

Personalisiertes Information-Brokering

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften
der Rheinisch-Westfälischen Technischen Hochschule Aachen
zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Diplom-Informatiker

Achim Nick

aus Boppard am Rhein

Berichter:

Universitätsprofessor Dr. rer. pol. Matthias Jarke
Universitätsprofessor Dr. rer. nat. Volker Wulf

Tag der mündlichen Prüfung: 10. Dezember 2002

Diese Dissertation ist auf den Internetseiten der Hochschulbibliothek online
verfügbar.

Kurzfassung

Information-Brokering ist die Vermittlung zwischen Informationsangeboten von Informationsanbietern und den Informationsbedürfnissen von Informationskonsumenten. Im Zentrum von Information-Brokering steht ein Modell von der Domäne, in der mit Informationen gehandelt wird. Bisherige Arbeiten aus dem Bereich ontologiebasierter Informationssysteme fokussieren auf die Repräsentation von Informationsangeboten in einer zentralen Domänenontologie, während Arbeiten aus dem Bereich adaptiver Informationssysteme sich auf die Personalisierung von Informationen aus mehr oder minder strukturierten Daten konzentrieren.

Diese Arbeit, die sich mit personalisiertem Information-Brokering beschäftigt, bringt beide Themenkreise zusammen und geht der Frage nach, wie das zentrale Modell von der Domäne strukturiert werden muss, damit es den Erfordernissen ontologiebasierter Informationssysteme genügt und gleichzeitig zur effizienten Personalisierung von Inhalten genutzt werden kann.

Ergebnis dieser Arbeit ist die Strukturierung des zentralen Domänenmodells mit Kategorien und Konzepten. Konzepte können, wie in Domänenontologien, zu einer Hierarchie geformt werden und können gleichzeitig beliebige Relationen zu weiteren Konzepten aufbauen. Konzepte repräsentieren die Informationstypen, mit denen in der konkreten Domäne gehandelt wird.

Kategorien können zu Kategoriehierarchien organisiert werden. Kategorien können im Gegensatz zu Konzepten keine Relationen untereinander bilden. Sie dienen der Abbildung von möglichen Interessen im Domänenmodell. Sie können zur Beschreibung von Eigenschaften von Konzepten eingesetzt werden.

Die Ideen dieser Arbeit wurden im Brokering-Werkzeugkasten Broker's Lounge umgesetzt. Broker's Lounge kann benutzt werden, um mit wenig Aufwand Brokeringssysteme in verschiedensten Domänen mit verschiedenen Brokeringmodellen umzusetzen. Zur Demonstration werden am Ende dieser Arbeit drei verschiedene Brokeringssysteme vorgestellt, die auf Basis von Broker's Lounge realisiert wurden.

Broker's Lounge ist domänenunabhängig und kann sich an die Informationsbedürfnisse von Benutzern anpassen. Broker's Lounge vereinigt damit die Vorteile von ontologiebasierten und adaptiven Informationssystemen und ist damit sowohl ein domänenunabhängiges, adaptives Informationssystem, wie es erst für die nächsten Jahre erwartet wurde, als auch ein personalisiertes, ontologiebasiertes Informationssystem, das einen wichtigen Beitrag für die Personalisierung von Inhalten im Semantic Web liefern kann.

Abstract

Information-Brokering deals with the brokering of information that is offered by information providers and desired by so called information consumers. A model of the domain where brokering takes place is central for information brokering. Existing works in the field of ontology-based information systems are focused on the representation of information offers, whilst works from the adaptive information systems community are centered on the personalization of information based on a domain model.

This thesis focuses on personalized Information-Brokering. The thesis brings together the two communities (ontology-based information systems and adaptive information systems) and enters into the question, how the central domain model should be structured, to allow an effective Mapping of information desires to the domain model.

The answer to this question (and the result of the thesis) is, that the domain model should be structured by Categories and Concepts. Concepts can be organized to an hierarchy of concepts and can hold relations to other Concepts. Concepts represent the type of information the broker deals with.

Categories can be organized to category hierarchies. Whilst Concepts can hold relations to other Concepts, Categories cannot build relations to other Categories. Categories are used for the representation of possible user interests. Category hierarchies usually describe characteristics of Concepts.

The ideas of this thesis has been implemented in the brokering toolkit Broker's Lounge. Broker's Lounge can be used to build up a brokering system for a new domain with minimal effort. To demonstrate this effect three applications of Broker's Lounge in three different domains were described.

Broker's Lounge ist domain independent and adaptive to the information desires of its users. Broker's Lounge pools the benefits of ontology-based information systems and adaptive information systems. Broker's Lounge can be classified as an domain independent, adaptive information system that is expected (by the adaptive information system community) to be realized in a couple of years. Broker's Lounge can also be characterized as an ontology-based information system that is user-adaptive. With its strong focus on adaptivity to user interests it contributes to the personalization of information in the Semantic Web.

Danksagung

Diese Arbeit entstand im Rahmen meiner Tätigkeit im Fraunhofer-Institut Angewandte Informationstechnik in Sankt Augustin. Ich bedanke mich bei all meinen Kollegen, die mich in den letzten Jahren meiner Tätigkeit begleitet haben und mit denen ich viele persönliche und fachliche Gespräche führen durfte.

Mein besonderer Dank gilt Prof. Dr. Jarke, der mich zu dieser Arbeit ermutigt und mir die Durchführung dieser Arbeit ermöglicht hat, sowie PD Dr. Volker Wulf, für das lebhaftes Interesse an meiner Arbeit und die Bereitschaft das Koreferat dieser Arbeit zu übernehmen.

Außerdem geht mein Dank an meine früheren Kollegen im Fraunhofer-Institut, Dr. Christoph G. Thomas, Dr. Jürgen Koenemann, Wolfgang Pohl, Ingo Schwab, Alexander Sigel, Thomas Geil und Andreas Brüggenthies, deren Anregungen in den früheren Phasen der Arbeit hilfreich waren.

Die Zusammenarbeit mit der Servicestelle für Forschungsförderung in Bochum (ELFI) hat mir wertvolle Einblicke in die Arbeit von Information-Brokern in der Domäne Forschungsförderung erlaubt. Diese Einblicke haben in dieser Arbeit breiten Niederschlag gefunden.

Zur Schärfung der Begrifflichkeiten in späteren Phasen haben die zahlreichen Diskussionen mit meinen Kollegen Andreas Zimmermann, Andreas Lorenz, Barbara Schmidt-Beltz, Prof. Dr. Reinhard Oppermann, Dr. Andreas Becks, Anke Meurer, Jens Rinne und Marc Spaniol sowie mit den Kollegen vom Lehrstuhl Informatik V der RWTH Aachen, Dr. Ralf Klamma und Christian Seeling beigetragen. Ihnen allen möchte ich an dieser Stelle danken. Nicht unerwähnt lassen möchte ich an dieser Stelle die Mitglieder der Doktoranden-Gruppe des Fraunhofer-Instituts, die mir in den letzten Jahren ein wichtiger Rückhalt waren.

Mein Dank geht auch an Roland Klemke, mit dem ich in letzten Jahren ungezählte Stunden der Diskussion über das gemeinsame Thema Information-Brokering verbracht haben.

Für die liebevolle Unterstützung möchte ich es nicht versäumen, mich bei meinen Eltern zu bedanken. Auch meinem Schwiegervater, der diese Arbeit leider nicht mehr in Händen halten kann, gilt mein Dank. Seine Wertschätzung akademischer Leistungen war für mich eine besondere Motivation.

Meine Liebe und Zuneigung gilt meiner Frau Elisabeth, die mit ihrer Liebe und ihrem Verständnis mich unterstützt hat und mich in schwierigen Phasen immer wieder aufgerichtet hat, sowie meiner Tochter Judith, die mir ein steter Quell der Freude ist. Ich hoffe, sie wird diese Zeilen und diese Arbeit eines Tages mit Interesse lesen.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Problembeschreibung	1
1.2	Forschungsansatz	2
1.3	Aufbau der Arbeit	6
2	Definitionen und Grundbegriffe	9
2.1	Ontologiebasierte Informationssysteme	9
2.1.1	Daten, Informationen und Wissen	9
2.1.2	Wissensmanagement	10
2.1.3	Ontologie	11
2.2	Adaptive Informationssysteme	13
2.2.1	Adaptivität und Adaptierbarkeit	13
2.2.2	Push und Pull	14
2.2.3	Domänenmodell	15
2.2.4	Overlay-Modell	16
2.2.5	Stereotypen	17
2.2.6	Interessen und Präferenzen	17
2.2.7	Feedback	18
2.2.8	Akquisitions- und Lernverfahren	19
2.2.9	Lernen an positiven Beispielen	20
2.3	Sonstige Begriffe	22
2.3.1	Vektorraummodell	22
2.3.2	Facettierungsproblem	24
2.4	Zusammenfassung	25
3	Information-Brokering	27
3.1	Aufgaben- und Rollenmodell	28
3.2	Kategorien und Konzepte	34
3.2.1	Funktion	35
3.2.2	Formale Definition	36
3.2.3	Beispiele	42

3.3	Profil	43
3.4	Dossier	46
3.5	Kategorisierungsprozesse im Information-Brokering	46
4	Adaptive Informationssysteme	49
4.1	Termbasierte adaptive Informationssysteme	51
4.2	Kollaborative Filtersysteme	55
4.3	Domänenmodellbasierte Benutzermodellierung	57
4.4	Lehr-Lernsysteme	58
4.5	Domänenmodellbasierte Informationsfilterungssysteme	60
4.6	Zusammenfassung	65
5	Ontologiebasierte Informationssysteme	71
5.1	Semantic-Web-Systeme	73
5.1.1	SHOE	75
5.1.2	Ontobroker	76
5.1.3	KnowWeb	77
5.1.4	CoMMA	77
5.2	Klassifikationssysteme	78
5.2.1	Subject Gateways	79
5.2.2	Yahoo!	79
5.2.3	FindUR	81
5.2.4	Planet-Onto	81
5.3	Ontologiesprachen	83
5.4	Modellierungstools für Ontologien	87
5.5	Positionierung von Broker's Lounge	89
5.5.1	Personalisierungsansätze	90
5.5.2	Brokeringmodelle	91
6	Information-Brokering-Domänenmodell	95
6.1	Modellierungsmittel in adaptiven und ontologiebasierten Informationssystemen	96
6.2	Basisstruktur des Domänenmodells	98
6.3	Beispieldomäne	108
6.4	Vergleich des IB-Domänenmodells mit OIL	109
7	Profilbildung	115
7.1	Instanzenfilter	117
7.2	Konzeptfilter	119
7.3	Kategorietypenfilter	120
7.4	Kategoriefilter	121

7.5	Zusammenspiel der verschiedenen Filter	122
7.6	Adaptiver Filtersupport	124
7.6.1	Konzeptfiltersupport	127
7.6.2	Kategoriefiltersupport	128
7.6.3	Relevanzfilter	129
7.6.4	Beschleunigung der Lernkurve	129
7.7	Push-Filter	131
8	Broker's Lounge	133
8.1	Positionierung von Broker's Lounge	134
8.2	Umsetzung des Aufgabenmodells	135
8.3	Implementierung des IB-Domänenmodells	140
8.4	Domänenmodelladministration	144
8.4.1	Kategorisierung	149
8.5	Quellenbeobachtung und Webrobots	152
8.6	Quellenbewertung und Kategorisierung	156
8.7	Profilschnittstelle	163
8.7.1	Konzeptfilter	167
8.7.2	Kategorie- und Kategorietypenfilter	168
8.7.3	Adaptiver Kategoriefiltersupport	168
8.7.4	Instanzenfilter	169
8.7.5	Zusammenspiel der verschiedenen Filter	171
8.7.6	Selektionsfilter	173
8.7.7	Push-Dienst	174
8.7.8	Stereotypen	176
8.7.9	Übersichtslisten und Detailansichten	177
8.7.10	Dossierverwaltung	179
8.7.11	Profil-Webschnittstelle	180
8.8	Architektur	180
9	Anwendungsbeispiele	185
9.1	ELFI	186
9.1.1	ELFI-Kontrollfluss und Brokeringmodell	187
9.1.2	Brokeringmodell von ELFI mit Broker's Lounge	191
9.1.3	ELFI-Domänenmodell	191
9.2	MarketMonitor	193
9.2.1	MarketMonitor-Domänenmodell	196
9.3	ScienceLounge	197
9.3.1	ScienceLounge-Domänenmodell	198
9.3.2	Einsatz Erfahrungen	202
9.4	Vergleich der Broker's Lounge-Anwendungen	202

10 Zusammenfassung und Ausblick	209
10.1 Ausblick	214
Literaturverzeichnis	218
A Lebenslauf	241

Abbildungsverzeichnis

1.1	Information-Brokering-Modell	3
2.1	Zusammenhang zwischen Domänen- und Overlay-Modell	16
2.2	Gegenüberstellung eines einfachen und eines facettierten Kategoriensystems	24
3.1	Information-Brokering-Modell	30
3.2	Quellenverwaltungsaufgaben	31
3.3	Domänenverwaltungsaufgaben	32
3.4	Kundenorientierte Aufgaben	33
3.5	Struktureller Unterschied zwischen Kategorien und Konzepten	35
3.6	Beispielstruktur mit Kategorien	38
3.7	Beispielstruktur mit Konzepten	40
3.8	Profil über einem Kategoriegraphen	45
4.1	Brokeringmodell für termbasierte, adaptive Informationsfilterungssysteme	55
4.2	Push-Broker-Modell für „normale“ Tutoringsysteme	59
4.3	Push-Broker-Modell für KBS-Hyperbook	60
4.4	Oyster-Aufgabenverteilung	65
5.1	Aufgabenverteilung in Semantic-Web-Systemen	74
5.2	CoMMA-Brokeringmodell	75
5.3	Brokeringmodell von Subject Gateways	80
5.4	Planet-Onto-Brokeringmodell	82
5.5	Protégé-Ontologie-Editor	88
5.6	OILEd-Editor	89
6.1	Basisstruktur für das IB-Domänenmodell	104
6.2	Ausschnitt aus dem ELFI-Domänenmodell	108
6.3	ELFI-Domänenmodell in OIL	114

7.1	Mengen im IB-Domänenmodell	116
7.2	Domänenmodellfilter	118
7.3	Konkatenation von Filtern	125
8.1	Information-Brokeringprozess in Broker's Lounge	137
8.2	Benutzeradministration in Broker's Lounge	139
8.3	Klassenhierarchie innerhalb des IB-Domänenmodells	140
8.4	Domänenmodellaufgaben	145
8.5	Schnittstelle zur Domänenmodelladministration	147
8.6	Eingabe von Kurzbeschreibung und Gewichtung von Informations- einheiten	149
8.7	Eingabe der Originalquelle	150
8.8	Eingabe der Klassifikation von Informationseinheiten	151
8.9	Eingabe der Relation zu anderen Informationseinheiten	153
8.10	Schnittstelle für die Robotadministration	154
8.11	Quellenübersicht im Quellenadministrator	157
8.12	Dokumentenbrowser – Übersicht	159
8.13	Dokumentenbrowser – Detailansicht mit zugehörigen Treffern . .	161
8.14	Dokumentenbrowser – Detailansicht mit zugehöriger Informations- einheit	162
8.15	Webschnittstelle des Dokumentenbrowsers – Übersichtstabelle . .	164
8.16	Webschnittstelle des Dokumentenbrowsers – Dokumentenansicht .	165
8.17	Profilbildungsschnittstelle	166
8.18	Aktualitätsfilter editieren	170
8.19	Eingabeschnittstelle für Relevanzfilter	172
8.20	Profilbildungsschnittstelle - Profilübersicht	175
8.21	Push-Dienst konfigurieren	176
8.22	Detailansicht einer Informationseinheit	179
8.23	Beispiel-Dossier in der Domäne Forschungsförderung	181
8.24	Profil-Webschnittstelle - Übersichtsliste	182
8.25	Architektur von Broker's Lounge	183
9.1	Kontrollfluss in der ELFI-Servicestelle	188
9.2	Personalisierungsschnittstelle des „alten“ ELFI-Systems	189
9.3	ELFI-Brokeringmodell	190
9.4	Brokeringmodell von ELFI bei der Umsetzung mit Broker's Lounge	192
9.5	Kontrollfluss in MarketMonitor	195
9.6	MarketMonitor-Brokeringmodell	196
9.7	ScienceLounge Brokeringmodell	199
9.8	Differenzsicht im Dokumentenbrowser von ScienceLounge	203

Tabellenverzeichnis

2.1	Featurevektoren	21
2.2	Termvektoren in einem termbasierten Informationssystem	23
2.3	Übersicht über die beschriebenen Begrifflichkeiten	25
4.1	Vergleich der verschiedenen domänenmodellbasierten Informationssysteme	66
5.1	Personalisierung in den verschiedenen ontologiebasierten Informationssystemen	91
5.2	Aufgabenverteilung bei den verschiedenen ontologiebasierten Informationssystemen	92
7.1	Konzeptfilter-Featurevektoren	128
7.2	Kategoriefilter-Featurevektoren	129
8.1	Aufgabenverteilung bei den verschiedenen Systemen	134
9.1	Relationen zwischen Konzepten, Kategorietypen und Attributen in ELFI	194
9.2	Relationen zwischen Konzepten, Kategorietypen und Attributen in MarketMonitor	198
9.3	Relationen zwischen Konzepten, Kategorietypen und Attributen in ScienceLounge	201
9.4	Vergleich der drei Broker's Lounge-Anwendungen	205

Kapitel 1

Einleitung

Das Wachstum des World-Wide Web (WWW) hat dazu geführt, dass Dokumente zu (fast) jedem erdenklichen Thema prinzipiell verfügbar sind. Die praktische Verfügbarkeit scheitert jedoch an der schieren Menge, der zur Verfügung stehenden Dokumente. Dieses Information-Overload-Problem ist eines der Haupthindernisse auf dem Weg zu einer wirklichen Wissensgesellschaft.

1.1 Problembeschreibung

Eine Möglichkeit dem entgegen zu wirken, besteht darin, den Zugriff auf Dokumente an persönliche Gegebenheiten und Interessen anzupassen oder die Dokumente mit Hilfe einer Ontologie zu strukturieren. Diese Ontologie beschreibt die Konzepte, die in der bearbeiteten Domäne von Bedeutung sind und die Dokumente werden mit den Konzepten der Ontologie annotiert. Dies ist Gegenstand des Semantic-Web-Projektes (vgl. [BLHL01]).

Der Ontologiebegriff in der Informatik, auf dem das Semantic-Web-Projekt aufbaut, ist jedoch viel älter und wurde von Wand und Weber (vgl. [WW89]) in die Informatik eingeführt. Der Begriff der „Ontologie“ sollte ursprünglich die wichtigsten Konzepte eines Informationssystems und die Beziehungen der Konzepte untereinander beschreiben. In der Folgezeit wurde der Begriff der „Ontologie“ auch in der Künstlichen Intelligenz verwendet. Prototypisch seien hier [Gru93] und [SBF98] genannt. Der Begriff „Ontologie“ stammt ursprünglich aus der Philosophie. „Ontologie“ ist die „Lehre vom Seienden“.

Während Personalisierungsaspekte in den letzten Jahren vor allem im Bereich Benutzermodellierung und adaptiver Informationssysteme eine große Rolle gespielt haben, wurden Aspekte der Domänenmodellierung im Bereich ontologiebasierter Informationssysteme thematisiert.

Bisher werden adaptive Systeme jeweils für eine Anwendungsdomäne maß-

geschneidert. Dies hat nach Fischer (vgl. [Fis01]) dazu geführt, dass die zusätzlichen Kosten für die Erstellung von adaptiven Systemeigenschaften in keinem Verhältnis zu dem dadurch erzielten Zuwachs an Bedienkomfort stehen. Um die Kosten für adaptive Systemeigenschaften zu senken, erwartet Brusilovsky (vgl. [Bru01]), dass in den nächsten Jahren adaptive Informationssysteme entstehen, die domänenunabhängig und damit nicht an eine bestimmte Domäne gebunden sind. Damit wird die Entwicklung von adaptiven Lehr-Lernsystemen nachvollzogen, bei denen schon seit einigen Jahren domänenunabhängige Systeme existieren. In dieser Arbeit wird versucht, diese Prognose von Brusilovsky zu erfüllen.

Arbeiten zu ontologiebasierten Informationssystemen fokussieren oft stark auf die Akquisition und die Repräsentation von Wissen. Die Bedeutung des persönlichen Zugangs zu Wissen wurde erst in jüngster Zeit erkannt und ist Gegenstand der Forschung (vgl. [Die00]).

1.2 Forschungsansatz

Information-Brokering bringt beide Ansätze zusammen und ist nach [Tim99] eines der innovativsten Geschäftsfelder im Elektronischen Handel im Internet. Information-Brokering vermittelt zwischen Informationsangeboten von Informationsanbietern und Informationsbedürfnissen von Informationskonsumenten (vgl. Abb. 1.1). Auf der einen Seite müssen die verschiedenen Informationsangebote (WebQuellen) beobachtet und auf ein gemeinsames Modell abgebildet werden. Dieses Modell beinhaltet sowohl eine Repräsentation der Quellen und der dort angebotenen Dokumente in einem Quellenmodell als auch eine Repräsentation der in diesen Dokumenten beschriebenen Inhalte in einem Information-Brokering-Domänenmodell¹. Auf der anderen Seite werden die Informationsbedürfnisse der Informationskonsumenten auf Basis des IB-Domänenmodells² befriedigt, indem die Interessen der Konsumenten auf das IB-Domänenmodell abgebildet werden.

In Abbildung 1.1 wird diese Abbildung durch die Einfärbung der verschiedenen Domänenmodellbausteine mit Grautönen symbolisiert. Um deutlich zu machen, dass zwischen dem Quellenmodell, in dem Meta-Informationen über Dokumente (z.B. das Datum der letzten Änderung) verwaltet werden, und dem IB-

¹Ich verwende den Begriff „Information-Brokering-Domänenmodell“ (kurz IB-Domänenmodell), um ihn so gegen den Begriff „Domänenmodell“, der im Kontext adaptiver Informationssysteme benutzt wird abzugrenzen. Der Begriff „Domänenmodell“ wird in Kap. 2.2.3 eingeführt. „Information-Brokering-Domänenmodell“ bezeichnet dagegen das zentrale Modell der Domäne, das in Information-Brokering-Systemen Verwendung findet.

²Wenn ein Information Broker einen kleinen Kundenkreis bedient, kann das IB-Domänenmodell auch ein implizites Modell sein, das nicht explizit in einer Datenstruktur beschrieben ist. Will ein Information Broker einen größeren Kundenkreis bedienen, sollte er ein explizites IB-Domänenmodell vorhalten.

Domänenmodell, in dem qualitätsgesicherte Inhalte enthalten sind, deutlich zu machen, wird das Quellenmodell in Abbildung 1.1 durch Kreise und das IB-Domänenmodell durch Dreiecke repräsentiert. Abbildung 1.1 wird im Verlauf dieser Arbeit immer wieder leicht variiert, um verschiedene Brokeringmodelle und die damit verbundenen Aufgabenverteilungen zu visualisieren.

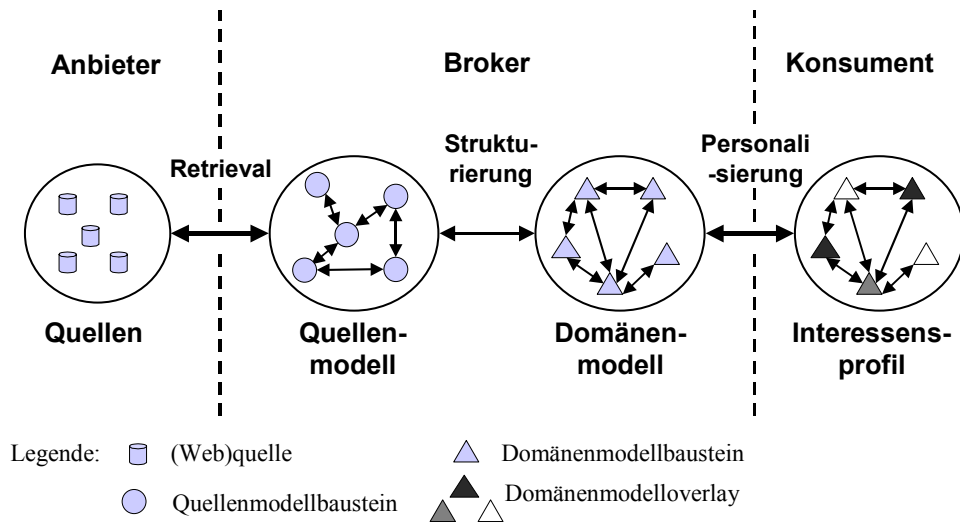


Abbildung 1.1: Information-Brokering-Modell

Es ist wichtig anzumerken, dass die drei verschiedenen Aufgabenbereiche Retrieval, Strukturierung und Personalisierung auf die drei beteiligten Parteien (Anbieter, Broker und Konsument) nahezu beliebig aufgeteilt werden können. Dies kann so weit gehen, dass die Rolle des Brokers ganz entfällt, da alle Aufgabenbereiche von den anderen beiden Parteien ausgeführt werden. Die verschiedenen Aufgaben können jeweils von zwei Parteien gemeinschaftlich ausgeführt werden. Je nach Aufgabenverteilung auf die verschiedenen Parteien, entstehen verschiedene Brokeringmodelle. Brokeringmodelle werden in dieser Arbeit benutzt, um die verschiedenen in dieser Arbeit beschriebenen Systeme unter Brokering-Gesichtspunkten zu klassifizieren. Neben der Unabhängigkeit von einer konkreten Domäne wird auch eine Unabhängigkeit vom gewählten Brokeringmodell angestrebt. Diese Unabhängigkeit vom Brokeringmodell bedingt, dass die verschiedenen Brokeringaufgaben, die in Kapitel 3 beschrieben werden, alle vom System unterstützt werden müssen und dass sie flexibel auf alle Benutzer des Systems verteilt werden können. Dadurch bekommt jeder Benutzer die seinen Aufgaben entsprechenden Rechte und Pflichten. Dies ist ein weiterer Nebenaspekt von Personalisierung.

Bisher fokussierten Arbeiten zu Information-Brokering (vgl. [Wie94], [JP96], [MKIS98], [FDES98] und für einen Überblick [JP98]) stark auf die Beziehung

zwischen Informationsanbietern und dem Broker. Sie decken meist die Prozesse Retrieval und Strukturierung ab. Die Rolle der Konsumenten erschöpft sich in diesen Ansätzen darin, dass sie einen einheitlichen Zugriff auf die Informationen der verschiedenen Anbieter bekommen. Das bedeutet, der Personalisierungsprozess wurde nicht explizit oder nur unzureichend abgebildet. Es ist kritisch anzumerken, dass das Problem des Information-Overload durch die Strukturierung (Abbildung auf das IB-Domänenmodell) gemildert aber nicht behoben werden kann, da das Problem seinen Ursprung in einem Überangebot von Informationen hat. Durch die Strukturierung von Informationen kann das Überangebot minimiert werden, da Informationen zu anderen Themen (die nicht zum IB-Domänenmodell gehören) ausgeschlossen werden können. Jedoch kann auch innerhalb einer eingegrenzten Domäne ein Überangebot an Informationen bestehen. Um diesem Phänomen entgegenzuwirken, sollten die aufbereiteten Informationen auf die Interessen von Konsumenten angepasst werden. Dadurch kann das Information-Overload-Problem weiter vermindert werden.

Forschungsfrage

In dieser Arbeit liegt der Schwerpunkt auf der Personalisierung von Inhalten für Konsumenten. Die angestrebte Personalisierung hat Rückwirkungen auf die Strukturierung des IB-Domänenmodells: Um eine effektive Abbildung von Konsumenteninteressen auf das IB-Domänenmodell zu ermöglichen, sollte das IB-Domänenmodell mögliche Interessen von Konsumenten berücksichtigen. Daraus leitet sich die zentrale Forschungsfrage dieser Arbeit ab:

Wie muss das IB-Domänenmodell strukturiert werden, damit die Interessen von Konsumenten effizient darauf abgebildet werden können.

Das IB-Domänenmodell muss mit einem Lernverfahren integriert werden, mit dem Benutzerinteressen gelernt werden, damit das System die Informationsbedürfnisse von Konsumenten erlernen und ihnen damit eine weitgehende Unterstützung anbieten kann. Das Lernverfahren wiederum muss mit der Benutzerschnittstelle (für Konsumenten) eng verzahnt werden, damit die Konsumenten ihr Interessenprofil jederzeit inspizieren und ggf. verändern können. Dieser Punkt ist wichtig für Information-Brokering, da ein fehlerhaftes Profil zu einem Informationsverlust für den Konsumenten führen kann.

An dieser Stelle wird das Verständnis von Adaptivität und Adaptierbarkeit deutlich, das dieser Arbeit zugrunde liegt: Die Kontrolle über Interessenprofile sollte beim Benutzer verbleiben, indem er sein Profil jederzeit ändern kann. Gleichwohl sollte das System versuchen, adaptiv zu sein, indem es die Interessen des Benutzers versucht zu erlernen und ggf. dem Benutzer Vorschläge zur Änderung des Profils unterbreitet. Die letzte Entscheidung über eine Profiländerung sollte beim Benutzer verbleiben. Dies reflektiert die Forderung von Fischer, der in

[Fis01] fordert, dass die volle Kontrolle beim Benutzer verbleiben muss. Da das System nur Tipps geben sollte und die Kontrolle beim Benutzer verbleibt, ist das System nicht in der Pflicht, ständig Vorschläge zu generieren, sondern kann sich auf die Rolle eines Korrektivs beschränken, das nur dann Vorschläge unterbreitet, wenn es hinreichend sicher ist. Adaptivität und Adaptierbarkeit sowie weitere Grundbegriffe und Definitionen aus den beiden bearbeiteten Gebieten adaptiver und ontologiebasierter Informationssysteme werden in Kapitel 2 eingeführt.

Adaptive Systeme zur Informationsfilterung basieren zu einem großen Teil auf Information-Retrieval-Methoden und sind deswegen textbasiert. Auch kollaborative Filterungssysteme finden dort Verwendung. Erst in jüngster Zeit ist ein deutlicher Trend zu reichhaltiger strukturierten Domänenmodellen erkennbar.

Konkrete Domänenmodelle wurden insbesondere von adaptiven Tutoringsystemen benutzt, die ein Modell des Lernalers aufbauen, das auf dem Domänenmodell basiert. Diese Domänenmodelle beschreiben jedoch bis auf seltene Ausnahmen nicht die Domänen-**inhalte**, die gelehrt werden sollen, sondern bilden die Struktur der Lernmaterialien (Kurse, Lerneinheiten, usw.) in einem Curriculum ab. Die Lerninhalte werden über den Umweg dieser Lernmaterialien implizit modelliert, indem mit einer Lerneinheit jeweils Dokumente assoziiert werden, die die zu lernenden Inhalte beinhalten.

In dieser Arbeit wird hingegen versucht, die **Inhalte** der Domäne, in der Information-Brokering stattfindet, in einem IB-Domänenmodell zu beschreiben. Bei diesem Domänenmodell handelt es sich um eine Ontologie im Sinne von [Gru93], die alle wichtigen Konzepte der beschriebenen Domäne beschreibt. Gleichzeitig wird versucht, die adaptiven Eigenschaften des Systems domänenunabhängig zu gestalten. Um dies zu erreichen, sollen die adaptiven Systemeigenschaften an die Basisbausteine des Domänenmodells gekoppelt werden, weil dann alle Domänen, die durch die Basisbausteine beschrieben werden können, bereits adaptive Systemeigenschaften aufweisen, ohne dass dafür zusätzliche Kosten anfallen.

Klemke fordert in [Kle02], dass Benutzer an ihren Kontext angepasste Informationen erhalten und der momentane Kontext zur Filterung von Informationen benutzt werden kann. Er beschreibt vier Kontextdimensionen. Diese Kontextdimensionen sind

1. der Ort, an dem sich ein Benutzer befindet,
2. die Zeit,
3. die Aufgabe, die ein Benutzer bearbeitet und
4. das Interesse des Benutzers an Informationen.

Die vorliegende Arbeit beschäftigt sich mit der Modellierung von Benutzerinteressen und der Frage, mit welchen Modellierungsmitteln die Modellierung von Benutzerinteressen geschehen kann. Diese Modellierungsmittel können dann auch zur Modellierung der übrigen Kontextdimensionen benutzt werden.

1.3 Aufbau der Arbeit

Die vorliegende Arbeit speist sich aus den beiden Themenkreisen adaptiver und ontologiebasierter Informationssysteme. Beide Themenkreise verwenden zum Teil eine ähnliche Terminologie. Um Missverständnisse entgegenzuwirken, führe ich in **Kapitel 2** die verschiedenen Begriffe ein, die in dieser Arbeit verwendet werden und aus den beiden Themenkreisen stammen.

Im Anschluss daran werde ich in **Kapitel 3** ein Aufgaben- und Rollenmodell für Information-Brokering einführen. Im Zentrum des Information-Brokeringmodells steht das sog. Information-Brokering-Domänenmodell (IB-Domänenmodell). Dieses IB-Domänenmodell ist eine Ontologie im Sinne ontologiebasierter Informationssysteme und kann gleichzeitig als Quelle für die Adaption an Benutzerinteressen benutzt werden.

Um dies bewerkstelligen zu können, werden in den folgenden Kapiteln adaptive und ontologiebasierte Informationssysteme auf ihre Modellierungsmittel untersucht. Adaptive Informationssysteme werden in **Kapitel 4** nach domänenunabhängigen Beschreibungsmitteln durchsucht. Ontologiebasierte Informationssysteme, die per Definition domänenunabhängig sind, werden im Gegenzug in **Kapitel 5** auf Personalisierungsaspekte untersucht. Alle Systeme werden, soweit möglich, auf ihr zugrunde liegendes Brokeringmodell untersucht. Adaptive Informationssysteme, die ein explizites Domänenmodell haben, benutzen fast alle *Kategoriehierarchien* zur Abbildung von Benutzerinteressen. Ontologiebasierte Informationssysteme benutzen *Konzepte*³, die zu einer Konzepthierarchie arrangiert und zueinander in Beziehung stehen können, um ein Modell von den Inhalten der Domäne zu repräsentieren. Ontologiebasierte Systeme, die eine Personalisierung der Inhalte anstreben, sind fast alle rein adaptierbare Systeme und bilden Interessenprofile als Overlay-Modelle⁴ über die Konzepthierarchie ab.

Als Konsequenz der Untersuchung adaptiver und ontologiebasierter Informationssysteme ergibt sich die Beschreibung des IB-Domänenmodells in **Kapitel 6**.

³Die Unterscheidung zwischen Kategorien und Konzepten ist essentiell für diese Arbeit. Das Verständnis von Kategorien und Konzepten, das dieser Arbeit zugrunde liegt, wird in **Kapitel 3.2** beschrieben.

⁴Overlay-Modelle werden in der Benutzermodellierung häufig benutzt, um Interessen- oder Benutzerprofile abzubilden. Overlay-Modelle bilden dabei in der Regel eine darunterliegende Struktur nach. In **Kapitel 2.2.4** werden Overlay-Modelle beschrieben.

Das IB-Domänenmodell ist zum einen eine Domänenontologie im Sinne ontologiebasierter Informationssysteme und ist aus verschiedenen Konzepten aufgebaut, die zueinander in Beziehung stehen können. Zu den Konzepten können jeweils Instanzen, sog. Informationseinheiten, gebildet werden, die einem Konzept zugeordnet werden und die Relationen zwischen Konzepten mit Leben füllen. Zu den verschiedenen Konzepten können Kategoriehierarchien definiert werden. Die Informationseinheiten eines Konzeptes können mit Kategorien aus den Kategoriehierarchien klassifiziert werden. Die Kategorien haben attributiven Charakter und können zur Gruppierung und Filterung von Informationseinheiten genutzt werden.

Dies wird bei der Profilbildung ausgenutzt, die in **Kapitel 7** beschrieben wird. Die Verbindungen zwischen Kategorien bzw. Kategoriehierarchien, Informationseinheiten und Konzepten werden benutzt, um Filter für die Inhalte des IB-Domänenmodells zu definieren. Kategorien, bzw. Kategoriefilter z.B. nutzen die Beziehung zwischen Informationseinheiten und Kategorien aus, um alle Informationseinheiten herauszufiltern, die mit einer bestimmten Kategorie klassifiziert wurden. In der Domäne Forschungsförderung kann ein Kategoriefilter z.B. benutzt werden, um aus allen verfügbaren Förderprogrammen die Förderprogramme herauszufiltern, die mit der Kategorie „Informatik“ in der Forschungsgebiete-hierarchie klassifiziert wurden. Eine Menge von Filtern, die miteinander kombiniert werden können, bilden zusammen ein Interessenprofil.

Kategoriefilter können mit einer adaptiven Tippkomponente ausgestattet werden. Als Lernverfahren für das Erlernen von Interessen wird das Lernen an positiven Beispielen benutzt (vgl. **Kap. 2.2.9**). Die adaptive Tippkomponente beobachtet den Benutzer und schlägt ihm bei signifikanten Abweichungen eine entsprechende Änderung seines Profils vor. Für jede Kategorie wird dabei ein Relevanzwert errechnet. Dieser Relevanzwert wird benutzt, um die Relevanz einzelner Informationseinheiten bewerten zu können, die sich aus der normalisierten Summe aller ihrer Kategorien errechnet. Mit Hilfe dieser Relevanzberechnung kann dem Benutzer ein adaptiver Relevanzfilter angeboten werden, der für alle Informationseinheiten einem Relevanzwert berechnet, und diese Werte zur Filterung der relevantesten Informationseinheiten benutzt.

Während in **Kapitel 6** das IB-Domänenmodell und in **Kapitel 7** die darauf basierende Filterung und Profilbildung beschrieben wird, enthält **Kapitel 8** eine Beschreibung der Funktionsweise von Broker's Lounge, in der das IB-Domänenmodell und die Profilbildung umgesetzt wurden. Broker's Lounge ist als Werkzeugkasten konzipiert, mit dem Brokeringsysteme für beliebige Domänen und beliebige Brokeringmodelle in kürzester Zeit erstellt werden können. In **Kapitel 8** werden alle Schnittstellen der Broker's Lounge zur Unterstützung der verschiedenen Brokeringaufgaben beschrieben. Besonderes Gewicht liegt dabei auf der Beschreibung der Personalisierungsschnittstelle, in der die verschiedenen in **Kapitel 7** beschriebenen Filter und die adaptive Tippkomponente umgesetzt wurden.

Mit dieser Schnittstelle können Interessenprofile erstellt und verwaltet werden.

In **Kapitel 9** werden Anwendungsbeispiele von Broker's Lounge beschrieben. Dort wird z.B. der Einsatz von Broker's Lounge in der Domäne Forschungsförderung beschrieben. Zur Verdeutlichung von Sachverhalten in dieser Arbeiten werde ich Beispiele aus dieser Domäne heranziehen. Schließlich werde ich in **Kapitel 10** abschließend die in dieser Arbeit gewonnen Erkenntnisse zusammenfassen, diskutieren und einen Ausblick auf weitere geplante Arbeiten geben, die auf den Ergebnissen dieser Arbeit aufbauen.

Kapitel 2

Definitionen und Grundbegriffe

In diesem Kapitel werden grundlegende Begriffe und Unterscheidungen eingeführt, die in dieser Arbeit von Bedeutung sind. Diese Begriffe und Unterscheidungen sind größtenteils aus den Bereichen ontologiebasierte Informationssysteme und adaptiver Informationssysteme entlehnt. Außerdem werden weitere Begrifflichkeiten eingeführt, die keinem der beiden Bereiche zugeordnet werden können. Um die Herkunft der Begrifflichkeiten deutlich zu machen, ist dieses Kapitel in die Bereiche ontologiebasierte (Kap. 2.1) und adaptive Informationssysteme (Kap. 2.2) sowie sonstige Begrifflichkeiten (Kap. 2.3) eingeteilt. Am Ende dieses Kapitels werden die Begrifflichkeiten anhand ihrer Herkunft zusammengefasst (Kap. 2.4).

2.1 Ontologiebasierte Informationssysteme

In diesem Abschnitt werden Grundbegriffe, die aus dem Bereich ontologiebasierter Informationssysteme stammen erläutert. Von besonderer Bedeutung ist dabei der Begriff der „Ontologie“, der in Kapitel 2.1.3 eingeführt wird.

2.1.1 Daten, Informationen und Wissen

Davenport und Prusack (vgl. [DP98]) grenzen Daten und Informationen voneinander ab, indem sie fordern, dass Informationen „mit Bedeutung und Zweck versehene Daten seien“. Sie fassen Informationen als Nachrichten auf, über deren Wert der Empfänger der Nachricht entscheidet. Eine empfangene Nachricht kann als Information aufgefasst werden, wenn sie für den Empfänger hilfreich ist. Ist dies nicht der Fall handelt es sich lediglich um Daten, obwohl dies vom Absender der Information nicht intendiert wurde.

Wissen lässt sich wiederum von Informationen unterscheiden, indem man wie Davenport und Prusack fordert, dass Wissen wesentlich mehr umfasst als Daten und Informationen. Sie fordern in ihrer Arbeitsdefinition unter anderem,

- dass Wissen zur Eingliederung und Bewertung von neuen Erfahrungen benutzt werden kann,
- dass Wissen in den Köpfen der Wissensträger entsteht und dort angewendet wird,
- und dass Wissen in vielen Organisationen nicht nur in Dokumenten und Speichern enthalten ist, sondern auch allmählich in organisatorische Routinen, Prozesse, Praktiken und Normen eingebettet wird.

Nach [ABH⁺98] und [AL99] kann Wissen nicht direkt zwischen verschiedenen Personen ausgetauscht werden, sondern muss zuvor in Informationen umgewandelt werden. Diese Informationen werden mittels eines Mediums als Daten zwischen Sender und Empfänger ausgetauscht. Damit der Empfänger der Information seinerseits in den Besitz des Wissens kommt, muss er aus den Informationen seinerseits Wissen erzeugen. Wissen entsteht damit im Kopf des Betrachters (vgl. [LR00]).

Betrachtet man den Wissensbegriff im Kontext von Knowledge-Engineering, so können darunter sicherlich gesicherte Fakten und Regeln verstanden werden. Das explizite Wissen wird im Knowledge-Engineering in der Regel in Ontologien (vgl. Kap. 2.1.3) beschrieben.

Im Rahmen von Information-Brokering und des Überangebots an Informationen z.B. im World-Wide Web, erfährt die Unterscheidung von Daten, Informationen und Wissen eine leicht andere Bedeutung: In vielen verfügbaren Dokumenten im World-Wide Web sind Informationen enthalten. Die verfügbaren Dokumente können auch als Nachricht des Autors an den Leser aufgefasst werden. Ist ein Dokument für einen Leser nicht hilfreich, so handelt es sich bei diesem Dokument für diesen Leser um reine Daten. Aufgabe von Information-Brokern ist es, seine Klienten mit „sinnvollen“ Informationen zu versorgen. Information-Broker benötigen dazu Wissen über die Informationsbedürfnisse ihrer Klienten. Dies muss auf die in den Dokumenten enthaltenen Informationen abgebildet werden.

Wenn eine Information den Empfänger in die Lage versetzt, ein konkretes Problem zu lösen, handelt es sich bei dieser Information um Wissen. Wissen kann damit als handlungsrelevante Information beschrieben werden.

2.1.2 Wissensmanagement

Wissensmanagement ist von zunehmender Bedeutung, da Wissen als Produktionsfaktor erkannt wurde (vgl. [NT97], [DP98], [AL99]). Wissensmanagement kann

als der strukturierte und organisierte Prozess des Erhebens, Organisierens und Kommunizierens von Wissen verstanden werden, um sie anderen Personen (Mitarbeitern) zugänglich zu machen, damit diese daraus einen Nutzen ziehen können. Im Wissensmanagement spielt Filterung eine wichtige Rolle, um die wirklich relevanten Informationen zu finden.

Benjamins beschreibt (vgl. [BFGP98a]) Wissensmanagement ebenfalls als einen Prozess, der aus den folgenden vier Schritten besteht:

1. Wissenssammlung (Knowledge gathering),
2. Wissensorganisation und -strukturierung (Knowledge organization and structuring),
3. Wissenspflege (Knowledge refinement) und
4. Wissensverteilung (Knowledge distribution)

Bei der Einführung von Wissensmanagement im Unternehmen muss man neben technischen auch soziale Aspekte berücksichtigen. Davenport und Prusack argumentieren in [DP98], dass technisches Wissensmanagement höchstens ein Drittel des Aufwands ausmachen sollte. Ein großes Problem im Wissensmanagement ist, dass nur formalisiertes Wissen auch wiederverwendet werden kann. Das formalisierte Wissen wird in einer Ontologie beschrieben. Verstecktes Wissen (tacit knowledge) kann nur schwer formalisiert werden.

Klamma beschreibt in [Kla00] das Management von Fehlerwissen und dessen Formalisierung. Das Management von Fehlerwissen basiert nach Klamma auf der Überwindung von Barrieren und der Eskalation von Verbesserungsprozessen innerhalb der Unternehmenshierarchie.

2.1.3 Ontologie

Der Begriff *Ontologie* stammt ursprünglich aus der Philosophie. Ontologie ist die „Lehre vom Seienden“. Eine Ontologie beschreibt in diesem Sinne die Dinge, die existieren. Im Laufe der Geschichte der Philosophie war der Begriff der Ontologie verschiedenen Wandlungen unterworfen, so dass heute eine eindeutige Definition des Begriffs Ontologie (in der Philosophie) nicht mehr existiert.

Der Begriff wurde Ende der 80er Jahre von Wand und Weber (vgl. [WW89]) in die Informatik eingeführt. Der Begriff der „Ontologie“ sollte ursprünglich die wichtigsten Konzepte eines Informationssystems und die Beziehungen der Konzepte untereinander beschreiben. In der Folgezeit wurde der Begriff der „Ontologie“ auch in der Künstlichen Intelligenz verwendet.

Im Rahmen des Wissensmanagements beschreiben Ontologien eine gemeinsame Wahrnehmung (Konzeptualisierung) der Welt oder einer bestimmten Domäne.

Obgleich es viele Definitionen von Ontologien im Wissensmanagement gibt, ist die Definition von Gruber die mit der größten Akzeptanz:

An ontology is a formal, explicit specification of a shared conceptualisation. [Gru93]

Diese Definition wird von [Fen01] aufgenommen. Nach Fensel kann eine Ontologie als „explizit“ aufgefasst werden, weil die benutzten Konzepte *explizit* beschrieben werden müssen. Ontologien sind „formal“, weil sie maschinenlesbar sein sollen. „Konzeptualisierung“ meint den Aufbau eines abstrakten Modells der relevanten Phänomene der Welt. Damit eine Ontologie sinnvoll eingesetzt werden kann, muss die gewählte Konzeptualisierung von möglichst allen Personen geteilt werden, die mit der Ontologie arbeiten sollen. Dies wird durch den Terminus „shared“ in Grubers Definition beleuchtet. In dieser Arbeit schließe ich mich diesem Verständnis von *Ontologie* weitestgehend an.

Der Begriff der „Ontologie“ ist eng verwandt mit dem Begriff eines „Terminologiesystems“, der in [von99] eingeführt wurde. Sowohl eine Ontologie als auch ein Terminologiesystem bieten einen Rahmen zur Bezeichnung von Gegenständen (der realen Welt).

In [SBF98] unterscheiden Studer, Benjamins und Fensel

1. Domänen-,
2. Applikations- und
3. Repräsentationsontologien, sowie
4. generische Ontologien.

Die verschiedenen Ontologietypen werden zu verschiedenen Zwecken benutzt. Generische Ontologien versuchen, generische Aspekte wie z.B. Zeit, die in vielen anderen Ontologien benutzt werden, zu beschreiben. Domänenontologien versuchen eine konkrete Domäne zu beschreiben und nutzen dazu ggf. auch generische Ontologien, die sie für die konkrete Domäne verfeinern. Repräsentationsontologien beschreiben die Modellierungsmittel, mit denen andere Ontologien, z.B. Domänenontologien, beschrieben werden können. Applikationsontologien beschreiben, wie konkrete Applikationen mit Domänenontologien arbeiten.

Die Nutzung verschiedener Ontologien zur Beschreibung verschiedener Aspekte wurde auch in CommonKADS (vgl. [SWdH⁺95]) zur Beschreibung des sog. Expertise-Modells benutzt. Dort wurde zwischen einer Domänenschicht zur Beschreibung der Domäne, einer Inferenzschicht, die beschreibt, welche Inferenzen ausgeführt werden können und einer Aufgabenebene unterschieden, die zur Beschreibung der auszuführenden Aufgaben benutzt wird.

Im Information-Brokering sind insbesondere Domänenontologien von großer Bedeutung, da ein Modell von der Domänen zentral für Information-Brokering ist, wie ich in Kapitel 3.1 erläutern werde. Das in dieser Arbeit erarbeitete IB-Domänenmodell kann auch als Applikationsontologie verstanden werden, weil es zur automatischen Interessenmodellierung benutzt werden kann. Um unabhängig von einer bestimmten Domäne zu sein, wurde für das IB-Domänenmodell in Broker's Lounge eine Repräsentationsontologie eingeführt, mit deren Hilfe IB-Domänenmodelle erstellt werden können. Diese Vorgehensweise entspricht der Frame-Ontologie in Ontolingua (vgl. [Gru93]).

2.2 Adaptive Informationssysteme

In diesem Abschnitt werden Begrifflichkeiten eingeführt, die aus dem Bereich adaptiver Informationssysteme stammen. Von besonderer Bedeutung ist dabei der Begriff des „Domänenmodells“ (vgl. Kap. 2.2.3), der auf den ersten Blick dem Begriff der „Ontologie“ aus dem Bereich ontologiebasierter Informationssysteme (vgl. 2.1.3) sehr ähnlich erscheint. Der Begriff „Domänenmodell“ im Kontext adaptiver Informationssysteme meint jedoch die Beschreibung des Adaptionsziels.

2.2.1 Adaptivität und Adaptierbarkeit

Die grundlegende Unterscheidung von adaptiven und adaptierbaren Systemen (vgl. [Opp94]) beschreibt, wer in der Interaktion zwischen dem System und dem Benutzer die Initiative ergreift. Bei adaptiven Systemen geht die Initiative vom System aus. Das System passt sich automatisch an den Benutzer an. Adaptierbare Systeme geben Benutzern die Möglichkeit, das System an ihre Bedürfnisse anzupassen. Adaptive Systeme haben grundsätzlich das Problem, dass sie den Benutzer und seine Bedürfnisse nicht kennen, sondern nur ein Modell von ihm haben. Adaptiven Systemen bleibt aus diesem Grund nur die Möglichkeit, den Benutzer zu fragen oder ihre Unsicherheit im Benutzermodell zu berücksichtigen. Adaptive Systeme sollten eine gute Balance zwischen adaptiven und adaptierbaren Systemeigenschaften haben, damit Benutzer auf der einen Seite nicht vom System bevormundet werden können und auf der anderen Seite die Informationsbedürfnisse befriedigt werden können, die der Benutzer nicht kennt oder nicht beschreiben kann (vgl. [PN99]).

Kuhlen (vgl. [Kuh95]) beschreibt die Diskrepanz zwischen subjektiven und objektiven Informationsbedarfen. Objektive Informationsbedarfe sind Informationsbedarfe, die ein Benutzer hat. Subjektive Informationsbedarfe sind die Informationsbedarfe, die ein Benutzer meint zu haben. Oftmals besteht nach Kuhlen

eine Lücke zwischen den objektiven und den subjektiven Informationsbedarfen. Adaptive Systeme sind geeignet, diese Lücke zwischen objektiven und subjektiven Informationsbedarfen von Benutzern aufzudecken.

Ein zweiter wichtiger Aspekt im Spannungsfeld von Adaptivität und Adaptierbarkeit ist die Kontrolle über das Benutzer- oder Interessenprofil. In adaptierbaren Systemen liegt die volle Kontrolle beim Benutzer, während in rein adaptiven Systemen die Kontrolle über das Profil allein beim System liegt. In Systemen, die sowohl adaptiv als auch adaptierbar sind, kann die Kontrolle zwischen dem Benutzer und dem System aufgeteilt werden. Fischer fordert in [Fis01], dass die volle Kontrolle beim Benutzer verbleibt und der Benutzer sein Profil jederzeit einsehen und ändern kann. Die Kontrolle sollte so weit gehen, dass ein Benutzer sämtliche adaptiven Systemeigenschaften abschalten kann. In dieser Arbeit schließe ich mich dieser Sicht von Adaptivität und Adaptierbarkeit an. Die Kontrolle des Profils durch den Benutzer wird dadurch bewerkstelligt, dass das System das Profil des Benutzers nur mit dessen Einverständnis ändern kann und die adaptiven Eigenschaften jederzeit abschaltbar sind.

2.2.2 Push und Pull

Informationssysteme können in Push- und Pullsysteme unterschieden werden (vgl. [Fis01]). Bei Pullsystemen werden die Benutzer selbst aktiv, um ihr Informationsbedürfnis zu stillen. Bei Pushsystemen hingegen wird das System aktiv und sendet Informationen an Benutzer.

Adaptive Pushsysteme haben gegenüber nicht adaptiven Pushsystemen den Vorteil, dass sie Informationen angepasst an die Interessen der jeweiligen Benutzer versenden können. So wird verhindert, dass Benutzer mit Informationen konfrontiert werden, die für sie nicht relevant sind. Zusätzlich können adaptive Systeme den Zeitpunkt einer Benutzerinformation an dessen Interessen und Präferenzen anpassen. So können z.B. hoch relevante Informationen sofort an Benutzer weitergeleitet werden, während alle anderen, relevanten Informationen zum verabredeten Zeitpunkt an den Benutzer weitergeleitet werden. Ein rein adaptives Pushsystem ist nicht möglich, da in Pushsystemen üblicherweise kein Feedback (siehe oben, Kap. 2.2.7) des Benutzers anfällt, das zur Anpassung an den Benutzer benötigt wird. Für den Benutzer haben Pushsysteme den Vorteil, dass sie aktuelle Informationen bekommen, ohne sich unmittelbar darum bemühen zu müssen.

Der oben beschriebene relative Vorteil von adaptiven Pushsystemen gegenüber nicht-adaptiven Push-Systemen ist für adaptive Pullsysteme gegenüber nicht adaptiven Pullsystemen weniger virulent, da die Information des Benutzers eine unmittelbare Reaktion auf die Anforderung durch den Benutzer ist. Der Vorteil von adaptiven gegenüber nicht adaptiven Pullsystemen liegt in der im vorhergegangenen Abschnitt (Kap. 2.2.1) beschriebenen Möglichkeit, Diskrepanzen zwi-

schen dem objektiven und dem subjektiven Informationsbedarfen des Benutzers aufzudecken.

In dieser Arbeit wird eine optimale Kombination von adaptiven Push- und Pull-Ansätzen angestrebt.

2.2.3 Domänenmodell

Im Kontext von adaptiven Systemen beschreibt ein Domänenmodell das Adaptionsziel, auf das adaptiert werden soll. In vielen Fällen handelt es sich dabei **nicht direkt** um ein Modell von der Domäne im Sinne einer Domänenontologie (siehe unten, Kap. 2.1.3), sondern um ein Modell von Beschreibungen der (Inhalts-)Domäne. Adaptive Lehr-Lernsysteme (vgl. Kap. 4.4) z.B. verwenden bis auf wenige Ausnahmen ein Curriculum als Domänenmodell. Das Curriculum wiederum kann in Lektionen, Sektionen, Lerneinheiten usw. unterteilt werden. Mit den verschiedenen Bestandteilen des Curriculums wiederum sind Texte, Grafiken oder ähnliches assoziiert, die die Domäne beschreiben, die gelehrt werden soll. Es handelt sich somit bei einem Curriculum um ein indirektes Domänenmodell, das die Domäne nur indirekt über assoziierte Dokumente oder ähnliches beschreibt. Die Interessen oder der Kenntnisstand eines Benutzers im Fall von adaptiven Lehr-Lernsystemen werden mit einem Overlay-Modell (siehe unten, Kap. 2.2.4) abgebildet. Eine tiefe Beschreibung von Curricula zum Einsatz in adaptiven Lehr-Lernsystemen findet sich in [Spe98].

Das Anliegen dieser Arbeit ist es, Domänenmodelle im Sinne adaptiver Informationssysteme und Domänenontologien im Sinne von ontologiebasierten Informationssystemen derart zusammen zu bringen, dass die Domäne durch eine Domänenontologie beschrieben wird und gleichzeitig als Grundlage für die Adaption von Interessen benutzt werden kann. Dieses Modell, dass sowohl als Domänenontologie als auch als Domänenmodell im Sinne adaptiver Informationssysteme verstanden werden kann, wird *IB¹-Domänenmodell* genannt. In Kapitel 6 werden IB-Domänenmodelle beschrieben und weiter gegen herkömmliche Domänenmodelle im Sinne adaptiver Systeme und gegen herkömmliche Domänenontologien abgegrenzt.

Ein Domänenmodell kann aus struktureller Sicht als einfacher Graph aufgefasst werden. Abb. 2.1 zeigt auf der linken Seite ein einfaches Domänenmodell. Die Knoten des Graphen könnten (um im oben aufgeführten Beispiel zu bleiben) z.B. für einzelne Lerneinheiten stehen. Die Relationen zwischen den Knoten beschreiben die Übergänge zwischen den Lerneinheiten.

¹IB steht für **I**nformation-**B**rokering

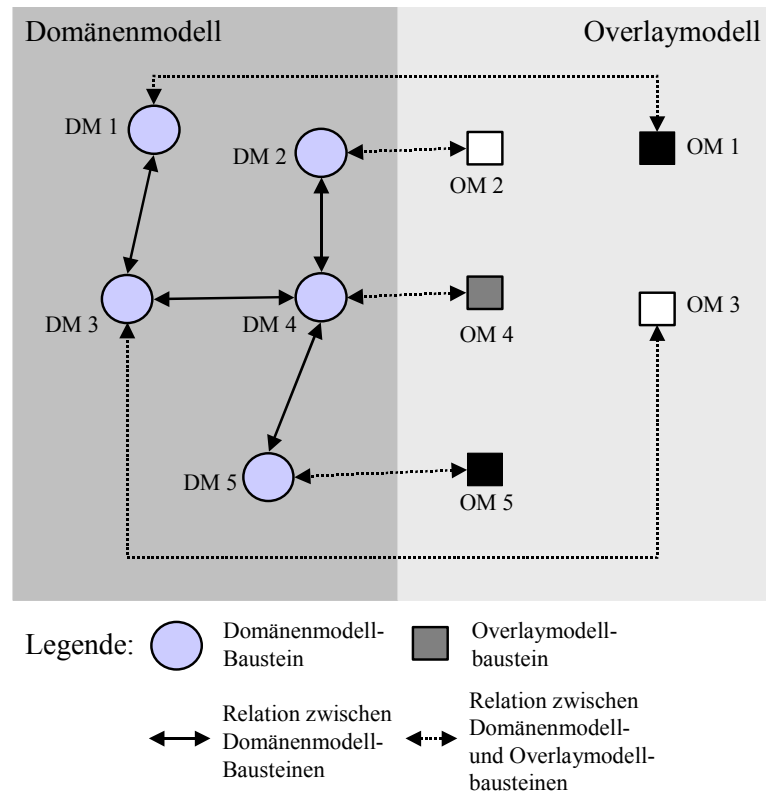


Abbildung 2.1: Zusammenhang zwischen Domänen- und Overlay-Modell

2.2.4 Overlay-Modell

Ein Overlay-Modell ist ein Modell, dass ein darunterliegendes Modell, z.B. ein Domänenmodell, in seiner Struktur nachbildet. Für jeden Bestandteil des darunterliegenden Modells enthält das Overlay-Modell eine Bewertung und eine Referenz auf diesen Bestandteil. Overlay-Modelle werden in der Benutzermodellierung verwendet, um Benutzermodelle aufzubauen. Die adaptiven Informationssysteme, die in dieser Arbeit beschrieben werden (vgl. Kap. 4), benutzen fast alle Overlay-Modelle über ein mehr oder minder strukturiertes Domänenmodell, um die Interessen von Benutzern abzubilden. Sie verwenden fast alle die Architektur für adaptive Informationssysteme, die in [KK96] beschrieben wurde und Interessenprofile als Overlay-Modell über einem zentralen Domänenmodell vorschlägt.

In dieser Arbeit wird diese Architektur aufgegriffen und Interessenprofile werden als Overlay-Modell über das Domänenmodell definiert. Der Zusammenhang zwischen Domänen- und Overlay-Modell wird in Abb. 2.1 verdeutlicht. Für jeden Knoten des Domänenmodells enthält das Overlay-Modell einen eigenen Knoten, der eine Relation zu diesem Knoten aufweist. Der Knoten des Overlay-Modells

enthält jeweils eine Interpretation des Domänenmodell-Knotens. Diese Interpretation enthält für adaptive Lehr-Lernsysteme z.B. den Kenntnisstand des Lernalters über die entsprechende Lerneinheit. Für adaptive Informationssysteme enthält die Interpretation das Interesse des Benutzers an dem entsprechenden Bestandteil des Domänenmodells. Die unterschiedlichen Interpretationen für die Domänenmodellbestandteile sind in Abb. durch die Einfärbung der Overlay-Modellbausteine visualisiert. Die Interpretationen des Overlay-Modells können genutzt werden, um eine Sicht auf das Domänenmodell zu generieren.

2.2.5 Stereotypen

Ein Grundproblem von adaptiven Systemen ist, dass sie eine gewisse Grundmenge an Wissen über den Benutzer brauchen, bevor sie effektiv adaptieren können. Dieses Problem ist besonders für neue Benutzer virulent, da hier kein Wissen über den Benutzer verfügbar ist. Einen Lösungsansatz der Benutzermodellierung stellt die Anwendung von Stereotypen dar. Dabei wird versucht, Benutzer bestimmten Benutzerklassen (Stereotypen) zuzuordnen. Auf Basis dieser Zuordnung wird das Benutzermodell des neuen Benutzers mit für diese Benutzerklasse typischen Werten initialisiert. Frühe adaptive Systeme wie Grundy (vgl. [Ric79]) basieren auf diesem Ansatz. Dort werden für jeden Stereotypen sog. Trigger definiert, die festlegen, wann ein Stereotyp aktiviert werden soll. Die Aktivierung eines Stereotypen bedeutet, dass das Profil des Benutzers mit den Werten des Stereotypen belegt wird. Die Vorbelegung des Profils kann sich auf das ganze Profil oder auf Teile des Profils beziehen. In neueren Systemen ist man von einem rein auf Stereotypen basierenden Ansatz abgekommen und nutzt Stereotypen hauptsächlich zur Initialisierung von leeren Benutzermodellen (vgl. [Bru01]). Das System arbeitet dann mit dem initialisierten Benutzermodell weiter.

Wenn Stereotypen zur Initialisierung benutzt werden, verbleibt das Problem, auf welcher Basis einem Benutzer ein bestimmter Stereotyp zugeordnet werden kann.

2.2.6 Interessen und Präferenzen

Interessen werden in adaptiven Informationssystemen benutzt, um Informationsbedürfnisse zu beschreiben. Dabei können Interessen zur Filterung relevanter Informationen benutzt werden. In vielen Fällen handelt es sich bei adaptiven Informationssystemen um *interessenmodellierende* und nicht um *benutzermodellierende* Systeme. Benutzermodellierende Systeme versuchen möglichst alle Aspekte eines Benutzers zu modellieren, während interessenmodellierende Systeme sich auf die Modellierung von Interessen beschränken. Personalisiertes Informations-Brokering beschränkt sich auf die Modellierung von Interessen, um Informations-

bedürfnisse effektiv stillen zu können. Interessen können nicht in sich selbst beschrieben werden, sondern nur in Relation zu einem Modell. Bei diesem Modell kann es sich z.B. um einen einfachen Termvektor oder um ein stark strukturiertes Domänenmodell (vgl. Kap. 2.2.3) handeln. Aus diesem Grund werden Interessenprofile oft als Overlay-Modelle der eigentlichen Inhalte oder des zu Grunde liegenden Domänenmodells verstanden.

Interessen können von Präferenzen unterschieden werden. Präferenzen beschreiben Eigenschaften des Benutzers, die nicht direkt in Beziehung zu den vermittelten Inhalten des Systems stehen, sondern die Interaktion des Benutzers mit dem System betreffen. Ein Beispiel für Präferenzen sind Darstellungsoptionen für die Benutzerschnittstelle oder die Zeitabstände, in denen ein Benutzer Informationen über den Push-Dienst eines Systems zu bekommen wünscht.

2.2.7 Feedback

Eine zentrale Frage für adaptive Systeme ist, wie sie Wissen über den Benutzer akquirieren (vgl. [Chi93]). Damit verbunden ist die Frage, wie valide dieses akquirierte Wissen ist. In vielen Fällen wird Feedback vom Benutzer benutzt, um das Benutzermodell zu aktualisieren. Es kann zwischen explizitem und implizitem Feedback unterschieden werden. Explizites Feedback wird vom System explizit angefordert oder der Benutzer ist gehalten, Feedback zu geben. Implizites Feedback wird durch die Beobachtung des Benutzers gewonnen. Dabei werden Aktionen des Benutzers interpretiert. Da es sich dabei um indirekt erhobene Daten handelt, ist implizites Feedback nicht so valide wie explizites Feedback. Im Gegensatz zu explizitem Feedback werden Benutzer bei implizitem Feedback nicht in ihrer momentanen Arbeit gestört. Benutzer sind oft unwillig, explizites Feedback zu geben, wenn sie dadurch nicht einen unmittelbaren Mehrwert erhalten. Die Aufforderung, (regelmäßig) explizites Feedback zu geben oder Fragen des Systems zu beantworten, kann dazu führen, dass der Benutzer des Systems überdrüssig wird und das System fortan nicht mehr benutzt (vgl. [Fis01]). Bei implizitem Feedback werden Aktionen des Benutzers im Hinblick auf das Adaptionsziel bewertet. In einem interessenmodellierenden System werden Aktionen, in denen ein Benutzer Interesse zeigt, indem er z.B. Detailinformationen anfordert, als positives Feedback bewertet, während Aktionen, in denen ein Benutzer Desinteresse zeigt, als negatives Feedback aufgefasst werden können. In der Regel kann negatives, implizites Feedback in interessenmodellierenden Systemen nur schwer beobachtet werden, da Aktionen, die als Desinteresse gewertet werden können, im Gegensatz zu positiven Aktionen den Benutzer nicht zu dem Ziel führen, sein Informationsbedürfnis zu stillen. Zur Abhilfe könnte implizites, negatives Feedback als Abwesenheit von implizitem, positiven Feedback verstanden werden.

In Broker's Lounge wird nur mit implizitem Feedback gearbeitet, da der Be-

nutzer dadurch nicht in seiner momentanen Arbeitssituation gestört wird. Broker's Lounge arbeitet nur mit positivem Feedback, da negatives, implizites Feedback wie zuvor beschrieben, nur schwer zu beobachten ist. Dies determiniert die Auswahl des Lernverfahrens (siehe unten, Kap. 2.2.8) und hat zur Wahl des Lernens an positiven Beispielen (vgl. Kap. 2.2.9) geführt.

2.2.8 Akquisitions- und Lernverfahren

Wie schon zuvor beschrieben, ist ein Grundproblem bei benutzer- bzw. interessenmodellierenden Systemen die Akquisition von Wissen über den Benutzer. Um zu Beginn möglichst schnell Wissen über den Benutzer (bzw. seine Interessen) zu generieren, arbeiten viele Systeme mit Stereotypen (siehe oben, Kap. 2.2.5). Es verbleibt das Problem der Zuordnung von Benutzern zu Stereotypen. Dies wird in vielen Systemen durch ein Eingangsinterview gelöst. Im AVANTI-System (vgl. [FKS97]) werden Benutzer zum Beispiel zu Beginn danach gefragt, ob sie Rollstuhlfahrer sind. Ist dies der Fall, wird ihnen der entsprechende Stereotyp zugeordnet und sie bekommen von nun an Informationen für Rollstuhlfahrer angezeigt. Kritisch anzumerken bleibt, dass Stereotypen in vielen Fällen Ergebnis persönlicher Einschätzungen sind, die zur Designzeit eines adaptiven Systems vorgenommen wurden. Sie werden oft nicht auf der Basis von empirischen Untersuchungen entwickelt.

Neben dem Eingangsinterview hat ein System grundsätzlich immer die Möglichkeit, den Benutzer zu befragen. Dies sollte nicht zu oft geschehen, da dadurch Benutzer in ihren Arbeitsabläufen gestört werden. Das System sollte möglichst nur mit implizitem Feedback arbeiten (vgl. Kap. 2.2.7 und [Fis01]). Alle Aktionen des Benutzers werden protokolliert und interpretiert. Da die Aktionen des Benutzers nur indirekt etwas über die Interessen des Benutzers aussagen, verbleibt ein gewisses Maß an Unsicherheit. Um mit dieser Unsicherheit umgehen zu können, setzen viele Ansätze darauf, diese Unsicherheit im Interessenmodell zu beschreiben (vgl. [Jam96]). Aus diesem Grund haben in den letzten Jahren viele Ansätze Machine-Learning-Methoden wie Bayes'sche Netze oder neuronale Netze benutzt, um Benutzermodelle zu lernen (vgl. [WPB01]). Problematisch dabei ist jedoch, dass diese Methoden eine gewisse Menge an Trainingsdaten benötigen, bevor sie verlässliche Vorhersagen über das Benutzerverhalten machen können. Die meisten Machine-Learning-Verfahren benötigen sowohl negatives als auch positives Feedback, um Anpassungsleistungen anbieten zu können. Negatives Feedback kann in der Regel nur durch explizites Feedback gewonnen werden (vgl. Kap. 2.2.7). Viele Systeme sehen daher explizites Feedback durch Benutzer vor, um so im Betrieb Trainingsdaten für das Lernen des Benutzermodells zu bekommen. Eine Ausnahme bildet hier das Lernen an positiven Beispielen, das im LaboUr-Projekt (vgl. [PN99]) entwickelt wurde (vgl. Kap. 2.2.9), da es nur posi-

tive Beispiele benötigt, um ein Benutzermodell aufzubauen (vgl. [SP99a]). Dieses Verfahren wird in Broker's Lounge benutzt. Aus diesem Grund werde ich es im nächsten Abschnitt (Kap. 2.2.9) beschreiben. Die Integration dieses Lernverfahrens mit dem IB-Domänenmodell und den Interessenprofilen wird in Kapitel 7 beschrieben.

Viele Ansätze, die ich in dieser Arbeit beschreiben werde, benutzen Bayes'sche Netze oder einen naiven Bayes-Klassifikator. Eine tiefe Beschreibung von Bayes'schen Netzen findet sich in [Pea88]. Ihre Verwendung im Rahmen von Benutzermodellierung wird in [Jam96], [MLJ⁺98] und [HN99b] beschrieben. Die Verwendung eines naiven Bayes-Klassifikators wird in [BP99] erklärt.

Die verschiedenen Machine-Learning-Ansätze zum Lernen von Interessen können in inhaltsbasierte und kollaborative Methoden unterschieden werden (vgl. [ZA01]). Inhaltsbasierte Methoden bilden die Interessen eines Benutzers auf ein Modell der Inhalte des Systems² ab, während kollaborative Methoden die Nutzung der Inhalte und nicht die Inhalte selbst in ein Modell integrieren. Kollaborative Ansätze benötigen im allgemeinen eine kürzere Trainingsphase als inhaltsbasierte Ansätze. Kollaborative Verfahren sind im Gegensatz zu inhaltsbasierten Methoden nicht in der Lage, Aussagen über neue Inhalte zu machen. Ein weiteres Problem von Machine-Learning-Verfahren besteht darin, dass die gelernten Benutzermodelle nicht direkt lesbar sind und sie somit nicht von Benutzern inspiziert werden können. Eine Lösung für dieses Problem kombiniert Machine-Learning-Verfahren mit Techniken der Wissensrepräsentation (vgl. [PN99]). Machine Learning wird verwendet, um Benutzermodelle (bzw. Interessenprofile) zu lernen, während die Benutzermodelle inspizierbar bleiben, da sie als Overlay-Modell eines Domänenmodells beschrieben werden. Um die Integration von Machine-Learning-Verfahren und Domänenmodellen zu ermöglichen, sollten die Beobachtungsdaten, auf deren Grundlage das Lernverfahren arbeitet, in einem Eingabeformat vorliegen, mit dem das Lernverfahren arbeiten kann. Außerdem sollte das Eingabeformat die Struktur des Domänenmodells abbilden, damit das Lernverfahren diese Struktur ausnutzen kann, damit das Lernverfahren die Struktur des Domänenmodells ausnutzen kann.

2.2.9 Lernen an positiven Beispielen

Ein Lernverfahren, das mit positivem impliziten Feedback arbeitet, wurde im LaboUr-Projekt (**L**earning **a**bout the **U**ser, [PN99]) entwickelt. Das Lernverfahren (vgl. [SP99a], [SP99b], [SPK00]) basiert auf dem χ^2 -Test (vgl. [Sie85]). Es beobachtet den Benutzer und sucht nach signifikanten Abweichungen vom Er-

²Die Inhalte des Systems können unterschiedlich sein. Es kann sich z.B. um Bücher in einem E-Commerce-System für Bücher oder um Nachrichten in einem Nachrichtenportal handeln.

wartungswert in seinem Verhalten. Das ursprüngliche Verfahren wurde von mir mit dem IB-Domänenmodell integriert, das in Kapitel 6 beschrieben wird. Dadurch kann die Struktur des Domänenmodells zur Beschleunigung der Lernkurve ausgenutzt werden. Dies reflektiert die Aussage von [Ped00], der eine enge Verzahnung von Domänenmodell und Lernverfahren fordert, weil nur das adaptiert werden kann, was zuvor im Domänenmodell modelliert wurde.

Bei diesem Lernverfahren werden alle Items³, für die eine Relevanz vom System errechnet werden soll, in Featurevektoren (siehe Tabelle 2.1) eingetragen. In diesen Featurevektoren sind neben allen Items auch alle möglichen Werte für die verschiedenen Attribute (der Items) eingetragen. Für jeden möglichen Wert wird eine eigenen Spalte (ein eigenes Feature) eingeführt. Diese Featurevektoren sind ein Standard-Eingabeformat, wie es auch von anderen Lernverfahren benutzt wird.

	$feature_1$	$feature_2$	$feature_3$	$\dots$	$feature_m$
$item_1$	-	-	x	$\dots$	-
$item_2$	x	-	x	$\dots$	x
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$item_l$	x	x	-	$\dots$	-

Tabelle 2.1: Featurevektoren

Für jedes Feature wird ein Erwartungswert für das Auftreten dieses Wertes bei der nächsten Itemselektion berechnet. Der Erwartungswert entspricht der Anzahl der Vorkommen k eines Features geteilt durch die Anzahl der Items (siehe Formel 2.1). Das LaboUr-Verfahren sucht im Benutzerverhalten nach signifikanten Abweichungen vom Erwartungswert. Hat der Benutzer signifikant mehr Items (als der Erwartungswert) eines Features ausgewählt, so ist der Benutzer wahrscheinlich an Items mit diesem Feature interessiert. Hat er dagegen signifikant weniger Items mit einem bestimmten Feature ausgewählt, so ist er wahrscheinlich an Items mit diesem Feature nicht interessiert. Die Signifikanz von Ereignissen kann mit der Formel 2.2 bei einem gegebenen Konfidenzintervall berechnet werden. Das Konfidenzintervall wird durch den kritischen Wert z ausgedrückt. Für eine Konfidenz von 95% hat z den Wert 1.96, bei einer Konfidenz von 99% beträgt er 2,58 (vgl. [Sie86]). Es ergeben sich eine untere Schwelle c_1 und eine obere Schwelle c_2 . Liegt die Anzahl der beobachteten Ereignisse unter der Schwelle c_1 , so hat der Benutzer signifikantes Desinteresse gezeigt, liegt die Zahl der Ereignisse über c_2 , so hat er signifikantes Interesse bekundet. Geht man davon aus, dass die Ereignisse normalverteilt sind, muss der Beobachtungszeitraum n mindestens

³Ein Item beschreibt den Sachverhalt, dessen Relevanz bewertet werden soll. Es kann sich dabei um Bücher in einem E-Commerce-System oder um Nachrichten in einem Nachrichtenportal.

30 betragen, anderenfalls sind mindestens 100 Ereignisse in die Betrachtungen mit einzubeziehen.

Um die Relevanz einzelner Items berechnen zu können, ist es wichtig, die Relevanz der einzelnen Features quantifizieren zu können. Diese wird durch Formel 2.3 berechnet. Liegt ein signifikantes Ergebnis für Desinteresse vor, so wird der zwischen 0 und der unteren Schranke c_1 verbliebene Bereich genutzt, um einen Wert für Desinteresse zwischen 0 und -1 zu errechnen. Ähnlich geht man vor, um signifikantes Interesse zu quantifizieren: Der Bereich zwischen der oberen Schranke c_2 und dem Beobachtungszeitraum wird auf das Intervall von 0 bis 1 normalisiert. Ausgehend von den Relevanzwerten für Features kann die Relevanz für Items berechnet werden (siehe Formel 2.4). Dabei werden die Werte für alle Features eines Items aufaddiert und auf die Anzahl der vorhandenen Features f normalisiert.

$$p = \frac{k}{l} \quad (2.1)$$

$$c_{1,2} = pn \pm z\sqrt{p(1-p)n}, n \geq 30 \quad (2.2)$$

$$featurerelevanz = \begin{cases} -\frac{c_1-o}{c_1}, & 0 \leq o < c_1 \\ 0, & c_1 \leq o \leq c_2 \\ 1 - \frac{n-o}{n-c_2}, & c_2 < o \leq n \end{cases} \quad (2.3)$$

$$itemrelevanz = \frac{1}{f} \sum_{i=0}^f featurerelevanz_i \quad (2.4)$$

2.3 Sonstige Begriffe

In diesem Abschnitt beschreiben wir mit dem Vektorraummodell (Kap. 2.3.1) und dem Facettierungsproblem (Kap. 2.3.2) zwei weitere wichtige Begrifflichkeiten, die nicht eindeutig den Begriffswelten adaptiver oder ontologiebasierter Informationssysteme zugeordnet werden können.

2.3.1 Vektorraummodell

Das Vektorraummodell ist ein Standard-Modell aus dem Information-Retrieval. Beim Vektorraum-Verfahren werden Dokumente durch Terme repräsentiert. Bei diesen Termen kann es sich um Wörter, Wortgruppen, Grundformen, Stammformen oder ähnliches handeln. Die Terme spannen einen Vektorraum auf. Nachdem die Dokumente in einen Term-Vektor umgewandelt wurden, können sie mit

anderen Dokumenten zu einer Dokumenten-Term-Matrix arrangiert werden. Die Terme lassen sich unterschiedlich gewichten. Unter den verschiedenen Gewichtsverfahren ist das TFIDF-Verfahren (**T**erm **F**requency * **I**nverse **D**ocument **F**requency, vgl. [SM83]) das bekannteste. Das TFIDF-Verfahren berechnet für jeden Term t einen Wert $TFIDF(t)$ (vgl. Formel 2.5), indem man die Häufigkeit $TF(t)$, mit der ein Term in einem Dokument vorkommt, multipliziert mit dem Logarithmus des Quotienten der Anzahl aller Dokumente D und der Anzahl der Dokumente $DF(t)$, in denen der Term w wenigstens einmal vorkommt. Der erste Faktor in Formel 2.5 beschreibt die Relevanz eines Terms innerhalb eines Dokumentes. Der zweite Faktor bewertet die Relevanz eines Terms innerhalb der gesamten Dokumentensammlung.

Alle Terme, die unter einer bestimmten Schranke liegen, werden nicht weiter betrachtet. Die Gewichtung der Terme erlaubt auf einfache Art und Weise die Bewertung der Ähnlichkeit von Dokumenten. Es gibt im Information-Retrieval eine reichhaltige Literatur über das TFIDF-Verfahren sowie Verfahren, die an TFIDF angelehnt sind. Eine gute Übersicht darüber findet sich in [Bec01].

$$TFIDF(t) = TF(t) * \log\left(\frac{D}{DF(t)}\right) \quad (2.5)$$

Das Vektorraummodell findet in termbasierten, adaptiven Informationssystemen (vgl. Kap. 4.1) Verwendung. Das Interessenmodell besteht in diesen Systemen aus einem Vektorraum. Der Vektorraum wird mit Dokumenten aufgebaut, für die der Benutzer Interesse gezeigt hat oder die der Benutzer bereits angeschaut hat. Zwischen diesem Interessenmodell und potentiellen Dokumenten wird eine Ähnlichkeitsbewertung vorgenommen. Aus den Termvektoren kann eine Matrix aufgebaut werden. Die Matrix aus Termvektoren ist in Abbildung 2.2 abgebildet. Diese Matrix dient als Standardeingabeformat für ein Machine-Learning-Verfahren. Es ist ähnlich zu der Matrix aus Featurevektoren in Abbildung 2.1, die als Eingabeformat für das Lernen an positiven Beispielen dient.

	WM 2002	Völler	Fußball	...	Japan	Korea
<i>Dokument₁</i>	-	-	x	...	-	-
<i>Dokument₂</i>	x	-	x	...	x	x
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
<i>Dokument_n</i>	x	x	-	...	-	-

Tabelle 2.2: Termvektoren in einem termbasierten Informationssystem

2.3.2 Facettierungsproblem

In dieser Arbeit spielen Hierarchien zur Beschreibung von Sachverhalten eine wichtige Rolle. Dies wird in der Beschreibung des Domänenmodells in Kapitel 6 deutlich. Das Facettierungsproblem tritt auf, wenn entlang einer Kategorienhierarchie mehrere Aspekte beschrieben werden sollen, die voneinander unabhängig sind. Um alle Kombinationsmöglichkeiten abbilden zu können, müssen alle Permutationen beider Aspekte in die Hierarchie integriert werden. Dies führt dazu, dass die Zahl der Hierarchieeinträge explodiert. Um dies zu verhindern beschreibt man die verschiedenen Aspekte jeweils in einer eigenen, unabhängigen Kategorienhierarchie. Die Items, die klassifiziert werden sollen, werden in einem facettierten Kategorienhierarchie jeweils von den verschiedenen Kategoriehierarchien klassifiziert.

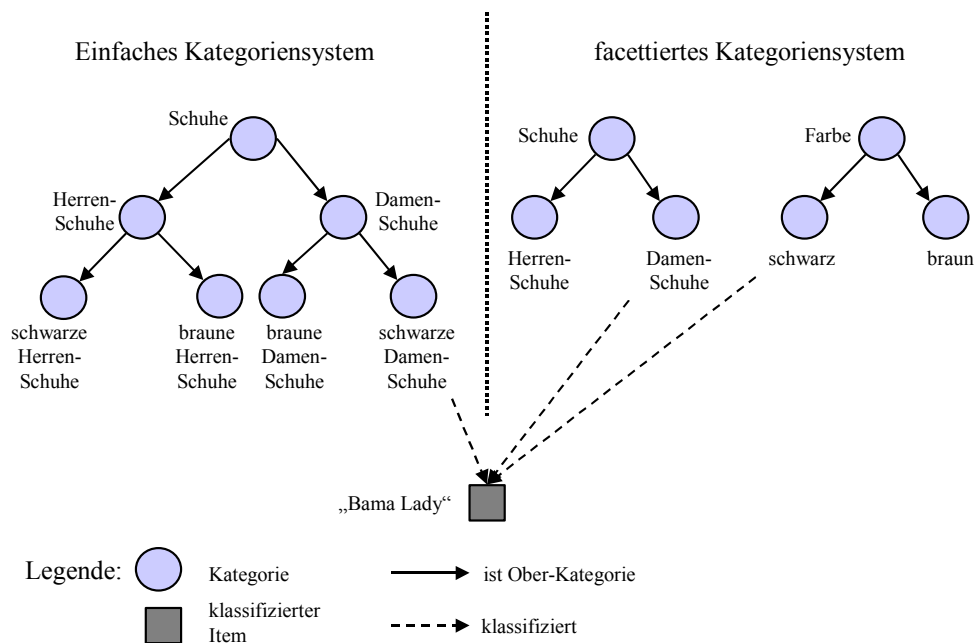


Abbildung 2.2: Gegenüberstellung eines einfachen und eines facettierten Kategoriensystems

Abbildung 2.2 stellt ein einfaches Kategoriensystem einem facettierten Kategoriensystem gegenüber. Dort wird ein kleines Kategoriensystem zur Klassifikation von Schuhen beschrieben. In dem einfachen Kategoriensystem, das auf der linken Seite dargestellt wird, werden für alle möglichen Kombinationen (schwarze und braune Herrenschuhe, sowie schwarze und braune Damenschuhe) jeweils eigene Kategorien in das Kategoriensystem integriert. In dem facettierten Kategoriensystem auf der rechten Seite werden stattdessen die beiden Aspekte (Damen-

Begriff	Bereich	Literaturquellen
Wissen	OIS	[DP98]
Wissensmanagement	OIS	[NT97], [DP98]
Ontologie	OIS	[WW89], [Gru93]
Adaptivität und Adaptierbarkeit	AIS	[Opp94], [Fis01]
Push und Pull	AIS	[Fis01]
Feedback	AIS	[Chi93]
Domänenmodell	AIS	[KK96], [Ped00]
Overlay-Modell	AIS	[KK96], [Ped00]
Stereotypen	AIS	[Ric79], [Bru01]
Bayes'sche Netze	AIS	[Pea88], [Jam96]
naiver Bayes-Klassifikator	AIS	[BP99]
Lernen an positiven Beispielen	AIS	[SPK00]
Vektorraummodell	Sonstige	[SM83], [Bec01]
Facettierungsproblem	Sonstige	[von99]

Tabelle 2.3: Übersicht über die beschriebenen Begrifflichkeiten

vs. Herrenschuhe) und die Farbe (schwarz vs. braun) in zwei getrennten Kategoriehierarchien verwaltet. In dem einfachen Kategoriensystem werden konkrete Schuhe (z.B. "Bama Lady") genau einer Kategorie zugeordnet. Im facettierten Kategoriensystem werden konkrete Schuhe jeweils mit einer Kategorie aus jeder Kategorienhierarchie klassifiziert.

Ein weiteres Beispiel für eine facettierte Struktur ist SNOMED. SNOMED ist ein Terminologiesystem für medizinische Fachbegriffe. Eine Beschreibung und Analyse von SNOMED und weiteren Terminologiesystemen vorwiegend aus der Medizin findet sich in [von99].

2.4 Zusammenfassung

In diesem Kapitel habe ich verschiedene Begrifflichkeiten aus den Bereichen adaptiver und ontologiebasierter Informationssysteme⁴ eingeführt, die in dieser Arbeit von Bedeutung sind. Ergänzend wurden die Begriffe des Facettierungsproblems und des Vektorraummodells eingeführt, die weder adaptiven noch ontologiebasierten Informationssystemen zugeordnet werden können. In Tabelle 2.3 werden diese Begriffe und entsprechende Literaturquellen aufgelistet und ihrem Gebiet zugeordnet.

⁴ Adaptive Informationssysteme werden in Tabelle 2.3 mit **AIS**, während ontologiebasierte Informationssysteme mit **OIS** abgekürzt werden.

Die Begrifflichkeiten aus dem Bereich ontologiebasierter Informationssysteme kreisen um den Begriff „Wissen“ und dessen Management. Fensel [Fen01] beschreibt Ontologien als Schlüsseltechnologie für Wissensmanagement.

Die Begrifflichkeiten aus dem Bereich adaptiver Informationssysteme kann in drei Gruppen aufgeteilt werden:

1. Die Unterscheidungen zwischen „Adaptivität und Adaptierbarkeit“ sowie zwischen „Push- und Pullsystemen“ beschreiben die Interaktion zwischen dem Benutzer und dem System. Diese Unterscheidungen werden in dieser Arbeit benutzt, um die verschiedenen Systeme gegeneinander zu vergleichen.
2. Die Begriffe „Domänen-“ und „Overlay-Modell“ sowie „Stereotypen“ gehören inhaltlich zusammen, da sie eine Architektur bedingen, in dessen Zentrum das Domänenmodell steht.
3. Die restlichen Begriffe aus dem Bereich adaptiver Informationssysteme beschäftigen sich mit dem Erlernen von Benutzerinteressen („Bayes’sche Netze“, „naiver Bayes-Klassifikator“, „Lernen an positiven Beispielen“). „Feedback“ beschreibt die Quelle auf dessen Basis die verschiedenen Lernverfahren arbeiten.

Das Facettierungsproblem bzw. dessen Lösung durch ein facettiertes Kategoriensystem findet Verwendung im IB-Domänenmodell für Broker’s Lounge, das in Kapitel 6 beschrieben wird. Es erlaubt eine reichhaltige Strukturierung und bietet darüber hinaus viele Möglichkeiten, an denen ein Lernverfahren ansetzen kann. Im Gegensatz dazu bietet ein Interessenprofil, das auf einem Vektorraummodell basiert, diese Möglichkeiten nicht, da das Vektorraummodell keine interne Strukturierung aufweist.

Kapitel 3

Information-Brokering

Information-Brokering kann nach [SMDP98] als der Prozess der Vermittlung zwischen Informationsangeboten und Informationsbedürfnissen angesehen werden. Es handelt sich um einen Wertschöpfungsprozess, da durch die Verdichtung von Information deren Wert (für den Informationskonsumenten) gesteigert wird (vgl. [Wor98]). In den Prozess sind drei verschiedene Rollen involviert. Informationsanbieter bieten Informationen an, Informationskonsumenten haben Informationsbedürfnisse, die der Information-Broker durch Vermittlung zwischen Informationsanbietern und Konsumenten zu befriedigen versucht. Dieses Verständnis von Information-Brokering basiert auf Erfahrungen aus dem ELFI- (vgl. [NKS98]) und dem COBRA-Projekt¹ (vgl. [SMDP98], [KK99]).

Dieses Kapitel beschreibt in den folgenden Abschnitten das grundlegende Aufgaben- und Rollenmodell, das als theoretische Fundierung von Broker's Lounge dient. Dieses Modell erweitert die Überlegungen zu Information-Brokering, die in der Einleitung (vgl. Kap. 1) kurz skizziert wurden. Danach werden grundlegende Begriffe des Information-Brokering wie Kategorien und Konzepte (vgl. Kap. 3.2), Profil (vgl. Kap. 3.3) und Dossier (vgl. Kap. 3.4) eingeführt. Besondere Bedeutung kommt dabei der Unterscheidung zwischen Kategorien und Konzepten zu, da diese Modellierungsmittel zur Bewertung der verschiedenen in dieser Arbeit vorgestellten Ansätze und Systeme benutzt wird. Abgeschlossen wird dieses Kapitel durch eine Betrachtung der verschiedenen Kategorisierungsprozesse im Information-Brokering (vgl. Kap. 3.5).

¹COBRA (Common open Brokerage Architecture) wurde von der Europäischen Kommission gefördert und hat eine allgemeine Information-Brokering-Architektur entwickelt, während ELFI (Elektronische Forschungsförder-Information) vom DFN-Verein aus Mitteln des Bundesministerium für Bildung und Forschung (BMBF) gefördert wurde und einen Informationsdienst für Forschungsförderung aufgebaut hat, der allen Wissenschaftlern in Deutschland zugänglich ist und einen schnellen Überblick über Fördermöglichkeiten gibt.

3.1 Aufgaben- und Rollenmodell

Information-Brokering kann nach [SMDP98] in drei Phasen unterschieden werden:

1. In einer ersten Rendezvousphase sollen die Interessen von Konsumenten mit den Angeboten der Anbieter zur Deckung gebracht werden.
2. In einer Transaktionsphase wird das Ergebnis durch den Broker geliefert und die Bezahlung durch den Kunden (Konsumenten) geregelt.
3. Eine abschließende Post-Transaktionsphase dient der Evaluierung der durch den Broker erbrachten Leistung.

Es ist zu beachten, dass in vielen Fällen die Informationsbeschaffung über einen Information-Broker und seine Bezahlung Bestandteil einer größeren Transaktion ist. Z.B. wird in der Domäne Forschungsförderung über Fördermöglichkeiten informiert. Die Information über Fördermöglichkeiten steht am Anfang eines Förderantrages. Die größere Transaktion, in die die Interaktion mit einem Information-Broker eingebettet ist, beschäftigt sich mit dem Stellen eines Förderantrages.

Ich will mich in dieser Arbeit auf die Rendezvousphase konzentrieren, da innerhalb der Rendezvousphase die Personalisierung von Inhalten stattfindet, die den Kern dieser Arbeit ausmacht. Die Transaktionsphase innerhalb des Information-Brokeringprozesses wurde z.B. schwerpunktmäßig vom MEMO-Projekt² (vgl. [SQ01]) bearbeitet.

Das hier beschriebene Aufgaben- und Rollenmodell findet sich in ähnlicher Form in der Dissertation von Roland Klemke (vgl. [Kle02]) wieder und beruht auf gemeinsamen Arbeiten. Das Aufgaben- und Rollenmodell dient in dieser Arbeit der Beschreibung der verschiedenen adaptiven und ontologiebasierten Informationssysteme (vgl. Kap. 4 und Kap. 5) als Brokeringssysteme sowie der Beschreibung der zentralen Rolle des IB-Domänenmodells im Information-Brokering.

Um seinen Kunden, den Konsumenten, gute Ergebnisse liefern zu können, muss der Broker deren Interessen und Informationsbedürfnisse kennen. Deswegen benötigt er ein möglichst genaues Modell ihrer Interessen. Da Broker üblicherweise eine größere Anzahl von Kunden zu bedienen haben, müssen sie diese Interessenmodelle möglichst weit formalisieren und auf die von den Informationsanbietern angebotenen Informationen abbilden. Um diesen Anforderungen gerecht zu werden, sollten sie ein IB-Domänenmodell aufbauen, das sowohl die verschiedenen Sichten der Informationsanbieter (von der Domäne) berücksichtigt

²<http://www.abnamro.com/memo/>

als auch effizient auf die Interessenmodelle der Kunden abgebildet werden kann. Damit der Broker seinen Kunden jederzeit aktuelle Informationen anbieten kann, müssen drei unabhängige Prozesse ausgeführt werden:

1. **Retrieval:** Veränderungen im Informationsangebot der Informationsanbieter müssen im Quellenmodell nachvollzogen werden. Neben dem einfachen Retrieval der Informationsangebote, müssen noch weitere, verwandte Aufgaben zur Verwaltung des Quellenmodells ausgeführt werden. Diese Aufgaben heißen daher zusammenfassend *Quellenverwaltungsaufgaben*.
2. **Strukturierung:** Die Informationsangebote der Informationsanbieter sind in vielen Fällen Träger der gewünschten Informationen. D.h., nicht die Informationsangebote (Dokumente), sondern die darin beschriebenen Informationen sind für den Broker und seine Klienten von Interesse. Diese Informationen müssen aus den Dokumenten herausgezogen und in das IB-Domänenmodell integriert werden. Neben der reinen Extraktion muss das IB-Domänenmodell ggf. angepasst werden, da unter Umständen neue Sachverhalte aufgetaucht sind, die bisher im Modell noch nicht beschrieben werden konnten. Diese Aufgaben werden *Domänenverwaltungsaufgaben* genannt.
3. **Personalisierung:** Die Inhalte des IB-Domänenmodell müssen auf die Bedürfnisse der verschiedenen Konsumenten angepasst werden. Dabei handelt es sich im Wesentlichen um eine Filterung der relevanten Inhalte des IB-Domänenmodells. Ziel ist es dabei, die Bedürfnisse des Konsumenten, der in diesem Fall Kunde des Brokers ist, adäquat zu befriedigen. Dieser Prozess wird als *kundenorientierter Prozess* bezeichnet.

Die Quellen- und Domänenverwaltungsaufgaben können zwischen Informationsanbietern und Brokern beliebig verteilt werden, während die kundenorientierten Aufgaben entsprechend zwischen Brokern und Konsumenten aufgeteilt werden können. Je nach Aufgabenverteilung auf die drei Rollen ergeben sich verschiedene Brokeringmodelle.

Abbildung 3.1 zeigt zum Beispiel eine paritätische Aufgabenverteilung der Quellen- und Domänenverwaltungsaufgaben zwischen Informationsanbietern und dem Broker sowie eine Teilung der kundenorientierten Aufgaben. Denkbar ist auch eine Aufgabenverteilung, bei der alle Aufgaben von einem Broker ausgeführt werden. D.h., das IB-Domänenmodell wird vom Broker an Veränderungen bei den Informationsanbietern angepasst und die Interessenmodelle der Kunden werden ebenfalls vom Broker verwaltet. Konsumenten stellen dann Anfragen an den Broker, die dieser auf Basis des jeweiligen Interessenmodells bearbeitet. Es ist außerdem möglich, alle Aufgaben zwischen Konsumenten und Anbietern zu

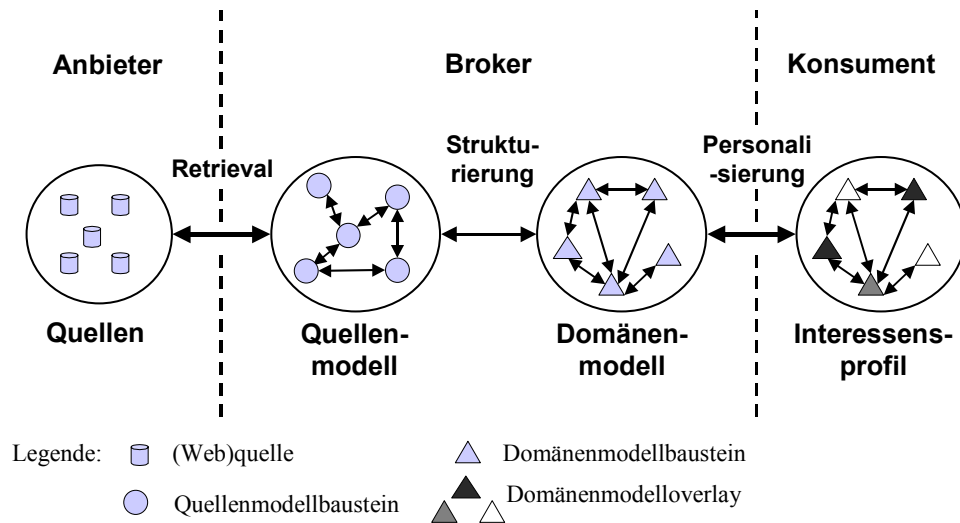


Abbildung 3.1: Information-Brokering-Modell

teilen. Der Broker entfällt, da Anbieter und Konsumenten diese Rolle dann zusammen ausfüllen. Bei diesem Modell sind zwei Extreme denkbar. Wenn der Anbieter alle Aufgaben ausführt, sprechen [SMDP98] von einem sog. Push-Broker-Modell. Werden dagegen alle Aufgaben vom Konsumenten selbst durchgeführt, sprechen sie von einem Pull-Broker-Modell.

Im Zentrum des Information-Brokering steht das sog. IB-Domänenmodell. Dies ist vergleichbar mit einem zentralen Enterprise-Modell in einem idealen Data-Warehouse-Modell, wie es in [JJQ⁺99] vorgeschlagen wird. Das Enterprise-Modell wird von verschiedenen OLAP³-Systemen genutzt, um multidimensionale Sichten (vgl. [FBSV99]) zu definieren. In einem idealen Data-Warehouse sind die verschiedenen Quellen ebenfalls Sichten auf das zentrale Data-Warehouse. Trifft diese idealisierte Annahme nicht zu, so wird zwischen den Quellen ein Wrapper benötigt, der die Daten aus den Quellen in eine mit dem Enterprise-Modell konforme Struktur übersetzt.

Während das IB-Domänenmodell mit dem Enterprise-Modell verglichen werden kann (siehe oben, letzter Abschnitt), entsprechen die Interessenprofile der Konsumenten den OLAP-Applikationen auf Client-Seite. Die (Web-)Quellen im Information-Brokeringprozess können mit den Quellen im Data-Warehouse-Modell zur Deckung gebracht werden. Das Quellenmodell kann mit den Wrappern eines Data-Warehouses verglichen werden. Dieser Zwischenschritt ist notwendig, da die (Web-)Quellen nicht automatisch auf das IB-Domänenmodell abgebildet werden können. In Semantic-Web-Systemen (vgl. Kap. 5.1) kann dagegen direkt

³Online Analytical Processing

auf annotierten Dokumenten gearbeitet werden.

Alle in dieser Arbeit beschriebenen adaptiven und ontologiebasierten Informationssysteme (Kapitel 4 und Kapitel 5) werden auf dieses Information-Brokering-Modell abgebildet, um sie anhand ihres Brokeringmodells klassifizieren zu können. Zur Verdeutlichung wird Abbildung 3.1 jeweils leicht modifiziert. Es zeigt sich, dass viele Systeme auch als Brokeringssysteme angesehen werden können. Mit den drei vorgestellten Rollen (Anbieter, Broker und Konsument) grenzt sich das vorgestellte Modell wesentlich gegen andere Information-Brokering-Ansätze (z.B. [FDES98], [KH95], [Wie94], [FEFP95], [JP96], eine Übersicht über Information-Brokering-Ansätze bietet [JP98]) ab, die sich hauptsächlich auf die Integration verschiedenster Datenquellen konzentrieren, um Benutzern eine einheitliche Meta-Schnittstelle anbieten zu können, mit der sie Anfragen stellen können.

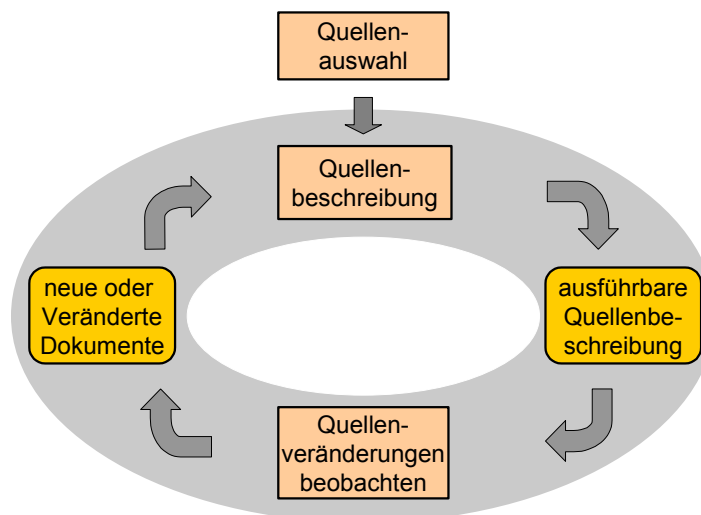


Abbildung 3.2: Quellenverwaltungsaufgaben

Die Quellenverwaltungsaufgaben (siehe Abb. 3.2) dienen der Synchronisation des Quellenmodells des Brokers und der Quellen der Informationsanbieter. Da der Broker nicht alle verfügbaren Quellen überwachen kann, ist eine gute Auswahl der relevanten Quellen (*Quellenauswahl*) wichtig, weil dadurch die gefundenen Ergebnisse verbessert werden. Wurde eine neue Quelle ausgewählt, so muss eine *Quellenbeschreibung* erstellt werden, die eine automatische Überprüfung der Quelle erlaubt (*Quellenveränderungen beobachten*) und die Suche auf die wichtigsten Bereiche einer Quelle fokussiert. Ergebnis der automatischen Quellenüberprüfung sind *neue oder veränderte Dokumente*, die dann auf das IB-Domänenmodell des Brokers abgebildet werden müssen. Handelt es sich bei den Quellen um Informationen, die schon strukturiert sind, so können sie ggf. direkt

auf das IB-Domänenmodell abgebildet werden. Entspricht diese Struktur schon der Struktur des IB-Domänenmodell des Brokers, wie das zum Beispiel im Ontobroker-Projekt (vgl. [FDES98]) der Fall ist, können diese neuen Informationen direkt in das IB-Domänenmodell integriert werden⁴. Liegen die Informationen dagegen in unstrukturierter Form vor, so müssen die relevanten Informationen noch in einem Zyklus der Domänenverwaltungsaufgaben auf das IB-Domänenmodell abgebildet werden.

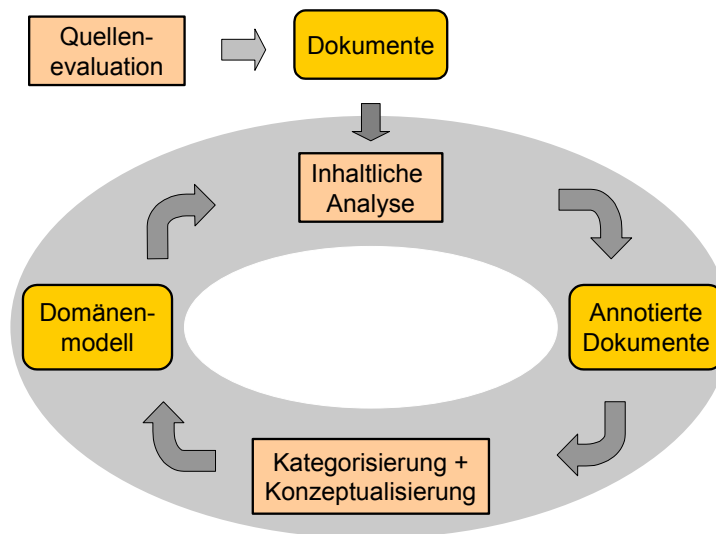


Abbildung 3.3: Domänenverwaltungsaufgaben

Zu den Domänenverwaltungsaufgaben gehören die *inhaltliche Analyse* und die *Konzeptualisierung/Kategorisierung* (siehe Abb. 3.3). Die bei der Quellen-evaluation gefundenen neuen oder veränderten Dokumente werden bei der inhaltlichen Analyse mit den Inhalten des IB-Domänenmodells verglichen, indem z.B. nach Schlüsselwörtern (für die Inhalte des IB-Domänenmodells) gesucht wird. Dies dient einer automatischen Vorkategorisierung der Dokumente. In Brokering-Szenarien, in denen der Broker mit annotierten Dokumenten handelt, wie z.B. in MarketMonitor (vgl. Kapitel 9.2), werden diese Dokumente nur weiter (manuell) kategorisiert, um das IB-Domänenmodell (und damit die Schlüsselwörter für die inhaltliche Analyse) zu erweitern. Wird mit Inhalten des IB-Domänenmodells gehandelt, so schließt sich an die inhaltliche Analyse eine obligatorische manuelle Konzeptualisierung/Kategorisierung der Dokumente an. Bei der Konzeptualisierung werden die Informationen in dem Dokument einem Konzept in der

⁴Wie man in Kap. 5.1 sehen kann, führen Informationsanbieter in diesem Fall die Domänenverwaltungsaufgaben weitestgehend selbstständig durch.

Konzepthierarchie zugeordnet. Ggf. werden sie einer bereits bekannten *Informationseinheit* (dieses Konzeptes) zugeteilt. Anschließend werden die gefundenen oder neuen Informationseinheiten nach einem Kategoriensystem *kategorisiert*. Damit wird die Unterscheidbarkeit der verschiedenen Informationseinheiten eines Konzeptes sichergestellt. Eine Trennung von Konzeptualisierung und Kategorisierung ist nicht möglich, da dies zu inkonsistenten Zuständen des zentralen IB-Domänenmodells führen würde.

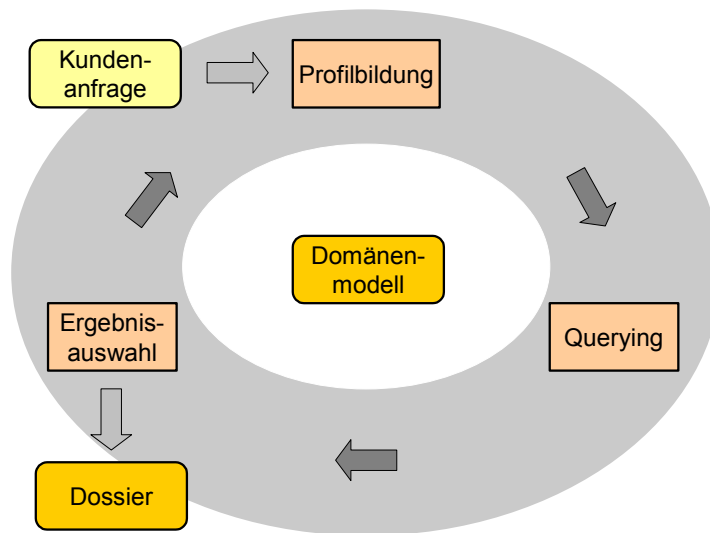


Abbildung 3.4: Kundenorientierte Aufgaben

Die kundenorientierten Aufgaben bestehen aus *Kundenanfrage*, *Profilbildung*, *Querying* und *Ergebnisauswahl* (siehe Abb. 3.4). Wenn ein Konsument eine Anfrage an den Broker stellt, wird diese aufgenommen (*Kundenanfrage*) und in ein Interessenprofil umgesetzt (*Profilbildung*). Ein Interessenprofil besteht aus verschiedenen Domänenmodellfiltern. Diese Filter werden dann als Anfrage auf das IB-Domänenmodell angewendet (*Querying*). Die Ergebnisse werden ggf. nochmals auf ihre Relevanz für den Kunden überprüft (*Ergebnisauswahl*) und dann zu einem *Dossier* zusammengefasst, das an den Kunden geliefert wird. Je nachdem, womit der Broker handelt, wird das Interessenprofil direkt auf das IB-Domänenmodell oder mittelbar (über das IB-Domänenmodell) auf das Quellenmodell angewendet. Im ersten Fall besteht das Ergebnis aus einer Menge von Informationseinheiten, die dem durch das Interessenprofil beschriebenen Bedingungen genügen. Im zweiten Fall wird zuerst ebenfalls eine Menge von Informationseinheiten aus dem IB-Domänenmodell mit Hilfe des Interessenprofils gefiltert. Die einzelnen Informationseinheiten sind wiederum mit einzelnen Dokumenten (in denen sie

beschrieben werden) assoziiert. Diese Verbindungen werden ausgewertet und die entsprechenden Dokumente sind das Ergebnis.

Im weiteren Verlauf der Arbeit will ich mich auf die Profilbildung innerhalb des kundenorientierten Prozesses konzentrieren, weil diese Aufgabe im Unterschied zu den anderen Tätigkeiten (im kundenorientierten Prozess) am schwierigsten zu automatisieren ist. Profilbildung ist eine intellektuelle Aufgabe, die den Broker fordert. Ziel ist es, dem Broker eine möglichst weitgehende Hilfestellung gewähren zu können. Da das Profil aus einer Menge von Domänenmodellfiltern besteht, ist das IB-Domänenmodell von zentraler Bedeutung für die Profilbildung und die angestrebte Unterstützungsleistung. Die effiziente Unterstützung der Profilbildung ist deswegen eine zentrale Anforderung an das IB-Domänenmodell.

Neben den Quellen- und Domänenmodellverwaltungsaufgaben und den Aufgaben im kundenorientierten Prozess gibt es eine Gruppe von Infrastrukturaufgaben, die den Betrieb eines Information-Brokering-Systems wie Broker's Lounge ermöglichen. Dazu gehört z.B. das Aufbauen eines initialen IB-Domänenmodells oder die Administration von Benutzern und ihrer Rechte. Auf diese Infrastrukturaufgaben werde ich an dieser Stelle nicht eingehen, sondern sie im Broker's Lounge-Kapitel (Kap. 8) kurz erklären.

3.2 Kategorien und Konzepte

In dieser Arbeit wird unterschieden zwischen *Kategorien* und *Konzepten*, obwohl diese Begriffe oft synonym verwendet werden. Diese Unterscheidung ist motiviert durch die beiden Themenfelder, die in dieser Arbeit zusammengebracht werden.

Während der Begriff der „Kategorie“ in domänenmodellbasierten, adaptiven Informationssystemen Verwendung findet, werden Ontologien im Wesentlichen aus Konzepten aufgebaut (siehe oben, Kap. 2.1.3).

Die Kategorien in adaptiven, domänenmodellbasierten Informationssystemen werden in der Regel hierarchisch organisiert. Die daraus entstandene Kategorienhierarchie wird als Domänenmodell benutzt und die Interessen der Benutzer werden auf dieses Modell abgebildet. Auch die Konzepte in ontologiebasierten Informationssystemen werden zu einer (Konzept-)Hierarchie arrangiert. Neben dieser hierarchischen Eigenschaft, die der hierarchischen Eigenschaft von Kategorien ähnelt, können Konzepte Relationen zu anderen Konzepten bilden. Dies wird in Abbildung 3.5 deutlich, die Kategorie- und Konzeptstrukturen zur Charakterisierung von Informationseinheiten gegenüberstellt.

Im nächsten Abschnitt (Kap. 3.2.1) wird zunächst die Funktion von Kategorien und Konzepten in adaptiven und ontologiebasierten Informationssystemen beschrieben. Daran schließt sich eine formale Beschreibung von Kategorien und Konzepten bzw. von Kategorie- und Konzeptgraphen an (Kap. 3.2.2). Zur Ver-

deutlichung der Funktion und Verwendung von Kategorien und Konzepten werden anschließend (Kap. 3.2.3) Beispiele für ihre Verwendung gegeben.

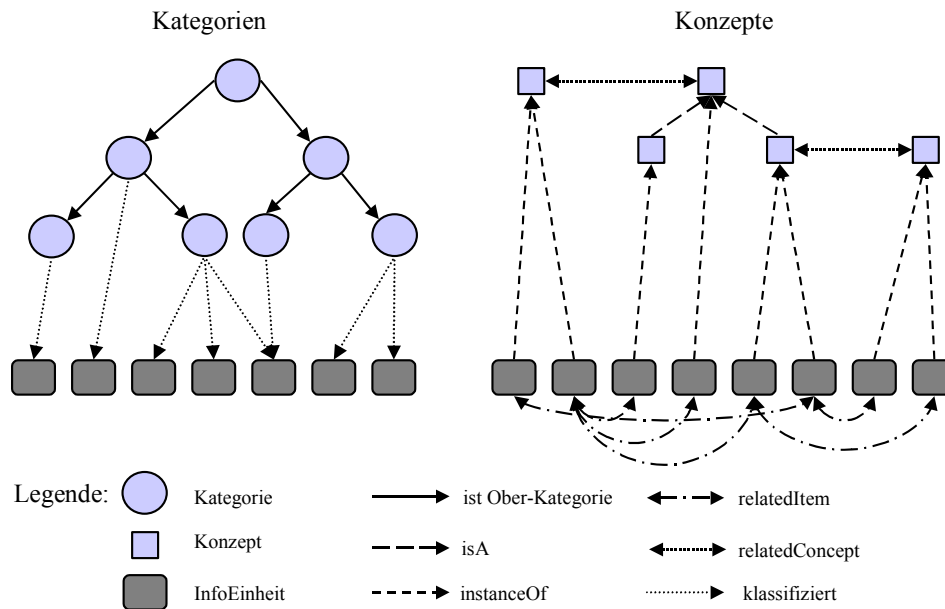


Abbildung 3.5: Struktureller Unterschied zwischen Kategorien und Konzepten

3.2.1 Funktion

Die Funktionen von Kategorien in adaptiven, domänenmodellbasierten Informationssystemen und von Konzepten in ontologiebasierten Informationssystemen unterscheiden sich:

- Kategorien werden in adaptiven, domänenmodellbasierten Systemen benutzt, damit Interessen von Benutzern darauf abgebildet werden können. Die Kategorien beschreiben Eigenschaften des Adaptionziels. In vielen adaptiven, domänenmodellbasierten Systemen (vgl. Kap. 26) werden Kategorien benutzt, um die thematische Eigenschaft von Dokumenten⁵ zu beschreiben. In allen vorgestellten adaptiven, domänenmodellbasierten Systemen werden Kategorien zur Klassifikation benutzt. In vielen Fällen werden dabei Dokumente klassifiziert.
- Konzepte werden in ontologiebasierten Informationssystemen benutzt, um relevante Phänomene der realen Welt zu beschreiben (vgl. [Gru93]). Die

⁵In diesem Fall stellen Dokumente das Adaptionziel dar.

Konzepte bilden zusammen eine Ontologie (siehe oben, Kap. 2.1.3). Konzepte können zueinander in Beziehung gesetzt werden, um Beziehungen (der in den Konzepten beschriebenen Phänomene) in der realen Welt nachzubilden. Neben einfachen Relationen zwischen Konzepten (relatedTo-Beziehungen) können auch Subsumptionsrelationen (isA-Beziehungen) zwischen Konzepten definiert werden. Zu den einzelnen Konzepten können jeweils Instanzen gebildet werden. D.h., es besteht eine Typisierungsrelation zwischen Konzepten und ihren Instanzen. Die zwischen den Konzepten definierte Struktur wird auf die Instanzen übertragen. Besteht z.B. eine Relation zwischen den Konzepten k_1 und k_2 , so kann diese Relation von den Instanzen i_1 und i_2 ⁶ mit Leben gefüllt werden. Die Subsumptionsrelation zwischen zwei Konzepten k_1 und k_2 bedeutet, dass alle Instanzen von k_1 auch gleichzeitig Instanzen von k_2 ⁷ sind. Außerdem erbt k_1 alle Relationen von k_2 . In beschreibungslogischen Systemen können Instanzen anhand ihrer Eigenschaften, bzw. ihrer Relationen automatisch einem Konzept in der Konzepthierarchie zugeordnet werden.

3.2.2 Formale Definition

In diesem Abschnitt werde ich Kategorien und Konzepte mit Hilfe von gerichteten Graphen beschreiben. Kategorien werden dabei im Rahmen eines Kategoriengraphen und Konzepte im Rahmen von Konzeptgraphen beschrieben. Die beiden Definitionen für Kategorie- und Konzeptgraphen werden in Kapitel 6 zur formalen Definition des IB-Domänenmodells zusammengefasst und erweitert.

Bei der Notation habe ich mich an die Graphentheorie angelehnt, die in [Kru98] formuliert wird. Dazu habe ich die Definitionen für einen gerichteten Graphen, Wege und Kreise in gerichteten Graphen übernommen und an die eigenen Bedürfnisse angepasst.

Definition 3.1 (Gerichteter Graph)

Ein **gerichteter Graph** (oder einfach **Graph**) ist ein Quadrupel $G = (V, R, \alpha, \omega)$ mit folgenden Eigenschaften:

1. V ist eine nicht leere Menge, die **Eckenmenge** des Graphen.
2. R ist eine Menge, die **Pfeilmenge** des Graphen.
3. Es gilt $V \cap R = \emptyset$.

⁶In diesem Fall ist i_1 eine Instanz von k_1 und i_2 eine Instanz von k_2

⁷In diesem Fall ist k_i ein Unterkonzept von k_2 .

4. $\alpha : R \rightarrow V$ und $\omega : R \rightarrow V$ sind Abbildungen. $\alpha(r)$ ist die Anfangsecke und $\omega(r)$ die Endecke des Pfeils r .

Der Graph G heißt **endlich**, wenn $|V \cup R| < +\infty$.

Definition 3.2 (Weg, Kreis)

Ein **Weg** w im Graphen G ist eine endliche Folge $w = (r_1, \dots, r_k) (k \geq 1)$ mit

1. $r_j \in R$ für $j = 1, \dots, k$ und
2. $\omega(r_j) = \alpha(r_{j+1})$ für $j = 1, \dots, k-1$.

Die **Startecke** des Weges w ist durch $\alpha(w) := \alpha(r_1)$ und die **Endecke** durch $\omega(w) := \omega(r_k)$ festgelegt. Die **Länge** $|w|$ des Weges w ist die Anzahl k der durchlaufenen Pfeile. Ein Weg heißt **Kreis**, wenn $\alpha(w) = \omega(w)$.

In den folgenden Definitionen für den Kategorie- und den Konzeptgraphen werden abweichend von der Definition für einen gerichteten Graphen mehrere disjunkte Ecken- und Pfeilmengen verwendet. Außerdem werden mehrere α - und ω -Abbildungen benutzt. Die verschiedenen Ecken- und Pfeilmengen können durch Vereinigung eine neue Ecken- bzw. Pfeilmenge definieren, die der ursprünglichen Definition eines gerichteten Graphen entspricht. Die verschiedenen α - und ω -Abbildungen können durch eine Fallunterscheidung jeweils eine α - und eine ω -Abbildung bilden.

Die folgende Definition eines *Kategoriengraphen* definiert ein Modell, dass auf Basis eines Kategoriensystems zur Charakterisierung von Informationseinheiten benutzt werden kann. Abbildung 3.6 zeigt z.B. einen Kategoriengraphen, der zur Beschreibung des Forschungsthemas von Förderprogrammen benutzt wird. In Kapitel 3.2.3 wird dieses Beispiel aufgegriffen und näher erläutert.

Definition 3.3 (Kategoriengraph)

Ein **Kategoriengraph** ist ein 8-Tupel $G_{Kat} = (Kat, I, O, Kla, \alpha O, \omega O, \alpha Kla, \omega Kla)$ mit folgenden Eigenschaften:

1. Kat ist eine nicht leere Menge von Kategorien. Kategorien sind Ecken des Graphen G_{Kat} .
2. I ist eine nicht leere Menge von Informationseinheiten. Informationseinheiten sind Ecken des Graphen G_{Kat} .

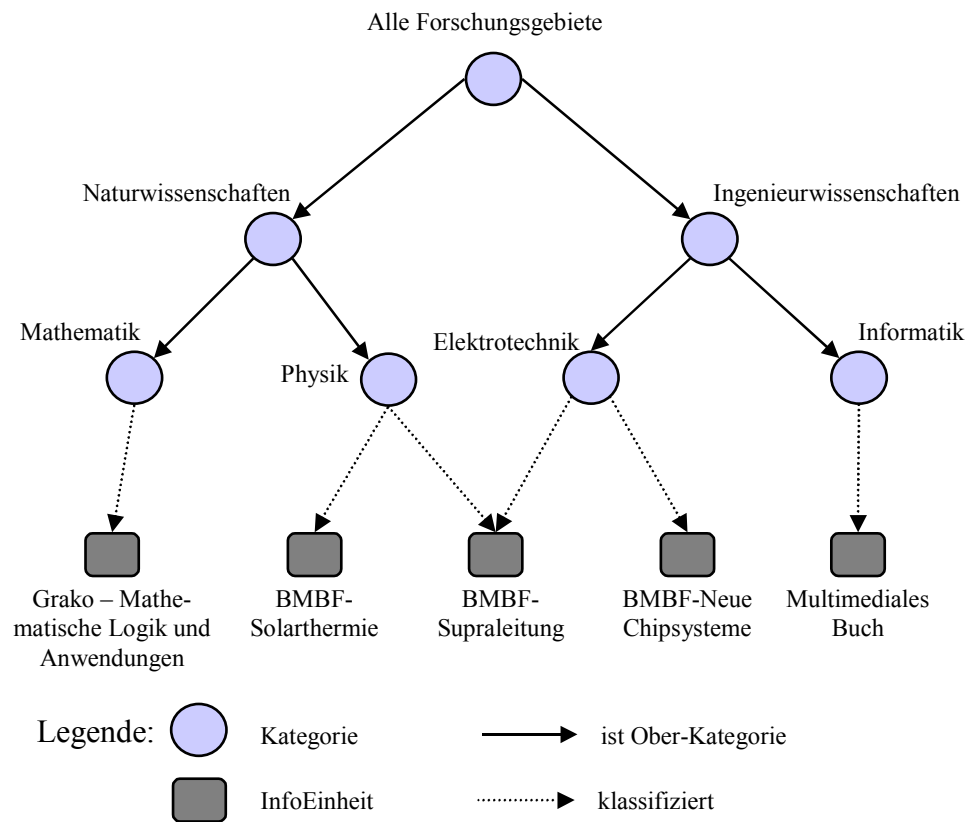


Abbildung 3.6: Beispielstruktur mit Kategorien

3. O ist eine Menge von *istOberkategorieVon*-Beziehungen (vgl. Abb. 3.6). Diese Beziehungen sind Pfeile des Graphen G_{Kat} .
4. Kla ist eine Menge von *Klassifikationsbeziehungen*. *Klassifikationsbeziehungen* sind Pfeile im Graphen G_{Kat} .
5. Es gilt $Kat \cap I = \emptyset \wedge Kat \cap O = \emptyset \wedge Kat \cap Kla = \emptyset \wedge I \cap O = \emptyset \wedge I \cap Kla = \emptyset \wedge O \cap Kla = \emptyset$.
6. $\alpha O : O \rightarrow Kat$ und $\omega O : O \rightarrow Kat$ sind Abbildungen. $\alpha O(o)$ heißt *Anfangsecke des Pfeils o* . $\omega O(o)$ heißt *Endecke des Pfeils o* .
7. $\alpha Kla : Kla \rightarrow Kat$ und $\omega Kla : Kla \rightarrow I$ sind Abbildungen. $\alpha Kla(kla)$ heißt *Anfangsecke des Pfeils kla* . $\omega Kla(kla)$ heißt *Endecke des Pfeils kla* .
8. Der Graph G_{Kat} ist *kreis- und schlingenfrei*. D.h., es existiert kein $o \in O$, für das $\alpha O(o) = \omega O(o)$ gilt.

9. Jede Informationseinheit i muss von mindestens einer Kategorie $kat \in Kat$ klassifiziert werden. $\forall i \in I : \exists kla \in Kla : \alpha Kla(kla) \in Kat \wedge \omega Kla(kla) = i$.
10. Eine Informationseinheit i kann nicht gleichzeitig von einer Kategorie kat_2 und einer Oberkategorie kat_1 dieser Kategorie klassifiziert werden:
 Wenn ein Weg w existiert mit $\alpha(w) = kat_1$ und $\omega(w) = kat_2$ und eine Klassifikationsbeziehung kla zwischen kat_2 und einer Informationseinheit i mit $\alpha Kla(kla) = kat_2$ und $\omega Kla(kla) = i$ existiert, dann existiert **keine** Klassifikationsbeziehung kla' mit $\alpha Kla(kla') = kat_1$ und $\omega Kla(kla') = i$.
11. Alle Kategorien bilden zusammen einen Baum mit einer eindeutigen Wurzel kat_{root} . D.h., es existiert ein Weg w von der Wurzel zu jeder Kategorie kat_i :
 $\exists kat_{root} \in Kat : \forall kat_i \in Kat : \exists w : \alpha(w) = kat_{root} \wedge \omega(w) = kat_i$

Die folgende Definition von **Konzeptgraphen** basiert auf Konzepten und dient ebenfalls der Charakterisierung von Informationseinheiten. Die Struktur von Informationseinheiten kann durch die Beschreibung von Relationen zwischen Konzepten determiniert werden. Ein Beispiel für einen Konzeptgraphen zeigt Abbildung 3.7. In Kapitel 3.2.3 wird dieses Beispiel aufgegriffen und erläutert.

Definition 3.4 (Konzeptgraph)

Ein **Konzeptgraph** ist ein 14-Tupel $G_{Kon} = (Kon, I, IsA, RelKon, InstanceOf, RelI, \alpha IsA, \omega IsA, \alpha RelKon, \omega RelKon, \alpha InstanceOf, \omega InstanceOf, \alpha RelI, \omega RelI)$ mit folgenden Eigenschaften:

1. Kon ist eine nicht leere Menge von Konzepten. Konzepte sind Ecken im Graphen G_{Kon} .
2. I ist eine nicht leere Menge von Informationseinheiten. Informationseinheiten sind Ecken des Graphen G_{Kon} .
3. IsA ist eine Menge von Generalisierungsbeziehungen zwischen Konzepten. IsA -Beziehungen sind Pfeile im Graphen G_{Kon} .
4. $RelKon$ ist eine Menge von einfachen Beziehungen zwischen Konzepten. $RelKon$ -Beziehungen sind Pfeile im Graphen G_{Kon} .
5. $InstanceOf$ ist eine Menge von Beziehungen zwischen Informationseinheiten und Konzepten. $InstanceOf$ -Beziehungen sind Pfeile im Graphen G_{Kon} .

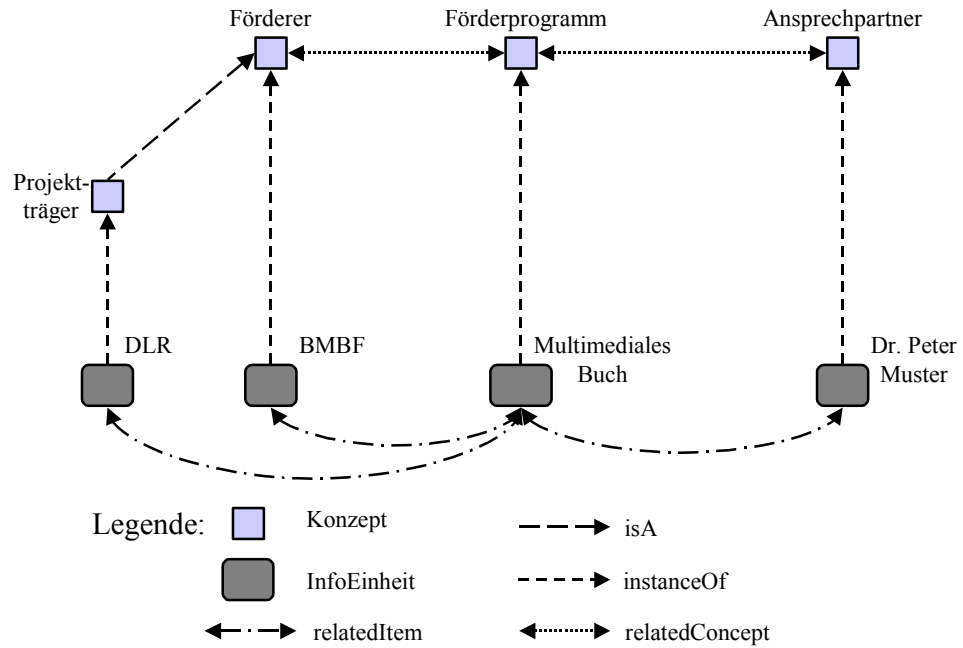


Abbildung 3.7: Beispielstruktur mit Konzepten

6. $RelI$ ist eine Menge von Beziehungen zwischen Informationseinheiten. Diese Beziehungen sind ebenfalls Pfeile im Graphen G_{Kon} und sind abhängig von den Beziehungen zwischen ihren jeweils zugeordneten Konzepten. Dieser Sachverhalt wird weiter unter dem Punkt 11. formal beschrieben.
7. Es gilt $Kon \cap I = Kon \cap IsA = Kon \cap RelKon = Kon \cap InstanceOf = Kon \cap RelI = I \cap IsA = I \cap RelKon = I \cap InstanceOf = I \cap RelI = IsA \cap RelKon = IsA \cap InstanceOf = RelKon \cap RelI = InstanceOf \cap RelI = \emptyset$.
8. $\alpha IsA : IsA \rightarrow Kon$ und $\omega IsA : IsA \rightarrow Kon$ sind Abbildungen. $\alpha IsA(isA)$ heißt Anfangsecke und $\omega IsA(isA)$ heißt Endecke des Pfeils isA . Der Graph $G'_{Kon} = (Kon, IsA, \alpha IsA, \omega IsA)$ ist kreis- und schlingenfrei. D.h., es existiert kein $isA \in IsA$ für das $\alpha IsA(isA) = \omega IsA(isA)$ gilt.
9. $\alpha RelKon : RelKon \rightarrow Kon$ und $\omega RelKon : RelKon \rightarrow Kon$ sind Abbildungen. $\alpha RelKon(relKon)$ und $\omega RelKon(relKon)$ heißen Anfangs- und Endecke⁸.

⁸An dieser Stelle habe ich den ungerichteten Graphen $G''_{Kon} = (Kon, RelTo, \gamma)$ als gerichteten Graphen $G'''_{Kon} = (Kon, RelTo, \alpha RelTo, \omega RelTo)$ beschrieben, um nicht einen gerichteten

10. $\alpha InstanceOf : InstanceOf \rightarrow I$ und $\omega InstanceOf : InstanceOf \rightarrow Kon$ sind Abbildungen. $\alpha InstanceOf(instanceOf)$ heißt Anfangsecke von $instanceOf$. $\omega InstanceOf(instanceOf)$ heißt Endecke von $instanceOf$. Jede Informationseinheit ist eindeutig genau einem Konzept zugeordnet. Für alle Informationseinheiten $i \in I$ gilt:

- $\exists instanceOf \in InstanceOf, kon \in Kon : \alpha InstanceOf(instanceOf) = i \wedge \omega InstanceOf(instanceOf) = kon$ und
- $\neg \exists instanceOf' \in InstanceOf, kon' \in Kon : \alpha InstanceOf(instanceOf') = i \wedge \omega InstanceOf(instanceOf') = kon'$.

11. $\alpha RelI : RelI \rightarrow I$ und $\omega RelI : RelI \rightarrow I$ sind Abbildungen. $\alpha RelI(relI)$ und $\omega RelI(relI)$ heißen Anfangs- und Endecke von $relI$. Die Beziehungen zwischen Informationseinheiten werden determiniert durch die Beziehungen zwischen den jeweiligen Konzepten. Für alle $relI \in RelI$ existieren zwei Wege w_1 und w_2 mit $w_1 = (instanceOf, relKon)$, $instanceOf \in InstanceOf$, $relKon \in RelKon$ und $w_2 = (instanceOf', relKon)$, $instanceOf' \in InstanceOf$, $relKon \in RelKon$ für die gilt:

- (a) $\alpha RelI(relI) = i_1$,
- (b) $\omega RelI(relI) = i_2$,
- (c) $\alpha InstanceOf(instanceOf) = i_1 \wedge \omega InstanceOf(instanceOf) = kon_1$,
- (d) $\alpha InstanceOf(instanceOf') = i_2 \wedge \omega InstanceOf(instanceOf') = kon_2$ und
- (e) $\exists relKon \in RelKon : \alpha RelKon(relKon) = kon_1 \wedge \omega RelKon(relKon) = kon_2$.

Wenn keine entsprechende $relKon$ -Beziehung zwischen den beiden Konzepten kon_1 und kon_2 (Bedingung (e)) existiert, so existiert ein Weg $w'_1 = (isA_1, \dots, isA_n)$, $n \geq 1$ und ein Weg $w'_2 = (isA'_1, \dots, isA'_m)$, $m \geq 0$ ⁹. Für w'_1 und w'_2 gilt:

- $\alpha(w'_1) = kon_1 \wedge \omega(w'_1) = kon'_1$,
- $\alpha(w'_2) = kon_2 \wedge \omega(w'_2) = kon'_2$ und
- $\exists relKon' \in RelKon : \alpha RelKon(relKon') = kon'_1 \wedge \omega RelKon(relKon') = kon'_2$.

und ungerichteten Graphen in einem Graphen (G_{Kon}) zu vereinen.

⁹Nach Definition 3.2, muss ein Weg mindestens die Länge 1 ($m \geq 1$) haben. Ich habe hier dennoch nur $m \geq 0$ gefordert, um eine Fallunterscheidung der verschiedenen Fälle $|w'_1| \geq 1 \wedge |w'_2| \geq 1$, $|w'_1| \geq 1$ oder $|w'_2| \geq 1$ zu vermeiden.

Der Graph $G_{RelI} = (I, RelI, \alpha RelI, \omega RelI)$ ist schlingenfrei. D.h., es existiert kein $relI \in RelI$ für das $\alpha RelI(relI) = \omega RelI(relI)$ gilt.

3.2.3 Beispiele

In den Abbildungen 3.6 und 3.7 sind Beispiele für einen Kategorien- und einen Konzeptgraphen gegeben.

Abbildung 3.6 zeigt einen Kategoriegraphen, der zur Klassifizierung von Förderprogrammen nach ihrem Forschungsthema dient. Nach Definition 3.3 besteht der Kategoriengraph aus den beiden Eckenmengen für Kategorien (Kat) und für Informationseinheiten (I). Die Kategorien sind zu einer Hierarchie geformt. Die Wurzelkategorie ist die Kategorie „Alle Forschungsgebiete“. Dies genügt der Bedingung 11 aus Definition 3.3. In Abbildung 3.6 besteht die Menge der Informationseinheiten I aus fünf Elementen („Grako Mathematische Logik und Anwendungen“, „BMBF-Solarthermie“, „BMBF-Supraleitung“, „BMBF-Neue Chipsysteme“ und „Multimediales Buch“). Diese fünf Informationseinheiten werden jeweils von einer Kategorie klassifiziert (Bedingung 9). Die Informationseinheit „BMBF-Supraleitung“ wird von zwei Kategorien, „Physik“ und „Elektrotechnik“, klassifiziert. Dies ist zulässig, da weder „Physik“ noch „Elektrotechnik“ eine Oberkategorie der jeweils anderen Kategorie ist. Damit ist die 10. Bedingung eingehalten. Unzulässig wäre die Klassifikation mit der Kategorie „Naturwissenschaften“ und mit der Kategorie „Physik“, da „Physik“ eine Unterkategorie von „Naturwissenschaften“ ist. Gleichwohl können alle Kategorien entlang der Hierarchie zur Klassifikation benutzt werden, solange Bedingung 10 eingehalten wird. Theoretisch könnte man fordern, dass jede Informationseinheit durch genau eine Kategorie klassifiziert wird. Im Falle der Informationseinheit „BMBF-Supraleitung“ müsste man dann eine neue Kategorie „physikalische Elektrotechnik“ als Unterkategorie von „Elektrotechnik“ und von „Physik“ definieren. Ich habe in Definition 3.3 bewusst darauf verzichtet, da sonst die Zahl der Kategorien stark steigen würde. Würde man dies fordern, so könnte man einen Kategoriegraphen auf einen Konzeptgraphen abbilden, indem man die Kategorien auf die Konzepte des Konzeptgraphen abbildet. Die Pfeilmenge O würde man auf die Pfeilmenge IsA abbilden. Dabei ist zu beachten, dass die Orientierung der Pfeile umzukehren ist. Die Pfeilmenge Kla wird auf die Pfeilmenge $InstanceOf$ abgebildet. Die Orientierung der Pfeile wird ebenfalls umgekehrt.

Abbildung 3.7 zeigt ein Beispiel für einen Konzeptgraphen aus dem Bereich Forschungsförderung. Der Konzeptgraph besteht aus 4 Konzepten („Förderprogramm“, „Ansprechpartner“, „Förderer“ und „Projekträger“). Für jedes der Konzepte existiert jeweils eine Informationseinheit („Multimediales Buch“, „Dr. Peter

Muster”, „BMBF” und „DLR”). Die verschiedenen Informationseinheiten werden jeweils durch einen *instanceOf*-Pfeil zu einem Konzept typisiert. Diese Zuordnung ist nach Bedingung 10 in Definition 3.4 eindeutig. Die (einfachen) Beziehungen zwischen Konzepten (*relKon*-Pfeile) haben nach Bedingung 11 Auswirkungen auf die Beziehung zwischen den zugeordneten Informationseinheiten. Z.B. bedeutet die Beziehung zwischen den Konzepten „Förderprogramm” und „Ansprechpartner”, dass die Informationseinheiten „Multimediales Buch” und „Dr. Peter Muster” eine Beziehung eingehen **können**. Dasselbe gilt für die Beziehung zwischen „Förderprogramm” und „Förderer”, die die Beziehung zwischen „Multimediales Buch” und „BMBF” erlaubt. Die Beziehung zwischen „Multimediales Buch” und „DLR” wird durch die *isA*-Beziehung zwischen „Projektträger” und „Förderer” legitimiert. Dadurch werden alle Beziehungen zwischen „Förderer” und anderen Konzepten auf das Konzept „Projektträger” vererbt. Dies wird im zweiten Teil der 11. Bedingung in Definition 3.4 beschrieben.

3.3 Profil

In diesem Abschnitt wird der Begriff des „Profils” definiert. Der Begriff „Profil” ist eng verwandt mit dem Begriff des „Overlay-Modells”, den ich in Kapitel 2.2.4 eingeführt habe. Ein „Profil” ist ein Overlay-Modell über einem Domänenmodell und beschreibt, wie interessant oder uninteressant einzelne Bestandteile des Domänenmodells für einen Benutzer sind. Dabei wird das Domänenmodell als gerichteter Graph aufgefasst. Bei dem Domänenmodell kann es sich um einen einfachen gerichteten Graphen (Definition 3.1) einen Kategoriengraphen (Definition 3.3) oder einen Konzeptgraphen (Definition 3.4) handeln. In Definition 3.5 wird ein Profil auf Basis eines einfachen gerichteten Graphen beschrieben. In Abbildung 2.1 wird der Zusammenhang zwischen Domänenmodell und Overlay-Modell beschrieben. Diese Abbildung kann als Visualisierung von Definition 3.3 verstanden werden.

Definition 3.5 (Profil)

Ein **Profil** des Graphen $G = (V, R, \alpha, \omega)$ ist ein 8-Tupel $Profil = (V, PItem, R, P, \alpha, \omega, \alpha P, \omega P)$ mit folgenden Eigenschaften:

1. V ist die Eckenmenge des Graphen G . V ist ebenfalls eine Eckenmenge in $Profil$.
2. $PItem$ ist eine Eckenmenge des Graphen $Profil$.
3. R ist die Pfeilmenge des Graphen G . R ist ebenfalls eine Pfeilmenge in $Profil$.

4. P ist eine Pfeilmenge des Graphen $Profil$.
5. Es gilt $V \cap PItem = V \cap R = V \cap P = PItem \cap R = PItem \cap P = R \cap P = \emptyset$.
6. $\alpha : R \rightarrow V$ und $\omega : R \rightarrow V$ sind Abbildungen in G und in $Profil$. $\alpha(r)$ ist die Startecke des Pfeils r . $\omega(r)$ ist die Endecke des Pfeils r .
7. $\alpha P : V \rightarrow PItem$ und $\omega P : V \rightarrow PItem$ sind Abbildungen in $Profil$. $\alpha P(p)$ ist die Startecke des Pfeils p und heißt **Baustein**. $\omega P(p)$ ist die Endecke des Pfeils p und wird **Profilwert** des Bausteins v genannt.
8. Für alle Ecken des Graphen G , existiert genau ein Profilwert. Für alle $v \in V$ gilt:
 - (a) $\exists p \in P : \alpha P(p) = v \wedge \omega P(v) = pItem, pItem \in PItem$.
 - (b) $\forall p_i, p_j : \alpha P(p_i) \neq \alpha P(p_j) \wedge \omega P(p_i) \neq \omega P(p_j), i \neq j$

Ein Profil über einem Graphen, der mehrere verschiedene Eckenmengen hat, definiert jeweils eine eigene Menge von Profil-Items für jede Eckenmenge. Für jede Profil-Itemmenge werden jeweils eigene Pfeilmengen und eigene Abbildungen definiert. Definition 3.6 definiert z.B. ein Profil über einem Kategoriegraphen. Ein Profil über einem Konzeptgraphen kann entsprechend definiert werden.

Definition 3.6 (Profil über einem Kategoriegraphen)

Sei $G_{Kat} = (Kat, I, O, Kla, \alpha O, \omega O, \alpha Kla, \omega Kla)$ ein Kategoriengraph. Ein 14-Tupel $Profil_{Kat} = (Kat, PItem_{Kat}, I, PItem_I, O, P_{Kat}, Kla, P_I, \alpha O, \omega O, \alpha P_{Kat}, \omega P_{Kat}, \alpha Kla, \omega Kla, \alpha P_I, \omega P_I)$ heißt **Profil über dem Kategoriegraphen** G_{Kat} . Es gelten neben den Eigenschaften für den Kategoriegraphen (siehe oben, Definition 3.3) folgende Eigenschaften:

1. $PItem_{Kat}$ ist eine Eckenmenge im Graphen $Profil_{Kat}$. $Profil_{Kat}$ beinhaltet die Profilwerte von Kategorien.
2. $PItem_I$ ist eine Eckenmenge im Graphen $Profil_{Kat}$. $Profil_I$ beinhaltet die Profilwerte von Informationseinheiten.
3. P_{Kat} und P_I sind Pfeilmengen des Graphen $Profil_{Kat}$.
4. Alle Ecken- und Pfeilmengen sind disjunkt.

5. $\alpha P_{Kat} : P_{Kat} \rightarrow Kat$ und $\omega P_{Kat} : P_{Kat} \rightarrow PItem_{Kat}$ sind Abbildungen. $\alpha P_{Kat}(p_{Kat})$ ist die Startecke des Pfeils p_{Kat} . $\omega P_{Kat}(p_{Kat})$ ist die Endecke des Pfeils p_{Kat} und wird Profilwert der Kategorie kat genannt.
6. $\alpha P_I : P_I \rightarrow I$ und $\omega P_I : P_I \rightarrow PItem_I$ sind Abbildungen. $\alpha P_I(p_I)$ ist die Startecke des Pfeils p_I . $\omega P_I(p_I)$ ist die Endecke des Pfeils p_I und wird Profilwert der Informationseinheit i genannt.

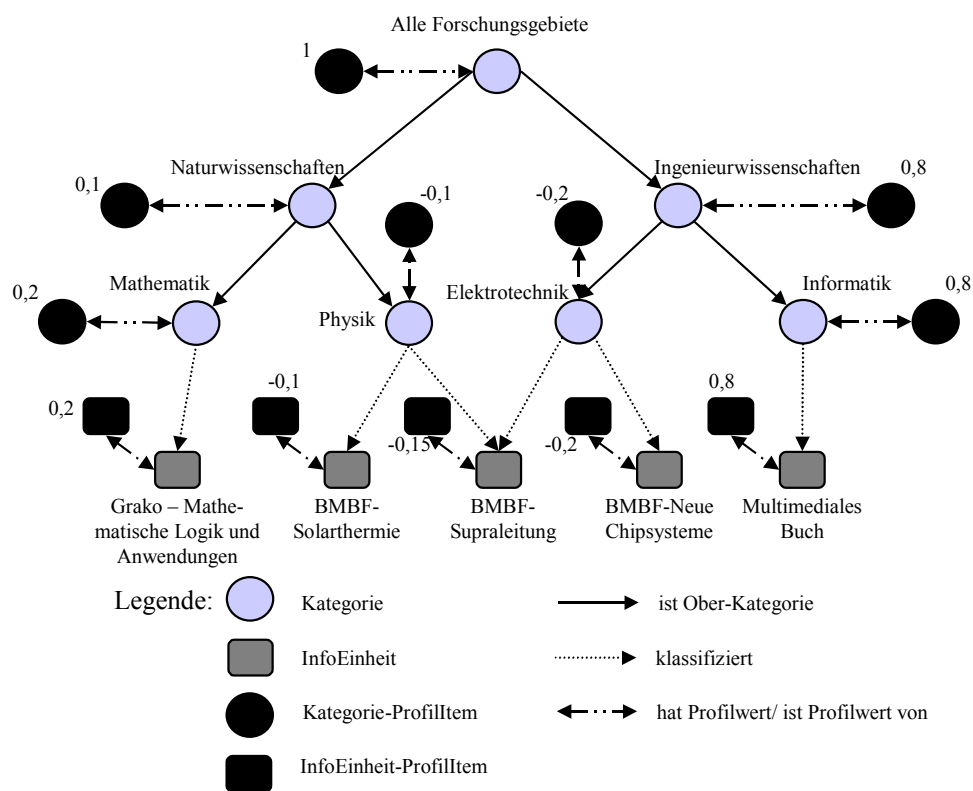


Abbildung 3.8: Profil über einem Kategoriegraphen

In Abbildung 3.8 ist das Beispiel aus Abbildung 3.6, das in Kapitel 3.2.3 erläutert wurde, als Profil dargestellt. Der Kategoriengraph aus Abbildung 3.6 wird dort um Ecken für jede Kategorie und jede Informationseinheit erweitert¹⁰. Bei diesen Ecken handelt es sich um Profil-Items, die mit einem Wert zwischen -1 und $+1$

¹⁰ Abweichend von der Darstellung von Interessenprofilen in den Abbildungen von Brokeringmodellen, werden in Abbildung 3.8 die verschiedenen Werte für einzelne Profilbausteine nicht durch verschiedene Grautöne, sondern durch verschiedene, annotierte Werte ausgedrückt.

belegt werden können¹¹. Das Intervall $[-1, 0[$ steht für uninteressante Items, $]0, 1]$ steht für interessante Items. Bei Profil-Items, die mit dem Wert 0 belegt wurden, ist keine Aussage darüber möglich, wie interessant der dazugehörige Baustein des Domänenmodells ist.

In ähnlicher Weise wie ein Profil über einem Kategoriegraphen kann auch ein Profil über einem Konzeptgraphen definiert werden, indem für jede Eckenmenge des Konzeptgraphen eine Menge von Profil-Items, eine Menge von Pfeilen zwischen den Ecken des ursprünglichen Graphen und den Profil-Items und entsprechende α - und ω -Abbildungen definiert werden.

3.4 Dossier

Der Begriff des „Dossiers“ wurde im Aufgaben- und Rollenmodell für Information-Brokering (vgl. Kap. 3.1) eingeführt. Er bezeichnet das Ergebnis des Brokering-Prozesses, das an den Konsumenten ausgeliefert wird (vgl. Abb. 3.4). Konsumenten sind üblicherweise an Informationseinheiten interessiert. Aus diesem Grund wird ein „Dossier“ als eine endliche Menge von Informationseinheiten definiert.

Definition 3.7 (Dossier)

Sei $Profil_{Kat} = (Kat, PItem_{Kat}, I, PItem_I, O, P_{Kat}, Kla, P_I, \alpha O, \omega O, \alpha P_{Kat}, \omega P_{Kat}, \alpha Kla, \omega Kla, \alpha P_I, \omega P_I)$ ¹² ein Profil über einem Kategoriegraphen. Die endliche Menge $I_{Dossier}$ wird **Dossier** genannt, wenn $I_{Dossier} \subseteq I$ gilt.

3.5 Kategorisierungsprozesse im Information-Brokering

Strukturierung und Personalisierung (siehe Kap. 3.1) sind spiegelbildliche Aufgaben: Bei der Strukturierung werden die Dokumente (und damit die Sicht der Anbieter auf die Domäne) auf das IB-Domänenmodell abgebildet. Bei der Personalisierung hingegen werden die Interessen der Konsumenten auf das IB-Do-

¹¹Die Interessenwerte für Kategorien in Abbildung 3.6 sind rein willkürlich gewählt und dienen lediglich der Verdeutlichung des Sachverhalts. Die Interessenwerte für Informationseinheiten werden aus den Interessenwerten der beteiligten Kategorien errechnet.

¹²Anstelle eines Profils über einem Kategoriegraphen hätte auch ein Profil über einem Konzeptgraphen gewählt werden können.

mänenmodell abgebildet. Kernaufgabe innerhalb des Strukturierungsprozesses ist die Konzeptualisierung und Kategorisierung. Bei der Konzeptualisierung/ Kategorisierung ist zuerst die Frage zu beantworten, ob diese Information in das IB-Domänenmodell gehört und wenn ja, um welche Art von Information es sich handelt¹³. Danach ist diese Information in das IB-Domänenmodell so einzuordnen, dass sie von Informationen desselben Typs unterscheidbar bleibt. Dies ist die Aufgabe der Kategorisierung. Alle Aufgaben im Strukturierungsprozess werden Domänenverwaltungsaufgaben genannt.

Kernaufgabe im Personalisierungsprozess ist die sogenannte Profilbildung (siehe oben, Kap. 3.1). Bei der Profilbildung muss zuerst geklärt werden, für welche Arten von Informationen sich ein Konsument interessiert. Dann können die Interessen für die verschiedenen Informationsarten genauer spezifiziert werden.

Die beiden Unteraufgaben innerhalb von Konzeptualisierung/Kategorisierung und Profilbildung sind untrennbar miteinander verbunden und entsprechen dem Kern- und Identifikationsprozess in der Kategorisierungstheorie von Smith und Medin (vgl. [SM81]). Im Kernprozess wird die Zugehörigkeit zu einem Konzept geklärt. Im Identifikationsprozess wird die Abgrenzung zu anderen Objekten desselben Konzeptes vorgenommen. Eine Trennung der beiden Prozesse lässt sich empirisch nicht belegen. Rosch (vgl. [Ros78]) hat nachgewiesen, dass Objekte mit typischen Eigenschaften von anderen eher zugeordnet werden können, als Objekte mit untypischen Eigenschaften. Außerdem können Objekte durch Abgrenzung zu typischen Gegenbeispielen ebenfalls schneller zugeordnet werden, als „normale“ Objekte. Daraus ist abzuleiten, dass die Eigenschaften von Informationen beschrieben werden sollten, die für die Endkonsumenten der Informationen möglichst typisch sind oder eine gute Abgrenzung untereinander zulassen.

Die jeweiligen Unteraufgaben haben im IB-Domänenmodell (siehe Kap. 6) zur Einführung von *Kategorien* und *Konzepten* geführt. Konzepte sind mit dem Kernprozess verbunden, während die Kategorien den Identifikationsprozess unterstützen.

¹³Dieser Bestandteil der Aufgabe wird als Konzeptualisierung beschrieben (siehe oben, Kap. 3.1).

Kapitel 4

Adaptive Informationssysteme

In diesem Kapitel stehen adaptive Informationssysteme im Mittelpunkt der Betrachtung. Adaptive Informationssysteme versuchen sich automatisch oder mit Hilfe des Benutzers an die Informationsbedürfnisse des Benutzers anzupassen. Dadurch wird versucht, dem Phänomen des Information Overload entgegenzuwirken. Das Problem des Information Overload hat durch die explosionsartige Entwicklung des World-Wide Web an Dringlichkeit zugenommen, da es der effizienten Nutzung der im WWW verfügbaren Informationen entgegensteht. Gleichzeitig ist die Bedeutung von Wissen gestiegen (vgl. [NT97], [DP98], [AL99]), das aus den verfügbaren Dokumenten gewonnen werden könnte.

Adaptive Informationssysteme versuchen möglichst viel über das Informationsbedürfnis ihrer Benutzer herauszufinden, damit sie diese Informationsbedürfnisse effektiv befriedigen können. Damit ist Benutzermodellierung (vgl. [Kob85]) zentraler Bestandteil von adaptiven Informationssystemen. Brusilovsky ([Bru01]) beschreibt, dass parallel zur Entwicklung des WWW die Anzahl der Ansätze und Systeme zur adaptiven Informationsversorgung von Benutzern in den letzten fünf Jahren deutlich gestiegen ist. Viele Ansätze nutzen Methoden aus dem Information-Retrieval (vgl. [SM83]), wie das Vektorraum-Modell (vgl. Kap. 2.3.1), um Benutzerinteressen abzubilden. Diese Ansätze verzichten auf eine strukturierte, modellbasierte Repräsentation von Interessen, um so dem Problem zu entgegen, dass Informationen im WWW üblicherweise über keine definierte Semantik verfügen. Erst in letzter Zeit gibt es Ansätze, die Inhalte von Informationssystemen in konkreten Domänenmodellen¹ zu modellieren und die Interessen von Be-

¹Der Terminus „Domänenmodell“, der in Kap. 2.2.3 definiert wurde ist nicht zu verwechseln mit dem Terminus „Domänenontologie“ oder dem Terminus „IB-Domänenmodell“. Ein *Domänenmodell* ist ein Modell von der Domäne, auf die adaptiert werden soll. Eine *Domänenontologie* ist ein Ontologie im Sinne von Grubers’ Definition (vgl. Kap. 2.1.3 und [Gru93]), die die Domäne beschreibt. Ein *IB-Domänenmodell* ist das zentrale Modell im Information-Brokeringmodell (vgl. Kap. 3) und ist sowohl ein Domänenmodell als auch eine Domänenontologie.

nutzern in Relation zu diesen Modellen zu beschreiben. Da Domänenmodelle für Information-Brokering von zentraler Bedeutung sind (siehe Kap. 3, bzw. Kap. 6), gilt diesen Systemen besondere Aufmerksamkeit. Innerhalb der adaptiven, domänenmodellbasierten Informationssysteme spielen adaptive Lehr-Lernsysteme eine besondere Rolle, da diese im Vergleich zu „normalen“, adaptiven, domänenmodellbasierten Informationssystemen weiter entwickelt sind und erwartet wird, dass „normale“ adaptive, domänenmodellbasierte Informationssysteme die Entwicklung von Lehr-Lernsystemen nachvollziehen werden (vgl. [Bru01]). Die Vorreiterrolle der Lehr-Lernsysteme wurde dadurch begünstigt, dass für Lehr-Lernsysteme im Unterschied zu „normalen“ Informationssystemen der Inhalt im Vorhinein bekannt ist. Dementsprechend können die Inhalte von Lehr-Lernsystemen für ein Domänenmodell aufgearbeitet werden.

Zur Zeit weisen adaptive Systeme eine ungünstige Kosten-Nutzen-Relation auf: Sie unterstützen die Benutzer in ihrer Arbeit, aber dem dadurch gewonnenen Vorteil stehen unverhältnismäßig hohe Kosten für die Realisierung von Adaptivität gegenüber (vgl. [Fis01]). Dies ist der Tatsache geschuldet, dass adaptive Systeme in vielen Fällen für eine bestimmte Domäne entwickelt werden und deswegen nicht leicht übertragbar sind auf andere Domänen. Um die Kosten für die Erstellung adaptiver Systeme zu senken, werden generische Systeme gebraucht, die auf verschiedenste Domänen mit wenig Aufwand angepasst werden können oder die zur leichten Erzeugung solcher adaptiver Systeme benutzt werden können. Im Bereich der adaptiven Lehr-Lernsysteme ist das Interbook-System (vgl. [BSW96]) prototypisch für diesen Ansatz. Mit Hilfe von Interbook können adaptive Lehr-Lernsysteme aus einem speziell annotierten Word-Dokument generiert werden. Ein ähnlich generisches System für adaptive, domänenmodellbasierte Informationssysteme existiert zur Zeit noch nicht, wird aber in den nächsten Jahren erwartet (vgl. [Bru01]). Diese Arbeit versucht mit Broker's Lounge diesen Anspruch einzulösen, indem die adaptiven Eigenschaften von Broker's Lounge an die Basisbausteine zur Beschreibung von Domänen gekoppelt werden und damit in allen Domänen zur Verfügung stehen, die mit diesen Basisbausteinen adäquat beschrieben werden können. Um möglichst viel über diese Basisbausteine zu lernen, werden existierende adaptive, domänenmodellbasierte Systeme unter diesem Aspekt untersucht.

Im weiteren Verlauf ist dieses Kapitel nach dem Strukturierungsgrad des benutzten Domänenmodells gegliedert. Im nächsten Abschnitt (Kap. 4.1) werden zunächst termbasierte Informationssysteme beschrieben, die im wesentlichen auf einem Vektorraummodell basieren. Die Domänenmodelle von termbasierten Informationssystemen weisen nur eine schwache Strukturierung auf. Im Anschluss werden kollaborative Filtersysteme (Kap. 4.2) beschrieben. Diese Systeme nutzen nicht direkt ein Domänenmodell, sondern die Dialoghistorie einer Vielzahl von Benutzern als Quelle für Anpassungsleistungen. Darauf folgt eine kurze Ein-

führung in domänenmodellbasierte Benutzermodellierung (vgl. Kap. 4.3). Als Vorreiter von domänenmodellbasierten, adaptiven Systemen werden dann in Kapitel 4.4 zunächst Lehr-Lernsysteme beschrieben. Lehr-Lernsysteme basieren auf einem strukturierten Domänenmodell, das in vielen Fällen das zu lernende Curriculum abbildet. Zwar versuchen Lehr-Lernsysteme in erster Linie den Wissensstand und nicht die Interessen des Benutzers zu modellieren, dennoch sind Lehr-Lernsysteme prägend für domänenmodellbasierte, adaptive Informationssysteme (vgl. [Bru01]). Ausgewählte domänenmodellbasierte Informationssysteme werden in Kapitel 4.5 beschrieben. Die Strukturierungsmittel, die in den beschriebenen domänenmodellbasierten Informationssystemen benutzt wurden, werden in der Zusammenfassung (Kap. 4.6) am Ende dieses Kapitels in einer Tabelle aufgelistet. Dort werden die verschiedenen Systeme und Ansätze gegenübergestellt, um Rückschlüsse auf Strukturierungsmittel für das IB-Domänenmodell, das in Kapitel 6 beschrieben wird, zu gewinnen.

4.1 Termbasierte adaptive Informationssysteme

Viele adaptive Informationssysteme kombinieren Methoden aus dem Information-Retrieval und der Benutzermodellierung. Diese Ansätze basieren auf dem Vektorraum-Modell (TFIDF-Verfahren, vgl. [SM83] und Kap. 2.3.1) oder einem ähnlichen Verfahren. Bei diesen Systemen werden Dokumente durch einen Vektor von Termen repräsentiert. Terme sind wichtige Wörter oder Wortgruppen. Die einzelnen Terme werden abhängig vom verwendeten Verfahren gewichtet. Die Benutzermodelle werden als Overlay-Modell über das Vektorraum-Modell gelegt oder die Nutzung von Dokumenten (durch den Benutzer) wird direkt in das Vektorraummodell integriert. Da diese Modelle für den Benutzer nicht inspizierbar sind, handelt es sich bei den Systemen fast ausnahmslos um rein adaptive Systeme. In allen Systemen haben die Benutzer keine direkte Kontrolle ihres Benutzerprofils, sondern können nur durch explizites Feedback oder die Nutzung des Systems Einfluss auf ihr Profil ausüben. Die adaptive Eigenschaft fast aller Systeme besteht darin, ihren Benutzern Dokumente vorzuschlagen, die sie interessieren könnten. Dies kann im Kontext eines konkreten Dokumentes geschehen, indem in diesem Dokument die Links durch eine entsprechende Markierung hervorgehoben werden, die zu für den Benutzer relevanten Dokumenten führen². Andere Systeme versuchen, das generelle Informationsbedürfnis ihrer Benutzer durch die Zuhilfenahme weiterer Quellen, wie z.B. Suchmaschinen, zu befriedigen.

Letizia (vgl. [Lie95]) ist ein System, das viele andere Forscher auf diesem Gebiet inspiriert hat. Es beobachtet den Benutzer beim Webbrowsen und erstellt aus

²Außerdem können auch irrelevante und indifferente Linke gekennzeichnet werden.

den Dokumenten, die ein Benutzer angeschaut hat, einen Dokumentenindex auf TFIDF-Basis. Wenn ein Benutzer sich ein neues Dokument anschaut, werden von einem Agenten alle Links dieses Dokumentes bzw. die dahinterliegenden Dokumente gegen den bestehenden Dokumentenindex geprüft und bewertet. Die relevantesten Links werden dem Benutzer empfohlen. Letizia kommt ohne explizites Feedback aus, da es nur auf Basis der vom Benutzer angeschauten Dokumente arbeitet. Inzwischen wurde Letizia um ein kollaboratives Szenario erweitert (vgl. [LDV99]), bei dem mehrere Benutzer vor ein und demselben Bildschirm sitzen und gemeinschaftlich browsen. Damit die einzelnen Benutzer identifiziert werden können, hat jeder Benutzer ein Abzeichen (engl. badge). Diese Abzeichen werden von einem Sensor identifiziert. Vor einer Session wird für jeden Benutzer ein Benutzerprofil angelegt. Dazu werden seine Homepage und die von dort referenzierten Seiten von einem Agenten mit dem TFIDF-Verfahren untersucht. Die 50 wichtigsten Begriffe bilden sein Profil. Bei einer Gruppensession werden die verschiedenen Profile gemischt. Das so entstehende Gruppenprofil wird zur Browsing-Unterstützung (ähnlich wie in der alten Letizia-Version) benutzt. D.h., ein Agent untersucht benachbarte Webseiten des aktuellen Dokumentes und bewertet sie anhand des Gruppenprofils. Findet der Agent relevante Seiten, so werden diese der Benutzergruppe vorgeschlagen. Zusätzlich wird eine Erklärung gegeben, warum der Vorschlag gemacht wurde und für wen innerhalb der Gruppe der Vorschlag interessant sein könnte.

Ein weiteres System, dass viele Forscher inspiriert hat, ist WebWatcher (vgl. [JFM97]). WebWatcher arbeitet als Proxy und schlägt auf einer Seite ggf. Links vor, die ein Benutzer als nächstes anwählen sollte. Dazu werden diese Links annotiert. Zu Beginn einer Session gibt ein Benutzer ein oder mehrere Suchworte an. Diese Suchworte werden dann mit den Texten der verschiedenen verfügbaren Links einer Seite in Beziehung gesetzt. WebWatcher basiert auf dem TFIDF-Algorithmus. Die daraus resultierende Bewertung führt zusammen mit einer Analyse der Linkstruktur und der Popularität der verschiedenen Links (wie oft wurde ein bestimmter Link von allen Benutzern ausgewählt) zu einer Bewertung der verschiedenen Links. Die Links mit der höchsten Bewertung, die über einem bestimmten Schwellwert liegen, werden vorgeschlagen. Es werden höchstens drei Links empfohlen.

Syskill & Webert (vgl. [PMB96]) benötigt im Gegensatz zu WebWatcher und Letizia explizites Feedback. In Syskill & Webert können für verschiedene Themen jeweils Benutzerprofile angelegt werden. Diese Benutzerprofile bestehen aus einer Menge von Schlüsselwörtern. Diese Schlüsselwörter sind für das jeweilige Thema die informativsten Wörter. D.h., diese Wörter kommen wesentlich häufiger in relevanten als in irrelevanten Dokumenten vor. Damit diese informativsten Wörter definiert werden können, müssen Benutzer zu Beginn eine Reihe von Dokumenten nach Relevanz bewerten. Diese Dokumentenbewertungen werden be-

nutzt, um die informativsten Wörter zu finden. Die informativsten Wörter werden nach Abschluss der Trainingsphase genutzt, um neue Dokumente zu bewerten, indem sie einen Termvektor bilden. Mit diesem Termvektor wird festgestellt, ob ein gegebenes Wort in einem Dokument enthalten ist. Dieser Termvektor ist ein Standardeingabeformat für verschiedenste Machine-Learning-Verfahren. In Syskill & Webert wird ein naiver Bayes-Klassifikator benutzt, um vorherzusagen, ob ein neues Dokument für einen Benutzer interessant ist oder nicht. Die Dokumentenbewertungen zu Beginn dienen neben dem Auffinden der informativsten Wörter auch der Initialisierung des naiven Bayes-Klassifikators mit den entsprechenden bedingten Wahrscheinlichkeiten.

NewsDude (vgl. [BP99], [Bil99]) basiert auf einem ähnlichen Mechanismus wie Syskill & Webert. NewsDude klassifiziert Nachrichten und liest sie seinen Benutzern vor. Die Benutzer sind gehalten, die vorgelesenen Dokumenten mit explizitem Feedback zu bewerten. NewsDude benutzt wie Syskill & Webert einen naiven Bayes-Klassifikator für Langzeitinteressen, während ein Kurzzeit-Profil auf einem Nearest-Neighbor-Algorithmus für das Vektorraum-Modell basiert.

Amalthea (vgl. [Mou97]) basiert auf einem gewichteten Vektorraummodell, das durch die Beobachtung der Webnutzung des Benutzers gefüllt wird. Im Unterschied zu den anderen hier vorgestellten Systemen hat Amalthea nicht genau ein Profil, sondern arbeitet mit einem genetischen Algorithmus. Das System hat eine Menge von Profilen, die von Filter-Agenten verwaltet werden. Die Profile werden mit einer Fitnessfunktion bewertet, Die „fittesten“ Profile dürfen dem Benutzer neue Dokumente vorschlagen. Außerdem dürfen die „fittesten“ Profile durch eine Cross-Over-Funktion neue Profile bilden. Die Fitnessfunktion basiert auf explizitem und implizitem Feedback des Benutzers.

Neben Systemen zur Browsing-Unterstützung gibt es auch einige Systeme, die als adaptive Benutzerschnittstellen zu Suchmaschinen konzipiert sind. Das HUMOS- (Hybrid User Modelling Subsystem) und WIFS-System (Web-oriented Information Filtering Subsystem, vgl. [ACM97]) bilden in Kombination ein solches adaptives Frontend zur Suchmaschine AltaVista. Das System arbeitet mit Stereotypen, die auf Basis eines fallbasierten Mechanismus den Benutzern zugeordnet werden. Für die Zuordnung zu Stereotypen müssen Benutzer bei der Anmeldung zusätzlich ein Eingangsinterview ausfüllen. Im weiteren Verlauf werden die Benutzerprofile mit einem neuronalen Netz bearbeitet. Das Benutzerprofil basiert auf den Seiten, die ein Benutzer über das System ausgewählt hat. Um das Benutzerprofil zu aktualisieren, geben die Benutzer explizites Feedback.

In neueren Publikationen ist ein deutlicher Trend zu mehr strukturierten Benutzermodellen für schlüsselwortbasierte, adaptive Informationssysteme erkennbar. So wurde z.B. SiteIF (vgl. [SS98]) um eine Strukturkomponente erweitert, die eine eindeutigere Zuordnung zu verschiedenen Domänen erleichtert. Dies verbessert sowohl die Präzision als auch die Abdeckung (Recall) des neuen Systems

gegenüber dem alten System (vgl. [MS01]). SiteIF lernt durch Beobachtung der Webnutzung des Benutzers. Die besuchten Webseiten werden analysiert. Die Ergebnisse werden als semantisches Netz abgelegt. Das semantische Netz besteht im wesentlichen aus verschiedenen Themen. Die Analyse und die Zuordnung zu bestimmten Themen wird durch die Benutzung von WordNet (vgl. [Mil90]) bewerkstelligt.

Das METIORE-System (vgl. [BD01]) ist ein Information-Retrieval-System, das sich an seine Benutzer anpasst. METIORE gibt Zugang zur Bibliothek in einem Forschungsinstitut. Benutzer können Anfragen an das System stellen. Diese Anfragen können einzelne Wörter und weitere Attribute z.B. Autor, Jahr usw. sein. Das System sucht auf Basis dieser Anfrage und bewertet die Ergebnisse mit einem naiven Bayes-Klassifikator, bei dem jeweils alle möglichen Ausprägungen eines Attributs benutzt werden. METIORE ist vergleichbar mit Syskill & Webert (vgl. [PMB96]), wenngleich der Featurevektor, der als Eingabe für den naiven Bayes-Klassifikator dient, anders ermittelt wird und nicht nur aus Wörtern besteht, sondern auch alle möglichen Ausprägungen für verschiedene Attribute als Features benutzt. Wie in Syskill & Webert benötigt das System explizites Feedback vom Benutzer.

Das METIOREW-System (vgl. [BCD01]) verwendet sowohl inhaltsbasierte als auch kollaborative Methoden zur Benutzermodellierung. Es ist eng verwandt mit dem METIORE-System. METIOREW ist im Gegensatz zu METIORE als Informationssystem für das WWW konzipiert. Da die Webinhalte im allgemeinen keine vorgegebene Inhaltsstruktur besitzen, die zur Interessenmodellierung ausgenutzt werden könnten, besteht der dem darunterliegenden Lernverfahren angebotene Featurevektor nur aus Schlüsselwörtern. Um die fehlende inhaltliche Struktur auszugleichen, benutzt METIOREW kollaborative Elemente. Zu Beginn einer Sitzung muss ein Benutzer sein Informationsbedürfnis, ähnlich wie in Web-Watcher (vgl. [JFM97]), in einigen Suchwörtern ausdrücken. Diese Suchwörter werden mit anderen ähnlichen Suchanfragen von anderen Benutzern verglichen. Ggf. werden dadurch neue Suchwörter in die Anfrage integriert. Die resultierende Anfrage wird an eine Suchmaschine weitergeleitet und die Ergebnisse mit dem Benutzerprofil bewertet. Diese Dokumente bekommt der Benutzer auf Anfrage mitgeteilt. Im Gegensatz zu WebWatcher und Letizia benötigt das System explizites Feedback durch den Benutzer.

Brokeringmodell von termbasierten, adaptiven Informationssystemen

Termbasierte, adaptive Informationsfilterungssysteme arbeiten mit Benutzerprofilen, die aus einer Reihe von häufig vorkommenden Wörtern bestehen. Diese Ansätze benötigen kein explizites Domänenmodell. Sie arbeiten stattdessen direkt auf dem Quellenmodell. Das Quellenmodell wird vom System selbst verwaltet.

Das Benutzerprofil wird als Overlay-Modell über das Quellenmodell realisiert.

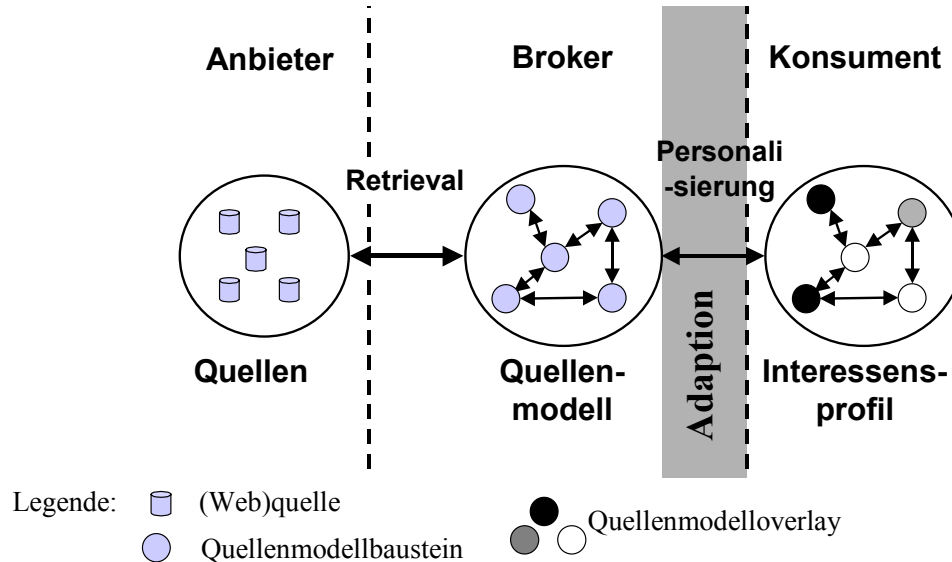


Abbildung 4.1: Brokeringmodell für termbasierte, adaptive Informationssysteme

Im kundenorientierten Prozess versuchen alle Ansätze, die Interessen der Benutzer zu erlernen, um dies zur Filterung einzusetzen. Das so erstellte Interessensprofil wird benutzt, um entweder Informationen auszufiltern, die als uninteressant für den Benutzer erkannt wurden oder besonders relevante Informationen zu empfehlen. In den meisten Fällen handelt es sich um rein adaptive Systeme, da die einzige Einflussmöglichkeit des Benutzers darin besteht, explizites Feedback zu geben. Dies bedeutet, dass die Profilbildung vollständig vom System übernommen wird. Insgesamt ergibt sich für termbasierte, adaptive Informationssysteme das Brokeringmodell, das in Abbildung 4.1 dargestellt wird.

4.2 Kollaborative Filtersysteme

Wie schon zuvor beschrieben, haben benutzermodellierende Systeme vor allem zu Beginn der Interaktion mit einem Benutzer Probleme, sich an ihn anzupassen, da das System kaum Informationen über den Benutzer hat. Dies kann zum einen durch die Nutzung von Stereotypen verbessert werden. Die andere Möglichkeit besteht darin, Informationen von ähnlichen Benutzern zu integrieren. Dieser kollaborative Ansatz erlaubt es dem System, schneller Vorschläge für relevante Items zu unterbreiten, lässt jedoch keine Aussage über neue zu bewertende Items zu

(vgl. [ZA01]). Dies bedeutet, dass kollaborative Lernverfahren nicht für Pushsysteme geeignet sind, da sie nicht direkt auf Neueintragungen in der Datenbasis reagieren können. Aus diesem Grund entfällt eine Darstellung des Brokeringmodells von kollaborativen Filtersystemen. Gleichwohl werden an dieser Stelle kollaborative Filtersysteme aufgeführt, da sie eine wichtige Stellung innerhalb adaptiver Informationssysteme einnehmen. Auch kommerziell haben kollaborative Filtersysteme eine breite Anwendung gefunden.

Kollaborative, adaptive System sind eng verwandt mit Recommender-Systemen, wie sie in [RV97] beschrieben werden. Im allgemeinen suchen adaptive Recommender-Systeme Benutzer mit ähnlichen Interessen oder einer ähnlichen Dialoghistorie (mit dem System) und vergleichen die Dialoghistorie des aktuellen Benutzers mit der Dialoghistorie von anderen Benutzern. Dokumente oder Items, die der aktuelle Benutzer noch nicht angeschaut hat, jedoch von ähnlichen Benutzern angeschaut wurden, werden diesem Benutzer empfohlen. Prototypisch für einen solchen Ansatz ist das Ringo-System, das Musik-CDs empfiehlt (vgl. [SM95]). Im E-Commerce finden solche adaptiven, kollaborativen Methoden Verwendung. Der Internet-Buchhändler Amazon³ benutzt ein ähnliches Verfahren zur Empfehlung von Büchern (vgl. [SKR01]).

GroupLens (vgl. [KMM⁺97]) ist ein Recommender-System für Usenet-News. Im Gegensatz zu Ringo benutzt GroupLens keine explizite Bewertung oder Empfehlung durch seine Benutzer. Stattdessen werden implizite Empfehlungen wie die Lesezeit benutzt.

Tapestry (vgl. [GNOT92]) ist ein kollaboratives Email- und Newsfilter-System. Benutzer können Mails und News annotieren. Ihre Annotationen werden zusammen mit allen Dokumenten in einer Datenbank abgelegt. Die Dokumente und Annotationen können von dort abgerufen werden. Benutzer können Filter definieren, die zur regelmäßigen Filterung benutzt werden. Diese Filter können auch soziale Aspekte einschließen: So kann man z.B. gezielt nach den Annotationen einzelner Benutzer filtern. Tapestry kann im Gegensatz zu GroupLens und Ringo nicht als adaptives System bezeichnet werden, da die Filter vom Benutzer selbst definiert werden müssen und er damit keine Informationen bekommt, die auf Initiative des Systems zurückgehen. Tapestry ist somit ein rein adaptierbares System.

Annotate (vgl. [GK99]) überträgt diesen Ansatz der sozialen Filterung auf Webdokumente. Mit Annotate können Webdokumente annotiert und bewertet werden. Diese Annotationen und Bewertungen können bei Anfragen an das System zur Filterung eingesetzt werden. Zur weiteren Filterung können soziale Filter definiert werden. Außerdem kann man mit (formalen) Metainformationen⁴ weiter

³<http://www.amazon.com>

⁴Formale Metainformationen sind z.B. Informationen über den Autor, oder das Datum der letzten Änderung. Formale Metainformationen sagen nichts über die **Inhalte** der beschriebenen Information aus, sondern nur über das Dokument, in dem die Inhalte beschrieben wurden.

filtern. Um Diskussionsprozesse in Gang zu bringen, sind verschiedene Stufen von Sichtbarkeit möglich.

Web Places (vgl. [MB98]) erweitert die soziale Filterung um den Aspekt der Awareness von ähnlichen Benutzern. Web Places arbeitet als Proxy und vergleicht darüber die Pfade, die Benutzer durch das WWW gewählt haben. Das System kann anzeigen, wenn Benutzer mit ähnlichen Interessen auf derselben Seite sind oder vor kurzem dort waren. Dies dient der Identifikation von ähnlichen Benutzern und kann zur Definition von sozialen Filtern genutzt werden. Außerdem können die Navigationspfade der verschiedenen Benutzer benutzt werden, um einzelne Links entsprechend zu annotieren.

4.3 Domänenmodellbasierte Benutzermodellierung

Wie schon in der Einleitung dieser Arbeit beschrieben, ist Domänenmodellierung ein wichtiger Aspekt von Information-Brokering. Um eine effiziente Personalisierung von Brokering-Inhalten durchführen zu können, sollte das Interessenmodell leicht auf das darunterliegende IB-Domänenmodell abbildbar sein. Aus diesem Grund beschäftigt sich der Rest dieses Kapitels mit domänenmodellbasierter Benutzermodellierung. Im Gegensatz zu termbasierten Ansätzen (vgl. Kap. 4.1), die üblicherweise eine flache Struktur von Einträgen im Benutzermodell aufweisen, benutzen domänenmodellbasierte Ansätze ein strukturiertes Domänenmodell, auf das die Interessen abgebildet werden können. Da im WWW lange (semantisch) unstrukturierte Informationen vorherrschten, hat die Arbeit an domänenmodellbasierten Informationssystemen für das WWW erst begonnen. Viel weiter fortgeschritten sind domänenmodellbasierte Ansätze für adaptive Lehr-Lernsysteme. In Lehr-Lernsystemen werden nicht Interessen von Benutzern, sondern deren Wissen modelliert. Lehr-Lernsysteme haben den Vorteil, dass die Lerninhalte im Vorhinein aufbereitet werden und sie somit mit einer Struktur versehen werden können, die zur Benutzermodellierung ausgenutzt werden kann.

Trotz der Unterschiede zwischen adaptiven Lehr-Lernsystemen und adaptiven Informationssystemen wird erwartet, dass adaptive, domänenmodellbasierte Informationssysteme die Entwicklung von Lehr-Lernsystemen mit einigem zeitlichen Abstand nachvollziehen werden (vgl. [Bru01]). Aus diesem Grund werden im nächsten Abschnitt zunächst adaptive Lehr-Lernansätze beleuchtet, bevor im darauf folgenden Abschnitt existierende Ansätze von adaptiven, domänenmodellbasierten Ansätzen beschrieben werden.

4.4 Lehr-Lernsysteme

Bei der Betrachtung der Domänenmodelle von Lehr-Lernsystemen ist zu berücksichtigen, dass die meisten Lehr-Lernsysteme in ihrem Domänenmodell nicht die **Inhalte** einer Domäne beschreiben, sondern stattdessen die Struktur der verwendeten Lehr-/Lernmaterialien oder eines Curriculums abbilden.

Stilbildend für viele weitere Ansätze ist Interbook (vgl. [BSW96]) sowie die darauf aufbauenden Ansätze von Specht (vgl. [Spe98]). Interbook ist kein reines Lehr-Lernsystem, sondern es handelt sich vielmehr um einen Generator, mit dessen Hilfe konkrete, adaptive Lehr-Lernsysteme erstellt werden können. Um ein neues Lehr-Lernsystem anzulegen, kodiert man alle Konzepte, die gelernt werden sollen, und die ihnen zugeordneten Inhalte in einem Word-Dokument. Zu jedem Konzept schreibt man ein eigenes Kapitel. Zu jedem Konzept (Kapitel) beschreibt man mit einer festgelegten Syntax die Vorbedingungen (Konzepte, die vorher gelernt werden müssen) und die Nachbedingungen (durch das Lesen des Dokumentes gelernte Konzepte). Um ein lauffähiges Tutoringsystem zu erhalten, interpretiert man das so entstandene Word-Dokument mit einem speziellen Parser. Das Domänenmodell besteht damit im Wesentlichen aus einer Menge von Konzepten, die über Vor- und Nachbedingungen miteinander verknüpft sind. Der Benutzer bekommt Unterstützung vom System, indem die Links des aktuellen Dokumentes auf Basis des Benutzermodells markiert oder ausgeblendet werden. Neben einfachen Textdokumenten, die zur Information der Benutzer benutzt werden können, führt Specht (vgl. [Spe98]) weitere Lernmaterialien ein und erlaubt die Definition von komplexeren Curricula. Er benutzt u.a. Tests, um zu überprüfen, ob ein konkretes Konzept vom Benutzer verstanden wurde. Dies erlaubt die validere Beschreibung des Wissensstands des Benutzers. Der individuelle Wissensstand kann dann durch die Abbildung auf das Curriculum visualisiert werden. Der Benutzer erhält dadurch die Möglichkeit, sich über seinen Kenntnisstand zu informieren und ggf. Fehlwahrnehmungen des Systems zu korrigieren. Das System ist damit sowohl ein adaptives als auch ein adaptierbares System, da es die Korrektur des Benutzerprofils erlaubt.

Da Silva et al. (vgl. [dvDDO98]) übernehmen die grundsätzliche Einteilung von Konzepten und Dokumenten von Interbook: Dokumente erklären Konzepte. Die Nachbedingungen, die gelten, nachdem ein Benutzer ein Dokument angeschaut hat, werden jedoch um einen „Level of expertise“ für das jeweilige Konzept erweitert. Jedes Dokument wird vom Autor mit einem „Level of difficulty“ annotiert. Dieser „Level of difficulty“ erhöht zum einem den „Level of expertise“ eines Benutzers, nachdem er das Dokument gelesen hat, und kann zum anderen benutzt werden, um ein an das Vorwissen des Benutzers angepasstes Dokument für ein zu lernendes Konzept auszuwählen.

Süß und Freitag (vgl. [SF99]) erlauben neben Vor- und Nachbedingungen die

Definition weiterer Relationstypen zur Beschreibung der Beziehung zwischen verschiedenen Lerninhalten. Die Relationstypen und Lerninhalte können in einer auf XML basierenden Modellierungssprache beschrieben werden.

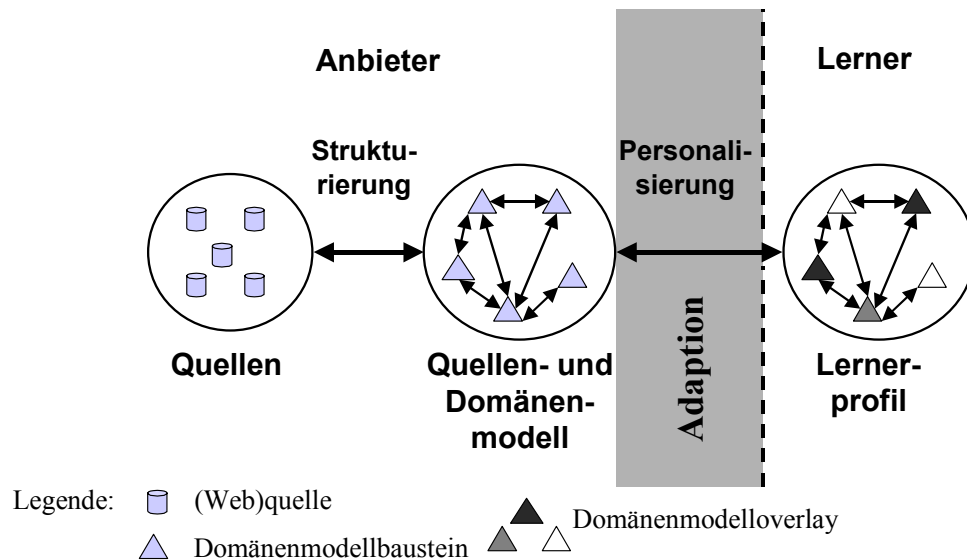


Abbildung 4.2: Push-Broker-Modell für „normale“ Tutoringsysteme

Während in den bisher vorgestellten Ansätzen das Curriculum und die zu vermittelnden Lerninhalte immer zusammenfielen, schlägt Henze in ihren Arbeiten (vgl. [HN99b], [HN99a], [Hen01b]) vor, ein pädagogisches Modell und ein Domänenmodell zu unterscheiden. Die verschiedenen Modelle wurden auf Basis von ConceptBase (vgl. [JGJS95]) realisiert. Das Domänenmodell besteht aus sog. Knowledge-Items, die miteinander in Beziehung stehen. Die Knowledge-Items sind mit Informationseinheiten verknüpft, die die jeweiligen Knowledge-Items beschreiben. Die Informationseinheiten haben wiederum Verbindungen zu anderen Informationseinheiten. Knowledge-Items und Informationseinheiten bilden zusammen das pädagogische Modell. Knowledge-Items und Informationseinheiten werden zu Projekten zusammengefasst, die jeweils verschiedene Lernziele verfolgen. Die Beziehung zwischen Projekten und Knowledge-Items ist zusätzlich durch eine Prozentrelation gekennzeichnet. Für jede Beziehung ist anzugeben, wie groß der Anteil des jeweiligen Knowledge-Items für den Lernerfolg des Projektes ist. Diese Relationen werden als Eingabe für ein Bayesches Netz benutzt, mit dem das Lernermodell verwaltet wird. Das Lernermodell ist als Overlay-Modell über den Knowledge-Items definiert. Wissen über den Lerner wird durch explizites Feedback gewonnen. Basierend auf der momentanen Belegung des Lernermodells, den Abhängigkeiten im pädagogischen Modell und den Beziehungen

zwischen den Projekten und den Knowledge-Items, können geführte Touren individuell berechnet werden.

Brokeringmodell von Lehr-Lernsystemen

Bei Lehr-Lernsystemen herrscht generell ein Push-Broker-Ansatz vor (siehe Abb. 4.2), da Broker und Anbieter zusammenfallen. Dies ist notwendig, da die Inhalte für die Informationskonsumenten, die in diesem Fall Lernende (kurz: Lerner) sind, aufbereitet werden müssen. Die kundenorientierten Aufgaben werden ebenfalls vollständig vom Lehr-Lernsystem ausgeführt, da es zu jedem Zeitpunkt versucht, den Kenntnisstand des Lerners mit einer adaptiven Komponente abzuschätzen und das Informationsangebot entsprechend anzupassen. In vielen Fällen gibt es keine Unterscheidung zwischen einem Inhaltsmodell, das die Inhalte beschreibt, die gelernt werden sollen und einem Quellenmodell, das beschreibt, wo die Inhalte zu finden sind. Eine Ausnahme bildet KBS-Hyperbook. Dort werden die Quellen und der Inhalt in verschiedenen Modellen gehalten. Es ergibt sich für KBS-Hyperbook das Brokeringmodell in Abbildung 4.3.

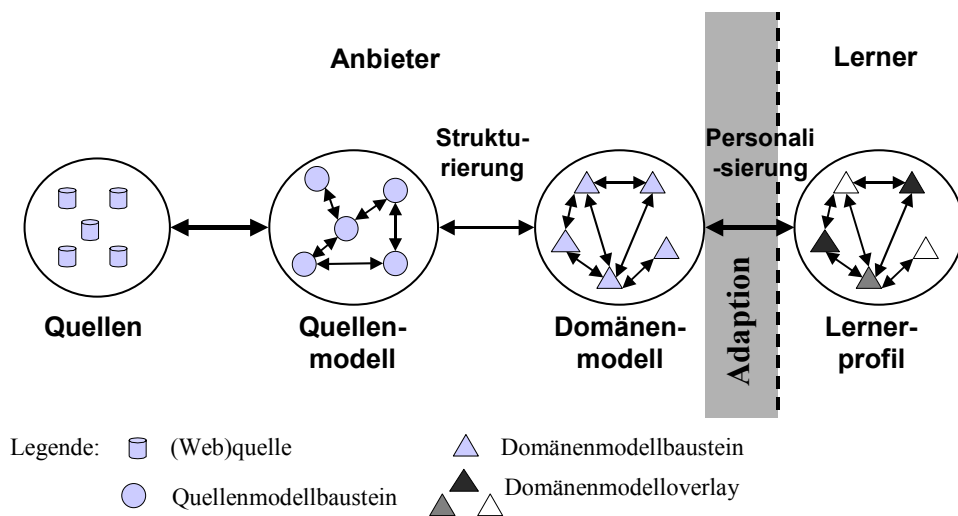


Abbildung 4.3: Push-Broker-Modell für KBS-Hyperbook

4.5 Domänenmodellbasierte Informationsfilterungssysteme

Die Grundarchitektur von domänenmodellbasierten Informationsfilterungssystemen besteht darin, auf dem Domänenmodell ein Interessenprofil als Overlay-

Modell zu beschreiben, das gleichzeitig zur effizienten Filterung benutzt werden kann. Diese Grundarchitektur wurde in [KK96] vorgeschlagen und seitdem von vielen anderen Ansätzen aufgegriffen. Die verschiedenen Ansätze unterscheiden sich durch

- die Struktur des Domänenmodells (und damit auch der Struktur des Interessenprofils),
- den Informationstypen, an den adaptiert wird (Adaptionsziel),
- die Art wie die Domänenmodelle mit Inhalt gefüllt werden und
- die Interaktion zwischen dem System und den Benutzern.

In Tabelle 4.1 am Ende dieses Kapitels werden die in diesem Abschnitt beschriebenen Ansätze und Systeme noch einmal zusammenfassend hinsichtlich der oben beschriebenen Dimensionen beschrieben.

Im EdInfo-Projekt (vgl. [RWH97], [HRW97]) besteht das Domänenmodell aus einer Reihe von sog. Buzzwords. Diese Menge der Buzzwords kann jederzeit erweitert werden. Editoren annotieren Dokumente, die sie in das System einstellen, mit diesen Buzzwords. Die Benutzer können ihre Interessenprofile auf der Basis der Buzzwords definieren und bekommen die Dokumente, die durch ihren Interessenfilter passen. Ein Demonstrator des EdInfo-Projektes ist ConCall (vgl. [WATR99]). ConCall verwaltet Informationen von Konferenzen und den entsprechenden Aufrufen zur Einreichung von Tagungsbeiträgen (kurz: Konferenzaufrufe). Die Editoren fügen Konferenzen und Konferenzaufrufe in das System ein und annotieren sie mit entsprechenden Buzzwords. Die Benutzer bekommen über ihr Profil Zugang zu den Konferenzaufrufen, die zu ihren Interessen passen. Unter den Benutzern von ConCall wurde eine Befragung über das System und gewünschte Erweiterungen gemacht. Ergebnisse der Benutzerbefragung waren, dass die Benutzer eine weitere Strukturierung der Buzzwords, verschiedene Typen von Buzzwords und deren Kategorisierung wünschen, um basierend darauf weitere Filter definieren zu können. Einige Benutzer sind bereit, auch Konferenzaufrufe zu editieren, allerdings fehlten zur damaligen Zeit (1999) gute Tools zum Editieren.

Ein großes Problem, mit dem viele adaptive Informationssysteme zu kämpfen haben, ist die Frage, wie man beliebige Informationen auf ein Domänenmodell und damit mittelbar auf Interessen abbilden kann. Um diesem Problem zu entgegen, beschränken sich einige Ansätze auf leicht erfassbare Meta-Daten, auf die adaptiert wird. So verwendet z.B. das MIRADOR-Projekt (vgl. [BCME00]) formale Metadaten der Dokumente (z.B. Autor, letzte Änderung,...) als Domänenmodell. Konsequenterweise werden diese Metainformationen zusammen mit den

Dokumenten angezeigt. Die Auswahl der angezeigten Metainformationen obliegt dem Benutzermodell, das hier nicht aus dem Inhalt der angezeigten Informationen, sondern aus den formalen Meta-Daten besteht. Neben der oben beschriebenen Anzeigefunktionalität kann das Benutzermodell auch zur Verfeinerung von Suchanfragen benutzt werden.

Timm et. al. (vgl. [TR98]) beschäftigen sich mit Elektronischer Produktberatung, die um eine Benutzermodellierungskomponente erweitert wurde. Aufgabe der Benutzermodellierungskomponente ist es, den Kunden, dem Benutzer, Produkte zum Kauf vorzuschlagen. Durch die Wahl dieses Szenarios entfällt die Notwendigkeit, Informationen auf das Domänenmodell abzubilden, da auf eine existierende Produktdatenbank zurückgegriffen werden kann. Das Domänenmodell sieht die Strukturierung von Produkten in hierarchisch organisierten Produktgruppen vor. Die einzelnen Produktgruppen haben jeweils Produktmerkmale. Zusätzlich wurden alle Produkte mit Kundenmerkmalen wie Umweltverträglichkeit und Preissegment annotiert, von denen man annimmt, dass sie eine Rolle für die Kaufentscheidung spielen. Um diese Kundenmerkmale quantifizieren zu können, werden sie mit einem Wert zwischen 0 und 1 abgelegt. Das Benutzermodell wird wiederum als Overlay-Modell über das Domänenmodell definiert. Das Benutzermodell kann entweder explizit (vom Benutzer) angegeben werden oder wird implizit aus der Kaufhistorie abgeleitet. Der Benutzer hat jederzeit die Möglichkeit, sein Benutzermodell zu inspizieren und zu verändern. Betrachtet man diesen Ansatz im Hinblick auf benutzte Basisbausteine für ein domänenunabhängiges Domänenmodell, so bleibt festzuhalten, dass auf der einen Seite eine hierarchische Struktur zur Beschreibung der Produktgruppen benutzt wurde und dass die existierenden Produktbeschreibungen um Kundenmerkmale erweitert wurden. In [HTS98] wird ein Ansatz beschrieben, der Produktberatung für einen Distributor von Hardware-Komponenten realisiert. Hier wurde jedoch auf Kundenmerkmale verzichtet⁵ und stattdessen der existierende Katalog des Distributors benutzt. Der Katalog basiert auf einer dreistufigen Kategorienhierarchie. Benutzer können im System Profile anlegen und editieren, die auf dieser Kategorienhierarchie basieren. Basierend auf den Profilen bekommen die Benutzer personalisierte Email-Newsletter zugesendet, in der Änderungen und Neueintragen in der Datenbasis, die zu ihrem Profil passen, erläutert werden. Die Inhalte der Newsletter werden von Zeit zu Zeit mit den tatsächlich genutzten Angeboten verglichen. Auf dieser Basis werden Empfehlungen für Profilveränderungen erarbeitet, die beim nächsten Login des Benutzer vorgeschlagen werden.

⁵Es ist anzunehmen, dass auf Kundenmerkmale verzichtet wurde, weil dies bedeutet hätte, dass alle Produkte mit diesen Kundenmerkmalen hätten annotiert werden müssen. Dies hätte die Kosten für die Umsetzung in die Höhe getrieben, da der Distributor eine große Menge von Produkten vertreibt.

Ardissono et. al. (vgl. [ACT99], [AG00], [AGP⁺01]) umgehen das Problem der Zuordnung von Dokumenten zu einem Domänenmodell ebenfalls, indem sie auf schon strukturierte Inhalte zugreifen. In ersten Fall (vgl. [ACT99]) handelt es sich um Neuigkeiten von einem Newsserver, die mit inhaltlichen Kategorien versehen sind, während es im zweiten Fall (vgl. [AG00]) um Produktdaten aus einer Produktdatenbank geht, die in Produktkategorien organisiert ist. Beide Ansätze benutzen eine Kategorienhierarchie zur Strukturierung der vorhandenen Informationen. Je nach Interesse bekommen die Benutzer nur den Titel, eine Zusammenfassung, einige Kapitel oder das ganze Dokument zu einem Produkt präsentiert. Die Benutzer haben jederzeit die Möglichkeit, sich die ausgeblendeten Informationen anzuschauen. Dies kann dazu führen, dass das Benutzermodell für diese Kategorie geändert wird. Im Falle der Produktdatenbank (vgl. [AG00]) werden die einzelnen Produkte nicht nur durch ihre Zuordnung zu einer Produktkategorie beschrieben, sondern haben jeweils ähnlich wie in [TR98] und [HTS98] eine Reihe von Merkmalen, die über die Hierarchie vererbt werden können. Bei diesen Merkmalen handelt es sich jedoch um flache Strukturen. Diese Merkmale sollen auf längere Sicht durch hierarchische Strukturen ersetzt werden, um so ein mächtigeres Beschreibungswerkzeug zu bekommen und bessere Adaptionsergebnisse zu erzielen.

Die angestrebte Struktur wird in ähnlicher Form im SWAN-Projekt benutzt (vgl. [GIK99], [GI00]). Das SWAN-Projekt ist ein nautisches Informationssystem, das von Seeleuten genutzt werden soll. Die verschiedenen im System verfügbaren Informationen werden durch sog. Konzepte gruppiert. Die verfügbaren Informationen zu einem Konzept können jeweils durch verschiedene Kategorien untereinander unterschieden werden. Neben dieser inhaltlichen Zuordnung der einzelnen Informationen, sind diese auch einer geographischen Position oder Region zugeordnet, die neben der inhaltlichen Filterung durch Konzepte und Kategorien eine räumliche Filterung erlaubt. Das System benutzt einen stereotypischen Ansatz, indem es die Benutzer in Benutzerklassen einteilt. Die Zuordnung zu einer Benutzerklasse impliziert bestimmte Aufgaben, die in einem Aufgabenmodell festgelegt sind. Diese Aufgaben führen zu bestimmten Informationsbedürfnissen, die zur Definition eines Informationsfilters führt. Die Filter werden auf Basis von Konzepten, Kategorien und einer räumlichen Region vorgenommen.

Bei Anwendungen, die keinen Zugriff auf strukturierte Daten haben, sondern vielmehr mit willkürlichen, unstrukturierten Informationen arbeiten, stellt sich das Problem der Zuordnung zum benutzten Domänenmodell. Während rein termbasierte Systeme (vgl. Kap. 4.1) das Problem umgehen, indem sie das Benutzermodell direkt auf die benutzten Wörter abbilden, gibt es auch neuere Ansätze (vgl. [CCS01], [dBRLEGN01] [Est01] und [KHK01]), die als grundlegendes Domänenmodell eine Kategorienhierarchie benutzen. Zu jeder Kategorie wird eine Reihe von Schlüsselwörtern definiert. Auf Basis dieser Schlüsselwörter erfolgt

die Zuordnung zu den verschiedenen Kategorien. Die Kategorien wiederum bilden das Rückgrat des Benutzermodells. Die Systeme beobachten die Nutzung des Benutzers und bilden die benutzten Dokumente auf die Kategorienhierarchie ab. Aus den relativen Häufigkeiten für die Kategorien lassen sich dann Interesse bzw. Desinteresse für die verschiedenen Kategorien herauslesen. Während Rodriguez et al. (vgl. [dBRLEGN01], [Est01]) im Wesentlichen dem oben beschriebenen Ansatz folgen, bieten Chen et al. (vgl. [CCS01]) zusätzlich noch einen sog. Personal View an, in dem alle interessanten Kategorien angezeigt werden. Mit diesem Personal View wird „Profilhygiene“ unterstützt, da nur interessante Kategorien in ihm enthalten sind. Fallen die relativen Häufigkeiten für eine Kategorie unter einen bestimmten Schwellwert, so wird sie mit ihrer Oberkategorie zusammengefasst. Umgekehrt können sehr interessante Kategorien auf mehrere Unterkategorien aufgeteilt werden. Die benutzte Kategorienhierarchie ist bei [CCS01] wie bei [dBRLEGN01] und [Est01] fest und wird vor dem Start des Systems festgelegt. Dagegen nutzen Kim et al. (vgl. [KHK01]) vorhandene, kategoriale Strukturen wie die Ordnerstruktur im Homeverzeichnis des Benutzers sowie seine Email- und Bookmark-Ordner zur Initialisierung der Kategorienhierarchie.

Müller et al. (vgl. [Mül99], [Mül01], [DEM⁺01]) verwenden ebenfalls eine Kategorienhierarchie zur Modellierung der Domäne, bewerkstelligen jedoch die Zuordnung von Dokumenten zu den verschiedenen Dokumenten nicht über eine feste Schlüsselwortliste, sondern benutzen stattdessen Wrapper-Induktion. Dabei wird das Feedback des Benutzers nicht nur benutzt, um sein Interesse zu modellieren, sondern auch, um die Zuordnung zu einer bestimmten Kategorie zu überprüfen und ggf. zu ändern. Dieser Ansatz wurde sowohl in der Meta-Suchmaschine OySTER (vgl. [Mül99], [Mül01]), als auch im Informationssystem Bikini (vgl. [DEM⁺01]) verwirklicht. In beiden Systemen werden die Interessen des Benutzers auf die Kategorienhierarchie abgebildet. Der Benutzer hat jederzeit Zugriff auf sein eigenes Profil und kann es jederzeit ändern. Gleichzeitig hat auch das System die Möglichkeit, das Profil zu ändern. Das Interessenprofil bildet die Interessen des Benutzers für jede Kategorie auf ein Kontinuum zwischen 0 und 1 ab. Die verschiedenen Belegungen für die Kategorien werden benutzt, um Dokumente anhand ihrer Kategorien, die vom System vergeben wurden, zu bewerten. Benutzer beider Systeme sind angehalten, explizites Feedback zu geben, damit das Interessenprofil aktualisiert werden kann.

Brokeringmodell von domänenmodellbasierten, adaptiven Informationssystemen

Domänenmodellbasierte Informationssysteme lassen sich danach unterscheiden, ob sie Quellen selbst auswerten, wie es z.B. beim Oyster-System der Fall ist, oder ob sie auf bereits strukturierten Daten aufbauen. Außerdem ist eine Unterschei-

dung danach möglich, ob die Systeme adaptiv, adaptierbar oder adaptiv und adaptierbar sind (vgl. Tab. 4.1). Handelt es sich um ein rein adaptierbares System, so wird der kundenorientierte Prozess weitestgehend vom Benutzer, in diesem Fall dem Konsumenten, alleine durchgeführt. Im anderen Extrem, wenn es sich um ein rein adaptives System handelt, wird der kundenorientierte Prozess vom System allein bewerkstelligt, ohne dass Benutzer eine Einflussmöglichkeit haben. Bei Systemen, die sowohl adaptierbar als auch adaptiv sind, werden die Aufgaben im kundenorientierten Prozess geteilt.

Wenn das System auf bereits existierenden Daten aufbaut, fallen Broker- und Anbieterrolle zusammen. Wenn das System die Quellen selbst auf das Domänenmodell abbildet, werden Quellen- und Domänenverwaltungsaufgaben vom System ausgeführt. In diesem Fall füllt das System die Rolle des Brokers aus. Prototypisch für Systeme, die sowohl adaptiv als auch adaptierbar sind und zudem noch Quellen- und Domänenverwaltungsaufgaben ausführen, ist das Oyster-System (vgl. [Mül99], [Mül01]). Die Aufgabenverteilung für das Oyster-System ist in Abbildung 4.4 dargestellt.

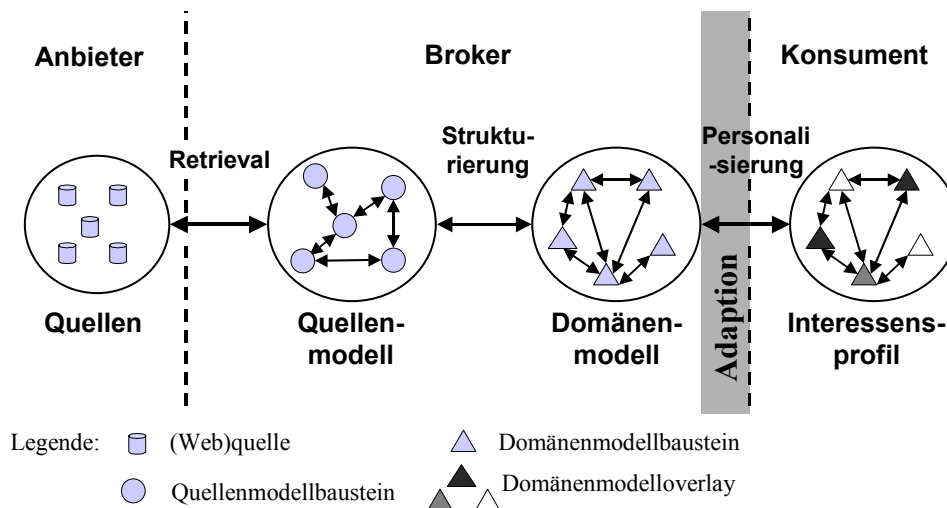


Abbildung 4.4: Oyster-Aufgabenverteilung

4.6 Zusammenfassung

Betrachtet man die in den vorhergegangenen Abschnitten beschriebenen Ansätze und Systeme zu adaptiven Informationssystemen im Hinblick auf das in der Einleitung kurz eingeführte und in Kapitel 3 weiter ausgeführte Modell von Information-Brokering, so kann man für termbasierte, adaptive Informationssysteme

Ansatz/System	Struktur des Domänenmodells	Adaptionsziel	Füllung des Domänenmodells	Benutzer-System-Interaktion
EdInfo/ConCall	flach	Konferenzaufrufe	manuell	adaptierbar
MIRADOR	Meta-Daten	Meta-Daten	automatisch	adaptiv
Timm et al.	Kategorien, Kundenmerkmale	Produktinformationen	vorhandene Datenbank	adaptiv, adaptierbar
Ardissono et al.	Kategorien, Dokumentstruktur	Dokumente, level of detail	vorhandene Strukturierung	adaptiv
SWAN	Konzepte, Kategorien	Dokumente	keine Angabe	adaptiv
Rodriguez et al.	Kategorien, Vektorraummodell	Dokumente	automatisch	adaptiv
Chen et al.	Kategorien, Vektorraummodell	Dokumente	automatisch	adaptiv, Profilhygiene
Kim et al.	vorhandene kategoriale Strukturen	Dokumente	automatisch	adaptiv
Müller et al.	Kategorien	Dokumente	automatisch	adaptiv, adaptierbar

Tabelle 4.1: Vergleich der verschiedenen domänenmodellbasierten Informationssysteme

festhalten, dass bei ihnen die Prozesse Retrieval, Repräsentation und Personalisierung zusammenfallen. Bei diesen Systemen ist das benutzte Domänenmodell durch das verwendete Retrieval-Verfahren determiniert. Meist handelt es sich dabei um ein Vektorraummodell oder ein Modell, das diesem sehr ähnlich ist, wie es beispielsweise von Pazzani (Syskill & Webert, vgl. [PMB96]) benutzt wird. Die Personalisierung wird durch ein Overlay-Modell auf das rudimentäre Domänenmodell bewerkstelligt.

Domänenmodellbasierte, adaptive Informationssysteme legen den Fokus auf die Repräsentation der Domäne im Domänenmodell und auf eine darauf basierende Personalisierung, die wiederum als Overlay-Modell über das (komplexere) Domänenmodell realisiert wird. In fast allen Fällen werden mit Kategorien hierar-

chische Strukturen zur Repräsentation der Domäne benutzt. Tabelle 4.1 gibt einen Überblick über die verschiedenen domänenmodellbasierten Informationssysteme, die in Kapitel 4.5 beschrieben wurden. In Tabelle 4.1 wurde auf die Auflistung der verschiedenen beschriebenen Lehr-Lernsystem (vgl. Kap. 4.4) verzichtet, da diese Systeme primär dem Zweck dienen, Lehrinhalte zu vermitteln. Das bedeutet, dass sie versuchen, das Wissen des Lernalers (des Benutzers) über bestimmte Lerninhalte, nicht aber sein Interesse sein an bestimmten Informationen oder Sachverhalten, wie es in den in Kapitel 4.5 beschriebenen Informationssystemen der Fall ist.

In den beschriebenen, domänenmodellbasierten Informationssystemen werden in vielen Fällen neben Kategorien weitere Strukturelemente zur Repräsentation benutzt. Bemerkenswert ist die explizite Beschreibung von Kundenmerkmalen, wie sie von Timm et al. (vgl. [TR98]) vorgeschlagen wurde, da sie die Rolle des Domänenmodells für adaptive Informationssysteme beleuchtet: Da das Interessenprofil des Benutzers durch eine Abbildung auf das Domänenmodell beschrieben wird, muss das Domänenmodell die Sicht der Benutzer auf die Domäne widerspiegeln. Für Information-Brokering ergibt sich die Forderung, dass das verwendete IB-Domänenmodell die möglichen Sichten der Konsumenten möglichst adäquat abbilden muss. Im SWAN-Projekt (vgl. [GIK99], [GI00]) wurde neben einer hierarchischen Beschreibung, die durch eine Konzepthierarchie realisiert wurde, eine weitere Unterteilung der Konzepte in Kategorien umgesetzt. Diese weitere Spezialisierung kann als Ausgestaltung der von Timm et al. eingesetzten Kundenmerkmale aufgefasst werden. In der Ausgestaltung des IB-Domänenmodells für Information-Brokering in Kapitel 6 wird dieser Ansatz aufgegriffen. Im Unterschied zu Broker's Lounge ist SWAN jedoch nur adaptiv. Dies bedeutet, dass die Benutzer keine Kontrolle über ihr eigenes Profil haben. Dadurch wird die Forderung von Fischer (vgl. [Fis01]) nicht erfüllt, der die volle Kontrolle des Benutzers über sein Profil fordert. Zudem wird mit einem stereotypischen Ansatz ein wenig flexibler Ansatz zum Erlernen von Benutzerinteressen benutzt.

Die domänenmodellbasierten, adaptiven Informationssysteme lassen sich in Systeme einteilen, die auf bereits strukturierte Daten zugreifen können und solche, die versuchen, beliebige Dokumente automatisch auf das verwendete Domänenmodell abzubilden. Diese Systeme beschäftigen sich zusätzlich zu den bereits oben beschriebenen Information-Brokeringprozessen Repräsentation und Personalisierung auch mit dem Retrieval der Inhalte. Es handelt sich bei diesen Systemen um vollautomatisierte Information-Brokeringsysteme. Es stellt sich jedoch die Frage, ob die mit automatischen Prozessen erreichbare Qualität bei der Zuordnung von Dokumenten zum IB-Domänenmodell für die Konsumenten ausreichend ist, oder ob es eines manuellen Auswertungsschritts bedarf, der gleichzeitig der Qualitätssicherung dient. In dieser Arbeit wird davon ausgegangen, dass dieser manuelle Qualitätssicherungsschritt notwendig ist, um hochqualitative Informationen anbieten zu können. Gleichwohl ist es angestrebt, die noch verblei-

bende manuelle Arbeit auf intellektuell anspruchsvolle Arbeiten zu beschränken. Deswegen können die in Kapitel 4.5 beschriebenen Retrieval-Verfahren nützliche Dienste zur Vorkategorisierung von Dokumenten leisten. Dies wurde bei der Bewertung von Dokumenten in der Broker's Lounge aufgegriffen und wird beschrieben in Kapitel 8.6.

Kollaborative Ansätze, wie sie in Kapitel 4.2 beschrieben wurden, nutzen nicht die Inhalte des Informationssystem zur Adaption, sondern deren Nutzung durch eine Gruppe von Benutzern. Mittels kollaborativer Techniken kann sich das System schneller an die Bedürfnisse von Benutzern anpassen, als mit rein inhaltsbasierten Ansätzen. Jedoch sind diese Techniken nicht auf neue Items in der Datenbasis anwendbar (vgl. [ZA01]). Da im Information-Brokering, sofern man von einer längerfristigen Beziehung zwischen Broker und Konsument ausgeht, regelmäßig neue Items entstehen, sind kollaborative Techniken nicht für die Filterung von neuen Informationen für Konsumenten einsetzbar. Gleichwohl können kollaborative Ansätze mit der Analyse von Brokeringprozessen verglichen werden, da auch dabei nicht der Inhalt, sondern dessen Nutzung im Vordergrund steht.

Aus den beschriebenen adaptiven Ansätzen lässt sich die Rolle des Systems im Personalisierungsprozess ableiten. Angestrebt wird, Konsumenten (Benutzern) eine möglichst weitgehende Unterstützung bei der Formulierung ihrer Interessen zu geben. Da die subjektiven Informationsbedarfe nicht notwendigerweise den objektiven Informationsbedarfen entsprechen (vgl. [Kuh95]) und die Benutzer ihre Interessen möglicherweise nicht adäquat mit den Mitteln des Systems beschreiben können (vgl. [PN99]), kann eine adaptive Komponente hier Abhilfe schaffen.

Auf der anderen Seite ist zu verhindern, dass den Benutzern für sie relevante Informationen vorenthalten werden. Da die Interessenprofile als Filter wirken, könnte eine vom System eigenmächtig durchgeführte Änderung des Profils und damit des Filters zu einem Vertrauensverlust in das System führen (vgl. [Mae94]). Dies könnte wiederum dazu führen, dass Benutzer das System nicht mehr benutzen. Um dies zu verhindern, sollte das Profil nur mit dem Einverständnis des Benutzers geändert werden können.

Andererseits darf die Angst vor einem Informationsverlust nicht soweit gehen, dass generell auf Filterung verzichtet wird, weil dies in letzter Konsequenz nicht zu einer Verringerung des Phänomens des Information Overload führt. Diese Angst des Benutzers kann reduziert werden, indem der Benutzer jederzeit einen Überblick über sein Profil hat und er es jederzeit inspizieren kann. Der Benutzer sollte jederzeit die Möglichkeit haben, sein Profil zu ändern und die Auswirkungen der Änderung sofort zu studieren. In den beschriebenen Systemen wurde diese Inspizierbarkeit des Profils nur von Timm et al. (vgl. [TR98]) und Müller et al. (vgl. [Mül01]) verwirklicht. Vorschläge zur Änderung des Profils sollten vom System nur dann gemacht werden, wenn sie mit hinreichender Sicherheit zutreffen, da sie sonst als Bevormundung des Systems gegenüber dem Benutzer verstanden

werden könnten.

Zusammenfassend kann man über das Verhältnis von Adaptivität und Adaptierbarkeit für die Personalisierung im Information-Brokeringprozess festhalten, dass die Benutzer die volle Kontrolle über ihr Profil behalten sollten und dass Adaptivität nur sparsam als Tippkomponente eingesetzt werden kann. Die Kontrolle des Systems durch den Benutzer sollte so weit gehen, dass Benutzer auch die Möglichkeit haben, die adaptive Komponente abzuschalten. Bildet man dieses Verständnis von Personalisierung auf das Information-Brokeringmodell ab (vgl. Kap. 1 und Kap. 3), so ergibt sich, dass der Personalisierungsprozess zwischen dem Konsumenten und dem System aufgeteilt wird. Wenn das System dem Benutzer (Konsumenten) Tipps gibt, nimmt das System die Rolle des Brokers ein. Dies bedeutet wiederum, dass ein menschlicher Broker in diesem Moment von der Personalisierungsaufgabe entlastet werden kann.

Kapitel 5

Ontologiebasierte Informationssysteme

In diesem Kapitel werden ontologiebasierte Informationssysteme beleuchtet. Ontologien (vgl. Kap. 2.1.3) spielen im Wissensmanagement (vgl. Kap. 2.1.2) eine wichtige Rolle, da sie zur Strukturierung von Wissen benutzt werden und somit die Basis für technisch gestütztes Wissensmanagement bilden (vgl. [Fen01]).

Wissensmanagement-Systeme können nach [BFGP98a] in vertikale und horizontale Systeme unterschieden werden:

- Vertikale Systeme sind für einen bestimmten Zweck, für eine bestimmte Domäne entwickelt worden und tun dort gute Dienste, sind jedoch nicht auf andere Domänen übertragbar.
- Horizontale Systeme sind domänenunabhängig¹ und übertragbar. Sie sind als Frameworks konzipiert, die auf die jeweilige Situation angepasst werden können. Horizontale Systeme sind in ihrer Erstellung deutlich komplexer als vertikale Systeme.

¹Benjamins et al. gehen in [BFGP98a] von einer engen Auslegung des Begriffs „Domäne“ aus. Sie bezeichnen mit „Domäne“ die Sachverhalte, die in einer Domänenontologie im System beschrieben werden können. Sie meinen mit horizontalen Systemen damit nicht Systeme, die in allen denkbaren Szenarien eingesetzt werden können. Broker's Lounge z.B. (vgl. Kap. 8 kann in vielen inhaltlichen Domänen benutzt werden. Broker's Lounge ist aber auf Brokering-Szenarien beschränkt. Trotz dieser Einschränkung kann es von vertikalen Systemen unterschieden werden, die nur in einer Domäne (und in einem Szenario) eingesetzt werden können. Als Beispiel sei hier die Entwicklung vom „alten“ ELFI-System zur ELFI-Umsetzung auf Basis von Broker's Lounge angeführt, die in Kapitel 9.1 beschrieben wird. Das alte ELFI-System konnte nur für das Brokering von Informationen über Forschungsförderung genutzt werden. Broker's Lounge und mit ihr die ELFI-Umsetzung mit Broker's Lounge kann zum Brokering beliebiger Informationen benutzt werden.

Ontologiebasierte Informationssysteme sind horizontale Systeme, da sie die Domäne in einer zentralen Ontologie beschreiben. Sie können dadurch auf verschiedene Domänen angepasst werden. Horizontale Systeme haben gegenüber vertikalen Systemen den Vorteil, dass die Kosten für ihre Erstellung deutlich niedriger sind als die Implementation eines neuen vertikalen Systems, da nur die Ontologie an die neue Domäne angepasst werden muss. Broker's Lounge ist als horizontales System konzipiert, das jeweils auf verschiedene Domänen angepasst wird.

Die Informationen, mit denen der Broker handelt, werden in einer Ontologie beschrieben. In kleineren überschaubaren Domänen ist es denkbar, dass Information-Broker mit einem internen Modell von der Domäne arbeiten. In diesem Fall überführen sie dieses interne Modell nicht in ein formales, explizites Modell. Wenn Information-Brokering technisch unterstützt werden soll, ist eine Ontologie zur Beschreibung der Domäne zentraler Bestandteil eines solchen Information-Brokering-Systems (siehe Kapitel 3 sowie insbesondere Kapitel 6).

Bei Information-Brokering-Systemen wie Broker's Lounge handelt es sich deswegen um ontologiebasierte Informationssysteme. In den nächsten Abschnitten werden ausgewählte Systeme und ihre jeweiligen Brokeringmodelle beschrieben. Die verschiedenen Systeme werden anhand ihres Brokeringmodells eingeteilt. Die Systeme können in Semantic-Web-Systeme (vgl. Kap. 5.1) und Klassifikationssysteme (vgl. Kap. 5.2) unterschieden werden. Bei Semantic-Web-Systemen wird die Strukturierung vom Informationsanbieter selbst durchgeführt, während Klassifikationssysteme die Strukturierung von einer zentralen Instanz durchführen lassen.

Neben der Betrachtung der Brokeringmodelle stehen die verwendeten Modellierungsmittel zur Beschreibung der Domäne im Vordergrund der Betrachtung. Fast alle vorgestellten Ansätze basieren auf Domänenontologien. Für die Formulierung und Bearbeitung von Ontologien sind Ontologiesprachen und Ontologietools von eminenter Bedeutung. Die verwendeten Ontologiesprachen lassen Rückschlüsse auf die verwendeten Modellierungskonstrukte zu. Die Analyse von Ontologietools zur Bearbeitung von Ontologien gibt Hinweise darauf, wie ein Editor zur Formulierung des Domänenmodells in Broker's Lounge aussehen könnte. Ontologiesprachen und Ontologietools werden im Anschluss in Kapitel 5.3 und Kapitel 5.4 beschrieben.

Abschließend werden die beschriebenen Systeme hinsichtlich ihrer Personalisierungsunterstützung untersucht. Basierend auf dieser Untersuchung wird Broker's Lounge, bzw. das angestrebte Leistungsspektrum von Broker's Lounge gegen die anderen Ansätze positioniert.

5.1 Semantic-Web-Systeme

Bei der bisherigen Entwicklung des World-Wide Web stand die Lesbarkeit der dargebotenen Informationen durch menschliche Benutzer im Vordergrund. Das explosionsartige Wachstum des WWW hat zwar eine große Menge von Informationen verfügbar gemacht, aber gleichzeitig wird die Orientierung in dieser Informationsfülle unmöglich. Das Semantic Web (vgl. [BLHL01]) setzt genau hier an und will diesen Informationen eine formale Semantik geben, damit sie automatisiert z.B. von Agenten (vgl. [Hen01a]) verarbeitet werden können. In [HH01] wird prototypisch beschrieben, wie ein von Agenten gesteuertes Semantic Web funktionieren könnte. Dabei setzt das Semantic Web auf die dezentrale Annotation von Dokumenten. Durch die dezentrale Organisation wird ein ähnlich explosionsartiges Wachstum wie bei herkömmlichen Dokumenten erwartet. Das Semantic Web wird durch Standards wie XML (EXtensible Markup Language) und RDF (Resource Description Framework, vgl. [Las98]) befördert.

Das Semantic Web baut auf den vorangegangenen Entwicklungen von Ontologiesprachen wie z.B. SHOE (vgl. [LSR96]) auf und versucht die Erfahrungen im OntoWeb-Netzwerk² zu bündeln. Dies hat unter anderem dazu geführt aus OIL (vgl. [FHH⁺00]) und DAML eine gemeinsame Sprache DAML+OIL (vgl. [Hor01]) zu entwickeln. Die Rolle von OIL, DAML und RDF und ihre Möglichkeiten im Semantic Web werden in [LM01] beschrieben. Im Abschnitt über Ontologiesprachen (Kap. 5.3) werde ich auf die verschiedenen Sprachen noch einmal kurz eingehen.

Semantic-Web-Systeme sind Systeme, bei denen Dokumente, bzw. Bestandteile von Dokumenten mit einer Ontologie bzw. einem Verweis auf eine Ontologie annotiert werden. Die Annotation der Dokumente wird in der Regel vom Anbieter der Information vorgenommen.

Bei dieser Klasse von Systemen dient eine zentrale Ontologie als Domänenmodell. Die Informationsanbieter werden verpflichtet, ihre eigenen Dokumente entsprechend der Ontologie in einer Ontologiesprache zu annotieren. Um die Annotation zu erleichtern, gibt es in vielen Fällen eine grafische Benutzeroberfläche, die bei der Annotation Unterstützung anbietet. Prototypisch für diese Klasse ist Ontobroker (vgl. [FDES98], Kap. 5.1.2) und seine Nachfolger. Ein ähnlicher, früherer Ansatz ist das SHOE-Projekt (vgl. [LSR96], [HH01], Kap. 5.1.1). Wie beschrieben, kommt den Informationsanbietern eine wichtige Rolle zu: Sie führen die Annotation und damit die Domänenverwaltungsaufgaben alleine aus. Im Unterschied zu Ontobroker und SHOE, bei denen die Annotationen zusammen mit dem Dokument verwaltet werden, werden in KnowWeb (vgl. [MSPK00], Kap. 5.1.3) Dokumente und Annotationen voneinander getrennt verwaltet.

²<http://www.ontoweb.org/>

Die Quellenverwaltungsaufgaben werden dagegen zwischen den Anbietern und dem Broker geteilt. Der Informationsanbieter muss seine neue Quelle beim Broker anmelden. Änderungen werden vom Broker automatisch überwacht. Im Unterschied zu allen anderen Brokeringsystemen, bei denen die Quellen unstrukturierte Dokumente³ enthalten, werden hier die Inhalte der Quellen direkt ohne weitere Aufarbeitung in das Domänenmodell integriert. Dies entlastet die zentrale Instanz, bei dem es sich in vielen Fällen um den Broker handelt, da kein manueller Arbeitseinsatz mehr erforderlich ist. Diese Aufgabe wird an die Informationsanbieter delegiert. Da keine zentrale manuelle Aufarbeitung notwendig ist, wird kein zentrales Quellenmodell benötigt. Stattdessen liegen die Dokumente schon in annotierter Form vor. Dies wird in Abbildung 5.1 deutlich, die das Brokeringmodell von Semantic-Web-Systemen zeigt.

Die kundenorientierten Prozesse werden weitestgehend von den Informationskonsumenten, den Benutzern des Brokeringsystems, selbst ausgeführt. Sie geben ihre Interessen an, indem sie entweder eine Anfrage (in einer Anfragesprache) an das System formulieren oder mit einer grafischen Benutzeroberfläche arbeiten, die entsprechende Anfragen an das System generiert. D.h., außer dem Querying werden alle Aktivitäten von den Konsumenten selbst ausgeführt.

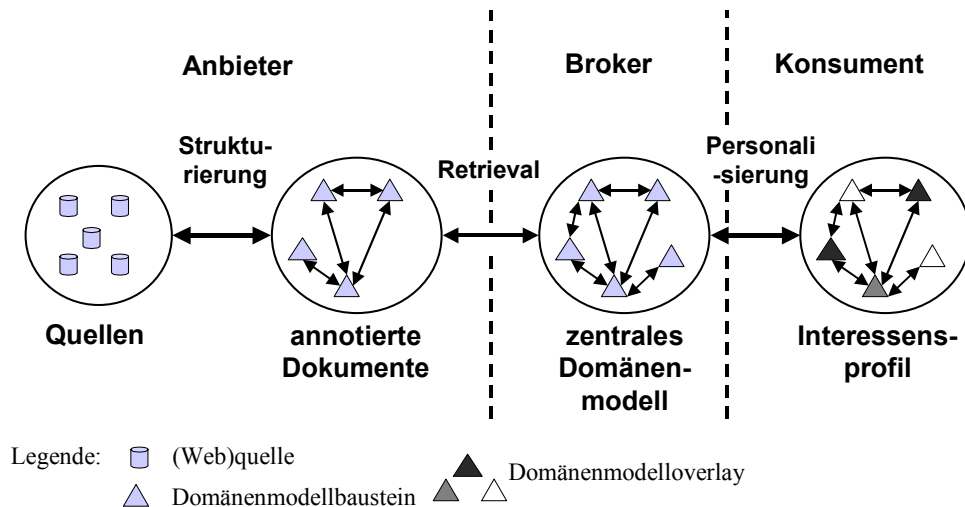


Abbildung 5.1: Aufgabenverteilung in Semantic-Web-Systemen

Eine Sonderstellung innerhalb der Semantic-Web-Systeme nimmt der agentenbasierte Ansatz ein, der im EU-Projekt CoMMA (vgl. [GDCG00], Kap. 5.1.4) verfolgt wird. Dort gibt es neben einem eher technisch orientierten Verbindungsagenten, der die Kommunikation zwischen den verschiedenen Agenten sicher-

³Unstrukturierte Dokumente sind in diesem Zusammenhang Dokumente, die noch nicht im Sinne des zentralen Domänenmodells aufgearbeitet wurden.

stellt noch Ontologie-, Dokumenten- und Benutzeragenten. Bildet man die letzteren drei Agenten auf das Brokeringmodell ab, so können die Dokumentenagenten als Informationsanbieter, der Ontologieagent als Broker und die Benutzeragenten als Konsumenten verstanden werden. Im CoMMA-Projekt wird von strukturierten Dokumenten ausgegangen. D.h. es handelt sich um RDF- oder XML-Dokumente, die nach einer zentralen, vom Ontologieagenten verwalteten Ontologie annotiert werden. Dies rechtfertigt die Einordnung in die Klasse der Semantic-Web-Systeme, da die Domänenverwaltungsaufgaben (wie in Semantic-Web-Systemen) von den Informationsanbietern alleine ausgeführt werden, während die Quellenverwaltungsaufgaben zwischen Broker und Informationsanbietern geteilt wird. In Abweichung von den anderen Semantic-Web-Systemen wird in CoMMA versucht, die Interessen der Benutzer durch Beobachtung des Benutzers zu erkennen. Damit kann den Konsumenten durch eine adaptive Komponente Unterstützung bei der Profilbildung angeboten werden. Es ergibt sich das Brokeringmodell, das in Abbildung 5.2 dargestellt ist.

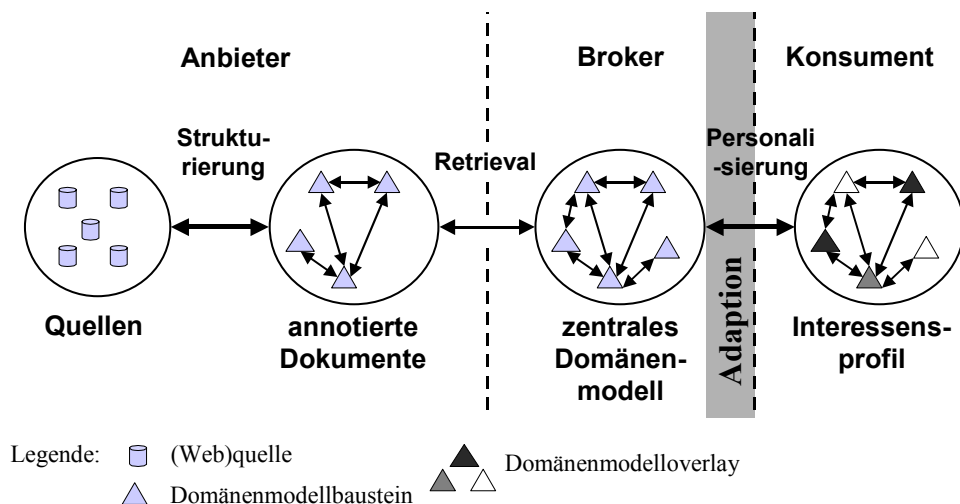


Abbildung 5.2: CoMMA-Brokeringmodell

5.1.1 SHOE

Einer der ersten Ansätze zur Annotation von Dokumenten mit Ontologien ist SHOE (vgl. [LSR96]). Bei diesem Ansatz wird auf eine zentrale Ontologie verwiesen und das Dokument entsprechend annotiert. Dazu wurden einige spezielle Tags definiert, die in einem HTML-Browser unsichtbar bleiben, jedoch einer zentralen Ontologie genügen. Diese Tags werden in normale HTML-Dokumente eingebettet und dienen der semantischen Repräsentation der Inhalte des Dokumentes.

Die SHOE-Tags können entweder mit einem normalen Texteditor erstellt werden oder man benutzt den sog. Knowledge Annotator, der bei der Annotation von bestehenden HTML-Dokumenten hilft und die Vollständigkeit der Annotationen sicherstellt. Die annotierten Dokumente werden von einem Webrobot eingelesen und die Ergebnisse in einer Wissensbank abgelegt. Die gefundenen Inhalte können mit einer Anfrageschnittstelle angefragt werden. Als Ergebnis werden wie bei einer Suchmaschine gefundene Webdokumente angeboten. Werden in der eigenen Wissensbank keine Ergebnisse gefunden, so wird die entsprechende Anfrage an eine Suchmaschine weitergeleitet. Das Zusammenspiel der verschiedenen SHOE-Komponenten zu einem Gesamtsystem als Prototyp für das Semantic Web wird in [HH01] beschrieben.

5.1.2 Ontobroker

Benjamins beschreibt in [FDES98] Ontobroker als Beispiel für ein horizontales Wissensmanagement-System. Bei Ontobroker wird die Ontologie über das Web verfügbar gemacht. Jeder kann am Aufbau der Ontologie teilnehmen. Inhalte (Fakten) kommen in das System, indem HTML-Seiten mit einer bestimmten Syntax annotiert werden. Diese Seiten müssen beim Ontobroker-System registriert werden, der diese Seiten regelmäßig besucht und sie mit der Ontologie abgleicht und ggf. einige Schlüsse zieht. Zugriff zu den gesammelten Informationen erhält man über ein Query-Interface. Eine Beispieldomäne ist $(KA)^2$. $(KA)^2$ (**K**nowledge **A**nnotation Initiative of the **K**nowledge **A**cquisition Community, vgl. [BFGP⁺98b]) beschreibt die Wissensmanagement-Community mit ihren Forschungsthemen sowie den beteiligten Personen und Gruppen. In der Zwischenzeit wurde Ontobroker unter dem Namen On2broker und On-to-knowledge ständig weiterentwickelt. Nutzte man im ursprünglichen Ontobroker noch eine Erweiterung des HTML-A-Tags zur Beschreibung der Inhalte eines Dokumentes (und greift damit die im Wesentlichen die SHOE zugrunde liegende Idee auf), so können die Inhalte jetzt auch in RDF beschrieben werden. Benutzer erhalten Zugriff auf das gesammelte Wissen in Ontobroker über einen hyperbolischen Baum, der die Ontologie visualisiert. Die Ontologie und die Inhalte werden mit dem FaCT-System (vgl. [Hor98]) verwaltet. Erfahrene Benutzer können Anfragen in der entsprechenden Anfragesprache stellen. Die Benutzerschnittstelle in Ontobroker übersetzt die Anfragen der Benutzer automatisch in FaCT-Anfragen. Benutzer können sich regelmäßig über Veränderungen zu ihren Anfragen mit einem Push-Dienst informieren lassen. Um die Anwendbarkeit von Ontobroker zu erweitern, wurde OIL (**O**ntology **I**nterchange **L**anguage) [FHH⁺00] entwickelt. OIL ist eine Beschreibungssprache für Ontologien. In Kapitel 5.3 werde ich OIL etwas genauer beschreiben.

5.1.3 KnowWeb

KnowWeb (vgl. [MSPK00]) ist ein Unternehmensgedächtnis, das auf einer Domänenontologie beruht. Die Domänenontologie wird zur Annotation und damit Strukturierung von Dokumenten benutzt. In KnowWeb wird eine Ontologie als eine Menge von Operatoren angesehen. Diese Operatoren erlauben die Definition von Klassen, die zu einer Hierarchie angeordnet werden können. Zu den Klassen können Relationen, Funktionen, Variablen und Konstanten definiert werden. Objekte können als Instanzen der verschiedenen Klassen definiert werden. Die Domänenontologie kann mit einem grafischen Editor editiert werden. Im System können HTML- und Word-Dokumente mit Operatoren aus der Domänenontologie annotiert werden. Handelt es sich um lokale Dokumente, so können auch einzelne Teile des Dokumentes (Kapitel, Abschnitte oder Sätze) annotiert werden. Die Annotationen werden durch farbliche Unterlegung kenntlich gemacht. Anfragen an das System können im Volltext und über die Konzepte gestellt werden. Bei Anfragen über Operatoren können zusätzlich auch noch benachbarte Operatoren in die Anfrage integriert werden.

5.1.4 CoMMA

Im Projekt CoMMA (vgl. [GDCG00], [PKD⁺]) wird ein Wissensmanagement-Ansatz verfolgt, der auf XML und RDF zur Repräsentation der Inhalte und einem Multi-Agenten-System zur Bereitstellung der Funktionalität beruht. Ziel ist unter anderem die Bereitstellung eines Unternehmensgedächtnisses. Es wird zwischen Ontologie-, Dokumenten-, Benutzer- und Verbindungsagenten unterschieden. Die Ontologieagenten verwalten die von allen Agenten benutzte Ontologie und bieten auf dieser Basis Brokeringdienste für die anderen Agenten an. Die Dokumentenagenten spielen die Rolle von Informationsanbietern und erlauben den Zugriff auf die von ihnen verwalteten Dokumente. Benutzeragenten ermöglichen menschlichen Benutzern den Zugriff auf die im System vorhandenen Informationen und versuchen gleichzeitig, die Interessen der Benutzer zu erkennen, um proaktiv seine Informationsbedürfnisse befriedigen zu können. Diese adaptive Systemeigenschaft beruht auf der Interpretation der Interaktionshistorie des Benutzers und dem Feedback, das ein Benutzer eventuell gegeben hat. Das Lernverfahren mit dem die Interessen des Benutzers gelernt werden, beruht auf einem Machine-Learning-Verfahren und wird in [KQ01] beschrieben.

Das CoMMA-Projekt unterscheidet sich durch die Betonung der Rolle von (End-)Benutzern von den bisher vorgestellten Ansätzen und folgt der in [DCGR98] formulierten Einsicht, dass den (End-)Benutzern eine entscheidende Rolle für den Erfolg von Wissensmanagement im Unternehmen zukommt. Das CoMMA-System reichert zum einen die Anfragen von Benutzern an das System mit ih-

rem Benutzerprofil an und verwendet dieses Profil zur adaptiven Sortierung (nach Relevanz) der Ergebnisse (Dokumente) der Anfrage. Das Benutzermodell ist ein Overlay-Modell des Domänenmodells. Um Dokumente nach inhaltlichen Interessen bewerten zu können, wurden diese vorher mit bestimmten Themen assoziiert. Abhängig von der Rolle, die ein Benutzer im Unternehmen innehat wird ihm zu Beginn ein bestimmtes Benutzerprofil zugeordnet, das in der Folgezeit überschrieben werden kann. Dies kann als Stereotypen-Mechanismus (vgl. Kap. 2.2.5) aufgefasst werden.

Eine aktive, adaptive Rolle des Systems wurde auch für Planet-Onto (vgl. [DM99], siehe Kap. 5.2.4) angestrebt. Dieser Anspruch wurde jedoch in neueren Publikationen wie z.B. [KDM⁺01] nur teilweise eingelöst.

5.2 Klassifikationssysteme

Klassifikationssysteme wählen einen anderen Weg, als die in den vorhergegangenen Abschnitten beschriebenen Semantic-Web-Systeme. Während Semantic-Web-Systeme davon ausgehen, dass die Informationsanbieter ihre Inhalte selbst annotieren und damit einem breiten Publikum zugänglich machen, gehen Klassifikationssysteme nicht von dieser optimistischen Annahme aus und führen diese Inhaltsanreicherung in einer zentralen Instanz durch. Dabei arbeiten sie wie die Annotationssysteme mit einer zentralen Ontologie oder einer einfacheren Kategorienhierarchie.

Die Dokumente und/oder Informationen der Informationsanbieter werden vom Broker mit einer Ontologie annotiert. D.h., die Domänenverwaltungsaufgaben werden vom Broker ausgeführt. Die verschiedenen Klassifikationssysteme können hinsichtlich der Komplexität des verwendeten Domänenmodells und nach der Aufgabenverteilung der Quellenverwaltungsaufgaben unterschieden werden. Im nächsten Abschnitt werden zunächst Subject Gateways (Kap. 5.2.1) beschrieben. Subject Gateways modellieren eine Domäne mit Hilfe einer Kategorienhierarchie. Yahoo! (Kap. 5.2.2) benutzt ebenfalls eine Kategorienhierarchie zur Modellierung. Im Gegensatz zu Subject Gateways beschränkt sich Yahoo! nicht auf die Beschreibung einer eng umgrenzten Domäne sondern versucht stattdessen das gesamte World-Wide Web abzubilden. Stärker strukturiert sind dagegen FindUR (Kap. 5.2.3) und Planet-Onto (Kap. 5.2.4), die auf einer Konzepthierarchie basieren. FindUR und Planet-Onto unterscheiden sich in der Bearbeitung der Quellenverwaltungsaufgaben. In FindUR werden die Quellenverwaltungsaufgaben vom System wahrgenommen, während in Planet-Onto die Quellenverwaltungsaufgaben von den Anbietern von Informationen wahrgenommen werden.

5.2.1 Subject Gateways

In [Wor98] beschreibt Emma Worsfold Subject Gateways. Subject Gateways sind von Bibliothekaren thematisch aufgearbeitete, katalogisierte Websites zu einem bestimmten Thema, z.B. Sozialwissenschaften. Analysen der eigenen katalogisierten Website zeigen, dass Benutzer diese thematische Aufbereitung mögen. Emma Worsfold betont, dass die Aufbereitung durch Menschen den Informationen einen Mehrwert hinzufügt, da gerade im Web keine Qualitätskontrolle vor der Veröffentlichung stattfindet. Die Qualitätskontrolle in Subject Gateways entspricht der Durchführung der Domänenverwaltungsaufgaben in einer zentralen Instanz. Auch die Auswahl der katalogisierten Quellen wird von einer zentralen Instanz durchgeführt. Damit werden auch die Quellenverwaltungsaufgaben von derselben zentralen Instanz durchgeführt.

Zugleich ist die Stärke von Subject Gateways, die menschliche Aufbereitung, auch die Schwäche des Ansatzes, da sie der Bearbeitungsmenge enge Grenzen setzt. Deswegen wird im DESIRE-Projekt⁴ die Zusammenarbeit von verschiedenen, verteilten Bibliothekaren in verschiedenen Ländern erprobt. Subject Gateways basieren auf der ROADS-Software, die im DESIRE-Projekt entwickelt wurde. Die Quellen werden geographisch verteilt katalogisiert und angefragt. Diese geographische Verteilung steht nicht im Widerspruch zum Information-Brokeringmodell, weil alle Bibliothekare zusammen die oben angesprochene Brokeringinstanz bilden. Die Benutzer können Volltext-Anfragen an Subject Gateways stellen oder die Kategorienhierarchie mit ihren Einträgen durchsuchen. Dies bedeutet, dass der kundenorientierte Prozess weitestgehend vom Benutzer, der in diesem Fall die Rolle des Konsumenten innehat, selbst ausgeführt wird. Es ergibt sich das Brokeringmodell, das in Abbildung 5.3 abgebildet ist. Dort nimmt der Broker breiten Raum ein, weil er sowohl die Quellen- als auch die Domänenverwaltungsaufgaben ausführt, während der kundenorientierte Prozess den Konsumenten überlassen wird.

Personalisierte Sichten auf Subject Gateways (vgl. [DC00]) werden als logische Weiterentwicklung derselben aufgefasst. Die Benutzerprofile basieren dabei auf der Kategorienhierarchie: Benutzer können Kategorien in ihr Profil aufnehmen. Eine Übersicht über weitere Subject Gateways wird in [MB00] gegeben.

5.2.2 Yahoo!

Das Internetportal Yahoo!⁵ arbeitet mit einer Kategorienhierarchie, die zur Klassifikation von Websites benutzt wird. Die Zuordnung von Websites zur Kategorienhierarchie wird von Mitarbeitern des Portals durchgeführt. Das bedeutet, die

⁴<http://www.desire.org>

⁵<http://www.yahoo.com/>

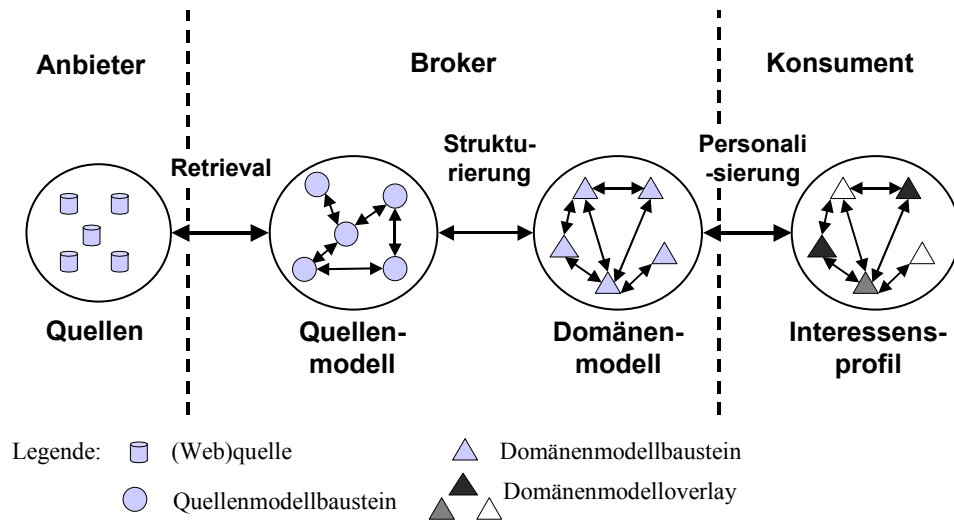


Abbildung 5.3: Brokeringmodell von Subject Gateways

Domänenverwaltungsaufgaben werden ebenso wie die Quellenverwaltungsaufgaben vom Broker durchgeführt. Die kundenorientierten Aufgaben werden weitestgehend von den Konsumenten selbst ausgeübt. Dazu sind die Kategorien jeweils mit Varianten (Synonymen) für die jeweilige Kategorie versehen. Suchanfragen (der Konsumenten) werden auf die Kategorien und die assoziierten Synonyme abgebildet. Ist eine Abbildung nicht möglich, wird die Anfrage an normale Suchmaschinen weitergeleitet. Das Brokeringmodell von Yahoo! entspricht damit dem Brokeringmodell von Subject Gateways (vgl. Abb. 5.3). Im Unterschied zu Subject Gateways werden die Websites von Mitarbeitern einer Organisation klassifiziert, statt von einer Menge von Bibliothekaren aus verschiedenen Organisationen. Außerdem unterscheiden sich Subject Gateways von Yahoo! dadurch, dass sie jeweils einen Themenkreis in aller Tiefe zu beschreiben versuchen, während Yahoo! versucht, das gesamte Web mit seiner Kategorienhierarchie zu beschreiben.

Fensel merkt in [Fen01] kritisch an, dass eine Beschreibung aller möglichen Sites im Web unmöglich ist, da die verwendete Kategorienhierarchie diese Komplexität nicht abdecken kann und die Kategorien keine Beziehungen untereinander außer der Subklassifizierung zulassen. Zusätzlich kritisch anzumerken ist, dass die benutzte Kategorienhierarchie unweigerlich mit dem Facettierungsproblem (vgl. [Ran62], Kap. 2.3.2) konfrontiert wird.

Yahoo! bietet mit MyYahoo eine personalisierte Schnittstelle an. In MyYahoo können Benutzer angeben, nach welchen Suchbegriffen gesucht werden soll. Zusätzlich können sie aus einer Kategorienhierarchie Kategorien auswählen, über die sie Informationen bekommen möchten. Diese Kategorienhierarchie ist nicht identisch mit der Kategorienhierarchie, die in Yahoo! benutzt wird. Es handelt

sich vielmehr um eine Kategorienhierarchie zur Klassifikation von Nachrichten.

Konsequenterweise werden in MyYahoo klassifizierte Nachrichten unter dieser Kategorienhierarchie angeboten. Daneben werden Suchergebnisse zu den definierten Volltext-Suchanfragen angeboten. Eine automatische Notifikation (per Email) über Änderungen in MyYahoo ist nicht vorgesehen.

5.2.3 FindUR

Das FindUR-System (vgl. [McG98]) sucht Dokumente mit Hilfe eines Domänenmodells. Das Domänenmodell basiert auf Beschreibungslogik (Description Logic). Zur Suche wird die Verity-Search Engine benutzt. Teile des Domänenmodells werden zu sog. Topic maps zusammengefasst und als Anfrage an die Suchmaschine gegeben. Das Domänenmodell kann über das Web kollaborativ gepflegt werden. Es besteht im Wesentlichen aus einer Konzepthierarchie. Zu einzelnen Konzepten können Synonyme angegeben werden, nach denen in Dokumenten gesucht wird. Außerdem gibt es noch sog. Evidences. Dies sind Wörter oder Phrasen, die keine Synonyme sind, aber trotzdem auf ein bestimmtes Konzept hinweisen. Zwischen den verschiedenen Konzepten können neben der isA-Relation, die durch die Konzepthierarchie ausgedrückt wird, auch noch weitere (einfache) Relationen definiert werden. Die Relationen werden in der Konzepthierarchie vererbt.

Für den Benutzer ist das FindUR-System eine Suchmaschine, in der er ganz „normale“ Volltextanfragen stellen kann. Zusätzlich können sie die Konzepte des Domänenmodells zur Einschränkung des Suchraums benutzen, um die Qualität der Ergebnisse zu erhöhen. Eine personalisierte Schnittstelle für das FindUR-System existiert nicht.

5.2.4 Planet-Onto

Planet-Onto (vgl. [DM99], [MSD00]) ist eine Sammlung von Tools, die um eine Ontologie herum gruppiert werden können, um so Menschen und ihr Wissen zusammenzubringen. Die Ontologie wird mit Web-Onto editiert. In Planet-Onto werden drei Benutzergruppen unterstützt. Diese Benutzergruppen lassen sich gut auf das Information-Brokeringmodell abbilden, das dieser Arbeit zugrunde liegt. Das resultierende Brokeringmodell für Planet-Onto zeigt Abbildung 5.4. Die drei Benutzergruppen von Planet-Onto umfassen:

1. Journalisten schreiben Newsstories, die sie per Email zum System weiterleiten. Die Journalisten sind somit die Anbieter von Informationen. Sie führen die Quellenverwaltungsaufgaben aus, indem sie die Informationen an das System schicken.

2. Wissenseditoren annotieren die Newsstories nach der gegebenen Ontologie oder erweitern diese. Dazu bekommen sie Unterstützung durch ein Tool namens KNOTE. Die Wissenseditoren sind die Broker in diesem System, da sie die Inhalte strukturieren und für die Leser zentralisiert zur Verfügung stellen.
3. Leser lesen die Stories und können ein Interessenprofil anlegen. Sie sind die Konsumenten der im System verfügbaren Informationen. Sie können mit dem form-basierten Tool LOIS Anfragen an das System stellen. Diese Anfragen ergeben zusammen das Interessenprofil.

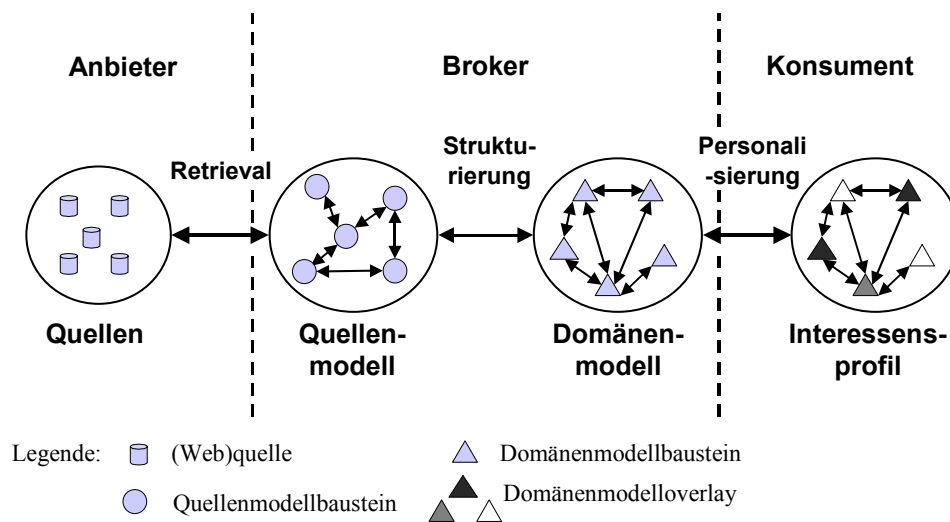


Abbildung 5.4: Planet-Onto-Brokeringmodell

Neue Stories werden über einen Push-Service an die Leser weitergeleitet, sofern die Stories in das Interessenprofil des Lesers passen. In [DM99] wurde angekündigt, dass die Analyse der LOIS-Anfragen zur Anpassung des Interessenprofils genutzt werden kann.

MyPlanet (vgl. [KDM⁺01]) ist eine Erweiterung der Planet-Onto-Architektur. In MyPlanet kann das Interessenprofil mit Interessenkategorien editiert werden. Zu den Kategorien können Schlüsselsequenzen angegeben werden, die zur automatischen Kategorisierung von neuen Stories genutzt werden. Wenn eine Story nicht in allen Slots automatisch belegt werden kann, können sie mit einem form-basierten Tool vervollständigt werden. Damit wird zur automatischen Kategorisierung ein ähnliches Verfahren benutzt, wie es in FaIRS (**F**aceted **I**nformation **R**etrieval **S**ystem, vgl. [Pri00]) Verwendung findet.

Wenn neue Stories in die Datenbasis hineinkommen und auf Interessengebiete von Benutzern passen, werden diese per Email an die entsprechenden Benutzer geschickt. Die Autoren nennen dies proaktiv. Es bleibt anzumerken, dass mit MyPlanet die zuvor angestrebte Adaptivität (s.o.) des Systems nicht eingelöst werden konnte, da das System nur hinsichtlich des Zeitpunktes, an dem er eine Nachricht erhält, für den Benutzer proaktiv erscheint, da er sich nicht aktiv um den Erhalt dieser Nachricht bemühen musste. Das System ist jedoch nicht proaktiv hinsichtlich der Interessen des Benutzers, da diese vom System unangetastet bleiben.

Neben MyPlanet gibt es im Umfeld von Planet-Onto weitere Bestrebungen, Personalisierung und Adaptivität anbieten zu können (vgl. [Nan01]). Dieser Ansatz versucht im Wesentlichen ein termbasiertes, adaptives Informationsfilterungssystem (vgl. Kap. 4.1) aufzubauen. Dieser Ansatz ist, wie viele vergleichbare Systeme, auf explizites Feedback durch den Benutzer angewiesen. Eine Integration in die übrige Planet-Onto-Umgebung ist nicht dokumentiert.

5.3 Ontologiesprachen

Ontologiesprachen dienen der Formulierung von Ontologien. Das IB-Domänenmodell in Broker's Lounge kann ebenfalls als Ontologie angesehen werden. In diesem Abschnitt sollen die in den vorgestellten Ansätzen benutzten Ontologiesprachen auf ihre Nützlichkeit für Information-Brokering untersucht werden. Dabei ist zu beachten,

- dass das IB-Domänenmodell im Allgemeinen von Domänenexperten bearbeitet wird. Diese Domänenexperten sind Experten in ihrer jeweiligen Domäne. Sie sind nicht notwendigerweise Experten für die benutzte Ontologiesprache. Das bedeutet, dass die Formulierung des IB-Domänenmodells auch ohne Spezialkenntnisse in einer Ontologiesprache möglich sein sollte. Diese Anforderung kann durch entsprechende Tools mit grafischer Benutzeroberfläche erfüllt werden. Diese Tools machen die Benutzung von Ontologiesprachen für den Benutzer transparent.
- Außerdem ist zu beachten, dass sich im laufenden Betrieb neue Anforderungen an das IB-Domänenmodell ergeben können, wenn das Angebot eines Brokers z.B. aufgrund einer aktuellen Anfrage thematisch erweitert werden soll.
- Schließlich sollen die Interessen der Konsumenten in einem späteren Schritt auf das IB-Domänenmodell abgebildet werden können. Dies bedeutet, dass die Ontologiesprache entsprechende Sprachkonstrukte bereithalten sollte, mit denen Interessen abgebildet werden können.

Die Rolle von Ontologiesprachen in ontologiebasierten Informationssystemen besteht in der Regel darin, eine Struktur oder ein Regelwerk vorzugeben für mögliche Instanzen oder Fakten. Aus diesem Grund lassen sich fast alle Systeme in eine Struktur- oder Typenebene und eine Instanzenebene einteilen. Die möglichen Instanzen werden durch die Typenebene determiniert. Im Unterschied zu diesen Ansätzen ist die Unterscheidung zwischen den beiden Ebenen in Broker's Lounge aufgeweicht, indem eine zusätzliche Strukturierungsebene zur Abbildung von Interessen eingeführt wird. Dies wird in Kapitel 6 beschrieben.

Nach [FM01] lassen sich Ontologiesprachen in

1. prädikatenlogische Ansätze erster Ordnung,
2. Frame-basierte
3. und beschreibungslogische Ansätze

einteilen.

Beispiele für prädikatenlogische Ansätze sind CycL und KIF (**K**nowledge **I**nterchange **F**ormat, vgl. [Gen91]). CycL ist im Cyc-Projekt (vgl. [LGP⁺90]) entstanden, das sich zum Ziel gesetzt hat, eine Ontologie des Alltagswissens zu erstellen und Alltagswissen somit für KI-Systeme nutzbar zu machen.

Ontolingua (vgl. [Gru93]) basiert auf KIF und erlaubt die Definition von Relationen zwischen Klassen und die Definition von Axiomen, die von den Instanzen erfüllt werden müssen. Die Modellierungsmittel von Ontolingua sind in einer eigenen Frame-Ontologie beschrieben. Bei Bedarf kann diese Frame-Ontologie um weitere Modellierungskonstrukte erweitert werden. Frame Logic ist eher an objekt-orientierten Systemen orientiert. Es erlaubt die Definition von Klassen, Attributen, Wertebereichen (und deren Einschränkung), Vererbung und Axiomen. Ontolingua ist primär für den Austausch von Ontologien konzipiert worden, da sie die Definition von Ontologien in vielen verschiedenen Ontologiesprachen und deren Übersetzung in jeweils andere Ontologiesprachen zulassen.

Zentrales Modellierungsmittel in framebasierten Ansätzen sind Klassen. Frame-basierte Modellierung ist eng verwandt mit objekt-orientierter Modellierung. Für die verschiedenen Klassen können jeweils Attribute definiert werden. Diese Attribute können in ihrem Wertebereich eingeschränkt werden. Die Attribute gelten nur für die Klasse(n), für die sie definiert wurden. Für jede Klasse können Instanzen definiert werden.

Beschreibungslogik (vgl. [BH91]) ist auch unter dem Namen terminologische Logik bekannt. Beschreibungslogik ist ein Fragment von Prädikatenlogik erster Ordnung, die aber noch entscheidbar ist und effiziente Inferenzmechanismen zur Verfügung stellt. Durch die Inferenzmechanismen kann für eine Instanz ihre Zugehörigkeit zu einem Konzept (Klasse) aufgrund ihrer Eigenschaften automatisch

inferiert werden. Dies bringt Vorteile, wenn eine Instanz nicht direkt zu einem Konzept zugeordnet werden kann oder eine Instanz unvollständig definiert wurde.

In einer Beschreibungslogik können Klassen, die hier Konzepte genannt werden, definiert werden. Konzepte können entweder durch eine Aufzählung von Instanzen oder intensional durch die Beschreibung ihrer Eigenschaften definiert werden. Diese Konzepte können eine Konzepthierarchie bilden. Konzepte, die nur ein Vaterkonzept haben, sind primitive Konzepte, während Konzepte mit mehreren Vätern sog. definierte Konzepte sind. Zwischen den Konzepten können zweistellige Relationen definiert werden. Dies kann auch als die Definition von Attributen aufgefasst werden. Attribute heißen hier jedoch Rollen. Die Werte für Rollen können sowohl hinsichtlich ihres Wertebereiches als auch hinsichtlich ihrer Kardinalität eingeschränkt werden. Als effiziente Implementierung von Beschreibungslogik steht z.B. das FaCT-System (vgl. [Hor98]) zur Verfügung.

OCML (**O**ntology **C**oncept **M**odelling **L**anguage) ist Ontolingua nachempfunden, jedoch im Gegensatz zu Ontolingua effizient entscheidbar. OCML findet im Planet-Onto-Projekt (vgl. [MSD00], siehe Kap. 5.2.4) zur Formulierung von Ontologien Verwendung. OCML-Ontologien können mit wenig Aufwand in Ontolingua überführt werden. OIL (vgl. [FHH⁺00]) ist ein Beispiel für eine Sprache, die Beschreibungslogik und frame-basierte Ansätze in sich vereint. Um das in dieser Arbeit vorgeschlagene IB-Domänenmodell in den Kontext der hier beschriebenen Ontologiesprachen einzuordnen, wird das IB-Domänenmodell in Kapitel 6.4 mit OIL verglichen. OIL wurde im On-To-Knowledge-Projekt⁶ entwickelt. Innerhalb des Semantic Web-Projektes wurde außerdem eine Vereinheitlichung mit DAML zu DAML+OIL erreicht (vgl. [Hor01]), so dass OIL eine weite Verbreitung gefunden hat.

OIL und DAML+OIL fokussieren auf die Beschreibung von Zusammenhängen zwischen den verschiedenen Konzepten auf der Schemaebene. Die Relationen zwischen den verschiedenen Konzepten werden später auf der Instanzenebene von konkreten Instanzen mit Leben gefüllt. Eine Unterstützung für die Arbeit mit konkreten Instanzen existiert in OIL nicht, wenngleich in DAML+OIL erste Schritte zum Umgang mit sog. Individuen (vgl. [Hor02]) unternommen wurden.

Dagegen unterscheidet Telos (vgl. [MBJK90]), bzw. O-Telos nicht zwischen Konzepten und Instanzen. Stattdessen können zwischen verschiedenen Objekten Instanziierungsbeziehungen aufgebaut werden. Dadurch können beliebig viele Beschreibungsebenen zur Metamodellierung aufgebaut werden. Auf der untersten Ebene werden konkrete Instanzen verwaltet. In der Ebene darüber werden Typen definiert, zu denen die Instanzen eine Instanziierungsbeziehung unterhalten. Über der Typenebene kann wiederum eine Meta-Ebene und darüber eine Meta-Meta-Ebene definiert werden. Als grundlegende Modellierungsmittel kennt

⁶<http://www.ontoknowledge.org/>

O-Telos, Objekte, annotierte Referenzen zwischen Objekten und die Instanziierungsbeziehung. Alle weiteren Sprachkonstrukte, wie z.B. Generalisierungsbeziehungen zwischen Objekten basieren auf diesen Grundkonstrukten (vgl. [JJNS98]). Neben der Möglichkeit beliebig viele Meta-Ebenen zu definieren, lässt O-Telos im Vergleich zu OIL auch die Formulierung von Integritätsbedingungen zu. ConceptBase ist eine Implementierung von O-Telos, die neben der Modellierung von O-Telos-Modellen auch die persistente Speicherung von Objekten (eines O-Telos-Modells) zulässt.

Zu der leichten Verständlichkeit von Ontologiesprachen gehört auch die leichte Bedienbarkeit und Erlernbarkeit der entsprechenden Editoren für Ontologiesprachen, auf die in Kapitel 5.4 eingegangen wird.

Neben der möglichst leichten Bedienbarkeit des IB-Domänenmodells (der Ontologie) müssen die Interessen von Informationskonsumenten auf das IB-Domänenmodell leicht abbildbar sein. Diese Abbildung sollte wie der Rest des Systems unabhängig von einer bestimmten Domäne möglich sein. Das bedeutet, dass auf der Ebene der Ontologiesprachen geeignete Modellierungskonstrukte zur Verfügung gestellt werden müssen, die zur Abbildung von Interessen geeignet sind.

Fasst alle Frame-basierten und beschreibungslogischen Ansätze sowie sich entwickelnde Webstandards wie RDF (vgl. [Las98]) weisen ähnliche Modellierungskonstrukte (vgl. [LM01]) auf. Diese sind im wesentlichen Konzepte (Klassen) und Relationen zwischen den Klassen. Diese Relationen können entweder zur Attributierung von Konzepten genutzt werden oder um die verschiedenen Konzepte mit einer Subsumptionsrelation zu einer Konzepthierarchie anzuordnen.

Subject Gateways setzen dagegen auf die Klassifizierung von Websites durch ein Kategoriensystem. Die verwendete Kategorienhierarchie kann mit der Konzepthierarchie der frame-basierten und beschreibungslogischen Ansätzen verglichen werden. Die Funktion ist jedoch unterschiedlich: Während die Kategorienhierarchie von Subject Gateways ausschließlich zur Beschreibung von Websites genutzt wird, beschreibt die Konzepthierarchie nach der Ontologie-Definition von Gruber (vgl. [Gru93] sowie Kap. 2.1.3) eine von der Nutzergemeinde geteilte Konzeptualisierung der modellierten Domäne. Die geteilte Konzeptualisierung umfasst damit alle Dinge, die in der modellierten Domäne für die Nutzergemeinde von Belang sind. Die Konzepthierarchie, die der Ontologie zugrunde liegt, wird benutzt, um die verschiedenen Konzepte voneinander unterscheiden zu können. Die Hierarchie wird damit zur Typisierung der verschiedenen Konzepte benutzt.

Die Kategorienhierarchie eignet sich gut zur Abbildung von Benutzerinteressen, da man voraussetzen kann, dass die Benutzer sich für Websites interessieren, wenn sie Subject Gateways nutzen. Die Kategorienhierarchie kann deswegen zur Beschreibung der thematischen Eigenschaften von Websites genutzt werden. Dies erlaubt eine Abbildung auf die thematischen Vorlieben von Benutzern. Es bleibt einzuwenden, dass sich Subject Gateways, wie auch der Name impliziert, auf die

Beschreibung des thematischen Aspektes von Websites beschränken und andere Aspekte ausblenden, für die Benutzer sich interessieren könnten.

Dagegen beschreiben die Konzepthierarchien in erster Linie verschiedene Dinge in der Welt, die sich in den verschiedenen Konzepten niedergeschlagen haben. Eine Abbildung der Benutzerinteressen auf die Konzepthierarchie ist in analoger Form zu den Subject Gateways möglich, hat aber eine geringere Aussagekraft, da dadurch nur eine Aussage möglich wird, für welche Konzepte (Dinge der realen Welt) sich Benutzer interessieren, nicht aber, wie diese Konzepte beschaffen sein müssen (welche Eigenschaften die Konzepte haben müssen), damit sich Benutzer für diese Konzepte interessieren. Man mag einwenden, dass man diese Eigenschaften durch spezielle Relationen beschreiben könnte. Man bewegt sich dann jedoch auf der Ebene eines konkreten Domänenmodells und hat die Meta-Ebene verlassen. Dadurch ist eine domänenunabhängige Beschreibung von Interessen nicht mehr möglich.

Subject Gateways nutzen zur Abbildung der Interessen dagegen direkt die Kategorien. Man könnte einwenden, dass man die Konzepthierarchie einfach um die Beschreibung von Eigenschaften erweitern könnte. Hier taucht das Problem der Aspektvermischung auf. D.h., die Subsumptionsrelation zwischen zwei Konzepten wird zur Beschreibung von unterschiedlichen Aspekten benutzt. Dies führt zwingend zum Facettierungsproblem (vgl. Kap. 2.3.2). D.h., die Konzepthierarchie wird stark aufgebläht, da alle Permutationen möglicher Eigenschaften in die Hierarchie mit aufgenommen werden müssten. Um diese Problematik zu entschärfen, kann man beide Ansätze miteinander verbinden: Die Konzepte werden zur Beschreibung der relevanten Dinge einer Domäne benutzt, während die Kategorien zur Beschreibung der Eigenschaften von Konzepten dienen. Dies wird in Kapitel 6 genauer ausgeführt.

5.4 Modellierungstools für Ontologien

Es gibt eine Vielzahl von Werkzeugen zur Erstellung und Modellierung von Ontologien. Die oben beschriebenen Systeme sind meist mit einer Ontologiesprache und damit mit einem Ontologieeditor assoziiert. So werden die Ontologien in Planet-Onto mit Web-Onto (vgl. [Dom98], [MSD00]) editiert. Web-Onto basiert auf OCML. Ontobroker-Ontologien (bzw. On-To-Knowledge-Ontologien) können mit OILed oder OntoEdit in OIL modelliert werden.

Ontolingua ist eine Umgebung, die nicht auf eine bestimmte Ontologiesprache festgelegt ist, sondern vielmehr die Integration verschiedenster Ansätze anstrebt, um so ein allgemeines Werkzeug bereitzustellen. Ein ebenfalls universeller Anspruch wird im Protégé-Ontologie-Editor (vgl. [EFHM99]) verfolgt. Protégé ist ein Framework zur Definition von Ontologien. Protégé wird eingesetzt, um Edi-

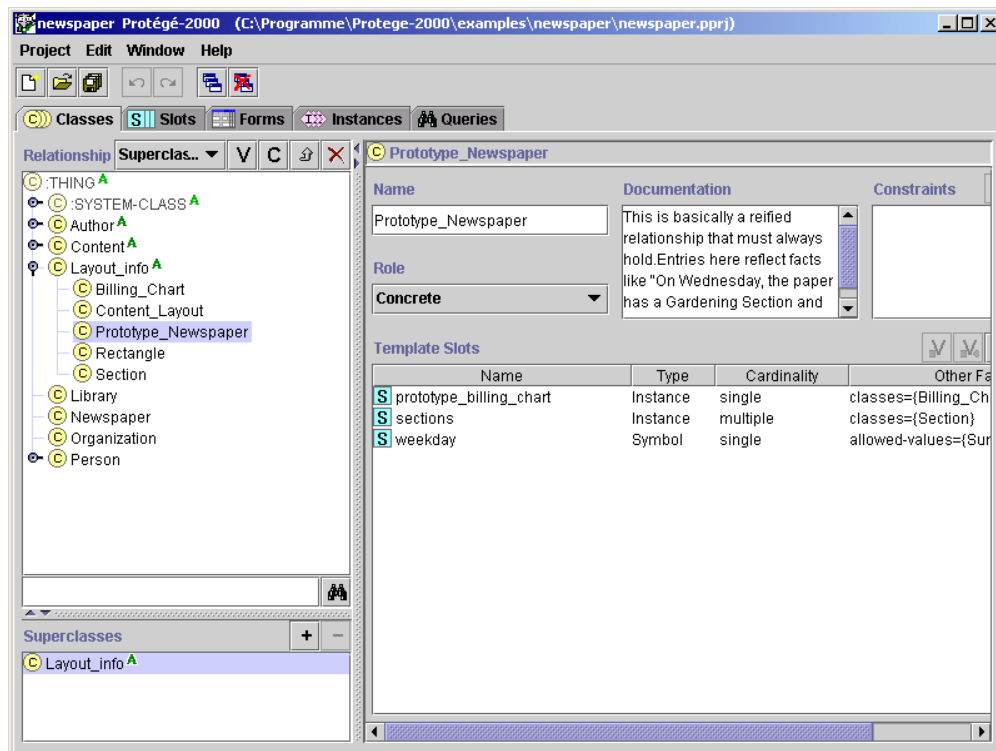


Abbildung 5.5: Protégé-Ontologie-Editor

toren zu generieren. Diese Editoren werden benutzt, um zuvor in Protégé definierte Ontologien mit konkreten Inhalten zu füllen. Das Editieren von Instanzen direkt in Protégé ist nicht möglich. Dies basiert auf der Idee, dass die Ontologie von einem Wissensingenieur aufgebaut wird, während die konkreten Instanzen von einem Domänenexperten eingegeben werden. Dies ist insbesondere in Domänen problematisch, in denen die Phasen der Ontologiedefinition und der Füllung des Systems mit einzelnen Fakten (Instanzen), nicht voneinander getrennt werden können. Wird die Domänenstruktur verändert, muss jeweils ein neuer Editor erzeugt werden.

ConceptBase (vgl. [JJNS98]) dagegen kennt die Unterscheidung zwischen Schemaebene und Instanzenebene nicht, da es zur Definition beliebig vieler Beschreibungsebenen benutzt werden kann. ConceptBase kann daher auf der einen Seite zum Editieren von einzelnen Instanzen genutzt werden. Auf der anderen Seite kann man mit ConceptBase auch jederzeit Änderungen auf den höheren Beschreibungsebenen vornehmen, die dann sofort wirksam werden. Zur Verdeutlichung der Zusammenhängen zwischen den verschiedenen Objekten dient ein graphischer Objektbrowser.

Die Benutzerschnittstelle von Protégé zur Ontologiedefinition basiert auf ei-

nen Baum auf der linken Seite, der die Klassenhierarchie visualisiert und einer Mehrzeilentabelle (multiple column list), zur Anzeige der jeweiligen Slots. Eine Ontologie besteht aus Klassen und Slots. Daneben gibt es außerdem Meta-Klassen und Meta-Slots. Protégé ist offen für Erweiterungen, spezielle Erweiterungen können jederzeit hinzugefügt werden.

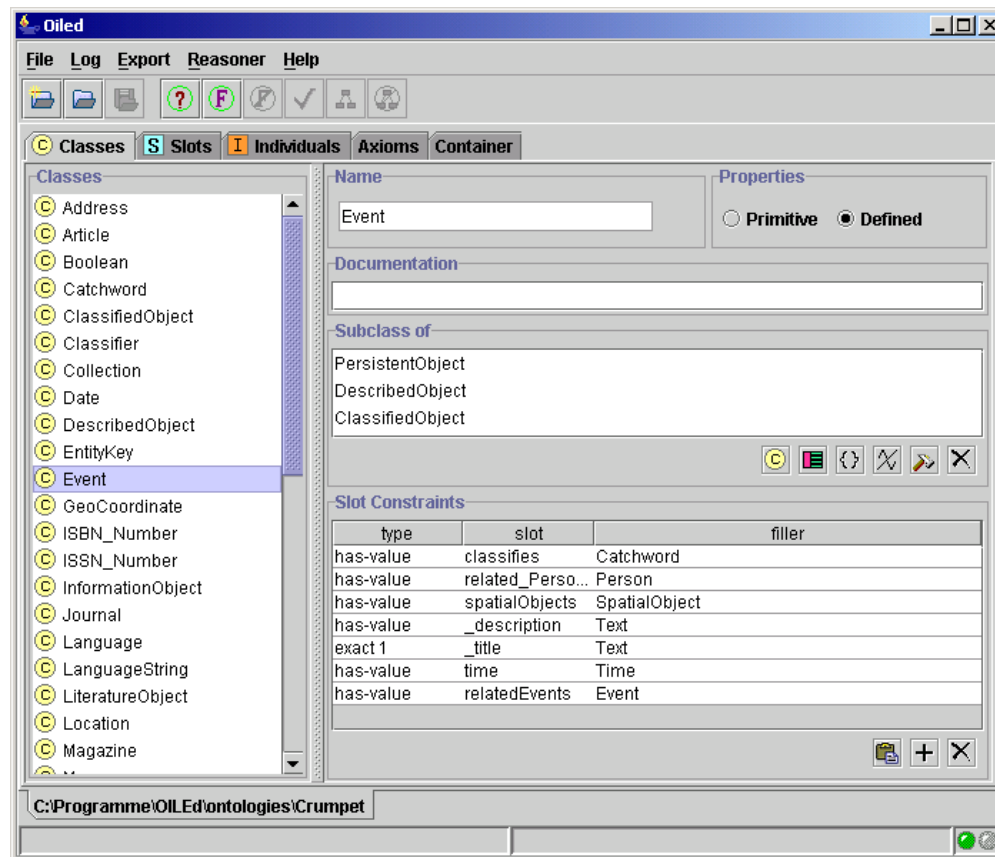


Abbildung 5.6: Oiled-Editor

Protégé und Web-Onto haben gegenüber Ontolingua den Vorteil, dass sie leichter zu erlernen sind (vgl. [DSW⁺99]). Sie kommen daher für einen Einsatz im Information-Brokering eher in Frage, als Ontolingua. Oiled und OntoEdit sind in ihrem Interface-Design im Wesentlichen Protégé nachempfunden und kommen daher ebenfalls in Frage.

5.5 Positionierung von Broker's Lounge

Broker's Lounge lässt sich in zweierlei Hinsicht gegen die hier vorgestellten ontologiebasierten Informationssysteme abgrenzen. Zum einen liegt bei Broker's

Lounge der Fokus auf der Unterstützung von Personalisierung. Darauf wird im nächsten Abschnitt und in Tabelle 5.1 eingegangen.

Zum anderen unterscheidet sich Broker's Lounge von alle anderen Ansätzen dadurch, dass verschiedenste Brokeringmodelle unterstützt werden können. Dies wird in übernächsten Abschnitt und in Tabelle 5.2 deutlich. Die formale Ausdrucksmächtigkeit des IB-Domänenmodells von Broker's Lounge wird erst in Kapitel 6.4 durch die Abbildung des IB-Domänenmodells auf OIL beschrieben, da das IB-Domänenmodell erst in Kapitel 6 eingeführt wird.

5.5.1 Personalisierungsansätze

In [Die00] und [DCGR98] wird die Wichtigkeit der Rolle der Benutzer betont. Eine prototypische Umsetzung dieser Ideen findet sich im CoMMA-Projekt (vgl. Kap. 5.1.4). Dort versucht das System, sich an die Interessen des Benutzers anzupassen. Es kann damit als adaptives System bewertet werden. Es geht aus den Veröffentlichungen leider nicht klar hervor, ob die angestrebte Adaptivität auch tatsächlich eingelöst werden konnte. Da Benutzer ihre Interessen auch selbst angeben können, ist das System zusätzlich adaptierbar. Benutzer können sich aktiv Informationen beim System besorgen (pull) und bekommen neue Einträge im System per Email zugesendet, wenn sie zu ihrem Interessenprofil passen (Push). Adaptivität wurde auch für Planet-Onto (vgl. Kap.5.2.4) angekündigt, bisher jedoch durch weitere Publikationen nicht voll eingelöst. Stattdessen stellt Planet-Onto mit MyPlanet eine personalisierte Schnittstelle zur Verfügung, die den personalisierten Zugriff auf die Inhalte unterstützt. Änderungen werden den Benutzern abhängig von ihrem Profil zugesendet. Auch Ontobroker stellt seinen Benutzern einen Push-Dienst und eine personalisierte Schnittstelle zur Verfügung.

ConceptBase⁷ erlaubt die Definition von Sichten (Views) auf die Inhalte des Modells (vgl. [SJ00]). Dadurch können persönliche Vorlieben auf das modellierte Modell abgebildet werden. Für die Sichten existiert ein Update-Mechanismus, der ein Update von materialisierten Sichten außerhalb des Servers erlaubt. Dadurch kann ein flexibler Push-Dienst für die Inhalte des Modells abhängig von den Interessen der Benutzer ermöglicht werden.

Das Internetportal Yahoo! bietet seinen Benutzern ebenfalls eine personalisierte Schnittstelle an. Es verfügt jedoch nicht über einen Push-Dienst. Im Gegensatz zu den anderen beschriebenen Systemen ist Yahoo! nur bedingt domänenunabhängig, da es nur eine Domäne, die ganze Welt, beschreibt. Subject Gateways sind zwar domänenunabhängig und werden für jeweilige Domänen angepasst, lassen jedoch eine personalisierte Anfrageschnittstelle vermissen und sind somit auch

⁷ConceptBase wurde nicht in die Tabellen 5.1 und 5.2 aufgenommen, da ConceptBase nur die Strukturierung und Personalisierung nicht aber den Brokeringprozess Retrieval unterstützt.

System	adaptierbar	adaptiv	Push-Dienst	domänen-unabhängig
CoMMA	x	x	x	x
Planet-Onto	x	-	x	x
Ontobroker	x	-	x	x
Yahoo!	x	-	-	o
Subject Gateways	x	-	-	x
FindUR	o	-	-	x
KnowWeb	-	-	-	x
SHOE	-	-	-	x
Broker's Lounge	x	x	x	x

Tabelle 5.1: Personalisierung in den verschiedenen ontologiebasierten Informationssystemen

nicht als adaptierbar zu beschreiben. Außerdem weisen sie als Konsequenz aus der fehlenden Personalisierung auch keinen Push-Dienst auf. Genau wie Subject Gateways unterstützt KnowWeb nicht die Personalisierung und ist damit weder adaptierbar, noch adaptiv und hat keinen Push-Dienst.

Die oben beschriebenen Personalisierungsansätze sind in Tabelle 5.1 zusammengefasst. Die Umsetzung meiner Ideen in Broker's Lounge (siehe Kapitel 8), ist am ehesten mit CoMMA zu vergleichen, da Broker's Lounge volle Personalisierungsunterstützung anbietet. D.h., Brokers's Lounge ist wie CoMMA sowohl ein adaptierbares als auch ein adaptives System.

Im Unterschied zu CoMMA bietet Broker's Lounge eine bessere Kontrolle des Benutzermodells durch die Benutzer, da sie ihr Profil jederzeit vor sich haben, es damit jederzeit ändern können und gravierende Änderungen nur mit Einverständnis des Benutzer gemacht werden können. Im Extremfall können Benutzer die adaptive Komponente abstellen, um so mit einem nunmehr adaptierbaren System zu arbeiten. Diese strikte Kontrolle der adaptiven Komponente löst die Forderungen von [Fis01] nach einem ausgewogenen Verhältnis von Adaptivität und Adaptierbarkeit und der vollen Kontrolle des Benutzerprofils durch die Benutzer ein.

5.5.2 Brokeringmodelle

Die verschiedenen in diesem Kapitel vorgestellten Ansätze und ihre Brokeringmodelle werden in Tabelle 5.2 gegenübergestellt. Alle Ansätze sind jeweils auf ein Brokeringmodell festgelegt. Die Festlegung auf ein Brokeringmodell ist größtenteils durch die gewählte Architektur determiniert. Sematic-Web-Systeme wie

Name	Quellenverwaltungs- aufgaben	Domänenverwal- tungsaufgaben	kundenorientierte Aufgaben
Semantic-Web- Systeme	Anbieter/Broker	Anbieter	Konsument
COMMA	Anbieter/Broker	Anbieter	System/Konsument
Planet-Onto	Anbieter	Broker	Konsument
Subject Gateways	Broker	Broker	Konsument
Broker's Lounge	beliebig	beliebig	System/beliebig

Tabelle 5.2: Aufgabenverteilung bei den verschiedenen ontologiebasierten Informationssystemen

z.B. Ontobroker sehen die Annotation von Dokumenten nach einer zentralen Ontologie vor. In diesem Fall werden die Domänenverwaltungsaufgaben vom Informationsanbieter durchgeführt. Die Quellenverwaltungsaufgaben werden zwischen dem Informationsanbieter und einem zentralen Broker geteilt, da der Anbieter neue Dokumente beim Broker anmeldet und diese Dokumente dann vom Broker regelmäßig auf Veränderungen untersucht werden. Die Annotation durch den Anbieter erlaubt die dezentrale Strukturierung von Dokumenten nach einer zentralen Ontologie.

Klassifikationssysteme gehen nicht von der optimistischen Annahme aus, dass die Anbieter von Informationen diese nach einer zentralen Ontologie annotieren und führen die Domänenverwaltungsaufgaben in einer zentralen Instanz durch. Broker's Lounge dagegen erlaubt die flexible Verteilung von Aufgaben zwischen den beteiligten Parteien und kann somit jeweils auf die Bedürfnisse der jeweiligen Domäne reagieren.

Alle vorgestellten Systeme lassen die kundenorientierten Aufgaben von den Konsumenten selbst durchführen. D.h., die Konsumenten passen das System an ihre Bedürfnisse an. Es handelt sich somit um adaptierbare Systeme (siehe Tab. 5.1). Das CoMMA-Projekt (vgl. Kap. 5.1.4) sieht zusätzlich die Unterstützung des Konsumenten durch eine adaptive und adaptierbare Komponente vor. Broker's Lounge ist ebenfalls als adaptierbares und adaptives System konzipiert. Die Unterstützung bei der Profilbildung im kundenorientierten Prozess durch das System bleibt jedoch nicht auf Konsumenten beschränkt, da die kundenorientierten Aufgaben, je nach Brokeringmodell, auch von Brokern oder Anbietern durchgeführt werden können.

In der Domäne Forschungsförderung, die in ELFI (vgl. Kap. 9.1) bearbeitet wurde, ist es Aufgabe von sog. Forschungsreferenten, die Wissenschaftler ihrer Institution über Möglichkeiten zur Förderung ihrer Forschungen zu informieren. In diesem Fall sind die Forschungsreferenten Broker für ihre Wissenschaftler. Sie

können eine Ausprägung von Broker's Lounge für die Domäne Forschungsförderung nutzen, um für ihre Wissenschaftler Profile über deren Interessen anzulegen.

Kapitel 6

Information-Brokering-Domänenmodell

Das IB-Domänenmodell nimmt in dem in Kapitel 3 beschriebenen Information-Brokeringmodell die zentrale Rolle ein. Das IB-Domänenmodell beinhaltet die Informationen, mit denen der Broker handelt. Auf der einen Seite müssen die Inhalte von Dokumenten mit den Inhalten des IB-Domänenmodells synchronisiert werden. Auf der anderen Seite müssen die Interessen der Konsumenten auf die Inhalte des IB-Domänenmodells abgebildet werden. Als weitere Anforderung kommt die leichte Erweiterbarkeit des IB-Domänenmodells zu jedem Zeitpunkt hinzu, um so das IB-Domänenmodell ggf. an neue Anforderungen anpassen zu können. Damit das IB-Domänenmodell auch von Laien auf dem Gebiet der Domänenmodellierung (bzw. auf dem Gebiet von Ontologiesprachen) bedient werden kann, sollte das IB-Domänenmodell ähnlich wie ontologiebasierte Informationssysteme mit einer grafischen Benutzeroberfläche editierbar sein. Dieser Aspekt wird in diesem Kapitel noch ausgespart, da die Verwaltung des IB-Domänenmodells im Broker's Lounge-Kapitel (Kap. 8) erläutert wird. Dort werden alle Schnittstellen für die verschiedenen Aufgaben im Information-Brokeringmodell beschrieben.

In diesem Kapitel stehen die Modellierungsmittel zur Abbildung der Inhalte (von Dokumenten) und zur Abbildung von Interessen im Vordergrund der Betrachtung. Im nächsten Abschnitt werden aus diesem Grund die Modellierungsmittel der in dieser Arbeit beschriebenen ontologiebasierten und adaptiven Informationssysteme beleuchtet. Daraus ergibt sich die Basisstruktur des IB-Domänenmodells. Die gewählte Basisstruktur und ihre Nutzung wird an einem Beispiel aus der Domäne Forschungsförderung illustriert. Um das IB-Domänenmodell gegen ontologiebasierte Informationssysteme abzugrenzen, wird das IB-Domänenmodell mit der Ontologiesprache OIL verglichen. OIL wurde zum Abgleich gewählt, weil es eine weite Verbreitung gefunden hat und die Sprachkonstrukte in den verschiedenen Ontologiesprachen jeweils weitestgehend vergleichbar sind (vgl. [LM01]).

6.1 Modellierungsmittel in adaptiven und ontologiebasierten Informationssystemen

In ontologiebasierten Informationssystemen spielen Konzepte eine wichtige Rolle, da Ontologien im Wesentlichen aus Konzepten bestehen. Es geht bei diesen Systemen um eine Abbildung der realen Welt in einem Modell (der Ontologie). Die Konzepte bilden zusammen eine Konzepthierarchie. Außerdem können die verschiedenen Konzepte in Relation zu anderen Konzepten gesetzt werden. Die Konzepte bilden eine Strukturebene, die die Strukturen der Instanzen in der Instanzenebene determinieren. Jede neue Instanz muss in einem Klassifikationsprozess eindeutig einem Konzept aus der Konzepthierarchie zugeordnet werden. Durch diese Zuordnung zu einem Konzept werden die Instanzen typisiert. Zusätzlich muss diese Instanz die Relationen des zugeordneten Konzeptes mit Leben füllen.

Bei adaptiven, domänenmodellbasierten Informationssystemen steht ebenfalls ein Klassifikationsprozess im Mittelpunkt: Für einzelne Items, die Benutzern angeboten werden sollen, ist zu entscheiden, ob dieses Item für den Benutzer interessant ist oder nicht und wenn ja, wie relevant dieses Item für den Benutzer ist. Um die Relevanz von Items bewerten zu können, werden dessen Eigenschaften beschrieben. Diese Eigenschaften sind zumeist thematischer Natur. Die verschiedenen, möglichen Eigenschaften wiederum können durch Kategorien abgebildet werden. Dies wird von fast allen domänenmodellbasierten, adaptiven Informationssystemen benutzt (siehe Kap. 4.5 und Tab. 4.1). Die Kategorien ermöglichen eine Vereinfachung der Relevanzbewertung: Die Relevanz von einzelnen Items wird durch die Relevanz der Kategorien berechnet, mit denen sie klassifiziert wurden. Die Kategorien können gleichzeitig zur Filterung benutzt werden, da sie eine leichte Gruppierung von Items ermöglichen.

Strukturell betrachtet bilden sowohl Kategorien als auch Konzepte jeweils eine Hierarchie (vgl. Kap. 3.2). Sie unterscheiden sich dadurch, dass Konzepte zusätzlich Relationen zu weiteren Konzepten ausbilden können.

Alle Systeme, die in Kap. 4.5 bzw. Tab. 4.1 beschrieben wurden, kennen jeweils nur einen Typ von Items. Dabei handelt es sich z.B. um Produktbeschreibungen oder Dokumente. Das hat den Vorteil, dass das Domänenmodell aus einer Kategorienhierarchie bestehen kann. Die Kategorienhierarchie wird dabei in ähnlicher Weise benutzt, wie die Konzepthierarchie in Ontologien: Die Zuordnung zu einer Kategorie dient gleichzeitig der Typisierung. Die Kategorienhierarchie kann damit auch als Ontologie zur Beschreibung einer Eigenschaft eines Itemtyps (z.B. von Dokumenten) aufgefasst werden.

Es bleibt kritisch anzumerken, dass dieser Ansatz nur solange funktioniert, wie das System nur mit einem Typ von Informationen (Items) handelt, weil nur

solange ein problemloser Wechsel der Modellierungsperspektive möglich ist, wie nur ein Informationstyp (Itemtyp) und eine Eigenschaft im Domänenmodell modelliert werden. In diesem Fall kann ein und dieselbe Kategorienhierarchie sowohl zur Typisierung als auch zur Beschreibung der (thematischen) Eigenschaft benutzt werden. Da die Interessen auf die Kategorienhierarchie abgebildet werden, ist es wichtig, dass die von der Hierarchie beschriebene Eigenschaft zur Beschreibung von Benutzerinteressen geeignet ist. Dies reflektiert die von Timm et al. [TR98] eingeführten Kundenmerkmale.

Ontologiebasierte Systeme, die Personalisierung unterstützen (siehe Kap. 5.5, bzw. Tab. 5.1), versuchen die Interessen ihrer Benutzer auf die Konzepthierarchie ihrer Ontologie abzubilden. Dies ist prinzipiell ähnlich zu bewerten wie das Vorgehen von adaptiven Systemen: Die Typisierungs-Kategorienhierarchie wird zur Relevanzbewertung oder zur Filterung von Informationen benutzt. D.h., es wird postuliert, dass der Benutzer, der sich für die angebotenen Konzepte interessiert die Relevanz der einzelnen angebotenen Informationen durch die Typisierungs-Kategorienhierarchie bewertet werden kann. In diesem Fall gibt es also ebenfalls einen Informationstyp (die Konzepte) und eine Kategorienhierarchie (die Konzepthierarchie, die als Typisierungs-Kategorienhierarchie aufgefasst wird) zur Relevanzbewertung.

An diesem Punkt ist anzumerken, dass Benutzerinteressen nicht notwendigerweise eindimensional sein müssen: Benutzerinteresse kann in mehreren Dimensionen bewertet werden, weil Benutzer ihre Entscheidungen (z.B. Kaufentscheidungen) an mehreren Eigenschaften (des Produktes) festmachen und sich nicht unbedingt nach einer einzelnen Eigenschaft, z.B. der Farbe (des Produktes) entscheiden. Deswegen ist für ein adaptives, domänenmodellbasiertes System zu fordern, dass es die Modellierung mehrerer, unabhängiger Eigenschaften zulässt, die für die Interessenmodellierung herangezogen werden können. Benutzer können sich nicht nur für mehrere Eigenschaften eines Informationstypen interessieren, sie können auch Informationen über mehrere möglicherweise miteinander verwandte Informationstypen wünschen. Um dies berücksichtigen zu können, ist eine Kombination von Konzepten und Kategorien nötig: Die Konzepte werden benutzt, um die Informationstypen zu beschreiben, während die Kategorien (bzw. die daraus aufgebauten Kategoriehierarchien) benutzt werden, um die Eigenschaften von Informationstypen (Konzepten) aus Benutzersicht zu beschreiben. Diese Kombination von Kategorien und Konzepten bildet das Fundament des IB-Domänenmodells, das im nächsten Abschnitt beschrieben wird.

6.2 Basisstruktur des Domänenmodells

Das Domänenmodell für Information-Brokering wird in Definition 6.1 formal definiert. Dort werden die Definitionen für den Konzept- (vgl. Definition 3.3) und den Kategoriegraphen (vgl. Definition 3.4) zusammengeführt, um so die Ansätze adaptiver und ontologiebasierter Informationssysteme zusammenzubringen. Um Interessen mehrdimensional abbilden zu können, werden die Kategorien zusätzlich jeweils einem sog. Kategorietypen zugeordnet. Durch die Bindung von Kategorien an jeweils genau einen Kategorietypen wird ein facettiertes Kategoriensystem möglich. Die Facetten werden dabei durch die Kategorietypen repräsentiert. Neben Konzepten und Kategorien werden außerdem noch Attribute (durch Attributtypen und Attributwerte) eingeführt, die zur Repräsentation von Attributen benutzt werden, die für die Interessenmodellierung von geringerer Bedeutung sind als Eigenschaften, die durch Kategorien bzw. Kategorietypen modelliert werden. Die Zusammenhänge zwischen Konzepten, Kategorietypen, Attributen, Kategorien und Informationseinheiten werden in Abbildung 6.1 verdeutlicht.

Definition 6.1 (IB-Domänenmodell)

Ein IB-Domänenmodell ist ein gerichteter Graph, der durch den 39 -Tupel $G_{Dom} = (Kon, KatTyp, AttTyp, I, Kat, Att, RelKon, IsA, RelKatTyp, RelAttTyp, InstanceOfKon, InstanceOfKatTyp, InstanceOfAttTyp, RelI, O, Kla, RelAtt, \alpha RelKon, \omega RelKon, \alpha IsA, \omega IsA, \alpha RelKatTyp, \omega RelKatTyp, \alpha RelAttTyp, \omega RelAttTyp, \alpha InstanceOfKon, \omega InstanceOfKon, \alpha InstanceOfKatTyp, \omega InstanceOfKatTyp, \alpha InstanceOfAttTyp, \omega InstanceOfAttTyp, \alpha RelI, \omega RelI, \alpha O, \omega O, \alpha Kla, \omega Kla, \alpha RelAtt, \omega RelAtt)$ beschrieben wird und folgende Eigenschaften hat:

1. *$Kon, KatTyp, AttTyp, I, Kat$ und Att sind Eckenmengen des Graphen G_{Dom} . Kon bezeichnet die Menge aller Konzepte, $KatTyp$ repräsentiert die Menge aller Kategorietypen und $AttTyp$ beschreibt die Menge aller Attributtypen. Kon ist eine nicht leere Menge ($Kon \neq \emptyset$). I bezeichnet die Menge aller Informationseinheiten, Kat repräsentiert die Menge aller Kategorien und Att beschreibt die Menge aller Attributwerte.*
2. *$RelKon, IsA, RelKatTyp, RelAttTyp, InstanceOfKon, InstanceOfKatTyp, InstanceOfAttTyp, RelI, O, Kla$ und $RelAtt$ sind Pfeilmengen in G_{Dom} :*
 - (a) *$RelKon$ bezeichnet die Menge aller einfachen Beziehungen zwischen Konzepten.*

- (b) *IsA* repräsentiert die Menge aller Generalisierungsbeziehungen zwischen Konzepten.
- (c) *RelKatTyp* beschreibt die Menge aller Klassifikationsbeziehungen zwischen Kategorietypen und Konzepten.
- (d) *RelAttTyp* steht für die Menge aller Relationen zwischen Konzepten und Attributtypen.
- (e) *InstanceOfKon* ist die Menge aller Beziehungen zwischen Informationseinheiten und Konzepten.
- (f) *InstanceOfKatTyp* bezeichnet die Menge aller Beziehungen zwischen Kategorien und Kategorietypen.
- (g) *InstanceOfAttTyp* bezeichnet die Menge aller Beziehungen zwischen Attributen und Attributtypen.
- (h) Die Menge aller Beziehungen zwischen Informationseinheiten wird durch die Pfeilmenge *RelI* ausgedrückt.
- (i) *O* bezeichnet die Menge aller ist-Oberkategorie-von-Beziehungen zwischen Kategorien.
- (j) *Kla* repräsentiert die Menge der Klassifikationsbeziehungen zwischen Kategorien und Informationseinheiten.
- (k) *RelAtt* beschreibt die Menge aller Beziehungen zwischen Informationseinheiten und Attributen.

3. Alle Pfeil- und Eckenmengen sind paarweise disjunkt.

4. $\alpha RelKon, \omega RelKon, \alpha IsA, \omega IsA, \alpha RelKatTyp, \omega RelKatTyp, \alpha RelAttTyp, \omega RelAttTyp, \alpha InstanceOfKon, \omega InstanceOfKon, \alpha InstanceOfKatTyp, \omega InstanceOfKatTyp, \alpha InstanceOfAttTyp, \omega InstanceOfAttTyp, \alpha RelI, \omega RelI, \alpha O, \omega O, \alpha Kla, \omega Kla, \alpha RelAtt$ und $\omega RelAtt$ sind Abbildungen. $\alpha X(x)$ heißt Anfangsecke des Pfeils x . $\omega X(x)$ heißt Endecke des Pfeils x .

5. $\alpha RelKon : RelKon \rightarrow Kon \wedge \omega RelKon : RelKon \rightarrow Kon$.

6. $\alpha IsA : IsA \rightarrow Kon \wedge \omega IsA : IsA \rightarrow Kon$. Der gerichtete Graph $G'_{Dom} = (Kon, IsA, \alpha IsA, \omega IsA)$ ist kreis- und schlingenfrei. D.h., es existiert kein $isA \in IsA$, für das $\alpha IsA(isA) = \omega IsA(isA)$ gilt.

7. $\alpha RelKatTyp : RelKatTyp \rightarrow KatTyp \wedge \omega RelKatTyp : RelKatTyp \rightarrow Kon$.

8. $\alpha RelAttTyp : RelKatTyp \rightarrow Kon \wedge \omega RelKatTyp : RelKatTyp \rightarrow AttTyp$.
9. $\alpha InstanceOfKon : InstanceOfKon \rightarrow I \wedge \omega InstanceOfKon : InstanceOfKon \rightarrow Kon$. Für alle Informationseinheiten $i \in I$ gilt:
 - $\exists instanceOfKon \in InstanceOfKon, kon \in Kon : \alpha InstanceOfKon(instanceOfKon) = i \wedge \omega InstanceOfKon(instanceOfKon) = kon$ und
 - $\neg \exists instanceOfKon' \in InstanceOfKon, kon' \in Kon : \alpha InstanceOfKon(instanceOfKon') = i \wedge \omega InstanceOfKon(instanceOfKon') = kon', kon \neq kon'$
10. $\alpha InstanceOfKatTyp : InstanceOfKatTyp \rightarrow Kat \wedge \omega InstanceOfKatTyp : InstanceOfKatTyp \rightarrow KatTyp$. Für alle Kategorien $kat \in Kat$ gilt:
 - $\exists instanceOfKatTyp \in InstanceOfKatTyp, katTyp \in KatTyp : \alpha InstanceOfKatTyp(instanceOfKatTyp) = kat \wedge \omega InstanceOfKatTyp(instanceOfKatTyp) = katTyp$ und
 - $\neg \exists instanceOfKatTyp' \in InstanceOfKatTyp, katTyp' \in KatTyp : \alpha InstanceOfKatTyp(instanceOfKatTyp') = i \wedge \omega InstanceOfKatTyp(instanceOfKatTyp') = katTyp', katTyp \neq katTyp'$
11. $\alpha InstanceOfAttTyp : InstanceOfAttTyp \rightarrow Att \wedge \omega InstanceOfAttTyp : InstanceOfAttTyp \rightarrow AttTyp$. Für alle Attribute $att \in Att$ gilt:
 - $\exists instanceOfAttTyp \in InstanceOfAttTyp, attTyp \in AttTyp : \alpha InstanceOfAttTyp(instanceOfAttTyp) = att \wedge \omega InstanceOfAttTyp(instanceOfAttTyp) = attTyp$ und
 - $\neg \exists instanceOfAttTyp' \in InstanceOfAttTyp, attTyp' \in AttTyp : \alpha InstanceOfAttTyp(instanceOfAttTyp') = i \wedge \omega InstanceOfAttTyp(instanceOfAttTyp') = attTyp', attTyp \neq attTyp'$
12. $\alpha RelI : RelI \rightarrow I$ und $\omega RelI : RelI \rightarrow I$. Die Beziehungen zwischen Informationseinheiten werden determiniert durch die Beziehungen zwischen den jeweiligen Konzepten. Für alle $relI \in RelI$ existieren zwei Wege w_1 und w_2 mit $w_1 = (instanceOfKon, relKon)$ und $w_2 = (instanceOfKon')$ für die gilt:
 - (a) $\alpha RelI(relI) = i_1$,

- (b) $\omega RelI(relI) = i_2$,
- (c) $\alpha InstanceOfKon(instanceOfKon) = i_1 \wedge \omega InstanceOfKon(instanceOfKon) = kon_1$,
- (d) $\alpha InstanceOfKon(instanceOfKon') = i_2 \wedge \omega InstanceOfKon(instanceOfKon') = kon_2$ und
- (e) $\exists relKon \in RelKon : \alpha RelKon(relKon) = kon_1 \wedge \omega RelKon(relKon) = kon_2$.

Wenn keine entsprechende *relKon*-Beziehung zwischen den beiden Konzepten kon_1 und kon_2 existiert, wie in Bedingung (e) gefordert, so existiert ein Weg $w'_1 = (isA_1, \dots, isA_n)$, $n \geq 1$ und ein Weg $w'_2 = (isA'_1, \dots, isA'_n)$, $n' \geq 0$ ¹. Für w'_1 und w'_2 gilt:

- $\alpha(w'_1) = kon_1 \wedge \omega(w'_1) = kon'_1$,
- $\alpha(w'_2) = kon_2 \wedge \omega(w'_2) = kon'_2$ und
- $\exists relKon' \in RelKon : \alpha RelKon(relKon') = kon'_1 \wedge \omega RelKon(relKon') = kon'_2$.

Der Graph $G_{RelI} = (I, RelI, \alpha RelI, \omega RelI)$ ist schlingenfrei. D.h., es existiert kein $relI \in RelI$ für das $\alpha RelI(relI) = \omega RelI(relI)$ gilt.

13. $\alpha O : O \rightarrow Kat \wedge \omega O \rightarrow Kat$. Der gerichtete Graph $G''_{Dom} = (KatTyp, Kat, InstanceOfKatTyp, O, \alpha InstanceOfKatTyp, \omega InstanceOfKatTyp, \alpha O, \omega O)$ ist kreis- und schlingenfrei. D.h., es existiert kein $o \in O$ für das $\alpha O(o) = \omega O(o)$ gilt.

Für alle $o \in O$, $kat_1 \in Kat$ und $kat_2 \in Kat$, für die ein Weg $w_o = (o_1, \dots, o_n)$, $n \geq 1$ mit $\alpha(w_o) = kat_1$ und $\omega(w_o) = kat_2$ existiert, gibt es einen Kategorietypen $katTyp \in KatTyp$ mit

- $\alpha InstanceOfKatTyp(instanceOfKatTyp_1) = kat_1 \wedge \omega InstanceOfKatTyp(instanceOfKatTyp_1) = katTyp$,
- $\alpha InstanceOfKatTyp(instanceOfKatTyp_2) = kat_2 \wedge \omega InstanceOfKatTyp(instanceOfKatTyp_2) = katTyp$ und
- $instanceOfKatTyp_1 \neq instanceOfKatTyp_2$.

¹Nach Definition 3.2, muss ein Weg mindestens die Länge 1 haben ($n \geq 1$) haben. Ich habe hier dennoch nur $n' \geq 0$ gefordert, um eine Fallunterscheidung der verschiedenen Fälle $|w'_1| \geq 1 \wedge |w'_2| \geq 1$, $|w'_1| \geq 1$ oder $|w'_2| \geq 1$ zu vermeiden.

Alle Kategorien $kat_i \in Kat, i = 1..n$, die einem Kategorietypp $katTyp$ zugeordnet sind haben eine eindeutige Wurzelkategorie kat_{root} .

Sei Kat_{katTyp} die Menge aller Kategorien, die einem Kategorietypp zugeordnet ist. D.h.,

$$\forall kat \in Kat_{katTyp} : \exists instanceOfKatTyp \in instanceOfKatTyp : \alpha InstanceOfKatTyp(instanceOfKatTyp) = kat \wedge \alpha InstanceOfKatTyp(instanceOfKatTyp) = katTyp.$$

Dann existiert eine eindeutige Wurzelkategorie $kat_{root} \in Kat_{katTyp}$. D.h., für alle $kat \in Kat_{katTyp}, kat \neq kat_{root}$ existiert ein Weg $w = (o_1, o_2, \dots, o_n), n \geq 1$ von der Wurzel kat_{root} zur Kategorie kat . D.h.,

$$\exists kat_{root} \in Kat_{katTyp} : \forall kat \in Kat_{katTyp}, kat \neq kat_{root} : \exists w : \alpha(w) = kat_{root} \wedge \omega(w) = kat$$

14. $\alpha Kla : Kla \rightarrow kat \wedge \omega Kla : Kla \rightarrow I$. Die Beziehungen zwischen Kategorien und Informationseinheiten werden determiniert durch die Beziehungen zwischen den entsprechenden Kategorietyppen und Konzepten. Für alle $kla \in Kla$ existieren zwei Wege w_1 und w_2 mit $w_1 = (instanceOfKatTyp, relKatTyp)$ und $w_2 = (instanceOfKon)$ für die gilt:

- (a) $\alpha Kla(kla) = kat$,
- (b) $\omega Kla(kla) = i$,
- (c) $\alpha InstanceOfKatTyp(instanceOfKatTyp) = kat \wedge \omega InstanceOfKatTyp(instanceOfKatTyp) = katTyp$,
- (d) $\alpha InstanceOfKon(instanceOfKon) = i \wedge \omega InstanceOfKon(instanceOfKon) = kon$ und
- (e) $\exists relKatTyp \in RelKatTyp : \alpha RelKatTyp(relKatTyp) = katTyp \wedge \omega RelKatTyp(relKatTyp) = kon$.

Wenn Bedingung (e) nicht erfüllt werden kann, existiert ein Weg $w'_2 = (isA_1, \dots, isA_n), n \geq 1$ für den gilt:

- $\alpha(w'_2) = kon$,
- $\omega(w'_2) = kon'$ und
- $\exists relKatTyp' \in RelKatTyp : \alpha RelKatTyp(relKatTyp') = katTyp \wedge \omega RelKatTyp(relKatTyp') = kon'$.

Jede Informationseinheit $i \in I$ muss für jeden mit dem Konzept kon assoziierten Kategorietypp $katTyp$ mit mindestens einer Kategorie kat , die dem Kategorietypp $katTyp$ zugeordnet ist, klassifiziert werden. D.h., für alle $i \in I$ existiert ein $kla \in Kla$ mit:

- $\alpha KLa(kla) = kat$ und
- $\omega KLa(kla) = i$

Um zu verhindern, dass eine Informationseinheit mit einer Ober- und einer Unterkategorie gleichzeitig klassifiziert werden kann, gilt zusätzlich: Wenn ein Weg $w_1 = (kla_1)$ mit $\alpha KLa(kla_1) = kat_1 \wedge \omega KLa(kla_1) = i$ und ein Weg $w_2 = (kla_2)$ mit $\alpha KLa(kla_2) = kat_2 \wedge \omega KLa(kla_2) = i$, dann existiert kein Weg $w_o = (o_1, \dots, o_n), n \geq 1$ mit $\alpha O(w_o) = kat_1 \wedge \omega O(w_o) = kat_2$.

15. $\alpha RelAtt : RelAtt \rightarrow I \wedge \omega RelAtt : RelAtt \rightarrow Att$. Die Beziehungen zwischen Attributen und Informationseinheiten werden durch die Beziehungen der zugeordneten Konzepte und Attributtypen determiniert. Für alle $relAtt \in relAtt$ existieren zwei Wege $w_1 = (instanceOfKon, relAttTyp)$ und $w_2 = (instanceOfAttTyp)$ für die gilt:

- (a) $\alpha RelAtt(relAtt) = i$,
- (b) $\omega RelAtt(relAtt) = att$,
- (c) $\alpha InstanceOfKon(instanceOfKon) = i \wedge \omega InstanceOfKon(instanceOfKon) = kon$,
- (d) $\alpha InstanceOfAttTyp(instanceOfAttTyp) = att \wedge \omega InstanceOfAttTyp(instanceOfAttTyp) = attTyp$ und
- (e) $\exists relAttTyp \in RelAttTyp : \alpha RelAttTyp(relAttTyp) = kon \wedge \omega RelAttTyp(relAttTyp) = attTyp$

Sollte Bedingung (e) nicht eingelöst werden können, so existiert ein Weg $w'_1 = (isA_1, \dots, isA_n), n \geq 1$, für den gilt:

- $\alpha(w'_1) = kon \wedge \omega(w'_1) = kon'$ und
- $\exists relAttTyp' \in RelAttTyp : \alpha RelAttTyp(relAttTyp') = kon' \wedge \omega RelAttTyp(relAttTyp') = attTyp$.

Im IB-Domänenmodell wird die Einteilung in Struktur- und Instanzenebene aufgegriffen, wie sie von ontologiebasierten Informationssystemen benutzt wird. Dies wird in Definition 6.1 durch die Pfeilmengen *InstanceOfKon*, *InstanceOfKatTyp* und *InstanceOfAttTyp* und die dazu gehörigen Abbildungen (vgl. Bedingung 9 - 11 in Def. 6.1) definiert. Die Strukturebene des IB-Domänenmodells besteht aus Konzepten, Kategorietypen und Attributtypen. Diese werden in Definition 6.1 durch die Eckenmengen *Kon*, *KatTyp* und *AttTyp* beschrieben. Die Instanzenebene des IB-Domänenmodells besteht dementsprechend

- aus Informationseinheiten (Eckenmenge I), die durch Konzepte typisiert werden,
- aus Kategorien (Eckenmenge Kat), die jeweils eindeutig einem Kategorie-typen zugeordnet werden und
- aus Attributwerten (Eckenmenge Att).

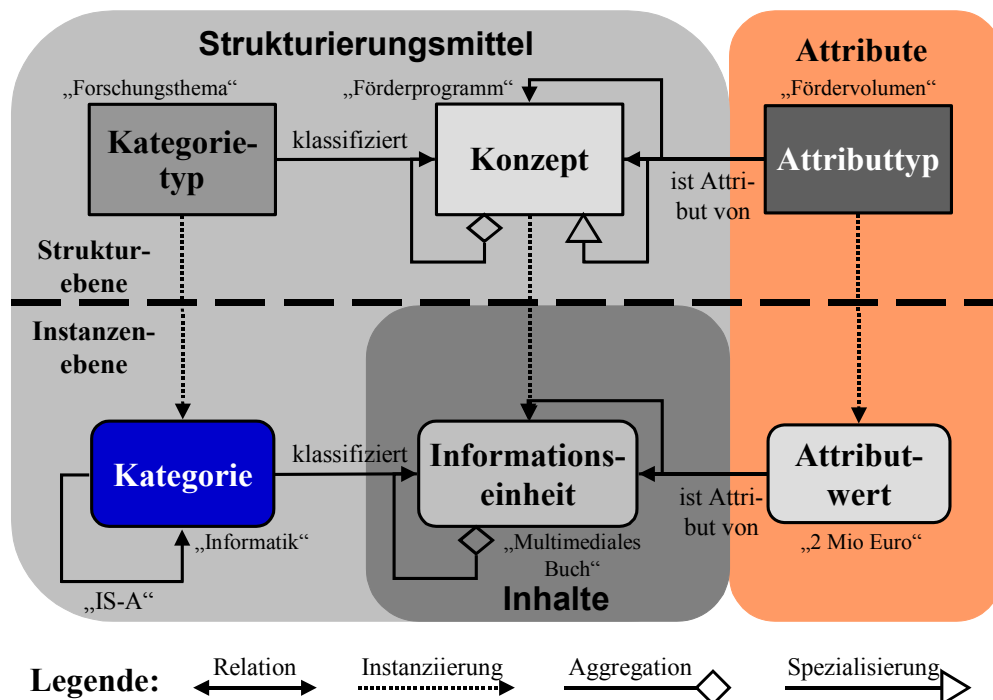


Abbildung 6.1: Basisstruktur für das IB-Domänenmodell

In ontologiebasierten Informationssystemen besteht die Strukturebene nur aus Konzepten. Die Konzepte im IB-Domänenmodell können wie in ontologiebasierten Informationssystemen Relationen zu anderen Konzepten ausbilden. Dies wird in der Definition durch die Pfeilmenge $RelKon$ und die Abbildungen α_{RelKon} und ω_{RelKon} bewerkstelligt. Bei diesen Relationen kann es sich auch um Aggregationen handeln. Die Konzepte können mit IsA-Relationen zu einer Konzepthierarchie arrangiert werden. Dabei werden die Relationen über die Hierarchie hinweg vererbt. Die IsA-Relationen werden in Definition 6.1 durch die Pfeilmenge IsA und ihre Abbildungen $(\alpha_{IsA}, \omega_{IsA})$ definiert. Die Vererbung der Relationen über die Konzepthierarchie wird nicht durch eine eigene Pfeilmenge sichergestellt, sondern hat sich bei der Beschreibung von Relationen in der Instanzenebene niedergeschlagen (vgl. Bedingung 12, 14 und 15). Konzepte im

IB-Domänenmodell werden benutzt, um die Informationstypen und ihre Relationen untereinander zu beschreiben, mit denen der Broker handelt. In der Domäne Forschungsförderung werden z.B. die Konzepte „Förderprogramm“, „Förderer“, „Projekträger“ und „Ansprechpartner“ benutzt.

Konkrete Informationseinheiten werden wie in ontologiebasierten Informationssystemen eindeutig einem Konzept zugeordnet. Alle Relationen des Konzeptes können von konkreten Informationseinheiten mit Leben gefüllt werden. D.h., wenn ein Konzept „Förderprogramm“ eine Relation zu einem anderen Konzept „Förderer“ aufweist, kann eine Informationseinheit „Multimediales Buch“, die dem Konzept „Förderprogramm“ zugeordnet wurde, eine Relation zu einer Informationseinheit „BMBF“ haben. Dabei wurde „BMBF“ dem Konzept „Förderer“ zugeordnet. In Definition 6.1 wird die Relation zwischen Informationseinheiten durch die Pfeilmenge $RelI$ beschrieben und ausgestaltet durch die Abbildungen $\alpha RelI$ und $\omega RelI$ (vgl. Bedingung 12).

Die Abbildung von Interessen geschieht wie in adaptiven, domänenmodellbasierten Informationssystemen über Kategoriehierarchien. Die Organisation der Kategorien zu einer Hierarchie in Definition 6.1 wird durch die Pfeilmenge O und ihre Abbildungen αO und ωO erreicht. Die Kategoriehierarchien im IB-Domänenmodell werden benutzt, um typische Eigenschaften, die für den Benutzer möglicherweise entscheidungsrelevant sind, zu beschreiben. Die Kategorien einer Kategoriehierarchie bilden dabei alle möglichen Werte für diese Eigenschaft ab. Dies folgt der Theorie der multi-attributiven Objektbewertung (vgl. [vWE86], [Sch01]), die davon ausgeht, dass die Auswahl aus einer Menge von Objekten durch die gewichtete Summe der Bewertung der verschiedenen Eigenschaften der Objekte geschieht. Das Interesse des Benutzers kann durch die Bewertung der einzelnen Eigenschaften errechnet werden. Da die Eigenschaften im IB-Domänenmodell jeweils durch Kategoriehierarchien beschrieben werden, kann das Interesse jeweils für eine Eigenschaft in ähnlicher Weise ermittelt werden, wie dies in den meisten adaptiven, domänenmodellbasierten Informationssystemen (vgl. Kap. 4.5, bzw. Tab. 4.1) bewerkstelligt wird. Die Relevanz von einzelnen Informationseinheiten kann dann als die gewichtete Summe aller Eigenschaften berechnet werden. Die Beschreibung aller Eigenschaften in einer einzigen Kategoriehierarchie ist nicht sinnvoll, da dies zum Facettierungsproblem (vgl. Kap. 2.3.2) führt und die Kategoriehierarchie aufblähen würde, um alle möglichen Kombinationen von Eigenschaften abbilden zu können.

Die Verbindung zwischen Kategoriehierarchien und Konzepten geschieht auf der Struktur-Ebene und wird in Definition 6.1 durch die Pfeilmenge $RelKatTyp$ und die Abbildungen $\alpha RelKatTyp$ und $\omega RelKatTyp$ definiert. Konzepte können durch Kategorietypen klassifiziert werden. Kategorietypen beschreiben die Eigenschaft, die jeweils durch eine Kategoriehierarchie beschrieben werden soll. Jede Kategorie ist eindeutig einem Kategorietyp zugeordnet. Die Kategorien kön-

nen zu einer Kategorienhierarchie organisiert werden, da sie jeweils Ober- und Unterkategorien aufweisen können. Jede Kategorienhierarchie sollte eine eindeutige Wurzelkategorie aufweisen. Beispiele für Kategorietypen aus der Domäne Forschungsförderung sind „Forschungsthemen“ und „Förderarten“. Beispiele für Kategorien sind „Informatik“ und „Stipendium“.

Die Verbindung zwischen Konzepten und Kategorietypen in der Strukturebene determiniert die Relation zwischen Kategorien und Informationseinheiten auf der Instanzebene: Für jeden Kategorietypp des zugeordneten Konzeptes **muss** eine konkrete Informationseinheit mit mindestens einer Kategorie des entsprechenden Kategorietypps klassifiziert werden. Ist eine Zuordnung (aus inhaltlichen Gründen) nicht möglich, sollte die Informationseinheit mit der jeweiligen Wurzelkategorie klassifiziert werden. Die Klassifikation ist zwingend, damit eine durchgängige Filterung auf Basis der Klassifikationen möglich wird. Die Klassifikation von Informationseinheiten und der Zusammenhang zur Beziehung zwischen den entsprechenden Konzepten und Kategorietypen (*RelKatTyp*) wird in Definition 6.1 in der Bedingung 14 mit den Abbildungen α_{Kla} und ω_{Kla} und der Pfeilmenge *Kla* beschrieben.

Die Beziehungen zwischen Ober- und Unterkategorien können als Relationen mit IsA-Charakter aufgefasst werden. D.h., alle Informationseinheiten, die mit einer Kategorie klassifiziert wurden, sind auch Informationseinheiten, die mit einer Oberkategorie klassifiziert werden könnten. Es handelt sich jedoch nicht um eine „echte“ IsA-Relation, da keine Eigenschaften zwischen den Kategorien vererbt werden. Quantitative Attribute können nicht direkt als Kategorien beschrieben werden, da Kategorien qualitative Eigenschaften beschreiben und diese Eigenschaften quantitativer Natur sind. Sind solche Eigenschaften von großer Wichtigkeit, so können sie indirekt als Kategorien modelliert werden, indem man verschiedene Wertebereiche zu einzelnen Kategorien gruppiert (z.B. Lebensalter: Säugling, 0-1 Jahr; Baby, 1-3 Jahre; Kind 3-10 Jahre;...). Diese künstlichen Kategorien müssen nicht notwendigerweise eine tiefe hierarchische Struktur aufweisen, können aber in ähnlicher Weise zu Gruppierung und Filterung von Informationseinheiten eingesetzt werden.

Weitere Eigenschaften, von denen man annimmt, dass sie nicht für die Auswahl des Benutzers von Bedeutung sind, aber trotzdem einen gewissen Mehrwert für ihn darstellen, werden als Attributtypen und Attributwerte modelliert. In Definition 6.1 ist dies durch die Eckenmengen *AttTyp* und *Att* reflektiert. Attributtypen werden auf der Strukturebene definiert. Sie werden in Relation zu Konzepten definiert (Pfeilmenge *RelAttTyp*). Konkrete Informationseinheiten können dann Relationen zu Attributwerten des jeweiligen Attributtyps etablieren. Dies wird in Definition 6.1 durch die Pfeilmenge *RelAtt* und ihre Abbildungen definiert. Die Unterscheidung von Kategorien und Attributen in entscheidungsrelevante und einfache Eigenschaften folgt der Beobachtung von Rosch (vgl. [Ros78]), dass typi-

sche Eigenschaften schneller erkannt werden als untypische Eigenschaften.

Abbildung 6.1 zeigt die Basisstruktur des IB-Domänenmodells. Als Notation wurde eine UML-nahe Notation (vgl. [FS00]) genutzt. Die durchgezogenen Pfeile stehen für Relationen zwischen den verschiedenen Bausteinen. Die gestrichelten Pfeile stehen für Instanziierungsbeziehungen. Zur Verdeutlichung wurde für jedes Modellierungsmittel jeweils eine Ausprägung für dieses Modellierungsmittel in die Abbildung integriert. Diese Ausprägungen werden im nächsten Abschnitt, in dem ein Beispieldomäne beschrieben wird, wieder aufgenommen.

Im Zentrum des IB-Domänenmodells stehen *Informationseinheiten*. Das sind die Informationen, mit denen der Broker handelt. Informationseinheiten sind eindeutig einem *Konzept* zugeordnet. Konzepte können mit anderen Konzepten in Relation stehen. Neben einfachen Assoziationen sind auch PartOf-Relationen zwischen Konzepten möglich². Die Relationen zwischen Konzepten werden von den Informationseinheiten (der jeweiligen) Konzepte mit Leben gefüllt. D.h., die möglichen Relationen zwischen Informationseinheiten werden durch die Relationen zwischen ihren Konzepten determiniert. Konzepte können zu einer Konzepthierarchie arrangiert werden, indem man für ein Konzept ein Unterkonzept definiert. Das Unterkonzept erbt alle Relationen seines Oberkonzeptes. Auch Mehrfachvererbung ist in der Konzepthierarchie möglich.

Die Kategorien und Konzepte können als Strukturierungsmittel zur Gruppierung und Filterung von Informationseinheiten eingesetzt werden, da sie eine direkte Verbindung zu Informationseinheiten aufweisen. Kategorietypen werden zur Koordination von Konzepten und Kategorien eingesetzt. Attributwerte könnten ebenfalls zur Gruppierung und Filterung eingesetzt werden. Attributtypen könnten dann wie Kategorietypen zur Koordination mit Konzepten genutzt werden. Es erscheint jedoch sinnvoll, sich zur Filterung und Gruppierung auf Kategorien und Kategorietypen zu konzentrieren, da diese zur Optimierung des Identifikationsprozesses modelliert werden. Würde man Attributwerte und Attributtypen gleichberechtigt zur Filterung und Gruppierung von Informationseinheiten einsetzen, würde man gleichzeitig die Bedeutung der Kategorien schmälern, da diese nur noch aufgrund ihrer hierarchischen Organisation eine besondere Rolle spielen würden, nicht aber aufgrund ihrer Funktion. Damit würden Kategorien und Attributwerte gleichberechtigt und die Modellierung von Typikalisierungseffekten ad absurdum geführt. Die (einfachen) Attribute sollten deswegen nur mittelbar zur Filterung eingesetzt werden, wenn die Filterung durch Kategorien und Konzepte nicht ausreicht. Um auch die Filterung durch quantitative Eigenschaften zu ermöglichen, kann dabei auf die Sortierung als eine weiche Form der Filterung zurückgegriffen werden. Wie schon in [Bru96] bzw. [Bru01] beschrieben, kann Sortierung zur

²Auf eine formale Beschreibung von PartOf-Relationen wurde verzichtet, da sich alle PartOf-Relationen auch als einfache Relationen modellieren lassen.

Filterung benutzt werden, da Benutzer sich oft auf die ersten Einträge einer Liste konzentrieren. Während Brusilovsky auf die Beschreibung des adaptiven Sortierens zu Filterungszwecken beschränkt, kann dieses weiche Filterverfahren auch allgemein angewendet werden. Konsequenterweise sind deswegen die Attributtypen noch einmal danach zu unterscheiden, wie ihre Attributwerte sortiert werden können. So ist zum Beispiel zwischen Attributen mit numerischen und alphanumerischen Wertebereichen zu unterscheiden.

6.3 Beispieldomäne

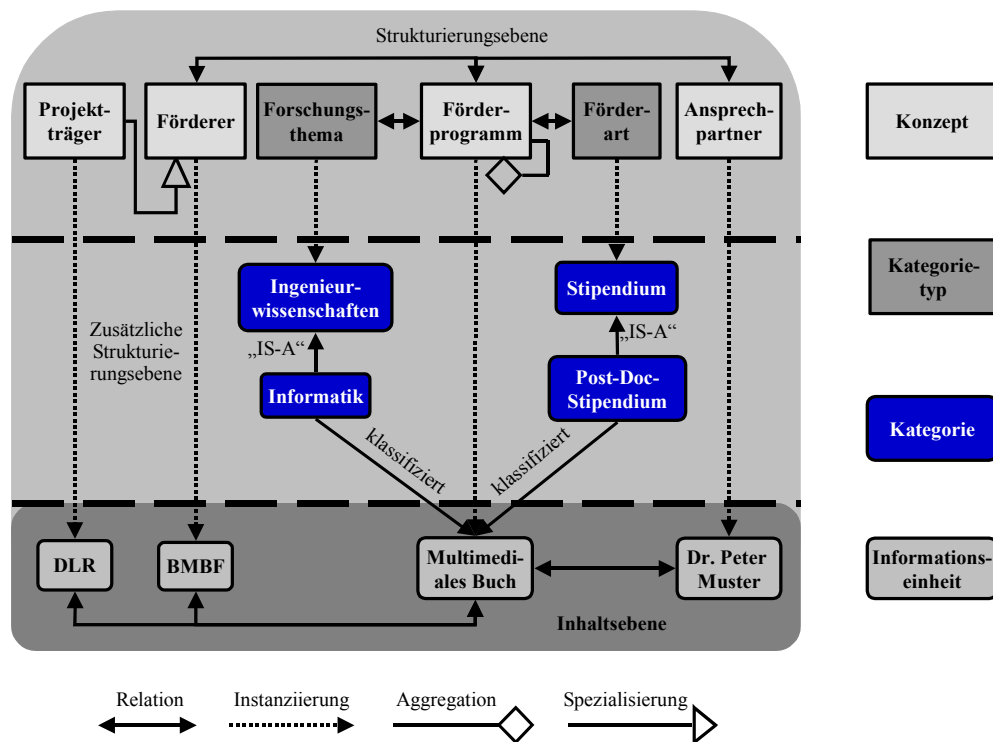


Abbildung 6.2: Ausschnitt aus dem ELFI-Domänenmodell

Information-Broker informieren typischerweise über eine Gruppe von verwandten Informationstypen. Dies wird im IB-Domänenmodell durch Konzepte modelliert. Abb. 6.2 zeigt einen Ausschnitt aus einem IB-Domänenmodell für Forschungsförderung. Dort sind beispielsweise neben Förderprogrammen auch deren Förderer und Ansprechpartner von Interesse. Deswegen wurden diese jeweils als Konzepte modelliert. Zusätzlich gibt es Projektträger, die die fachliche Betreuung von Förderprogrammen eines Förderers überwachen und praktisch aus-

gestalten. Projektträger können als spezielle Förderer angesehen werden. Im IB-Domänenmodell ist dies dadurch berücksichtigt, dass das Projektträgerkonzept vom Konzept Förderprogramm abgeleitet wurde. Förderprogramme haben einen Förderer (ggf. auch einen Projektträger) und einen Ansprechpartner. Förderprogramme können unter dem Aspekt der Forschungsgebiete, die gefördert werden und unter dem Aspekt der Förderart kategorisiert werden³. Die Kategorietypen in der Strukturierungsebene werden von Kategorien mit Leben gefüllt. In Abb. 6.2 sind lediglich jeweils zwei Kategorien für jeden Kategorietyp aufgeführt, um ihre hierarchische Organisation anzudeuten. Die Kategorien bilden zusammen eine zusätzliche Strukturierungsebene, die zur zusätzlichen Filterung von Informationseinheiten genutzt werden können. Darauf wird in Kapitel 7 in größerer Tiefe eingegangen. Die Kategorien werden zur Klassifikation von Informationseinheiten genutzt⁴. So ist z.B. das Förderprogramm „Multimediales Buch“ ein Post-Doc-Stipendium, das für das Forschungsgebiet Informatik ausgeschrieben wurde. An diesem Beispiel lässt sich die mittelbare Subsumptionsrelation zwischen Ober- und Unterkategorien illustrieren: Das Förderprogramm „Multimediales Buch“ ist ein Post-Doc-Stipendium, da es in der Förderarten-Dimension mit der Kategorie „Post-Doc-Stipendium“ klassifiziert wurde. Gleichzeitig handelt es sich auch um ein Stipendium, da „Post-Doc-Stipendium“ eine Unterkategorie von „Stipendium“ ist. Die Beziehungen zwischen Konzepten werden auf ihre jeweiligen Informationseinheiten übertragen. Im Beispiel wird das Förderprogramm „Multimediales Buch“ vom Förderer „BMBF“ gefördert, die Ausgestaltung übernimmt der Projektträger „DLR“ und der Ansprechpartner für das Programm ist „Dr. Peter Muster“.

6.4 Vergleich des IB-Domänenmodells mit OIL

In diesem Abschnitt wird das IB-Domänenmodell gegen andere Ontologiesprachen abgegrenzt, die auch im Semantic-Web-Projekt (vgl. [BLHL01]) benutzt werden. Ich verzichte auf eine Abgrenzung zu reinen prädikatenlogischen Ansätzen, da diese auch in anderen ontologiebasierten Informationssystemen keine große Rolle spielen (vgl. Kap. 5.3). Das IB-Domänenmodell ist eine Ontologie im Sinne von [Gru93], da es sich um ein gemeinsames, geteiltes Verständ-

³Eine genauere Beschreibung der Domäne Forschungsförderung findet sich am Ende dieser Arbeit in der ELFI-Fallstudie in Kap. 9.1. Hier wird nur ein Ausschnitt des ELFI-Domänenmodells beschrieben, um die Verwendung der verschiedenen Bausteine des IB-Domänenmodells zu illustrieren. Auf Attributtypen und Attributwerte (von Konzepten und Informationseinheiten) wird in der Darstellung vollkommen verzichtet, da diese im Prinzip ähnlich zu Beziehungen zwischen Konzepten sind.

⁴Es ist anzumerken, dass alle Kategorien entlang einer Kategorienhierarchie zur Klassifikation genutzt werden können.

nis von der Domäne handelt, in der Information-Brokering stattfindet. Das IB-Domänenmodell wird zwar in der Regel vom Broker verwaltet, aber der Broker kann nur erfolgreich sein, wenn er im Sinne der anderen Beteiligten (Konsumenten und Anbieter) handelt, da er von ihnen abhängig ist. Insofern kann man davon ausgehen, dass sie sein Modell von der Domäne teilen. Bei dem IB-Domänenmodell handelt es sich auch um ein formales Modell, da es auf Basis von Definition 6.1 über eine klare Semantik verfügt.

Um das IB-Domänenmodell gegen Beschreibungslogik, frame-basierte Logik und die Bestrebungen im Semantic-Web-Projekt abzugrenzen, grenze ich es gegen OIL (**O**ntology **I**nterchange **L**anguage, vgl. [FHH⁺00], [FvHH⁺01]) ab, da es sich bei OIL sowohl um einen frame-basierten, als auch um einen beschreibungslogischen Ansatz handelt (vgl. [Fen01]), der zudem breiten Niederschlag im Semantic-Web-Projekt gefunden hat. Außerdem lässt sich OIL mit wenig Aufwand in andere Ontologiesprachen wie DAML (**D**APRA **A**gent **M**arkup **L**anguage, vgl. [Hen01a]) übersetzen, wie die Entwicklung der gemeinsamen Sprache DAML+OIL (vgl. [Hor01] und [Hor02]) zeigt, für die auch schon Ontologieeditoren existieren bzw. bestehende wie z.B. Protégé erweitert wurden (vgl. [NSD⁺01]). OIL ist damit prototypisch für eine große Menge von Ontologiesprachen und die Notwendigkeit zur Abgrenzung gegen weitere Ontologiesprachen entfällt weitestgehend.

OIL besteht im Wesentlichen aus Konzepten. Diese Konzepte bilden eine Konzepthierarchie, die Mehrfachvererbung zulässt. Es werden primitive und definierte Konzepte unterschieden. Konzepte sind primitiv, wenn sie nur ein Vaterkonzept haben. Sie sind definiert, wenn sie mehr als ein Vaterkonzept aufweisen. Zwischen den verschiedenen Konzepten können Relationen definiert werden, die an die Kinderkonzepte vererbt werden. Die Relationen werden als sogenannte „Slots“ definiert. Die Slots werden genutzt, um die Rollen der Relationen für die einzelnen Konzepte zu beschreiben. Die verschiedenen Slots können zu einer Slothierarchie arrangiert werden. Die Slots müssen getypt werden. D.h., die Werte eines Slots müssen entweder Instanzen eines bestimmten Konzeptes sein, oder sie müssen aus einer Reihe von Individuen ausgewählt werden. Individuen sind Konstanten in der Ontologie. Für die verschiedenen Slots kann jeweils eine Kardinalität der Werte bestimmt werden. Neben der All- und Existenzquantifizierung können Mindest- und Maximalkardinalitäten sowie exakte Kardinalitäten festgelegt werden. Für Konzepte, Slots und Individuen können Axiome von allgemeiner Gültigkeit aufgestellt werden. Durch die Möglichkeit zur Bestimmung der Kardinalität unterscheidet sich OIL vom vorgestellten IB-Domänenmodell, da dort alle Beziehungen zwischen Konzepten bzw. zwischen Informationseinheiten, die durch die Beziehungen zwischen Konzepten determiniert werden, fakultativer Natur sind⁵,

⁵Dies schlägt sich im OIL-Beispiel in Abb. 6.3 dadurch nieder, dass Beziehungen zwischen

während die Beziehungen zwischen Informationseinheiten und Kategorien, die durch die Beziehungen zwischen Konzepten und Kategorien determiniert werden, verpflichtend sind⁶. Eine engere Beschreibung der Kardinalität von Beziehungen ist im IB-Domänenmodell zur Zeit nicht möglich.

Vergleicht man die Konzepte im IB-Domänenmodell mit den Konzepten in OIL, so stellt man fest, dass sie prinzipiell ähnlicher Natur sind. Sie können zu einer Konzepthierarchie arrangiert werden. Es können Relationen zwischen Konzepten definiert werden. Die Relationen zwischen Konzepten werden auf Kinderkonzepten vererbt. PartOf-Relationen werden im IB-Domänenmodell explizit als solche gekennzeichnet, während die entsprechenden Slots in OIL mit dem jeweils inversen Slot gekennzeichnet werden können. Auch wenn die Konzepte im IB-Domänenmodell und in OIL im Prinzip ähnliche Eigenschaften haben, so ist die Bedeutung von Konzepten im IB-Domänenmodell doch eine andere. Während die Konzepte in OIL nur als Grundbausteine für die Ontologie fungieren, markieren die Konzepte im IB-Domänenmodell diejenigen Informationstypen, die von Interesse sind und mit denen gehandelt wird.

Der Unterschied zwischen dem IB-Domänenmodell und OIL zeigt sich auch darin, dass OIL weder explizite Attribute noch Kategorien kennt. Gleichwohl lassen sich diese (Attribute und Kategorien) in OIL modellieren. Attribute lassen sich in OIL relativ leicht abbilden, indem man für jeden Attributtyp einen eigenen Slot definiert und für die verschiedenen Wertebereiche jeweils ein eigenes Konzept anlegt. Handelt es sich bei dem Wertebereich um eine abzählbare Menge, z.B. eine Ordinalskala, so können für die einzelnen Werte Individuen eingeführt werden.

Schwieriger gestaltet sich die Modellierung von Kategorien in OIL, da es hier mehrere Möglichkeiten gibt und das den Kategorien implizite Merkmal der Hierarchie inhärent mit der Konzepthierarchie verbunden ist. Will man die hierarchische Anordnung von Kategorien in OIL übertragen, so modelliert man die verschiedenen Kategoriehierarchien des IB-Domänenmodells als Konzepthierarchien. Ähnliches schlägt auch [Pri99] zur Umsetzung eines facettierten Thesaurus in Beschreibungslogik vor. Die Zuordnung zu den ursprünglichen Konzepten (im Sinne des IB-Domänenmodells) geschieht über einen Slot, der den jeweiligen Kategorietyphen ersetzt. Um eine durchgängige Kategorisierung zu sichern, wird für den Slot eine minimale Kardinalität gefordert. Da die Kategorien (im IB-Domänenmodell) zur Filterung und Gruppierung genutzt werden, weisen sie üblicherweise keine weiteren Relationen und Attribute auf. Das bedeutet, dass die Kategorien, die in OIL zu Konzepten wurden, außer dem Slot zum ursprünglichen⁷

Konzepten durch `has-value` gekennzeichnet werden.

⁶Im Beispiel in Abb. 6.3 werden die entsprechenden Slots durch `value-type` gekennzeichnet.

⁷Ich bezeichne Konzepte (in der OIL-Ontologie) als *ursprünglich*, wenn sie Konzepten im IB-Domänenmodell entsprechen. Dagegen bezeichne ich Konzepte als *neu*, wenn sie erst bei der

Konzept keine weiteren Slots aufweisen. Innerhalb einer Kategorienhierarchie haben alle neuen Konzepte denselben Slot. Auf seiten der ehemaligen Kategorie ist der Slot existenzquantifiziert, da nicht jede Kategorie notwendigerweise mit einer Instanz eines Konzeptes (einer Informationseinheit im IB-Domänenmodell) assoziiert sein muss. Für jedes (neue) Konzept sollte genau eine Instanz angelegt werden.

Abbildung 6.3 zeigt die Definition der in Kapitel 6.3 und Abbildung 6.2 beschriebenen Beispieldomäne als OIL-Ontologie. Diese Ontologie wurde mit dem OilEd-Editor erstellt und folgt der oben beschriebenen Vorgehensweise, um ein IB-Domänenmodell in eine OIL-Ontologie umzuwandeln.

```
begin-ontology
ontology-container
title ELFI Ontology
creator Achim Nick
identifizier "id"
language "OIL"

ontology-definitions

class-def Förderart
slot-constraint förderprogramme has-value Förderprogramm
// SUB: Stipendium

class-def Förderprogramm
slot-constraint forschungsthemen value-type Forschungsthema
slot-constraint förderarten value-type Förderart
slot-constraint ansprechpartner has-value Ansprechpartner
slot-constraint förderer has-value Förderer
slot-constraint projektträger has-value Projektträger

class-def Stipendium subclass-of
Förderart
// Förderart
// SUB: Post-Doc-Stipendium

class-def Förderer
slot-constraint förderprogramme has-value Förderprogramm
// SUB: Projektträger
```

Übersetzung eingeführt wurden.

```
class-def Post-Doc-Stipendium subclass-of
Stipendium
// Stipendium

class-def Projektträger subclass-of
Förderer
// Förderer

class-def Ansprechpartner
slot-constraint förderprogramme has-value Förderprogramm

class-def Ingenieurwissenschaften subclass-of
Forschungsthema
// Forschungsthema
// SUB: Informatik

class-def Forschungsthema
slot-constraint förderprogramme has-value Förderprogramm
// SUB: Ingenieurwissenschaften

class-def Informatik subclass-of
Ingenieurwissenschaften
// Ingenieurwissenschaften

slot-def ansprechpartner
domain Ansprechpartner

slot-def forschungsthemen
domain Forschungsthema

slot-def förderer
domain Förderer

slot-def projektträger
domain Projektträger

slot-def förderarten
domain Förderart

slot-def förderprogramme
domain Förderprogramm
```

end-ontology

Abbildung 6.3: ELFI-Domänenmodell in OIL

Will man verhindern, dass für jedes (neue) Konzept in OIL eine Instanz angelegt werden muss, so modelliert man für jeden Kategorietypen ein Konzept in OIL, welches neben dem Slot für das ursprüngliche Konzept einen Slot aufweist, der jeweils mit Instanzen des eigenen Konzeptes gefüllt werden muss. Zusätzlich formuliert man ein Axiom, das sicherstellt, dass dadurch keine Kreise in dieser Hierarchie entstehen können. Außerdem benötigt man ein weiteres Axiom, welches garantiert, dass die (ursprünglichen) Konzepte nicht mit einer Kategorie und einer Kinderkategorie derselben Kategorie gleichzeitig assoziiert werden können. Diese Modellierungsmöglichkeit hat den Nachteil, dass die im IB-Domänenmodell explizit beschriebene Kategorienhierarchie durch eine implizite Struktur ersetzt wird und die Hierarchie dadurch nicht mehr direkt als solche wahrgenommen werden kann.

Die dritte Möglichkeit der Übersetzung der Kategorienhierarchie aus dem IB-Domänenmodell in OIL ist die Betonung der attributiven Eigenschaft der Kategorien. Sie beschreiben Aspekte und damit Eigenschaften von Konzepten. In OIL ersetzt man die Kategorien durch Individuen. Die explizite hierarchische Eigenschaft der Kategorien geht dabei in OIL verloren. Man könnte jedoch den Pfad durch die Kategorienhierarchie im Namen der Individuen widerspiegeln und eine Reihe von Axiomen definieren, die sicherstellen, dass die Auswirkungen auf die Kategorisierung der (ursprünglichen) Konzepte dieselben bleiben, indem man die Axiome so definiert, dass eine gleichzeitige Auswahl von Kategorien und deren Kinderkategorien unmöglich bleibt.

Bei dem IB-Domänenmodell handelt es sich nach der Einteilung von Studer et al. (vgl. Kap. 2.1.3, [SBF98]) um eine Applikationsontologie, die mit Zusatzwissen für die Unterstützung der Profilbildung oder ähnlicher Zwecke des Information-Brokering angereichert wurde. Das IB-Domänenmodell kann damit nicht (in vernünftiger Art und Weise) für die allgemeine Repräsentation von Wissen benutzt werden, weil das IB-Domänenmodell für die Zwecke des Information-Brokering und insbesondere die Profilbildung aufbereitet wurde. Umgekehrt macht OIL wenige bis keine Annahmen über die Nutzung der Ontologie und ist deswegen universeller einsetzbar, wenngleich es Information-Brokering nicht in dem Maße unterstützt, wie es durch das IB-Domänenmodell möglich ist.

Kapitel 7

Profilbildung

Die Profilbildung ist ein wichtiger Bestandteil im kundenorientierten Prozess (vgl. Kap. 3.1). Es geht bei der Profilbildung darum, die Interessen des Benutzers (des Informationskonsumenten) in einem Interessenprofil auf das IB-Domänenmodell (des Brokers) abzubilden. Die Profilbildung sollte vom Brokeringsystem und dem Benutzer gemeinsam durchgeführt werden. Dabei ist darauf zu achten, dass ein ausgewogenes Verhältnis von Adaptivität und Adaptierbarkeit gewahrt bleibt (vgl. Kap. 2.2.1 und [Fis01]). Dies ist insbesondere deshalb von großer Bedeutung, da das Interessenprofil zur Filterung eingesetzt wird und falsche Profileinstellungen dazu führen können, dass relevante Informationen verloren gehen. Aus diesem Grund kommt einer adaptiven Systemkomponente eine Assistentenrolle zu: Das System beobachtet den Benutzer und macht ggf. Vorschläge zur Änderung des Profils, führt diese Änderungen aber niemals eigenmächtig, sondern nur mit dem Einverständnis des Benutzers durch. Durch die Beobachtung des Benutzers können Widersprüche zwischen seinen subjektiven und objektiven Informationsbedarfen aufgedeckt werden (vgl. [Kuh95]).

Neben dieser adaptiven Funktionalität spielt die Adaptierbarkeit des Interessenprofils eine wichtige Rolle. Damit das Informationsbedürfnis des Benutzers befriedigt werden kann, müssen seine Interessen in einem Interessenprofil abbildbar sein. Dazu wird ein möglichst reichhaltiges Instrumentarium benötigt. Interessenprofile bestehen aus einer Reihe von verschiedensten Filtern, die miteinander kombinierbar sind, um eine möglichst feingranulare Filterung zu ermöglichen. Ziel der Definition der verschiedenen Filter ist die Filterung von relevanten Informationseinheiten. Aus diesem Grund werden alle Filter, die nicht direkt zur Filterung von Informationseinheiten benutzt werden können, *Hilfsfilter* genannt. Filter, die direkt zur Filterung von Informationseinheiten benutzt werden können, heißen *Zielfilter*.

In diesem Kapitel werden zunächst die auf Basis des IB-Domänenmodells (vgl. Kap. 6) grundsätzlich möglichen Filter (Kap. 7.1 - Kap. 7.4) und deren Kom-

bination (Kap. 7.5) beschrieben. Die möglichen Filter geben den Rahmen vor, der zur Filterdefinition genutzt werden kann. Da der Benutzer die volle Kontrolle über sein Interessenprofil haben soll, spielt die Visualisierung des Profils und seiner Filter eine große Rolle. Ebenso sollten die Filter jederzeit editierbar sein, damit der Benutzer die volle Kontrolle behält. Wegen der Forderung nach voller Benutzerkontrolle geben die möglichen Filter den Rahmen für die Adaptierbarkeit des Systems¹ vor. Die Benutzerschnittstelle für die Definition von Profilen wird in Kapitel 8.7 als Bestandteil von Broker's Lounge beschrieben.

Um dem Benutzer eine möglichst weitgehende Unterstützung durch eine adaptive Assistenzkomponente zu geben, muss diese Komponente ihn bei der Definition der verschiedenen Filter beraten. Dieser adaptive Filtersupport wird in Kapitel 7.6 eingeführt.

Filter lassen sich anhand ihrer Definition grundsätzlich in intensionale und extensionale Filter einteilen. *Intensionale Filter* beschreiben *Eigenschaften*, die Elemente erfüllen müssen, um durch diesen Filter zu „passen“, während *extensionale Filter* durch die *Aufzählung aller Elemente* definiert werden, die durch den Filter „passen“ sollen. Intensionale Filter haben gegenüber extensionalen Filtern den Vorteil, dass sie auch auf bisher noch unbekannte Elemente angewendet werden können.

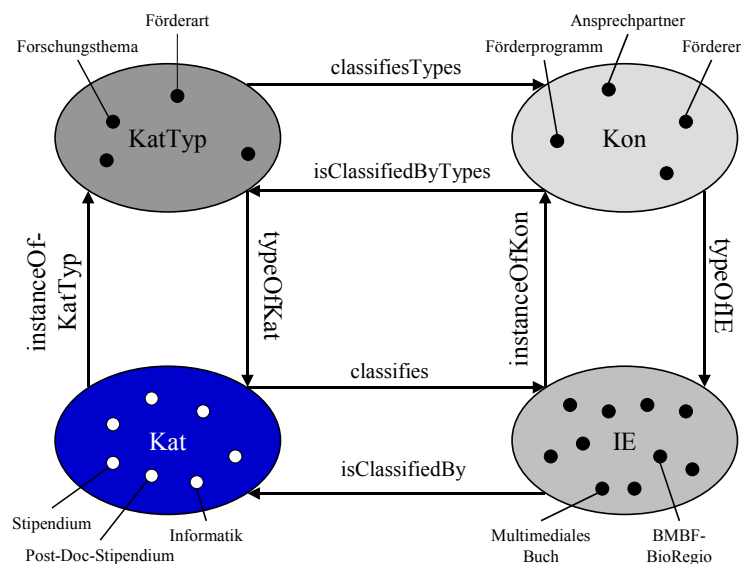


Abbildung 7.1: Mengen im IB-Domänenmodell

Um Informationseinheiten aus dem IB-Domänenmodell zu filtern, bietet es

¹Gemeint ist hier die Adaptierbarkeit des Profils und nicht die Adaptierbarkeit der Oberfläche an bestimmte Präferenzen, wie z.B. Schriftarten usw.

sich an, für die vier Bausteine (Kategorien und Kategorietypen, sowie Informationseinheiten und Konzepte) des IB-Domänenmodells jeweils eigene Filter zu definieren und diese dann zu einem gemeinsamen Domänenmodellfilter zu integrieren. Dabei werden Kategorien, Kategorietypen, Konzepte und Informationseinheiten als Mengen verstanden (siehe Abb. 7.1). Ziel ist es, die Menge der Informationseinheiten zu beeinflussen, die durch die verschiedenen Filter passen. Die Menge der Filterelemente wird *Filtermenge* genannt, während die Menge, auf die ein Filter angewendet wird, *Anwendungsmenge* heißt. Die verschiedenen Filter werden größtenteils extensional formuliert², jedoch intensional benutzt. D.h., die Filtermenge, durch die der Filter extensional definiert wurde, wird auf Elemente eines anderen Typs angewendet. Dabei werden Beziehungen zwischen den Elementen der Filtermenge und Elementen der Anwendungsmenge zur Filterung ausgenutzt. In den nächsten Abschnitten werde ich dies anhand von Beispielen erläutern. Ich benutze dabei eine Notation für Filter, bei der die Filter- und die Anwendungsmenge jeweils angegeben werden. Die Filtermenge wird tiefgestellt und die Anwendungsmenge hochgestellt angegeben. Die Menge aller Informationseinheiten wird durch *IE*, die Menge aller Konzepte durch *KON*, die Menge aller Kategorietypen durch *KatTyp* und die Menge aller Kategorien wird durch *Kat* repräsentiert. Die Filtermenge wird jeweils mit einem tiefgestellten f markiert.

$${}^{\text{Anwendungsmenge}}_{\text{Filtermenge}} \text{filter}$$

Wenn die Filtermenge eine Teilmenge der Anwendungsmenge ist oder die Filtermenge der Anwendungsmenge entspricht und der Filter extensional definiert wurde, so handelt es sich um einen *trivialen Filter*. Triviale Filter sind für alle Basisbestandteile³ des IB-Domänenmodells definiert.

7.1 Instanzenfilter

Extensional definierte Instanzenfilter werden durch eine Aufzählung von Informationseinheiten definiert, die durch den Filter passen. Im Gegensatz zu den anderen

²Eine Filtermenge kann nur in solchen Fällen angegeben werden, in denen ein Filter extensional formuliert wird, da ein intensional definierter Filter keine konkreten Elemente, sondern Eigenschaften zur Filterung benutzt. Ein Beispiel für einen intensional definierten Filter ist ein intensionaler Instanzenfilter (vgl. Kap. 7.1).

³Die Basisbestandteile des IB-Domänenmodells wurden in Kapitel 6.2 beschrieben. Die Basisbestandteile sind Konzepte, Kategorietypen, Attributtypen, Informationseinheiten, Kategorien und Attributwerte. Beim Aufbau des Interessenprofils wird auf Attribute (Attributtypen und -werte) verzichtet, da diese durch ihre Modellierung als weniger relevant für den Filterprozess eingestuft wurden.

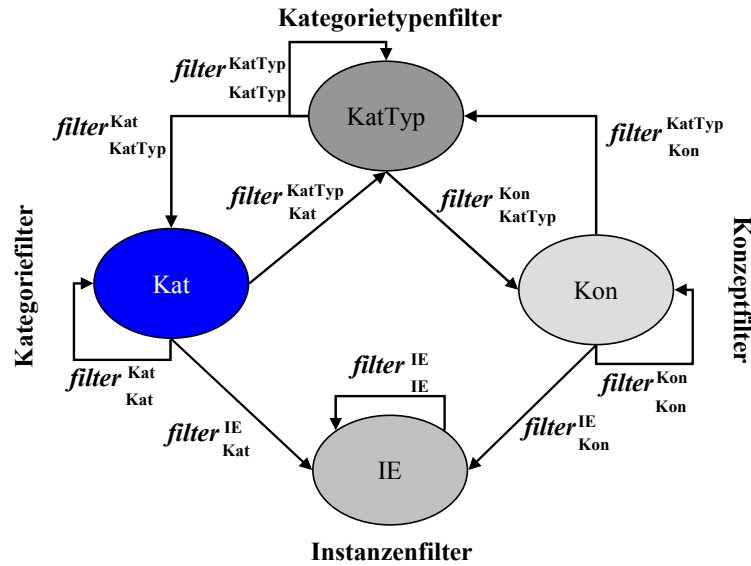


Abbildung 7.2: Domänenmodellfilter

Domänenmodellfiltern können extensional definierte Instanzenfilter nicht intensional genutzt werden, da Filter- und Anwendungsmenge auf derselben Wertebereich definiert werden (siehe Def. 7.1). Ein Beispiel für einen extensionalen Instanzenfilter ist ein Filter, der wichtige Förderprogramme (Informationseinheiten) enthält, für die ein Benutzer einen Förderantrag vorbereitet. Er kann diesen Instanzenfilter benutzen, um über Veränderungen in diesen Förderprogrammen informiert zu werden. Intensionale Instanzenfilter nutzen Eigenschaften von Informationseinheiten aus. Ein Aktualitätsfilter z.B. nutzt den Zeitpunkt der letzten Änderung von Informationseinheiten zur Filterung.

Definition 7.1 (extensionaler, intensionaler Instanzenfilter)

Eine Abbildung $filter_{IE_f}^{IE} : IE_f \times IE \rightarrow IE'$ heißt **extensionaler Instanzenfilter** wenn für alle $ie \in IE'$ gilt: $ie \in IE_f$. Extensionale Instanzenfilter sind triviale Instanzenfilter.

Eine Abbildung $filter_{IE}^{IE} : IE \times IE \rightarrow IE'$ heißt **intensionaler Instanzenfilter**. Instanzenfilter sind Zielfilter, weil sie direkt zur Filterung von Informationseinheiten benutzt werden können.

7.2 Konzeptfilter

Konzeptfilter filtern Informationseinheiten zu einem bestimmten Konzept. Konzeptfilter werden üblicherweise extensional definiert. Sie werden intensional genutzt, da sie zur Filterung von Informationseinheiten eingesetzt werden (siehe Def. 7.2). Dabei wird die Zugehörigkeit von Informationseinheiten zu den jeweiligen Konzepten ausgenutzt. D.h. alle Informationseinheiten passen durch den Filter, deren Konzepte im Konzeptfilter enthalten sind. In der Domäne Forschungsförderung werden z.B. alle Informationseinheiten des Konzepts Förderprogramm gefiltert, da sich ein Benutzer unter Umständen nur für Informationseinheiten dieses Typs (konkrete Förderprogramme) interessiert, während Informationseinheiten eines anderen Typs, z.B. Förderer, ihn nicht interessieren.

Definition 7.2 (Konzeptzielfilter)

Eine Abbildung $filter_{Kon_f}^{IE} : Kon_f \times IE \rightarrow IE'$ heißt **Konzeptzielfilter**, wenn für alle $ie \in IE'$ gilt: $typeOfIE(ie) \in Kon_f$.

Außerdem können Konzepte zur Filterung von Kategorietypen (siehe Def. 7.3) benutzt werden. Dieser Filter kann zwar nicht direkt zur Filterung von Informationseinheiten eingesetzt werden, findet aber in Kombination mit anderen Domänenmodellfiltern (siehe Kap. 7.5) zur Filterung von Informationseinheiten Verwendung.

Definition 7.3 (Konzepthilfsfilter)

Eine Abbildung $filter_{Kon_f}^{KatTyp} : Kon_f \times KatTyp \rightarrow KatTyp'$ heißt **Konzepthilfsfilter**, wenn für alle $katTyp \in KatTyp'$ gilt: $classifiesTypes(katTyp) \cap Kon_f \neq \emptyset$.

Ein trivialer Konzeptfilter (Definition 7.4) kann ebenfalls in Verbindung mit anderen Filtern zur Filterung von Informationseinheiten eingesetzt werden. Dabei werden Konzepte minimiert und damit mittelbar über andere Filter die Menge der Informationseinheiten verringert, die durch den Filter passen.

Definition 7.4 (trivialer Konzeptfilter)

Eine Abbildung $filter_{Kon_f}^{Kon} : Kon_f \times Kon \rightarrow Kon'$ heißt **trivialer Konzeptfilter**, wenn für alle $kon \in Kon'$ gilt, dass $kon \in Kon_f$.

7.3 Kategorietypenfilter

Ein extensionaler Kategorietypenfilter listet die Kategorietypen (des entsprechenden Konzepts) auf. Kategorietypenfilter können zur Filterung von Konzepten oder Kategorien eingesetzt werden. Eine Filterung von Informationseinheiten ist nur in Verbindung mit Konzept- und Kategoriefiltern möglich (siehe Kap. 7.5). Dabei nehmen Kategorietypenfilter die Rolle eines Bindegliedes zwischen Konzept- und Kategoriefiltern wahr.

Definition 7.5 (Kategorietypenfilter für Kategorien)

Eine Abbildung $filter_{KatTyp_f}^{Kat} : KatTyp_f \times Kat \rightarrow Kat'$ heißt **Kategorietypenfilter für Kategorien**, wenn $\forall kat \in Kat' : typeOfKat(kat) \in KatTyp_f$ gilt.

Definition 7.5 formuliert den Kategorietypenfilter für Kategorien. Dabei wird die $typeOfKat$ -Beziehung zwischen Kategorien und Kategorietypen ausgenutzt: Jede Kategorie ist eindeutig einem Kategorietypen zugeordnet. Definition 7.6 definiert den Filter für Konzepte. Er nutzt die Beziehung zwischen Kategorietypen und Konzepten aus.

Definition 7.6 (Kategorietypenfilter für Konzepte)

Eine Abbildung $filter_{KatTyp_f}^{Kon} : KatTyp_f \times Kon \rightarrow Kon'$ heißt **Kategorietypenfilter für Konzepte**, wenn für alle $kon \in Kon'$ gilt: $isClassifiedByTypes(kon) \cap KatTyp_f \neq \emptyset$.

Ein trivialer Kategorietypenfilter (Definition 7.7) wird zur Minimierung von Kategorietypen eingesetzt. Er wird zusammen mit den anderen Kategorietypenfiltern benutzt.

Definition 7.7 (trivialer Kategorietypenfilter)

Eine Abbildung $filter_{KatTyp_f}^{KatTyp} : KatTyp_f \times KatTyp \rightarrow KatTyp'$ heißt **trivialer Kategorietypenfilter**, wenn für alle $katTyp \in KatTyp'$ gilt, dass $katTyp \in KatTyp_f$.

7.4 Kategoriefilter

Kategoriefilter können zur Filterung von Informationseinheiten (siehe Definition 7.8) oder zur Filterung von Kategorietypen (Definition 7.9) benutzt werden. Sie werden üblicherweise durch Aufzählung von Kategorien entlang von Kategoriehierarchien (der entsprechenden Kategorietypen) definiert.

Definition 7.8 (Kategoriezielfilter)

Eine Abbildung $filter_{Kat_f}^{IE} : Kat_f \times IE \rightarrow IE'$ heißt **Kategoriezielfilter**, wenn für alle $ie \in IE'$ gilt:
 $isClassifiedBy(ie) \cap Kat_f \neq \emptyset$.

Kategoriezielfilter filtern Informationseinheiten anhand ihrer Klassifikation für den jeweiligen Kategoriety. Eine Informationseinheit passt durch den Filter, wenn die Informationseinheit mit einer Kategorie aus dem Kategoriefilter klassifiziert wurde. Es ist dabei zu beachten, dass die verschiedenen Kategorien im Filter nicht unabhängig voneinander sind. D.h., wenn eine Kategorie in den Filter aufgenommen wird, so werden automatisch auch alle ihre Oberkategorien und Unterkategorien in den Filter aufgenommen. Wenn ein Kategoriezielfilter z.B. die Kategorien „Mathematik“ und „Maschinenbau“ enthält, gehören auch deren Oberkategorien „Naturwissenschaften“ und „Ingenieurwissenschaften“ und die Unterkategorien von „Mathematik“ und „Maschinenbau“ zum Kategoriefilter. D.h., alle Förderprogramme (Informationseinheiten) die mit „Mathematik“ oder „Maschinenbau“ oder einer der anderen Kategorien klassifiziert wurden, passen durch den Filter.

Definition 7.9 (Kategoriehilfsfilter)

Eine Abbildung $filter_{Kat_f}^{KatTyp} : Kat_f \times KatTyp \rightarrow KatTyp'$ heißt **Kategoriehilfsfilter**, wenn für alle $katTyp \in KatTyp'$ eine Kategorie $kat \in Kat$ existiert, für die $typeOfKat(kat) = katTyp_f$ gilt.

Während der Kategoriezielfilter direkt zur Filterung von Informationseinheiten eingesetzt werden kann, wird der Kategoriehilfsfilter nur mittelbar zur Filterung von Informationseinheiten genutzt. Ebenso kann ein trivialer Kategoriefilter (Definition 7.10) nur in Verbindung mit anderen Filtern zur Filterung von Informationseinheiten eingesetzt werden.

Definition 7.10 (trivialer Kategoriefilter)

Eine Abbildung $filter_{Kat_f}^{Kat} : Kat_f \times Kat \rightarrow Kat'$ heißt **trivialer Kategorie-**

filter, wenn für alle $kat \in Kat'$ gilt:
 $kat \in Kat_f$.

7.5 Zusammenspiel der verschiedenen Filter

Wie man aus Abbildung 7.2 ersehen kann, können Kategorie- ($filter_{Kat_f}^{IE}$) und Konzeptzielfilter ($filter_{Kon_f}^{IE}$) (vgl. Def. 7.8 und Def. 7.2) direkt zur Filterung von Informationseinheiten eingesetzt werden. Die anderen Kategorie- und Konzeptfilter können dagegen nicht direkt zur Filterung von Informationseinheiten benutzt werden. Stattdessen können sie in zusammengesetzten Filtern (vgl. Def. 7.11) benutzt werden, um die Filtermengen von anderen Filtern einzuschränken.

Definition 7.11 (zusammengesetzter Filter)

Ein Filter heißt **zusammengesetzter Filter**, wenn zwei Filter $filter_{A_f}^B$ und $filter_{B_f}^C$ derart miteinander kombiniert werden, dass die Ergebnismenge des ersten Filters der Filtermenge des zweiten Filters entspricht. Es gilt

1. $filter_{A_f}^B : A_f \times B \rightarrow B'$,
2. $filter_{B_f}^C : B_f \times C \rightarrow C'$ und
3. $B' = B_f$.

Zusammengesetzte Filter werden durch die Angabe der Ergebnismenge und der benutzten Filtermengen notiert, indem die Anwendungsmenge hochgestellt und die Filtermengen tiefgestellt dargestellt werden. Der aus den beiden Filtern zusammengesetzte Filter wird wie folgt notiert:

$filter_{A+B}^C$

Um mit den verschiedenen Hilfsfiltern Einfluss auf die Filterung von Informationseinheiten zu bekommen, kann man sie zu einem zusammengesetzten Filter arrangieren, der die Filtermenge von Zielfiltern beschränkt. Die Filtermenge eines Kategoriezielfilters einzuschränken, kann man einen Filter $filter_{Kon+KatTyp}^{Kat}$ aus einem Konzepthilfsfilter und einem Kategorietypenfilter für Kategorien zusammensetzen. Diesen zusammengesetzten Filter kombiniert man nun mit dem Kategoriezielfilter $filter_{Kat_f}^{IE}$ zu dem neuen zusammengesetzten Zielfilter $filter_{Kon+KatTyp+Kat}^{IE}$. Um eine weitere Einflussmöglichkeit auf den zusammengesetzten Filter zu bekommen könnte auch ein trivialer Kategorietypenfilter

benutzt werden, um die Menge der Kategorietypen einzuschränken, die auf den Kategorietypenfilter für Kategorien einwirkt.

Als Beispiel lässt sich ein kombinierter Filter zur Filterung von Stipendien für Mathematiker aufführen: Aus allen Kategorietypen des Konzepts „Förderprogramm“ werden die Kategorietypen „Forschungsthemen“ und „Förderarten“ ausgewählt. In den beiden Kategoriedimensionen werden die Kategorien „Stipendium“ und „Mathematik“⁴ in den Kategoriefilter aufgenommen. dadurch werden alle Förderprogramme gefiltert, die mit „Stipendium“ (in der Förderarten-Dimension) und „Mathematik“ (in der Forschungsgebiete-Dimension) klassifiziert wurden.

Um mit den verschiedenen Hilfsfiltern Einfluss auf einen Konzeptzielfilter zu bekommen, kombiniert man einen Kategoriehilfsfilter mit einem Kategorietypenfilter für Konzepte und ggf. mit verschiedenen trivialen Filtern zum zusammengesetzten Filter $filter_{Kat+KatTyp}^{Kon}$. Diesen zusammengesetzten Filter wendet man auf den Konzeptzielfilter $filter_{Kon}^{IE}$ an und erhält den zusammengesetzten Zielfilter $filter_{Kat+KatTyp+Kon}^{IE}$.

Ein Filter zur Filterung von Stipendien für Mathematiker würde ausgehend von den gewählten Kategorien die beiden Kategorietypen „Forschungsgebiete“ und „Förderarten“ ergeben. Diese als Kategorietypenfilter angewendet, ergeben das Konzept „Förderprogramm“, das wiederum auf die verfügbaren Informationseinheiten angewendet wird.

Filter-Algebra

Die verschiedenen Zielfilter (Konzept-, Kategoriezielfilter und zusammengesetzte Zielfilter) lassen sich beliebig konkatenieren. Bei der Konkatenation zweier Zielfilter wird die Ergebnismenge des ersten Filters als Anwendungsmenge des zweiten Filters benutzt.

Definition 7.12 (Konkatenation von Zielfiltern)

Die Relation $\circ$ zweier Zielfilter $filter_A^{IE}$ und $filter_B^{IE}$ ($filter_A^{IE} \circ filter_B^{IE}$) heißt **Konkatenation**, wenn gilt:

1. $filter_A^{IE_1} : A \times IE_1 \rightarrow IE'$,
2. $filter_A^{IE_2} : A \times IE_2 \rightarrow IE''$ und
3. $IE' = IE_2$ und $IE' \subseteq IE''$.

⁴und damit auch die Kategorie „Naturwissenschaften“ und die Unterkategorien von „Mathematik“

◦ ist kommutativ und assoziativ. D.h., es gilt:

- $filter_A^{IE} \circ filter_B^{IE} = filter_B^{IE} \circ filter_A^{IE}$ und
- $(filter_A^{IE} \circ filter_B^{IE}) \circ filter_C^{IE} = filter_A^{IE} \circ (filter_B^{IE} \circ filter_C^{IE})$.

Die zusammengesetzten Zielfilter werden typischerweise zusammen mit den jeweils komplementären Zielfilterfiltern⁵ konkateniert. So wird zum Beispiel $filter_{Kon+KatTyp+Kat}^{IE}$ mit dem Konzeptzielfilter $filter_{Kon_f}^{IE}$ konkateniert, weil dabei die Filtermenge Kon_f sowohl zur Filterung von Informationseinheiten, als auch zur Filterung von Kategorietypen im zusammengesetzten Filter benutzt werden kann. Um möglichst viele Möglichkeiten zur Filterung zu haben, sollten möglichst viele verschiedene Zielfilter zum Einsatz kommen. Jeder Zielfilter reduziert die Anzahl der Informationseinheiten, die durch den Filter passen. Diese Menge bildet jeweils den Input für die Anwendungsmenge des nächsten Filters. In Abbildung 7.3 wird dies durch die jeweils abnehmende Menge von Informationseinheiten symbolisiert. In Abbildung 7.3 wird die Konkatenation des zusammengesetzten Filters $filter_{Kon+KatTyp+Kat}^{IE}$ mit dem Konzeptzielfilter $filter_{Kon}^{IE}$ und einem abschließenden Instanzenfilter visualisiert. Auf dieselbe Art und Weise können noch weitere Instanzenfilter (z.B. ein Aktualitätsfilter) angehängt werden, um so die Menge der Informationseinheiten weiter zu verkleinern.

7.6 Adaptiver Filtersupport

Die Definition der verschiedenen Filter (Def. 7.1 - 7.4) obliegt dem Benutzer. Es erscheint sinnvoll, ihm durch das System Unterstützung bei der Einstellung der Filter zu geben (vgl. [PN99]), da es möglich ist, dass

- er seine wahren Interessen nicht kennt,
- er sie nicht mit dem vom System vorgegebenen Formalismus ausdrücken kann oder
- sich seine Interessen im Laufe der Zeit geändert haben und dies nicht in den Filtern nachvollzogen wurde.

⁵Der komplementäre Zielfilter des zusammengesetzten Filters $filter_{Kon+KatTyp+Kat}^{IE}$ ist der Konzeptzielfilter $filter_{Kon}^{IE}$. Der komplementäre Zielfilter von $filter_{Kat+KatTyp+Kon}^{IE}$ ist der Kategoriezielfilter $filter_{Kat}^{IE}$.

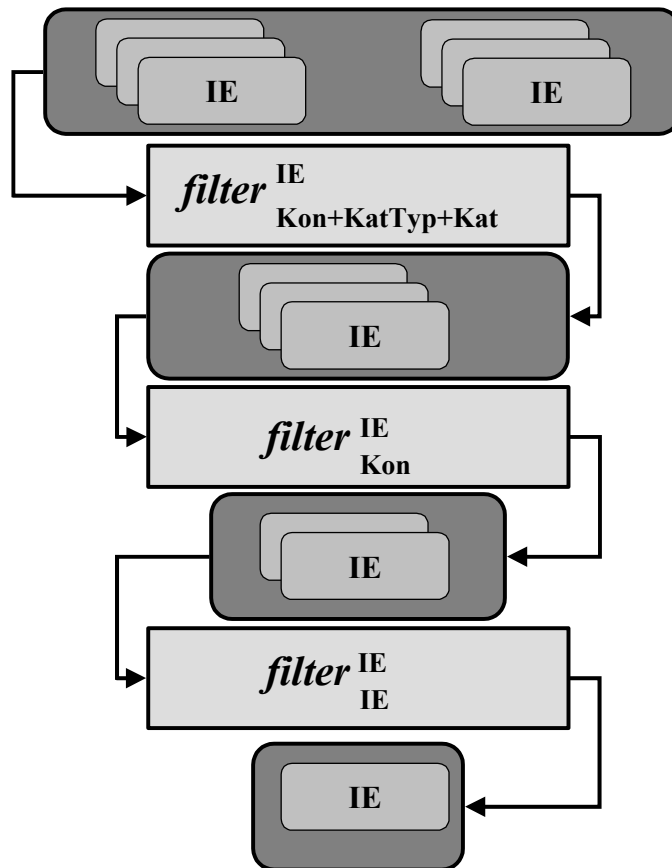


Abbildung 7.3: Konkatenation von Filtern

Durch eine adaptive Vorschlagskomponente können, wie schon zuvor beschrieben (vgl. Kap. 2.2.1, sowie [Kuh95]), Diskrepanzen zwischen den subjektiven und objektiven Informationsbedarfen erkannt werden. Die Auflösung dieser Diskrepanzen dient der Profilhygiene (vgl. [CCS01]). Der Benutzer bekommt genau die Informationen, die er braucht. Eine Überversorgung mit Informationen, die durch einen zu allgemeinen Filter verursacht wird, führt tendenziell erneut zum Phänomen des Information Overload, während ein zu spezieller Filter zu einer Unterversorgung führt, die das Informationsbedürfnis des Benutzers möglicherweise nicht stillen kann und ihm potenziell wichtige Information vorenthält.

Damit den Benutzern adäquate Vorschläge für die Einstellung ihrer Filter gemacht werden können, muss das System versuchen, die Interessen der Benutzer zu erkennen. Damit die Benutzer in ihrer normalen Arbeit nicht gestört werden, sollte dabei auf direktes Feedback (siehe Kap. 2.2.7) verzichtet werden. Stattdes-

sen sollte aus der Systembenutzung auf die Interessen von Benutzern geschlossen werden. Um zu vermeiden, dass die Benutzer das System als inkompetent bezüglich ihrer Interessen (vgl. [Mae94]) wahrnehmen, sollten nur Vorschläge gemacht werden, die mit hinreichender Sicherheit zutreffen. Außerdem sollten Filter nicht eigenmächtig vom System geändert werden, da dies dazu führen kann, dass Benutzern wichtige Informationen vorenthalten werden. Dies kann zu einem Vertrauensverlust (vgl. [Mae94]) gegenüber dem System führen.

Für die Auswahl des Lernverfahrens (vgl. Kap. 2.2.8) ergibt sich, dass das Verfahren nur auf indirektem Feedback beruhen darf. Es kommt hinzu, dass das indirekte Feedback nur positive Beispiele umfasst. D.h., Benutzer zeigen Interesse an bestimmten Informationseinheiten, indem sie z.B. Zusatzinformationen zu dieser Informationseinheit anfordern. Sie zeigen jedoch nicht explizit Desinteresse an Informationseinheiten, die sie nicht näher betrachten. Desinteresse wäre nur durch explizite Nachfrage durch das System herauszufinden, was jedoch vermieden werden sollte, um den Benutzer in seiner Arbeit mit dem System nicht zu stören (siehe oben). Außerdem ist zu beachten, dass dem System in seiner Rolle als Assistent hauptsächlich die Funktion eines Korrektivs zukommt. Dies dient der Profilhygiene. Im Gegensatz zu rein adaptiven Systemen muss die adaptive Assistenzkomponente in ihrer Rolle als Korrektiv nicht zu jedem Zeitpunkt Hinweise geben. WebWatcher (vgl. [JFM97]) annotiert z.B. in jedem Webdokument, das ein Benutzer ausgewählt hat, die drei interessantesten Links. Die Assistenzkomponente (in Broker's Lounge) braucht im Gegensatz dazu nur dann initiativ zu werden, wenn sie hinreichend sicher ist, dass ein Filter geändert werden sollte.

Ein Verfahren, das alle diese Anforderungen erfüllt, ist das Lernen an positiven Beispielen (vgl. Kap. 2.2.9). Dieses Verfahren basiert auf dem χ^2 -Test (vgl. [Sie85]) und wurde im LaboUr-Projekt entwickelt und von mir an das IB-Domänenmodell angepasst und um einem Strukturfaktor (siehe unten) erweitert, der die Struktur des IB-Domänenmodells für die Adaption ausnutzt. Eine Evaluation des ursprünglichen Verfahrens findet sich in [SPK00]. Dies trägt der in [Ped00] formulierten Einsicht Rechnung, dass nur adaptiert werden kann, was zuvor modelliert wurde. Pednault fordert außerdem eine enge Verzahnung von Lernverfahren und interner Repräsentation, wie es in dieser Arbeit umgesetzt wurde.

Das Lernverfahren benötigt als Input jeweils Featurevektoren. Die Spalten des Featurevektors beschreiben Features der beschriebenen Items, die entweder erfüllt sind oder nicht. Features sind zweiwertig. Sie werden von einem zu beschreibenden Item⁶, z.B. einer Informationseinheit entweder erfüllt oder nicht erfüllt.

⁶Ich habe bewusst die Begriffe „Feature“ und „Item“ gewählt, um deutlich zu machen, dass das Lernverfahren unabhängig von einem verwendeten Datenmodell ist. In dieser Arbeit werden als Items Informationseinheiten benutzt, weil auf Informationseinheiten adaptiert werden soll. Als Features werden Konzepte und Kategorien benutzt, da diese zur Charakterisierung und Filterung von Informationseinheiten benutzt werden können (vgl. Def. 7.2 und Def. 7.8). Ein Feature ist

Erfüllte Features werden jeweils durch ein Kreuz dargestellt, während unerfüllte Features jeweils durch einen Strich symbolisiert werden (vgl. Tab. 7.1 und Tab. 7.2). Ein Featurevektor bildet jeweils einen Item auf die Features ab. Zusätzlich werden die Aufrufe des Benutzers für die verschiedenen Items gezählt.

Das Lernverfahren kann dann zur Anwendung kommen, wenn aus einer größeren Menge von Items einzelne Items aufgrund ihrer Eigenschaften ausgewählt und signifikante Abweichungen vom „normalen“ Verhalten deutlich gemacht werden sollen. Signifikante Abweichungen können mit einer vorzugebenden Sicherheit (einem vordefinierten Konfidenzintervall) beobachtet werden. Das bedeutet, dass das LaboUr-Verfahren sich insbesondere als Korrektiv des Benutzers eignet.

Es ist wichtig zu notieren, dass auf die einzelnen Features nur diskrete Werte abgebildet werden können. Ein Kontinuum von Werten kann auf die Features abgebildet werden, indem man sie zu diskreten Intervallen zusammenfasst. Der Featurevektor sollte nicht zu spärlich besetzt sein. D.h., die Features sollten zur Gruppierung von Items genutzt werden können. Überträgt man das LaboUr-Verfahren auf das IB-Domänenmodell, so entsprechen die Items den Informationseinheiten des IB-Domänenmodells. Die Ereignisse die vom System beobachtet werden, sind Informationen über Informationseinheiten, die ein Benutzer abrufen. Es ist zu beachten, dass als Itemmenge nur die Informationseinheiten in die Berechnung mit einfließen sollten, die im Moment auch tatsächlich angeboten werden und nicht durch einen der verschiedenen, hintereinander geschalteten Filter herausgefiltert wurde. Werden Filter verändert und wirkt sich dies auf die Menge der angebotenen Informationseinheiten aus, so ist eine Neuberechnung der verschiedenen Relevanzwerte durchzuführen. Als Features kommen Konzepte und Kategorien in Frage, da sie beide zur Gruppierung von Informationseinheiten genutzt werden können. Das bedeutet, dass sowohl für Kategorie- als auch für Konzeptfilter eine adaptive Filterunterstützungskomponente angeboten werden kann.

7.6.1 Konzeptfiltersupport

Für eine adaptive Konzeptfilterunterstützung dienen die verschiedenen Konzepte als Features. Es ergeben sich Featurevektoren, wie sie in Tabelle 7.1 dargestellt sind. Die Erwartungswerte (gemäß Formel 2.1) werden berechnet, indem man die Anzahl der Informationseinheiten eines Konzepts durch die Anzahl aller Informationseinheiten dividiert. Liegt eine Konzepthierarchie vor, wenn z.B. ein Konzept eine Spezialisierung eines anderen Konzeptes ist, wird dies im Featurevektor dadurch berücksichtigt, dass alle Items, die dem spezialisierten Konzept zugeordnet

erfüllt, wenn eine Verbindung zwischen dem beschriebenen Item und dem Feature besteht. Werden z.B. als Items Informationseinheiten beschrieben und als Features Konzepte benutzt wie in Tabelle 7.1, dann ist das Feature mit dem Konzept der Informationseinheit erfüllt.

werden, auch gleichzeitig dem generalisierten Feature zugeordnet werden. Die Unterstützungsleistung durch das System könnte darin bestehen, dem Benutzer vorzuschlagen, ein Konzept aus einem Konzeptfilter zu entfernen, wenn das System ein signifikantes Desinteresse des Benutzers festgestellt hat. Im umgekehrten Fall, wenn ein Konzept nicht im Filter enthalten ist und trotzdem ein signifikantes Interesse registriert wurde, könnte das System den Benutzer nahelegen, dieses Konzept in den Filter wieder aufzunehmen.

	$Konzept_1$	$Konzept_2$	$\dots$	$Konzept_m$
$Informationseinheit_1$	-	x	$\dots$	-
$Informationseinheit_2$	x	-	$\dots$	-
$Informationseinheit_3$	-	-	$\dots$	x
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$Informationseinheit_n$	x	x	$\dots$	-

Tabelle 7.1: Konzeptfilter-Featurevektoren

7.6.2 Kategoriefiltersupport

Für die adaptive Kategoriefilterunterstützung wird das Verfahren leicht abgewandelt. Da jede Informationseinheit mindestens mit einer Kategorie für jede Kategoriedimension klassifiziert werden muss, werden die verschiedenen Kategoriedimensionen getrennt voneinander behandelt (siehe Tabelle 7.2). Die Itemmenge ist wie zuvor die Menge aller Informationseinheiten. Wegen der Trennung nach Kategorietypen gibt es verschiedene Featuremengen: Die Kategorien einer Kategoriedimension (eines Kategorietyps) bilden eine Featuremenge. Um die Relevanz einer Kategorie (Featurerelevanz) zu errechnen, werden nur die Verteilungen für die Kategorien mit ein und demselben Kategorietypen betrachtet. D.h., es handelt sich um verschiedene Featurevektoren, die nach einem einheitlichen Verfahren behandelt werden. Um die hierarchische Eigenschaft der Kategorien im Featurevektor abzubilden, erhalten Informationseinheiten, die für eine Kategorie einen Eintrag im Featurevektor haben, automatisch für alle Oberkategorien weitere Einträge im Featurevektor. Ähnlich wie bei den Konzeptfiltern kann das System dem Benutzer vorschlagen, Kategorien in einen Kategoriefilter aufzunehmen, wenn diese vom System als relevant erkannt wurde und diese Kategorien bisher nicht im Filter enthalten sind. Umgekehrt wird dem Benutzer nahegelegt, irrelevante Kategorien, die im Filter enthalten sind, aus dem Filter zu entfernen.

	<i>Kategorietyp₁</i>			<i>Kategorietyp₂</i>			...	<i>Kategorietyp_k</i>
	<i>Kat₁</i>	...	<i>Kat_m</i>	<i>Kat_{m+1}</i>	...	<i>Kat_{m+l}</i>	...	<i>Kat_{m+...+k}</i>
<i>IE₁</i>	-	...	x	x	...	-	...	x
<i>IE₂</i>	x	...	x	-	...	x	...	-
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
<i>IE_n</i>	x	...	-	x	...	-	...	-

Tabelle 7.2: Kategoriefilter-Featurevektoren

7.6.3 Relevanzfilter

Basierend auf der adaptiven Unterstützung für Konzept- und Kategoriefilter kann das System die Relevanz von Informationseinheiten bewerten. Dabei wird die Formel für die Berechnung der Itemrelevanz (siehe Formel 2.4) angewendet. Die Relevanz von Informationseinheiten (siehe Formel 7.2) wird durch die Kombination der Konzeptfilterrelevanz und der Relevanz für die verschiedenen Kategorietypen berechnet. Da Informationseinheiten in einer Kategoriedimension mehr als einer Kategorie zugeordnet werden können, müssen die verschiedenen Kategorien (in einer Kategoriedimension) normalisiert werden (siehe Formel 7.1) bevor sie zur Berechnung der Relevanz von Informationseinheiten herangezogen werden können. Die Relevanz von Informationseinheiten kann zur Filterung von Informationseinheiten eingesetzt werden. Dabei passen nur die Informationseinheiten durch den Filter, die einen Mindestrelevanzwert aufweisen. Es handelt sich dabei um einen intensionalen Instanzenfilter, da eine Eigenschaft von Informationseinheiten, nämlich ihre Relevanz für einen einzelnen Benutzer, für die Filterung ausgenutzt wird. Die Relevanzwerte können zur adaptiven Sortierung (vgl. [Bru96]) von Informationseinheiten benutzt werden. Dabei werden die Informationseinheiten nach ihrer Relevanz geordnet dem Benutzer angeboten.

$$\text{Kategorietyprelevanz} = \frac{1}{m} \sum_{i=0}^m \text{Kategorierelevanz}_i \quad (7.1)$$

$$\text{IE-Relevanz} = \frac{\text{Konzeptrelevanz} + \sum_{i=0}^n \text{Kategorietyprelevanz}_i}{n+1} \quad (7.2)$$

7.6.4 Beschleunigung der Lernkurve

Wie schon zuvor beschrieben, beruht das hier angewandte Verfahren aus dem LaboUr-Projekt auf einer statistischen Signifikanzanalyse. Um mit gegebener Sicherheit (z.B. 99%) Aussagen über den Benutzer machen zu können, muss bei diesem Verfahren eine größere Anzahl von Ereignissen abgewartet werden, bevor

man zu schlüssigen Aussagen über den Benutzer kommen kann. Um diesen Effekt zu mindern, kann man einerseits auf Stereotypen (vgl. Kap. 2.2.5) zurückgreifen und andererseits die Struktur des IB-Domänenmodells zur Berechnung der Relevanz ausnutzen.

Stereotypen

Die Nutzung von Stereotypen in Benutzermodellierungssystemen hat lange Tradition. Stereotypen werden benutzt, um das Akquisitionsverfahren zu vereinfachen und abzukürzen. Dabei werden Benutzer einem oder mehreren Stereotypen zugeordnet. Die Stereotypen werden benutzt, um das Interessenprofil des Benutzers zu initialisieren. Danach arbeitet das System mit dem modifizierten Profil weiter.

Übertragen auf das IB-Domänenmodell und die verschiedenen Filter, kann ein Stereotyp entweder durch eine Festlegung der Relevanzwerte für Konzepte und Kategorien oder einfacher durch die Angabe von Informationseinheiten definiert werden, die von Benutzern, die diesem Stereotyp zugeordnet werden sollen, typischerweise ausgewählt werden. Aus dieser Sequenz von Informationseinheiten können dann die Relevanzwerte für die verschiedenen Konzepte und Kategorien errechnet werden. Daraus werden wiederum die Relevanzwerte für alle Informationseinheiten berechnet.

Strukturfaktor

Das IB-Domänenmodell kann als Graph aufgefasst werden (vgl. Def. 6.1). Die Knoten in diesem Graphen sind die Basisbestandteile des IB-Domänenmodells (Konzepte, Kategorietypen, Kategorien und Informationseinheiten). Die Kanten sind die verschiedenen Relationstypen (typeOf, isA, classifies, relatedTo, partOf). Kategorie- und Konzeptfilterunterstützung nutzen die typeOf-, isA- und classifies-Beziehungen aus. Um den Lernprozess zu beschleunigen, kann man auch die Assoziationen (relatedTo-Beziehungen) und die partOf-Beziehungen in die Berechnung einbeziehen. Dabei geht man davon aus, dass Benutzer sich auch für im Domänenmodellgraphen benachbarte Informationseinheiten interessieren und propagiert das Ereignis an die benachbarten Informationseinheiten. Dabei wird die Wertigkeit der Ereignisse mit jeder Propagierung abgeschwächt. Ich will dies an einem Beispiel aus der Domäne Forschungsförderung verdeutlichen. Wenn ein Benutzer Informationen über das Förderprogramm „DFG-Normalverfahren“ anfordert, interessiert er sich mit hoher Wahrscheinlichkeit auch für den Förderer (die Informationseinheit „DFG“) und den entsprechenden Ansprechpartner für dieses Förderprogramm. Bei einem Propagationsschritt werden die Zähler für die benachbarten Informationseinheiten um den Propagationsfaktor (z.B. 0, 5) erhöht. Dies hat ggf. Rückwirkungen auf die assoziierten Konzept- und Kategoriefilter. Es

ist zu beachten, dass zur Propagierung nur part-of- und related-To-Beziehungen benutzt werden, da die anderen Beziehungen schon im bisherigen Verfahren Berücksichtigung gefunden haben.

7.7 Push-Filter

Durch die verschiedenen miteinander kombinierbaren Filter eines Profils kann eine Sicht auf die Inhalte des IB-Domänenmodells definiert werden. Diese Profile können benutzt werden, um die verschiedenen Benutzer, deren Interessen durch die Profile repräsentiert werden, über regelmäßig mit Hilfe eines Push-Dienstes über Veränderungen in ihren Profilen zu informieren (vgl. Kap. 8.7.7). Dieser regelmäßige Push-Dienst geschieht im Batch-Modus. D.h., in regelmäßigen Abständen werden die verschiedenen Profile auf die Inhalte des IB-Domänenmodells angewendet und die Veränderungen werden an den Benutzer weitergeleitet.

Neben diesem Batch-Szenario ist auch ein On-line-Szenario denkbar, bei dem nicht regelmäßig alle Profile gegen den Inhalt des IB-Domänenmodells getestet werden, sondern nur solche, die von den Veränderungen auch betroffen sind. Dies entspricht im Wesentlichen der Vorgehensweise, wie sie in [SJ00] für das View-Management in ConceptBase beschrieben wurde.

Um die Vorgehensweise zu erläutern, gehe ich von dem Fall aus, dass eine neue Informationseinheit $infoUnit_{new}$ in das IB-Domänenmodell eingefügt wurde. $infoUnit_{new}$ wird durch das Konzept kon typisiert und durch die Kategorien kat_1 und kat_2 in den Kategoriedimension $katTyp_1$ und $katTyp_2$ klassifiziert.

Um den Aufwand für die Ermittlung der in Frage kommenden Profile möglichst gering zu halten, werden die verschiedenen Zielfilter nacheinander angewendet. D.h., dass zuerst alle Profile gefiltert werden, die den Konzeptfilter für das Konzept kon aufweisen. Aus den verbleibenden Profilen werden diejenigen ausgefiltert, die einen Kategoriefilter für die Kategorien kat_1 und kat_2 in den entsprechenden Kategoriedimensionen haben. Auf die restlichen Profile werden die verschiedenen Instanzenfilter angewendet.

Eine Ausnahme bilden dabei die Profile, die einen Relevanzfilter für das Konzept kon beinhalten, weil in diesem Falle eine einfache Filterung wie z.B. für Konzeptfilter nicht möglich ist, da zum Einen die Relevanz der neuen Informationseinheit $infoUnit_{new}$ in den jeweiligen Profilen errechnet werden muss und zum Anderen die Integration einer neuen Informationseinheit zu einer Veränderung der Berechnungsgrundlage für die Relevanzberechnung führt.

Dies bedeutet, dass die Relevanzberechnung für alle Informationseinheiten der verbliebenen Profilen neu durchgeführt werden muss. Will man dagegen nur die Relevanz der neuen Informationseinheit $infoUnit_{new}$ berechnen, weil man z.B. nur dann einen Benutzer unmittelbar benachrichtigen will, wenn eine neue Infor-

mationseinheit erzeugt wurde, die einen Relevanzwert von mindestens 80 Punkten (von 100 Punkten) hat, so genügt es, die entsprechenden Werte anzupassen, die für die Berechnung benötigt werden.

Zur Berechnung der Relevanz von Informationseinheit $infoUnit_{new}$ wird nach Formel 7.2, die Relevanz der beteiligten Kategorietypen $katTyp_1$, $katTyp_2$ und des Konzeptes kon benötigt. Die Relevanz des Konzeptes kon errechnet sich nach den Formeln 2.1 – 2.4. Dabei trägt die neue Informationseinheit $infoUnit_{new}$ zu einer Veränderung des durch Formel 2.1 berechneten Erwartungswertes für das Konzept kon bei.

Die Berechnung der Kategorietypenrelevanz für die Kategorietypen $katTyp_1$ und $katTyp_2$ basiert auf Formel 7.1. Dazu müssen zunächst die Relevanzwerte für die beteiligten Kategorien kat_1 und kat_2 errechnet werden. Die Relevanzwerte für die beiden Kategorien werden ebenfalls mit den Formeln 2.1 – 2.4 ermittelt, wobei die neue Informationseinheit $infoUnit_{new}$ zu einer Veränderung der Erwartungswerte für die Kategorien kat_1 und kat_2 gemäß Formel 2.4 führen.

Wurde hingegen die Informationseinheit $infoUnit_{update}$ lediglich verändert und wurde dabei die Kategorisierung und die Zugehörigkeit zu einem Konzept nicht verändert, kann eine Neuberechnung der Relevanz $infoUnit_{update}$ sowie der übrigen Informationseinheiten unterbleiben, da die Berechnungsgrundlage nicht verändert wurde. Wurde die Kategorisierung geändert, so genügt eine Anpassung der veränderten Erwartungswerte für die neuen Kategorien zur Berechnung des Relevanzwertes von $infoUnit_{update}$. Zur Berechnung der Relevanz aller Informationseinheiten muss zusätzlich der Erwartungswert für die ehemaligen Kategorien von $infoUnit_{update}$ angepasst werden. In diesem Fall müssen alle Informationseinheiten, die mit einer neuen oder einer alten Kategorie von $infoUnit_{update}$ oder mit einer Oberkategorie kategorisiert wurden neu berechnet werden.

Wenn eine Informationseinheit $infoUnit_{delete}$ aus dem IB-Domänenmodell gelöscht wurde, macht eine Berechnung der Relevanz von $infoUnit_{delete}$ wenig Sinn. Gleichwohl ändern sich die Erwartungswerte zur Berechnung von Konzept- und Kategoriefiltern. Dadurch können sich die Relevanzwerte für die beteiligten Informationseinheiten ändern.

Die Neuberechnung der Relevanzwerte für Informationseinheiten kann auf die Informationseinheiten eingeschränkt werden, die mit einer Kategorie klassifiziert wurden, deren Relevanzwert sich tatsächlich ändert oder die mit einem Konzept typisiert wurden, dessen Relevanzwert verändert wurde. Hintergrund dieser Einschränkung ist, dass der Relevanzwert für viele Kategorien bei 0 liegt, da keine signifikante Abweichung vom Erwartungswert festgestellt werden konnte. Auch nach einer geringfügigen Änderung des Erwartungswertes führt in vielen Fällen nicht zu einer Veränderung der Relevanz der entsprechenden Kategorie. Ähnliches gilt für die Bewertung der Relevanz von Konzepten.

Kapitel 8

Broker's Lounge

Broker's Lounge ist ein Werkzeugkasten, der es erlaubt, Information-Brokering-Systeme zu erstellen. Broker's Lounge orientiert sich an dem in Kapitel 3.1 beschriebenen Aufgaben- und Rollenmodell für Information-Brokering. Die Werkzeuge sind für die verschiedenen im Information-Brokeringprozess anfallenden Aufgaben ausgelegt und können beliebig kombiniert werden. Dadurch können viele verschiedene Brokeringmodelle (vgl. Tab. 8.1) abgebildet werden.

Broker's Lounge wurde zur Unterstützung von Information-Brokering konzipiert. In der Tradition der Balance von Adaptivität und Adaptierbarkeit (vgl. Kap. 2.2.1) ist es nicht Ziel, den Broker überflüssig zu machen und durch das System zu ersetzen, sondern ihn soweit wie möglich zu unterstützen. In Broker's Lounge werden Benutzer¹ des Systems dadurch unterstützt, dass ihnen Routinetätigkeiten soweit wie möglich abgenommen werden, damit mehr Zeit für intellektuell mehr fordernde Tätigkeiten bleibt, mit denen ein Mehrwert geschaffen werden kann. So ermöglicht z.B. die automatische Beobachtung und Bewertung von Quellen dem Broker mehr Zeit für die Kategorisierung von relevanten Dokumenten aufzuwenden. Durch die Kategorisierung wird der Wert der kategorisierten Information für die Konsumenten gesteigert (vgl. [Wor98]).

Im kundenorientierten Prozess ist die Abbildung der Interessen der Konsumenten (Profilbildung) von zentraler Bedeutung. Der Aufbau eines Interessenprofils durch einen Broker setzt die genaue Kenntnis der Informationsbedürfnisse seines Kunden voraus. Dadurch wird die Zahl der Kunden limitiert, die ein Broker bedienen kann.

Die Unterstützung bei der Profilbildung durch eine adaptive Systemkomponente ermöglicht es, die Brokeringdienstleistungen einem größeren Kundenkreis anzubieten, da eine genaue Kenntnis der Informationsbedürfnisse des Konsumenten

¹An dieser Stelle ist anzumerken, dass die Benutzer des Systems nicht notwendigerweise Informationskonsumenten sein müssen. Es kann sich auch um Broker und Informationsanbieter handeln, da alle im Information-Brokeringprozess beteiligten Personen unterstützt werden sollen.

Name	Quellenverwaltungs- aufgaben	Domänenverwal- tungsaufgaben	kundenorientierte Aufgaben
Semantic-Web-Systeme	Anbieter/Broker	Anbieter	Konsument
COMMA	Anbieter/Broker	Anbieter	System/Konsument
Planet-Onto	Anbieter	Broker	Konsument
Subject Gateways	Broker	Broker	Konsument
Lehr-Lernsysteme	-	Anbieter	System
KBS-Hyperbook	Anbieter	Anbieter	System
termabsierte In- formationssysteme	-	Broker	System
Oyster	Broker	Broker	System/Konsument
Broker's Lounge	beliebig	beliebig	System/beliebig

Tabelle 8.1: Aufgabenverteilung bei den verschiedenen Systemen

ten durch den Broker nicht notwendig ist, da dies vom System übernommen wird. Konsumenten erhalten dadurch Unterstützung bei der Profilbildung, ohne dass der Broker eingreifen muss. Wird die Profilbildung durch den Broker durchgeführt, so bekommt der Broker Unterstützung bei der Profilbildung durch das System.

8.1 Positionierung von Broker's Lounge

Die verschiedenen Brokeringmodelle der in dieser Arbeit beschriebenen Ansätze und System, die in Tabelle 8.1 beschrieben wurden, lassen den Schluss zu, dass die Ansätze aus dem Bereich ontologiebasierter Informationssysteme in der Beziehung zwischen Anbietern und Broker variieren, während die Beziehung zwischen Brokern und Konsumenten in fast allen Systemen derart gestaltet ist, dass die Interessen von den Konsumenten selbst angegeben werden müssen. Einige Ansätze wie Ontobroker (vgl. [FDES98]) und MyPlanet² (vgl. [KDM⁺01]) sprechen zwar von Proaktivität, gemeint ist damit jedoch nur, dass Benutzer neue Informationen zugesendet bekommen, die durch ihr Interessenprofil passen, wenn diese in das IB-Domänenmodell aufgenommen werden. Es handelt sich bei diesen Systemen damit um Push-Systeme und nicht um adaptive Systeme. Im Gegensatz dazu wird im CoMMA-Projekt Proaktivität hinsichtlich des Benutzerprofils angestrebt. D.h., das System versucht sich selbst an die Benutzerinteressen anzupassen.

Dies wird von fast allen Benutzermodellierungssysteme angestrebt, die in die-

²MyPlanet ist eine Profilbildungsschnittstelle, die in der Planet-Onto-Architektur zum Einsatz kommt.

ser Arbeit vorgestellt wurden. Lehr-Lernsysteme und termbasierte Informationssysteme gestalten den gesamten kundenorientierten Prozess selbst. Bei domänenmodellbasierten Informationsfilterungssystemen wird eine Teilung der kundenorientierten Aufgaben zwischen dem System und dem Benutzer angestrebt. Während die meisten termbasierten Informationsfilterungssysteme direkt auf den Quellen (Dokumenten) arbeiten, basieren domänenmodellbasierte Informationssysteme, wie z.B. das Oyster-System, auf einem expliziten Domänenmodell.

In Broker's Lounge wird zum einem angestrebt, möglichst verschiedene Brokeringmodelle realisieren zu können (vgl. [JKN01]). D.h., Broker's Lounge soll die verschiedenen Brokeringmodelle abbilden können, die von den vorgestellten Systemen benutzt werden, ohne dabei auf ein Modell festgelegt zu sein. Zum anderen soll es sich im kundenorientierten Prozess wie ein adaptives Informationsfilterungssystem mit einem expliziten Domänenmodell verhalten, um so möglichst weitestgehende Unterstützung bei der Profilbildung anbieten zu können. D.h., dass die Profilbildung nicht mehr vom Broker in Zusammenarbeit mit dem Benutzer ausgeführt werden muss, da das System sich an den Benutzer (den Konsumenten) anpassen und ihn so bei der Definition seines Interessenmodells unterstützen kann.

In keinem Fall ist angestrebt, Broker's Lounge als ein rein adaptives System zu gestalten, da dann nicht sichergestellt werden kann, dass die Informationsbedürfnisse von Konsumenten wirklich befriedigt werden, da das System die Informationsbedürfnisse des Benutzers möglicherweise falsch einschätzt und ihm als Folge davon wichtige Information vorenthält. Stattdessen ist eine Teilung der Aufgaben zwischen dem System und dem Benutzer (dem Konsumenten) angestrebt, wobei der Benutzer die volle Kontrolle über sein Profil behält und das System lediglich Vorschläge zur Änderung des Profils macht.

8.2 Umsetzung des Aufgabenmodells

Im Information-Brokeringprozess (vgl. Kap. 3.1) fallen im Wesentlichen drei Aufgabengruppen an:

1. Quellenverwaltungsaufgaben,
2. Domänenverwaltungsaufgaben und
3. Personalisierung.

Broker's Lounge bildet diesen Prozess nach (vgl. Abb. 8.1).

Innerhalb der Quellenverwaltungsaufgaben wird ein Modell von den verschiedenen, relevanten Quellen aufgebaut. Dieses Quellenmodell enthält Meta-Informationen über die in den Quellen verfügbaren Dokumente und die Struktur der

Quellen. Damit der Broker jederzeit auf dem neusten Stand ist, muss er die verschiedenen Quellen regelmäßig untersuchen, um neue oder veränderte Dokumente zu finden und sein Informationsangebot entsprechend anzupassen.

Diese Aufgabe wird in Broker's Lounge von Webrobots übernommen, die auf eine oder mehrere (Web-)Quellen angesetzt werden, um zu untersuchen, ob diese Quelle neue oder veränderte Dokumente enthält. Damit Veränderungen von Dokumenten festgestellt werden können, bauen die Robots ein eigenes Archiv von bearbeiteten Dokumenten auf, mit dem sie jederzeit Veränderungen in einzelnen Dokumenten aufdecken können. Durch die Webrobots wird der Broker entlastet, da er nur noch die verschiedenen Robots für die Quellen konfigurieren muss. Die Quellenadministrations-Schnittstelle zur Konfiguration von Webrobots und deren Funktionsweise werden in Kapitel 8.5 beschrieben.

Alle neuen oder veränderten Dokumente müssen kategorisiert werden. Dies kann in zwei Schritten geschehen:

1. Zuerst muss geprüft werden, ob ein Dokument relevante Informationen für die Domäne enthält. Ist dies der Fall, muss
2. geprüft werden, ob das IB-Domänenmodell diese Information bereits enthält. Ist dies nicht der Fall, muss das IB-Domänenmodell entsprechend angepasst werden.

Der erste Schritt dient der Einschränkung der Dokumente, die weiter kategorisiert werden müssen und wird in Broker's Lounge automatisiert (inhaltliche Analyse, vgl. Abb. 8.1), indem die neuen oder veränderten Dokumente gegen die Inhalte des IB-Domänenmodells geprüft werden. Übereinstimmungen zwischen einem Dokument und Teilen des IB-Domänenmodells werden in der inhaltlichen Analyse benutzt, um das Dokument mit einem Domänenwert zu annotieren. Für jede Übereinstimmung mit dem IB-Domänenmodell bekommt ein Dokument eine Punktzahl. Die Punkte eines Dokumentes werden aufaddiert und auf die Länge des Dokumentes normalisiert. Je höher der so errechnete Domänenwert ist, umso relevanter ist das Dokument für die Domäne, die das IB-Domänenmodell beschreibt. Die inhaltlichen Übereinstimmungen zwischen Dokumenten und dem IB-Domänenmodell können als Richtschnur für die Einordnung in das IB-Domänenmodell im zweiten Schritt benutzt werden.

Während der erste Schritt der Kategorisierung automatisiert werden kann, ist der zweite Schritt eine klassische, intellektuelle Aufgabe, die eine hohe Qualität der kategorisierten Inhalte sicherstellt. Um den zweiten Schritt zu unterstützen, bietet Broker's Lounge einen Dokumentenbrowser an, der neben der Visualisierung von Dokumenten, die Übereinstimmungen mit dem IB-Domänenmodell und weitere Meta-Informationen zur Kontextualisierung visualisiert. Der Dokumentenbrowser, die Bewertung von Dokumenten durch die inhaltliche Analyse und

die intellektuelle Aufbereitung von Dokumenten werden in Kapitel 8.6 beschrieben.

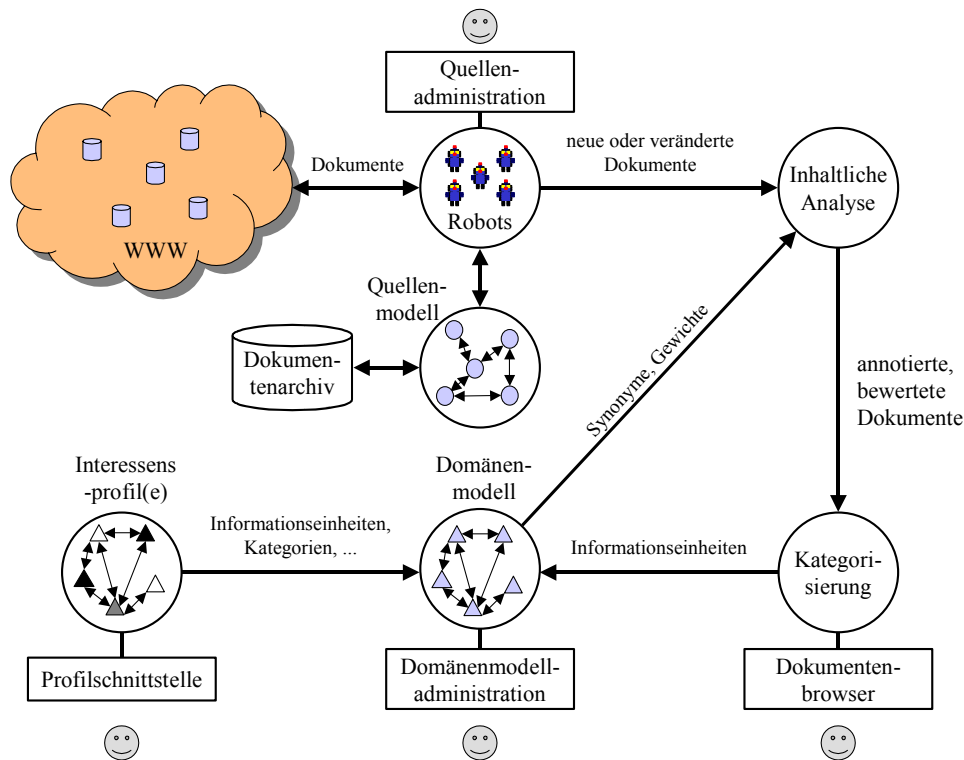


Abbildung 8.1: Information-Brokeringprozess in Broker's Lounge

Um die Wahrscheinlichkeit von Übereinstimmungen zwischen Dokumenten und Einträgen im IB-Domänenmodell zu erhöhen, können für jede Informationseinheit im IB-Domänenmodell Synonyme angegeben werden. Alle Informationseinheiten im IB-Domänenmodell können darüber hinaus mit einer Gewichtung versehen werden, die in die Berechnung des Domänenwerts einfließt. Ebenso können alle Kategorien mit einer Gewichtung belegt werden, um so Schwerpunkte im IB-Domänenmodell zu setzen. Neben der Nutzung des IB-Domänenmodells zur inhaltlichen Analyse dient das IB-Domänenmodell der Beschreibung und Verwaltung der Inhalte, mit denen der Broker handelt. Die Verwaltung des IB-Domänenmodells wird in Kapitel 8.4 beschrieben.

Die zentrale Rolle des IB-Domänenmodells wird dadurch untermauert, dass es zur Abbildung von Konsumenteninteressen benutzt wird (vgl. Abb. 8.1). Interessenprofile werden, wie in Kapitel 7 beschrieben, durch eine Kombination von Domänenmodellfiltern definiert. Interessenprofile können in der Profilschnittstelle (vgl. Kap. 8.7) verwaltet werden. Die Filter werden zur Reduktion der Inhal-

te des IB-Domänenmodells auf die relevanten Informationseinheiten benutzt und können als Suchanfrage aufgefasst werden. Um eine Integration von Suchen und Browsen zu ermöglichen, können die Auswirkungen einer Filteränderung sofort beobachtet werden. Dadurch können die Filter besser an die Informationsbedürfnisse angepasst werden. Außerdem kann die Profilschnittstelle zum Browsen der Inhalte des IB-Domänenmodells benutzt werden.

Um die Hemmschwelle für die Benutzung von Broker's Lounge für Endkonsumenten möglichst niedrig zu halten, wird für die Profilschnittstelle eine Webschnittstelle angeboten. Die Profil-Webschnittstelle ist in ihrem Funktionsumfang gegenüber der Profilschnittstelle des „normalen“ Broker's Lounge-Client reduziert, um den ungeübten Benutzer nicht zu verwirren. Die Profil-Webschnittstelle wird im Rahmen der Profilbildungsschnittstelle in Kapitel 8.7.11 beschrieben. In manchen Fällen handelt der Broker nicht mit Informationseinheiten aus dem Domänenmodell, sondern mit Dokumenten, die Informationseinheiten beschreiben. Dies ist z.B. in MarketMonitor (vgl. Kap.9.2) der Fall. In diesem Fall dient der Dokumentenbrowser als Schnittstelle für Informations-Endkonsumenten. Um hier ebenfalls die Hemmschwelle zu senken, ist in Broker's Lounge eine Webschnittstelle für den Dokumentenbrowser integriert. Die Webschnittstelle für den Dokumentenbrowser wird in Kapitel 8.6 beschrieben.

Broker's Lounge unterstützt die verschiedenen Aufgaben, die im Brokeringprozess anfallen. Die Aufgaben können von einem Administrator (vgl. Abb. 8.2) beliebig an die verschiedenen Benutzer des Systems vergeben werden. Abhängig von den zugewiesenen Rechten bekommt der Benutzer die Werkzeuge und Schnittstellen angeboten, die er zur Erfüllung seiner Aufgaben benötigt. Durch diese flexible Aufgabenverteilung können verschiedenste Brokeringmodelle abgebildet werden. Dies löst die Kapitel 8.1 und Tabelle 8.1 postulierte Position von Broker's Lounge ein.

Das Semantic-Web-Brokeringmodell, bei dem die Inhalte von den Anbietern selbst strukturiert werden, kann dadurch abgebildet werden, dass man den verschiedenen Anbietern das Recht zur Kategorisierung zubilligt. In diesem Fall entfällt die Aufgabe der Quellenadministration, da Quellen- und IB-Domänenmodell zusammenfallen. Die Profilbildung kann entweder von einem Broker oder von den Konsumenten selbst ausgeführt werden. Ein Schwachpunkt des Semantic-Web-Brokeringmodells ist, dass es die Kooperation der Anbieter voraussetzt. Mit Broker's Lounge kann auf die mangelnde Kooperation von Anbietern reagiert werden, indem die Kategorisierung vom Broker selbst ausgeführt wird. D.h., Broker's Lounge kann nicht nur verschiedene Brokeringmodelle abbilden, Broker's Lounge kann auch mehrere Brokeringmodelle miteinander verbinden.

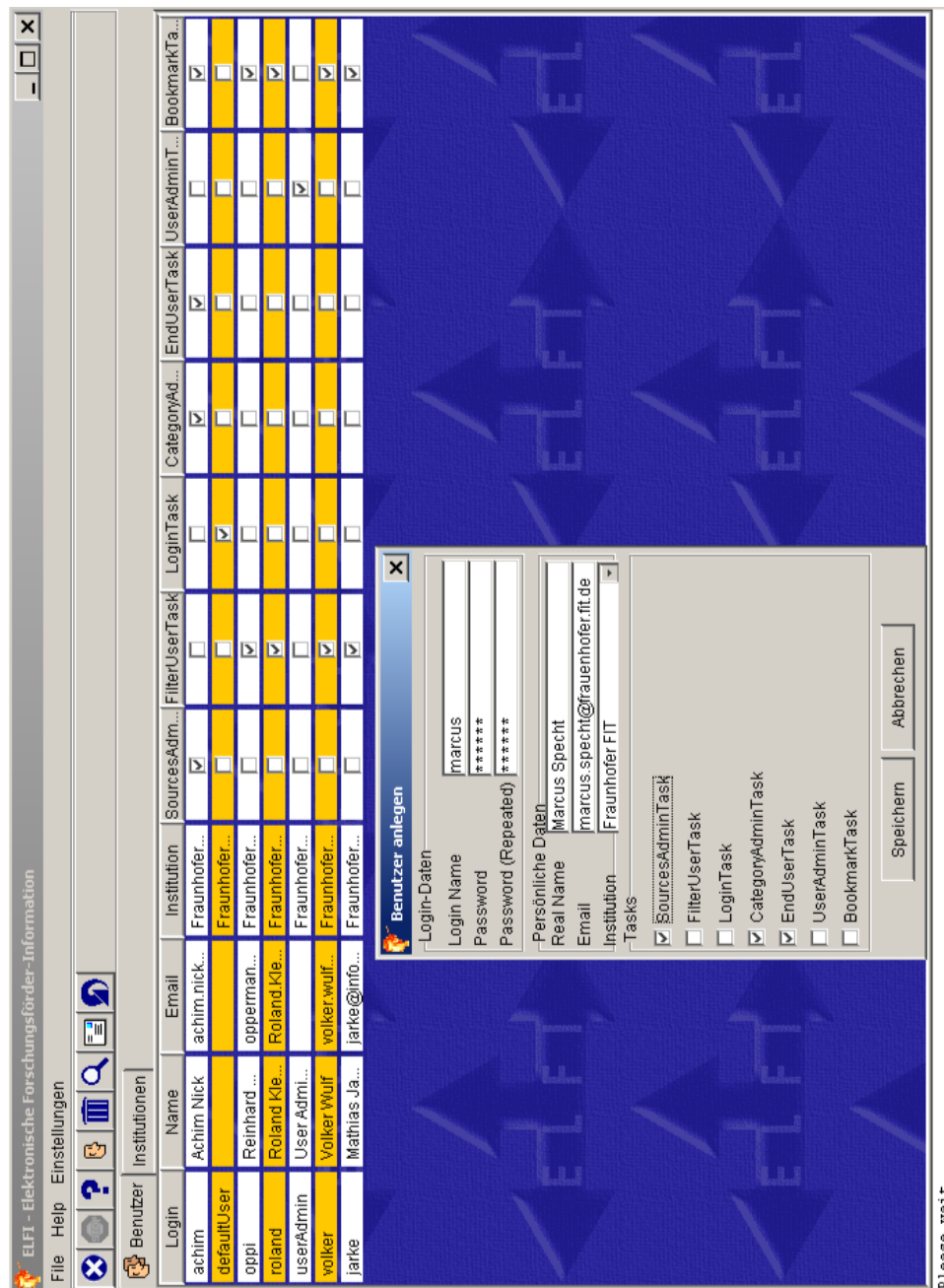


Abbildung 8.2: Benutzeradministration in Broker's Lounge

senhierarchien in der Instanzen- und Typenebene. Beide sind über eine Assoziation miteinander verbunden, die durch die Unterklassen jeweils verfeinert wird.

- `AlphaNumericComponent` und `AlphaNumericComponentType` werden benutzt, um einfache Textattribute abzubilden.
- `URLComponent` und `URLComponentType`³ werden benutzt, um URLs als einfache Attribute abzubilden.
- `OntologyContainer` und `OntologyContainerType` können zur Gruppierung einfacher Attribute genutzt werden. Ein Anwendungsbeispiel für einen Container ist die Modellierung einer Adresse, die aus einigen einfachen Attributen, wie Straße, Postleitzahl usw. besteht. Außerdem können Container mit anderen Containern in Beziehung gesetzt werden, da es sich bei Containern ebenfalls um Komponenten (`Component`) handelt. Dabei werden die Beziehungen auf der Typenebene festgelegt und in der Instanzenebene mit Leben gefüllt. Es können drei Typen von Beziehungen unterschieden werden:
 1. Einfache Assoziationen bilden einfache Beziehungen zwischen zwei Containern ab.
 2. `PartOf`-Beziehungen bilden Beziehungen ab, bei denen der eine Partner der Relation Teil des anderen Beziehungspartners ist. Ein Beispiel für eine `PartOf`-Relation (zwischen Konzepten bzw. der entsprechenden Informationseinheiten) ist die Beziehung zwischen Teil- und Oberprogrammen in der Domäne Forschungsförderung. Durch die Unterscheidung zwischen einfachen Assoziationen und `PartOf`-Relationen wird dem Strukturfaktor des Lernverfahrens die Möglichkeit gegeben, diese Relationen im Strukturfaktor (vgl. Kap. 7.6.4) unterschiedlich zu bewerten, indem z.B. das Teilprogramm eines Förderprogramms bei der Propagation einen höheren Wert erhält als der Förderer des Förderprogramms.
 3. Bei Spezialisierungs/Generalisierungsbeziehungen zwischen Containern erbt der speziellere Container alle Beziehungen des Super-Containers. Ein Beispiel für eine Spezialisierungsrelation sind Projektträger, da es sich bei diesen um eine spezielle Form von Förderern handelt.

³Neben den in Abb. 8.3 dargestellten einfachen Attributen (`AlphaNumericComponent` und `URLComponent`) können beliebige weitere einfache Attribute definiert werden, die hier nicht weiter erläutert werden sollen.

- **NamedContainer** und **NamedContainerType** haben im Unterschied zu **Container** bzw. **ContainerType** einen Eigennamen. Dieser Eigennamen kann in der inhaltlichen Analyse, bei der Dokumente gegen die Inhalte des Domänenmodells geprüft werden, benutzt werden. Um eine größere Abdeckung zu erzielen, können zu dem Eigennamen Synonyme definiert werden. Zusätzlich können **NamedContainer** mit einer Gewichtung versehen werden.
- **InformationItem** und **Concept** erweitern **NamedContainer** und **NamedContainerType** um drei weitere Eigenschaften:
 1. Um den Rückbezug von Informationseinheiten zu Dokumenten im WWW herstellen zu können, können Informationseinheiten (**InformationItem**) mit einer URL annotiert werden. Zusätzlich können Informationseinheiten mit einer weiteren URL mit einem Bild (z.B. einem Firmenlogo) assoziiert werden. Diese Informationen werden den Konsumenten angezeigt, wenn sie sich eine Detailansicht einer Informationseinheit anschauen (vgl. Kap. 8.7 und Abb. 8.22) und dient der Sicherstellung der Authentizität der angebotenen Information, durch den Rückbezug auf die Originalquelle.
 2. Die Informationseinheiten dienen der schnellen Information von Konsumenten. Aus diesem Grund hat **InformationItem** ein zusätzliches Attribut, mit dem eine Kurzbeschreibung für jede Informationseinheit angegeben werden kann.
 3. Wie schon zuvor mehrfach beleuchtet (vgl. Kap. 6 und Kap. 7), sollen die Inhalte des IB-Domänenmodells, die Informationseinheiten, auf die Interessen von Konsumenten effizient abbildbar sein. Aus diesem Grund können Informationseinheiten mit Kategorien klassifiziert werden. Die Kategorien sind durch die Kategorietypen determiniert, die eine Klassifikationsbeziehung zum entsprechenden Konzept haben. Für jede Beschreibungsdimension, die durch die Kategorietypen vorgegeben werden, muss mindestens eine Kategorie aus der entsprechenden Kategorienhierarchie ausgewählt werden.
- **Category** und **CategoryType** unterscheiden sich von **NamedContainer** und **NamedContainerType** dadurch, dass sie Konzepte bzw. Informationseinheiten klassifizieren können (s.o.). Außerdem können Kategorien zu Hierarchien angeordnet werden. Dabei ist zu beachten, dass eine Kategorie mehrere Vaterkategorien haben kann. Alle Kategorien sind unter einer (eindeutigen) Wurzelkategorie subsumiert. Der Kategorietyp hält eine Referenz auf diese Wurzelkategorie.

Anhand der Klassenhierarchie lassen sich die Unterschiede zu herkömmlichen Systemen, die Beschreibungslogik implementieren, sichtbar machen. `ConceptBase` (vgl. [JGJS95]) z.B. teilt sich in eine Typen- und eine Instanzenebene auf. In der Typenebene kann zwischen Konzepten und Relationen zwischen Konzepten unterschieden werden. Ein Konzept hat einen Namen und kann zu anderen Konzepten über Relationen in Beziehung gesetzt werden. Außerdem können Konzepte zu einer Subsumptionshierarchie angeordnet werden.

Das IB-Domänenmodell von `Broker's Lounge` kennt keine expliziten Relationen. Stattdessen werden die Relationen in den `Containertypen` selbst verwaltet. `Containertypen` können ebenfalls zu einer Subsumptionshierarchie angeordnet werden. Da `Containertypen` in `Broker's Lounge` keinen Namen haben, sind sie nicht gleichwertig zu den Konzepten von `ConceptBase`. Anstelle von `Container` kann man `NamedContainer` mit den Konzepten aus `ConceptBase` vergleichen.

Konzepte, Informationseinheiten, Kategorietypen, Kategorien sowie Attributtypen und Attribute markieren den Unterschied zu `ConceptBase`. Prinzipiell könnte jede Domäne, die in einem IB-Domänenmodell von `Broker's Lounge` beschrieben wurde, auch in ein Modell von `ConceptBase` überführt werden, indem man nur noch `NamedContainer` verwendet.

Während Modelle in `ConceptBase` eine Domäne möglichst wertfrei und objektiv zu beschreiben versuchen, ist die Aufteilung in Attributtypen, Konzepte und Kategorietypen nicht wertfrei, sondern nimmt eine Gewichtung der Relevanz von einzelnen Aspekten innerhalb des IB-Domänenmodells für die verschiedenen Konsumenten vor. Bei der Strukturierung der Domäne muss die Relevanz für die Konsumenten antizipiert werden, da die Inhalte des IB-Domänenmodells später auf die Interessen der Konsumenten abgebildet werden. Dabei kommt den Konzepten die größte Bedeutung zu, da Konzepte die Informationstypen beschreiben, mit denen der Broker handelt und für die sich die Konsumenten interessieren. Durch Kategorietypen und Kategorien werden Eigenschaften von Konzepten bzw. von den Informationstypen beschrieben. Es sollten die Eigenschaften ausgewählt werden, die von den Konsumenten benutzt werden, um die entsprechenden Konzepte zu klassifizieren und zu unterscheiden. Eine Kategorienhierarchie beschreibt ein Konzept unter einem bestimmten Blickwinkel. Dieser Blickwinkel sollte entlang der Kategorienhierarchie beibehalten werden. Wenn Konzepte unter mehreren Blickwinkeln klassifiziert werden, ist dies durch mehrere voneinander unabhängige Kategoriehierarchien abzubilden. Eine Kategorienhierarchie kann auch als Ontologie zur Beschreibung von Konzepten bzw. der durch den Kategorietypen beschriebenen Eigenschaft aufgefasst werden. Ggf. kann eine Kategorienhierarchie zur Fortsetzung der von den Konzepten aufgespannten Konzepthierarchie benutzt werden. Quantitative Eigenschaften können als Kategorien beschrieben werden, indem verschiedene Wertebereiche zusammengefasst und diese als Kategorien aufgefasst werden (vgl. Kap. 6). Bei Kategorien handelt

es sich um eine besondere Art von Attributen:

- Die Werte für einen Kategorietypen sind durch die Kategorienhierarchie (für diesen Kategorietypen) vorgegeben. Es handelt sich bei den Kategorietypen somit um Aufzählungstypen.
- Die Kategorietypen und Kategorien werden zur Abbildung von Interessen benutzt, indem über die verschiedenen Kategorien ein Overlay-Modell gelegt wird, das für jede Kategorie die Relevanz für den Benutzer angibt (vgl. Kap. 7 und Kap. 8.7).

(Einfache) Attribute unterscheiden sich von Kategorien bzw. Kategorietypen dadurch, dass die in ihnen beschriebenen Informationen für die Konsumenten von Bedeutung sind, jedoch nicht (wie Kategorien) zur Unterscheidung und Bewertung von Informationseinheiten eingesetzt werden.

Der Strukturierung der Domäne kommt neben der Abbildung der Domäne auch die Aufgabe zu, die verschiedenen Bestandteile des IB-Domänenmodells in Bezug auf die Relevanz für die Konsumenten zu beurteilen. Die Domäne muss aus der Sicht der Konsumenten modelliert werden. Die Zuordnung zu Konzepten, Attribut- und Kategorietypen resultiert in einer unterschiedlichen Behandlung in der Filterung und der adaptiven Unterstützungsleistung.

Bei der Kategorienadministration müssen die Kategoriehierarchien für die verschiedenen Kategorietypen mit Leben gefüllt werden. D.h. die Konzepte müssen jeweils unter verschiedenen Gesichtspunkten klassifiziert werden. Da die Kategoriehierarchien vorrangig zur Abbildung von Interessen benutzt werden und Konsumenten ihre Interessen auch selbst angeben können sollten, müssen die Kategorien die Sprache der Konsumenten verwenden und die Kategoriehierarchie sollte nur den Ausschnitt beschreiben, der für die Konsumenten von Bedeutung ist.

8.4 Domänenmodelladministration

Das IB-Domänenmodell ist das zentrale Modell in Broker's Lounge. Die Qualität des IB-Domänenmodells entscheidet über die Qualität der Dienstleistungen, die angeboten werden können, da das IB-Domänenmodell in fast alle Aufgaben (vgl. Abb. 8.1) einbezogen wird. Um zu beurteilen, ob ein Dokument für die modellierte Domäne relevant ist, wird es mit dem IB-Domänenmodell bzw. den Inhalten des IB-Domänenmodells verglichen. Die Inhalte des IB-Domänenmodells, die Informationseinheiten, werden im Kategorisierungsprozess so aufbereitet, dass die Interessen von Konsumenten effizient darauf abgebildet werden können.

Damit Informationseinheiten in das IB-Domänenmodell eingefügt werden können, muss ein entsprechendes IB-Domänenmodell existieren. D.h., die grundlegenden Strukturen des IB-Domänenmodells müssen vorhanden sein, bevor mit

der Kategorisierung begonnen werden kann. Das IB-Domänenmodell von Broker's Lounge besteht im Gegensatz zu beschreibungslogischen Ansätzen, die nur zwischen einer Strukturierungs- und Inhaltsebene unterscheiden,

- aus einer Strukturierungsebene, die aus Konzepten und Kategorietypen aufgebaut ist,
- einer zusätzlichen Strukturierungsebene aus den verschiedenen Kategoriehierarchien, die für die Abbildung von Interessen benutzt werden und
- aus einer Inhaltsebene, die aus den verschiedenen Informationseinheiten besteht, mit denen der Broker handelt (vgl. Kap. 6).

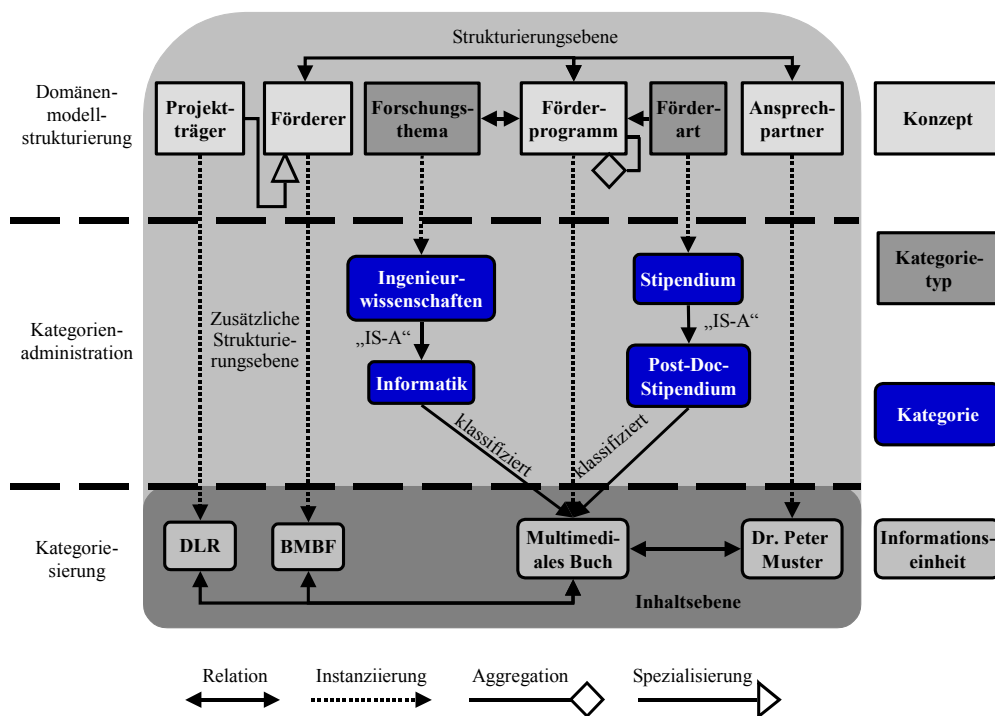


Abbildung 8.4: Domänenmodellaufgaben

Aus den drei Ebenen des IB-Domänenmodells ergeben sich drei unterschiedliche Aufgaben (vgl. Abb. 8.4):

1. Innerhalb der *Domänenmodellstrukturierung* wird die Domäne mit Konzepten und Kategorietypen strukturiert.
2. Die Kategorietypen müssen in der zusätzlichen Strukturierungsebene mit Kategoriehierarchien gefüllt werden. Der Aufbau und die Pflege dieser Hierarchien wird *Kategorienadministration* genannt.

3. Das Füllen des IB-Domänenmodells mit konkreten Informationseinheiten ist Gegenstand der *Kategorisierung*.

Domänenmodellstrukturierung und Kategorienadministration sind Infrastrukturaufgaben, die vor allem zu Beginn der Arbeit mit Broker's Lounge von Bedeutung sind. Änderungen der Domänenmodellstruktur können unmittelbare Auswirkungen auf die Kategorienadministration haben, wenn z.B. ein neuer Kategorietyt eingeführt wurde, der dann mit einer Kategorienhierarchie gefüllt werden muss. Die Kategorisierung ist ebenfalls von diesen Änderungen betroffen, da die veränderte Domänenmodellstruktur in der Inhaltsebene für konkrete Informationseinheiten mit Leben gefüllt werden muss. Änderungen in der zusätzlichen Strukturierungsebene, z.B. das Einfügen einer neuen Kategorie, führen dazu, dass bei der Kategorisierung diese veränderte Kategorienhierarchie verwendet werden kann, um die Belegung des entsprechenden Kategorietyps für eine Informationseinheit festzulegen.

Änderungen der Domänenmodellstruktur sind dann sinnvoll, wenn der Broker sein Angebot erweitern oder verändern will. Wenn z.B. in der Domäne Forschungsförderung neben Informationen über Förderer, Förderprogramme und Ansprechpartner auch Informationen über bereits geförderte Projekte angeboten werden sollen, wird in das IB-Domänenmodell ein neues Konzept „Projekt“ eingeführt. Wenn sich ein neues Forschungsgebiet entwickelt, das gefördert wird, kann in der Domäne Forschungsförderung darauf reagiert werden, indem eine neue Kategorie für dieses Forschungsgebiet (in der Forschungsgebietehierarchie) eingeführt wird.

Dies markiert einen gravierenden Unterschied zu Protégé (vgl. [EFHM99]). Protégé und mit ihm viele weitere Ansätze, die von Protégé inspiriert wurden, werden benutzt, um eine Domänenmodellstruktur zu definieren. Diese Struktur wird dann zur Erzeugung eines Editors für die Instanzen der definierten Domänenmodellstruktur benutzt. Wenn die Domänenmodellstruktur geändert werden soll, muss der gesamte Prozess wiederholt werden.

Broker's Lounge dagegen erlaubt die Änderung der Strukturebenen⁴ und die Bearbeitung von Instanzen in **einer** Schnittstelle. Alle drei verschiedenen Domänenmodellaufgaben (siehe oben, Abb. 8.4) können mit ein und derselben Schnittstelle ausgeführt werden, die in Abbildung 8.5 dargestellt ist. Die drei Aufgaben beinhalten einander: Die Aufgabe der Kategorienadministration, beinhaltet auch gleichzeitig die Aufgabe der Kategorisierung. D.h., ein Benutzer, der das Recht hat, Kategorien zu administrieren, hat gleichzeitig auch alle Rechte, die ein Kategorisierer hat. Ebenso beinhaltet das Recht der Domänenmodellstrukturierung auch das Recht zur Kategorienadministration und zum Kategorisieren.

⁴Im Gegensatz zu beschreibungslogischen Ansätzen werden in der Broker's Lounge zwei Ebenen zur Strukturierung von Inhalten benutzt.

Name	Syn...	Homepage	Teil...	Ob...	Forschun...	Que...	De...	Ins...	Re...	An...	Fö...	Do...
BauA - Formul...	BA...	http://www...										
BauA - Eiprob...	BA...	http://www...										
BMBF - "Gestal...	B...	http://www...										
BMBF - "IT-For...	B...	http://www...										
BMBF - "Kontin...	B...	http://www...										
BMBF - ATMin...	B...	http://www...										
BMBF - Gradia...	B...	http://www...										
BMBF - Inform...	B...	http://www...										
BMBF - Kompo...	B...	http://www...										
BMBF - Mikros...	B...	http://www...										
BMBF - Multi-m...	B...	http://www...										
BMBF - Neuart...	B...	http://www...										
BMBF - Neue ...	B...	http://www...										
BMBF - Schlüs...	B...	http://www...										
BMBF - Smart ...	B...	http://www...										
BMBF - Syste...	B...	http://www...										
BMBF-Display...	B...	http://www...										
BMBF-Mikrore...	B...	http://www...										
BMWI - MobilM...	B...	http://www...										
CDG - Praxistr...	C...	http://www...										
DFG - Aktionsp...	DF...	http://www...										
DFG-SP-Optis...	DF...	http://www...										
DFG-SP-Quant...	DF...	http://www...										
DFG-SP1140 ...	DF...	http://www...										
EU - GROWTH...	E...	http://www...										
EU - IST Benut...	E...	http://www...										
EU - IST Forc...	E...	http://www...										
EU - IST Gener...	E...	http://www...										
EU - IST Letak...	E...	http://www...										
EU - IST Letak...	E...	http://www...										
EU - IST Letak...	E...	http://www...										

Abbildung 8.5: Schnittstelle zur Domänenmodelladministration

Eine Trennung der verschiedenen Rechte ist sinnvoll, da Änderungen der Domänenmodellstruktur und der Kategorien weitreichende Folgen für die Qualität des IB-Domänenmodells haben können und deswegen nur von geübten Benutzern

ausgeführt werden sollten.

Die Rechte der verschiedenen Rollen lassen sich an Abbildung 8.5 gut verdeutlichen:

1. Kategorisierer haben die Möglichkeit, die konkreten Inhalte des IB-Domänenmodells zu ändern. Dies entspricht in der Schnittstelle zur Domänenmodelladministration den Einträgen in der Tabelle auf der rechten Seite. Diese Tabelle zeigt immer die Informationseinheiten an, die zur momentan ausgewählten Kategorie oder dem momentan ausgewählten Konzept gehören. Dies sind alle Informationseinheiten, die mit der ausgewählten Kategorie oder mit einer ihrer Unterkategorien klassifiziert wurden. Wenn ein Konzept ausgewählt wurde, werden alle Informationseinheiten eines Konzeptes aufgelistet.
2. Der Kategorienadministrator kann die Kategorien administrieren, indem er den Kategorienbaum auf der linken Seite editiert. Dort kann er Kategorien hinzufügen, löschen oder umbenennen.
3. Die Domänenmodellstruktur kann durch das Einfügen, Umbenennen oder Löschen von Konzepten, Kategorietypen und Attributtypen geschehen. Außerdem können Relationen zwischen Konzepten definiert werden. Wurde ein neues Konzept eingefügt, führt dies dazu, dass im Baum ein neuer Konzeptknoten⁵ angezeigt wird. Wenn ein neues Attribut oder eine neue Relation zu einem Konzept hinzugefügt wurde, schlägt sich dies in einer neuen Spalte für die Tabellen dieses Konzeptes nieder. Handelt es sich bei der neu hinzugefügten Relation um einen Kategorietypen, bedeutet dies, dass dem Konzept eine neue Beschreibungsdimension hinzugefügt wurde. Dies führt im Baum dazu, dass für den entsprechenden Konzeptknoten eine neue Kategorienhierarchie eingefügt wird. Wenn es sich um einen neuen Kategorietypt handelt, wird nur ein Wurzelknoten für die neue Kategorienhierarchie eingefügt. Wird ein Kategorietypt mit einer existierenden Hierarchie wiederverwendet, wird die entsprechende Kategorienhierarchie unter den entsprechenden Konzeptknoten kopiert.

Um jederzeit eine durchgängige Filterung mit Kategorien gewährleisten zu können, müssen alle Informationseinheiten in ihren jeweiligen Kategoriedimensionen mit mindestens einer Kategorie klassifiziert sein. Damit kein inkonsistenter Zustand entsteht, wenn eine neue Kategoriedimension (ein neuer Kategorietypt) eingeführt wurde, werden alle Informationseinheiten

⁵Konzepte werden im Baum durch ein kleines grünes Klassensymbol visualisiert. In Abb. 8.5 sind mit Förderern, Personen, Dokumenten und Förderprogrammen vier Konzepte vorhanden.

eines Konzeptes mit der Wurzelkategorie der neu eingeführten Kategorienhierarchie klassifiziert. Diese Klassifikationen können von einem Kategorisierer jederzeit wieder geändert werden.

EU - GROWTH Förderung der Forschungsinfrastruktur editieren

Förderprogramm bearbeiten

Name: EU - GROWTH Förderung der Forschungsinfrastruktur

erzeugt am: Dienstag, 26. März 2002

letzte Änderung: Dienstag, 26. März 2002

☒ Datum der letzten Änderung aktualisieren

Förderprogramm

- Homepage
- Bild
- Synonyme
- Teilprogramme
- Oberprogramme
- Forschungsthemen
- Querschnittsthemen
- Deadline
- Förderarten
- Institutionsarten
- Regionen
- Ansprechpartner
- Förderer
- Dokument
- Sonstige Attribute

Gewichtung: 0 25 50 75 100

Haltbarkeitsdatum:

Neues Datum: 14 März 2002

Kurzbeschreibung:

Konkrete Maßnahmen sind erforderlich, um die optimale Nutzung der vorhandenen Forschungsinfrastrukturen und Großanlagen in enger Abstimmung mit den Leitaktionen zu fördern. Die gemeinsame Nutzung von Ressourcen sowie der grenzüberschreitende Zugang zu industriellen Forschungszentren, Testeinrichtungen für den Verkehr, Windkanälen, Rechenzentren der industriellen Forschung oder Prüfeinrichtungen für Werkstoffe und Bauteile ist Voraussetzung für eine bessere Nutzung der neuen Technologien, die einen Zugang zu aufwendigen Infrastrukturen und Ausrüstungen verlangen. Durch Vernetzung bestehender Einrichtungen muß Doppelarbeit weitmöglichst vermieden, die Komplementarität gesteigert und die Kompatibilität beim Datenaustausch gewährleistet werden. Da der Entwicklungsstand der nationalen Einrichtungen in der Gemeinschaft unterschiedlich ist, muß die Zusammenarbeit auch in Bereichen wie Konformitätsbewertung, Ausbildung und Know-how-Transfer verstärkt werden. Bei der Unterstützung der Forschungsinfrastrukturen werden folgende Ziele verfolgt:

Optimale Nutzung mittlerer und großer Forschungseinrichtungen an unterschiedlichen Standorten; umgehende Weitergabe vorhandener oder ergänzender FTE-Ergebnisse an die industriellen Bedarfsträger und Verbesserung der Interoperabilität gemeinsamer Protokolle und des Datenaustauschs.

Maßnahmen:

- Virtuelle Institute

Speichern Abbrechen

Abbildung 8.6: Eingabe von Kurzbeschreibung und Gewichtung von Informationseinheiten

8.4.1 Kategorisierung

Die Kategorisierung dient dem Füllen des IB-Domänenmodells mit konkreten Informationseinheiten (vgl. Abb. 8.4). Da das IB-Domänenmodell auch zur Identifikation von relevanten Dokumenten benutzt wird, erwächst der Kategorisierung eine weitere Aufgabe darin, die Inhalte des IB-Domänenmodells so aufzubereiten, dass sie für die Identifikation von relevanten Dokumenten benutzt werden können. Aus diesem Grund können Informationseinheiten (vom Kategorisierer)

mit Synonymen und einer Gewichtung versehen werden. Diese Synonyme gehen in die Bewertung von Dokumenten ein (vgl. Abb. 8.1). Neben der Gewichtung einzelner Informationseinheiten, können auch Kategorien mit einer Gewichtung versehen werden. Die Gewichtung der Kategorien kann benutzt werden, um die verschiedenen Bestandteile eines IB-Domänenmodells unterschiedlich zu bewerten, während die Gewichtung einzelner Informationseinheiten der Hervorhebung einzelner Informationseinheiten gegenüber anderen, ähnlichen Informationseinheiten dient. Die Verwendung der verschiedenen Gewichtungen für die Dokumentenbewertung wird in Kapitel 8.6 erläutert.

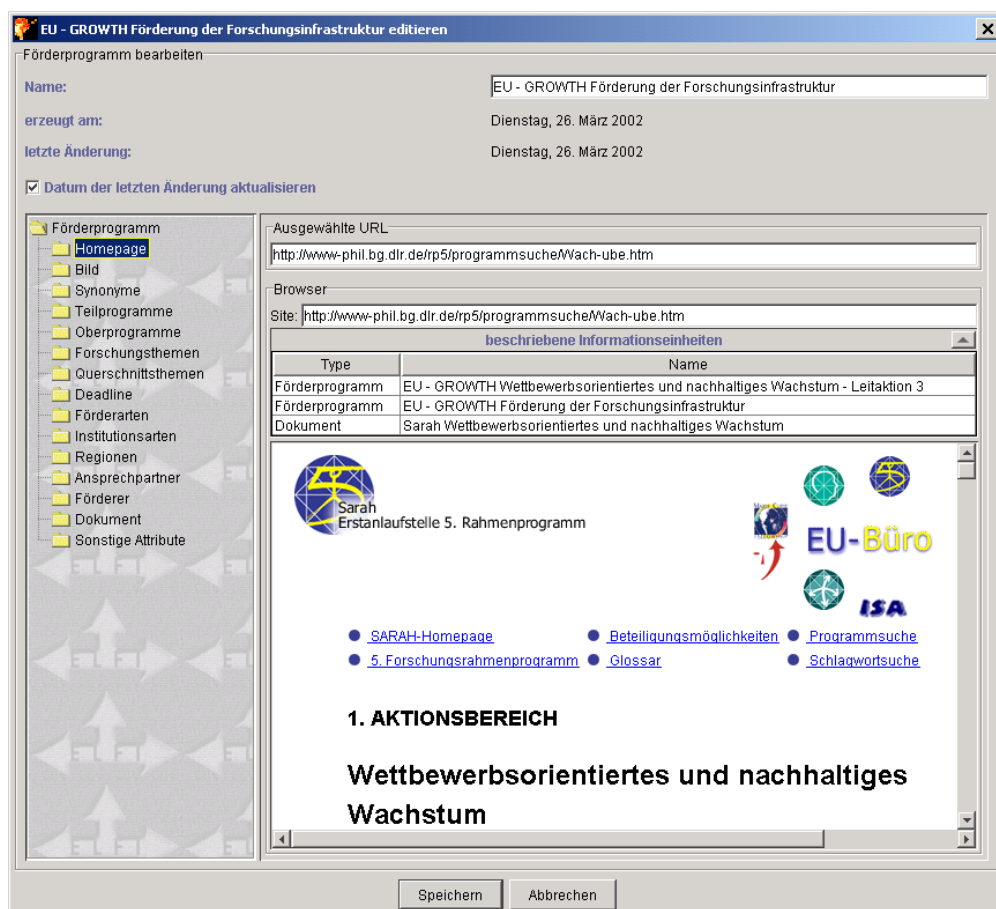


Abbildung 8.7: Eingabe der Originalquelle

Das IB-Domänenmodell dient der Beschreibung von Inhalten, die für die Konsumenten von Bedeutung sind. Aufgabe des Brokers ist es, seinen Kunden (den Konsumenten) einen Überblick über die für sie relevanten Informationen zu geben. Dazu fertigt der Broker zu jeder Informationseinheit eine Kurzbeschreibung an. Um die Authentizität zu wahren, werden alle Informationseinheiten mit ei-

nem Verweis auf die Originalquelle versehen, die von dieser Informationseinheit beschrieben wird. Die Abbildungen 8.6 - 8.9 zeigen die Eingabeschnittstelle für Informationseinheiten. Die Schnittstelle ist zweigeteilt:

- Die linke Seite ist durch die Struktur des entsprechenden Konzeptes determiniert. Neben einigen Standardeinträgen, die für jedes Konzept existieren, werden alle Relationen zu anderen Konzepten und Kategorietypen angezeigt. Alle (einfachen) Attribute werden unter dem Punkt „Sonstige Attribute“ zusammengefasst.
- Durch die Auswahl auf der linken Seite wird die rechte Seite determiniert. In Abbildung 8.6 ist z.B. die Eingabe von Kurzbeschreibung und Gewichtung abgebildet.

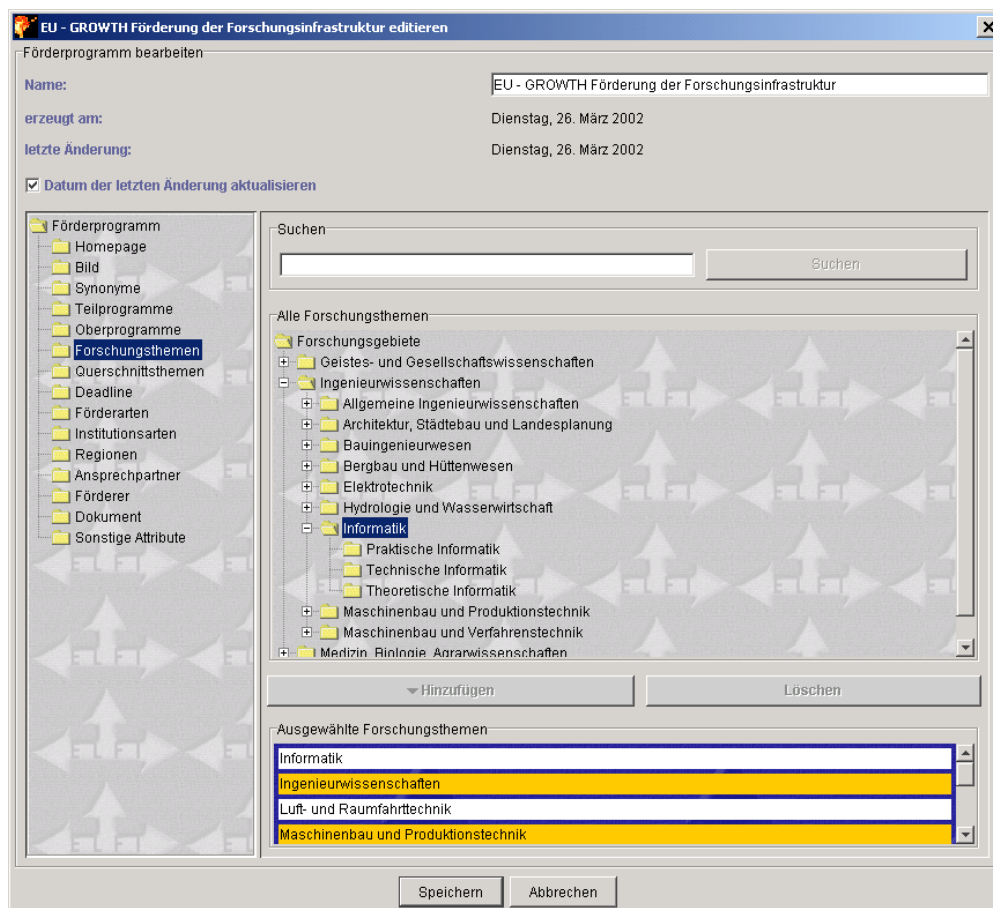


Abbildung 8.8: Eingabe der Klassifikation von Informationseinheiten

Es gilt, die verschiedenen Relationen und Attribute sukzessive abzarbeiten. Unter dem Punkt „Synonyme“ können die Synonyme für diese Informationsein-

heit editiert werden. Unter dem Punkt „Homepage“ verbirgt sich die Referenz auf die Originalquelle (vgl. Abb. 8.7). Abbildung 8.8 zeigt die Eingabe der Klassifikation einer Informationseinheit innerhalb einer Kategoriedimension „Forschungsthemen“. Im oberen Teil der rechten Seite ist die Kategorienhierarchie zu der entsprechenden Kategoriedimension abgebildet, aus der die konkreten Klassifikationen für diese Informationseinheit ausgewählt werden können. Für jede Kategoriedimension (jeden Kategorietypen) muss mindestens eine Kategorie ausgewählt werden, damit später eine effiziente Filterung auf Basis der Klassifikationen durchgeführt werden kann. Sollte keine der angezeigten Kategorien zur Klassifikation geeignet erscheinen, so muss entweder eine entsprechende, neue Kategorie eingeführt werden oder diese Informationseinheit muss mit der Wurzelkategorie klassifiziert werden, da eine differenziertere Beschreibung nicht möglich ist. Außerdem wird durch die Schnittstelle sichergestellt, dass nicht zugleich Vater- und Kinderkategorien zur Klassifikation benutzt werden, da dies Nachteile für die Filterung hat.

Die Eingabe von Relationen zu anderen Informationseinheiten (vgl. Abb. 8.9) ähnelt der Eingabe von Klassifikationen: In der oberen Hälfte sind alle Informationseinheiten des verwandten Konzeptes aufgelistet. Aus dieser Liste können die passenden Informationseinheiten ausgewählt werden, zu denen Relationen aufgebaut werden sollen. Sollte die verwandte Informationseinheit, zu der eine Relation aufgebaut werden soll, noch nicht definiert sein und deswegen nicht in der oberen Liste enthalten sein, so kann an dieser Stelle über den Button „Neu“ diese neue Informationseinheit definiert werden. Im Gegensatz zu den Klassifikationen sind die Relationen zu anderen Informationseinheiten fakultativer Natur. Die Relationen werden nicht direkt zur Filterung genutzt. Eine Informationseinheit gilt als definiert und kann in das IB-Domänenmodell eingefügt werden, sobald sie einen Namen hat und in all ihren Kategoriedimensionen mit mindestens einer Kategorie klassifiziert wurde.

8.5 Quellenbeobachtung und Webrobots

Der Retrievalprozess in Broker's Lounge wird durch den Einsatz von automatisierten Webrobots unterstützt. Diese Webrobots untersuchen automatisch einen Ausschnitt des WWW auf veränderte oder neue Dokumente. Die Ergebnisse der Robots werden sowohl im Quellenmodell als auch im Dokumentenindex abgelegt. Im Quellenmodell wird die Struktur der untersuchten Quellen verwaltet, während der Volltext der behandelten Dokumente im Dokumentenindex abgelegt wird. Der Dokumentenindex kann benutzt werden, um Änderungen in einzelnen Dokumenten zu identifizieren und sichtbar zu machen. Nachdem die Dokumente in den Dokumentenindex kopiert wurden, werden sie der inhaltlichen Analyse zugeführt.

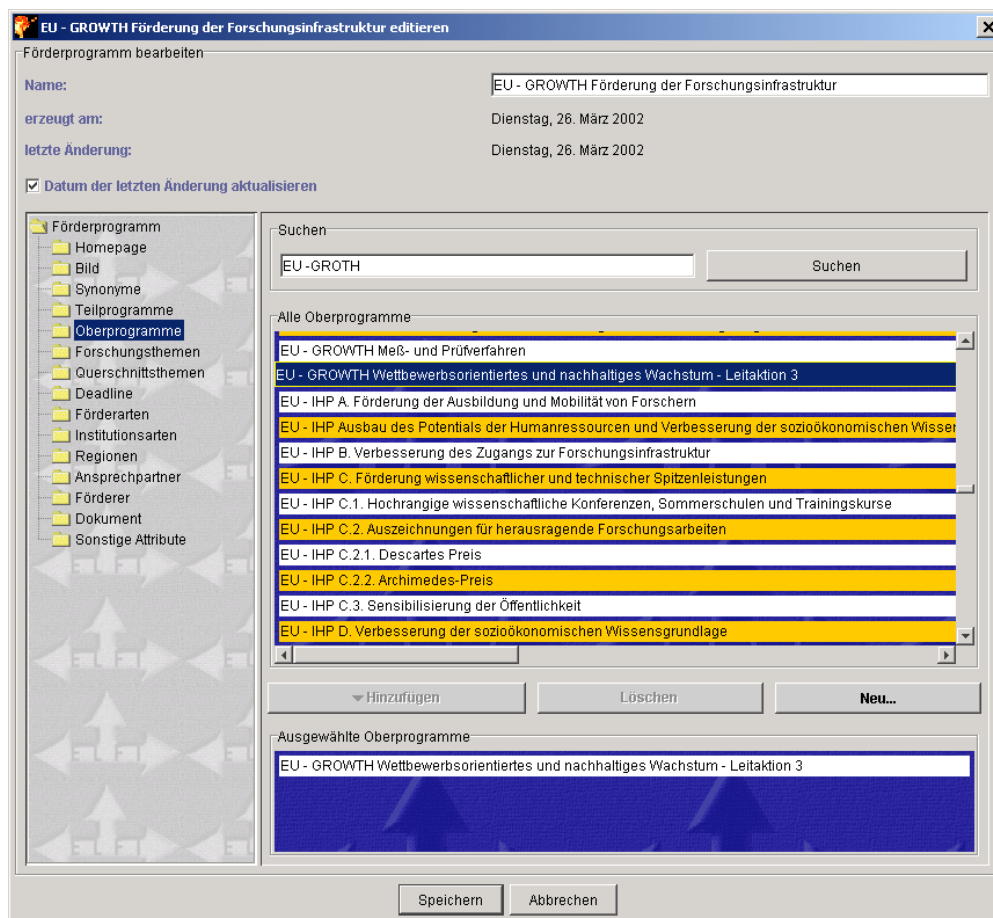


Abbildung 8.9: Eingabe der Relation zu anderen Informationseinheiten

Die automatische Quellenbeobachtung setzt die Konfiguration der Webrobots voraus. Die Webrobots werden dadurch auf bestimmte Bereiche (z.B. einzelne WWW-Sites) beschränkt, um so sicherzustellen, dass nur relevante Bereiche untersucht und irrelevante Bereiche ignoriert werden. Dadurch kann die Qualität der Ergebnisse verbessert werden.

Abbildung 8.10 zeigt die Schnittstelle zur Administration und Konfiguration der Webrobots. Zur Organisation der Robots und der von ihnen bearbeiteten Quellen können sie verschiedenen Quellentypen zugeordnet werden. Die Quellentypen sind hierarchisch organisiert. Die Schnittstelle ist zweigeteilt:

- Die linke Seite zeigt die verschiedenen Quellentypen. Die Hierarchie der Quellentypen kann dort in ähnlicher Art und Weise wie die Kategorien des IB-Domänenmodells bearbeitet werden.
- Durch die Auswahl eines Quellentyps wird die Anzeige der rechten Sei-

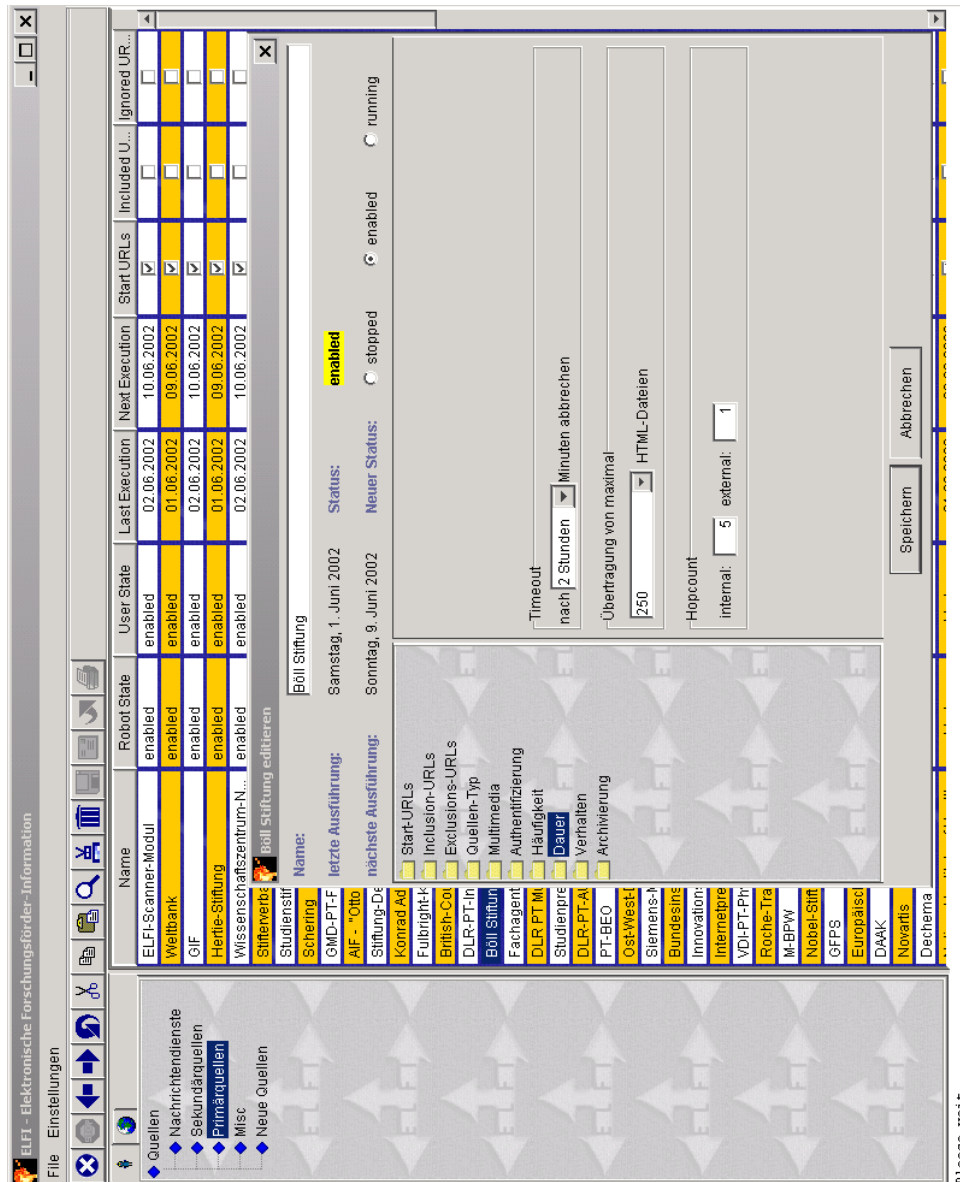


Abbildung 8.10: Schnittstelle für die Robotadministration

te determiniert. Dort wird eine Tabelle mit allen Robots des selektierten Quellentyps angezeigt. Dort können die Robots gestartet, gestoppt oder in Bereitschaft versetzt werden.

Einzelne Robots können durch die Auswahl aus der Robotliste editiert werden (vgl. Abb. 8.10). Die Administration unterteilt sich in folgende Bestandteile:

- *Start-URLs* sind die Einstiegspunkte des Robots. Dort starten die Robots ih-

re Suche. Robots, bei denen keine Start-URL definiert wurde, sind ungültig.

- *Inclusion-URLs* beschreiben Verzeichnisse, die in die Bearbeitung durch den Robot einbezogen werden sollen.
- *Exclusion-URLs* enthalten URLs, die vom Robot explizit nicht bearbeitet werden sollen. Dadurch können irrelevante Bereiche von der Bearbeitung ausgeschlossen werden.
- Der Robot muss eindeutig einem *Quellentyp* zugeordnet werden. Die Ergebnisse des Robots werden ebenfalls diesem Quellentyp zugeordnet. Dadurch können die bearbeiteten Dokumente nach Quellentypen unterschieden werden.
- Die Robots legen ein eigenes Dokumentenarchiv⁶ an. Für jeden Robot muss unter dem Punkt *Multimedia* entschieden werden, ob neben den reinen Textdokumenten auch die dort referenzierten Multimedia-Inhalte in das eigene Archiv kopiert werden sollen.
- Für Quellen, die eine Authentifizierung bedingen, können unter dem Punkt *Authentifizierung* der Login-Namen und das zugehörige Kennwort angegeben werden, damit der Robot sich bei dieser Quelle authentifizieren kann.
- Damit Quellen regelmäßig beobachtet werden können, kann man unter dem Punkt *Häufigkeit* die Periodizität der Robots angeben.
- Unter dem Punkt *Dauer*, der in Abbildung 8.10 dargestellt ist, können verschiedene Abbruchkriterien definiert werden. Der Robot stellt seine Arbeit nach einem festzulegenden Zeitraum ein. Wenn er zuvor einen bestimmten Umfang an Dokumenten bearbeitet hat, bricht er ebenfalls ab. Der *Hopcount* bezeichnet die Tiefe des zu bearbeitenden Dokumentenraums. Robots bearbeiten Webquellen ausgehend von den Start-URLs in Breitensuche. Durch den Hopcount kann die Tiefe des Suchraums festgelegt werden. Die Unterscheidung in internen und externen Hopcount dient der Einschränkung des Suchraums. Der interne Hopcount bezeichnet die Tiefe der Suche innerhalb der bearbeiteten (Web-)Quelle, während der Hopcount die maximale Tiefe des Suchraums außerhalb der bearbeiteten Quelle widerspiegelt. Ein externer Hopcount „1“ bedeutet, dass alle externen Dokumente⁷, von denen innerhalb der Quelle verwiesen wird, vom Robot bearbeitet werden,

⁶Die Begriffe „Dokumentenarchiv“ und „Dokumentenindex“ werden in dieser Arbeit synonym verwendet.

⁷Ein Dokument wird als externes Dokument beschrieben, wenn dieses Dokument nicht auf dem durch den Robot bearbeiteten Webserver (der Quelle) liegt.

jedoch die Dokumente, auf die innerhalb des externen Dokumentes verwiesen wird, nicht mehr bearbeitet werden. Alle internen Dokumente werden, wie oben beschrieben, dem Quellentyp des Robots zugeordnet. Alle externen Quellen werden dem Quellentyp „Neue Quellen“ zugewiesen. Diese Quellen sind Kandidaten für die Definition neuer Robots. In Kapitel 8.6 wird darauf eingegangen.

- Unter dem Punkt *Verhalten* kann gesteuert werden, ob und wie ein Robot die Regeln der Netiquette (vgl. [She94]) beachtet.
- Der abschließende Punkt *Archivierung* steuert die Interaktion mit dem Dokumentenarchiv. Da die Kapazität des Archivs nicht unendlich ist, können Einschränkungen für die Ablage im Dokumentenarchiv definiert werden. So können z.B. nur solche Dokumente in das Archiv kopiert werden, die bei der inhaltlichen Analyse einen Mindestwert bekommen haben. Außerdem können Dokumente mit einer zeitlichen Einschränkung versehen werden. Dies bedeutet, dass Dokumente nach einem festgelegten Zeitraum aus dem Dokumentenarchiv entfernt werden.

Das Quellenmodell beinhaltet sowohl die Konfigurationen der verschiedenen Robots als auch die Ergebnisse der Robots. Dies sind die bearbeiteten Dokumente und die Struktur der beobachteten Quellen. Während Abbildung 8.10 die Sicht auf die verschiedenen Robots und ihre Konfiguration illustriert, zeigt Abbildung 8.11 die Struktur der bearbeiteten Quellen und die gefundenen Dokumente. Die Quellen sind ebenfalls nach Quellentypen geordnet. Für die Quellen wird jeweils die Verzeichnisstruktur⁸ angezeigt. Wenn ein Verzeichnis ausgewählt wurde, werden auf der rechten Seite alle Dokumente angezeigt, die im Dokumentenarchiv vorhanden sind und in diesem Verzeichnis gefunden wurden. Mit Hilfe dieser Sicht kann der Quellenadministrator die Struktur der bearbeiteten Quellen analysieren, um die Konfiguration der Robots anzupassen, indem z.B. Verzeichnisse, die keine relevanten Verzeichnisse enthalten, in die Liste der Exclusion-URLs eingetragen werden.

8.6 Quellenbewertung und Kategorisierung

Damit der Broker immer auf dem neusten Stand ist, muss er alle Quellen regelmäßig beobachten und sein IB-Domänenmodell ggf. entsprechend anpassen. Die Robots liefern alle neuen oder veränderten Dokumente, die sie finden. Dadurch wird der Broker entlastet, weil er die Quellen nicht mehr selbst beobachten muss

⁸Die Verzeichnisstruktur wird nur soweit angezeigt, wie sie dem System bekannt ist.

Diese Funktion wird in Broker's Lounge von der inhaltlichen Analyse ausgefüllt. Dabei werden die neuen und veränderten Dokumente mit den Inhalten des IB-Domänenmodells verglichen und gewichtet. Die Inhalte des IB-Domänenmodells werden durch die im IB-Domänenmodell vorhandenen Synonyme repräsentiert. Es handelt sich dabei um Synonyme von Kategorien und Informationseinheiten⁹. Es ist zu beachten, dass es sich bei den Synonymen nicht notwendigerweise um einzelne Worte handeln muss. Es kann sich dabei auch um Wortgruppen handeln. Ein Beispiel dafür ist z.B. „Deutsche Forschungsgemeinschaft“ als Synonym für „DFG“.

Wenn in einem Dokument einzelne Synonyme gefunden wurden, werden sie markiert. Die markierten Treffer eines Dokumentes bilden die Grundlage für die Bewertung der Dokumente. Für alle Synonyme von Kategorien wird zum Domänenwert des Dokumentes die Gewichtung der entsprechenden Kategorie hinzu addiert. Für Synonyme von Informationseinheiten wird die Gewichtung der entsprechenden Informationseinheit hinzugefügt. Außerdem wird die Gewichtung aller Kategorien, mit denen diese Informationseinheit klassifiziert wurde, hinzugenommen. Um längere Dokumente nicht gegenüber kürzeren Dokumenten zu bevorzugen, wird der so errechnete Domänenwert auf die Länge des Dokumentes logarithmisch normalisiert. Um eine Vergleichbarkeit der daraus resultierenden Werte herbeizuführen, wird der Domänenwert abschließend auf einen Bereich von null bis hundert Punkten normalisiert.

Der errechnete Domänenwert kann dazu benutzt werden, irrelevante Dokumente, die keine Treffer aufweisen und deswegen einen Domänenwert von null Punkten haben, von der weiteren Bearbeitung durch den Broker auszuschließen, indem sie nicht in das Dokumentenarchiv gelangen bzw. direkt wieder von dort gelöscht werden. Der Domänenwert kann außerdem dazu benutzt werden, die gefundenen Dokumente nach ihrer Relevanz für die Domäne zu sortieren. Dies ermöglicht es dem Broker, die wichtigsten Dokumente zuerst zu bearbeiten. Abbildung 8.12 zeigt den Dokumentenbrowser, mit dem man die Dokumente im Dokumentenarchiv anschauen kann. Die Schnittstelle ist, wie die anderen Schnittstellen auch¹⁰, in eine linke und eine rechte Seite unterteilt. Die linke Seite wird benutzt, um Anfragen an das Dokumentenarchiv zu generieren. Folgende Anfragekriterien können benutzt werden:

- Volltextanfragen: Es werden nur Dokumente angezeigt, in denen der angegebene Begriff vorkommt. Zusätzlich werden alle Dokumente mit einem Anfragewert (Query Score) versehen, der angibt, wie relevant ein Doku-

⁹NamedContainer kommen in der Praxis selten vor. Sie wurden in der Klassenhierarchie hauptsächlich als gemeinsame Superklasse von Category und InformationItem eingeführt, damit beide einen Namen und damit assoziierte Synonyme haben.

¹⁰Eine Ausnahme bildet hier die Schnittstelle für die Benutzeradministration (vgl. Abb. 8.2).

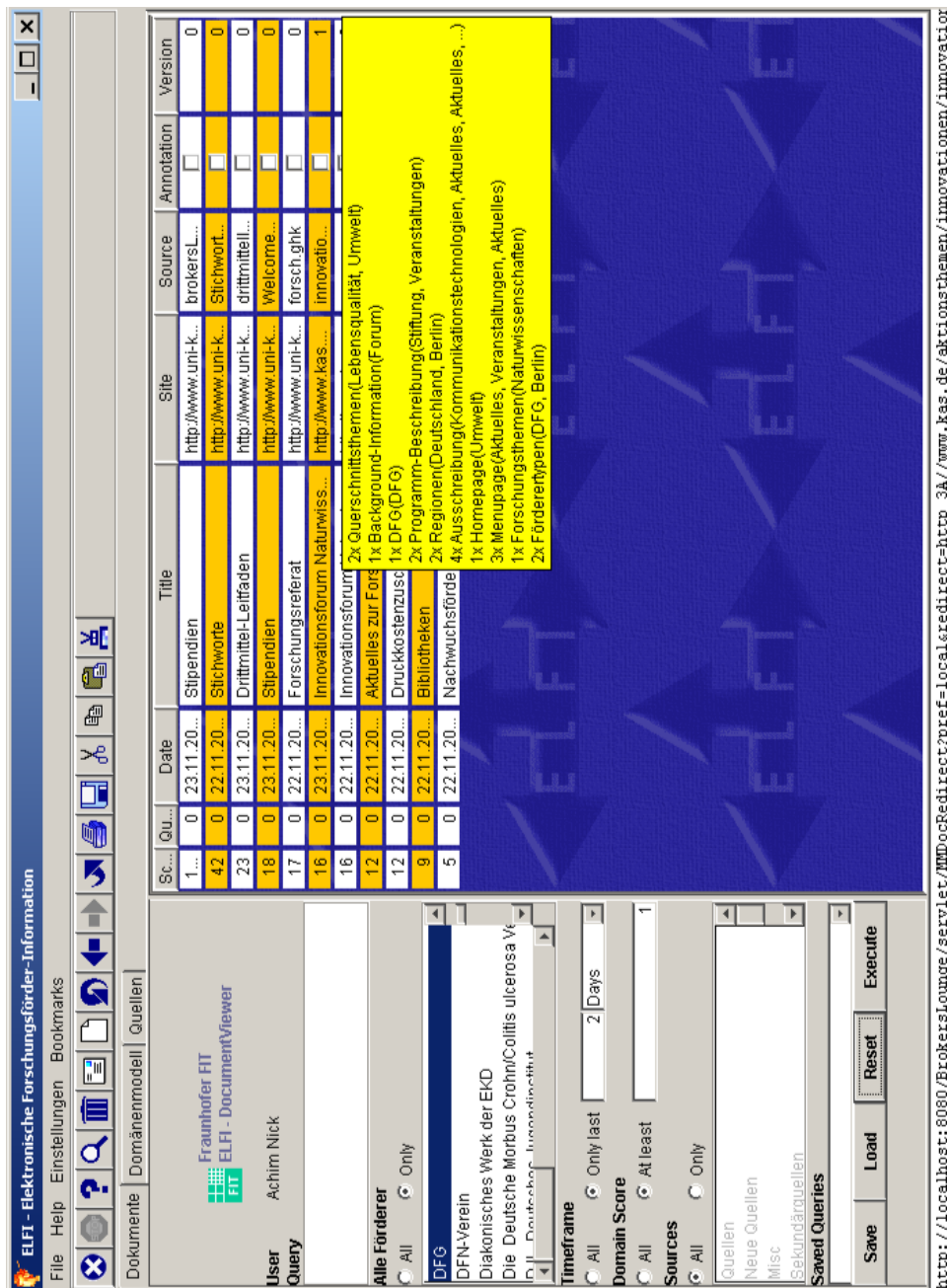


Abbildung 8.12: Dokumentenbrowser – Übersicht

ment für den angefragten Begriff ist. Dieser Anfragewert wird in ähnlicher Weise ermittelt, wie der Domänenwert.

- Informationseinheiten: Das Ergebnis kann auf solche Dokumente einge-

schränkt werden, die Treffer der angegebenen Informationseinheiten aufweisen. In Abbildung 8.12 wurde die Suche z.B. auf Dokumente eingeschränkt, die Synonyme der Informationseinheit „DFG“ enthalten.

- **Zeitraumen:** Es werden nur solche Dokumente angezeigt, die innerhalb des angegebenen Zeitrahmens erstellt oder verändert wurden.
- **Domänenwert:** Es werden nur Dokumente angezeigt, die mindestens den angegebenen Domänenwert aufweisen.
- **Quellentyp:** Die Ergebnisse können auf solche Dokumente eingeschränkt werden, die aus Quellen eines bestimmten Quellentyps stammen.

Alle Anfragekriterien werden miteinander verbunden. In Abbildung 8.12 wird z.B. nach allen Dokumenten gefragt, in denen DFG oder eines der Synonyme von DFG vorkommt, die nicht älter als zwei Tage sind und mit mindestens einem Punkt bewertet wurden.

Die rechte Seite zeigt jeweils die Ergebnisse der Anfrage. Neben dem Titel des Dokumentes wird die Quelle angezeigt, von der das Dokument stammt, das Datum der letzten Änderung, die Versionsnummer des Dokumentes sowie der Domänenwert des Dokumentes. Wenn eine Volltextanfrage gestellt wurde, wird zusätzlich noch ein Anfragewert für jedes Dokument berechnet. Damit ein schneller Überblick über die Relevanz der einzelnen Dokumente möglich ist, werden die Treffer eines jeden Dokumentes in einem vergrößerten Tooltip angezeigt, wenn der Benutzer eine Zeit lang auf einem Dokument verweilt. Dadurch können irrelevante Dokumente, die zwar Treffer aufweisen, aber trotzdem nicht zur Domäne gehören, identifiziert und gelöscht werden. Wenn ein Benutzer sich entscheidet, ein Dokument genauer anzuschauen, wählt er dieses Dokument aus und bekommt das Dokument in einem HTML-Browser angezeigt (vgl. Abb. 8.13). Dieser HTML-Browser ist in die Schnittstelle integriert und kontextualisiert zur besseren Beurteilung Zusatzinformationen, die an die Aufgaben des Benutzers angepasst sind. In diesem Fall werden die formalen Metadaten des Robots (vgl. Abb. 8.13, „Robot-Daten“) angezeigt, da der Benutzer gleichzeitig für die Quellenadministration zuständig ist. Kommt der Benutzer zu dem Schluss, dass auf Grund dieses Dokumentes die Konfiguration des entsprechenden Robots geändert werden muss, so bekommt er Zugriff auf die entsprechende Robotkonfiguration, indem er den Robot im Kontextfenster anwählt.

Unabhängig von seinen weiteren Aufgaben bekommt er die Treffer für das aktuelle Dokument angezeigt. Außerdem kann er das Dokument annotieren. Diese Annotation kann von allen anderen Benutzern eingesehen werden. Wenn er zugleich das Recht hat, Informationseinheiten anzulegen und er auf Basis des Dokumentes eine neue Informationseinheit anlegen möchte, wählt er das Konzept der

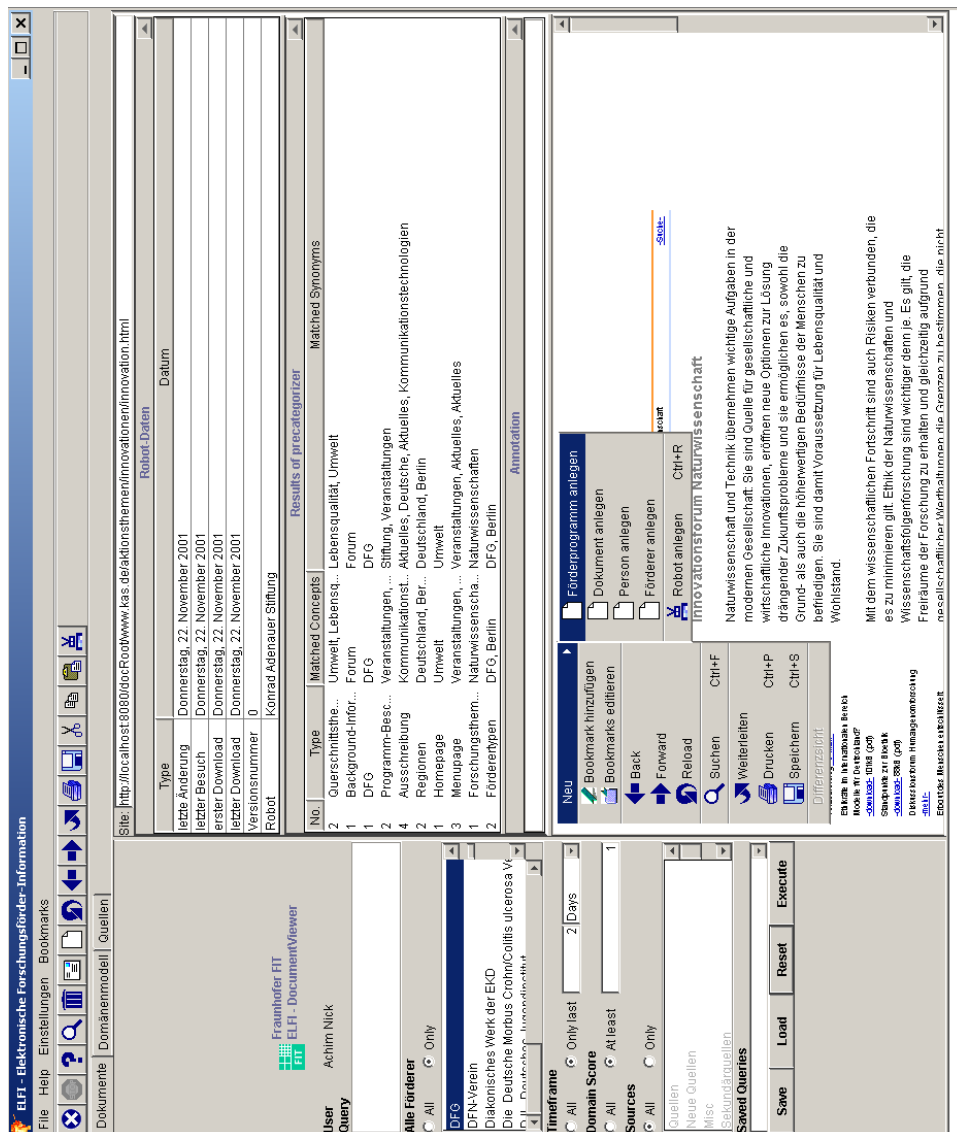


Abbildung 8.13: Dokumentenbrowser – Detailansicht mit zugehörigen Treffern

neuen Informationseinheit aus. Er kann dann eine neue Informationseinheit auf Basis der in Kapitel 8.4.1 beschriebenen Schnittstelle erzeugen. Die momentane Seite wird als Homepage für die neue Informationseinheit angenommen.

Wenn der Benutzer das Recht hat, Informationseinheiten zu editieren und das momentane Dokument der Homepage einer oder mehrerer Informationseinheiten entspricht, werden diese in einem weiteren Kontextfenster („beschriebene Informationseinheiten“, vgl. Abb. 8.14) angezeigt. Durch Auswahl einer Informationseinheit gelangt der Benutzer zur Schnittstelle zum Editieren von Informationsein-

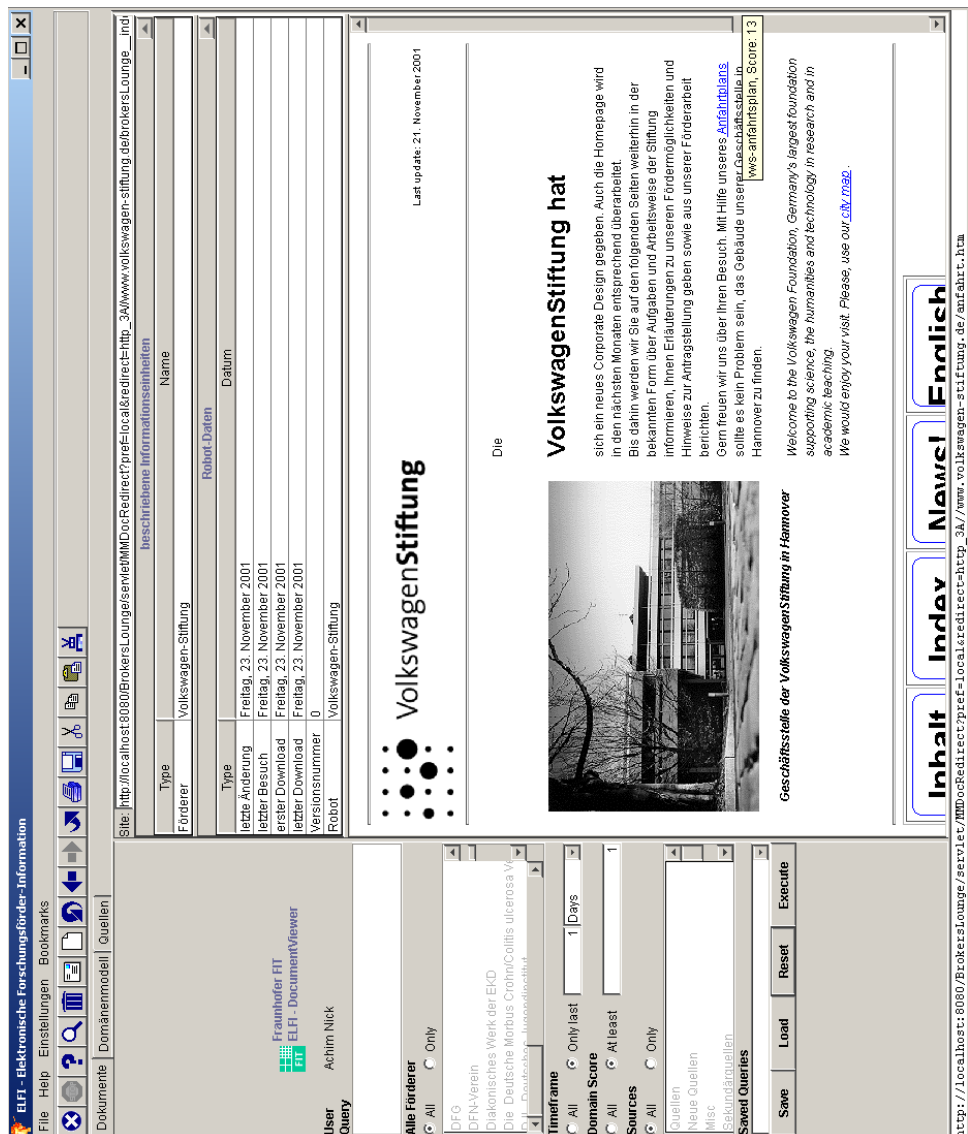


Abbildung 8.14: Dokumentenbrowser – Detailansicht mit zugehöriger Informationseinheit

heiten (vgl. Kap. 8.4.1). Dadurch kann der Benutzer Änderungen in einem Dokument direkt in die entsprechenden Informationseinheiten übertragen. Änderungen in einem Dokument können durch eine Differenzsicht sichtbar gemacht werden. Wurde zum Beispiel in einem Förderprogramm, bzw. dem Dokument, in dem das Förderprogramm beschrieben wird, die Deadline geändert, so wird der Benutzer durch die Differenzsicht darauf aufmerksam gemacht, indem die neuen Bestandteile, ebenso wie die gelöschten Bestandteile des Dokumentes farblich markiert

werden. Dadurch wird dem Benutzer die Identifikation der veränderten Bereiche erleichtert und er muss nicht das gesamte Dokument lesen, um die Änderungen zu erkennen.

Webschnittstelle für den Dokumentenbrowser

Wie schon zuvor beschrieben (vgl. Kap. 8.2), bietet Broker's Lounge eine Webversion des Dokumentenbrowsers an (vgl. Abb. 8.15). Die Webschnittstelle dient im Unterschied zum Dokumentenbrowser des Broker's Lounge-Clients **nicht** der Kategorisierung von Dokumenten, sondern kann benutzt werden, um die Dokumente im Dokumentenarchiv anzuschauen.

Die Webschnittstelle entspricht in ihren Anfragemöglichkeiten dem im vorangegangenen Abschnitt beschriebenen Dokumentenbrowser. Um einen schnellen Überblick über die gefundenen Dokumente zu bekommen, werden für alle Dokumente in der Tabelle Tooltips mit den gefundenen Treffern angezeigt.

Die Detailansichten (vgl. Abb. 8.16) von einzelnen gefundenen Dokumenten entsprechen ebenfalls weitestgehend den Detailansichten des Dokumentenbrowsers (vgl. Abb. 8.13 - 8.14), haben jedoch keine optionalen Kontextfenster, in denen die Robotdaten und die assoziierten Informationseinheiten eines Dokumentes visualisiert werden. Stattdessen werden die gefundenen Treffer für dieses Dokument angezeigt. Die Annotation von Dokumenten ist ebenfalls möglich.

8.7 Profilschnittstelle

Die kundenorientierten Aufgaben im Information-Brokeringprozess werden mit Hilfe der Profilschnittstelle (vgl. Abb. 8.17) unterstützt. Die Interessen des Konsumenten werden mit dieser Schnittstelle in einem oder mehreren Interessenprofilen abgebildet. Mehrere Interessenprofile können benutzt werden, um verschiedene Zwecke abzubilden, für die ein Benutzer Informationen benötigt. Die Interessenprofile definieren jeweils eine Sicht des Konsumenten auf die Inhalte des IB-Domänenmodells. Interessenprofile sind als Overlay-Modell (vgl. Kap. 2.2.4) über das IB-Domänenmodell konzipiert und bestehen aus einer Kombination von Domänenmodellfiltern, wie sie in Kapitel 7 beschrieben wurden.

Grundsätzlich sind zwei Nutzungen der Profilschnittstelle von Broker's Lounge möglich: Entweder bilden Konsumenten ihre Interessen selbst mit Hilfe der Profilbildungsschnittstelle auf das IB-Domänenmodell ab, oder diese Aufgabe wird ihnen von einem Broker abgenommen, der ihre Interessen und Informationsbedürfnisse kennt. In diesem Fall ist der (End-)Konsument der Informationen außerhalb des Broker's Lounge-Systems.

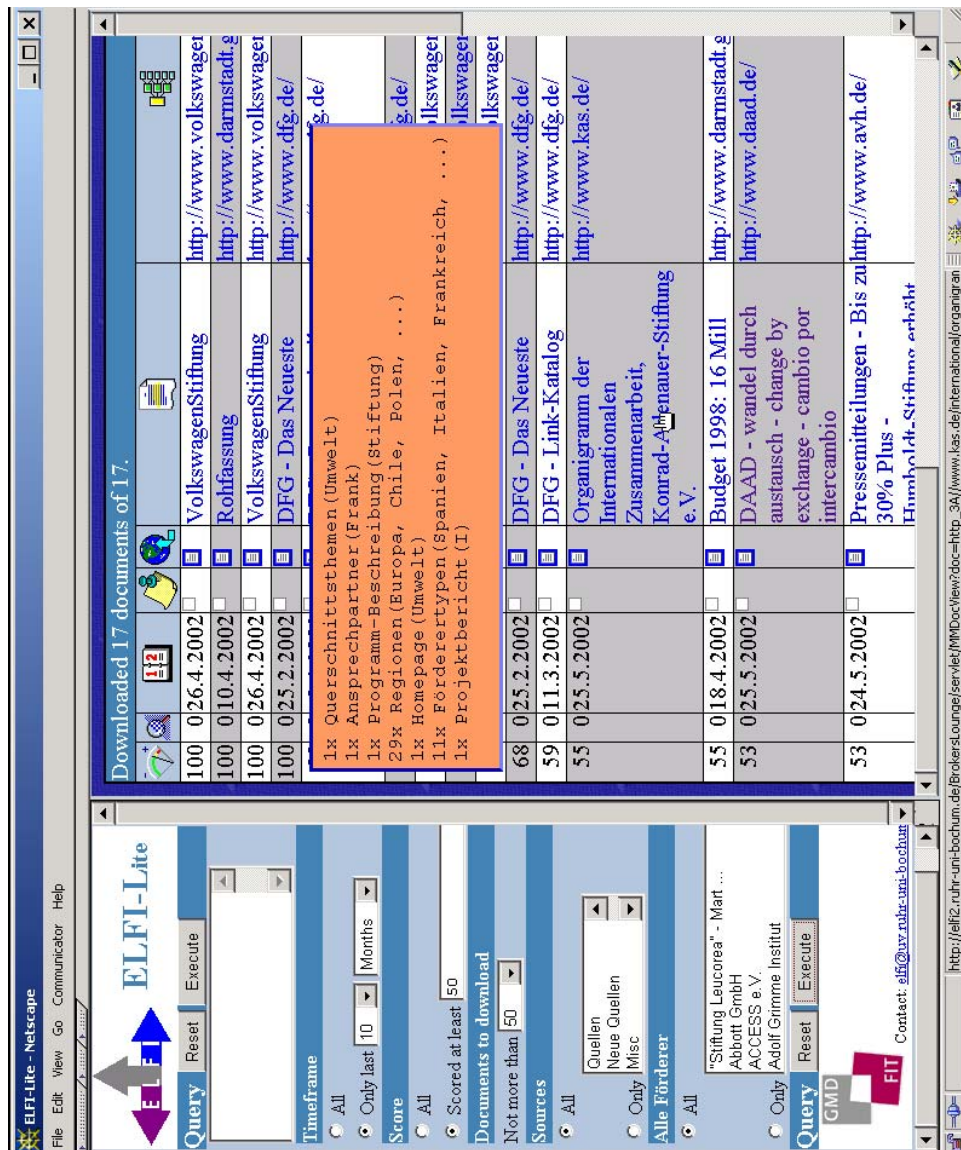


Abbildung 8.15: Webschnittstelle des Dokumentenbrowsers – Übersichtstabelle

Der kundenorientierte Prozess besteht nach dem Aufgaben- und Rollenmodell für Information-Brokering, das in Kapitel 3.1 bzw. Abbildung 3.4 beschrieben wurde, aus

- der *Profilbildung*, bei der die Konsumenteninteressen auf das IB-Domänenmodell abgebildet werden,
- dem *Querying*, bei dem das definierte Interessenprofil auf die Inhalte des IB-Domänenmodells angewendet wird und als Ergebnis alle diejenigen In-

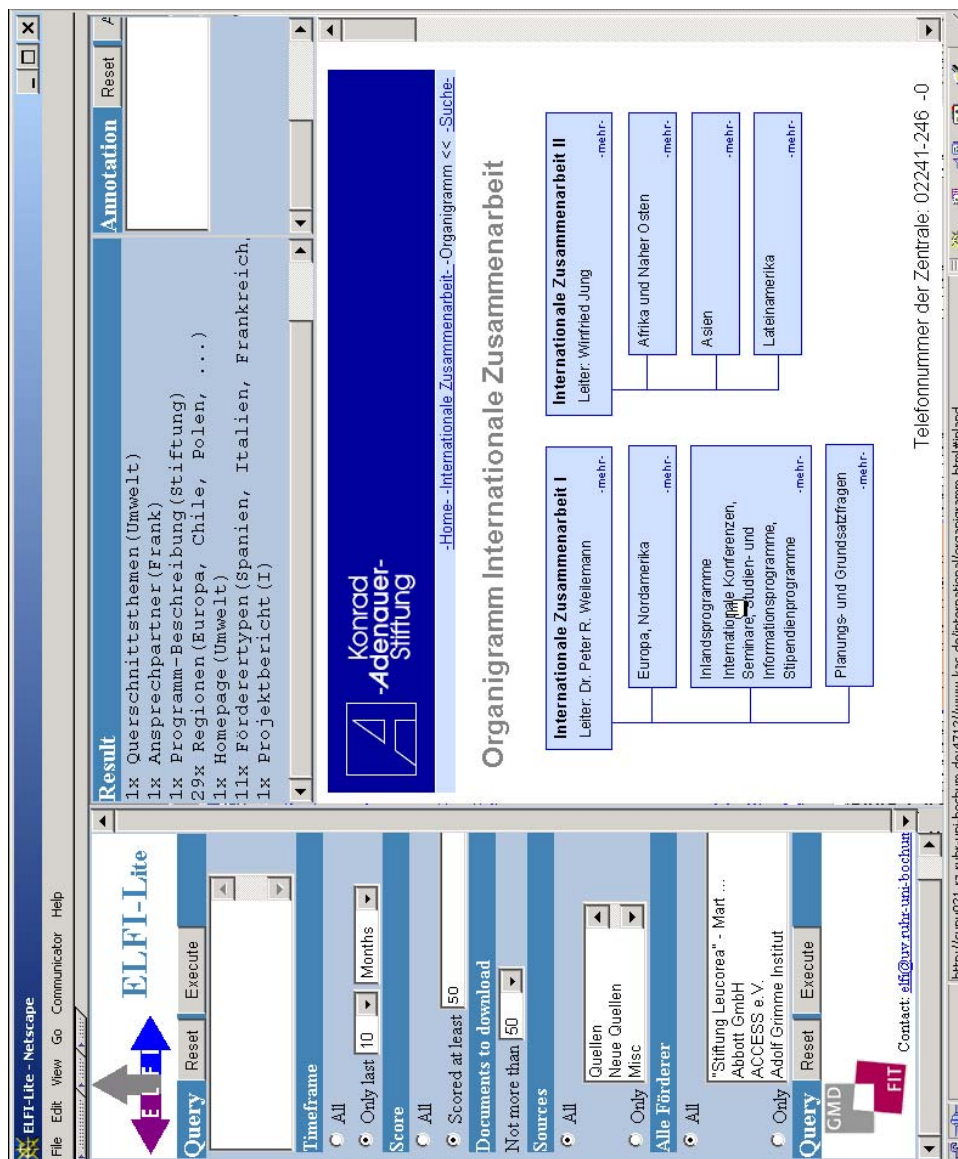


Abbildung 8.16: Webschnittstelle des Dokumentenbrowsers – Dokumentenansicht

formationseinheiten liefert, die den durch das Interessenprofil definierten Bedingungen entsprechen.

- In einem abschließenden Schritt der *Ergebnisauswahl* werden die Ergebnisse des Querying noch einmal überarbeitet, und die für den Konsumenten relevantesten Ergebnisse werden ausgewählt und ggf. in einem sog. Dossier zusammengestellt.

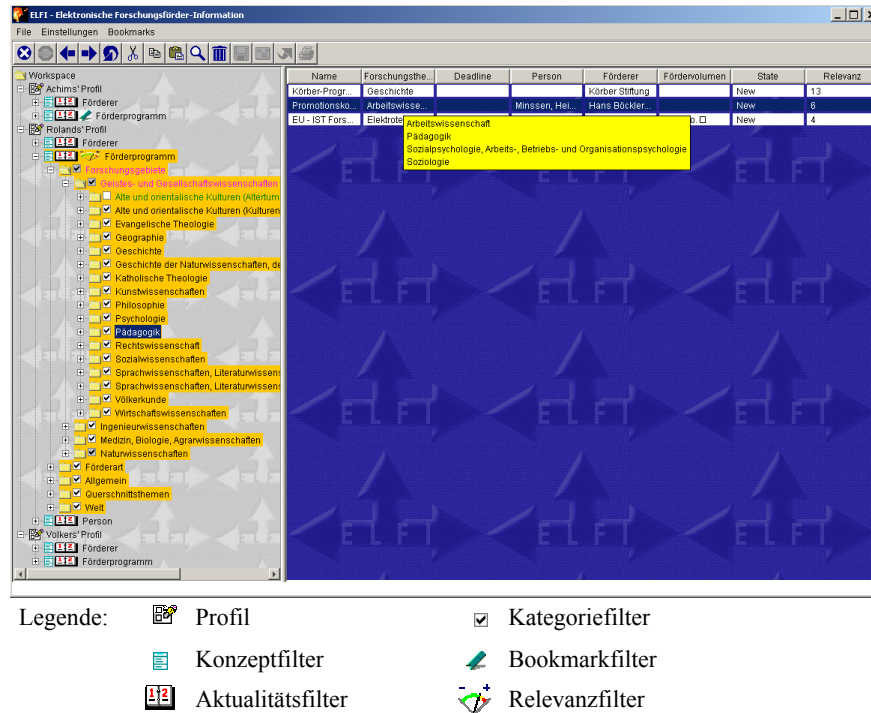


Abbildung 8.17: Profilbildungsschnittstelle

In der Profilbildungsschnittstelle werden die Profilbildung und das Querying miteinander integriert: Sobald innerhalb der Profilbildung ein Profil geändert wurde, wird dieser veränderte Filter auf das IB-Domänenmodell angewendet und das Ergebnis angezeigt. Dadurch bekommt der Benutzer jederzeit eine unmittelbare Rückmeldung über die Auswirkungen einer vorgenommenen Profilveränderung.

Die Profilbildungsschnittstelle ist zweigeteilt:

- Die linke Seite der Schnittstelle visualisiert die momentan definierten Interessenprofile und unterstützt damit die Profilbildung. Die verschiedenen Interessenprofile können dort jederzeit geändert werden.
- Die rechte Seite zeigt jeweils die Ergebnisse des auf der linken Seite ausgewählten Interessenprofils. Die Ergebnisse bestehen aus den Informationseinheiten, die durch den Filter des momentan ausgewählten Interessenprofils passen. Aus der Ergebnisliste können in einem weiteren Schritt die für einen Konsumenten relevanten Informationseinheiten ausgewählt und zu einem Dossier zusammengefasst werden.

In Abbildung 8.17 wurden drei Profile definiert (Achim's Profil, Roland's Profil und Volker's Profil). Diese Interessenprofile bestehen nach den Überlegungen

in Kapitel 7 aus einer Reihe von Domänenmodellfiltern. Diese Domänenmodellfilter arbeiten typischerweise zusammen, um ihre Filterwirkung zu maximieren. Im einzelnen werden die Filter in Kapitel 7 in

- Konzept-,
- Kategorietypen-,
- Kategorie- und
- Instanzenfilter

unterschieden. Alle Filtertypen mit Ausnahme von Kategorietypenfiltern können in der Profilbildungsschnittstelle direkt beobachtet und vom Benutzer editiert werden.

Wie in Kapitel 7.6 beschrieben, können Konzeptfilter und Kategoriefilter durch eine adaptive Komponente unterstützt werden. Diese adaptive Komponente beobachtet den Benutzer und versucht daraus auf die Interessen des Benutzers zu schließen. Deckt das System Diskrepanzen zwischen dem aus der Beobachtung des Benutzers gewonnenen Modell und dem tatsächlich vom Benutzer angegebenen Modell auf, so gibt die adaptive Komponente dem Benutzer den Tipp, sein Profil entsprechend zu ändern. Dies dient der Profilhygiene. Der Benutzer behält die volle Kontrolle. Zur Zeit ist in Broker's Lounge die adaptive Unterstützung von Kategoriefiltern, nicht aber die adaptive Unterstützung von Konzeptfiltern umgesetzt.

8.7.1 Konzeptfilter

Durch die Auswahl von Konzeptfiltern in der Profilbildungsschnittstelle kann der Benutzer angeben, für welche Informationstypen (Konzepte) er¹¹ sich interessiert. In Abbildung 8.17 enthalten „Achim's" und „Volker's Profil" jeweils die Konzeptfilter „Förderprogramm" und „Förderer", während in „Roland's Profil" zusätzlich noch der Konzeptfilter „Person" ausgewählt wurde. D.h., dass Volker und Achim sich für Informationen über Förderprogramme und Förderer interessieren. Roland interessiert sich zusätzlich für Informationen über Personen (z.B. über Ansprechpartner von Förderprogrammen). In einem initialen Profil sind zunächst Konzeptfilter für alle Konzepte definiert, die im IB-Domänenmodell enthalten sind. Diese Konzeptfilter können jederzeit aus dem Profil gelöscht werden. Um Konzeptfilter wieder in das Profil aufzunehmen, kann man das ursprüngliche Profil wiederherstellen.

¹¹In diesem Fall ist der Benutzer auch gleichzeitig der Konsument im Informations-Brokeringsprozess. Wenn der Benutzer ein Profil für einen Dritten anlegt, passt er das Profil an dessen Interessen an.

8.7.2 Kategorie- und Kategorietypenfilter

Wenn im darunterliegenden IB-Domänenmodell mit Konzepten Kategorietypen assoziiert sind, werden diese im Profil durch Hierarchien von Kategoriefiltern abgebildet. Diese Hierarchie bildet die Kategorienhierarchie im IB-Domänenmodell nach, die mit dem entsprechenden Kategorietypen verbunden ist. In Abbildung 8.17 sind mit dem Konzeptfilter „Förderprogramm“ fünf Kategoriefilterhierarchien („Forschungsgebiete“, „Förderart“, „Allgemein“, „Querschnittsthemen“ und „Welt“) assoziiert. Dies entspricht dem ELFI-Domänenmodell, das in Kapitel 9.1 genauer ausgeführt wird. Kategorietypenfilter können aus einem Profil entfernt werden, indem die entsprechende Kategorienhierarchie aus dem Profil gelöscht wird. Diese Hierarchie kann zu jedem Zeitpunkt wiederhergestellt werden.

Jede Kategorie innerhalb der Hierarchie kann vom Benutzer durch einen Haken in das Profil integriert und durch das Löschen des Hakens aus dem Profil entfernt werden. Wenn für eine Kategorie ein oder mehrere Unterkategorien ausgewählt und ein oder mehrere Unterkategorien abgewählt sind, wird dies durch einen grauen Haken repräsentiert. In Abbildung 8.17 weisen die Kategorien „Forschungsgebiete“, „Geistes- und Gesellschaftswissenschaften“ und „Naturwissenschaften“ graue Haken auf. Wenn eine Kategorie in den Filter aufgenommen wird, werden alle Unterkategorien ebenfalls in den Filter aufgenommen. Wenn eine Kategorie aus dem Filter entfernt wird, geschieht dies ebenfalls für alle ihre Unterkategorien.

Kategoriefilter dienen der Filterung von Informationseinheiten (vgl. Kap. 7.4). Dabei werden alle Kategorien einer Kategorienhierarchie zusammen ausgewertet: Eine Informationseinheit passt durch die gewählten Filter, wenn diese Informationseinheit mit mindestens einer Kategorie aus dem Filter klassifiziert wurde.

Wenn ein neues Profil initialisiert wurde, sind mit den jeweiligen Konzeptfiltern alle (durch das IB-Domänenmodell) möglichen Kategoriehierarchien assoziiert. Alle Kategorien sind in den jeweiligen Filter integriert.

8.7.3 Adaptiver Kategoriefiltersupport

Die adaptive Unterstützung von Kategoriefiltern beruht auf dem Verfahren, das in Kapitel 7.6.2 beschrieben wurde. Für jede Kategorie wird mit einem Zähler die Anzahl der Informationseinheiten gezählt, die vom Benutzer angeschaut wurden und die mit dieser Kategorie oder einer ihrer Unterkategorien klassifiziert wurden. Außerdem wird ein Erwartungswert berechnet, der angibt, wie viele Aufrufe von Informationseinheiten über einen festgelegten Zeitraum im Mittel beobachtet werden sollten. Der Erwartungswert errechnet sich aus den Informationseinheiten, die mit dieser Kategorie (oder einer Unterkategorie) klassifiziert wurden und der Grundgesamtheit aller Informationseinheiten (vgl. Formel 2.1), die durch den

momentanen Filter passen. Mit einem gegebenen Konfidenzintervall (z.B. 95%) kann daraus mit Formel 2.2 eine untere und eine obere Signifikanzschwelle errechnet werden. Liegt der Zähler einer Kategorie oberhalb der oberen Schwelle und hat der Benutzer diese Kategorie abgewählt, schlägt das System dem Benutzer vor, diese Kategorie in das Profil mit aufzunehmen. Dies wird dem Benutzer signalisiert, indem die entsprechende Kategorie grün eingefärbt wird. In Abbildung 8.17 wird dem Benutzer z.B. vorgeschlagen, die Kategorie „Alte und orientalische Kulturen (Altertumswissenschaft)“ in sein Profil aufzunehmen. Liegt der Zähler unterhalb der unteren Schranke und ist eine Kategorie im Profil enthalten, schlägt das System dem Benutzer vor, diese Kategorie aus dem Profil zu entfernen, indem es diese Kategorie rot einfärbt. Tipps zum Entfernen einzelner Kategorien aus dem Profil können dem Benutzer erst gegeben werden, nachdem der (minimale) Beobachtungszeitraum abgelaufen ist, weil man erst dann sicher sein kann, dass der Zähler nicht aus mangelnder Gelegenheit unter der unteren Schranke geblieben ist.

Alle Oberkategorien von grün oder rot markierten Kategorien werden in magenta eingefärbt, um dem Benutzer zu signalisieren, dass sich unterhalb dieser Kategorie ein Tipp für eine Profiländerung befindet. In Abbildung 8.17 sind z.B. die Kategorien „Geistes- und Gesellschaftswissenschaften“ und „Forschungsgebiete“ magenta eingefärbt, da die Unterkategorie „Alte und orientalische Kulturen (Altertumswissenschaft)“ grün eingefärbt wurde.

8.7.4 Instanzenfilter

Instanzenfilter nutzen zur Filterung von Informationseinheiten entweder Eigenschaften von Informationseinheiten aus, die erfüllt sein müssen oder sie zählen alle Informationseinheiten auf, die durch den Filter passen. In Broker's Lounge sind zur Zeit drei verschiedene Instanzenfilter realisiert, die in den nächsten Abschnitten beschrieben werden. Instanzenfilter werden jeweils mit Konzeptfiltern kombiniert. D.h., sie werden jeweils für die Informationseinheiten eines Konzeptes definiert. Instanzenfilter können editiert werden, indem das entsprechende Symbol (siehe Legende in Abb. 8.17) angewählt wird.

Aktualitätsfilter

Aktualitätsfilter filtern Informationseinheiten aufgrund der Aktualität von Informationseinheiten. In Abbildung 8.17 sind alle Konzeptfilter mit einem Aktualitätsfilter verknüpft. Aktualitätsfilter können benutzt werden, um alle nicht aktuellen Informationseinheiten heraus zu filtern. Um die Aktualität von Informationseinheiten bewerten zu können, wird ein Referenzpunkt benötigt. Dieser Referenzpunkt kann entweder das Datum der letzten Änderung einer Informationsein-

heit oder der Zeitpunkt des letzten Besuchs durch den entsprechenden Benutzer sein. Im ersten Fall wird Aktualität aus Sicht des Systems bewertet. Im zweiten Fall wird Aktualität aus Sicht der persönlichen Dialoghistorie bewertet: Eine Informationseinheit gilt solange als „neue“ Informationseinheit, bis diese Informationseinheit von dem entsprechenden Benutzer besucht wurde. Wird eine Informationseinheit verändert, nachdem sie von dem entsprechenden Benutzer angeschaut wurde, gilt diese Informationseinheit als „verändert“, bis sie erneut angeschaut wurde.

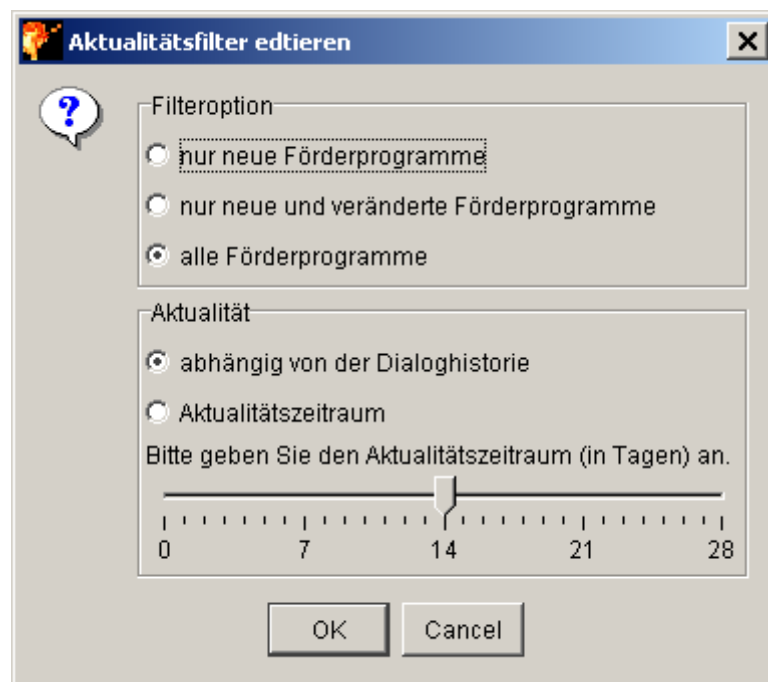


Abbildung 8.18: Aktualitätsfilter editieren

Abbildung 8.18 zeigt die Schnittstelle zur Einstellung von Aktualitätsfiltern. Man kann zwischen den beiden Aktualitätsmodi (persönliche Dialoghistorie oder Systemsicht) auswählen. Wenn man sich für die Systemsicht entscheidet, kann man einen Aktualitätszeitraum angeben, in dem Informationseinheiten als neu oder verändert gelten. Gibt man z.B. einen 14-tägigen Aktualitätszeitraum an, werden alle Informationseinheiten als „neu“ oder „verändert“ markiert, die in den letzten 14 Tagen definiert oder verändert wurden. Der berechnete Aktualitätsstatus wird in der Tabelle auf der rechten Seite jeweils in der Spalte „State“ angezeigt. Neben dem Aktualitätsmodus kann der Filter auf die verschiedenen Aktualitätszustände eingeschränkt werden. Wurde „nur neue Förderprogramme“ ausgewählt, passieren nur neue Informationseinheiten (des Konzeptes Förderprogramm) den Filter. Der Filter kann außerdem auf neue und veränderte Informationseinheiten

eingestellt oder ganz ausgeschaltet werden. Wenn der Filter ganz ausgeschaltet wird, werden keine Informationseinheiten gefiltert. Die Filterung kann in diesem Fall der Benutzer implizit durchführen, indem er die Informationseinheiten in der Tabelle auf der rechten Seite nach ihrem Aktualitätszustand sortiert.

Bookmarkfilter

In Broker's Lounge haben Benutzer die Möglichkeit, wie in herkömmlichen Web-Browsern zu Web-Dokumenten Lesezeichen (Bookmarks) anzulegen. Außerdem können Benutzer Lesezeichen zu einzelnen Informationseinheiten anlegen. Bookmarkfilter nutzen diese Informationseinheiten, für die ein Lesezeichen-Eintrag vorgenommen wurde, als extensionalen Filter. Das bedeutet, dass alle Informationseinheiten durch den Filter passen, für die zuvor ein Lesezeichen angelegt wurde. Bookmarkfilter eignen sich besonders, um Veränderungen bei besonders relevanten Informationseinheiten beobachten, für die man einen Lesezeichen-Eintrag vorgenommen hat. Im Beispiel in Abbildung 8.17 ist der Förderprogramm-Konzeptfilter in „Achim's Profil“ mit einem Bookmarkfilter assoziiert.

Relevanzfilter

Relevanzfilter nutzen die Relevanz von einzelnen Informationseinheiten zur Filterung. Die Relevanz von Informationseinheiten wird aus der Relevanz der Kategorien berechnet, mit denen eine Informationseinheiten klassifiziert wurde. Die Berechnung der Relevanz von Kategorien und Informationseinheiten wurde in Kapitel 7.6.3 beschrieben. Die Relevanz von Informationseinheiten wird in einer Relevanz-Spalte in der Tabelle auf der rechten Seite visualisiert (vgl. Abb. 8.17). Diese Spalte kann benutzt werden, um die Informationseinheiten (in der Tabelle) nach ihrer Relevanz zu sortieren.

Die Relevanz von Informationseinheiten kann außerdem genutzt werden, um alle Informationseinheiten unterhalb einer bestimmten Relevanzschwelle auszufiltern. Diese Relevanzschwelle kann mit der Eingabeschnittstelle angegeben werden, die in Abbildung 8.19 dargestellt ist. Wenn der Filter mit der Filteroption „alle Förderprogramme“ deaktiviert wird, werden nur die Relevanzwerte berechnet und in der Tabelle dargestellt.

8.7.5 Zusammenspiel der verschiedenen Filter

In Kapitel 7.5 wurden zwei Optionen für das Zusammenspiel der verschiedenen Filter zur Filterung von Informationseinheiten identifiziert:

1. Die erste Option $filter_{Kon_f+KatTyp_f+Kat_f}^{IE}$ nutzt Konzepte zur Filterung von Informationseinheiten. Die Konzepte werden ebenfalls benutzt,

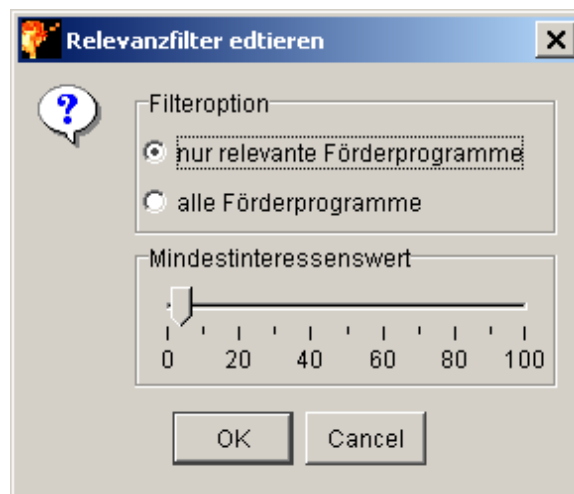


Abbildung 8.19: Eingabeschnittstelle für Relevanzfilter

um Kategorietypen auszuwählen, für die jeweils Kategoriefilterhierarchien angelegt werden können. Diese Kategoriefilterhierarchien können ebenfalls zur Filterung von Informationseinheiten benutzt werden. Informationseinheiten passieren den Filter, wenn sie sowohl den Konzept- als auch die Kategoriefilter passieren.

Die erste Option bedeutet, dass Benutzer zuerst nach Konzepten (Informationstypen) einschränken, bevor sie die Kategoriehierarchien benutzen, um die Informationseinheiten des ausgewählten Konzeptes untereinander zu unterscheiden. In der Domäne Forschungsförderung gibt ein Benutzer z.B. zuerst an, dass er sich für Informationen über Förderprogramme (Konzept Förderprogramm) interessiert, bevor er die Forschungsgebiete angibt, mit denen aus der Menge aller Förderprogramme die für ihn relevanten Förderprogramme gefiltert werden können.

2. Die zweite Option $filter_{Kat_f + KatTyp_f + KonTyp_f}^{IE}$ wählt den entgegengesetzten Weg: Ausgehend von einem Kategoriefilter, mit dem auch Informationseinheiten gefiltert werden können, wird die Menge von Kategorietypen eingeschränkt. Diese eingeschränkte Menge von Kategorietypen wird wiederum benutzt, um die Menge von Konzepten zu bestimmen, mit denen Informationseinheiten gefiltert werden sollen. Die so ermittelten Konzepte können als Konzeptfilter Informationseinheiten filtern. Dieser Konzeptfilter wird dann mit dem anfänglichen Kategoriefilter kombiniert.

Bei der zweiten Option geben Benutzer zuerst die Eigenschaften an, die Informationseinheiten erfüllen müssen, bevor sie die Informationstypen angeben, für die sie Informationen wünschen. In der Domäne Forschungsför-

derung würde ein Benutzer in diesem Fall z.B. angeben, dass er sich für Informationen zu einer bestimmten Kategorie, z.B. „Informatik“, bevor er einschränkt, dass er sich nur für Informationen des Typs „Förderprogramme“ interessiert.

In Broker's Lounge wurde die erste Option gewählt, da sich diese Option an der Schnittstelle leichter umsetzen lässt. Bei der ersten Option (vgl. Abb. 8.17) werden ausgehend von den Konzeptfiltern die jeweiligen Kategoriefilterhierarchien angezeigt. Wollte man die zweite Option in ähnlicher Art und Weise umsetzen, so würde man mit der Visualisierung der Kategoriefilterhierarchien beginnen. Die in diesem Fall abschließende Filterung nach Konzepten müsste dann für jede Kategorie einzeln geschehen oder ganz entfallen.

Der Fall, in dem auf eine abschließende Filterung nach Konzepten verzichtet wird, führte z.B. in der Domäne Forschungsförderung dazu, dass man nur noch nach Kategorien (z.B. „Informatik“) einschränken kann. D.h., man bekommt alle Informationseinheiten, die mit dieser Kategorie klassifiziert wurden. Eine weitere Unterscheidung nach dem Typen (Konzept) der Informationseinheiten ist nicht mehr möglich. Dies führt in Fällen, in denen eine Kategorienhierarchie zur Klassifikation von mehreren Konzepten benutzt wird, zu einem Verlust an Mächtigkeit für die Filterbeschreibung.

Will man die nachgelagerte Filterung mit Konzepten an der Oberfläche visualisieren, kann man dies tun, indem man jeden Kategoriefilter mit einem eigenen Unterknoten für die jeweiligen Konzepte versieht, oder die Kategoriefilter(-knoten) jeweils mit den gefilterten Konzepten annotiert. Beide theoretischen Möglichkeiten führen zu einer größeren Unübersichtlichkeit der Schnittstelle.

8.7.6 Selektionsfilter

In Kapitel 8.7.5 wurde beschrieben, wie die verschiedenen vom Benutzer definierten Filter zusammenwirken. Um dem Benutzer anzuzeigen, welche Filter zur Zeit benutzt werden, wird der Geltungsbereich der momentan aktiven Filter farblich unterlegt (vgl. Abb. 8.17). Diese Filter definieren die Menge aller Informationseinheiten, für die ein Benutzer sich interessiert. Damit Benutzer sich innerhalb dieses Raumes orientieren können, haben sie die Möglichkeit, die Profil-, Konzept- und Kategorienstruktur zur Navigation des Raumes der gefilterten Informationseinheiten zu nutzen. Durch die Selektion wird jeweils ein weiterer Filter definiert, der zusammen mit den anderen Filtern zur weiteren Einschränkung genutzt wird.

Kategorie Selektion

In Abbildung 8.17 hat der Benutzer z.B. die Kategorie „Pädagogik“ ausgewählt. Das bedeutet, dass zusätzlich zu den anderen Filtern (Kategorie-, Konzept- und Instanzenfiltern) ein weiterer Kategoriefilter für die Kategorie „Pädagogik“ definiert wird. Im Unterschied zu den anderen Kategoriefiltern, die jeweils innerhalb einer Kategorienhierarchie disjunktiv wirken, wird der Selektionsfilter konjunktiv benutzt. D.h., auf der rechten Seite werden alle Informationseinheiten angezeigt, die durch den momentanen Filter passen **und** zusätzlich mit der Kategorie „Pädagogik“ klassifiziert wurden.

Konzept Selektion

Wird ein Konzept (ein Konzeptfilterknoten) ausgewählt, wird dieser Konzeptfilter als weiterer Filter zusätzlich zu den anderen Filtern (Konzept-, Kategorie- und Instanzenfilter) angewendet. Da derselbe Konzeptfilter schon in den bisher verwendeten Filtern enthalten ist, bewirkt der selektierte Konzeptfilter keine weitere Filterung von Informationseinheiten. Stattdessen liefert die Selektion eines Konzeptes einen Überblick über alle Informationseinheiten, die zu diesem Konzept gehören und durch den momentanen Filter passen.

Profil Selektion

Ähnlich wie bei der Selektion eines Konzeptknotens dient die Selektion eines Profils dazu, sich einen schnellen Überblick über das selektierte Profil zu verschaffen. Dabei werden die Inhalte des Profils in Relation zur persönlichen Dialoghistorie des Benutzers gesetzt. Dadurch bekommt der Benutzer einen schnellen Überblick darüber, was sich innerhalb seines Profils seit seinem letzten Login verändert hat. In Abbildung 8.20 sind z.B. für „Volker's Profil“ fünf neue Förderprogramme aufgelistet.

8.7.7 Push-Dienst

Die Benutzer haben mit der Profilbildungsschnittstelle die Möglichkeit, sich aktiv über die abgebildete Domäne und die Veränderungen in dieser Domäne zu informieren. Der Push-Dienst ermöglicht es dem Benutzer, sich per Email über die Veränderungen in seinen Profilen informieren zu lassen. Damit Benutzer nur die Informationen bekommen, für die sie sich interessieren, werden nicht alle Veränderungen innerhalb des IB-Domänenmodells an den Benutzer gesendet, sondern nur die Informationen, die auf mindestens eines seiner Profile passen. Der Benutzer erhält für jedes seiner Profile eine Übersicht über die Veränderungen seit seinem

Name	Typ	Letzte Änder...	Letzter Besuch	Status
Grako "Bioinformatik"	Förderprogramm	24.01.2002		New
BMBF - "Komponenten und Systeminnovationen der Silizium-Leitu...	Förderprogramm	24.01.2002		New
BMBF - Ekompass	Förderprogramm	24.01.2002		New
Grako "Betriebswirtschaftliche Aspekte lose gekoppelter Systeme ...	Förderprogramm	24.01.2002		New
BMBF - Service-Engineering und Service-Design	Förderprogramm	24.01.2002		New
Grako "Betriebswirtschaftliche Aspekte lose gekoppelter Systeme im Zeitalter neuer Medien"				

Workspace

- Achims' Profil
- Rolands' Profil
- Volkers' Profil

Grako "Betriebswirtschaftliche Aspekte lose gekoppelter Systeme im Zeitalter neuer Medien"

Abbildung 8.20: Profilbildungsschnittstelle - Profilübersicht

letzten Login bzw. seit der letzten Mail, die er erhalten hat. Die Profil-Übersichten entsprechen im Wesentlichen der Übersicht, die bei der Profilselektion (vgl. Kap. 8.7.6) angezeigt werden. Die Email des Push-Dienstes enthält Verweise auf die

Detailansichten der veränderten Informationseinheiten. Der Benutzer kann sich die Detailansichten in seinem Web-Browser anschauen, indem er die jeweiligen Verweise benutzt. Er nutzt dazu die in Kapitel 8.7.11 beschriebene Webschnittstelle für Profile.

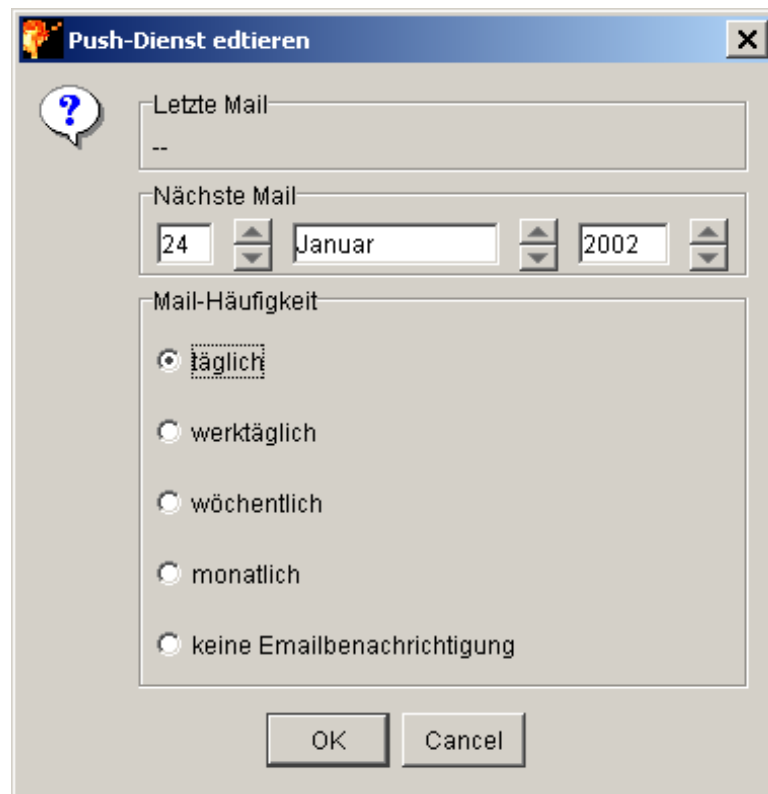


Abbildung 8.21: Push-Dienst konfigurieren

Um zu verhindern, dass Benutzer sich durch zu häufige Emails des Push-Dienstes gestört fühlen, können sie das Intervall, in dem sie vom Push-Dienst informiert werden, selbst bestimmen oder den Push-Dienst ggf. abbestellen. Die Schnittstelle zur Konfiguration des Push-Dienstes ist in Abbildung 8.21 dargestellt.

8.7.8 Stereotypen

Das Lernen an positiven Beispielen (vgl. Kap. 2.2.9), das in Broker's Lounge eingesetzt wird, benötigt einen gewissen Vorrat an Ereignissen ($n \geq 30$, vgl. Kap. 2.2.9), bevor das Lernverfahren zu validen Ergebnissen kommen kann. Dies bedeutet, dass ein Benutzer erst dann Hilfestellungen durch das System bekommen kann, wenn er eine gewisse Anzahl von Detailansichten angeschaut hat, weil das

Anschauen von Detailansichten den Input für das Lernverfahren liefert. Um diese Phase abzukürzen, können in Broker's Lounge Stereotypen verwendet werden. Stereotypen werden in adaptiven Informationssystemen benutzt, um Benutzer- oder Interessenprofile zu initialisieren, damit das System schneller Annahmen über den Benutzer oder seine Interessen machen kann (vgl. Kap. 2.2.5).

In Broker's Lounge bestehen Profile aus einer Reihe von Filtern sowie einer Historie an selektierten Detailansichten, die für die Bewertung von Informationseinheiten durch einen Aktualitätsfilter und/oder einen Relevanzfilter benötigt werden. Ein Stereotyp in Broker's Lounge entspricht einem Profil. Um ein neues Profil mit einem Stereotyp zu initialisieren, werden alle Filter des Stereotyps kopiert und in das neue Profil eingefügt. Die Historie des Stereotyps wird ebenfalls kopiert. Die kopierte Historie wird im neuen Profil ausschließlich für die Bewertung von Informationseinheiten durch einen Relevanzfilter benutzt, um zu verhindern, dass Informationseinheiten durch einen Aktualitätsfilter als bereits gesehen bewertet werden, obwohl sie innerhalb des Profils noch nicht angeschaut wurden. Alle folgenden Aktionen des Benutzers werden vom Aktualitätsfilter in die Betrachtung miteinbezogen.

Ein Problem in adaptiven Informationssystemen ist die Frage, wie einem neuen Benutzer ein Stereotyp zugeordnet werden kann, da zu diesem Zeitpunkt wenig Wissen über den Benutzer vorliegt (vgl. Kap. 2.2.5). Diese Problematik wird in Broker's Lounge dadurch umgangen, dass diese Zuordnung von den Benutzern selbst ausgeführt wird. Um die Auswahl eines geeigneten Stereotyps zu unterstützen, werden alle zur Auswahl stehenden Stereotypen mit einer Kurzbeschreibung versehen. Die Auswahl eines Stereotyps erfolgt, wenn ein Benutzer ein neues Profil anlegt.

8.7.9 Übersichtslisten und Detailansichten

Durch die verschiedenen Filter und Selektionsmöglichkeiten kann der Benutzer den Umfang der Tabelle auf der rechten Seite der Profilbildungsschnittstelle beeinflussen. Die Tabelle besteht aus der Menge der Informationseinheiten, die durch die zu diesem Zeitpunkt definierten Filter passen. Die Tabelle zeigt jeweils den Namen der Informationseinheit sowie alle Querbeziehungen zu anderen Informationseinheiten, alle Kategorien, mit denen diese Informationseinheit klassifiziert wurde, sowie die Werte aller Attribute an. Der Benutzer kann auswählen, welche Spalten er angezeigt haben möchte. In Abbildung 8.17 werden zum Beispiel die Einträge für die Forschungsthemen (des Förderprogramms) angezeigt, während für die anderen Kategoriehierarchien keine Spalten ausgewählt wurden. Die Inhalte jeder Tabellenzeile werden durch einen vergrößerten Tooltip angezeigt, um so sicherzustellen, dass die Benutzer auch mehrzeilige Zeileinträge einsehen können.

Durch die Konfigurationsmöglichkeiten der Tabelle und die Leseunterstützung durch die Tooltips soll sichergestellt werden, dass die Benutzer sich möglichst gut über Informationseinheiten informieren können, bevor sie sich entscheiden, für eine Informationseinheit eine Detailansicht anzufordern, indem sie die entsprechende Informationseinheit in der Tabelle auswählen. Die gute Information des Benutzers, bevor er eine Detailansicht auswählt, ist wichtig, da auf Basis der ausgewählten Detailansichten die adaptive Unterstützungskomponente das Interessenmodell des Benutzers aufbaut.

Detailansichten (vgl. Abb. 8.22) sind immer in gleicher Weise aufgebaut:

- Der Name der Informationseinheit sowie das Konzept, zu dem die Informationseinheit gehört, wird im blauen Balken zu Beginn der Detailansicht angezeigt.
- Wenn die Informationseinheit mit einem Web-Dokument assoziiert ist, wird ein Link zu diesem Web-Dokument etabliert, der mit einem Globus markiert wird, um deutlich zu machen, dass dieser Link zu einem Dokument außerhalb des Systems führt. Der Link trägt den Namen der Informationseinheit. Wurde die Informationseinheit nicht mit einem Web-Dokument assoziiert, wird stattdessen der Namen der Informationseinheit abgebildet.

Wenn die Informationseinheit zusätzlich mit einem Bild oder Logo assoziiert ist (z.B. dem Logo des Förderers in der Domäne Forschungsförderung), wird dieses Bild oder Logo links von dem Namen (Link) der Informationseinheit angezeigt.

- Es folgen die Querbeziehungen zu anderen Informationseinheiten. In Abbildung 8.22 sind z.B. Querbeziehungen zum Förderer und zum Ansprechpartner des Förderprogramms aufgeführt. Diese Links können verfolgt werden, um zu den entsprechenden Detailansichten zu gelangen. Im Unterschied zu dem Link zum assoziierten Web-Dokument sind diese Links nicht mit einem Globus-Icon gekennzeichnet, da diese Links Verweise innerhalb des Systems darstellen.
- Unter den Querbeziehungen werden auf der linken Seite die Kategorien aufgeführt, mit denen die Informationseinheit klassifiziert wurde.
- Auf der rechten Seite werden die zusätzlichen Attribute sowie die Kurzbeschreibung der Informationseinheit angezeigt.
- Die Detailansicht wird abgerundet durch die Angabe des Zeitpunktes der letzten Änderung der beschriebenen Informationseinheit.

Site: <http://localhost:8080/Erokers/Lounge/Service/ConceptRedirect?conceptID=901543276710>

Förderprogramm Stipendien zur Erforschung der europäischen Kulturgeschichte vom 18. bis zum 20. Jahrhundert

Stipendien zur Erforschung der europäischen Kulturgeschichte vom 18. bis zum 20. Jahrhundert

Teilprogramm: [Einrich, Lothar, Dr.](#)

Oberrprogramm: [Stiftung Weimarer Klassik](#)

Person:

Förderer:

Dokument:

Forschungsthemen: Kunstwissenschaften Neuere Geschichte Sprachwissenschaften, Literaturwissenschaften und Volkskunde (Gruppe A)	Deadline: 2000 DM
Querschnittsthemen: Querschnittsthemen	Beschreibung: Die Stiftung Weimarer Klassik fördert durch Stipendien wissenschaftliche Projekte, die europäische Kulturgeschichte von der frühen Neuzeit bis zur Gegenwart zum Inhalt haben und unter Nutzung der in Weimar vorhandenen Materialien zu realisieren sind. Es sind besonders Vorhaben erwünscht, die zur intensiven Erforschung der Sammlungen des Goethe- und Schiller-Archivs, des Goethe-Nationalmuseums und der Herzogin Anna Amalia Bibliothek beitragen. Weimarer-Stipendien werden vornehmlich zu folgenden Schwerpunkten vergeben: Kultur der Weimarer Klassik und der Jenaer Romantik/Wirkungsgeschichte der Weimarer Klassik Goethe im kulturellen Transfer der Völker und Nationen Wirkungsgeschichte Friedrich Nietzsches
Förderarten: Reisekostenzuschuss Stipendium	
Institutionsarten: Öffentliche Forschung	
Regionen: Welt	

Letzte Änderung: 10.12.2001

All Rights Reserved

Abbildung 8.22: Detailansicht einer Informationseinheit

8.7.10 Dossierverwaltung

Wie schon zuvor beschrieben, sind zwei Nutzungen der Profilschnittstelle möglich:

1. Entweder wird die Profilschnittstelle vom Informationskonsumenten selbst bedient oder

2. die Profilschnittstelle wird von einem Broker bedient, der Informationen für einen Informationskonsumenten zusammenstellt.

Broker, die Informationen für Informationskonsumenten zusammenstellen, werden in Broker's Lounge unterstützt, indem sie Detailansichten zu einem sog. Dossier zusammenfassen können. Ein Dossier entspricht einem HTML-Dokument, das aus einer Menge von Detailansichten (von Informationseinheiten) besteht. Informationseinheiten können aus den Übersichtstabellen kopiert und in ein Dossier eingefügt werden. Abbildung 8.23 zeigt z.B. ein Dossier, das aus sechs verschiedenen Förderern und Förderprogrammen zusammengestellt wurde. Dieses Dossier kann entweder zur Weiterverarbeitung mit einem HTML-Editor auf die eigene Festplatte kopiert oder direkt per Email weitergeleitet werden.

8.7.11 Profil-Webschnittstelle

Um die Hemmschwelle für die Nutzung der Profilschnittstelle für Informationskonsumenten möglichst niedrig zu halten, bietet Broker's Lounge eine vereinfachte Webschnittstelle der Profilbildungsschnittstelle an. Diese vereinfachte Webschnittstelle (vgl. Abb. 8.24) kann nur zur Navigation des Profils und der Inhalte des IB-Domänenmodells genutzt werden. Eine Veränderung der verschiedenen Filter ist nicht möglich. Wenn Filter verändert werden sollen, geschieht dies auf Basis der „normalen“ Profilbildungsschnittstelle.

In der Webschnittstelle wird auf der linken Seite, ähnlich wie in der Profilbildungsschnittstelle (vgl. Abb. 8.17), der Profilbaum angezeigt. Eine Entsprechung für die angelegten Dossiers in der Profilbildungsschnittstelle existiert in der Webschnittstelle nicht. Die rechte Seite wird jeweils von einer Tabelle eingenommen. Die angezeigten Spalten entsprechen den in der Profilbildungsschnittstelle ausgewählten Spalten. Eine Veränderung der angezeigten Spalten ist in der Webschnittstelle nicht möglich. Von den Tabellen kann zu Detailansichten von Informationseinheiten verzweigt werden. Die Detailansichten entsprechen den Detailansichten der Profilbildungsschnittstelle.

8.8 Architektur

Broker's Lounge ist als 3-Schichten-Architektur konzipiert (vgl. Abb. 8.25) und wurde vollständig in Java¹² implementiert. Die unterste Schicht bildet eine objektorientierte Datenbank, die über einen Datenbank-Server zugänglich ist. Als Datenbank wurde POET 5.1 verwendet. Über der Datenbankschicht liegt die Applikationsschicht. Die Applikationsschicht beinhaltet alle zentralen Komponenten

¹²<http://java.sun.com/>

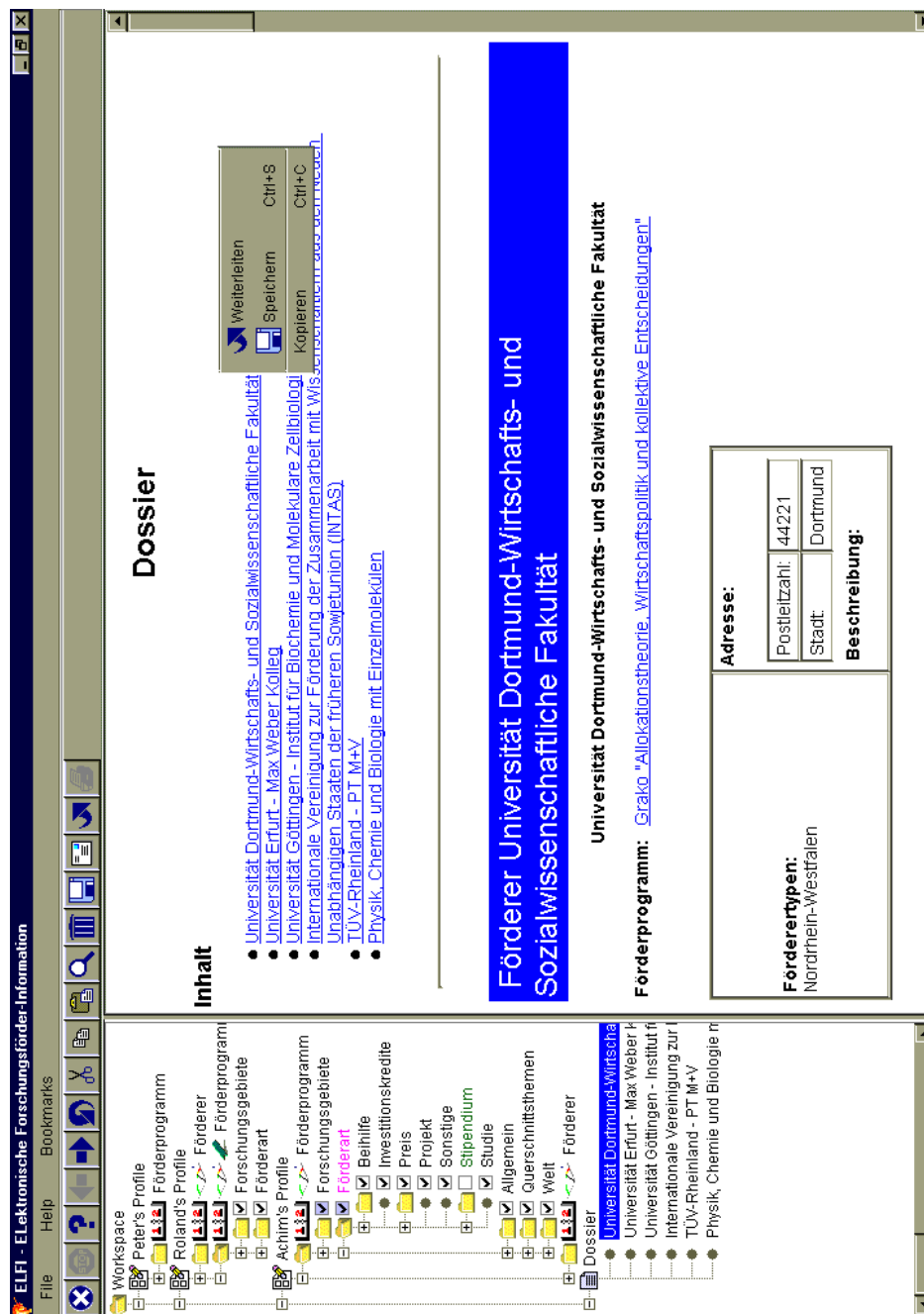


Abbildung 8.23: Beispiel-Dossier in der Domäne Forschungsförderung

zum Betrieb von Broker's Lounge. Die Client-Schicht kommuniziert mit der Applikationsschicht über CORBA. Datenbank- und Applikationsschicht sind in der Regel beim Betreiber von Broker's Lounge angesiedelt, während die Clientschicht

The screenshot shows a Netscape browser window with the address bar displaying 'http://localhost:8080/brokerslounge/land/elfi/brokerslounge'. The main content area is titled 'ELFI - Servicestelle für Elektronische Forschungsförder-Informationen' and displays a table of six funding programs under the heading '6 Förderprogramme zu Pädagogik'.

Name	Teilprogramme	Oberprogramme	Forschungsthemen	Deadline	Person	Förderer	Fördervolumen	Status	Relevanz
Körber-Programme on Historical Culture and Politics in Europe			Geschichte, Pädagogik			Körber Stiftung		New	17
Promotionskolleg "Innovation von und in Organisationen"			Arbeitswissenschaft, Pädagogik, Sozialpsychologie, Arbeits- Betriebs- und Organisationspsychologie, Soziologie		Müssen, Heiner, Prof. Dr.	Hans Böckler Stiftung, Universität Bochum - Institut für Arbeitswissenschaft			12
VWS - Neue Informations- und Kommunikationstechniken in Wirtschaft, Medien und Gesellschaft			Informatik, Pädagogik, Rechtswissenschaft, Sozialwissenschaften, Wirtschaftswissenschaften		Junkers, Helga, Dr.	Volkswagen-Stiftung		New	7
MSWWF-NRW-Qualität der Lehre			Pädagogik			Ministerium für Schule, Wissenschaft und Forschung NRW (MSWF)		New	5
EU - IST Forschungsinfrastruktur- Vernetzung			Elektrotechnik, Erziehungswissenschaft und Bildungsforschung, Informatik, Nachrichtentechnik, Praktische Informatik, Theoretische Informatik			Rosenberg, Europäische Kommission	161 Mio. €	New	5
EU - IHP C.2.2. Archimedes-Preis			Informatik, Lern-Lehr-Forschung, Didaktik und Hochschuldidaktik, Mathematik, Nachrichtentechnik, Wissenschaft von der Politik		Stramm, Kathrin	Fulbright-Kommission	durchschnittlich 50.000 €	New	4

At the bottom of the browser window, there is a 'filter's profile' section with a tree view of categories like 'Achims Profil', 'Rolands Profil', 'Förderer', 'Förderprogramm', 'Forschungsgebiete', etc. To the right of the table, there is a 'Push-Dienst' section with a 'Letzte Mail:' field showing '01.01.70' and a 'Nächste Mail:' field with radio button options for 'täglich', 'wöchentlich', 'monatlich', and 'keine Email'. A 'Speichern' button is also present.

Abbildung 8.24: Profil-Webschnittstelle - Übersichtsliste

beim Nutzer von Broker's Lounge ausgeführt wird. Damit ein Benutzer Broker's Lounge benutzen kann, muss er einen entsprechenden Broker's Lounge-Client auf seinem Computer installieren. Tut er dies nicht, hat er lediglich Zugriff auf die

Webschnittstellen mit verringertem Funktionsumfang, die in Kap. 8.6 und Kap. 8.7.11 beschrieben wurden. Der Broker's Lounge-Client ist als Java-Applikation konzipiert, der sich über CORBA mit dem Applikationsserver in der Applikationsschicht verbindet. Als CORBA-Implementierung wurde Visibroker 3.4 von Inprise benutzt.

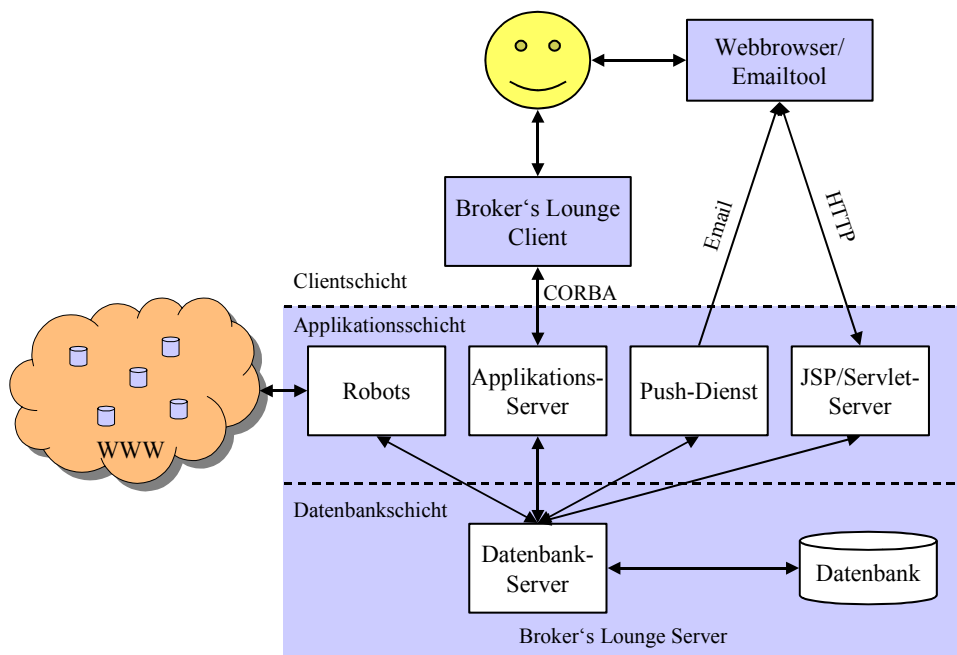


Abbildung 8.25: Architektur von Broker's Lounge

Der Applikationsserver übersetzt CORBA-Aufrufe des Clients in Datenbank-Transaktionen, die er an den Datenbank-Server weiterleitet. Durch die Verwendung von CORBA kann Broker's Lounge auch weitere Clients unterstützen, die in anderen Programmiersprachen geschrieben wurden, wenn diese Clients das CORBA-Protokoll implementieren.

Die Webschnittstellen werden über einen Webserver angeboten, der sich wie der Applikationsserver mit der Datenbankschicht verbindet. Die Webschnittstellen von Broker's Lounge wurden als Java-Servlets¹³ in Verbindung mit Java-Server-Pages (JSP)¹⁴ realisiert. Bei dem verwendeten Webserver handelt es sich daher um einen Servlet- und JSP-Container, wie er von Sun Microsystems spezi-

¹³Das Whitepaper zu Java-Servlets von Sun Microsystems ist unter <http://java.sun.com/products/servlet/whitepaper.html> verfügbar

¹⁴Das Whitepaper zu Java Server pages von Sun Microsystems ist unter <http://java.sun.com/products/jsp/whitepaper.html> verfügbar

fiziert wurde. Als Webserver wurde Tomcat 4.0¹⁵ verwendet.

Die Robots zur Beobachtung von verschiedenen Quellen werden von der Robotkomponente gesteuert. Diese Robotkomponente fügt die gefundenen Dokumente in das Dokumentenarchiv ein und führt anschließend die inhaltliche Analyse mit Hilfe des IB-Domänenmodells durch.

Der Push-Dienst wird von einer weiteren Komponente durchgeführt. Diese Push-Dienst-Komponente verbindet sich in regelmäßigen Abständen zum Datenbank-Server und überprüft die Profile aller Benutzer, die den Push-Dienst abonniert haben, auf veränderte und neue Informationseinheiten, die durch die definierten Filter passen. Werden solche Informationseinheiten gefunden, erhält der Benutzer eine Email, die Verweise auf alle diese Informationseinheiten enthält. Diese Verweise führen zu den Detailansichten der entsprechenden Informationseinheiten in der Webschnittstelle.

Die weiteren Komponenten der Applikationsschicht (Robots, Push-Dienst und Webserver) verbinden sich aus Effizienzgründen direkt zum Datenbank-Server. In der Regel werden die weiteren Komponenten vom Betreiber von Broker's Lounge auf eigenen Computern betrieben. Für alle Komponenten, die mit dem Datenbank-Server verbunden sind, ist eine Client-Lizenz für diese Datenbank nötig. Dagegen entfällt diese Lizenz für alle Clients, die über den Applikationsserver mit der Datenbank kommunizieren. Ggf. könnten die weiteren Komponenten (Robots, Push-Dienst und Webserver) sich auch über CORBA zum Applikationsserver verbinden. In der Domäne Forschungsförderung z.B. könnte jede Universität ihren eigenen Webserver betreiben, über den sie ihren Mitarbeitern Informationen über Forschungsförderung anbietet.

Alle Komponenten mit Ausnahme der Datenbank sind in Java implementiert. Die Datenbank ist auf allen gängigen Plattformen mit Ausnahme von MacOS verfügbar. Das bedeutet, dass der Broker's Lounge-Server, zu dem die Datenbank, der Datenbank-Server und alle Komponenten innerhalb der Applikationsschicht (vgl. Abb. 8.25) gehören, auf allen gängigen Plattformen außer MacOS verfügbar ist. Der Broker's Lounge-Client kann auf allen java-fähigen Plattformen betrieben werden.

¹⁵Tomcat ist verfügbar unter <http://jakarta.apache.org/tomcat/>

Kapitel 9

Anwendungsbeispiele

In diesem Kapitel werden ausgewählte Anwendungen von Broker's Lounge in verschiedenen Domänen mit jeweils verschiedenen Brokeringmodellen beschrieben, um die Unabhängigkeit von Broker's Lounge hinsichtlich der Domäne und des gewählten Brokeringmodells zu demonstrieren.

Für die drei ausgewählten Anwendungen ELFI (vgl. Kap. 9.1), MarketMonitor (vgl. Kap. 9.2) und ScienceLounge (vgl. Kap. 9.3) wurden jeweils Brokering-Systeme auf der Basis von Broker's Lounge realisiert. ELFI bearbeitet die Domäne Forschungsförderung und informiert über Forschungsfördermöglichkeiten. ELFI wird von der Servicestelle für Forschungsförderinformation¹ an der Ruhr-Universität Bochum betrieben. Alle Beispiele, die in dieser Arbeit in den vorangegangenen Kapitel beschrieben wurden, stammen aus der ELFI-Domäne. In ELFI wird mit Informationseinheiten über Förderprogramme, Förderer und Ansprechpartnern von Förderprogrammen gehandelt. D.h., die Kunden von ELFI erhalten Informationseinheiten als Ergebnis ihrer Anfragen.

MarketMonitor wird zur Markt- und Konkurrenzbeobachtung in der Stahldomäne eingesetzt. Im Unterschied zu ELFI wird in MarketMonitor nicht mit Informationseinheiten gehandelt. Stattdessen bekommen die Konsumenten (von der inhaltlichen Analyse) annotierte und bewertete Dokumente vom System geliefert. Der Grund dafür liegt darin, dass bei der Markt- und Konkurrenzbeobachtung die zeitnahe Information der Konsumenten von besonderer Bedeutung ist. Diese zeitnahe Information kann bei einer manuellen Aufbereitung nicht immer gesichert werden.

ScienceLounge ist eine Selbstanwendung der Gruppe, die Broker's Lounge entwickelt hat. Sie dient der Beobachtung des Ausschnitts der Scientific Community, der für Broker's Lounge und die daraus entstehenden wissenschaftlichen Arbeiten als relevant bewertet wurde. Das Brokeringmodell von ScienceLounge

¹<http://www.elfi.ruhr-uni-bochum.de/>

ist eine Mischung aus den Modellen, die bei ELFI und MarketMonitor angewendet wurden.

Die drei Anwendungsbeispiele werden in Kapitel 9.4 abschließend miteinander verglichen.

9.1 ELFI

ELFI wird seit Ende 1997 als Informationsdienst für Forschungsförderung von der Servicestelle für Forschungsförderinformation betrieben. Die Nutzung des Dienstes war zunächst kostenlos. Seit Anfang 2002 ist die Nutzung des Dienstes kostenpflichtig. Ende Januar 2002 hatten ca. 24% aller deutschen Universitäten diesen Dienst für ihre Wissenschaftler abonniert.

Die Servicestelle wurde im Rahmen eines Pilotprojektes errichtet. Das Pilotprojekt wurde vom DFN-Verein (Verein zur Förderung des Deutschen Forschungs-Netzes) aus Mitteln des BMBF gefördert. Im Rahmen des Pilotprojektes wurde die Software zum Betrieb der Servicestelle von der Ruhr-Universität und dem Institut für Angewandte Informationstechnik der GMD² erstellt. An das Pilotprojekt schloss sich ein weiteres Projekt ELFI-PRO an, das die existierende ELFI-Software durch eine neue Software ersetzen sollte, die leichter zu betreiben sein sollte. Im Rahmen von ELFI-PRO³ und weiteren Projekten ist Broker's Lounge entstanden. Die Aufteilung des IB-Domänenmodells in Konzepte und Kategorien greift Erfahrungen aus dem ELFI-Projekt auf und verallgemeinert sie. Aus diesem Grund werde ich im nächsten Abschnitt (Kap. 9.1.1) die Entstehung des ELFI-Projektes sowie die Grundzüge des „alten“ ELFI-Systems beschrieben. Zur Zeit stellt die Servicestelle den Betrieb auf Broker's Lounge-Software um.

Der Anstoß zum ELFI-Projekt ging von den Forschungsreferenten der deutschen Universitäten aus (vgl. [Aea96]). Forschungsreferenten sind Informations-Broker, die an den verschiedenen Universitäten angestellt sind. Ihre Aufgabe ist die Information der Wissenschaftler ihrer Institution über Fördermöglichkeiten sowie deren Beratung und aktive Unterstützung bei der Antragsstellung. In der Praxis können viele Forschungsreferenten ihrem Beratungsauftrag kaum nachkommen, da sie zuviel Zeit benötigen, um sich selbst über Fördermöglichkeiten zu informieren.

Dies wurde durch eine Umfrage unter den deutschen Forschungsreferenten bestätigt, die zu Beginn des ELFI-Projektes durchgeführt wurde. Im Rahmen dieser Umfrage wurden Fragebögen an alle deutschen Hochschulen (außer theolo-

²In der Zwischenzeit ist die GMD in der Fraunhofer-Gesellschaft aufgegangen. Das Institut für Angewandte Informationstechnik heißt nun Fraunhofer-Institut Angewandte Informationstechnik (FhG-FIT)

³ELFI-PRO wurde ebenfalls vom DFN aus Mitteln des BMBF gefördert.

gischen Hochschulen und Kunsthochschulen) und Großforschungseinrichtungen versendet. Der Rücklauf belief sich auf ca. 40%. Es ergab sich unter anderem,

- dass Forschungsreferenten nur ein Viertel ihrer Zeit für die Beratung ihrer Wissenschaftler aufwenden können, da sie den Großteil ihrer Zeit für die Informationssuche (51%) und die Information ihrer Wissenschaftler aufwenden,
- dass sie oft Informationen nicht rechtzeitig erhalten und diese nicht umfangreich genug sind,
- dass sie aufbereitete Informationen gegenüber nicht aufbereiteten Informationen vorziehen,
- dass sie fast alle über alle Förderer informieren und beraten und
- dass sie oft keinen Überblick über die Forschungsinteressen ihrer Wissenschaftler haben und deswegen eine punktgenaue Information nur selten möglich ist, statt dessen erstellt ein Großteil (80%) einen Rundbrief, der an alle Wissenschaftler verteilt wird.

Eine Umfrage zu Beginn des ELFI-PRO-Projektes (Anfang 1999) hat ergeben, dass das ELFI-System und die damit verbundene Dienstleistung insgesamt positiv bewertet werden. Während ELFI sich zu Beginn ausschließlich an Forschungsreferenten wendete, wurde der Dienst später auch für Wissenschaftler zugänglich gemacht.

Durch die Etablierung der Servicestelle hat sich die Rolle der Forschungsreferenten verändert. Die Zeit, die für die Suche und Aufbereitung von relevanten Informationen aufgewendet werden muss, ist deutlich geringer geworden, da diese Aufgabe nun von der Servicestelle übernommen wird. Es verbleibt dadurch mehr Zeit für die Beratung von Wissenschaftlern. Da Wissenschaftler sich außerdem mit dem ELFI-System über Fördermöglichkeiten informieren können, wird die Bedeutung von Forschungsreferenten als Multiplikator von Fördermöglichkeiten geringer.

9.1.1 ELFI-Kontrollfluss und Brokeringmodell

Die ursprüngliche ELFI-Software beobachtet die Webserver der verschiedenen Förderer. Dies geschieht wie in Broker's Lounge (vgl. Abb. 8.1) auf der Basis von Webrobots. Die Webrobots liefern als Ergebnis neue und veränderte Dokumente (vgl. Abb. 9.1). Diese neuen und veränderten Dokumente werden mit dem sog. Mastertool kategorisiert.

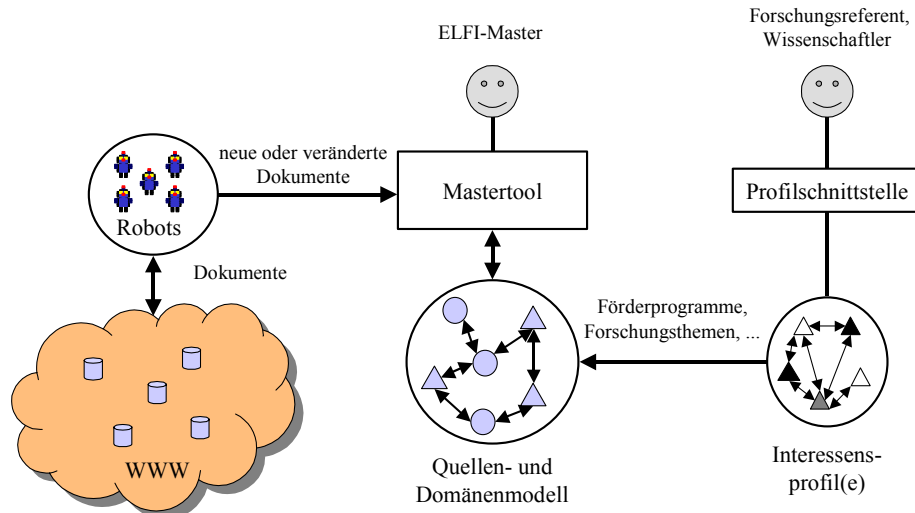


Abbildung 9.1: Kontrollfluss in der ELFI-Servicestelle

Das Mastertool umfasst alle Quellen- und Domänenmodellverwaltungsaufgaben. Das Mastertool ist als monolithischer Block konzipiert. Eine Teilung der Aufgaben wie in Broker's Lounge ist nicht möglich. Das Mastertool wird von einem sog. ELFI-Master bedient und arbeitet auf einem Quellen- und Domänenmodell. In diesem Quellen- und Domänenmodell sind Dokumente und Domäneninhalte eng verzahnt. Das Quellen- und Domänenmodell ist als vertikales System konzipiert, das auf die Domäne Forschungsförderung abgestimmt wurde. D.h., das Quellen- und Domänenmodell kennt keine Konzepte und Kategorien, sondern besteht aus Förderprogrammen, Förderern, Forschungsthemen, Förderarten usw. Eine Erweiterung des Domänenmodells ist nicht ohne weitere Programmierung möglich.

Forschungsreferenten und Wissenschaftler haben Zugang zum ELFI-System über eine Personalisierungsschnittstelle (vgl. Abb. 9.2). Diese Schnittstelle ist als Java-Applet realisiert. Die Benutzer können ihr Profil jederzeit ändern, indem sie die entsprechenden Haken an den verschiedenen Kategorien löschen oder hinzunehmen. Die Funktionalität des Java-Applets entspricht im Wesentlichen der Funktionalität der Personalisierungsschnittstelle von Broker's Lounge. Im Gegensatz zu Broker's Lounge ist das alte ELFI-System ein rein adaptierbares System, da die Benutzer vom System keinerlei Unterstützung bei der Verwaltung ihrer Interessen bekommen. Die Erfahrungen im ELFI-Projekt und die Umfrage zu Beginn des ELFI-PRO-Projektes haben gezeigt, dass die Benutzer einen möglichst einfachen Zugang zum System wünschen und Schwierigkeiten beim Erler-

The screenshot shows a Netscape browser window titled "ELFI - Personal View - Netscape". The address bar shows "http://www.elfi.uni-bochum.de/elfi_1_0/". The browser interface includes a menu bar (File, Edit, View, Go, Communicator, Help), a toolbar with icons for bookmarks, messages, and other functions, and a status bar at the bottom indicating "Applet EWS running".

The main content area is divided into two sections. The left section, titled "Elfi Profile Ansicht", shows a profile for "Achim Nick, 31.8.19100" with the status "Modus: Aktuelles eigenes Profil". Below this is a tree view of research areas with checkboxes:

- ☒ Forschungsgebiete
- ☒ Geistes- und Gesellschaftswissenschaften
- ☒ Medizin, Biologie, Agrarwissenschaften
- ☒ Naturwissenschaften
- ☒ Ingenieurwissenschaften
- ☒ Hydrologie und Wasserwirtschaft
- ☒ Allgemeine Ingenieurwissenschaften
- ☒ Architektur, Städtebau und Landesplanung
- ☒ Bauingenieurwesen
- ☒ Bergbau und Hüttenwesen
- ☒ Elektrotechnik
- ☒ Informatik
- ☒ Theoretische Informatik
- ☒ Praktische Informatik
- ☒ Technische Informatik
- ☒ Maschinenbau und Produktionstechnik
- ☒ Maschinenbau und Verfahrenstechnik

The right section, titled "20 allgemeine und 8 speziellere Förderprogramme zu Informatik", contains a table with the following columns: Förderer, Förderprogramm, and Deadline.

Förderer	Förderprogramm	Deadline
BMBF	BMBF-Dienstleistungen für die Zukunft	unbekannt
BMBF	BMBF-Zusammenarbeit mit Chile	unbekannt
DFG	DFG-SP - Algorithmen großer und komplexer Netzwerke	1. November 2000
DFG	DFG-SP - Kooperierende Teams mobiler Roboter in dynamischen Umgebungen	20. Oktober 2000
DFG	DFG-SP - Medizinische Navigation und Robotik	1. April 2001
DFH	DFH-Sipendium Frankreich	unbekannt
DFN Verein	DFN Verein - Entwicklungsprogramm	unbekannt
ESF-PhD	ESF-PhD studentship	unbekannt
GI	GI-Dissertationspreis	unbekannt
Gerda Tschura Stiftung	Gerda Tschura Stipendium für alleinerziehende Mütter und Väter	unbekannt
Grako	Grako - Effiziente Algorithmen und Mehrskalermethoden	31. Januar 1998
Grako	Grako-Anwendungen der Supraleitung	unbekannt
Grako	Grako Betriebswirtschaftliche Aspekte lose gekoppelter Systeme im Zeitalter neuer Medien	offen
Grako	Grako - Aufgabenorientierte Kommunikation	offen
Grako	Grako - Kognitionswissenschaft	unbekannt
Infincon	Infincon Stipendien für M. Sc. Programm Microelectronics	unbekannt
Innovationspreis	Innovationspreis Klinische Informationstechnologie	23. Oktober 2000

Abbildung 9.2: Personalisierungsschnittstelle des „alten“ ELFI-Systems

nen der Personalisierungsschnittstelle hatten. Aus diesem Grund bietet Broker's Lounge mit der Profil-Webschnittstelle einen einfachen Zugang für diese Benut-

zer an. Da die ursprüngliche Interaktionsform ein sehr mächtiges Werkzeug zur Beschreibung von Interessen bietet, wurde diese Interaktionsform in der Personalisierungsschnittstelle des Broker's Lounge-Clients übernommen. Die Profil-Webschnittstelle wendet sich an den gelegentlichen Benutzer, während die Personalisierungsschnittstelle des Broker's Lounge-Clients für häufige Benutzer konzipiert wurde.

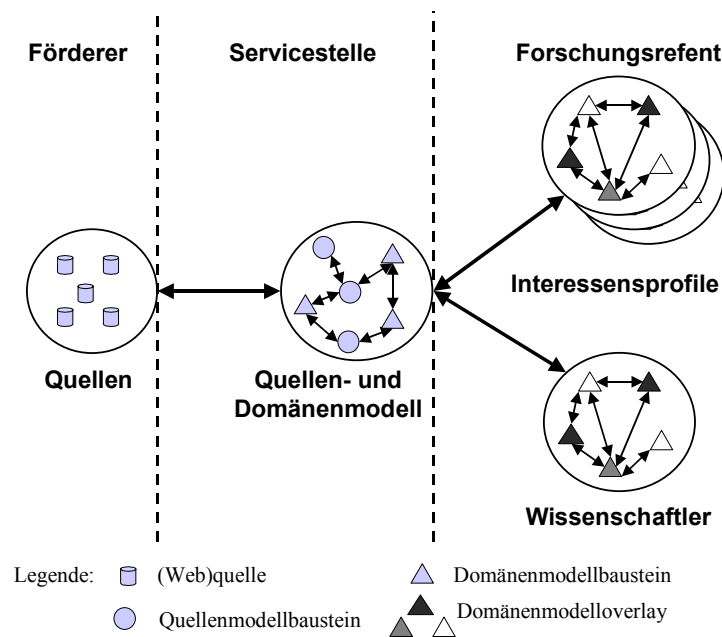


Abbildung 9.3: ELFI-Brokeringmodell

Aus dem Kontrollfluss in der Servicestelle und der verwendeten Software ergibt sich das Brokeringmodell in Abbildung 9.3. Die Domänen- und Quellenverwaltungsaufgaben werden von der Servicestelle übernommen. Die kundenorientierten Aufgaben werden vollständig von den Kunden der Servicestelle ausgeführt. Die Kunden der Servicestelle können in Forschungsreferenten und Wissenschaftler unterschieden werden. Wissenschaftler verwalten ihr eigenes Interessenprofil und sind die Endkonsumenten der aufbereiteten Informationen über Förderprogramme und Förderer. Forschungsreferenten benutzen Interessenprofile, um die Interessen von Wissenschaftlern (oder Gruppen von Wissenschaftlern) abzubilden. Forschungsreferenten sind in diesem Fall nicht die Endkonsumenten von Informationen, sondern fungieren als lokale Broker, die ihren Kunden, den Wissenschaftlern ihrer Institution, die Abbildung ihrer Interessen auf die ELFI-Inhalte abnimmt.

9.1.2 Brokeringmodell von ELFI mit Broker's Lounge

Bei der Implementierung des ELFI-Services auf Basis von Broker's Lounge kann der Kontrollfluss innerhalb der Servicestelle beibehalten werden, da der bestehende Kontrollfluss dem in Abbildung 8.1 beschriebenen Kontrollfluss von Broker's Lounge entspricht.

Der Kontrollfluss in der Servicestelle wird um die inhaltliche Analyse von neuen oder veränderten Dokumenten erweitert. Die Aufteilung der verschiedenen Quellen- und Domänenverwaltungsaufgaben auf die Schnittstellen zur Quellen- und Domänenmodelladministration und den Dokumentenbrowser erlaubt eine Entflechtung der verschiedenen Aufgaben. Diese Aufgaben könnten auch räumlich verteilt ausgeführt werden. Zum Beispiel könnten die Ansprechpartner von Förderprogrammen ihre Förderprogramme selbst in das IB-Domänenmodell eintragen. Außerdem könnten die Webrobots für die Websites von Förderern von den Administratoren dieser Websites in das Quellenmodell eingefügt werden. Broker's Lounge bietet außerdem einen Push-Dienst an, der benutzt werden kann, um Wissenschaftler und Forschungsreferenten vom System automatisch über Änderungen in ihren Profilen zu informieren.

Broker's Lounge ist als adaptives und adaptierbares System konzipiert. D.h., die Benutzer der Profilbildungsschnittstelle bekommen eine Hilfestellung bei der Erstellung und der Verwaltung ihrer Profile. Dies hat Rückwirkungen auf das Brokeringmodell. Der kundenorientierte Prozess wird nun von Wissenschaftlern und Forschungsreferenten in Zusammenarbeit mit dem System ausgeführt. Die Beziehung zwischen den Förderern und der Servicestelle bleibt zunächst unverändert, wenn man davon absieht, dass die Förderer ihre Förderprogramme selbst eintragen könnten. Die Quellenverwaltungsaufgaben werden von der Servicestelle ausgeführt. Auch die Kategorisierung und mithin die Domänenverwaltungsaufgaben werden von der Servicestelle durchgeführt. Es ergibt sich das Brokeringmodell, das in Abbildung 9.4 abgebildet ist.

9.1.3 ELFI-Domänenmodell

Das IB-Domänenmodell von ELFI wurde in Ausschnitten in dieser Arbeit zur Verdeutlichung von einzelnen Sachverhalten benutzt. Das ELFI-Domänenmodell besteht zur Zeit aus vier verschiedene Konzepte:

1. Förderprogramme beschreiben Fördermöglichkeiten,
2. Förderer bezeichnen die Zuwendungsgeber,
3. Ansprechpartner repräsentieren die Kontaktpersonen für Förderprogramme und

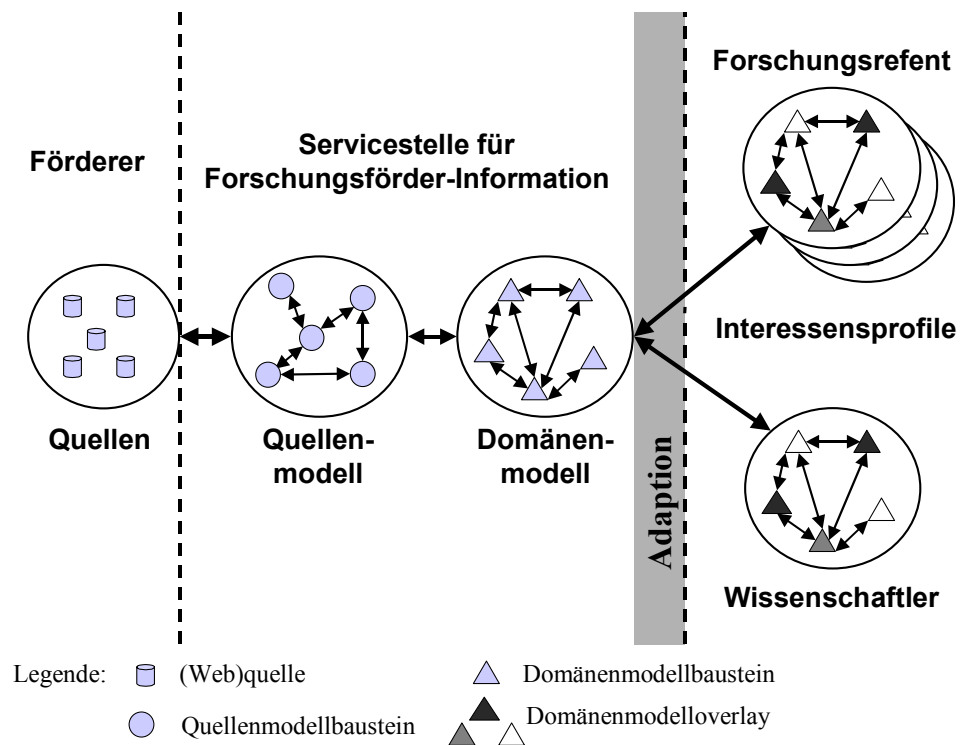


Abbildung 9.4: Brokeringmodell von ELFI bei der Umsetzung mit Broker's Lounge

4. Dokumente enthalten Zusatzinformationen zu einzelnen Web-Dokumenten.

Relationen zwischen Konzepten

Förderprogramme, Förderer und Ansprechpartner haben jeweils eine Relation zu Dokumenten. Förderprogramme weisen zusätzlich eine Relation zu Förderern, und Ansprechpartnern auf. Förderer haben eine entsprechende Relation zu Förderprogrammen und sind mit Ansprechpartnern assoziiert, die bei ihnen beschäftigt sind. Die Ansprechpartner weisen eine spiegelbildliche Verbindung zu den Förderern auf, bei denen sie beschäftigt sind und zu den Förderprogrammen, für die sie verantwortlich sind.

Förderprogramme sind mit anderen Förderprogrammen über eine PartOf-Relation verbunden. Dadurch werden Teil- und Oberprogramme von Förderprogrammen abgebildet.

Relationen zwischen Konzepten und Kategorietypen

Förderer werden durch den Kategorietypp Förderertypen klassifiziert. Förderertypen sind z.B. nationale Stiftungen, Bundesministerien oder internationale Förderer. Dokumente werden durch ihren Dokumententyp klassifiziert. Dokumententypen sind z.B. Ausschreibungen und Formulare. Förderprogramme werden mit insgesamt fünf Kategorietypen klassifiziert:

1. Forschungsthemen beschreiben die Themen, die von einem Förderprogramm gefördert werden sollen. Die Forschungsthemenhierarchie orientiert sich an der Fachgutachterwahlliste der DFG.
2. Querschnittsthemen beschreiben interdisziplinäre Themen, die orthogonal zu den Forschungsthemen liegen. Beispiele für Querschnittsthemen sind „Umwelt“ und „Informationsgesellschaft“.
3. Förderarten beschreiben die Art der Förderung des Förderprogramms. Typische Förderarten sind „Stipendium“ und „Projektförderung“.
4. Institutionsarten bezeichnen die Institutionen, die zur Förderung zugelassen werden. Institutionsarten sind z.B. „Forschungseinrichtungen“ und „Kleine und mittlere Unternehmen (KMU)“.
5. Regionale Zulassungsbeschränkungen zu Förderprogrammen werden durch den Kategorietypp Regionen abgedeckt. Beispiele für Regionen sind „Deutschland“ oder „Rheinland-Pfalz“.

Attribute von Konzepten

Förderer und Ansprechpartner sind jeweils mit einer Adresse assoziiert. Adressen setzen sich aus einer Reihe von alphanumerischen Attributen zur Beschreibung von Anschrift, Telefon- und Faxnummern zusammen. Förderprogramme sind mit einem Attribut zur Beschreibung der Deadline für die Einreichung von Projektvorschlägen und zur Beschreibung des ausgeschriebenen Fördervolumens verbunden.

Die Relationen zwischen Konzepten, Kategorietypen und Attributen sind in Tabelle 9.1 zusammengefasst.

9.2 MarketMonitor

MarketMonitor ist eine Implementierung von Broker's Lounge zur Markt- und Konkurrenzbeobachtung in der Stahldomäne. MarketMonitor wurde von einem

	Förderprogramm	Förderer	Ansprechpartner	Dokument
Förderprogramm	x	x	x	x
Förderer	x	-	x	x
Ansprechpartner	x	x	-	x
Dokument	x	x	x	-
Förderertypen	-	x	-	-
Dokumententypen	-	-	-	x
Forschungsthemen	x	-	-	-
Querschnittsthemen	x	-	-	-
Förderarten	x	-	-	-
Institutionsarten	x	-	-	-
Regionen	x	-	-	-
Adresse	-	x	x	-
Deadline	x	-	-	-
Fördervolumen	x	-	-	-

Tabelle 9.1: Relationen zwischen Konzepten, Kategorietypen und Attributen in ELFI

Stahlunternehmen zu diesem Zweck eingesetzt. Anders als bei ELFI stehen bei MarketMonitor nicht die aufbereiteten Informationseinheiten des IB-Domänenmodells, sondern die gewichteten und mit Treffern aus dem IB-Domänenmodell annotierten Dokumente im Mittelpunkt des Interesses.

Die Rolle des IB-Domänenmodells ändert sich dadurch. Das IB-Domänenmodell wird in MarketMonitor nicht zur Sammlung aller relevanten Informationseinheiten benutzt, für die sich Benutzer interessieren könnten. Stattdessen wird es genutzt, um das Bild des Unternehmens von der zu beobachtenden Domäne (in diesem Fall der Stahlmarkt) abzubilden. Für die einzelnen Einträge können Synonyme angegeben werden, um die Wahrscheinlichkeit von Treffern innerhalb der inhaltlichen Analyse zu erhöhen. Um die unterschiedlichen Bereiche des IB-Domänenmodells unterschiedlich bewerten zu können, kann jeder Eintrag im IB-Domänenmodell mit einer Gewichtung versehen werden. Diese Gewichtung fließt in die Bewertung einzelner Dokumente ein (vgl. Kap. 8.6).

Der Dokumentenbrowser wird als Schnittstelle zu den Endbenutzern von MarketMonitor benutzt. Dort können Benutzer einen Überblick über die relevantesten Dokumente in ihrer Domäne bekommen. Der Dokumentenbrowser kann auch weiterhin zur Kategorisierung von Dokumenten benutzt werden. Die Kategorisierung von Dokumenten ist in MarketMonitor von geringerer Bedeutung, da eine Erweiterung des IB-Domänenmodells mit neuen Informationseinheiten nur dann

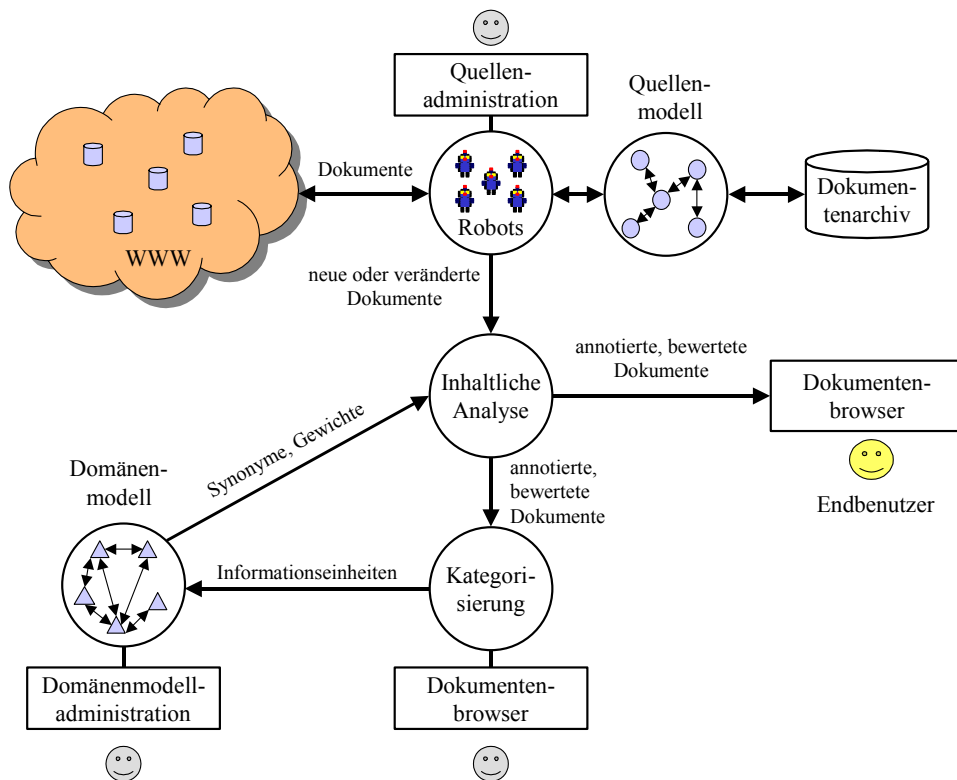


Abbildung 9.5: Kontrollfluss in MarketMonitor

Sinn macht, wenn diese Informationseinheiten die veränderte Sicht des Unternehmens auf die zu beobachtende Domäne reflektiert. Dies kann z.B. der Fall sein, wenn ein neuer Wettbewerber in den Markt eintritt oder ein neues Produkt angeboten wird.

Da bei MarketMonitor mit annotierten und gewichteten Dokumenten und nicht mit den Inhalten des IB-Domänenmodells gehandelt wird, verzichtet MarketMonitor auf die Profilbildungsschnittstelle. Benutzer können stattdessen im Dokumentenbrowser einzelne Anfragen abspeichern, um diese Anfragen später erneut zu benutzen.

Das Brokeringmodell von MarketMonitor (vgl. Abb. 9.6) ist ein klassisches Pull-Brokermodell im Sinne von [SMDP98], da alle relevanten Aufgaben vom Unternehmen ausgeführt werden, das MarketMonitor betreibt. Innerhalb des Unternehmens kann zwischen einem internen Broker und internen Informationskonsumenten unterschieden werden. Die Brokeringdienstleistung wird im Unterschied zu ELFI im Wesentlichen auf automatischer Basis vom System durch die inhaltliche Analyse durchgeführt. Ggf. wird das IB-Domänenmodell verän-

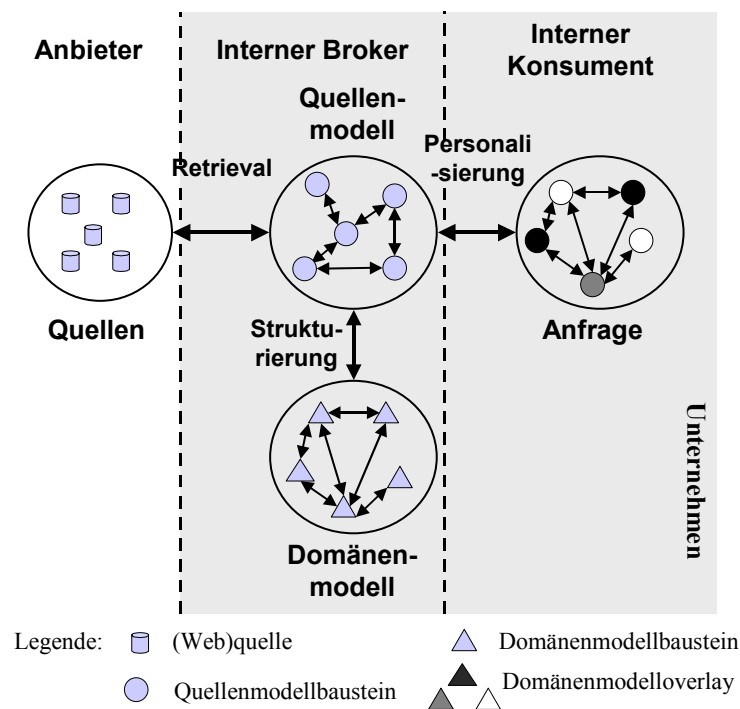


Abbildung 9.6: MarketMonitor-Brokeringmodell

dert, um eine veränderte Sicht auf die Domäne abzubilden. Die Abbildung der Interessen von Konsumenten wird von diesen selbst durchgeführt. Die Interessen werden nicht auf das IB-Domänenmodell, sondern auf das Quellenmodell abgebildet, indem die Konsumenten mit Hilfe des Dokumentenbrowsers Anfragen an das System stellen.

9.2.1 MarketMonitor-Domänenmodell

Das IB-Domänenmodell von MarketMonitor basiert auf drei Konzepten:

1. Organisationen beschreiben alle relevanten Organisationen, die im beobachteten Markt aktiv sind.
2. Produkte beschreiben alle Produkte, die im Markt angeboten werden und vom Unternehmen, das MarketMonitor betreibt, als relevant erachtet werden.
3. Personen beschreiben alle relevanten Personen, die in einem Markt aktiv sind.

Relationen zwischen Konzepten

Die Aufgabe von MarketMonitor ist die Beobachtung der Akteure, die auf dem zu beobachtenden Markt aktiv sind. Diese Akteure sind im IB-Domänenmodell (von MarketMonitor) durch Organisationen abgebildet. Organisationen weisen Relationen zu Produkten auf, die von der jeweiligen Organisation angeboten werden. Außerdem besteht eine Beziehung zwischen Personen und Organisationen.

Relationen zwischen Konzepten und Kategorietypen

Produkte werden durch den Kategorietypp Geschäftsfelder klassifiziert. Die Hierarchie der Geschäftsfelder in MarketMonitor bildet nicht alle möglichen Geschäftsfelder ab, sondern reflektiert die Sicht des Unternehmens, das MarketMonitor betreibt, auf den beobachteten Markt. Ein Beispiel für ein Geschäftsfeld ist „Rost-Säure-Hitze beständiger Stahl“.

Diese Geschäftsfelder-Hierarchie kann ebenfalls zur Klassifikation von Organisationen benutzt werden. Die Geschäftsfelder repräsentieren in diesem Fall die Geschäftsfelder, in denen eine Organisation aktiv ist. Organisationen können außerdem durch ihren Organisationstyp klassifiziert werden. Beispiele für Organisationstypen sind „Firma“ oder „Verband“.

Personen können durch ihre Rolle im Unternehmen klassifiziert werden. Ein Beispiel für eine Rolle ist „CEO“.

Zusammenfassung des IB-Domänenmodells

Die Relationen zwischen den verschiedenen Konzepten und Kategorietypen sind in Tabelle 9.2 zusammengefasst. Neben den in den vorausgegangenen Abschnitten beschriebenen Konzepten und Kategorietypen haben Organisationen und Personen jeweils ein Attribut, das die Anschrift der Organisation bzw. der entsprechenden Person beschreibt. Dies ist ebenfalls in Tabelle 9.2 berücksichtigt.

9.3 ScienceLounge

ScienceLounge ist eine Selbstanwendung von Broker's Lounge und dient der Beobachtung der Scientific Community in den Bereichen Kontext, ontologiebasierter und adaptiver Informationssysteme. Mit Hilfe von ScienceLounge können Wissenschaftler, Konferenzen und Publikationen mit wenig Aufwand beobachtet werden.

In ScienceLounge wird sowohl mit Informationseinheiten als auch mit annotierten Dokumenten gehandelt. Informationseinheiten beschreiben z.B. Veröffentlichungen von Wissenschaftlern, Konferenzen. Interessante Dokumente können

	Organisation	Produkt	Person
Organisation	-	x	x
Produkt	x	-	-
Person	x	-	-
Geschäftsfelder	x	x	-
Organisationstypen	x	-	-
Rollen	-	-	x
Anschrift	x	-	x

Tabelle 9.2: Relationen zwischen Konzepten, Kategorietypen und Attributen in MarketMonitor

z.B. Publikationslisten von Wissenschaftlern sein. Das gehandelte Gut (Informationseinheiten und Dokumente) vereinigt die Ansätze miteinander, die in ELFI und in MarketMonitor gewählt wurden.

Alle Benutzer in ScienceLounge sind gleichberechtigt. D.h., alle Benutzer können Quellen administrieren oder Dokumente kategorisieren. Daraus ergibt sich, dass alle Benutzer jeweils Zugriff auf alle verschiedenen Arbeitsbereiche bekommen. Das bedeutet, dass eine eindeutige Trennung zwischen Brokern und Konsumenten nicht möglich ist. Vielmehr werden alle Brokeringaufgaben von den Endkonsumenten ausgeführt. Lediglich bei der Personalisierung von Inhalten des IB-Domänenmodells mit der Profilbildungsschnittstelle bekommen die Benutzer von ScienceLounge Unterstützung durch das System. Die Anbieter von Informationen verhalten sich in diesem Modell passiv. Es handelt sich damit nach [SMDP98] um ein Pull-Brokermodell.

Der Kontrollfluss von ScienceLounge entspricht dem Kontrollfluss von Broker's Lounge, der in Kapitel 8.2 bzw. Abbildung 8.1 beschrieben wurde. Da alle Aufgaben von den Konsumenten selbst durchgeführt werden, können alle Schnittstellen als Endbenutzerschnittstellen aufgefasst werden. Eine besondere Bedeutung kommt dabei dem Dokumentenbrowser und der Profilbildungsschnittstelle zu, da über beide Schnittstellen Zugang zu neuen Informationen gegeben werden kann. Über den Dokumentenbrowser kann der Benutzer Zugang zu neuen oder veränderten Dokumenten erlangen, während er über die Profilbildungsschnittstelle Zugang zu neuen oder veränderten Informationseinheiten bekommen kann, die von den Benutzern des Systems angelegt oder verändert wurden.

9.3.1 ScienceLounge-Domänenmodell

Das IB-Domänenmodell von ScienceLounge basiert auf sieben Konzepten:

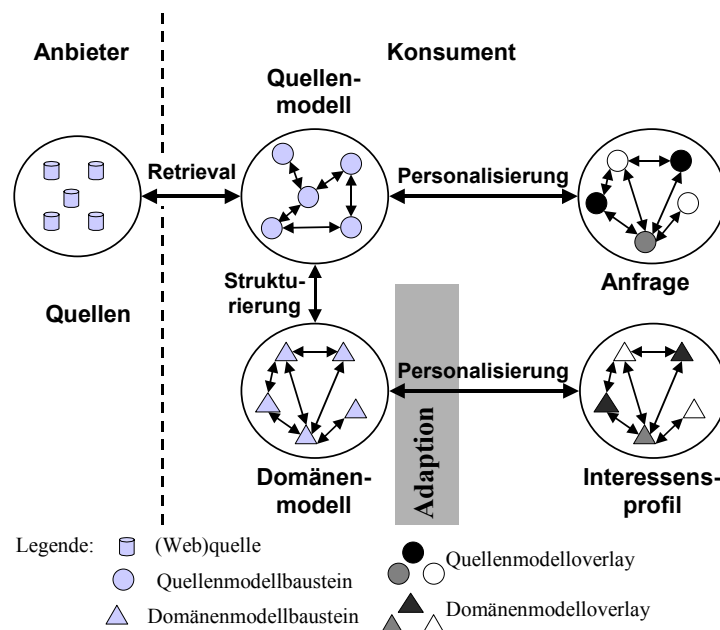


Abbildung 9.7: ScienceLounge Brokeringmodell

1. Konferenzen,
2. Organisationen, die Forschung betreiben,
3. Veröffentlichungen (Paper),
4. Publikationsmöglichkeiten, in denen Forscher ihre Arbeiten veröffentlichen können,
5. Personen, die in der Scientific Community eine Rolle spielen,
6. relevante Projekte und
7. Prototypen.

Relationen zwischen Konzepten

Die zentrale Stellung innerhalb der verschiedenen Konzepte nimmt das Konzept „Person“ ein, da es Querbeziehungen zu fast allen anderen Konzepten aufweist. Personen haben Relationen zu

- Konferenzen, bei denen sie im Programmkomitee vertreten sind,
- Organisationen, bei denen sie beschäftigt sind,

- Veröffentlichungen (Paper), an denen sie mitgewirkt haben und zu
- Projekten und Prototypen, an denen sie beteiligt sind.

Veröffentlichungen weisen eine weitere Relation zu Publikationsmöglichkeiten auf. Diese Relation beschreibt, wo einzelne Papiere veröffentlicht wurden. Ein Beispiel für eine Publikationsmöglichkeit sind die „Communications of the ACM“.

Prototypen haben neben der Relation zu den beteiligten Personen auch eine Relation zu dem Projekt (den Projekten), in dessen Rahmen der Prototyp entstanden ist.

Relationen zwischen Konzepten und Kategorietypen

Zentral für die Scientific Community, die mit ScienceLounge beobachtet werden soll, ist die Klassifikation nach verschiedenen Forschungsgebieten. Die Forschungsgebiete werden genutzt, um Konferenzen, Papiere, Personen und Publikationen zu klassifizieren. Diese starke Bedeutung einer Kategorienhierarchie hebt sich von der Rolle anderer Kategorienhierarchien in den vorgestellten Beispielen ab.

Konferenzen können neben der thematischen Dimension auch nach ihrem Typ unterschieden werden. Dies geschieht durch den Kategorietyt „Konferenztypen“. Beispiele für Konferenztypen sind „nationaler Workshop“ oder „internationale Konferenz“.

Organisationen können entsprechend nach ihrer Organisationsart klassifiziert werden. Typische Organisationsarten sind „Universität“ oder „Forschungseinrichtung“.

Personen können nach ihrer Rolle in der Scientific Community unterschieden werden. Sie können z.B. als „Programm-Chair“ oder „Projektleiter“ eingestuft werden.

Publikationsmöglichkeiten werden ebenfalls nach ihrem Typ klassifiziert. Beispiele für Publikationstypen sind „Journal“ oder „Proceedings“.

Zusammenfassung des IB-Domänenmodells

Die verschiedenen Konzepte und Kategorietypen und ihre Relationen untereinander sind in Tabelle 9.3 zusammengefasst. Die verschiedenen Konzepte weisen eine nur spärliche Attributierung auf. Für Konferenzen wird deren Ort und Jahr angegeben. Für Personen wird ihre jeweilige Emailadresse eingetragen.

Wie schon zuvor beschrieben, nehmen Personen und Forschungsgebiete zentrale Stellungen innerhalb des IB-Domänenmodells ein. Personen spielen vermutlich deswegen eine große Rolle, weil innerhalb der Scientific Community Ideen

	Konferenz	Organisation	Paper	Publikation	Person	Projekt	Prototyp
Konferenz	-	-	-	-	X	-	-
Organisation	-	-	-	-	X	-	-
Paper	-	-	-	X	X	-	-
Publikationsmöglichkeiten	-	-	-	X	-	-	-
Person	X	X	X	-	-	X	X
Projekt	-	-	-	-	X	-	X
Prototyp	-	-	-	-	X	X	-
Forschungsgebiete	X	-	X	X	X	-	-
Organisationsarten	-	X	-	-	-	-	-
Publikationstypen	-	-	-	X	-	-	-
Rollen	-	-	-	-	X	-	-
Ort	X	-	-	-	-	-	-
Jahr	X	-	-	-	-	-	-
Emailadresse	-	-	-	-	X	-	-

Tabelle 9.3: Relationen zwischen Konzepten, Kategorietypen und Attributen in ScienceLounge

oft mit einzelnen Personen oder Gruppen von Personen assoziiert sind. Die Scientific Community definiert sich oft über Themen, an denen geforscht wird. Die verschiedenen Communities grenzen sich über die Themen, die sie bearbeiten, voneinander ab. Dies rechtfertigt die Rolle des Kategorietyps „Forschungsgebiete“

Das ScienceLounge-Domänenmodell kann mit der DBLP-Computer Science Bibliography der Universität Trier⁴ verglichen werden. DBLP bietet eine Übersicht über Veröffentlichungen aus dem Bereich der Informatik. DBLP strukturiert Veröffentlichungen nach Autoren, Journals, Konferenzen, Serien (z.B. LNAI, LNCS) und Büchern. Außerdem werden die Veröffentlichungen nach Themen strukturiert. Es werden jedoch nur wenige Themen zur Strukturierung benutzt. Die Veröffentlichungen in DBLP entsprechen dem Konzept „Paper“ in ScienceLounge. Autoren in DBLP können mit dem Konzept „Person“ in ScienceLounge verglichen werden. Konferenzen in DBLP werden in ScienceLounge durch das Konzept

⁴<http://www.informatik.uni-trier.de/ley/db/index.html>

„Konferenz“ ausgedrückt. Journals, Serien und Bücher können in ScienceLounge als Publikationsmöglichkeiten mit einen jeweils verschiedenen Publikationstyp beschrieben werden. Konferenzen in DBLP können auch als Publikationsmöglichkeit (im ScienceLounge-Domänenmodell) angesehen werden, da für DBLP Veröffentlichungen im Vordergrund der Betrachtung stehen. Somit sind Konferenzen bei DBLP gleichbedeutend mit den Proceedings der Konferenz. „Proceedings“ könnten als weiterer Publikationstyp im ScienceLounge-Domänenmodell modelliert werden. Die Themen von DBLP können mit den „Forschungsgebieten“ in ScienceLounge verglichen werden.

9.3.2 Einsatzerfahrungen

Im täglichen Einsatz hat sich ScienceLounge als sehr wertvolles Hilfsmittel zur Beobachtung der Scientific Community und von relevanten Literaturquellen erwiesen, da Informationen über neue Veröffentlichungen oder neue Prototypen sehr zeitnah entdeckt werden konnten. Als besonders hilfreich hat sich dabei die Differenzsicht im Dokumentenbrowser erwiesen. Die Differenzsicht visualisiert die Veränderungen eines Dokumentes gegenüber vorangegangenen Versionen des Dokumentes, indem die neuen Einträge grün markiert und die gelöschten Wörter in rot markiert werden. Gelöschte Wörter werden zusätzlich noch durchgestrichen markiert.

Abbildung 9.8 zeigt zum Beispiel eine Differenzsicht auf eine Publikationsseite. Durch die Differenzsicht wird deutlich, welche Veröffentlichungen in die Liste aufgenommen, verändert oder gelöscht wurden.

Als problematisch hat sich die paritätische Aufgabenverteilung auf alle Benutzer von ScienceLounge erwiesen, da dadurch kein Anreiz zur Kategorisierung von Informationseinheiten gegeben war. Neue Informationseinheiten wurden nur zögerlich in das System eingebracht. Dies deckt sich mit der Erkenntnis, dass für erfolgreiches Wissensmanagement Anreize für das Teilen von Wissen gegeben werden müssen (vgl. [DP98]).

9.4 Vergleich der Broker's Lounge-Anwendungen

Die drei vorgestellten Anwendungen von Broker's Lounge ELFI, MarketMonitor und ScienceLounge zeigen deutlich die verschiedenen Möglichkeiten von Broker's Lounge.

In ELFI wird mit Informationseinheiten über konkrete Förderprogramme, Förderer und Ansprechpartner gehandelt. ELFI weist eine klare Aufgabenverteilung auf: Die Domänen- und Quellenverwaltungsaufgaben werden zur Zeit von der

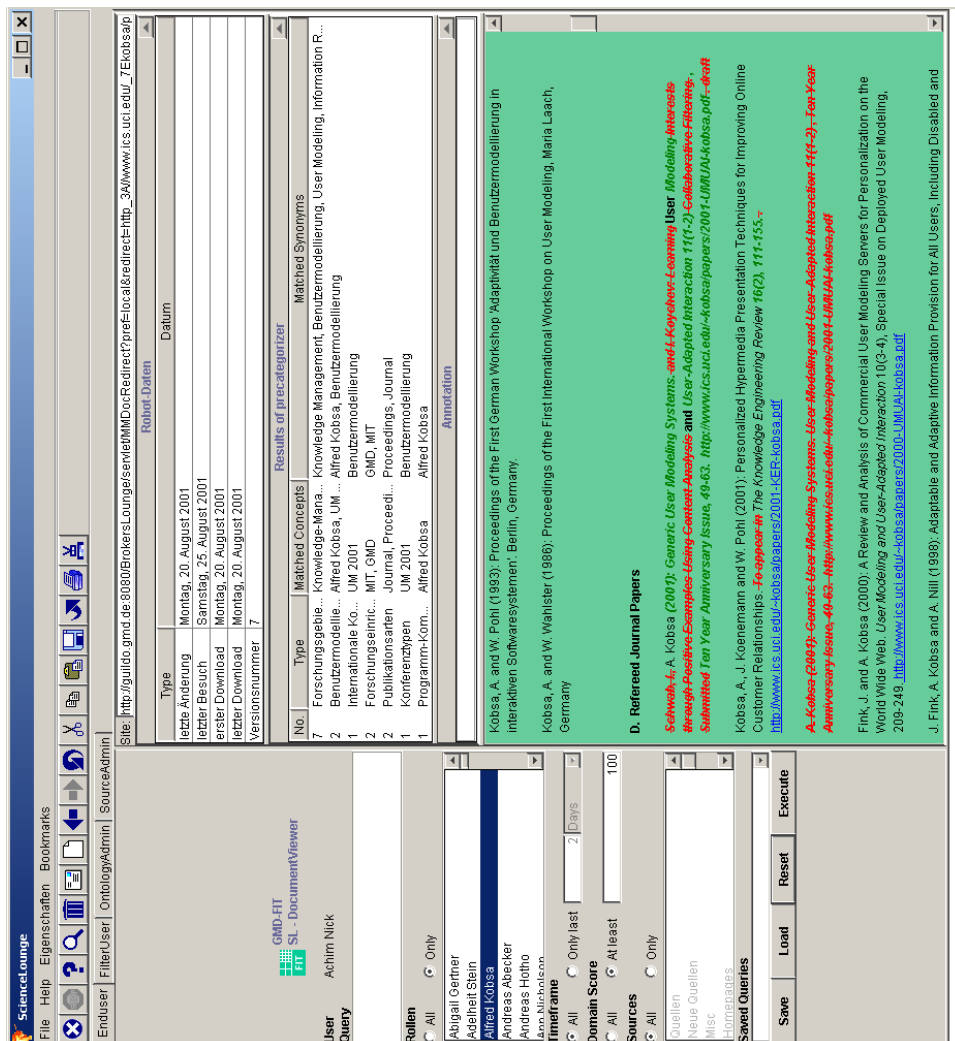


Abbildung 9.8: Differenzsicht im Dokumentenbrowser von ScienceLounge

Servicestelle für Forschungsförderung durchgeführt. Die kundenorientierten Aufgaben werden entweder von Wissenschaftlern, die ein konkretes Informationsbedürfnis haben, selbst durchgeführt oder lokale Forschungsreferenten tun dies im Auftrag und in Kenntnis der Interessen der entsprechenden Wissenschaftler. Sowohl Wissenschaftler als auch Forschungsreferenten bekommen Unterstützung bei der Erstellung und Verwaltung ihrer Informationsbedürfnisse durch das System. Das Brokeringmodell von ELFI kann als neutraler Broker bezeichnet werden, da die Servicestelle zwischen den Förderern, die die Rolle der Informationsanbieter ausfüllen und den Wissenschaftlern und Forschungsreferenten vermittelt. Die Servicestelle ist weder mit den Förderern noch mit den Wissenschaftlern und

Forschungsreferenten direkt assoziiert und kann aus diesem Grund ihre Rolle als neutraler Broker ausfüllen. Im Zentrum des IB-Domänenmodells von ELFI steht das Konzept „Förderprogramm“. Die weiteren Konzepte „Förderer“, „Ansprechpartner“ und „Dokument“ sind mit dem Konzept „Förderprogramm“ assoziiert und beschreiben Informationen im Umfeld von Förderprogrammen. Förderprogramme können mit fünf verschiedenen Kategoriehierarchien klassifiziert werden. „Forschungs-“ und „Querschnittsthemen“ beschreiben thematische Aspekte von Förderprogrammen, während die anderen drei Kategoriehierarchien Eigenschaften und Teilnahmekriterien von Förderprogrammen beschreiben. Die Umfrage zu Beginn des ELFI-Pro-Projektes und die Analyse der Logfiles des ELFI-Systems zeigen, dass von allen fünf Kategoriehierarchien die Forschungsthemenhierarchie mit Abstand am häufigsten benutzt wird.

In MarketMonitor wird im Gegensatz zu ELFI nicht mit Informationseinheiten, sondern mit annotierten, bewerteten Dokumenten gehandelt. Die Annotation und Bewertung wird vom System innerhalb der inhaltlichen Analyse durchgeführt. MarketMonitor wird typischerweise von einem Unternehmen zur Marktbeobachtung eingesetzt. Da die Kooperation der beobachteten Konkurrenten nicht vorausgesetzt werden kann, werden alle Quellen- und Domänenverwaltungsaufgaben vom Unternehmen ausgeführt, das MarketMonitor betreibt. Die kundenorientierten Aufgaben werden von den Informationskonsumenten innerhalb des Unternehmens durchgeführt. Da alle Brokeringaufgaben innerhalb des Unternehmens ausgeführt werden, handelt es sich bei MarketMonitor um ein Pull-Brokermodell. Zentrales Konzept innerhalb des IB-Domänenmodells ist das Konzept „Organisation“, da MarketMonitor der Beobachtung der verschiedenen Marktteilnehmer dient. Die verschiedenen Marktteilnehmer werden im IB-Domänenmodell als Organisationen repräsentiert. Das Konzept „Organisation“ steht mit den Konzepten „Person“ und „Produkt“ in Beziehung. Zentraler Kategorietyt sind „Geschäftsfelder“ zur Klassifikation von Produkten und Organisationen. Diese Hierarchie von Geschäftsfeldern ist eine thematische Beschreibung von Produkten und Organisationen, während weitere Kategoriehierarchien wie Organisationstypen und Rollen den Typ oder die Eigenschaft ihrer klassifizierten Konzepte bzw. Informationseinheiten beschreiben.

In ScienceLounge wird sowohl mit Informationseinheiten als auch mit annotierten, bewerteten Dokumenten gehandelt. Dies ist begründet in der paritätischen Aufgabenverteilung in ScienceLounge. Alle verschiedenen Brokeringaufgaben werden von allen Benutzern des Systems durchgeführt. Alle Benutzer des Systems sind gleichzeitig Konsumenten und Broker. Es handelt sich um ein klassisches Pull-Brokermodell. Zentrales Konzept im IB-Domänenmodell von ScienceLounge ist das Konzept „Person“, das beobachtete Personen repräsentiert. Das Konzept hat Relationen zu fast allen anderen verwendeten Konzepten. Fast alle Konzepte werden durch den Kategorietyt „Forschungsgebiete“ klassifiziert.

	ELFI	MarketMonitor	ScienceLounge
gehandeltes Gut	Informationseinheiten	annotierte Dokumente	Informationseinheiten und annotierte Dokumente
Brokeringmodell	neutraler Broker	Pull-Broker	Pull-Broker
zentrales Konzept	Förderprogramm	Organisation	Person
zentraler Kategorietyt	Forschungsthemen	Geschäftsfelder	Forschungsgebiete
adaptive Unterstützung	ja	nein	ja

Tabelle 9.4: Vergleich der drei Broker's Lounge-Anwendungen

Die Forschungsgebiete-Kategorienhierarchie ist damit die zentrale Hierarchie zur Klassifikation von Informationseinheiten in ScienceLounge. Forschungsgebiete beschreiben eine thematische Dimension ihrer klassifizierten Informationseinheiten, wohingegen die weiteren Kategorietyten den Typ oder die Art ihrer klassifizierten Informationseinheiten beschreiben. Der Kategorietyt „Organisationsarten“ beschreibt z.B. den Typ einer Organisation.

Die drei beschriebenen Anwendungen von Broker's Lounge sind in Tabelle 9.4 zusammengefasst. Die drei Anwendungen bearbeiten verschiedene Domänen. Sie divergieren im Brokeringmodell und im Gut, mit dem sie handeln. Neben dem Handel mit Informationseinheiten (ELFI) kann auch mit den von der inhaltlichen Analyse annotierten und bewerteten Dokumenten (MarketMonitor) oder mit Informationseinheiten und annotierten, bewerteten Dokumenten (ScienceLounge) gehandelt werden. In ELFI und MarketMonitor sind die Brokering-Aufgaben klar auf die beteiligten Personen verteilt. In ScienceLounge gibt es keine Aufteilung der verschiedenen Aufgaben. Stattdessen werden alle Aufgaben von jeder beteiligten Person ausgeführt. Es sind Mischformen zwischen der rigiden Aufgabenverteilung in ELFI und MarketMonitor und der freien Aufgabenverteilung in ScienceLounge möglich. Dadurch können verschiedenste Brokeringmodelle in einer Anwendung unterstützt werden. Das in Kapitel 8 formulierte Ziel, möglichst verschiedene Brokeringmodelle unterstützen zu können, wird dadurch übererfüllt.

Die verschiedenen IB-Domänenmodelle haben jeweils ein zentrales Konzept, um das herum weitere Konzepte gruppiert wurden. In ELFI z.B. werden um das zentrale Konzept „Förderprogramm“ die weiteren Konzepte „Förderer“ und „Ansprechpartner“ gruppiert. Bei der Modellierung der Domäne wurde davon aus-

gegangen, dass die Benutzer sich im Kern für Förderprogramme interessieren, da sie dadurch ihre Forschungsarbeiten finanziell absichern können. Für Förderer und Ansprechpartner von Förderprogrammen interessieren sich die Benutzer, so die Annahme, weil durch die Kenntnis des Umfeldes die Wahrscheinlichkeit einer Förderung ihrer Anträge gesteigert werden kann. In MarketMonitor steht das Konzept „Organisation“ im Mittelpunkt des Interesses, da mit MarketMonitor das Handeln dieser Organisationen auf dem beobachteten Markt beobachtet werden soll. Die weiteren Konzepte „Produkt“ und „Person“ beschreiben das Umfeld von Organisationen. Im IB-Domänenmodell von ScienceLounge nimmt das Konzept „Person“ die zentrale Rolle ein. Damit ist ScienceLounge ein Werkzeug zur Beobachtung von Personen in der Scientific Community. Dies reflektiert, dass Wissenschaft im Wesentlichen von Wissenschaftlern und ihren Ideen vorangebracht wird. Die weiteren Konzepte im IB-Domänenmodell repräsentieren das Wirken von einzelnen Personen, wie die Veröffentlichung von einzelnen Papieren (Konzept „Paper“), die Mitwirkung bei Konferenzen oder Workshops (Konzept „Konferenz“) oder die Erstellung von Prototypen (Konzept „Prototyp“) und Produkten (Konzept „Produkt“). Organisationen beschreiben die Organisation, bei der Wissenschaftler arbeiten.

Den zentralen Konzepten in den beschriebenen IB-Domänenmodellen, die das zentrale Informationsbedürfnis der Konsumenten des Systems beschreiben, steht jeweils ein zentraler Kategorietypp gegenüber. In ScienceLounge ist der Kategorietypp „Forschungsgebiete“ von zentraler Bedeutung, da fast alle Konzepte bzw. deren Informationseinheiten von Forschungsgebieten klassifiziert werden. In ELFI ist der Kategorietypp „Forschungsthemen“ ebenfalls zentral, da er am häufigsten von den Benutzern genutzt wird. Würde man das IB-Domänenmodell um weitere Konzepte wie z.B. „Forschungsprojekt“ oder „Forschungsbericht“ erweitern, so würden diese ebenfalls mit dem Kategorietypp „Forschungsthemen“ klassifiziert werden. In MarketMonitor wird die zentrale Rolle für die Klassifikation vom Kategorietypp „Geschäftsfelder“ eingenommen. Alle zentralen Kategorietyppen in den beschriebenen IB-Domänenmodellen sind thematischer Natur. Die weiteren Kategorietyppen werden meist zur Typisierung von Informationseinheiten benutzt. Im ELFI-Domänenmodell wird z.B. der Kategorietypp „Förderarten“ benutzt, um zu beschreiben, ob es sich bei einem Förderprogramm um eine Projektförderung oder ein Stipendium handelt. In ScienceLounge werden z.B. Konferenzen mit dem Kategorietypp „Konferenztypen“ typisiert.

Die thematischen Kategorietyppen ähneln in ihrer Anlage und Verwendung der Verwendung von Kategoriehierarchien in Subject Gateways (vgl. Kap. 5.2.1) und den in Kapitel 4.5 beschriebenen domänenmodellbasierten, adaptiven Informationssystemen. Die Typisierung von Konzepten durch Kategorietyppen kann mit wenig Aufwand in eine herkömmliche Ontologie überführt werden, wie sie z.B. von Ontobroker (vgl. Kap. 5.1.2) verwendet wird. Statt den Typ eines Konzeptes in einer

getrennten Kategorienhierarchie zu beschreiben, werden dazu die Konzepte selbst zu einer Konzepthierarchie arrangiert.

Die adaptive Unterstützung von Konsumenten ist zur Zeit an die Profilbildungsschnittstelle gekoppelt. D.h., dass nur die Konsumenten vom System eine adaptive Unterstützungsleistung angeboten bekommen, die die Profilbildungsschnittstelle zur Befriedigung ihres Informationsbedürfnisses benutzen. Da in MarketMonitor mit Dokumenten und nicht mit Informationseinheiten gehandelt wird, bekommen die Konsumenten in MarketMonitor keine adaptive Unterstützungsleistung vom System. In den anderen beiden Systemen (ELFI und ScienceLounge) bekommen die Konsumenten hingegen über die Profilbildungsschnittstelle adaptive Unterstützungsleistungen angeboten. Adaptive Unterstützung für die Benutzer des Dokumentenbrowsers ist angedacht und wird im Ausblick (vgl. Kap. 10.1) kurz beschrieben.

Kapitel 10

Zusammenfassung und Ausblick

Information-Brokering wird in dieser Arbeit als der Prozess der Vermittlung zwischen Informationsangeboten und Informationsbedürfnissen verstanden. Es handelt sich um einen Wertschöpfungsprozess, da durch die Verdichtung von Information deren Wert (für den Informationskonsumenten) gesteigert wird (vgl. [Wor98]). Timmers (vgl. [Tim99]) beschreibt Information-Brokering als eines der interessantesten neuen Geschäftsfelder im E-Commerce.

In den Information-Brokeringprozess sind drei verschiedene Rollen involviert. Informationsanbieter bieten Informationen an, Informationskonsumenten haben Informationsbedürfnisse, die der Information-Broker durch Vermittlung zwischen Informationsanbietern und Konsumenten zu befriedigen versucht. Der Information-Brokeringprozess kann in drei Teilprozesse unterteilt werden:

1. Die Informationsquellen müssen erfasst und zugänglich gemacht werden.
2. Die in den Quellen enthaltenen Dokumente werden interpretiert und nach einem IB-Domänenmodell strukturiert, um
3. schließlich angepasst an die Informationsbedürfnisse der Informationskonsumenten diesen zugänglich gemacht zu werden.

In früheren Arbeiten zu Information-Brokering ([Wie94], [JP96], [MKIS98], [FDES98] und für einen Überblick [JP98]) wurde der Schwerpunkt auf die Integration möglichst verschiedener Quellen und deren Interpretation gelegt. Die Benutzer bekommen Zugang zu den integrierten Informationen. Ontologiebasierte Informationssysteme wie Ontobroker (vgl. [FDES98] und Kap. 5.1.2) greifen diesen Ansatz auf und integrieren die Informationen aus den verschiedenen Quellen in einer Domänenontologie. In vielen Fällen werden die Dokumente in den Quellen in einer Ontologiesprache kodiert. Das bedeutet, dass die Strukturierung von den Informationsanbietern selbst durchgeführt wird. Die vorstrukturierten Informationen können ohne weiteren Aufwand in das zentrale IB-Domänenmodell

integriert werden. Dies entspricht dem Brokeringmodell des Semantic-Web-Ansatzes (vgl. [BLHL01] und Kap. 5.1), das die dezentrale Annotation von Dokumenten mit Annotationssprachen wie z.B. RDF vorsieht. Bei diesem Brokeringmodell wird die Kooperation der Informationsanbieter mit einem zentralen Broker vorausgesetzt.

Diese Arbeit hat sich dagegen mit personalisiertem Information-Brokering beschäftigt und hat den Schwerpunkt auf die Beziehung zwischen Broker und Informationskonsumenten gelegt. Damit ein Broker die Informationsbedürfnisse von Informationskonsumenten adäquat befriedigen kann, benötigt er eine genaue Kenntnis seiner Informationsbedürfnisse. Dies limitiert die Zahl der Konsumenten, die ein Broker bedienen kann. Die Abbildung der Konsumenteninteressen kann alternativ von den Konsumenten selbst vorgenommen werden. Da man nicht davon ausgehen kann, dass die Konsumenten ihre Informationsbedürfnisse adäquat auf die Inhalte des zentralen IB-Domänenmodells abbilden können (vgl. [PN99]), bekommen die Konsumenten Unterstützung durch das System. Das System beobachtet den Benutzer und versucht, dadurch Rückschlüsse auf die Interessen des Benutzers zu ziehen. Auf Basis des beobachteten Interessenmodells gibt das System dem Benutzer Tipps zur Änderung seines Anfrageprofils und gewichtet einzelne Einträge im IB-Domänenmodell nach den Interessen des Benutzers. Wichtig ist in diesem Zusammenhang, dass die Benutzer nicht vom System bevormundet werden und die volle Kontrolle über ihr Interessenprofil behalten, da eine unbefugte Änderung des Interessenprofils durch das System zu einem Informationsverlust für den Benutzer führen kann. Die volle Benutzerkontrolle bei gleichzeitiger adaptiver Unterstützungsleistung verwirklicht ein ausgewogenes Verhältnis von Adaptivität und Adaptierbarkeit wie es von [Opp94] und [Fis01] gefordert wurde.

Zentrale Datenstruktur für den vorgestellten Ansatz von Information-Brokering ist ein IB-Domänenmodell, das die verschiedenen von den Informationsanbietern erhältlichen Informationen enthält. Auf dieses IB-Domänenmodell werden die Informationsbedürfnisse und Interessen der Informationskonsumenten abgebildet. Die Struktur des IB-Domänenmodells erlaubt eine effiziente Abbildung von Interessen auf die Struktur des IB-Domänenmodells. Zur Strukturierung des IB-Domänenmodells werden neben Konzepten, die zu einer Konzepthierarchie arrangiert werden können auch mehrere unabhängige Kategoriehierarchien zur Klassifikation einzelner Informationen benutzt. Die Verwendung von Konzepten und Kategorien reflektiert die beiden Strömungen, die in personalisiertem Information-Brokering zusammenfließen:

1. Ontologiebasierte Informationssysteme bilden ein Modell der Domäne in einer Ontologie ab. Diese Ontologien bestehen aus Konzepten, die zu einer Konzepthierarchie organisiert werden können. Zwischen den verschiedenen

Konzepten können Relationen definiert werden.

2. Kategoriehierarchien werden in adaptiven, domänenmodellbasierten Informationssystemen benutzt, um Interessenprofile abzubilden. Im Unterschied zu Konzepten, die zusätzlich zu ihrer hierarchischen Organisation noch Relationen zu anderen Konzepten aufbauen können, können Kategorien nur zu einer Hierarchie organisiert werden.

Im zentralen IB-Domänenmodell werden Konzepte benutzt, um die verschiedenen Informationstypen abzubilden, die von den Informationsanbietern angeboten werden. In der Domäne Forschungsförderung werden z.B. die Konzepte „Förderprogramm“, „Förderer“ und „Ansprechpartner“ benutzt. Kategoriehierarchien werden genutzt, um relevante Eigenschaften abzubilden. Beispiele für Kategoriehierarchien in der Domäne Forschungsförderung sind „Forschungsthemen“ und „Förderarten“. Auf diese Kategoriehierarchien können die Interessen der verschiedenen Benutzer abgebildet werden. Die Kategorien, die die verschiedenen Kategoriehierarchien bilden, werden benutzt, um Informationseinheiten zu klassifizieren. Informationseinheiten sind eindeutig je einem Konzept zugeordnet. Die Relationen der Informationseinheiten sind durch die Relationen des zugeordneten Konzeptes determiniert. Ein Beispiel für eine konkrete Informationseinheit ist das Förderprogramm „Multimediales Buch“.

Auf Basis des IB-Domänenmodells können verschiedene Filter zur Filterung von Informationseinheiten definiert werden. Diese Filter können beliebig kombiniert werden. Informationseinheiten können durch Instanzen-, Konzept- und Kategoriefilter gefiltert werden. Instanzenfilter nutzen Eigenschaften von Informationseinheiten wie z.B. das Datum der letzten Änderung der Informationseinheit, zur Filterung aus. Die Filterwirkung von Konzeptfiltern beruht auf der eindeutigen Zuordnung von Informationseinheiten zu Konzepten. Konzeptfilter können z.B. aus der Menge aller Informationseinheiten die Menge der Informationseinheiten filtern, die dem Konzept „Förderprogramm“ zugeordnet sind. Kategoriefilter basieren auf der Beziehung zwischen Kategorien und Informationseinheiten. Mit einem Kategoriefilter können alle Informationseinheiten gefiltert werden, die einer oder mehreren Kategorien zugeordnet wurden. Ein Kategoriefilter könnte z.B. alle Informationseinheiten filtern, die mit den Kategorien „Mathematik“ und „Informatik“, klassifiziert wurden. Neben allen Informationseinheiten, die mit den beiden Kategorien klassifiziert wurden, passen alle Informationseinheiten durch den Filter, die mit einer der Unterkategorien, z.B. mit „Theoretische Informatik“ oder „Theoretische Mathematik“ klassifiziert wurden.

Durch die Nutzung verschiedenster Kategoriehierarchien zur Beschreibung unabhängiger Aspekte von Konzepten, können sehr feingranulare Filter definiert werden. Das Konzept „Förderprogramm“ wird z.B. durch fünf unabhängige Kate-

goriehierarchien beschrieben. Dadurch wird z.B. ein Filter möglich, der alle Projektförderungen für das Forschungsthema „Informatik“ und das Querschnittsthema „Umwelt“ liefert, an dem Hochschulen aus Nordrhein-Westfalen teilnehmen dürfen und die in den letzten zwei Wochen verändert wurden. Durch die Verwendung beliebig vieler (Kategorie-)Hierarchien unterscheidet sich der hier vorgestellte Ansatz wesentlich von den Ansätzen, die in ontologiebasierten Informationssystemen gewählt wurden, da diese Ansätze nur eine Konzepthierarchie kennen. Eine Abbildung der verschiedenen Kategoriehierarchien in eine Konzepthierarchie ist möglich, hat aber den Nachteil, dass alle Permutationen der verschiedenen unabhängigen Kategoriehierarchien in diese Konzepthierarchie aufgenommen werden müssten. Dadurch steigt die Anzahl der zu beschreibenden Konzepte stark an und die Ausdruckskraft der Konzepthierarchie sinkt, weil mehrere unabhängige Aspekte, die zuvor durch jeweils eine Kategorienhierarchie beschrieben wurden, nun durch eine Hierarchie beschrieben werden. Aus diesem Grund ist nicht klar, welcher Aspekt durch die Konzepthierarchie beschrieben werden soll.

Der gewählte Ansatz hebt sich durch die beliebig vielen Kategoriehierarchien ebenfalls wesentlich von adaptiven, domänenmodellbasierten Informationssystemen ab, da diese Systeme in der Regel nur auf einer Kategorienhierarchie basieren. Diese Kategorienhierarchie beschreibt in der Regel den thematischen Aspekt von Dokumenten.

Die Konzept- und Kategoriehierarchien im IB-Domänenmodell können von einem adaptiven Lernverfahren benutzt werden, um die Interessen von Benutzern zu erlernen. Das Lernverfahren basiert nur auf positiven Beispielen und benötigt kein explizites Feedback durch den Benutzer. Dieser wird dadurch nicht in seiner momentanen Arbeit gestört. Für jede Kategorie (jedes Konzept) kann das Lernverfahren das Interesse des Benutzers an Informationseinheiten, die mit dieser Kategorie klassifiziert (diesem Konzept typisiert) wurde, auf ein Intervall von -1 bis +1 abbilden. Das benutzte Lernverfahren basiert auf dem χ^2 -Test. Werden signifikante Abweichungen in negativer Hinsicht festgestellt, so wird das Desinteresse des Benutzers auf das Intervall von -1 bis 0 abgebildet. Bei einer signifikanten Abweichung nach oben, wird das Interesse des Benutzers auf den positiven Bereich zwischen 0 und 1 abgebildet. Wenn keine signifikanten Abweichungen festgestellt werden konnten, wird diese Kategorie (dieses Konzept) mit dem Interessenwert 0 belegt. Dieses Verfahren betont die Extreme (extremes Interesse bzw. extremes Desinteresse). Da für viele Kategorien und Konzepte keine Quantifizierung von Interessenwerten möglich ist, weil keine signifikanten Abweichungen beobachtet wurden, ist dieses Verfahren für Ansätze nicht geeignet, die zu jeder Zeit eine Quantifizierung von Interesse durch das System benötigen. In WebWatcher z.B. werden auf jeder Seite jeweils drei relevante Links vorgeschlagen. Diese Dienstleistung wäre auf Basis des verwendeten Lernverfahrens nicht möglich. Umgekehrt ist die Treffsicherheit anderer Verfahren deutlich geringer. WebWatcher z.B.

hat eine Treffsicherheit seiner Vorschläge von ca. 48,9% (vgl. [JFM97]). Im vorgestellten Verfahren lässt sich ein Konfidenzintervall von z.B. 95 oder 99% vorgeben. Zusammenfassend lässt sich sagen, dass andere Lernverfahren öfter Tipps generieren, sich beim gewählten Verfahren jedoch die Treffsicherheit der Tipps bestimmen lässt. Das Lernverfahren ist damit insbesondere als Korrektiv für den Benutzer geeignet, um grobe Abweichungen von durchschnittlichen Benutzern mit großer Sicherheit zu erkennen.

Stellt das System Abweichungen zwischen dem eigenen Modell der Interessen des Benutzers und seinen tatsächlich ausgewählten Filtern fest, bekommt der Benutzer einen Tipp, die entsprechende Kategorie (das entsprechende Konzept) aus dem jeweiligen Filter zu löschen oder es hinzuzufügen. Das bedeutet, dass das System in diesem Fall die Rolle eines Korrektivs einnimmt. Wichtig ist, dass der Benutzer die volle Kontrolle über „sein“ Profil behält, da eine unbefugte Änderung des Profils zu einem Informationsverlust für den Benutzer führen kann. Da die Tipps mit einem hohen Konfidenzintervall versehen werden können, werden selten „unsinnige“ Vorschläge vom System generiert, so dass das Vertrauen der Benutzer in die Leistungsfähigkeit des Systems nicht beeinträchtigt wird. Auf Basis der vom System berechneten Relevanzwerte von Kategorien und Konzepten kann die Relevanz von Informationseinheiten errechnet werden. Die Relevanzwerte für die verschiedenen Informationseinheiten können zur Sortierung nach Relevanz und zur Filterung benutzt werden, indem z.B. nur die Informationseinheiten angezeigt werden, die eine bestimmte Relevanzschwelle übersteigen.

Die in dieser Arbeit beschriebenen Ideen wurden in Broker's Lounge implementiert. Broker's Lounge ist ein Werkzeugkasten, mit dem Brokering-Systeme für verschiedenste Domänen mit jeweils unterschiedlichen Brokeringmodellen sehr schnell umgesetzt werden können. Broker's Lounge kann sowohl als ontologiebasiertes als auch als adaptives, domänenmodellbasiertes Informationssystem verstanden werden, da ein Domänenmodell, das IB-Domänenmodell, von zentraler Bedeutung ist und sich das System an den Benutzer anpassen kann. Die Vorteile beider Ansätze ergänzen sich in Broker's Lounge: Durch die Beschreibung der jeweiligen Anwendungsdomäne im IB-Domänenmodell ist Broker's Lounge in vielen Domänen einsetzbar. Die adaptive Komponente in Broker's Lounge ist an die Basisbausteine des IB-Domänenmodells gekoppelt, so dass die Adaptivität in allen Domänen zur Verfügung steht, die durch diese Basisbausteine beschrieben werden können. Das bedeutet, dass Broker's Lounge mit wenig Aufwand auf neue Domänen übertragen werden kann. Die Kosten für die Erstellung eines adaptiven Informationssystems sinken dadurch drastisch, so dass die ungünstige Kosten-Nutzen-Relation von adaptiven Features in Informationssystemen, wie sie von [Fis01] beschrieben wurde, deutlich günstiger gestaltet werden kann. Die Bindung von adaptiven Features an die Basisbausteine des IB-Domänenmodells führt zu einem domänenunabhängigen Informationssystem, wie es von [Bru01] für die

nächsten Jahren erwartet wurde.

Durch die adaptiven Unterstützungsleistungen, die Broker's Lounge anbietet, hebt Broker's Lounge sich wesentlich von anderen ontologie-basierten Informationssystemen wie Ontobroker (vgl. [FDES98] und Kap. 5.1.2) oder Planet-Onto (vgl. [DM99], [MSD00] und Kap. 5.2.4) ab, da diese Systeme sich nicht an den Benutzer anpassen können. Diese Systeme sehen nur die Anpassung des Systems durch den Benutzer vor. Im CoMMA-Projekt (vgl. [GDCG00] und Kap. 5.1.4) wird eine agentenbasierte Unterstützung der Benutzer angestrebt. Broker's Lounge hebt sich von CoMMA durch die Unterstützung beliebiger Brokeringmodelle ab, während CoMMA die drei am Brokeringprozess beteiligten Parteien in jeweils eigenen Agenten abbildet. In Broker's Lounge können verschiedenste Brokeringmodelle abgebildet werden, indem die verschiedenen Brokeringaufgaben individuell an die verschiedenen Benutzer verteilt werden können. Dabei kann die individuelle Aufgabenzuteilung als ein weiterer Aspekt von Personalisierung verstanden werden.

Die Strukturierung des IB-Domänenmodells mit Kategorien und Konzepten hebt sich von den Ansätzen in anderen ontologiebasierten Systemen ab: Im Gegensatz zu diesen Arbeiten geht es im IB-Domänenmodell von Broker's Lounge nicht darum, die Domäne möglichst zweckfrei zu beschreiben, damit sie in vielen Kontexten sinnvoll eingesetzt werden kann, sondern darum, die Domäne so aufzubereiten, dass die Inhalte möglichst einfach zur Abbildung von Konsumenteninteressen benutzt werden können. Das IB-Domänenmodell wird somit für einen bestimmten Zweck aufbereitet. Die Benutzung von Kategorien und Konzepten reflektiert diesen Zweck. Konzepte repräsentieren die Informationstypen, für die Konsumenten sich interessieren. Kategoriehierarchien beschreiben Eigenschaften von Konzepten und werden zur Klassifikation von konkreten Informationseinheiten genutzt, die wiederum eindeutig einem Konzept zugeordnet sind. Die verschiedenen Kategorien, aus denen eine Kategorienhierarchie aufgebaut ist, repräsentieren die verschiedenen Werte für die von der Hierarchie beschriebene Eigenschaft.

10.1 Ausblick

Bestehende Sprachen zur Beschreibung von Ontologien wie z.B. OIL könnten um die Modellierungsmittel Kategorie und Kategorietyt erweitert werden, um die Abbildung von Benutzerinteressen zu ermöglichen und damit einen weiteren Schritt zu einem personalisierten, semantischen Netz zu machen, wie es in [Hor02] angedeutet wird. Bestehende Ontologien könnten personalisiert werden, indem man bestehende Ontologie-Strukturen auf Kategorien und Kategorietypten versucht abzubilden.

Mit dieser Arbeit und der Implementierung der Ideen in Broker's Lounge wurde ein Fundament geschaffen, auf dem viele weitere Brokeringanwendungen implementiert werden können. Drei Anwendungen von Broker's Lounge wurden in dieser Arbeit in Kapitel 9 beschrieben. Ziel ist es, viele weitere Applikationen auf Basis von Broker's Lounge zu erstellen, um mehr über Brokering- und IB-Domänenmodelle zu lernen. Außerdem soll die Funktionalität von Broker's Lounge sukzessive erweitert werden.

Die adaptive Komponente in Broker's Lounge basiert zur Zeit auf einem reinen inhaltsbasierten Ansatz, da inhaltsbasierte Ansätze sich schneller an neue Einträge im IB-Domänenmodell anpassen können (vgl. [ZA01]). Das inhaltsbasierte Lernverfahren soll um kollaborative Elemente erweitert werden, da die Qualität des Lernverfahrens dadurch deutlich gesteigert werden kann (vgl. [HBD00]). Dieses Lernverfahren kann eingesetzt werden, um Benutzer mit ähnlichen Interessen ausfindig zu machen. In der Tradition von Adaptivität und Adaptierbarkeit kann der Einführung des adaptiven, kollaborativen Lernverfahrens die bessere Sichtbarkeit der im Brokeringprozess beteiligten Personen gegenüberstehen, indem zum Beispiel jeweils deutlich gemacht wird, wer die verschiedenen Informationseinheiten und Kategorien im IB-Domänenmodell erzeugt und verändert hat. Ggf. kann auch aufgezeigt werden, wer die verschiedenen Informationseinheiten und Kategorien genutzt hat, sofern dies nicht gegen einschlägige Datenschutzbestimmungen verstößt.

Neben der Erweiterung des angewendeten Lernverfahrens um kollaborative Elemente kann der Einsatz des Lernverfahrens in Broker's Lounge neben der Profilbildungsschnittstelle (vgl. Kap. 8.7) auf den Dokumentenbrowser angewendet werden, indem man die Assoziation von Dokumenten mit Inhalten des IB-Domänenmodells ausnutzt. Neben dem Einsatz des Lernverfahrens zur Personalisierung von Schnittstellen innerhalb von Broker's Lounge, kann der Nucleus aus IB-Domänenmodell und darauf basierenden Filter- und Personalisierungsdienstleistungen gekapselt werden. Dieser Nucleus kann in anderen Systemen, deren Inhalte sich gut auf ein Broker's Lounge-Domänenmodell abbilden lassen, als Personalisierungs-Plugin verwendet werden. Dieser Nucleus könnte z.B. verwendet werden, um in einem bestehenden Portal Personalisierungsdienstleistungen anzubieten.

Ein weiteres interessantes Erweiterungsfeld besteht in der Auswertung der Nutzung von Informationseinheiten durch die verschiedenen Benutzer, da nicht nur die absolute Zahl der Aufrufe einer Informationseinheit erhoben werden kann, sondern diese Aufrufe mit den konkreten Interessenprofilen der Benutzer assoziiert werden können. Dies kann genutzt werden,

- damit Information-Broker eine Gesamtübersicht über die Nutzung ihres Angebotes haben und sie ihr Angebot an die Interessen der Konsumenten an-

passen können,

- damit Informationsanbieter einen Überblick über die Nutzung „ihrer“ Informationseinheiten bekommen und sich gegenüber anderen Informationsanbietern positionieren können und
- damit Konsumenten sich gegenüber anderen Konsumenten positionieren können, indem sie Konsumenten mit ähnlichen Interessen identifizieren können.

In der Domäne Forschungsförderung z.B. könnte der Überblick über die Nutzung des ELFI-Systems (vgl. Kap. 9.1) und der darin beschriebenen Förderprogramme vom Forschungsministerium genutzt werden, um Lücken und Überangebote in der Forschungsförderlandschaft zu identifizieren. Dieser Überblick kann von der Servicestelle für Forschungsförderinformation, die das ELFI-System betreibt, ebenfalls genutzt werden, um das eigene Angebot zu überprüfen.

Die Förderer, die in der Domäne Forschungsförderung die Rolle der Informationsanbieter innehaben, können mit Hilfe der Nutzungsinformationen des ELFI-Systems die Nutzung ihrer eigenen Förderprogramme überprüfen. Außerdem können sie sich gegenüber den anderen Förderern positionieren.

Die Wissenschaftler und Forschungsreferenten, die in der ELFI-Domäne die Rolle der Informationskonsumenten einnehmen, könnten mit Hilfe der Nutzungsstatistik sich selbst in Relation zu den anderen Wissenschaftlern positionieren. Außerdem kann dieses Feature genutzt werden, um Personen mit ähnlichen Interessen zu finden.

Neben der Einführung kollaborativer Elemente in Broker's Lounge steht in Zukunft auch die Einbeziehung weiterer Kontextdimensionen im Mittelpunkt. Dabei wird das Interesse des Benutzers als eine Dimension von Kontext definiert. Weitere Dimensionen von Kontext sind Ort, Zeit und Aufgaben. Die weiteren Dimensionen von Kontext werden in ähnlicher Weise behandelt wie Interessen derzeit in Broker's Lounge. Dies bedeutet, dass die weiteren Kontextdimensionen im IB-Domänenmodell ähnlich wie Interessen mit Konzepten und Kategorien im IB-Domänenmodell modelliert werden. Um die verschiedenen Dimensionen im IB-Domänenmodell abbilden zu können, kommt eine weitere Beschreibungsebene hinzu, die der Beschreibung der jeweiligen Dimension dient. Erste Ansätze dazu finden sich dazu in [Spa02]. Eine sehr ausführliche Beschreibung von Kontext, Kontextdimensionen und Kontextmodellen findet sich in der Dissertation von Roland Klemke (vgl. [Kle02]).

Ein Beispiel für die Anreicherung von Broker's Lounge mit weiteren Kontextdimensionen entsteht zur Zeit im Rahmen des SAiMotion-Projekt¹. Ziel des

¹SAiMotion steht für **S**ituation **A**wareness in **M**otion. Weitere Informationen zum SAiMotion-Projekt sind erhältlich unter <http://www.saimotion.de>

SAiMotion-Projektes ist die Erstellung eines personalisierten Messeführers, der seinen Benutzern angepasst an ihre momentane Situation unterstützt und ihnen zielgerichtet Vorschläge z.B. über relevante Exponate und Aussteller unterbreitet. Die Situation kann als Kontext aufgefasst werden, der mit den oben genannten Dimensionen Zeit, Ort, Aufgabe und Interesse beschrieben werden kann.

Literaturverzeichnis

- [ABH⁺98] ABECKER, ANDREAS, ANSGAR BERNARDI, KNUT HINKELMANN, OTTO KÜHN und MICHAEL SINTEK: *Toward a Technology for Organizational Memories*. IEEE Intelligent Systems & Their Applications, 13(3), Mai/Juni 1998.
- [ACM97] AMBROSINI, LEONARDO, VINCENZO CIRILLO und ALESSANDRO MICARELLI: *A Hybrid Architecture for User-Adapted Information Filtering on the World Wide Web*. In: JAMESON, ANTHONY, CÉCILE PARIS und CARLO TASSO (Herausgeber): *User Modeling: Proceedings of the Sixth International Conference, UM97*, Seiten 59–61. Springer-Verlag, Wien, New York, 1997. Verfügbar unter <http://um.org>.
- [ACT99] ARDISSONO, LILIANA, LUCA CONSOLE und ILARIA TORRE: *Strategies for personalizing the access to news servers*. In: ROGERS, SETH und WAYNE IBA (Herausgeber): *Adaptive User Interfaces – Papers from 1999 AAAI Spring Symposium*, Seiten 7–12. AAAI Press, 1999. ISBN 1-57735-107-X.
- [Aea96] ADAMCZAK, WOLFGANG und ANKE BACKER ET AL.: *Konzept zur koordinierten Nutzung elektronischer Informationsdienste für die Forschungsförderung*. erhältlich unter <http://www.elfi.ruhr-uni-bochum.de/elfi/vorlauf/konzept.html>, 1996.
- [AG00] ARDISSONO, LILIANA und ANNA GOY: *Dynamic Generation of Adaptive Web Catalogs*. In: BRUSILOVSKY, P., O. STOCK und C. STRAPPARAVA (Herausgeber): *Adaptive Hypermedia and Adaptive Web-Based Systems - International Conference, AH 2000*, Seiten 5–16, Trient, Italien, August 2000. Springer-Verlag, Lecture Notes in Computer Science (LNCS 1892). Verfügbar unter <http://ah2000.itc.it/>.

- [AGP⁺01] ARDISSONO, L., A. GOY, G. PETRONE, M. SEGNAU und P. TORASSO: *Tailoring the recommendation of tourist information to heterogeneous user groups*. In: DE BRA, PAUL (Herausgeber): *Workshop "Adaptive Hypertext and Hypermedia" held at Eighth International Conference on User Modeling, UM 2001*, Sonthofen, Juli 2001. Verfügbar unter <http://wwwis.win.tue.nl/ah2001/proceedings.html>.
- [AL99] ALAVI, MARYAM und DOROTHY LEIDNER: *Knowledge Management Systems: Emerging Views and Practices from the field*. In: *32nd International Conference on System Sciences, HICSS-32*, Hawaii, USA, Januar 1999. Verfügbar unter <http://computer.org/proceedings/hicss/0001/0001toc.htm>.
- [BCD01] BUENO, DAVID, RICARDO CONEJO und AMOS A. DAVID: *An Objective Oriented Content Based and Collaborative Recommending System*. In: DE BRA, PAUL (Herausgeber): *Workshop "Adaptive Hypertext and Hypermedia" held at Eighth International Conference on User Modeling, UM 2001*, Sonthofen, Juli 2001. Verfügbar unter <http://wwwis.win.tue.nl/ah2001/proceedings.html>.
- [BCME00] BENTAL, DIANA, ALISON CAWSEY, PATRICK MCANDREW und BRUCE EDDY: *What Does the User Want to Know About Web Resources? A User Model for Metadata*. In: BRUSILOVSKY, P., O. STOCK und C. STRAPPARAVA (Herausgeber): *Adaptive Hypermedia and Adaptive Web-Based Systems - International Conference, AH 2000*, Seiten 268–271, Trient, Italien, August 2000. Springer-Verlag, Lecture Notes in Computer Science (LNCS 1892). Verfügbar unter <http://ah2000.itc.it/>.
- [BD01] BUENO, DAVID und AMOS A. DAVID: *METIORE: A Personalized Information Retrieval System*. In: BAUER, MATTHIAS, PIOTR J. GMYTRASIEWICZ und JULITA VASSILEVA (Herausgeber): *User Modeling 2001 - Eighth International Conference on User Modeling*, Band 2109 der Reihe *Lecture Notes in Artificial Intelligence (LNAI)*, Seiten 168–177, Sonthofen, Juli 2001. Springer-Verlag, Berlin. Verfügbar unter <http://www.um.org/conferences.html>.
- [Bec01] BECKS, ANDREAS: *Visual Knowledge Management with Adaptable Document Maps*. Dissertation, RWTH Aachen, 2001.

- [BFGP98a] BENJAMINS, V. RICHARD, DIETER FENSEL und ASUNCIÓN GÓMEZ PÉREZ: *Knowledge Management through Ontologies*. In: REIMER, ULRICH (Herausgeber): *Second International Conference on Practical Aspects of Knowledge Management (PAKM98)*, Seiten 5.1–5.12. Basel, Schweiz, 1998. Verfügbar unter <http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-13/>.
- [BFGP⁺98b] BENJAMINS, V. RICHARD, DIETER FENSEL, ASUNCION GOMEZ-PEREZ, STEFAN DECKER, MICHAEL ERDMANN, ENRICO MOTTA und MARK MUSEN: *The Knowledge Annotation Initiative of the Knowledge Acquisition Community (KA)²*. In: GAINES, B. und M. MUSEN (Herausgeber): *Proceedings of Eleventh Workshop on Knowledge Acquisition, Modeling and Management (KAW 98)*, Banff, Kanada, April 1998. SRDG Publications. Verfügbar unter <http://ksi.cpsc.ucalgary.ca/KAW/KAW98/KAW98Proc.html>.
- [BH91] BAADER, FRANZ und BERNHARD HOLLUNDER: *A Terminological Knowledge Representation System with Complete Inference Algorithms*. In: *Processing Declarative Knowledge*, Band 567 der Reihe *Lecture Notes in Artificial Intelligence*, Seiten 67–86, Kaiserslautern, Juli 1991. Springer-Verlag.
- [Bil99] BILLSUS, DANIEL: *Improving User Model Aquisition from Labeled Text Documents*. In: BRUSILOVSKY, PETER, PAUL DE BRA und ALFRED KOBSA (Herausgeber): *Workshop "Adaptive Systems and User Modeling on the World Wide Web" held at Seventh International Conference on User Modeling, UM99*, Banff, Kanada, Juni 1999.
- [BLHL01] BERNERS-LEE, TIM, JAMES HENDLER und ORA LASSILA: *The Semantic Web*. Scientific American, (5), Mai 2001. Verfügbar unter <http://www.sciam.com/2001/0501issue/0501berners.lee.html>.
- [BP99] BILLSUS, DANIEL und MICHAEL J. PAZZANI: *A Hybrid User Model for News Story Classification*. In: KAY, JUDY (Herausgeber): *Seventh International Conference on User Modeling, UM99*, Banff, Kanada, Juni 1999. Springer-Verlag, Wien, New York. Verfügbar unter <http://www.um.org/conferences.html>.

- [Bru96] BRUSILOVSKY, PETER: *Methods and Techniques of Adaptive Hypermedia*. User Modeling and User-Adapted Interaction, 6(2-3):87–129, 1996.
- [Bru01] BRUSILOVSKY, PETER: *Adaptive Hypermedia*. User Modeling and User-Adapted Interaction, 11(1-2):87–110, 2001.
- [BSW96] BRUSILOVSKY, PETER, ELMAR SCHWARZ und GERHARD WEBER: *A Tool for Developing Adaptive Electronic Textbooks on WWW*. In: MAURER, HERMANN (Herausgeber): *World Conference on the WWW and Internet WebNet'96*, 1996. Verfügbar unter <http://aace.virginia.edu/aace/conf/webnet/proc96index.html>.
- [CCS01] CHEN, CHIEN CHIN, MENG CHANG CHEN und YEALI SUN: *A Web Document Personalization User Model and System*. In: ABBATTISTA, AYSE GOKER und FABIO, ROSS WILKINSON und GIOVANNI SEMERARO (Herausgeber): *Workshop "Machine Learning, Information Retrieval and User Modeling" held at Eighth International Conference on User Modeling, UM 2001*, Sonthofen, Juli 2001. Verfügbar unter <http://www.scms.rgu.ac.uk/um2001-ws/>.
- [Chi93] CHIN, DAVID N.: *Acquiring user models*. Artificial Intelligence Review, 7:185–197, 1993.
- [dBRLEGN01] BUENAGA RODRÍGUEZ, MANUEL DE, MANUEL J. MAÑA LÓPEZ, ALBERTO DÍAZ ESTEBAN und PABLO GERVÁS GÓMEZ-NAVARRO: *A User Model Based on Content Analysis for the Intelligent Personalization of a News Service*. In: BAUER, MATHIAS, PIOTR J. GMYTRASIEWICZ und JULITA VASSILEVA (Herausgeber): *User Modeling 2001 - Eighth International Conference on User Modeling*, Band 2109 der Reihe *Lecture Notes in Artificial Intelligence (LNAI)*, Seiten 216–218, Sonthofen, Juli 2001. Springer-Verlag, Berlin. Verfügbar unter <http://www.um.org/conferences.html>.
- [DC00] DESIRE-CONSORTIUM: *DESIRE Information Gateways Handbook*, 2000. <http://www.desire.org/handbook/>.
- [DCGR98] DIENG, ROSE, OLIVIER CORBY, ALAIN GIBOIN und MYRIAM RIBIÈRE: *Methods and Tools for Corporate Knowledge Management*. In: GAINES, B. und M. MUSEN (Herausgeber): *Proceedings of Eleventh Workshop on Knowledge Acquisition, Modeling and Management (KAW 98)*, Banff,

- Kanada, April 1998. SRDG Publications. Verfügbar unter <http://ksi.cpsc.ucalgary.ca/KAW/KAW98/KAW98Proc.html>.
- [DEM⁺01] DÖRFLER, A., S. EILERT, A. MENTRUP, M. E. MÜLLER, R. ROLF, C.-R. ROLLINGER, F. SIEVERTSEN und F. TRENKAMP: *Bikini: User Adaptive News Classification in the World Wide Web*. In: SCHÄFER, RALPH, MARTIN E. MÜLLER und S. A. MACSKASSY (Herausgeber): *Workshop "Machine Learning for User Modeling" held at Eighth International Conference on User Modeling, UM 2001*, Sonthofen, Juli 2001.
- [Die00] DIENG, ROSE: *Knowledge Managment and the Internet*. IEEE Intelligent Systems & their applications, Knowledge Management and the Internet, 15(3):14–17, Mai/Juni 2000.
- [DM99] DOMINGUE, JOHN und ENRICO MOTTA: *A Knowledge-Based News Server Supporting Ontology-Driven Story Enrichment and Knowledge Retrieval*. In: FENSEL, DIETER und RUDI STUDER (Herausgeber): *Knowledge Aquisition, Modeling and Management – 11th European Workshop, EKAW'99 (LNAI 1621)*, Seiten 103–120. Springer-Verlag Berlin, Heidelberg, New York, Schloß Dagstuhl, 1999.
- [Dom98] DOMINGUE, JOHN: *Tadzebao and WebOnto: Discussing, Browsing and Editing Ontologies on the Web*. In: GAINES, B. und M. MUSEN (Herausgeber): *Proceedings of Eleventh Workshop on Knowledge Acquisition, Modeling and Management (KAW 98)*, Banff, Kanada, April 1998. SRDG Publications. Verfügbar unter <http://ksi.cpsc.ucalgary.ca/KAW/KAW98/KAW98Proc.html>.
- [DP98] DAVENPORT, THOMAS H. und LAURENCE PRUSACK: *Wenn ihr Unternehmen wüßte, was es alles weiß...: das Praxisbuch zum Wissensmanagement*. mi, Verl. Moderne Industrie, Landsberg/Lech, 2. Auflage, 1998.
- [DSW⁺99] DUINEVELD, A. J., R. STOTER, M. R. WEIDEN, B. KENEPÄ und V. R. BENJAMINS: *WonderTools? A comparative Study of ontological engineering tools*. In: *12th Knowledge Aquisition for Knowledge-based Systems Workshop*, Banff, Kanada, 1999. Verfügbar unter <http://sern.ucalgary.ca/KSI/KAW/KAW99/papers.html>.

- [dvDDO98] DA SILVA, DENISE PILAR, RAFAËL VAN DURM, ERIK DUVALL und HENK OLIVÉ: *Concepts and documents for adaptive educational hypermedia: a model and a prototype*. In: BRUSILOVSKY, PETER und PAUL DE BRA (Herausgeber): *Second Workshop on Adaptive Hypertext and Hypermedia held in conjunction with The Ninth ACM International Hypertext Conference Hypertext'98*, 1998. Verfügbar unter <http://www.wis.win.tue.nl/ah98/Proceedings.html>.
- [EFHM99] ERIKSSON, HENRIK, RAYMOND W. FERGERTSON, YUVAL HAHAR und MARK A. MUSEN: *Automatic Generation of Ontology Editors*. In: *12th Knowledge Acquisition for Knowledge-based Systems Workshop*, Banff, Kanada, 1999. Verfügbar unter <http://sern.ucalgary.ca/KSI/KAW/KAW99/papers.html>.
- [Est01] ESTEBAN, ALBERTO DÍAZ: *Integrating Multilingual Text Classification Tasks and User Modeling in Personalized Newspaper Services*. In: BAUER, MATHIAS, PIOTR J. GMYTRASIEWICZ und JULITA VASSILEVA (Herausgeber): *User Modeling 2001 - Eighth International Conference on User Modeling*, Band 2109 der Reihe *Lecture Notes in Artificial Intelligence (LNAI)*, Seiten 268–270, Sonthofen, Juli 2001. Springer-Verlag, Berlin. Verfügbar unter <http://www.um.org/conferences.html>.
- [FBSV99] FRANCONI, ENRICO, FRANZ BAADER, ULRIKE SATTLER und PANOS VASSILIADIS: *Multidimensional Data Models and Aggregation*. In: JARKE, MATTHIAS, MAURIZIO LENZERINI, YANNIS VASSILIOU und PANOS VASSILIADIS (Herausgeber): *Fundamentals of Data Warehouse*, Kapitel 5, Seiten 87–105. Springer-Verlag, Berlin, 1999. ISBN 3-540-65365-1.
- [FDES98] FENSEL, DIETER, STEFAN DECKER, MICHAEL ERDMANN und RUDI STUDER: *Ontobroker: Or How to Enable Intelligent Access to the WWW*. In: GAINES, B. und M. MUSEN (Herausgeber): *Proceedings of Eleventh Workshop on Knowledge Acquisition, Modeling and Management (KAW 98)*, Banff, Kanada, April 1998. SRDG Publications. Verfügbar unter <http://ksi.cpsc.ucalgary.ca/KAW/KAW98/KAW98Proc.html>.
- [FEFP95] FIKES, RICHARD, ROBERT ENGELMORE, ADAM FARQUHAR und WANDA PRATT: *Network-Based Information Brokers*. the AAAI Spring Workshop on Information Gathering from Distributed, Heterogeneous Environments, 1995.

- [Fen01] FENSEL, DIETER: *Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce*. Springer-Verlag, Heidelberg, 2001. ISBN 3-540-41602-1.
- [FHH⁺00] FENSEL, D., I. HORROCKS, F. VAN HARMELEN, S. DECKER, M. ERDMANN und M. KLEIN: *OIL in a Nutshell*. In: DIENG, ROSE (Herausgeber): *Knowledge Acquisition, Modeling, and Management, Proceedings of the European Knowledge Acquisition Conference (EKAW-2000), Lecture Notes in Artificial Intelligence (LNAI)*. Springer-Verlag, Oktober 2000. Verfügbar unter <http://www.cs.vu.nl/ontoknow/oil/download/oilnutshell.pdf>.
- [Fis01] FISCHER, GERHARD: *User Modelling in Human-Computer Interaction*. User Modeling and User-Adapted Interaction, 11(1-2):65–86, 2001.
- [FKS97] FINK, JOSEF, ALFRED KOBISA und JÖRG SCHRECK: *Personalized Hypermedia Information Provision through Adaptive and Adatable System Features: User Modeling, Privacy and Security Issues*. In: BRUSILOVSKY, PETER, JOSEF FINK und JUDY KAY (Herausgeber): *Workshop "Adaptive Systems and User Modeling on the World Wide Web" held at Sixth International Conference, UM97*, Sardinien, Italien, Juni 1997. Verfügbar unter <http://zeus.gmd.de/fink/um97/proceedings.html>.
- [FM01] FENSEL, DIETER und MARK A. MUSEN: *The Semantic Web: A Brain for Humankind*. IEEE Intelligent Systems Issue on the Semantic Web, 16(2):25–24, März/April 2001. Guest Editor's Introduction.
- [Fow97] FOWLER, MARTIN: *Analysis Patterns: Reusable Object Models*. Addison Wesley, Menlo Park, Kalifornien, USA, 1997.
- [FS00] FOWLER, MARTIN und KENDALL SCOTT: *UML Distilled: a brief guide to the standard object modelling language*. Addison-Wesley, Boston, USA, 2. Auflage, Juli 2000.
- [FvHH⁺01] FENSEL, DIETER, FRANK VAN HARMELEN, IAN HORROCKS, DEBORAH M. MC GUINNESS und PETER F. PATEL-SCHNEIDER: *OIL: An Ontology Infrastructure for the Semantic Web*. IEEE Intelligent Systems Issue on the Semantic Web, 16(2):38–45, März/April 2001.

- [GDCG00] GANDON, FABIEN, ROSE DIENG, OLIVIER CORBY und ALAIN GIBOIN: *A Multi-Agent System to Support Exploiting an XML-based Corporate Memory*. In: REIMER, ULRICH (Herausgeber): *Third International Conference on Practical Aspects of Knowledge Management (PAKM2000)*, Seiten 10.1–10.12. Basel, Schweiz, 2000. Verfügbar unter <http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-34/>.
- [Gen91] GENESERETH, MICHAEL R.: *Knowledge Interchange Format*. In: *Principles of Knowledge Representation and Reasoning- Proceedings of the Second International Conference (KR91)*, Seiten 599–600, Cambridge Massachusetts, April 1991. Morgan Kaufmann Publishers, San Matteo, Kalifornien, USA.
- [GI00] GARLATTI, SERGE und SÉBASTIEN IKSAL: *Concept Filtering and Spatial Filtering in an Adaptive Information System*. In: BRUSILOVSKY, P., O. STOCK und C. STRAPPARAVA (Herausgeber): *Adaptive Hypermedia and Adaptive Web-Based Systems - International Conference, AH 2000*, Trient, Italien, August 2000. Springer-Verlag, Lecture Notes in Computer Science (LNCS 1892). Verfügbar unter <http://ah2000.itc.it/>.
- [GIK99] GARLATTI, S., S. IKSAL und P. KERVELLA: *Adaptive On-Line Information System by means of a Task Model and Spatial Views*. In: BRUSILOVSKY, PETER, PAUL DE BRA und ALFRED KOBSA (Herausgeber): *Workshop "Adaptive Systems and User Modeling on the World Wide Web" held at Seventh International Conference on User Modeling, UM99*, Banff, Kanada, Juni 1999.
- [GK99] GINSBURG, MARK und AJIT KAMBIL: *Annotate: A Web-based Knowledge Management Support System for Document Collections*. In: *32nd International Conference on System Sciences, HICSS-32*, Hawaii, USA, Januar 1999. Verfügbar unter <http://computer.org/proceedings/hicss/0001/0001toc.htm>.
- [GNOT92] GOLDBERG, DAVID, DAVID NICHOLS, BRIAN M. OKI und DOUGLAS TERRY: *Using Collaborative Filtering to Weave an Information Tapestry*. *Communications of the ACM*, 35(12):61–70, Dezember 1992.
- [Gru93] GRUBER, THOMAS R.: *A Translation Approach to Portable Ontology Specifications*. *Knowledge Aquisition*, 5(2), 1993.

- [HBD00] HIRSH, HAYM, CHUMKI BASU und BRIAN D. DAVIDSON: *Learning to Personalize*. Communications of the ACM, 43(8):102–106, August 2000.
- [Hen01a] HENDLER, JAMES: *Agents and the Semantic Web*. IEEE Intelligent Systems Issue on the Semantic Web, 16(2):30–37, März/April 2001.
- [Hen01b] HENZE, NICOLA: *Adaptive Hyperbooks: Adaptation for Project-Based Learning Resources*. Dissertation, Universität Hannover, April 2001.
- [HH01] HEFLIN, JEFF und JAMES HENDLER: *A Portrait of the Semantic Web in Action*. IEEE Intelligent Systems Issue on the Semantic Web, 16(2):54–59, März/April 2001.
- [HN99a] HENZE, NICOLA und WOLFGANG NEJDL: *Adaptivity in the KBS Hyperbook System*. In: BRUSILOVSKY, PETER, PAUL DE BRA und ALFRED KOBSA (Herausgeber): *Workshop "Adaptive Systems and User Modeling on the World Wide Web" held at Seventh International Conference on User Modeling, UM99*, Banff, Kanada, Juni 1999.
- [HN99b] HENZE, NICOLA und WOLFGANG NEJDL: *Bayesian Modeling for Adaptive Hypermedia Systems*. In: 7. GI-Workshop "Adaptivität und Benutzermodellierung in Interaktiven Softwaresystemen", Magdeburg, September 1999. Verfügbar unter <http://www-mmt.inf.tu-dresden.de/joerding/abis99/proceedings.html>.
- [Hor98] HORROCKS, IAN: *The FaCT System*. In: DE SWART, H. (Herausgeber): *Automated Reasoning with Analytic Tableaux and Related Methods: International Conference Tableaux'98*, Nummer 1397 in *Lecture Notes in Artificial Intelligence*, Seiten 307–312, Oisterwijk, Niederlande, Mai 1998. Springer-Verlag, Berlin.
- [Hor01] HORROCKS, IAN: *The Generation of DAML+OIL*. In: GOBLE, CAROLE, DEBORAH L. MCGUINNESS, RALF MÖLLER und PETER F. PATEL-SCHNEIDER (Herausgeber): *Working Notes of the 2001 International Description Logics Workshop (DL-2001)*, Stanford, USA, August 2001. Verfügbar unter <http://SunSITE.Informatik.RWTH-Aachen.DE/Publications/CEUR-WS/Vol-49/>.

- [Hor02] HORROCKS, IAN: *DAML+OIL: A Reason-able Web Ontology Language*. In: JENSEN, CHRISTIAN S., KEITH G. JEFFERY, JAROSLAV POKORNÝ, SIMONAS SALTENIS, ELISA BERTINO, KLEMENS BÖHM und MATTHIAS JARKE (Herausgeber): *Advances in Database Technology - EDBT 2002, 8th International Conference on Extending Database Technology*, LNCS 2287, Seiten 2–13, Prag, Tschechien, März 2002. Springer-Verlag Berlin.
- [HRW97] HÖÖK, KRISTINA, ÅSA RUDSTRÖM und ANNIKA WÆRN: *Edited Adaptive Hypermedia: Combining Human and Machine Intelligence to Achieve Filtered Information*. In: MILOSAVLJEVIC, MARIA, PETER BRUSILOVSKY, JOHANNA D. MOORE, JON OBERLANDER und OLIVIERO STOCK (Herausgeber): *Flexible Hypertext Workshop held in conjunction with The Eight ACM International Hypertext Conference Hypertext'97*, 1997. Verfügbar unter <http://www.mri.mq.edu.au/mariam/flexht/>.
- [HTS98] HORSTMANN, RALF, ULF J. TIMM und PETRA SCHUMANN: *Internet-based Infobroker for Electronic Commerce*. In: *8. European Multimedia, Microprocessor Systems and Electronic Commerce Conference*, Bordeaux, Frankreich, 1998.
- [Jam96] JAMESON, ANTHONY: *Numerical Uncertainty Management in User and Student Modeling: An Overview of Systems and Issues*. *User Modeling and User-Adapted Interaction*, 5(3-4):193–251, 1996.
- [JFM97] JOACHIMS, THORSTEN, DAYNE FREITAG und TOM MITCHELL: *WebWatcher: A Tour Guide for the World Wide Web*. In: *Proceedings of Fifteenth International Joint Conference on Artificial Intelligence IJCAI 97*, Nagoya, Aichi, Japan, August 1997. Morgan Kaufmann Publishers.
- [JGJS95] JARKE, MATTHIAS, RAINER GALLERSDÖRFER, MANFRED A. JEUSFELD und MARTIN STAUDT: *ConceptBase - A Deductive Object Base for Meta Data Management*. *Journal of Intelligent Information Systems*, 4(2):167–192, 1995.
- [JJNS98] JEUSFELD, M.A., M. JARKE, H.W. NISSEN und M. STAUDT: *ConceptBase: Managing conceptual models about information*

- systems. In: P.BERNUS, K.MERTINS und G.SCHMIDT (Herausgeber): *Handbook of Information Systems*, Seiten 265–285. Springer-Verlag, 1998.
- [JJQ⁺99] JARKE, MATTHIAS, MANFRED A. JEUSFELD, CHRSTOPH QUIX, TIMOS SELLIS und PANOS VASSILIADIS: *Metadata Data Warehouse Quality*. In: JARKE, MATTHIAS, MAURIZIO LENZERINI, YANNIS VASSILIOU und PANOS VASSILIADIS (Herausgeber): *Fundamentals of Data Warehouse*, Kapitel 7, Seiten 123–158. Springer-Verlag, Berlin, 1999. ISBN 3-540-65365-1.
- [JKN01] JARKE, M., R. KLEMKE und A. NICK: *Brokers's Lounge - an Environment for Multi-Dimensional User Adaptive Knowledge Management*. In: *34th International Conference on System Sciences, HICSS-34*, Hawaii, USA, Januar 2001.
- [JP96] JEUSFELD, MANFRED A. und MIKE PAPAZOGLOU: *Information Brokering: Design, Search and Transformation*. Technischer Bericht, RWTH Aachen, 1996.
- [JP98] JEUSFELD, MANFRED A. und MICHAEL PAPAZOGLOU: *Information Brokering*. In: KRÄMER, BERND, MICHAEL PAPAZOGLOU und HEINZ-W. SCHMIDT (Herausgeber): *Information Systems Interoperability*, Kapitel 10, Seiten 265–301. Research Studies Press Ltd., Taunton, Somerset, England, 1998. ISBN 0-86380-228-1.
- [KDM⁺01] KALFOGLOU, YANNIS, JOHN DOMINGUE, ENRICO MOTTA, MARIA VARGAS-VERA und SIMON BUCKINGHAM SHUM: *my-Planet: an ontology-driven Web-based personalised news service*. In: *Workshop on "Ontologies and Information Sharing" held at Seventh International Joint Conference on Artificial Intelligence, IJCAI'01*, Seattle, USA, August 2001. Verfügbar unter <http://www.tzi.de/bauer/um-ws/>.
- [KH95] KUOKKA, DANIEL und LARRY HARADA: *Supporting Information Retrieval via Matchmaking*. In: *AAAI Spring Symposium on Information Gathering*, 1995.
- [KHK01] KIM, SANGHEE, WENDY HALL und ANDY KEANE: *Using Document Structures for Personal Ontologies*. In: BAUER,

- MATHIAS, PIOTR J. GMYTRASIEWICZ und JULITA VASSILEVA (Herausgeber): *User Modeling 2001 - Eighth International Conference on User Modeling*, Band 2109 der Reihe *Lecture Notes in Artificial Intelligence (LNAI)*, Seiten 240–242, Sonthofen, Juli 2001. Springer-Verlag, Berlin. Verfügbar unter <http://www.um.org/conferences.html>.
- [KK96] KAY, JUDY und BOB KUMMERFELD: *User model based filtering and customisation of web pages*. In: KAY, JUDY und BOB KUMMERFELD (Herausgeber): *Workshop "User Modelling for Information Filtering on the World Wide Web" held at Fifth International Conference on User Modeling, UM96*, Kailua-Kona, Hawaii, Januar 1996. Verfügbar unter <http://www.cs.su.oz.au/~bob/um96-workshop.html>.
- [KK99] KLEMKE, R. und J. KOENEMANN: *Supporting Information Brokers with an Organisational Memory*. In: *5th German Conf. on Knowledge-Based Systems, Workshop on Knowledge Management, Organizational Memory and Reuse, Internal Report*, Würzburg University, März 1999.
- [Kla00] KLAMMA, RALF: *Vernetztes Verbesserungsmanagement mit einem Unternehmensgedächtnis-Repository*. Dissertation, RWTH Aachen, 2000.
- [Kle02] KLEMKE, ROLAND: *Modelling Context in Information Brokering Processes*. Dissertation, RWTH Aachen, 2002.
- [KMM⁺97] KONSTAN, JOSEPH A., BRADLEY N. MILLER, DAVID MALTZ, JONATHAN L. HERLOCKER, LEE R. GORDON und JOHN RIEDL: *GroupLens: Applying Collaborative Filtering to Usenet News*. *Communications of the ACM*, 40(3):77–87, März 1997.
- [Kob85] KOBZA, ALFRED: *Benutzermodellierung in Dialogsystemen*. Springer-Verlag, Berlin, 1985. ISBN 0-387-16074-4.
- [KQ01] KISS, ADORJAN und JOËL QUINQUETON: *Multiagent Cooperative Learning of User Preferences*. In: STUMME, GERD, ANDREAS HOTH und BETTINA BERENDT (Herausgeber): *Workshop "Semantic Web Mining" held at ECML/PKDD-2001 12th European Conference on Machine Learning (ECML'01)*

- 5th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'01)*, Freiburg, September 2001. Verfügbar unter <http://semwebmine2001.aifb.uni-karlsruhe.de/online.html>.
- [Kru98] KRUMKE, SVEN OLIVER: *Graphentheoretische Konzepte und Algorithmen*. Vorlesungsskript, März 1998. Verfügbar unter <http://www.zib.de/krumke/Postscript/graphen-skript.pdf>.
- [Kuh95] KUHLEN, RAINER: *Informationsmarkt: Chancen und Risiken der Kommerzialisierung von Wissen*. Konstanz : UVK, Universitätsverlag, 1995. ISBN 3-87940-528-X.
- [Las98] LASSILA, ORA: *Web Metada: A Matter of Semantics*. IEEE Internet Computing, 2(4):30–37, Juli/August 1998.
- [LDV99] LIEBERMAN, HENRY, NEIL VAN DYKE und ADRIANA VIVACQUA: *Let's Browse: A Collaborative Web Browsing Agent*. In: *International Conference on Intelligent User Interfaces 1999 IUI'99*, Los Angeles, USA, Januar 1999. Verfügbar unter <http://lieber.www.media.mit.edu/people/lieber/Lieberary/Lets-Browse/Lets-Browse-Intro.html>.
- [LGP⁺90] LENAT, DOUGLAS B., RAMANATHAN V. GUHA, KAREN PITTMAN, DEXTER PRATT und MARY SHEPHERD: *CYC: Toward Programs with Common Sense*. Communications of the ACM, 33(8):30–49, August 1990.
- [Lie95] LIEBERMAN, HENRY: *Letizia: An Agent That Assists Web Browsing*. In: *International Joint Conference on AI (IJCAI-95)*, Montreal, Kanada, August 1995.
- [LM01] LASSILA, ORA und DEBORAH MCGUINNESS: *The Role of Frame-Based Representation on the Semantic Web*. Linköping Electronic Articles in Computer and Information Science, 6(1), 2001. ISSN 1401-9841, Verfügbar unter <http://www.ida.liu.se/ext/etai/received/semaweb/sframe.html>.
- [LR00] LUEG, CHRISTOPHER und REINHARD RIEDL: *How Information Technology Could benefit from Modern Approaches to Knowledge Management*. In: REIMER, ULRICH (Herausgeber): *Third International Conference on Practical Aspects of Knowledge Management (PAKM2000)*, Seiten 15.1–15.7. Basel,

- Schweiz, 2000. Verfügbar unter <http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-34/>.
- [LSR96] LUKE, SEAN, LEE SPECTOR und DAVID RAGER: *Ontology-Based Knowledge Discovery on the World-Wide Web*. In: FRANZ, A. und H. KITANO (Herausgeber): *Working Notes of the Workshop on Internet-Based Information Systems at the 13th National Conference on Artificial Intelligence (AAAI-96)*, Seiten 96–102. AAAI Press, 1996.
- [Mae94] MAES, P.: *Agents that Reduce Work and Information Overload*. Communications of the ACM, 37(7):31–40, Juli 1994.
- [MB98] MAGLIO, PAUL P. und ROB BARRETT: *Adaptive Communities and Web Places*. In: BRUSILOVSKY, PETER und PAUL DE BRA (Herausgeber): *Second Workshop on Adaptive Hypertext and Hypermedia held in conjunction with The Ninth ACM International Hypertext Conference Hypertext'98*, 1998. Verfügbar unter <http://wwwis.win.tue.nl/ah98/Proceedings.html>.
- [MB00] MILLER, LIBBY und DAN BRICKLEY: *ImeshTk: Subject Gateway Review Literature Review*, 2000. <http://www.ilrt.bris.ac.uk/discovery/2000/09/imesh/>.
- [MBJK90] MYLOPOULOS, J., A. BORGIDA, M. JARKE und M. KOUBARAKIS: *Telos: Representing knowledge about information systems*. ACM Transactions on Information Systems, 8(4):325–362, Oktober 1990.
- [McG98] MCGUINNESS, DEBORAH L.: *Ontological Issues for Knowledge-Enhanced Search*. In: GUARINO, NICOLA (Herausgeber): *Proceedings of International Conference - Formal Ontologies in Information Systems, FOIS 98*, Seiten 302–316, Trient, Italien, 1998. IOS Press.
- [Mil90] MILLER, G.: *An On-Line Lexical Database*. International Journal of Lexicography, 13(4):235–312, 1990.
- [MKIS98] MENA, E., V. KASHYAP, A. ILLARRAMENDI und A. SHETH: *Domain Specific Ontologies for Semantic Information Brokering on the Global Information Infrastructure*. In: GUARINO, NICOLA (Herausgeber): *Proceedings of International Conference - Formal Ontologies in Information Systems, FOIS 98*, Seiten 269–283, Trient, Italien, 1998. IOS Press.

- [Mül99] MÜLLER, MARTIN E.: *Inducing Conceptual User Models*. In: 7. GI-Workshop "Adaptivität und Benutzermodellierung in Interaktiven Softwaresystemen", Magdeburg, September 1999. Verfügbar unter <http://www-mmt.inf.tu-dresden.de/joerding/abis99/proceedings.html>.
- [Mül01] MÜLLER, MARTIN E.: *Inducing Content Based User Models with Inductive Logic Programming Techniques*. In: SCHÄFER, RALPH, MARTIN E. MÜLLER und S. A. MACSKASSY (Herausgeber): *Workshop "Machine Learning for User Modeling" held at Eighth International Conference on User Modeling, UM 2001*, Sonthofen, Juli 2001. Verfügbar unter <http://www.dfki.de/rafer/um2001-ml4um-ws/>.
- [MLJ⁺98] MEHLMANN, OLAF, JENS LANDVOGT, ANTHONY JAMESON, THOMAS RIST und RALPH SCHÄFER: *Einsatz Bayes'scher Netze zur Identifikation von Kundenwünschen im Internet*. *Künstliche Intelligenz*, 12(3):43–48, 1998.
- [Mou97] MOUKAS, A. G.: *User Modeling in a MultiAgent Evolving System*. In: BAUER, MATHIAS, WOLFGANG POHL und GEOFF WEBB (Herausgeber): *Workshop "Machine Learning for User Modeling" held at Sixth International Conference on User Modeling, UM97*, Sardinien, Italien, Juni 1997. Verfügbar unter <http://www.dfki.de/bauer/um-ws/>.
- [MS01] MAGNINI, BERNARDO und CARLO STRAPPARAVA: *Improving User Modelling with Content-Based Techniques*. In: BAUER, MATHIAS, PIOTR J. GMYTRASIEWICZ und JULITA VASILEVA (Herausgeber): *User Modeling 2001 - Eighth International Conference on User Modeling*, Band 2109 der Reihe *Lecture Notes in Artificial Intelligence (LNAI)*, Seiten 74–83, Sonthofen, Juli 2001. Springer-Verlag, Berlin. Verfügbar unter <http://www.um.org/conferences.html>.
- [MSD00] MOTTA, ENRICO, SIMON BUCKINGHAM SHUM und JOHN DOMINGUE: *Ontology-Driven Document Enrichment: Principles, Tools and Applications*. *International Journal of Human-Computer Studies*, 52(5):1071–1109, 2000.
- [MSPK00] MACH, MARIAN, TOMAS SABOL, JAN PARALIC und ROBERT KENDE: *Knowledge Modelling in Support of Knowledge Management*. In: *Knowledge at work - research information for socie-*

- ty – 5th Conference on Current Research Information Systems in Europe (CRIS'2000). Helsinki, Finnland, 2000. Verfügbar unter <http://www.cordis.lu/cris2000/home.html>.
- [Nan01] NANAS, NIKOLAOS: *Research Proposal: An Adaptive, Evolutionary User profile for Knowledge Management*. Technical Report KMI-TR-114, Open University, Knowledge Media Institute, Juni 2001. Verfügbar unter <http://kmi.open.ac.uk/publications/techreports-text.cfm>.
- [NKS98] NICK, A., J. KOENEMANN und E. SCHALÜCK: *ELFI: Information brokering for the domain of research funding*. Computer Networks and ISDN Systems, 30:1491–1500, 1998.
- [NSD⁺01] NOY, NATALYA F., MICHAEL SINTEK, STEFAN DECKER, RAY W. FERGERSON MONICA CRUBÉZY und MARK A. MUSEN: *Creating Semantic Web Contents with Protégé-2000*. IEEE Intelligent Systems Issue on the Semantic Web, 16(2):60–71, März/April 2001.
- [NT97] NONAKA, IKUJIRO und HIROTAKA TAKEUCHI: *Die Organisation des Wissens: Wie japanische Unternehmen eine brachliegende Resource nutzbar machen*. Campus-Verlag, Frankfurt am Main, 1997. Aus dem Engl. von Friedrich Mader.
- [Opp94] OPPERMAN, R.: *Adaptively supported Adaptability*. International Journal of Human-Computer Studies, 40(3):455–472, 1994.
- [Pea88] PEARL, JUDEA: *Probabilistic Reasoning in Intelligent systems: Networks of plausible Inference*. Morgan Kaufmann Publishers, San Matteo, Kalifornien, USA, 1988.
- [Ped00] PEDNAULT, EDWIN P. D.: *Representation is Everything*. Communications of the ACM, 43(8):80–83, August 2000.
- [PKD⁺] PEREZ, PHILLIPE, HERVÉ KARP, ROSE DIENG, OLIVIER CORBY, ALAIN GIBOIN, FABIEN GANDON, JOEL QUINQUETON, AGOSTINO POGGI, GIOVANNI RIMASSA, CLAUDIO FIETTA, JÜRGEN MÜLLER und JOACHIM HACKSTEIN: *Corporate Memory Management through Agents*. <http://citeseer.nj.nec.com/381229.html>.

- [PMB96] PAZZANI, MICHAEL, JACK MURAMATSU und DANIEL BILL-SUS: *Syskill & Webert: Identifying interesting web sites*. In: *Proceedings of the National Conference on Artificial Intelligence*, Portland, USA, 1996.
- [PN99] POHL, WOLFGANG und ACHIM NICK: *Machine Learning and Knowledge Representation in the LaboUr Approach to User Modeling*. In: KAY, JUDY (Herausgeber): *Seventh International Conference on User Modeling, UM99*, Banff, Kanada, Juni 1999. Springer-Verlag, Wien, New York. Verfügbar unter <http://www.um.org/conferences.html>.
- [Pri99] PRISS, UTA: *Description Logic and Faceted Knowledge Representation*. In: *International Description Logic Workshop*, Linköping, Sweden, August 1999. Verfügbar unter <http://SunSITE.Informatik.RWTH-Aachen.DE/Publications/CEUR-WS/Vol-22/>.
- [Pri00] PRISS, UTA: *Facetted Information Representation*. In: STUMME, GERD (Herausgeber): *Working with Conceptual Structures. Proceedings of the 8th International Conference on Conceptual Structures*, Seiten 84–94, Aachen, 2000. Shaker Verlag.
- [Ran62] RANGANATHAN, S. R.: *Elements of library classification*. Asia Publishing House, Bombay, Indien, 1962.
- [Ric79] RICH, ELAINE: *User Modeling via Stereotypes*. *Cognitive Science*, 3:329–354, 1979.
- [Ros78] ROSCH, ELEANOR: *Principles of Categorization*. In: ROSCH, ELEANOR und BARBARA B. LLOYD (Herausgeber): *Cognition and Categorization*, Kapitel 2, Seiten 27–48. Lawrence Erlbaum Associates, Hillsdale, New Jersey, USA, 1978. ISBN 0-470-26377-6.
- [RV97] RESNICK, PAUL und HAL R. VARIAN: *Recommender Systems*. *Communications of the ACM*, 40(3):56–58, März 1997.
- [RWH97] RUDSTRÖM, A., A. WAERN und K. HÖÖK: *Interactive adaptation of Intranet newsletters*. In: BRUSILOVSKY, PETER, JOSEF FINK und JUDY KAY (Herausgeber): *Workshop "Adaptive Systems and User Modeling on*

- the Worl Wide Web” held at Sixth International Conference, UM97, Sardinien, Italien, Juni 1997. Verfügbar unter <http://zeus.gmd.de/fink/um97/proceedings.html>.*
- [SBF98] STUDER, RUDI, V. RICHARD BENJAMINS und DIETER FENSEL: *Knowledge Engineering: Principles and Methods*. Data and Knowledge Engineering, 25(1-2), März 1998.
- [Sch01] SCHÄFER, RALPH: *Rules for Using Multi-Attribute Utility Theory for Estimating a User’s Interests*. In: HENZE, NICOLA (Herausgeber): *Proceedings des 9. GI-Workshops: ABIS-Adaptivität und Benutzermodellierung in interaktiven Softwaresystemen*, Dortmund, Oktober 2001. Verfügbar unter http://www.kbs.uni-hannover.de/henze/ABIS_Workshop2001/ABIS_2001.html.
- [SF99] SÜSS, CHRISTIAN und BURKHARD FREITAG: *Meta-Daten als Grundlage für Adaption in Lernsystemen*. In: 7. GI-Workshop ”Adaptivität und Benutzermodellierung in Interaktiven Softwaresystemen”, Magdeburg, September 1999. Verfügbar unter <http://www-mmt.inf.tu-dresden.de/joerding/abis99/proceedings.html>.
- [She94] SHEA, VIRGINIA: *The core rules of netiquette*. Albion Books, San Francisco, USA, 1994. verfügbar unter <http://www.albion.com/netiquette/corerules.html>.
- [Sie85] SIEGEL, SIDNEY: *Nichtparametrische statistische Methoden*. Fachbuchhandlung für Psychologie, Eschborn, 2. Auflage, 1985.
- [Sie86] SIEBER, H.: *Mathematische Formeln mit Formelsammlung E*. Ernst Klett Verlag, Stuttgart, 1986.
- [SJ00] STAUDT, MARTIN und MATTHIAS JARKE: *View Management Support in Advanced Knowledge Base Servers*. Journal of Intelligent Information systems, 15:253–285, 2000.
- [SKR01] SCHAFER, J.B., J. KONSTAN und J. RIEDL: *Electronic Commerce Recommender Applications*. Journal of Data Mining and Knowledge Discovery, 5(1/2):115–152, 2001. Verfügbar unter <http://www.cs.umn.edu/Research/GroupLens/research.html>.
- [SM81] SMITH, EDWARD E. und DOUGLAS L. MEDIN: *Categories and Concepts*, Band 4 der Reihe *Cognitive Science Series*. Harvard

- University Press, Cambridge, Massachusetts, 1981. ISBN 0-674-10275-4.
- [SM83] SALTON, GERARD und MICHAEL J. MACGILL: *Information Retrieval - Grundlegendes für Informationswissenschaftler*. MacGraw-Hill, 1983. ISBN 3-89028-051-X.
- [SM95] SHARDANAND, UPENDRA und PATTIE MAES: *Social Information Filtering: Algorithms for Automating "Word of Mouth"*. In: ROBERTS, TERRY und SCOTT ROBERTSON (Herausgeber): *1995 Conference on Human Factors in Computing Systems CHI'95*, Denver, USA, Mai 1995.
- [SMDP98] STRENS, R. M., M. MARTIN, J. E. DOBSON und S. PLAGEMANN: *Business and Market Models of Brokerage in Network-Based Commerce*. In: *Proceedings of 5th International Conference on Intelligence in Services and Networks (IS&N'98)*, LNCS, Antwerpen, Belgien, Mai 1998. Springer-Verlag.
- [SP99a] SCHWAB, INGO und WOLFGANG POHL: *Learning Information Interest from Positive Examples*. In: BAUER, MATTHIAS, PIOTR GMYTRASIEWICZ und WOLFGANG POHL (Herausgeber): *Workshop "Machine Learning for User Modeling" held at Seventh International Conference on User Modeling, UM99*, Banff, Kanada, Juni 1999. Verfügbar unter <http://www.dfki.de/bauer/um99-ws/>.
- [SP99b] SCHWAB, INGO und WOLFGANG POHL: *Learning User Profiles from Positive Examples*. In: PAPATHEODOROU, DR. CHRISTOS (Herausgeber): *Workshop "Machine Learning in User Modeling" held at ECCAI Advanced Course on Artificial Intelligence for 1999 (ACAI-99)*, Chania, Griechenland, Juli 1999. Verfügbar unter <http://www.iit.demokritos.gr/skel/eetn/acai99/>.
- [Spa02] SPANIOL, MARC: *Situationsgesteuerte Inhaltspräsentation unter Verwendung mobiler Assistenten: Modellierung und Bewertung von Situationsmustern*. Diplomarbeit, RWTH Aachen, 2002.
- [Spe98] SPECHT, MARCUS: *Adaptive Methoden in computerbasierten Lehr/Lernsystemen*. GMD Research Series. GMD Forschungszentrum Informationstechnik, Sankt Augustin, 1998. Dissertation Universität Trier.

- [SPK00] SCHWAB, INGO, WOLFGANG POHL und IVAN KOYCHEV: *Learning to Recommend from Positive Evidence*. In: LIEBERMAN, HENRY (Herausgeber): *ACM Intelligent User Interfaces Conference (IUI-2000)*, New Orleans, USA, Januar 2000. Verfügbar unter <http://lieber.www.media.mit.edu/people/lieber/IUI/>.
- [SQ01] SCHOOP, MAREIKE und CHRISTOPH QUIX: *DOC.COM: a framework for effective negotiation support in electronic marketplaces*. *Computer Networks*, 37(2):153–170, 2001.
- [SS98] STEFANI, ANNA und CARLO STRAPPARAVA: *Personalizing Access to Web Sites: The SiteIF Project*. In: BRUSILOVSKY, PETER und PAUL DE BRA (Herausgeber): *Second Workshop on Adaptive Hypertext and Hypermedia held in conjunction with The Ninth ACM International Hypertext Conference Hypertext'98*, 1998. Verfügbar unter <http://wwwis.win.tue.nl/ah98/Proceedings.html>.
- [SWdH⁺95] SCHREIBER, A. TH., B. J. WIELINGA, R. DE HOOG, H. AKKERMANS und W. VAN DE VELDE: *CommonKADS: A Comprehensive Methodology for KBS Development*. *IEEE Expert*, Seiten 28–37, Dezember 1995.
- [Tim99] TIMMERS, PAUL: *Electronic Commerce: Strategies and Models for Business-to-Business Trading*. Wiley Series in Information Systems. John Wiley & Sons, New York, 1999.
- [TR98] TIMM, ULF J. und MARTIN ROSEWITZ: *Benutzermodellierung in der Elektronischen Produktberatung - Konzept und prototypische Realisierung in einer Online-Umgebung*. In: 6. GI-Workshop "Adaptivität und Benutzermodellierung in Interaktiven Softwaresystemen", Erlangen, Oktober 1998. Verfügbar unter <http://www.forwiss.uni-erlangen.de/timm/abis-98/abis-98.htm>.
- [von99] VON BUOL, BETTINA: *Qualitätsgestützte, kooperative Terminologiarbeit*. Dissertation, RWTH Aachen, September 1999.
- [vWE86] WINTERFELD, D. VON und W. EDWARDS: *Decision Analysis and Behavioral Research*. Cambridge University Press, Cambridge, England, 1986.
- [WATR99] WÆRN, ANNIKA, CHARLOTTE AVERMAN, MARK TIERNEY und ÅSA RUDSTRÖM: *Information Services Based on User Pro-*

- file Communication*. In: KAY, JUDY (Herausgeber): *Seventh International Conference on User Modeling, UM99*, Banff, Kanada, Juni 1999. Springer-Verlag, Wien, New York. Verfügbar unter <http://www.um.org/conferences.html>.
- [Wie94] WIEDERHOLD, GIO: *Interoperation, Mediation, and Ontologies*. In: *Proceedings International Symposium on Fifth Generation Computer Systems (FGCS94), Workshop on Heterogeneous Knowledge Bases*, Band W3, Seiten 33–48, Tokyo, Dezember 1994.
- [Wor98] WORSFOLD, EMMA: *Subject Gateways: Fulfilling the DESIRE for knowledge*. *Computer Networks and ISDN Systems*, 30:1479–1489, 1998.
- [WPB01] WEBB, GEOFFREY I., MICHAEL PAZZANI und DANIEL BILL-SUS: *Machine Learning for User Modeling*. *User Modeling and User-Adapted Interaction*, 11(1-2):19–29, 2001.
- [WW89] WAND, YAIR und RON WEBER: *A Model Of Systems Decomposition*. In: DEGROSS, JANICE I., JOHN C. HENDERSON und BENN R. KONSYNSKI (Herausgeber): *Proceedings of the Tenth International Conference on Information Systems*, Boston, USA, Dezember 1989.
- [ZA01] ZUCKERMAN, INGRID und DAVID W. ALBRECHT: *Predictive Statistical Models for User Modeling*. *User Modeling and User-Adapted Interaction*, 11(1-2):5–18, 2001.

Anhang A

Lebenslauf

Persönliche Daten

Achim Nick, geboren am 24. Juli 1969 in Boppard am Rhein.

Schulausbildung

1976-1980	Sankt Nikolaus Grundschule Spay am Rhein
1980-1989	Bischöfliches Cusanus-Gymnasium Koblenz
1989	Abschluss mit Abitur

Zivildienst

1989-1990	Förderverein für Individuelle Schwerstbehinderten-Betreuung und Mobilen Sozialen Hilfsdienst in Koblenz und Umgebung e.V.
-----------	--

Universitäre Ausbildung

1990-1997	Universität Koblenz-Landau, Abt. Koblenz
1997	Abschluss als Diplom-Informatiker

Beruflicher Werdegang

1994-1995	Praktikant und Werkstudent am Fraunhofer-Institut FIT ¹
1995-1997	Studentischer Mitarbeiter bei Fraunhofer-FIT
1997-2002	Wissenschaftlicher Mitarbeiter bei Fraunhofer-FIT
seit 2002	Entwicklungsleiter der bureau42 GmbH

¹Das Fraunhofer-Institut FIT ist durch die Fusion der GMD mit der Fraunhofer-Gesellschaft aus dem gleichnamigen GMD-Institut hervorgegangen