluation.

Kombination und Filterung aller Ergebnisse erhöht die Menge richtig gefundener Gesichter Die Installation combineAll hat zu 810 richtig erkannten Gesichtern geführt. Dies sind 64 mehr richtig erkannte Gesichter im Vergleich zum Gesichtserkennung von face.com alleine. Allerdings ist auch die Menge an falsch erkannten Gesichtern von 504 auf 564 gestiegen. Im Vergleich zum Gesichtserkennung face.com konnte so die Anzahl eingesparter Arbeitsschritte immerhin von 40,7 Prozent auf 42,1 Prozent gesteigert werden. Allerdings ist dieses Ergebnis im Vergleich zum Gesichtserkennung IIS (57,5 Prozent) deutlich schlechter.

CombineAll als Verfahren kann somit das Gesamtergebnis verbessern, allerdings nur dann, wenn beide Gesichtserkennung deutlich mehr Gesichter richtig erkennen als Fehler produziert werden. Eine sehr geeignete Kombination wäre zum Beispiel ein Gesichtserkennung, der spezialisiert ist auf die Erkennung von Gesichtern als Seitenprofil mit einem Erkennung, der Gesichter in der Frontalansicht erkennt.

Da der Gesichtserkennung von face.com deutlich mehr Fehler produziert hat als der Gesichtserkennung von IIS, wurde dadurch das Gesamtergebnis schlechter als das alleinige Ergebnis von IIS. Eine generelle Forderung an die Hersteller von Objekterkennung wäre also, die Menge an Fantomgesichtern zu reduzieren. Dann könnten

Objekt- erkenner	Richtig erkannte Bildteile	Nicht erkannte Bildteile	Phantome	Falsch erkannt	Falsch zugeordnet	Arbeitsschritte manuelle Erstellung	Nötige Korrekturschritte	Summe Arbeitsschritte	Ersparnis Prozent
CombineAll	810	208	564	0	0	624	1128	1752	42,1
CombineSame	665	343	86	0	0	1029	172	1201	60,2
Combine- Hierarchical- Same	204	804	113	0	0	2412	226	2638	12,8
Combine- Hierarchical- Same (Lauf 2)	523	485	113	0	0	1455	226	1681	44,4
Combine- TextFace- identification	576	432	0	0	154	1296	462	1758	41,7

Tabelle 13.1: Übersicht über die Ergebnisse der Kombinationsverfahren (Basis: Tabelle 12.2)

13.7. LABORUNTERSUCHUNG 2: MESSUNG DER KOMBINIERTEN ERGEBNISSE VON OBJE

die gezeigten Maßnahmen zur Kombination gleichartiger Ergebnisse zu einer umfangreichen Verbesserung des Gesamtergebnisses führen.

Kombination und Filterung aller Ergebnisse verringert die Menge falscher Resultate und verbessert das Gesamtergebnis combineSame hat zu 665 Bildteilen geführt, die beide Gesichtserkennung über einstimmend erkannt haben. So hat sich zwar das Ergebnis richtiger Stellen zwar um etwa 100 Bildteile verringert im Vergleich zu Gesichtserkennung face.com, allerdings konnten die falsch erkannten Stellen auf nur noch 86 Fehler reduziert werden.

Dieses Verfahren würde sich also dann anbieten, wenn die eingesetzten, gleichartigen Objekterkennung eine hohe Menge an richtig erkannten Stellen finden und gleichzeitig aber auch viele Fantomgesichter produzieren. Dann würde die Filterung wie gewünscht diese Fehler reduzieren. CombineSame hat somit in der durchgeführten Evaluation immerhin dazu geführt, dass die Einsparung der benötigten Arbeitsschritte von 57,5 Prozent bei Fraunhofer IIS alleine auf 60,2 Prozent reduziert werden konnte. Da Gesichtserkennung eingesetzt wurden, die zum Stand der durchgeführten Evaluation zu den am Markt besten Erkennern gehörten, ist aktuell keine weitere Verbesserung hinsichtlich der Erkennung von Gesichtern mit diesem Verfahren möglich.

Beim Einsatz von ImageNotion mit Benutzern stellten wir innerhalb des IMAGINATION Projektes fest, dass diese sich an falsch erkannten Bildteilen oft stören. So kann man folgern, dass subjektiv ein Ergebnis auch dann als besser wahrgenommen, wenn zwar im Gesamtergebnis etwas weniger Bildteile richtig beschrieben worden sind, *dafür aber deutlich weniger Bildteile falsch beschriftet wurden*.

Zur Erhöhung der Präzision des Gesamtergebnisses von gleichartigen Objekterkennung empfiehlt sich daher das Verfahren combineSame, da es gleichzeitig die falschen Bildteile entfernt und weiterhin den Arbeitsaufwand zur Markierung von Bildteilen senken kann.

Kombination hierarchisch verwandter Ergebnisse senkt die Anzahl an Fehlern Wie das Ergebnis zu CombineHierarchicalSame zeigt, konnte die Menge falscher Stellen deutlich von 438 auf 113 verbleibende Fehler reduziert. Die Menge an richtig erkannten Bildteilen wurde allerdings ebenfalls deutlich reduziert. Im Vergleich mit dem Personenerkennung alleine blieben von ursprünglich 304 Bildteilen nur noch 204 übrig. Ausgehend vom Gesichtserkennung blieben gar von 699 richtigen Bildteilen nur 204 Teile übrig.

Die Kombination hierarchisch verwandter Ergebnisse erhöht somit die Präzision für die gefundenen Bildteile und reduziert deutlich die Menge an Fehlern. Allerdings gehen für den hierarchisch unter dem anderen Objekterkennung liegenden Gesichtserkennung zu viele richtige Ergebnisse verloren.

Eine mögliche Lösung ist, vom Gesichterkenner markierte Bildteile mit einer hohen Wahrscheinlichkeit immer zu behalten. Diese Lösung (Combine-Hierarchical-Same (Lauf 2) in Tabelle Tabelle 13.1 führte zu 523 richtig erkannten Bildteilen. Es wurden hierbei vom Gesichterkenner nur die Vorschläge übernommen, zu denen mindestens 50 Prozent Wahrscheinlichkeit für Korrektheit angegeben wurden. So konnten im Bildarchiv Personen und Gesichter automatisch beschriftet werden und gleichzeitig die Fehlermenge gerade für den Personenerkenner deutlich herab gesetzt werden.

Sollen in einem Bildarchiv also ganze Personen und Gesichter markiert werden, lohnt sich der Einsatz von combineHierarchicalSame, um für präzise Ergebnisse und deutlich weniger Fehler sorgen zu können.

Textinterpretierer verbessern die Gesichtsidentifikation Im Filmarchiv befinden sich oft Szenen mit einem Schauspieler alleine oder zwei Schauspielern, wobei dann oft eine der Personen weiblich und die andere Person männlich ist. Entsprechend konnte die Steuerung oft die Ergebnisse von Textinterpretierer und Gesichtsidentifikation geeignet kombinieren. Entsprechend gelang es auch ohne Trainingsdaten, dass Personen richtig erkannt wurden. So konnten insgesamt 576 Personen richtig erkannt werden, was mehr als der Hälfte an Bildteilen, die Personen zeigen, im Bildarchiv entspricht.

Falls Textinterpretierer vorhanden sind, sollte das genannte Verfahren zur automatischen Beschriftung der Namen von Gesichtern auf Bildteilen eingesetzt werden.

13.7.3 Fazit

Die Verfahren combineSame und combineHierarchicalSame haben gezeigt, dass die Kombination automatischer Verfahren zu präziseren Ergebnissen führen kann, wodurch weniger falsch markierte Bildteile im Gesamtergebnis auftreten. Somit können diese Verfahren die Präzision der Objekterkenner erhöhen, reduzieren aber auch den Recall. So können auch die nötigen Arbeitsschritte reduziert werden. Abbildung 13.7 zeigt einen Überblick über die einzelnen Verfahren und vergleicht die dabei nötigen Arbeitsschritte. Hinsichtlich der Personenerkenner hat combineSame das beste Gesamtergebnis erzielt und zu der größten Reduktion an Arbeitsschritten geführt. Weiterhin hat combineHierarchicalSame im zweiten Lauf das bessere Ergebnis erzielt.

Die aufgestellte Frage *Die Kombination der Ergebnisse automatischer Verfahren zur Beschriftung von Bildteilen verbessert die Qualität des entstehenden Gesamtergebnisses* kann somit basierend auf der gezeigten Evaluation generell mit ja beantwortet werden. Verbesserungen konnten sowohl für gleichartige Objekterkenner

13.7. LABORUNTERSUCHUNG 2: MESSUNG DER KOMBINIERTEN ERGEBNISSE VON OBJE

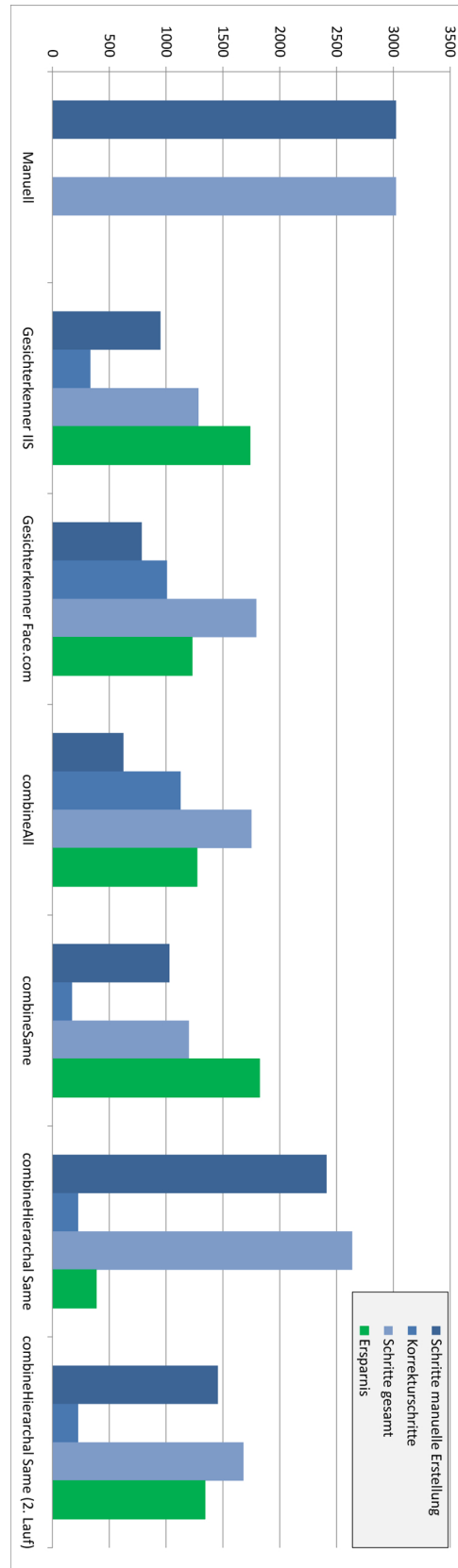


Abbildung 13.7: Vergleich richtig und falsch erkannter Bildteile und der resultierenden Ersparnis an Arbeitsschritten

wie auch für hierarchisch verbundene Objekterkenner durchgeführt werden.

Die Antwort auf Frage 3 dieser Arbeit lautet daher:

Frage 3

Verbessert die Kombination der Ergebnisse automatischer Verfahren zur Beschriftung von Bildteilen das Gesamtergebnis?

Antwort

Ja, das Gesamtergebnis der automatischen Verfahren kann durch Erstellung einer Hierarchie von Objekterkennern verbessert werden. Dadurch kann eine Steuerung zur Kombination aller Ergebnisse erkennen, welche Objekterkenner gleichartige und welche hierarchisch verbundenen Ergebnisse liefern. Durch die Kombination gleichartiger Ergebnisse kann die Anzahl an falsch erkannten Ergebnissen gesenkt werden. Durch die Kombination hierarchisch verwandter Ergebnisse kann die Wahrscheinlichkeit erhöht werden, dass der Inhalt einzelner Bildteile richtig erkannt wurde.

Anzumerken bleibt an dieser Stelle, dass wir sowohl im IMAGINATION Projekt wie auch ich in den eigenen Untersuchungen davon ausgegangen sind, dass zum Einen die Einzelergebnisse der eingesetzten Objekterkenner deutlich besser sind, sowie auch die Kombination dieser Ergebnisse zu einem deutlich besseren Gesamtergebnis führen können. Die Evaluation hat zwar gezeigt, dass die Kombination der Objekterkenner das Gesamtergebnis verbessern kann. Es wurde aber auch gezeigt, dass der Gesamtbeitrag zur automatischen Erstellung von Beschriftungen von Bildteilen noch deutlich hinter den gestellten Erwartungen durch Bildarchivare und auch den eigenen Erwartungen liegt. In der Erkennung von Bildteilen liegt somit noch ein hohes Verbesserungspotential, welches durch präzisere Objekterkenner ausgeschöpft werden sollte.

13.8 Zusammenfassung

Mit der Kombination automatischer Verfahren kann wie gezeigt und gewünscht die Menge an falsch erkannten Bildteilen reduziert werden.

Leider ist die Qualität vieler Objekterkenner noch nicht ausreichend gut genug, um eine Vielzahl von Objekten zu erkennen. Bereits das Erkennen von Personen, Autos und Flugzeugen war schon sehr unbefriedigend vom Gesamtergebnis.

Aufgrund der guten Ergebnisse der Kombination von Gesichtserkennern können in der Praxis der Bildagenturen somit aktuell Gesichter am besten erkannt werden.

Kapitel 14

Hochwertige Suchergebnisse mit ImageNotion

Wie in Kapitel 11 gezeigt, wurden innerhalb des IMAGINATION Projekts semantische Bildbeschriftungen für ein Filmarchiv erstellt. Die dafür nötigen Ontologien wurden komplett von den Bildarchivaren selbst erstellt. Innerhalb dieses Kapitels soll durch entsprechende Suchanfragen anschaulich gezeigt werden, dass damit tatsächlich die gewünschten hochwertigen Suchergebnisse dieser Arbeit möglich sind.

Der Beweis, dass die semantische Bildsuche tatsächlich zu besseren Suchergebnissen führt hinsichtlich Präzision und Vollständigkeit im Vergleich zur textbasierten Bildsuche, muss nicht mehr gezeigt werden. Entsprechende Laboruntersuchungen wie in [Nag07] und [Hol06] haben dieses eindeutig belegt. Daher wird stattdessen ein entsprechender Durchlauf durch die ImageNotion Anwendung mittels der Formulierung semantischer Suchanfragen gezeigt.

14.1 Bildsuche in einem Filmarchiv

In Sektion 2.6 wurde ein Szenario zur Bildsuche in einem Filmarchiv mittels der Schauspielerin „Elizabeth Taylor“ gezeigt. Hierbei haben in einem Workshop mit Bildsuchern im IMAGINATION Projekt die Teilnehmer semantische Suchanfragen wie spezifische Filmtitel, aber auch generelle Anfragen wie „Schauspieler England“ als geeignete Suchanfragen formuliert. Generelle Anfragen erfordern wie in Sektion 3.3.1 gezeigt den Einsatz entsprechender Möglichkeiten zur *semantischen Anfrageerweiterung*.

Im Filmarchiv zeigen 184 Bilder die Schauspielerin *Uma Thurman*. Da für *Elizabeth Taylor* nur sechs Bilder im Archiv vorhanden sind, wird zur Illustration der

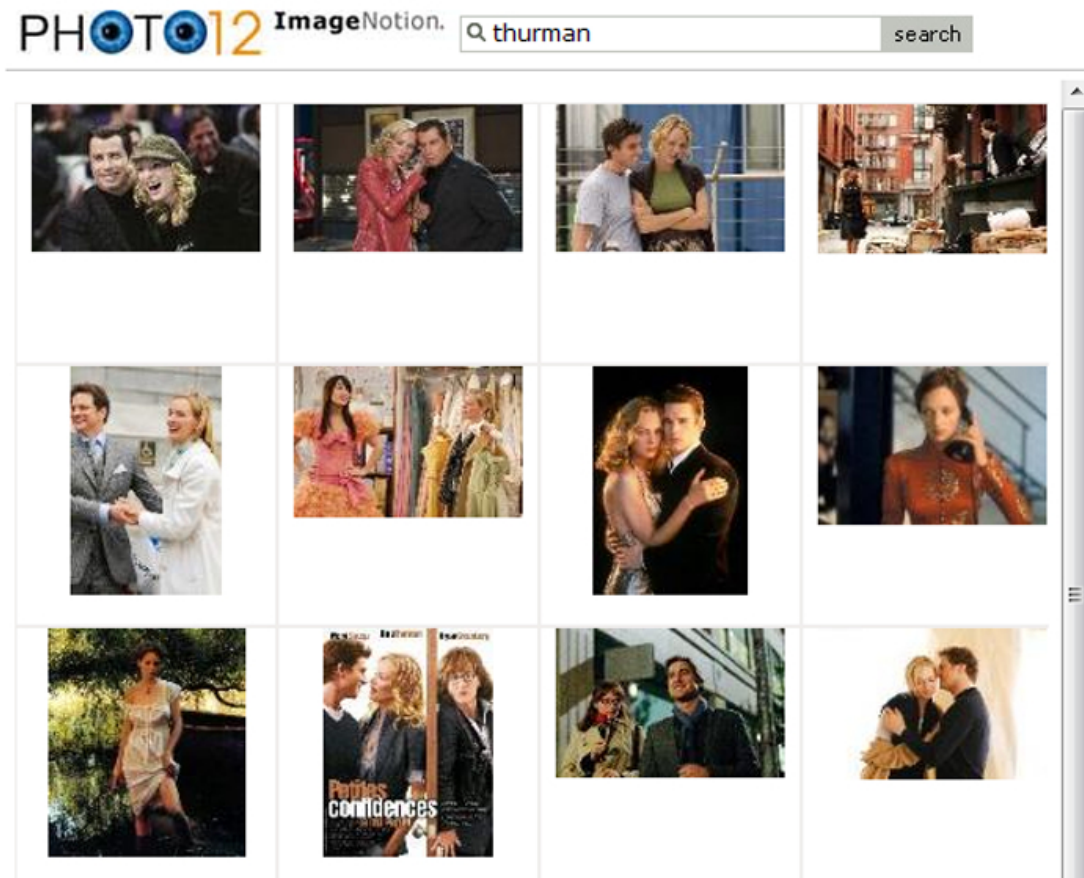


Abbildung 14.1: Ergebnis der textbasierten Bildsuche „Thurman“

Umsetzung semantischer Suchanfragen in ImageNotion daher *Uma Thurman* und entsprechende Anfragen verwendet. Entsprechende Suchanfragen und deren Umsetzungen sollen im Folgenden gezeigt werden.

14.2 Textbasierte Bildsuche „Thurman“

Diese Arbeit begann mit der Untersuchung von textbasierten Verfahren zur Bildsuche und dessen Schwächen. Abbildung 14.1 zeigt das Suchergebnis der textbasierten Bildsuche in ImageNotion. Da nur eine Person mit dem Stichwort „Thurman“ im Archiv vorhanden ist, zeigen zwar alle Bilder die Schauspielerin Thurman. Aber wie man sieht ist das Suchergebnis nicht nach Inhalt geordnet. Bildteile, die Uma Thurman sehr deutlich zeigen, werden daher nicht wie gewünscht bevorzugt.

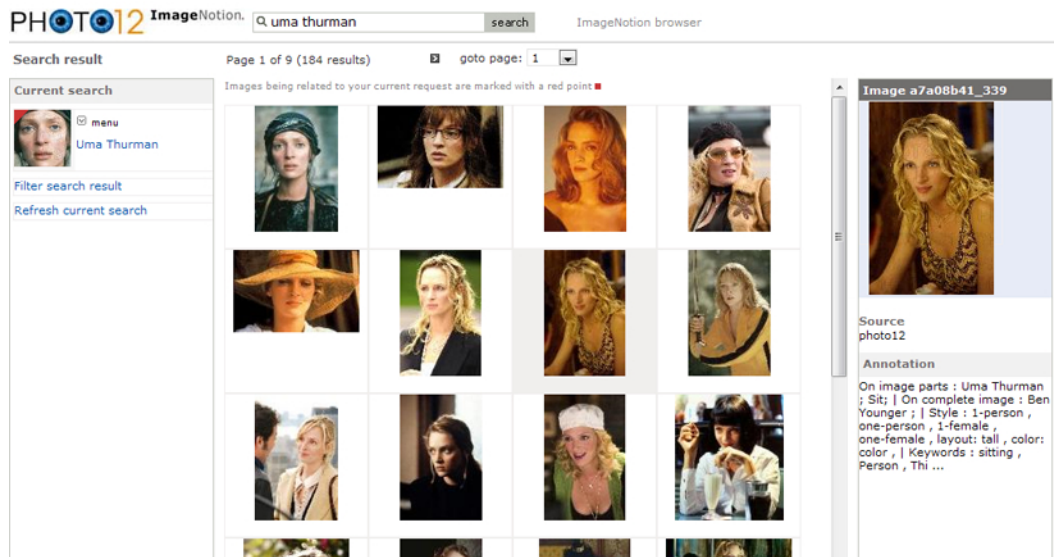


Abbildung 14.2: Ergebnis der semantischen Bildsuche *Uma Thurman*

14.3 Semantische Suche „Uma Thurman“

Abbildung 14.2 zeigt im Vergleich dazu die erste Suchergebnisseite der semantischen Anfrage nach *Uma Thurman*. Wie man sieht ist das Ergebnis in ImageNotion nach Größe der Bildteile bezüglich der semantischen Suchanfrage geordnet. Entsprechend sind Porträtbilder dieser Schauspielerin auf den ersten Plätzen dieser Suchanfrage.

Die Onlineevaluation in Kapitel 6 hat gezeigt, dass die meisten Teilnehmer ein so geordnetes Suchergebnis als besseres Suchergebnis eingestuft haben.

14.4 Semantische Suche „Quentin Tarantino“

Abbildung 14.3 zeigt das Ergebnis der semantischen Suchanfrage nach *Quentin Tarantino*, einem amerikanischen Regisseur und Schauspieler. Auch hier ist die Ordnung nach Bildteilen wichtig, weiterhin das Ausnutzen von Hintergrundwissen.

Die ersten vier Bilder zeigen Quentin Tarrantino direkt abgebildet. Diese Bilder haben somit eine höhere Relevanz als die weiteren Bilder, die Filmausschnitte zeigen, bei denen Quentin Tarrantino Regie hatte. ImageNotion konnte aufgrund der semantischen Metadaten diesen Unterschied ermitteln. Es werden zum Einen Bildbeschriftungen für Bildteile bevorzugt und zum Anderen Beschriftungen, bei denen die Metadaten direkt die Beschriftung *Quentin Tarrantino* enthalten.

Die restlichen im Suchergebnis gezeigten Bilder konnte ImageNotion aufgrund

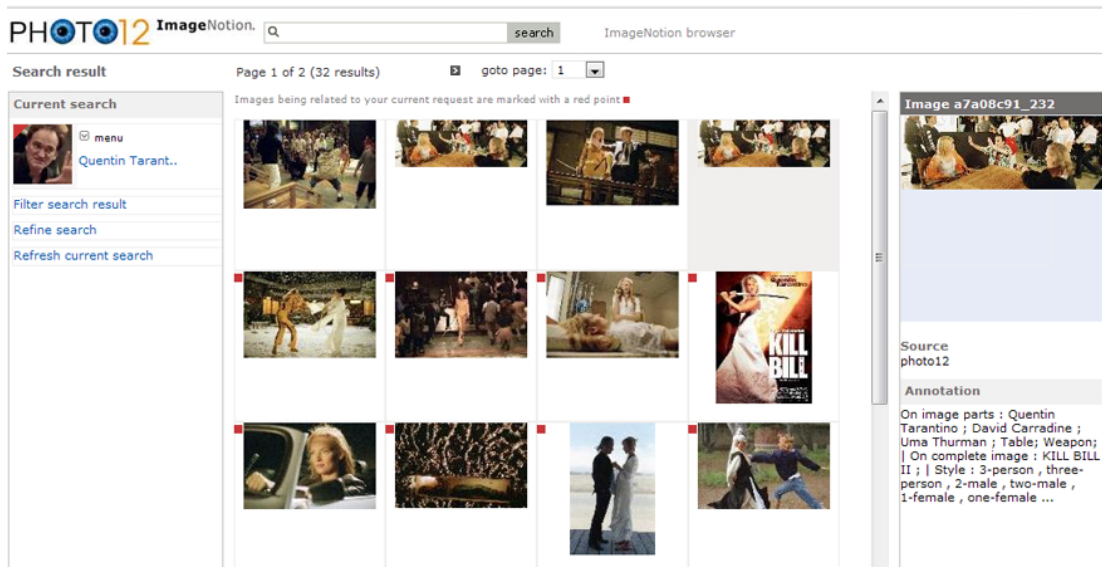


Abbildung 14.3: Ergebnis der semantischen Bildsuche *Quentin Tarrantino*

des Hintergrundwissens in den semantischen Bildbeschriftungen zu dem jeweiligen Filmtitel ermitteln, mit denen die Filmszenen beschriftet sind. Die dazu gehörigen semantischen Elemente wie *Pulp Fiction* und *Kill Bill* enthalten nämlich die Relation „hat Regisseur“ *Quentin Tarrantino*.

14.5 Interpretation der Suchanfrage „Weibliche Schauspieler“

Ein Beispiel für eine generelle und oft an ein Filmarchiv gestellte Suchanfrage ist „weibliche Schauspieler“. Eine solche Suchanfrage muss entsprechend interpretiert werden. Es müssen alle Bilder ermittelt werden, die Schauspieler und vom Geschlecht weiblich sind.

ImageNotion muss daher die Anfrage *Schauspieler* auflösen können zu „ermittle alle Bilder, die Schauspieler zeigen“. Anschließend muss diese Menge relevanter Bilder gefiltert werden nach dem Geschlecht der abgebildeten Schauspieler. Übrig bleiben alle Bilder, die weibliche Schauspieler zeigen. Abbildung 14.4 zeigt das dadurch in ImageNotion entstandene Suchergebnis.

Das gezeigte Suchergebnis war möglich durch das Ausnutzen des Hintergrundwissens in den semantischen Elementen. Abbildung 14.5 zeigt die Relationen des semantischen Elements *Uma Thurman*, wie diese mit Hilfe der ImageNotion Metho-

14.5. INTERPRETATION DER SUCHANFRAGE „WEIBLICHE SCHAUSPIELER“219

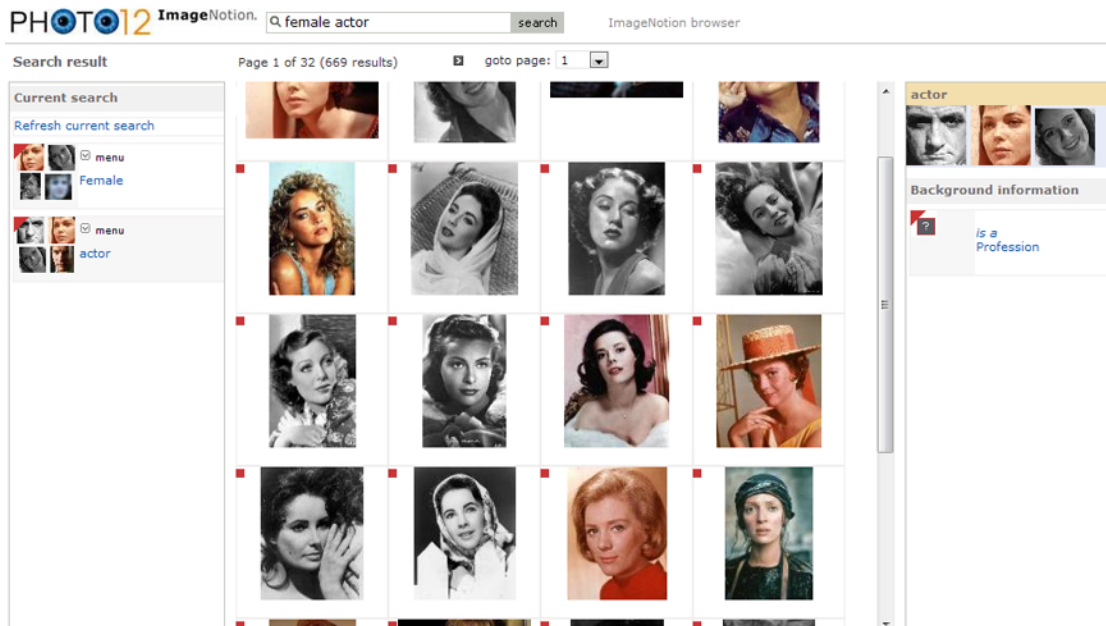


Abbildung 14.4: Ergebnis der semantischen Bildsuche *Weibliche Schauspieler*

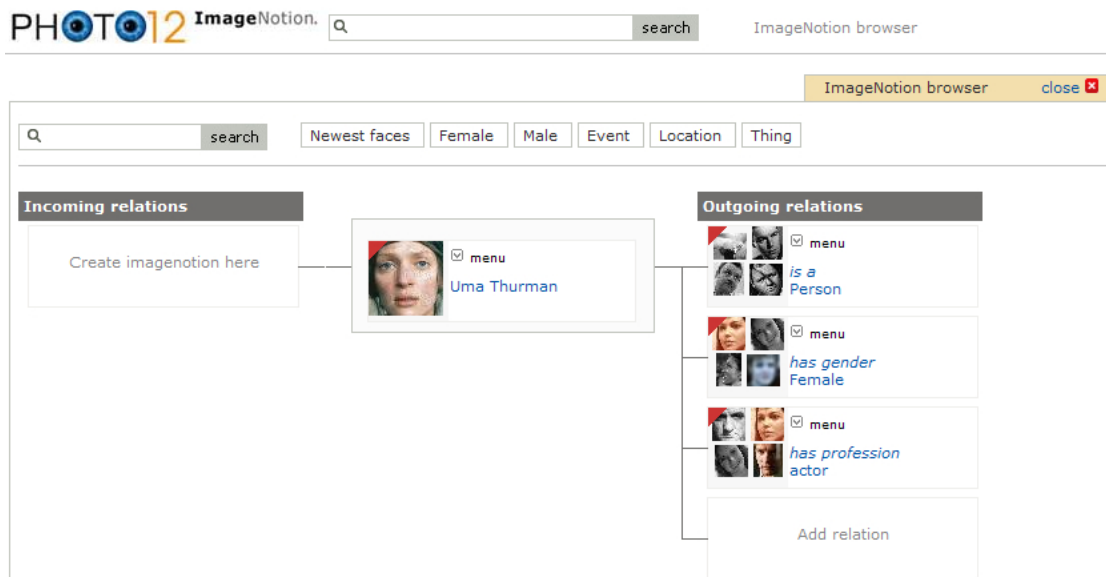


Abbildung 14.5: Hintergrundwissen des semantischen Element *Uma Thurman*

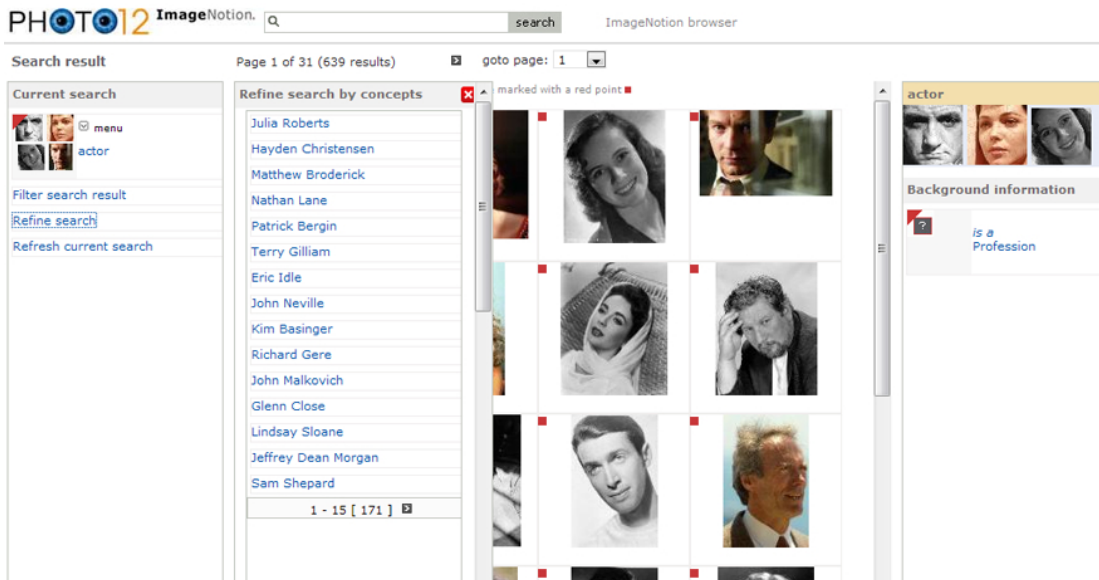


Abbildung 14.6: Vorschläge zur strukturierten Verfeinerung der Suchanfrage *Schauspieler*

de durch die Bildarchivare erstellt wurden. So wurde angegeben, dass Uma Thurman „eine Person ist“, „vom Geschlecht weiblich ist“ und „von Beruf Schauspielerin“ ist. Diese Hintergrundinformationen zu den semantischen Elementen aller Schauspieler wurde von ImageNotion ausgenutzt, um das in Abbildung 14.4 gezeigte Suchergebnis zu erstellen.

14.6 Strukturierte Verfeinerung der Suchanfrage „Schauspieler“

Die Onlineevaluation hat ergeben, dass Benutzer die strukturierte Verfeinerung von semantischen Suchanfragen verstehen und gerne in eine Bildsuchmaschine integriert hätten. ImageNotion enthält daher ein entsprechendes Verfahren. Abbildung 14.6 zeigt Vorschläge zur Verfeinerung der Suchanfrage *Schauspieler*. Aufgrund des in ImageNotion vorhandenen Hintergrundwissen können dabei Vorschläge wie *Julia Roberts* oder *John Malkovich* als Verfeinerung der generellen Suchanfrage *Schauspieler* hin zu spezifischen Anfragen einzelner Schauspieler geben.

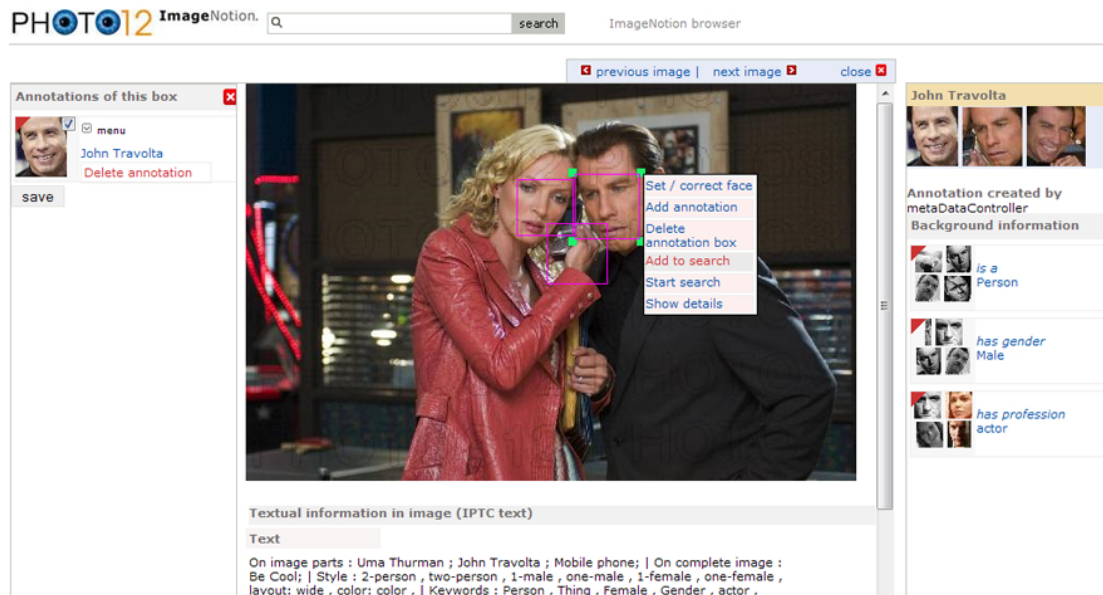


Abbildung 14.7: Navigation durch das Bildarchiv mit Bildteil *John Travolta*

14.7 Navigation durch das Bildarchiv

Wie Abbildung 14.7 zeigt, erlauben beschriftete Bildteile die Angabe darüber, was auf einem Bildteil abgebildet ist. So zeigt der Bildteil rechts in der Abbildung *John Travolta*. ImageNotion hat somit das *Wissen über den Inhalt von Bildteilen*. Dieses Wissen kann ausgenutzt werden, um durch das Bildarchiv zu navigieren. Jeder Bildteil hat ein Menü zum Stellen einer Suchanfrage. Wie die Onlineevaluation ergab, soll es möglich sein, bei einem Bildteil eine neue Suchanfrage zu starten oder die aktuelle Suchanfrage mit zu berücksichtigen.

Die aktuelle Suchanfrage war *Uma Thurman*. Mit dem Klick auf den Link „Add Search“ kann die Suchanfrage erweitert werden. Abbildung 14.8 zeigt das so durch Klick auf Bildteile entstandene Suchergebnis. Es zeigt nur diejenigen Bilder des Archivs, die *Uma Thurman* und *John Travolta* enthalten.

14.8 Zusammenfassung

In Kapitel 2 wurden als Ansätze für hochwertige Suchergebnisse die drei Bedingungen formuliert, dass eine Bildsuchmaschine *Semantik der Suchanfrage*, *Semantik der Bilder* und *Semantik von Bildteilen* benötigt. Entsprechende Ansätze wurden in dieser Arbeit ermittelt und durch umfangreiche empirische Untersuchungen und

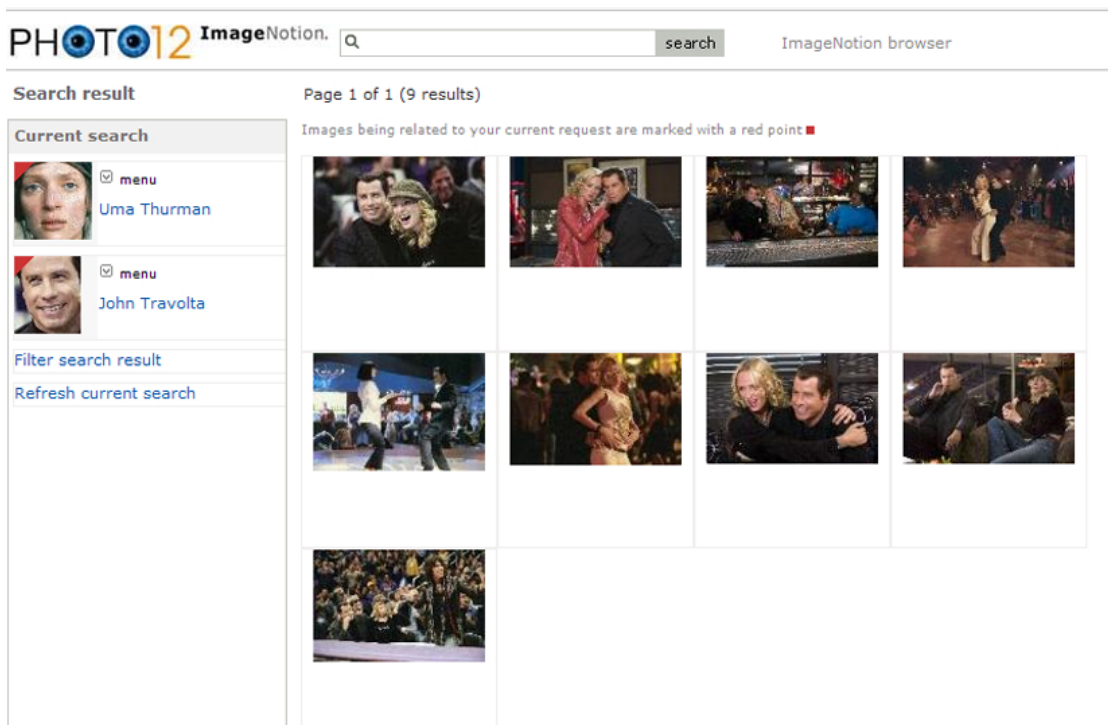


Abbildung 14.8: Suchergebnis nach Klick auf Bildteil *John Travolta*

Evaluationen so geeignete Lösungen und Verfahren entwickelt, die in der ImageNotion Anwendung umgesetzt worden sind.

Die Beispiele in diesem Kapitel und die dabei entstandenen hochwertigen Suchergebnisse zu den jeweiligen semantischen Suchanfragen haben dies anschaulich gezeigt.

Kapitel 15

Zusammenfassung und Ausblick

Eine Zusammenfassung der Ergebnisse dieser Arbeit und einen Ausblick zu dieser Arbeit gibt das folgende, abschließende Kapitel.

15.1 Zusammenfassung

Das formulierte Ziel dieser Arbeit war der Aufbau eines Gesamtsystem mit dem Namen *ImageNotion* [WN07a], welches hochwertige Suchergebnisse auf der Basis von semantischen Verfahren und automatischen Verfahren zur Beschriftung von Bildteilen ermöglichen soll.

Zur genauen Ermittlung der Umsetzung wurde eine Reihe empirischer Untersuchungen und Evaluationen durchgeführt, um vier Fragen geeignet beantworten zu können:

1. Wie sollen semantische Verfahren eingesetzt werden?

Es sollte geklärt werden, wie semantische Verfahren innerhalb einer Suchmaschine eingesetzt werden müssen, ob diese überhaupt in der Bildsuche benötigt werden und ob semantische Verfahren im Einsatz in einer Bildsuchmaschine verstanden wurden.

Die Analyse bestehender textbasierter Bildsuchmaschinen sowie die Befragung von Geschäftsführern von fünf Bildagenturen führten aufgrund der von diesen geforderter Möglichkeit zum flexiblen Übergang von einer textbasierten zu einer semantischen Bildsuchmaschine zur folgenden Entscheidung:

Frage 1 (a): An welcher Stelle in einer Bildsuchmaschine sollen semantische Verfahren eingesetzt werden?

Grundlegende Entwurfsentscheidung: *ImageNotion muss als eine Hybridlösung aus textbasierten und semantischen Metadaten betrieben werden.* Der Einsatz semantischer Verfahren muss sowohl an der Stelle der Anfrageerweiterung wie im Bereich der Erstellung semantischer Metadaten erfolgen.

Die Analyse von mehr als 100.000 Suchanfragen an fotomarktplatz.de, die von mehr als 6.000 registrierten Benutzern nach Bildmaterial zur Nutzung in Verlagen, Werbeagenturen, Firmen und Zeitungen gestellt wurden, ergab die Notwendigkeit für semantische Verfahren in der Bildsuche:

Frage 1 (b)

Werden semantische Verfahren zur Verbesserung der Bildsuche benötigt?

Antwort

Ja, semantische Verfahren zur Verbesserung der Bildsuche werden benötigt. Sie sollen die hohe Anzahl an Suchanfragen pro Sitzung und damit den Zeitaufwand für die Benutzer reduzieren. Hierzu eignen sich die Verfahren *lexikalische Anfrageerweiterung*, *semantische Anfrageerweiterung*, *Beschriftung von Bildteilen* sowie die *Neuformulierung von Suchanfragen*.

Schließlich konnte durch die Durchführung einer Onlineevaluation mit 137 Teilnehmern die Frage nach dem praktischen Verständnis für semantische Verfahren durch Benutzer mit einem Prototyp der ImageNotion Anwendung geklärt werden:

Frage 1 (c)

Sind die entstehenden Möglichkeiten zur Bildsuche für Nutzer in der Praxis verständlich und wie sollen diese eingesetzt werden?

Antwort

Benutzer verstehen die Idee der semantischen Suche, können semantische Suchanfragen formulieren und nachvollziehen und empfanden weiterhin die entstehenden Ergebnisse als besser im Vergleich zum textbasierten Suchergebnis.

Als Verfahren zur Formulierung von Suchanfragen sind *Vorschlag semantischer Elemente* und *Auflösung einer semantischen Suchanfrage* zu verwenden.

2. Wie und von wem werden semantische Hintergrundinformationen erstellt? Da der Erfolg semantischer Anwendungen maßgeblich von der *Umsetzbarkeit* und der *Akzeptanz* durch die Benutzer abhängt [Hep07], wurde untersucht, wie Bildarchivare mit den bestehenden Ansätzen zur Erstellung von Ontologien und Nut-

zung in semantischen Anwendungen zurecht kommen. Das Ergebnis eines Workshops mit sechs Bildarchivaren im IMAGINATION Projekt ergab das folgende Ergebnis:

Frage 2 (a)

Sind getrennt erstellte, vorgefertigte Ontologien ausreichend und praktisch anwendbar?

Antwort

Nein, der getrennte Ansatz der Ontologieerstellung ist für Bildarchivare *weder ausreichend noch praktisch durchführbar*, denn dieser Ansatz ist zu umständlich und unpraktisch und weiterhin zu teuer. Hinzu kommt, dass die Elemente einer getrennt erstellten Ontologie oft nicht verstanden werden oder Elemente vorhanden sind, die nicht benötigt werden.

Die Suche nach einer geeigneten Möglichkeit, damit Benutzer selbst benötigte Ontologien zur Erstellung semantischer Metadaten erstellen können, führte zur Formulierung eines geeigneten Phasenmodells, dem *Ontology Maturing Process Model*. Es beschreibt, wie Benutzer selbst Ontologien entwickeln können. Auf der Grundlage dieses Modells wurde die ImageNotion Methodologie entwickelt:

Frage 2 (b)

Wie können Bildredakteure in diesen Prozess eingebunden werden und selbst Hintergrundwissen erstellen?

Antwort

Die ImageNotion Methode ermöglicht auf Basis des *Ontology Maturing Process Model* die nötigen Schritte, damit Benutzer die zur Erstellung semantischer Metadaten benötigten leichtgewichtigen Ontologien erstellen können. Fehlen semantische Elemente für die Bildbeschriftung, werden diese von einem Benutzer initial hinzugefügt. Anschließend werden diese *imagenotions* durch das Hinzufügen von textuellen Informationen zu einem kontrollierten Vokabular konsolidiert. Das Hinzufügen von Relationen ermöglicht schließlich den flexiblen Übergang zu Taxonomien und Instanzenmodellen. Die durch Benutzer erstellten *imagenotions* und Ontologien erlauben so einen praxistauglichen Ablauf zur Erstellung semantischer Metadaten.

Um zu ermitteln, ob eine semantische Bildbeschriftung in der Praxis überhaupt hinsichtlich des entstehenden Zeitaufwands möglich ist, wurde ein während des IMAGINATION Projekt mit semantischen Metadaten versehenes Filmarchiv entsprechend untersucht:

Frage 2 (c)

Reduzieren semantische Verfahren den Zeitaufwand zur Erstellung von Metadaten?

Antwort

Die Auswertung des Zeitaufwands zur Beschriftung eines Filmarchivs mit 722 Bildern hat gezeigt, dass semantische Verfahren die Bildbeschriftung beschleunigen. Im Vergleich zur textbasierten Bildbeschriftung führen die Nutzung einer Taxonomie zu einer Zeitersparnis von 34 Prozent und die Nutzung eines Instanzenmodells sogar zu einer Zeitersparnis von 50,1 Prozent.

3. Verbessert die Kombination der Ergebnisse automatischer Verfahren zur Beschriftung von Bildteilen das Gesamtergebnis? Die Evaluation von Objekterkennern mit diesem Filmarchiv ergab, dass die Objekterkennung viele Fehler produzieren, hauptsächlich durch die Markierung von Bildteilen, die nicht das angegebene Objekt abbilden. Entsprechend wurde die Idee ausformuliert und umgesetzt, verschiedene Objekterkennung auf der Basis semantischer Beschriftungen und Informationen miteinander zu kombinieren:

Frage 3

Verbessert die Kombination der Ergebnisse automatischer Verfahren zur Beschriftung von Bildteilen das Gesamtergebnis?

Antwort

Ja, das Gesamtergebnis der automatischen Verfahren kann durch Erstellung einer Hierarchie von Objekterkennern verbessert werden. Dadurch kann eine Steuerung zur Kombination aller Ergebnisse erkennen, welche Objekterkennung gleichartige und welche hierarchisch verbundenen Ergebnisse liefern. Durch die Kombination gleichartiger Ergebnisse kann die Anzahl an falsch erkannten Ergebnissen gesenkt werden. Durch die Kombination hierarchisch verwandter Ergebnisse kann die Wahrscheinlichkeit erhöht werden, dass der Inhalt einzelner Bildteile richtig erkannt wurde.

Aufgrund meiner langjährigen Tätigkeit und Erfahrung im Bereich der Bildsuche war es mir in dieser Arbeit sehr wichtig, Lösungen und Umsetzungen zu ermitteln, die sich praktisch umsetzen lassen. Daher wurden die Fragestellungen wann immer möglich mit Hilfe von Befragungen und Workshops mit geeigneten Benutzern begleitet. So gelang es, von der Analyse über die Prüfung der Umsetzbarkeit bis hin zur Lösungssuche die gestellten Fragen dieser Arbeit durch die Auswertung der empirischen Untersuchungen und durch Evaluationen beantworten zu können.

Die gewonnen Erkenntnisse der Analyse von Geschäftsführern von Bildagentu-

ren ermöglichte eine entsprechende Erstellung der Systemarchitektur von ImageNotion. Die Analyse von mehr als 100.000 Suchanfragen von mehr als 6.000 Benutzern aus fotomarktplatz.de zeigte deutlich, dass semantische Verfahren bei der Bildsuche benötigt werden, um hochwertige Suchergebnisse zu ermöglichen. Die Onlineevaluation mit mehr als 130 Teilnehmern zeigte, dass die möglichen semantischen Verfahren auch von Benutzern in der Bildsuche verstanden und gewollt werden. Mit Hilfe von sechs Bildarchivaren in einem Workshop zur Erstellung von semantischen Metadaten wurde analysiert, inwieweit diese mit vorgefertigten Ontologien umgehen können und welche Anforderungen diese stellen, um semantische Metadaten sinnvoll erstellen zu können. Die Analysephase hat somit zum Einen die praktische Notwendigkeit semantischer Verfahren belegt, weiterhin konnten auch die Anforderungen aus der Realwelt selbst ermittelt und abgeleitet werden.

Zur Suche von Lösungen wurden ebenfalls die Benutzer in Workshops mit einbezogen. Ihre Ideen zu den nötigen Abläufen für eine praktisch umsetzbare Lösung für die Erstellung semantischer Metadaten führten zur ImageNotion Methode. Die anschließende empirische Untersuchung dieser Methode mit Fachbenutzern hat gezeigt, dass dadurch die Erstellung von semantischen Metadaten deutlich besser möglich wurde. So entstand eine in der Praxis nutzbare Lösung für die Erstellung semantischer Metadaten. Ebenfalls konnte durch die Zeitmessung für die Erstellung von semantischen Metadaten belegt werden, dass diese Form der Bildbeschriftung nicht nur eine deutlich höhere Qualität hat (wie bereits aus Labormessungen bekannt war), sondern auch der Zeitaufwand gleichzeitig um 50 Prozent reduzieren konnte. In der Praxis bedeutet dies, dass die Erstellung semantischer Metadaten umsetzbar ist, da damit die Kosten für die Bildbeschriftung für die Betreiber deutlich reduziert werden können und gleichzeitig sogar eine höhere Qualität der Bildbeschriftungen entsteht.

Ebenso wurde untersucht, wie gut Objekterkenner operieren, wenn diese in einem Filmarchiv praktisch eingesetzt werden. Aufgrund der vom Ergebnis her eher ernüchternden Resultate führte dies zu Ideen, wie man das Gesamtergebnis durch die Kombination der Ergebnisse verbessern kann. Eine entsprechende Laboruntersuchung hat gezeigt, dass die Kombination der Ergebnisse tatsächlich oftmals das Gesamtergebnis verbessern können.

Zusammengefasst hat diese Arbeit somit gezeigt, wie man „rapid prototyping“ betreiben kann, um daraus ein Gesamtsystem erstellen zu können, welches praktisch nutzbar ist und gleichzeitig das gestellte Ziel erfüllen kann, hochwertige Suchergebnisse generieren zu können.

15.2 Ausblick

Mit dem Ausblick soll zum Einen gezeigt werden, wie die in dieser Arbeit erarbeiteten Ergebnisse bereits eingesetzt werden. Weiterhin werden offene wissenschaftliche Fragestellungen, die während der Erstellung dieser Arbeit nicht gelöst wurden.

15.2.1 Einsatz von ImageNotion

PRIMO: Privacy on Image Objects

PRIMO ist ein am IPD Karlsruhe entwickeltes Projekt [BWBB08]. Es soll den Schutz der eigenen Persönlichkeit auf Bildobjekten unterstützen. Hierzu wurde ein System mit dem Namen PRIMO entwickelt, das in sozialen Netzwerkseiten wie Facebook oder Flickr neu aufgeladene Bilder lädt und diese durch die Objekterkennung in ImageNotion mit semantischen Metadaten versieht.

Nutzer von PRIMO können zusätzlich Trainingsbilder aufladen, die ihr Gesicht zeigen, um den Objekterkennung zur Erkennung von Gesichtern zu trainieren. Weiterhin können die Benutzer Regeln formulieren, wie sie benachrichtigt werden wollen, wenn PRIMO Bilder entdeckt, auf denen der Benutzer abgebildet ist. Der Benutzer kann so auf Bilder reagieren, die er nicht in einer sozialen Netzwerkseite veröffentlicht sehen will, indem er zum Beispiel die Löschung des Bilds beauftragt.

Das ImageNotion Startup

Aufgrund dem großen Interesse durch Betreiber von Bildagenturen auf die ImageNotion Anwendung während des IMAGINATION Projekts zum Beispiel bei Konferenzen und Ausstellungen wie der Photokina in Köln haben sich Gabor Nagypal und Andreas Walter zur Fortführung der ImageNotion Entwicklung mit der Gründung der ImageNotion UG entschlossen.

Die ImageNotion Anwendung soll hierbei kommerziell an Bildagenturen vertrieben werden, damit diese von den Einsparmöglichkeiten auf Seiten der Bildbeschriftung und deren Kunden von den Verbesserungen in der Bildsuche profitieren können. Abbildung 15.1 zeigt die Innenseite der hierfür entwickelten Werbebroschüre. Erste Bildagenturen konnten bereits gefunden werden, die ImageNotion in Testbetrieben einsetzen.

ImageNotion. Effiziente Bildverwaltung mit Intelligenz www.imagenotion.com

Wie können Sie Ihre **Bildsuche verbessern** und gleichzeitig **Zeit und Kosten** bei der Bildbeschriftung **sparen?**

Bild, semantisches Modell und Beschriftungen entstammen dem Cinema-Archiv der Bildagentur photo12, Paris, www.photo12.com

Bessere Qualität der Suchergebnisse
Suche **Schauspieler**: ImageNotion weiß selbst, dass dann Bilder, die **Uma Thurman** oder **Angelina Jolie** enthalten, mit zu zeigen sind.

Schnellere und bessere Beschriftung
ImageNotion kennt die Bedeutung von Beschriftungen und fügt selbst wichtige Stichwörter hinzu, z.B. **E-Type**: **Jaguar, Auto, Oldtimer** und **Automodell**.

Abbildung 15.1: Innenansicht des ImageNotion Broschüre 2010

15.2.2 Offene wissenschaftliche Fragestellungen

Umfangreiches Ausnutzen von Relationen bei der Bildsuche

Wie in der Onlineevaluation in Kapitel 6 gezeigt, scheint ein generelles Verständnis für Relationen zwischen semantischen Elementen zu bestehen. In dieser Arbeit wurde aber nicht ermittelt, inwieweit hierdurch die Möglichkeiten zur Eingabe von Suchanfragen die Suchergebnisse für Benutzer verbessern.

Ein Beispiel zur Eingabe von Suchanfragen mit Hilfe von Relationen wäre „Finde alle Bilder, bei denen ein Bildteil australische Schauspielerinnen zeigt und ein anderer Bildteil männliche Schauspieler“. Eine Umsetzung einer solchen Suchanfrage würde eine entsprechende Benutzerschnittstelle für die Eingabe solcher Suchanfragen anbieten, welche den Inhalt einzelner Bildteile spezifizieren lässt. Der Benutzer könnte dann angeben, dass er für Bildteil 1 „weibliche Schauspielerinnen“ finden möchte. Ein entsprechend aufgebauter Index müsste dann semantische Elemente anbieten, die sich auf weiteren Bildteilen befinden, also zum Beispiel „männliche Schauspieler“.

Benutzerworkshops müssten dann zum Einen ermitteln, welche Möglichkeiten solcher auf Relationen basierenden Suchanfragen von Benutzern verstanden werden und wie die Benutzerschnittstelle geeignet aufgebaut werden muss.

Kontextabhängige Umsetzung von Suchanfragen und Anzeige von Ergebnissen

Die Onlineevaluation aus Kapitel 6 ergab weiterhin, dass sich mehr als 50 Prozent der Teilnehmer an einer zu großen Menge an beschrifteten Bildteilen stören würden. Aus dieser Erkenntnis kann somit die Fragestellung aufgeworfen werden, wie eine kontextabhängige Umsetzung von Suchanfragen und die Anzeige der Suchergebnisse aussehen müsste. Eine kontextabhängige Suche würde die bisher gestellten Suchanfragen einer Sitzung mit betrachten und könnte so zum Beispiel ermitteln, dass ein Benutzer nach Filmtitel sucht. Entsprechend könnte bei der Umsetzung von Suchanfragen das Ergebnis so geordnet werden, dass bei Suchanfragen Bilder bevorzugt werden, die sich auf Filmtitel beziehen. Bei der Suchanfrage *Nicole Kidman* könnten dadurch erst Bilder aus Filmen gezeigt werden, bevor Bilder von Presseveranstaltungen oder Preisverleihungen gezeigt werden.

Entsprechend könnten die Bildbeschriftungen bei der Anzeige im Suchergebnis aufgrund des Kontextes der Suchanfrage selbst gefiltert werden. Ein Beispiel hierbei wäre, bei der Suchanfrage *Schauspieler* nur die Bildteile auf den Bildern anzuzeigen, die Personen zeigen.

Entsprechende Verfahren zur Berücksichtigung eines Benutzerkontextes müssten für diese Fragestellung entsprechend konzeptioniert und umgesetzt werden sowie mit Hilfe von Benutzern bezüglich der genauen Umsetzung und Verständlichkeit untersucht werden.

Zeitliche Veränderung von Inhalten in Ontologien für die Bildsuche

Innerhalb des IMAGINATION Projekts ist aufgefallen, dass sich die Inhalte von Ontologien für die Bildsuche zeitlich verändern beziehungsweise sich auf bestimmte Zeiträume beziehen müssen. Konkret fiel dies bei der Beschriftung von Bildern der aktuellen Bundeskanzlerin Angela Merkel auf. Von dieser Politikerin lagen Bilder vor in ihrer Zeit als Umweltministerin wie auch als Bundeskanzlerin. Korrekterweise sollte die Bildsuche Angela Merkel dann wissen, dass sie aktuell Bundeskanzlerin ist, um diese Bilder im Suchergebnis bevorzugen zu können.

Ideen zur Umsetzung dieser zeitlichen Veränderung von Inhalten in Ontologien wären zum Einen, zu jeder Relation einen zeitlichen Bezug zu geben oder Facetten einzelner semantischer Elemente zu ermöglichen. Facetten für Angela Merkel wären

dann „hat Beruf“ „Umweltministerin“ (mit Datumsangabe) und „hat Beruf“ „Bundeskanzlerin“ mit Datumsangabe. Entsprechend könnten dann diese Facetten für die Beschriftung der Bilder verwendet werden.

Die hauptsächlichen Fragestellungen sind hierbei also die genaue Umsetzung der Berücksichtigung zeitlicher Aspekte wie auch eine entsprechende Indizierung und Anzeige der zeitlichen Unterschiede der Ergebnisse auf der Benutzerschnittstelle.

Aufbau eines semantischen Fotomarktplatzes

Die Nutzung der in der Systemarchitektur von ImageNotion (siehe Sektion 3.4) vorgesehenen *Web Service Schnittstelle* ermöglicht das Zusammenschalten mehrerer Systeme. Dies ermöglicht die Erstellung eines *semantischen Fotomarktplatzes*. Ein Benutzer könnte so mittels einer semantischen Suchanfrage parallel über mehrere Bildsuchmaschinen hinweg Suchergebnisse erhalten.

Die dabei auftretenden Fragen sind ähnliche Fragestellungen, wie diese sich auch im Semantic Web [SBLH06] ergeben. Hierzu gehört der einheitliche Austausch semantischer Informationen. Wie können also Bildbeschriftungen geeignet ausgetauscht und einheitlich interpretiert werden? Dies erfordert die geeignete Verbindung der einzelnen Ontologien in den Systemen und die Erstellung eines einheitlich zu verwendeten Austauschformats für die semantischen Metadaten.

Die *ImageNotion Methode* war in dieser Arbeit auf kleine Gruppen ausgelegt, die zusammen eine Ontologie für eine Bildsuchmaschine erstellen. Hinsichtlich eines aus mehreren Systemen verbundenen Systems ergäben sich hierbei Fragestellungen der Kollaboration großer Gruppen, also zum Beispiel wie sich eine große Gruppe von Bildarchivaren untereinander auf zu verwendende semantische Elemente einigen kann.

Anhang A

Datenbasen der Untersuchungen

Die Datenbasen der durchgeführten Untersuchungen sind zu umfangreich zum Abdruck als Anhang in diesem Buch.

Diese sind daher auf entsprechende Anfrage in elektronischer Form erhältlich.

Kontakt

Andreas Walter, research@imagenotion.com

ImageNotion Installation für wissenschaftliche Arbeiten Kostenlose ImageNotion Installation zur Durchführung wissenschaftlicher Arbeiten sind möglich. Bitte nehmen Sie hierzu ebenfalls mit mir Kontakt auf.

Literaturverzeichnis

- [AHAQI⁺07] Omara Abdul Hamid, Muhammad Abdul Qadir, Nadeem Iftikhar, Mohib Ur Rehman, Mobin Uddin Ahmed, and Imran Ihsan. Generic multimedia database architecture based upon semantic libraries. *Informatica*, 18(4):483–510, 2007.
- [BF06] Dunja Mladenic Blaz Fortuna, Marko Grobelnik. Semi-automatic data-driven ontology construction system. In *Proc. of the 9th International multi-conference Information Society IS-2006, Ljubljana*, 2006.
- [BP98] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.*, 30(1-7):107–117, 1998.
- [BSW⁺07] Simone Braun, Andreas Schmidt, Andreas Walter, Gabor Nagypal, and Valentin Zacharias. Ontology maturing: a collaborative web 2.0 approach to ontology engineering. In *Proceedings of the Workshop on Social and Collaborative Construction of Structured Knowledge at the 16th International World Wide Web Conference (WWW 07), Banff, Canada*, 2007.
- [BSWZ07] Simone Braun, Andreas Schmidt, Andreas Walter, and Valentin Zacharias. The ontology maturing approach to collaborative and work-integrated ontology development: Evaluation results and future directions. In *International Workshop on Emergent Semantics and Ontology Evolution (ESOE), 6th International Semantic Web Conference (ISWC 2007), Busan, Korea, Nov. 11-15 2007*, 2007.
- [BWBB08] Thorben Burghardt, Andreas Walter, Erik Buchmann, and Klemens Böhm. Primo - towards privacy aware image sharing. In *Web Intelligence/IAT Workshops*, pages 21–24, 2008.

- [BYRN99] Ricardo A. Baeza-Yates and Berthier Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1999.
- [CCC10] Ming-Yen Chen, Hui-Chuan Chu, and Yuh-Min Chen. Developing a semantic-enable information retrieval mechanism. *Expert Syst. Appl.*, 37(1):322–340, 2010.
- [CDG⁺05] Nick Crofts, Martin Doerr, Tony Gill, Stephen Stead, and Matthew Stiff. Definition of the cidoc conceptual reference model version 4.2. In *CIDOC CRM Special Interest Group*, 2005.
- [CGSW03] Edward Chang, Kingshy Goh, Gerard Sychay, and Gang Wu. Cbsa: Content-based soft annotation for multimodal image retrieval using bayes point machines. *IEEE Transactions on Circuits and Systems for Video Technology*, 13:26–38, 2003.
- [DJLW08] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv.*, 40(2):1–60, 2008.
- [FLGP02] Mariano Fernández-López and Asun Gómez-Pérez. A survey on methodologies for developing, mainaining, integrating, evaluating and reengineering ontologies. Deliverable 1.4, EU IST Project IST-2000-29243 OntoWeb, 2002.
- [Fre92] Gottlob Frege. Über sinn und bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, NF 100:25–50, 1892.
- [FTN⁺00] C. Olivia Frost, Bradley Taylor, Anna Noakes, Stephen Markel, Deborah Torres, and Karen M. Drabenstott. Browse and search patterns in a digital image database. *Inf. Retr.*, 1(4):287–313, 2000.
- [Gru93] Thomas R. Gruber. A translation approach to portable ontology specifications. *Knowl. Acquis.*, 5(2):199–220, 1993.
- [Gua98] Nicola Guarino. Formal ontology and information systems. In Nicola Guarino, editor, *Proceedings of the 1st International Conference on Formal Ontologies in Information Systems (FOIS'98)*, pages 3–15, Trento, Italy, June 1998. IOS Press.
- [Gud95] Venkat N. Gudivada. Content-based image retrieval systems (panel). In *CSC '95: Proceedings of the 1995 ACM 23rd annual conference on Computer science*, page 274, New York, NY, USA, 1995. ACM.

- [GW01] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2001.
- [Hep07] Martin Hepp. Possible ontologies: How reality constrains the development of relevant ontologies. In *IEEE Internet Computing, 11 2007*, 2007.
- [HKS06] Ian Horrocks, Oliver Kutz, and Ulrike Sattler. The even more irresistible sroiq. In Patrick Doherty, John Mylopoulos, and Christopher A. Welty, editors, *KR*, pages 57–67. AAAI Press, 2006.
- [HM06] Eero Hyvönen and Eetu Mäkelä. E.: Semantic autocompletion. In *In: Proceedings of the first Asia Semantic Web Conference (ASWC 2006)*, pages 4–9. Springer-Verlag, 2006.
- [Hol06] Laura Hollink. *Semantic annotation for retrieval of visual resources*. PhD thesis, Vrije Univierity of Amsterdam, 2006.
- [Int05] International Press Telecommunications Council. IPTC Core Schema for XMP Version 1.0 Specification document, 2005.
- [JJ05] Corinne Jörgensen and Peter Jörgensen. Image querying by image professionals: Research articles. *J. Am. Soc. Inf. Sci. Technol.*, 56(12):1346–1359, 2005.
- [JJfC00] Alejandro Jaimes, Ro Jaimes, and Shih fu Chang. A conceptual framework for indexing visual information at multiple levels. In *in proceedings of SPIE Internet Imaging 2000*, pages 2–15, 2000.
- [KE06] Christian Küblbeck and Andreas Ernst. Face detection and tracking in video sequences using the modifiedcensus transformation. *Image Vision Comput.*, 24(6):564–572, 2006.
- [KPJ09] P.S. Karthik, C. Pulla, and C.V. Jawahar. Incremental on-line semantic indexing for image retrieval in dynamic databases. In *SLAM 09*, pages 9–16, 2009.
- [KSH⁺04] Yan Ke, Rahul Sukthankar, Larry Huston, Yan Ke, and Rahul Suktankar. Efficient near-duplicate detection and sub-image retrieval. In *In ACM Multimedia*, pages 869–876, 2004.
- [KVA04] Konstantinos Kotis, George A. Vouros, and Jerónimo Padilla Alonso. HCOME: A Tool-Supported Methodology for Engineering Living Ontologies. In Christoph Bussler, Val Tannen, and Irimi Fundulaki,

- editors, *Semantic Web and Databases. Second International Workshop - SWDB 2004*, volume 3372 of *LNCS*, pages 155–166, Berlin Heidelberg, Germany, August 2004. Springer-Verlag.
- [Lon02] Philip M. Long. Minimum majority classification and boosting. In *Eighteenth national conference on Artificial intelligence*, pages 181–186, Menlo Park, CA, USA, 2002. American Association for Artificial Intelligence.
- [LW06] Jia Li and James Z. Wang. Real-time computerized annotation of pictures. In *MULTIMEDIA '06: Proceedings of the 14th annual ACM international conference on Multimedia*, pages 911–920, New York, NY, USA, 2006. ACM.
- [LZLM05] Ying Liu, Dengsheng Zhang, Guojun Lu, and Wei-Ying Ma. Region-based image retrieval with high-level semantic color names. In *MMM '05: Proceedings of the 11th International Multimedia Modelling Conference*, pages 180–187, Washington, DC, USA, 2005. IEEE Computer Society.
- [Mar05] J. Martinez. Mpeg-7 overview (version 10). <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>, Oct. 2005.
- [MS00] Marjo Markkula and Eero Sormunen. End-user searching challenges indexing practices in the digital newspaper photo archive. *Inf. Retr.*, 1(4):259–285, 2000.
- [Nag07] Gábor Nagypál. *Possibly imperfect ontologies for effective information retrieval*. PhD thesis, University of Karlsruhe, 2007.
- [PNNJ05] A. Powell, M. Nilsson, A. Naeve, and P. Johnston. Dublin core meta-data initiative - abstract model, 2005. White Paper.
- [Ram01] Ariel Ortiz Ramirez. Three-tier architecture. *Linux Journal*, page 7, 2001.
- [SBLH06] Nigel Shadbolt, Tim Berners-Lee, and Wendy Hall. The semantic web revisited. *IEEE Intelligent Systems*, 21(3):96–101, 2006.
- [Sch05] Andreas Schmidt. Knowledge Maturing and the Continuity of Context as a Unifying Concept for Knowledge Management and E-Learning. In *Proceedings of I-KNOW '05, Special Track on Integrating Working and Learning*, July/August 2005.

- [SEA⁺02] York Sure, Michael Erdmann, Juergen Angele, Steffen Staab, Rudi Studer, and Dirk Wenke. OntoEdit: Collaborative ontology development for the semantic web. In *Proceedings of the first International Semantic Web Conference 2002*, Sardinia, Italia, June 9-12 2002. Springer, LNCS 2342.
- [SGCD09] William Robson Schwartz, Raghuraman Gopalan, Rama Chellappa, and Larry S. Davis. Robust human detection under occlusion by integrating face and person detectors. In *ICB '09: Proceedings of the Third International Conference on Advances in Biometrics*, pages 970–979, Berlin, Heidelberg, 2009. Springer-Verlag.
- [SH07] Katharina Siorpaes and Martin Hepp. myontology: The marriage of ontology engineering and collective intelligence. In *Bridging the Gap between Semantic Web and Web 2.0 (SemNet 2007)*, pages 127–138, 2007.
- [SH08] Katharina Siorpaes and Martin Hepp. Ontogame: weaving the semantic web by online games. In *ESWC'08: Proceedings of the 5th European semantic web conference on The semantic web*, pages 751–766, Berlin, Heidelberg, 2008. Springer-Verlag.
- [Sha86] Sara Shatford. Analyzing the subject of a picture: A theoretical approach. *Cataloging & Classification Quarterly*, 6:39 – 62, 1986.
- [Smi08] Barry Smith. Ontology (science). In *FOIS*, pages 21–35, 2008.
- [SSSS01] Steffen Staab, Rudi Studer, Hans-Peter Schnurr, and York Sure. Knowledge processes and ontologies. *IEEE Intelligent Systems*, 16(1):26–34, 2001.
- [STMA09] Evaggelos Spyrou, Giorgos Toliás, Phivos Mylonas, and Yannis Avrithis. Concept detection and keyframe extraction using a visual thesaurus. *Multimedia Tools Appl.*, 41(3):337–373, 2009.
- [Tay68] Robert S. Taylor. The process of asking questions. In *American Documentation 13*, pages 391–396, 1968.
- [TCZZ06] Xiaoyang Tan, Songcan Chen, Zhi-Hua Zhou, and Fuyan Zhang. Face recognition from a single image per person: A survey. In *Pattern Recognition*, volume 39, pages 1725–1745, New York, NY, USA, 2006. Elsevier Science Inc.

- [TN07] Tania Tudorache and Natasha Fridman Noy. Collaborative protégé. In *Proceedings of the Workshop on Social and Collaborative Construction of Structured Knowledge (CKC 2007) at the 16th International World Wide Web Conference (WWW2007)*, 2007.
- [VFaC05] David Vallet, Miriam Fernández, and Pablo Castells. An ontology-based information retrieval model. *Lecture Notes in Computer Science : The Semantic Web: Research and Applications*, pages 455–470, 2005.
- [VKV⁺06] Max Völkel, Markus Krötzsch, Denny Vrandečić, Heiko Haller, and Rudi Studer. Semantic Wikipedia. In *Proceedings of the 15th international conference on World Wide Web (WWW'06)*, pages 585–594. ACM Press, 2006.
- [vOTSP07] Jacco van Ossenbruggen, Raphael Troncy, Giorgos Stamou, and Jeff Z. Pan. Image Annotation on the Semantic Web. W3C working draft, W3C, March 2007.
- [Wel99] Welty, C., Lehmann, F., Gruninger, G., and Uschold, M. Ontology: Expert Systems All Over Again? Invited panel at AAAI-99: The National Conference on Artificial Intelligence. Austin, Texas., 1999.
- [WHT09] Lior Wolf, Tal Hassner, and Yaniv Taigman. The one-shot similarity kernel. In *International Conference on Computer Vision (ICCV)*, 2009.
- [WN07a] Andreas Walter and Gabor Nagypal. Imagenotion - methodology, tool support and evaluation. In *GADA/DOA/CoopIS/ODBASE 2007 Confederated International Conferences DOA, CoopIS and ODBASE, Proceedings, LNCS*. Springer, 2007.
- [WN07b] Andreas Walter and Gábor Nagypál. Semantical interoperability with IMAGINATION content using standardized ontologies. In *Workshop on Ontology-Driven Interoperability for Cultural Heritage Objects, Joint DELOS-MultiMatch event*, Tirrenia, Pisa (Italy), February 15 2007.
- [WN07c] Andreas Walter and Gabor Nagypal. The utility of specific relations in ontologies for image archives. In *Proceedings of the third International Conference on Semantic Systems (I-Semantics)*, pages 17–24, 2007.
- [WN08] Andreas Walter and Gabor Nagypal. The combination of techniques for automatic semantic image annotation generation in the imagination application. In *5th European Semantic Web Conference, ESWC*

2008, Tenerife, Canary Islands, Spain, June 1-5, 2008 Proceedings, 2008.

- [WNN08] Andreas Walter, Gabor Nagypal, and Khaled Nagi. Evaluating semantic techniques for the exploration of image archives on the example of the imagenotion system. In *Alexandria Engineering Journal (AEJ)*, Vol. 47, No. 4. ISSN: 1110-0168 Springer, 2008.
- [Yoo09] JungWon Yoon. Effectiveness of query expansion using flickr related tags. In *DCMI '09: Proceedings of the 2009 International Conference on Dublin Core and Metadata Applications*, pages 133–134. Dublin Core Metadata Initiative, 2009.
- [ZBS09] Valentin Zacharias, Simone Braun, and Andreas Schmidt. Social semantic bookmarking with soboleo. In San Murugesan, editor, *Handbook of Research on Web 2.0, 3.0 and X.0: Technologies, Business, and Social Applications*, pages 225–241. IGI Global, 2009.