

Konzeption und Implementierung eines Abonnementdienstes für die SWING-Suchmaschine

Studienarbeit

Universität Rostock, Fachbereich Informatik



vorgelegt von: Beate Porst

Betreuer: Prof. Dr. Andreas Heuer
Dr. Ing. Holger Meyer

Inhaltsverzeichnis

1	Einführung	4
1.1	Ziel der Arbeit	5
1.2	Aufbau der Arbeit	5
2	Die SWING-Suchmaschine	6
2.1	Besondere Merkmale von SWING	6
2.2	Ausnutzung der Suchmaschine durch den Abonnementdienst . . .	7
3	Grundlagen	8
3.1	Datenbanken und WWW	8
3.1.1	HyperText Transfer Protocol (HTTP)	8
3.1.2	Uniform Ressource Identifier (URI)	8
3.1.3	Common Gateway Interface (CGI)	9
3.1.4	HyperText Markup Language (HTML)	9
3.1.5	Datenbankanbindung an das World Wide Web	10
3.2	Agenten im Internet	10
3.2.1	Definition	11
3.2.2	Charakteristische Eigenschaften	11
3.2.3	Klassifikation von Agenten	13
3.3	Abonnementdienste	15
4	Architekturmodell eines Abonnementdienstes	20
4.1	Probleme der Abonnementdienste	20
4.2	Anforderungen an Abonnementdienste	23
4.3	Das Architekturmodell	24

4.3.1	Der Personal Digital Assistant	26
4.3.2	Der Informationsagent	27
4.3.3	Der Informationsanbieter	29
4.3.4	Klassifikation des Abonnementdienstes	30
4.3.5	Zusammenfassung	30
5	Evaluierung ausgewählter Abonnementdienste	32
5.1	Kriterien für Abonnementdienste	32
5.2	Ariadne	34
5.3	The Informant	36
5.4	Pointcast Business Network 2.0	36
5.5	Elsevier Science - ContentsDirect	38
5.6	Übersicht und Zusammenfassung	38
6	Der SWING-Abonnementdienst	41
6.1	Konzeption	41
6.1.1	Zielstellung	41
6.1.2	Analyse	42
6.1.3	Funktionalität	45
6.1.3.1	Anfragebearbeitung	46
6.1.3.2	URL-Reminder	47
6.1.3.3	Newschannel	48
6.1.3.4	Themenkatalog	49
6.1.4	Datenbankentwurf des SWING-Abonnementdienstes	49
6.1.5	Architektur	52
6.1.6	Ausblick auf zusätzliche Funktionalität	53
6.2	Implementierung	53
6.2.1	Umsetzung des ER-Schemas in das relationale Schema	54
6.2.2	Konfliktauflösung	54
6.2.3	Beschreibung der Implementierung	55
7	Zusammenfassung und Ausblick	58
7.1	Zusammenfassung	58

7.2	Bewertung der Ergebnisse	59
7.3	Ausblick	59
	Literaturverzeichnis	61
	Abbildungsverzeichnis	63
A	Das SWING-Datenbankschema	64
B	Konfliktauflösung im SWING-Schema	68

Kapitel 1

Einführung

Der Begriff Informationsgesellschaft als Gesellschaftsform des neuen Jahrhunderts nimmt einen immer breiter werdenden Raum ein. Die Ursache ist eine dramatische Entwicklung der letzten Jahren, die dazu führte, daß dem stetig steigenden Informationsbedarf des Menschen eine unüberschaubare Menge an Informationsangeboten gegenübergestellt wurde.

Ein Grund dafür war die Einführung des World Wide Web (WWW) als ein riesiger, digital verteilter Informationsraum. Als Austauschort von Forschungsdokumenten auf dem Gebiet der Nuklearphysik von CERN¹ zu Beginn der neunziger Jahre entwickelt, hat es nunmehr seinen Siegeszug durch alle Bereiche angetreten. Seit dieser Zeit ist ein enormer Anstieg des Informationsangebotes zu verzeichnen. Reichte es anfänglich vielleicht noch aus, sich von Seite zu Seite "hangelnd" die gewünschten Informationen zusammen zu suchen, so mußte diese Form der Informationsrecherche schon bald von leistungsfähigen Suchmaschinen übernommen werden, um ein akzeptables Anfrageergebnis zu erzielen. Inzwischen hat das WWW eine so große Akzeptanz erlangt, daß täglich hunderte neuer Nutzer und damit potentielle Informationsquellen hinzukommen. Laut einer Analyse von Netcraft² wurden im Juli 1998 2.6 Millionen Web-Server registriert. Geht man davon aus, daß pro Web-Server wiederum einige hundert Internetseiten etabliert sind, kann man von mehreren hundert Millionen zur Verfügung stehenden Internetseiten ausgehen. Für eine befriedigende Informationsgewinnung durch das WWW kommt erschwerend hinzu, daß jeder Anbieter seine Dokumente beliebig verschieben oder gar löschen kann, unabhängig von bestehenden Verweisen auf diese Informationsquellen. Eine referentielle Integritätssicherung, wie man sie aus dem Datenbankbereich kennt, existiert nicht. Diese Faktoren lassen eine professionell betriebene Informationssuche, selbst unter Verwendung leistungsfähiger Suchmaschinen, zur zeit- und nervenaufreibenden Arbeit werden. Abhilfe könn-

¹CERN - European Laboratory for Particle Physics

²<http://www.netcraft.com/survey/> - Web-Server-Statistik

ten hier Profil- bzw. Abonnementdienste schaffen, die mit Hilfe von Agententechnologien gezielt nach Informationen suchen, diese entsprechend den Wünschen des Nutzers aufbereiten und in periodischer Form zur Verfügung stellen. Der Nutzer würde dadurch zum einen von der aufwendigen Informationssuche befreit, zum anderen auch mit einem qualitativ höheren Informationsgehalt versorgt werden.

Für die SWING-Suchmaschine soll im Rahmen dieser Studienarbeit ein solcher Abonnementdienst konzipiert und implementiert werden.

1.1 Ziel der Arbeit

Das Ziel der Studienarbeit ist die Konzeption und Implementierung eines Abonnementdienstes für die SWING Suchmaschine. Der Schwerpunkt der Arbeit wird dabei nicht auf der vollständigen Implementierung aller Komponenten des Abonnementdienstes liegen, sondern vielmehr in der konzeptionellen Ausarbeitung und datenbanktechnischen Vorbereitung. Desweiteren soll anhand eines Architekturmodells gezeigt werden, welche Leistungskraft unter idealen Voraussetzungen in einem Abonnement-Agentensystem stecken kann.

1.2 Aufbau der Arbeit

Im zweiten Kapitel werden zunächst die besonderen Merkmale der SWING-Suchmaschine herausgestellt. Die notwendigen Grundbegriffe, die für das Verständnis der Arbeit erforderlich sind, liefert das dritte Kapitel. Einen Großteil dieses Kapitels wird dabei die Erläuterung des Begriffes *Agent* beziehungsweise *Agentensystem* einnehmen. Im darauffolgenden Kapitel wird ein zukünftiges Architekturmodell eines Informationsabonnementdienstes aufgestellt. Dabei werden die Funktionsweise als auch die notwendigen Voraussetzungen zur Umsetzung näher beschrieben. In Kapitel fünf werden verschiedene existierende Profil- und Abonnementdienste auf ihre Funktionalität hin untersucht und gegenübergestellt. Das Kapitel sechs bildet neben dem Kapitel vier den zweiten Schwerpunkt dieser Arbeit. Hier werden die aus der Untersuchung in Kapitel fünf gewonnenen Erkenntnisse in ein Konzept für einen Abonnementdienst umgesetzt. Desweiteren erfolgt eine kurze Beschreibung des bereits erfolgten Implementierungsteils. Den Abschluß der Arbeit bildet eine Bewertung der Ergebnisse der Studienarbeit sowie ein Ausblick auf mögliche Erweiterungen beziehungsweise Veränderungen.

Kapitel 2

Die SWING-Suchmaschine

Das Ziel des Projektes SWING¹ (Suchdienst für WWW-basierte Informationssysteme der **n**ächsten **G**eneration) ist die Entwicklung einer Suchmaschine, die speziell an das Regionale Informationssystem MV-Info angepaßt ist. Dabei sollen von der Suchmaschine hauptsächlich Informationen gefunden werden, die einen Bezug zu Mecklenburg Vorpommern haben. Diese Informationen können auf verschiedensten WWW-Servern weltweit gespeichert sein. Ausschlaggebend für das Wiederauffinden eines Dokumentes ist nur, daß dem System die Adresse bekannt ist. Die Funktionalität von SWING umfaßt neben den herkömmlichen Suchfunktionen (Suche in HTML, Text, PostScript-Dateien) auch die Möglichkeit der Integration von Informationen aus Datenbanken. Die Auswahl und Indizierung der Informationen erfolgt dezentral.

2.1 Besondere Merkmale von SWING

Neben den klassischen Ausprägungen einer Suchmaschine, wie zum Beispiel Volltextsuche wurden in SWING diverse neue Features realisiert. Nachfolgend soll nun ein Abriß der exklusiven Merkmale von SWING skizziert werden. Für ausführlichere Informationen lese man in [HMDL97] nach.

- Besonderes Merkmal von SWING ist die Integration von Datenbankinhalten in die Suche. Der Vorteil ist, daß sich die Ergebnisse einer Datenbankfrage wesentlich besser in den Kontext einer Nutzeranfrage einpassen.
- Im Gegensatz zu den meisten international verfügbaren Suchmaschinen ist die SWING Suchmaschine auf Informationen aus und über Mecklenburg-Vorpommern beschränkt.

¹SWING ist erreichbar unter der URL <http://swing.informatik.uni-rostock.de>

- SWING ist die erste Suchmaschine, die eine Konsistenzsicherung der bereitgestellten Informationen ermöglicht. Dieser Konsistenzcheck erfolgt bei der Aufbereitung des Suchindexes. Die dabei festgestellten Konsistenzprobleme werden den Betreibern per elektronischer Post mitgeteilt.
- Neben den in einem üblichen HTML-Text vorkommenden Wörtern, werden von SWING eine Reihe spezifischer META-Tags ausgewertet. Mit Hilfe dieser Tags kann eine bessere Bewertung des Suchergebnisses auf die Benutzerwünsche vorgenommen werden.
- SWING unterstützt die Suchanfrage durch die Anwendung von linguistischem Wissen. Der Einsatz des Begriffslexikons erfolgt zum einen transparent als auch im Dialog mit dem Nutzer.
- Im Gegensatz zu vielen anderen Suchmaschinen erfolgt in SWING eine verteilte, parallele Suche auf verschiedenen Rechnern des Landes. Dadurch kann dem Nutzer ein weitaus schnelleres Anfrageergebnis präsentiert werden.

Zu diesen im Wesentlichen bereits realisierten Merkmalen, werden sich in nächster Zeit weitere Teile dazugesellen. Dazu gehört auch der, in dieser Studienarbeit konzipierte Abonnementdienst.

2.2 Ausnutzung der Suchmaschine durch den Abonnementdienst

Der zu entwickelnde Abonnementdienst ist von der Konzeption her unabhängig von einer speziellen Suchmaschine, das bedeutet, daß es sich überwiegend um eine eigenständige Anwendung handelt. Jedoch wurde dieser Abonnementdienst für die SWING-Suchmaschine entwickelt. Insofern nutzt er zur Anfragebearbeitung über eine Schnittstelle die von der Suchmaschine bereitgestellten Dokumentenreferenzen und Ranking-Ergebnisse aus.

Kapitel 3

Grundlagen

In den vorangegangenen Kapiteln wurden wiederholt Begriffe wie Agent, Internetseite, HTML u.ä. verwandt, ohne jedoch näher auf ihre Bedeutung einzugehen. Für Fachleute mögen diese und die Begriffe des nachfolgenden Kapitels selbstverständlich sein. Um aber auch anderen interessierten Lesern einen Überblick zu verschaffen, sollen die wichtigsten Grundlagen hier aufgeführt werden.

3.1 Datenbanken und WWW

3.1.1 HyperText Transfer Protocol (HTTP)

HTTP ist das Übertragungsprotokoll des WWW für Hypertextdokumente. Die Kommunikation zwischen Client und Server setzt dabei das Transportprotokoll TCP/IP voraus. HTTP selbst ist ein verbindungs- und zustandsloses Protokoll, das heißt, daß nachdem der Client einen Request an den Server gesandt hat, wird eine Verbindung zwischen diesen beiden aufgebaut. Der Server überträgt das angeforderte Dokument an den Client, anschließend wird die Verbindung beendet und der Server gibt seine Ressourcen frei.

Die Vorteile des zustandslosen Protokolls, alle Ressourcen freigeben zu können und dadurch eine Geschwindigkeitssteigerung auf Seiten des Servers zu erreichen, bilden gleichzeitig die wesentlichen Nachteile für größere Anwendungen im WWW, da keine Möglichkeit besteht, den Server in einen Zustand zu versetzen, auf den in einer späteren Anforderung reagiert werden kann.

3.1.2 Uniform Resource Identifier (URI)

Um eine Verbindung zwischen Client und Server aufzubauen, wird eine Netzwerkadresse benötigt. Im WWW wird das allgemeine Konzept zur einheitlichen

Adressierung beliebiger Internet-Ressourcen verwandt. Durch dieses Konzept lassen sich im WWW auch andere Dienste wie zum Beispiel FTP (File Transfer Protocol) und WAIS (Wide Area Information Servers) nutzbar machen.

Eine URI ist somit eine Internetadresse, die wiederum in URN (Uniform Resource Names) und URL (Uniform Resource Locator) unterteilt wird. Da URNs bisher noch nicht spezifiziert wurden, findet zur Zeit nur der URL Anwendung. Er hat im allgemeinen die Form: Protokoll://Host:Port/Pfad/File.

3.1.3 Common Gateway Interface (CGI)

Anfangs implementierten die Server-Entwickler die Mechanismen zum Aufruf externer Programme und deren Parameterübergabe nach eigenem Gutdünken. Dies führte jedoch recht bald zu Problemen, da Skripte beziehungsweise Programme nur auf den eigenen Servern liefen. Um die Ausführung eines Skriptes auch auf einem anderen Server zu ermöglichen, wurde von den damals führenden Server-Entwicklern von CERN und NCSA ein Standard festgelegt, der vorallem die Belegung einer Reihe von Umgebungsvariablen festschrieb. Über diese werden dem CGI-Skript Informationen über den Server und den vom Client empfangenen Request mitgeliefert. Die Umgebungsvariablen kann man grob in Client-Request-unabhängige und Client-Request-abhängige Variablen einteilen.

Zur Zeit bilden CGI-Skripte noch die größte Gruppe ausführbarer Programme im WWW, jedoch werden sich in den kommenden Jahren mehr und mehr andere Formen der Programmausführung durchsetzen (wie z.B. Java).

3.1.4 HyperText Markup Language (HTML)

HTML ist das Dokumentenformat im WWW schlechthin. Sie ist konform zu der von Goldfarb entwickelten SGML (Standard Generalized Markup Language) nach ISO-Norm 8879. SGML fungiert als Metasprache, mit deren Hilfe Dokumentensprachen spezifiziert werden können. Die Definition der spezifizierten Sprache wird in einer Document Type Definition (DTD) festgeschrieben. HTML ist eine Dokumentensprache, deren Strukturelemente in der HTML-DTD festgelegt sind.

Obwohl HTML 2.0 den Internetstandard im Hypertextbereich bildet, ist nicht gewährleistet, daß jedes HTML-Dokument das gleiche Erscheinungsbild auf unterschiedlichen WWW-Browsern aufweist. Die Ursache ist, daß viele der zur Zeit eingesetzten WWW-Browser Strukturelemente unterstützen, die über den HTML-Standard hinausgehen (Versionen 3.0, 3.2, 4.0).

3.1.5 Datenbankbindung an das World Wide Web

Um Datenbanken an das WWW anzubinden, müssen grundsätzlich ein Datenbankserver und ein WWW-Server bereitgestellt werden. Der WWW-Server stellt dabei die HTML-Seiten bereit. Der Datenbankserver ermöglicht den Zugriff auf die Daten in der Datenbank. Es existieren nun verschiedene Möglichkeiten, wie über das WWW auf die Datenbank zugegriffen werden kann [Loe97].

1. Auf die Datenbank wird mittels selbsterstellter Skripte oder Programme (Python, C, Perl) zugegriffen. Der Aufruf erfolgt, indem dem WWW-Server eine URL in Form eines CGI Programms mit optionalen Kommandozeilenparametern übergeben wird. Dieses Programm liefert als Antwort eine in der Regel fest kodierte HTML-Seite an den WWW-Server zurück, die dieser an den Client weiterleitet. Soll auf diese Antwort wiederum mit einer Datenbankabfrage reagiert werden, so muß sowohl ein neues CGI-Modul gestartet als auch eine neue Verbindung zum Datenbankserver aufgebaut werden.
2. Anstelle der vielen einzelnen Module wird hier nur ein einziges CGI-Programm aufgerufen und in den Adreßraum des WWW-Servers verlagert. Dieser Mechanismus ermöglicht es, einen ständigen Kontakt zur Datenbank aufrecht zu erhalten.
3. Wird die Ausführung der beiden vorangegangenen Methoden vom Server vorgenommen, so erfolgt bei dieser Methode eine lokale Programmabarbeitung beim Client. Dies geschieht mittels der Programmiersprache Java. Zu diesem Zweck wird die gesamte Datenbankschnittstelle als Applet vom HTTP-Server an den Client übergeben. Vorteil ist, daß der WWW-Client direkt mit dem Datenbankserver kommunizieren kann.

Die Vorteile der beiden letzten Methoden, eine Datenbankverbindung über einen längeren Zeitraum zu etablieren, werden aber durch bestehende Sicherheitslücken bei der Authorisierung von Datenbankbenutzern erheblich gemindert.

3.2 Agenten im Internet

Um einen möglichst hohen Nutzen aus den Möglichkeiten die sich durch das Internet bieten ziehen zu können, bedarf es einer neuen Kategorie von Werkzeugen. Eine dieser neuen Kategorien entsteht durch die Entwicklung von intelligenten Agenten.

3.2.1 Definition

Leider gibt es bis heute in der Literatur keine allgemein akzeptierte Definition eines intelligenten Agenten. So werden in [FG96] allein 11 verschiedene Definitionen von intelligenten Agenten aufgeführt und diese Zahl ließe sich mühelos erhöhen. Die Schwierigkeit einer einheitlichen theoretischen Basis liegt in dem interdisziplinären Charakter der Agenten. So stellt der Wissenschaftsbereich Künstliche Intelligenz andere Anforderungen als der der Informations- und Kommunikationssysteme.

Trotz der Probleme einer einheitlichen Definition existieren Gemeinsamkeiten die von [MRK96] folgendermaßen beschrieben wurden:

Definition: Ein Agent ist ein Softwareelement, das Teile eines Programms selbstständig ausführt. Typischerweise handelt er im Interesse des Nutzers, arbeitet weitestgehend autonom und kann mit anderen Agenten oder Nutzern kommunizieren. Mobile Agenten können sich zudem von einem System zu einem anderen bewegen, um auf entfernte Ressourcen zuzugreifen oder andere Agenten zu treffen und mit ihnen kommunizieren.

Agenten kann man prinzipiell den Kategorien menschliche Agenten, Hardwareagenten und Softwareagenten zuordnen. An dieser Stelle soll aber die Einschränkung getroffen werden, nur eine Betrachtung der Softwareagenten vorzunehmen.

3.2.2 Charakteristische Eigenschaften

Einige Eigenschaften intelligenter Softwareagenten sind bereits in der Definition angeklungen. Diesen sollen weitere folgen, um eine Abgrenzung gewöhnlicher Softwareprogramme von Softwareagenten vornehmen zu können. Es sei angemerkt, daß kein Agent alle genannten Eigenschaften besitzen muß. Vielmehr existiert eine breite Palette von Agenten mit einfacher Struktur bis hin zu hochkomplexen Agentensystemen.

[BZW98] unterteilen die Eigenschaften der Softwareagenten in interne und externe. Zu den internen zählen diejenigen, die die inneren Vorgänge des Agenten bestimmen. Die externen Charakteristika beschreiben das Zusammenspiel von Agenten untereinander beziehungsweise die Kommunikation von Mensch und Agent (vgl. Abbildung 3.1).

- **Reaktivität:** Sie ist eine der Grundanforderungen an intelligente Agenten und sollte daher von jedem Agenten in einer gewissen Form unterstützt werden. Als Reaktivität wird dabei die Fähigkeit eines Agenten bezeichnet, auf Einflüsse aus seiner Umwelt in angemessener Form reagieren zu können. Dazu sollte er über Sensoren oder ein internes Modell seiner Umwelt verfügen, aus dem er selbständig Rückschlüsse ziehen kann.

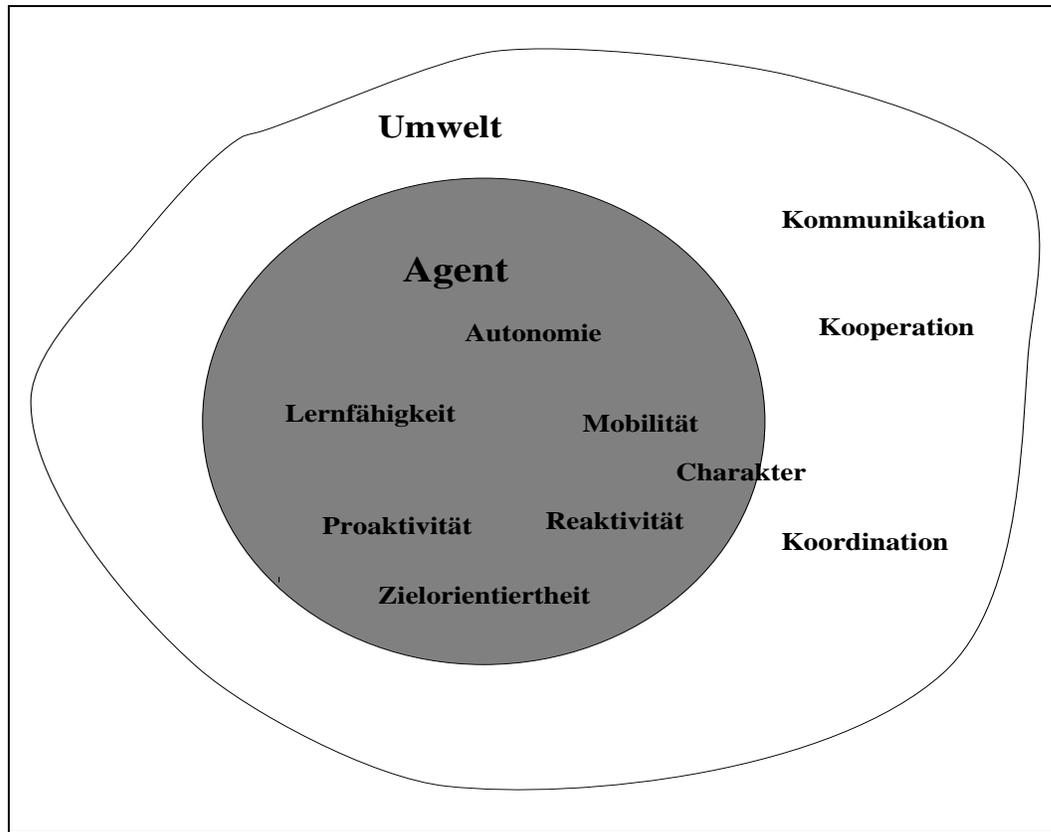


Abbildung 3.1: Charakteristische Eigenschaften intelligenter Agenten

- **Proaktivität/Zielorientiertheit:** Die Proaktivität geht über die Reaktivität hinaus, indem der Agent in bestimmten Situationen selbständig die Initiative ergreift, um auf ein bestimmtes Ziel hinzuwirken.
- **Schlußfolgerungs-/Lernfähigkeit:** Um überhaupt den Namen Agent verdienen zu können, muß ein Softwaresystem über ein gewisses Maß an Intelligenz verfügen. Die Spanne reicht hier von wenig intelligenten bis zu hochintelligenten Agenten. Sie setzt sich aus der internen Wissensbasis des Agenten, der Fähigkeit, Schlußfolgerungen aus den Inhalten der Basis ziehen zu können und der Möglichkeit, sich der Umwelt anzupassen, zusammen.
- **Autonomes Handeln:** Ein wesentliches Unterscheidungsmerkmal zwischen Agenten und anderen Softwareprogrammen ist ihre Fähigkeit zum autonomen Handeln. Sie besitzen die Gabe, ihr Ziel in einem gewissen Rahmen ohne Anweisung aus der Umwelt zu verfolgen und erreichen dadurch eine Entlastung des Anwenders.

- **Mobilität:** Der grundsätzliche Unterschied mobiler Agenten zu ihren stationären Entsprechungen ist die Fähigkeit, sich in einem Netzwerk frei von einem Ort zu einem anderen zu bewegen, um mit anderen Agenten zu kommunizieren oder um einen dort zur Verfügung stehenden Service in Anspruch zu nehmen. In der mobilen Umgebung erfolgt kein Austausch von Request/Response Nachrichten über RPC¹. Vielmehr wird die aufrufende Prozedur des Client direkt zum in Frage kommenden Server übertragen und dort lokal ausgeführt (Remote Programming).
- **Kommunikation/Kooperation:** Die Wahrnehmung seiner Aufgaben erfordert von dem Agenten unter Umständen die Interaktion mit seiner Umwelt. Dazu zählen zum einen die menschlichen Benutzer. Existiert desweiteren eine Agenten-Kommunikationssprache, so ist auch der Austausch von Informationen zwischen den Agenten möglich. Ist die Menge der Anfragen und Antworten genau spezifiziert, so spricht man von Kommunikation. Einen Schritt weiter geht die Kooperation, da hier ein Dialog zwischen Agenten stattfindet mit dem Ziel, eine gemeinsame Aufgabe zu lösen. Zu diesem Zweck muß die Kommunikationssprache um den Austausch von Zielvorstellungen, Absichten und Wissensbeständen der beteiligten Agenten erweitert werden.
- **Charakter:** Eine nicht zu unterschätzende Eigenschaft von Agenten ist ihr Charakter, dabei insbesondere Ehrlichkeit, Vertrauenswürdigkeit, Zuverlässigkeit und emotionales Verhalten. Kein Benutzer wird einem Agenten wichtige Aufgaben übertragen, wenn Zweifel an seiner Vertrauenswürdigkeit bestehen und dadurch zu befürchten ist, daß Informationen an unbefugte Agenten beziehungsweise Personen weitergeleitet werden. Um bei der Interaktion mit einem menschlichen Nutzer die Identität eines Softwareprogramm zu verbergen, schlüpfen Agenten häufig in die Rolle virtueller Benutzer. Um diese Rolle aber richtig auszufüllen, müssen sie Emotionen ausdrücken können.

Um nicht den Eindruck zu erwecken, daß intelligente Agenten tatsächlich Gefühle ausdrücken könnten, sei angemerkt, daß sie nur durch programmierte Reaktionen auf die Aktionen der Nutzer reagieren.

3.2.3 Klassifikation von Agenten

Mit Hilfe der im vorigen Abschnitt beschriebenen charakteristischen Eigenschaften kann eine Klassifikation vorgenommen werden. Dies kann in tabellarischer Form erfolgen, indem den Agentensystemen die zutreffenden Kriterien zugeordnet

¹Remote Procedure Call - Client/Server Kommunikation über den Aufruf von entfernten Prozeduren

werden. [BZW98] vertreten die Ansicht, daß der praktische Wert dieser Tabelle aufgrund der begrenzten Anschaulichkeit der Darstellung beschränkt ist. Dort erfolgt die Klassifikation in einem dreidimensionalen Koordinatensystem. Den drei gewählten Kriterien Mobilität, Intelligenzgrad und Agentenanzahl wird jeweils eine Koordinatenachse zugeordnet. Abbildung 3.2 zeigt die Darstellung, in der sich alle Agentensysteme klassifizieren lassen. Um nicht die Anschaulichkeit der dreidimensionalen Darstellungsweise durch eine zu komplexe Einteilung der Achsen zu verlieren, werden zur Beschreibung der Kriterien jeweils nur zwei Methoden verwendet. So wird der Intelligenzgrad eines Agenten in einfaches oder komplexes Verhalten untergliedert. Das Klassifikationskriterium Anzahl der Agenten wird durch die Anzahl der im System beteiligten Agenten gebildet. Die Mobilität eines Agentensystems ist gegeben, wenn die beteiligten Agenten die Fähigkeit besitzen sich in einem Netzwerk frei zu bewegen.

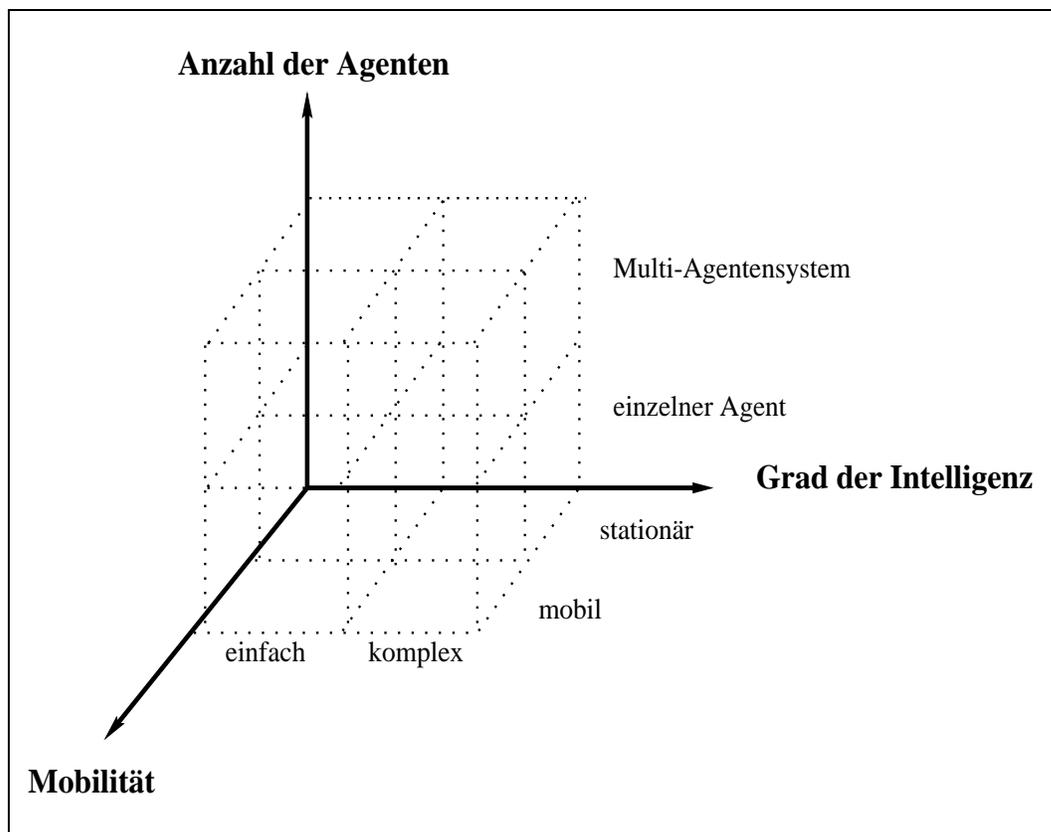


Abbildung 3.2: Klassifikationsmatrix nach [BZW98]

Zum Abschluß dieses Abschnittes soll am Beispiel eines Informationsagenten gezeigt werden, wie sich dieses Agentensystem in die Klassifikationsmatrix einfügt. Informationsagenten (vgl. Abbildung 3.3) besitzen in der Regel einen geringen

Intelligenzgrad und fallen daher in den Bereich einfacher Agenten. Bereits existierende Informationsagenten agieren bei ihrer Suche nach Informationen weitestgehend allein und sind überwiegend stationärer Natur.

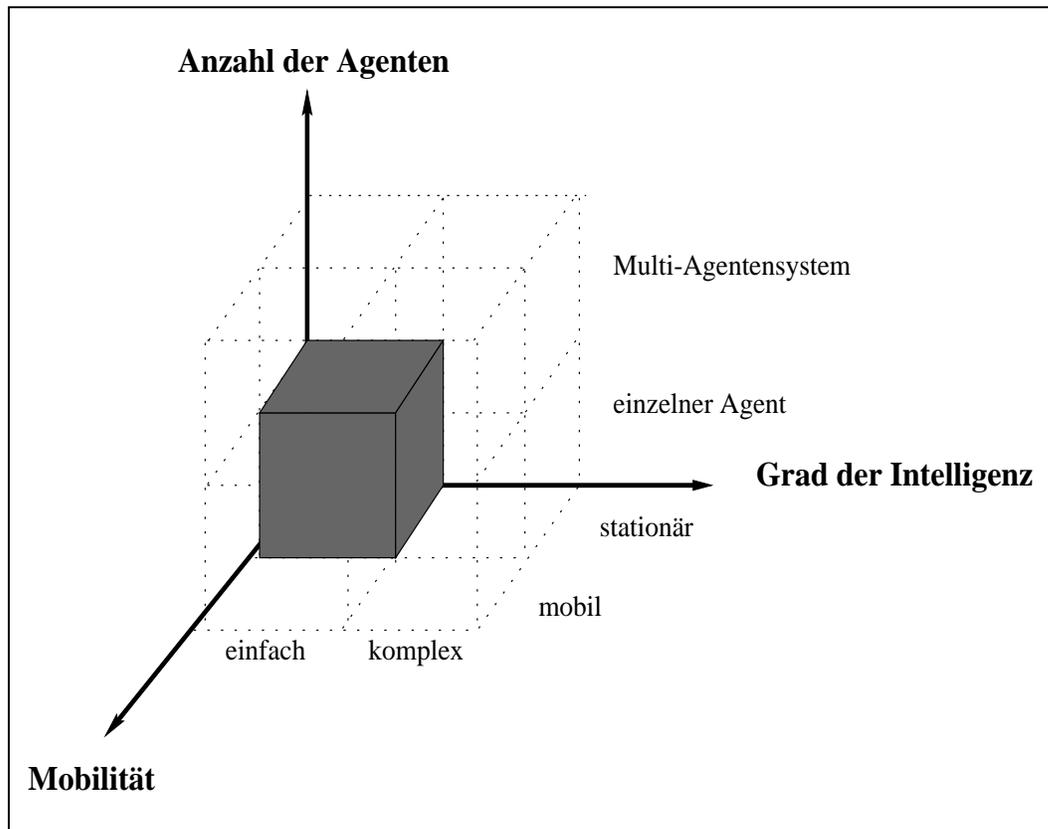


Abbildung 3.3: Einordnung von Informationsagenten in die Klassifikationsmatrix

3.3 Abonnementdienste

Im klassischen Sinne wird durch den Begriff Abonnement ein fest vereinbarter, regelmäßiger Bezug von Zeitschriften o.ä. beschrieben.

Fast jeder hat solch einen Abonnementservice bereits in irgendeiner Form in Anspruch genommen. Es stellt sich nun die Frage, welchen Vorteil ein Abonnementdienst im WWW bringen kann. Zur Beantwortung der Frage soll ein Beispiel über die Fernsehgewohnheiten herangezogen werden.

Möchte man regelmäßig über das Fernsehprogramm informiert sein, wird man wahrscheinlich eine Programmzeitschrift abonnieren. Nun ist es erfahrungsgemäß so, daß ein Zuschauer nicht alle Sendungen konsumieren wird, sondern nur

diejenigen, die seinen Interessen, Neigungen und Zeitvorstellungen entsprechen. Die Fernsehzeitschrift liefert aber Informationen, die über den für den Zuschauer relevanten Anteil hinausgehen. Um keine interessante Sendung zu verpassen, wird der Zuschauer die Programmzeitschrift nach seinen Interessen durchsuchen, was bei derzeit über 30 deutschen Fernsehkanälen bereits einen beträchtlichen Zeitaufwand mit sich bringt. Die Arbeit des Suchens und Filterns von TV-Sendungen durch den Konsumenten könnte durch den Einsatz eines intelligenten Fernsehagenten entfallen.

Woher weiß der Fernsehagent, für welche Sendungen ich mich speziell interessiere? Jeder Mensch besitzt bestimmte Interessen und Neigungen, denen er in irgendeiner Form nachgeht. Damit nun der Fernsehagent die Lieblingssendungen eines Zuschauers herausfiltern kann, muß er dessen Interessen und Fernsehgewohnheiten kennen. Diese Informationen bezeichnet man als das Profil eines Fernsehkonsumenten. In diesem Profil könnte beispielsweise festgehalten sein, daß sich ein Zuschauer für Sport und Musik interessiert und in der Regel zwischen 20.00 Uhr und 22.00 Uhr fernsieht. Mit Hilfe dieser Informationen wird der Fernsehagent diejenigen Sendungen selektieren, die über Sport bzw. Musik berichten und in der gewählten Zeit laufen. Je feiner dabei die Granularität des Benutzerprofils ist, desto zutreffender wird auch das Ergebnis des Fernsehagenten sein.

Im Abschnitt 3.2 wurde über die charakteristischen Eigenschaften von Agenten gesprochen. Obwohl der Fernsehagent bisher recht einfach gestrickt ist, besitzt er schon einige dieser Eigenschaften. So ist er sowohl reaktiv als auch autonom, da er selbständig und in regelmäßigen Abständen gefilterte Fernsehinformationen an den Konsumenten weiterleitet. Weiterhin muß er über kommunikative Fähigkeiten verfügen, um Informationen vom Benutzer entgegen nehmen zu können.

In einem nächsten Schritt könnte der Fernsehagent dahingehend erweitert werden, daß er aufgrund von Vergleichen mit anderen Nutzerprofilen oder durch das Sammeln von Erfahrungen dem Nutzer individuelle Empfehlungen für interessante Sendungen ausspricht.

Nimmt man eine Klassifikation des Fernsehagenten entsprechend der im Abschnitt 3.2.3 aufgestellten Matrix vor, so besitzt ein Fernsehagentensystem das Sendungen exakt nach Profilvergaben selektiert vorwiegend einfache Intelligenz. Die Anzahl der beteiligten Agenten wird von System zu System unterschiedlich sein. So wird es Fernsehagenten geben, die allein operieren, wiederum werden auch Multi-Agentensysteme existieren, die sich ihre Informationen durch Kommunikation mit anderen Agenten beschaffen. Diese Agenten werden dabei aber überwiegend stationär arbeiten. Daraus resultierend ergibt sich Abbildung 3.4.

Nun sind Abonnementdienste in der Praxis nichts Neues, so daß an dieser Stelle kurz auf die Unterschiede bestehender Systeme eingegangen werden soll. Betrachtet man Abonnementdienste, so kann man die Systeme nach zwei Kriterien klassifizieren. Zum einen handelt es sich dabei um die Form des Dienstes, das

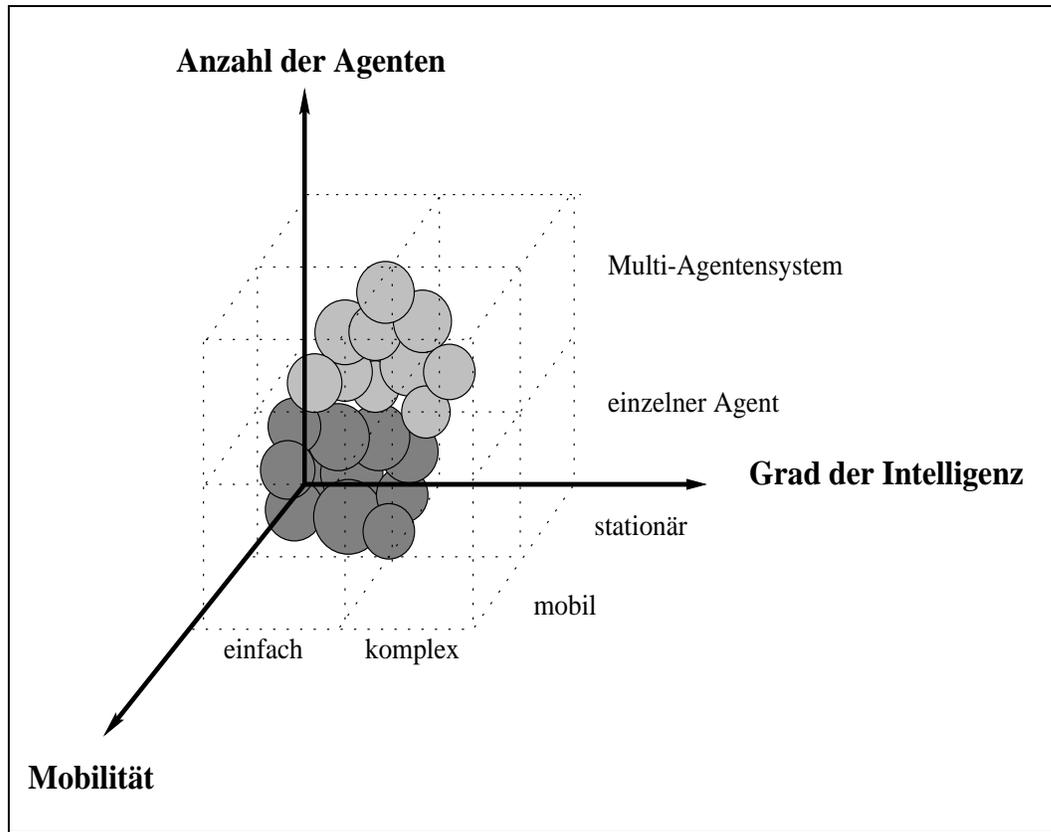


Abbildung 3.4: Klassifikationsmatrix eines Fernsehagentensystems

andere Kriterium gibt die Organisation des Services an.

- Form des Abonnementdienstes²
 - **Alertingdienst:** In einem förderativen, serviceorientierten System zur Verbreitung und Nutzung von Informationen treten Ereignisse auf, von denen nicht bekannt ist, wann sie auftreten, die jedoch für den Nutzer von Interesse sind. Ein Alertingdienst, ist wie der Name schon besagt ein Benachrichtigungsdienst, der den interessierten Nutzer über das Eintreten von Ereignissen informiert. Die Ereignisse werden dabei immer anbieterseitig ausgelöst.
 - **Profildienst:** Ein Profildienst filtert auftretende Ereignisse hinsichtlich nutzerspezifischer Interessen und informiert den Benutzer über relevante Ereignisse. Die Definition des erforderlichen Nutzerprofils kann auf verschiedene Arten erfolgen, zum Beispiel durch Beobachtung des Benutzerverhaltens, anhand von Benutzereigenschaften

²entnommen aus [Glo98]

oder durch Beispielvorgaben. Ein Profildienst kann reaktiv sein, das heißt er kann auftretende Ereignisse hinsichtlich der definierten Profile filtern. Er kann aber auch aktiv nach auftretenden Ereignissen bei Diensteanbringern fragen.

- Organisationsform des Abonnementdienste
 - Abonnementdienst ist beim Client (Anwender) organisiert.
 - Abonnementdienst ist beim Server (Diensteanbieter) organisiert.

Beide Formen haben sowohl Vor- als auch Nachteile. An dieser Stelle wird darauf verzichtet näher auf die Probleme der beiden Formen einzugehen, da dies im nächsten Kapitel vorgenommen wird.

Eine etwas andere Klassifikation für Systeme zur Verbreitung von Informationen, zu denen auch Abonnementdienste zählen, wird in [FZ96] vorgenommen. Dort wurden die Informationsbereitstellungsmechanismen miteinander verglichen.

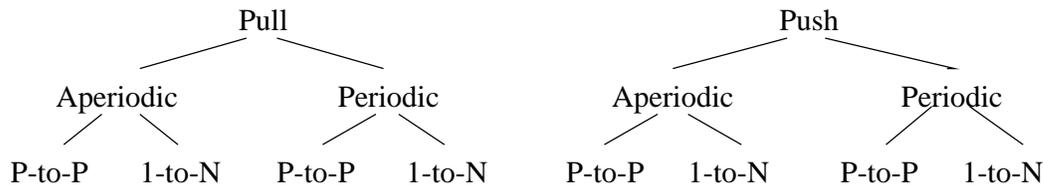


Abbildung 3.5: Mechanismen der Informationsbereitstellung

Entsprechend Abbildung 3.5 werden folgende Merkmale in Relation gesetzt:

- Push vs. Pull
- Periodic vs. Aperiodic
- Point to Point (P-to-P) vs. One to N (1-to-N)

1. **Push vs. Pull:** Wird der Transfer der Informationen auf eine Anfrage durch den Client initiiert, so spricht man vom von einem pull-basierten Ansatz. Wird in Kontrast dazu der Informationstransfer zum Client (u.U. auch mehrerer Clients) durch den Server initiiert, handelt es sich um einen push-basierten Ansatz.

Durch pull-basierte Anforderungen wird der Server in seiner Arbeit unterbrochen, um auf diese Anforderung zu reagieren. Er kann dabei nur sehr begrenzt Einfluß auf die Reihenfolge der Bereitstellung der Informationen beim Client nehmen. Auf eine pull-basierte Anfrage kann der Server nur

mit der augenblicklich zur Verfügung stehenden Informationsmenge reagieren. Neue Datenelemente oder Änderungen existierender Dokumente können vom Client nur durch periodisches Pollen ermittelt werden.

Die push-basierte Informationsbereitstellung vermeidet viele Nachteile pull-basierter Anfragen. So kann der Client in einem Profil eine Beschreibung der für ihn interessanten Daten ablegen. Diese Informationen stellen eine Anfrage des Client an den Server dar. Diese Anfrage wird wiederkehrend mit den vorhandenen Daten verglichen. Stimmen neue bzw. geänderte Daten mit dem Profil eines Clients überein, so werden diese Datenelemente dem Client bereitgestellt.

2. **Periodic vs. Aperiodic:** Sowohl push- als auch pull-basierte Anfragen können in einer zyklischen oder azyklischen Form ablaufen. Periodische Datenbereitstellung erfordert einen festes Zeitintervall, während für die unregelmäßige Datenübertragung kein Zeitplan benötigt wird. Beim periodischen Push beispielsweise werden die Daten aufgrund eines vorher festgelegten Zeitplanes übertragen. Demgegenüber erfolgt die Datenübertragung beim azyklische Push aufgrund des Vorhandensein neuer bzw. geänderter Datenelemente.
3. **P-to-P vs. 1-to-N:** Bei der Nutzung einer Punkt zu Punkt Kommunikation werden Daten von einer Quelle zu einem Zielsystem gesendet. Die 1 zu N Kommunikation erlaubt mehreren Zielmaschinen Daten zu empfangen, die durch eine Datenquelle ausgesendet wurden (bspw. Broadcast, Multicast). Mailing-Listen stellen beispielsweise eine 1 zu N-Multicast-Datenübertragung dar. Eine an die Mailing-Liste gesendete Nachricht wird an alle eingetragenen Abonnenten dieser Liste weitergeleitet.

Bezieht man die Charakteristiken von [FZ96] auf den SWING-Abonnementdienst, so stellt dieser einen push-basierten Ansatz dar, der in periodischer Form eine Punkt zu Punkt Verbindung zu den Nutzern dieses Dienstes aufbaut. Dabei wurde außer acht gelassen, daß der Abonnementdienst auch als reine Suchmaschine fungieren kann. Im Gegensatz dazu werden dem Nutzer der SWING-Suchmaschine Informationen über eine Punkt zu Punkt-Verbindung, pull-basiert und azyklisch bereitgestellt.

Kapitel 4

Architekturmodell eines Abonnementdienstes

In der Einleitung zur Studienarbeit wurde bereits angedeutet, daß Suchmaschinen nicht das Allheilmittel zur Bewältigung der Informationsflut im Internet darstellen. Aus diesem Grund gehen immer mehr Informationsanbieter dazu über, Informations-Abonnementdienste anzubieten. Dabei entwickelt jeder seine eigenen Benutzerschnittstellen, obwohl beim Vergleich der Systeme Gemeinsamkeiten zu erkennen sind. Der Grund sind fehlende Standards. Es existieren zwar Bestrebungen, Standards zu verwirklichen, diese sind aber hauptsächlich auf einer niederen Ebene angesiedelt (Kommunikationsprotokolle, Syntax der Anfragesprache). In diesem Kapitel soll ein Architekturmodell eines Abonnementdienstes vorgestellt werden, das zum einen auf bekannte Techniken wie dem Informationsaustausch durch mobile Agenten zurückgreift, aber auch neue Kommunikationsmöglichkeiten schafft, durch die Voraussetzung von Standards auf der Anwendungsebene.

Im Abschnitt 4.1 sollen zunächst die Probleme der Abonnementdienste aufgedeckt werden. Daraus resultierend werden im Abschnitt 4.2 die Anforderungen an ein Architekturmodell beschrieben. Abschnitt 4.3 stellt dann das allgemeine Architekturmodell im Konkreten vor.

4.1 Probleme der Abonnementdienste

Im Abschnitt 3.3 wurden bereits die Bedeutung und die Unterscheidungskriterien von Abonnementdiensten beschrieben. In diesem Abschnitt sollen nun speziell die Probleme existierender Dienste herauskristallisiert werden, um diese bei dem Entwurf eines neuen Architekturmodells zu umgehen.

Der wesentliche Aufbau eines Abonnementdienstes, unabhängig von seiner Orga-

nisationsform entspricht der Abbildung 4.1 [Her96]. Der Agent bildet die zentrale Rolle in dieser Drei-Schichten-Architektur. Die Nutzer haben einen bestimmten Informationsbedarf zu deren Deckung sie sich eines Agenten bedienen. Der Agent verfügt nun nicht unbedingt selbst über die nachgefragte Information und bedient sich wiederum Informationsanbietern. Das Anfrageergebnis der Informationsanbieter wird vom Agenten unter Umständen noch einer Filterung unterzogen, bevor es dem Nutzer bereitgestellt wird.

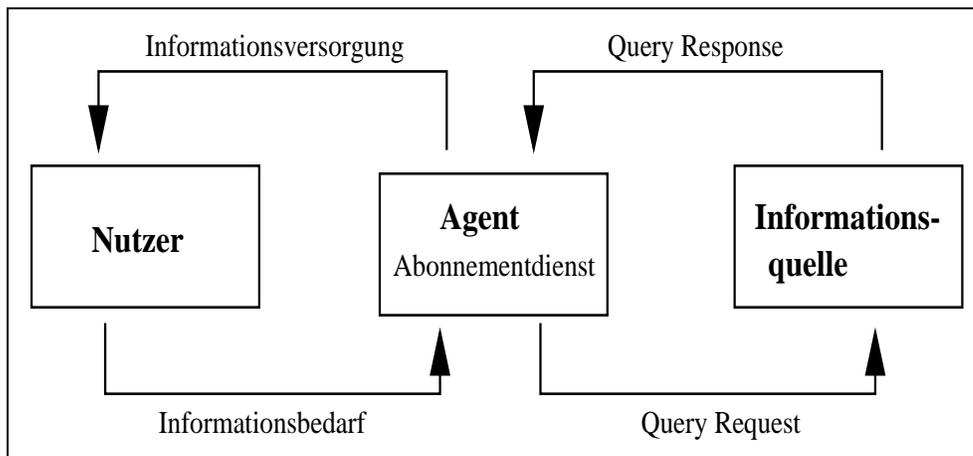


Abbildung 4.1: Aufbau eines Abonnementdienstes

Aufgrund der zentralen Rolle des Agenten sollen die Probleme

- zwischen dem Nutzer und dem Agenten
- zwischen dem Agenten und der Informationsquelle
- und des Agenten selbst

analysiert werden.

1. Probleme zwischen Nutzer und Agent

- Allgemein
 - Jedem Abonnementdienst ist die Verwendung eines Nutzerprofils gemein. Zwar werden je nach Komplexität mehr oder weniger Präferenzen festgelegt, jedoch hat der Nutzer in der Regel wiederkehrende Fragen zu beantworten. Ändern sich seine Präferenzen, die in mehr als einem Abonnementdienst gespeichert sind, so müssen diese Änderungen in jedem Dienst vollzogen werden.

- Clientbasierte Systeme
 - Die Verbesserung der Informationsgüte durch Lernalgorithmen ist nur durch Beobachtung des Nutzers und durch die interne Wissensbasis des Agenten möglich, nicht jedoch durch Vergleich mit anderen Nutzerprofilen und deren Verhaltensweisen.
 - Die Installation neuer Versionen eines Abonnementdienstes und die damit verbundenen Probleme obliegen dem Anwender.
 - Abonnementdienste existieren nicht für alle Betriebssysteme, so daß von vornherein eine Reihe von Nutzern ausgeschlossen wird.
- Serverbasierte Systeme
 - Je nach System entstehen bei einigen nur Kommunikationskosten für die Dauer der Updatesitzung des Nutzerprofils, die abonnierte Information wird beispielsweise direkt als Email an den Nutzer übertragen. Andere System informieren den Benutzer zwar über neue Informationen, die Bereitstellung erfolgt aber direkt beim Abonnementsservice, so daß sämtliche Kommunikationskosten zu Lasten des Nutzer gehen.
 - Obwohl der Abonnementdienst für den Nutzer einen seriösen Eindruck erweckt, ist nicht ausgeschlossen, daß mit den Daten des Nutzerprofils vom Dienstanbieter Mißbrauch betrieben wird.
 - Im Gegensatz zu den clientbasierten Systemen kann dieser Abonnementdienst von allen WWW-Teilnehmern genutzt werden. Aufgrund der Heterogenität der Hard- und Software der Anwender wird aber kaum auf deren Bedürfnisse eingegangen.

2. Probleme zwischen Agent und Informationsquelle

- Agent greift auf eigene Informationsquelle zurück
 - Je nach Art der Information kann der Umfang unter Umständen nicht ausreichend sein.
 - Mit Zunahme der Informationsmenge wächst auch die Gefahr eines schlechten Aktualisierungsgrades.
- Agent fragt vorgegebene Informationsquellen ab
 - Es wird eine größere Auswahl an Informationen getroffen, allerdings werden nur die Informationsquellen erreicht, zu denen die Schnittstellen bekannt sind.

3. Probleme des Agenten selbst

- Um eine gute Qualität der gefilterten Information und eine optimale Anpassung an den Nutzer zu erreichen, müssen Agenten entwickelt

werden, die über die Anwendung fester Regeln hinausgehen. In der Praxis sind aber Agenten, die Methoden der Künstlichen Intelligenz anwenden bisher kaum realisiert.

4.2 Anforderungen an Abonnementdienste

Aus den Problemen des letzten Abschnittes können nun die Anforderungen abgeleitet werden, die bei der Umsetzung in ein Architekturmodell eines Abonnementdienstes Berücksichtigung finden sollen.

- Das Architekturmodell darf den Agentenentwicklern nicht die Verwendung einer bestimmten Programmiersprache aufzwingen, vielmehr muß durch die Verwendung einer einheitlichen Schnittstellendefinition die Kommunikation und Kooperation erreicht werden.
- Die verwendete Agentensprache muß Plattformunabhängigkeit und ein hohes Maß an Sicherheit gewährleisten.
- Jeder Nutzer besitzt genau ein Benutzerprofil, auf das jeder Agent, der Informationen aus diesem Profil benötigt, zugreifen kann. Dies erfordert nicht nur eine standardisierte Schnittstelle zwischen dem Nutzer und dem Dienstanbieter, sondern auch Richtlinien für das Nutzerprofil selbst.
- Der Abonnementdiensteanbieter kann sich die Informationen an beliebigen Stellen des Internets zusammensuchen, da alle Informationsanbieter über standardisierte Schnittstellen verfügen.
- Da verschiedene Anbieter von Abonnementdiensten auf dasselbe Nutzerprofil zugreifen können, müssen hohe Sicherheitsmechanismen realisiert werden, die einen Datenmißbrauch verhindern beziehungsweise minimieren.
- Das Nutzerprofil enthält neben persönlichen Präferenzen auch Angaben zum verwendeten System, so daß vom Dienstanbieter auf diese Gegebenheiten Rücksicht genommen werden kann.
- Da sich der für die Informationsfilterung zuständige Agent beim Dienstanbieter befindet, können Rückschlüsse sowohl aus dem Verhalten des Nutzers selbst als auch durch Präferenzenvergleich mit anderen Nutzern gezogen werden.
- Der Service sollte dem Nutzer die Informationen in einer übersichtlichen Form bereitstellen, ohne das für den Nutzer zusätzliche Kosten entstehen.

4.3 Das Architekturmodell

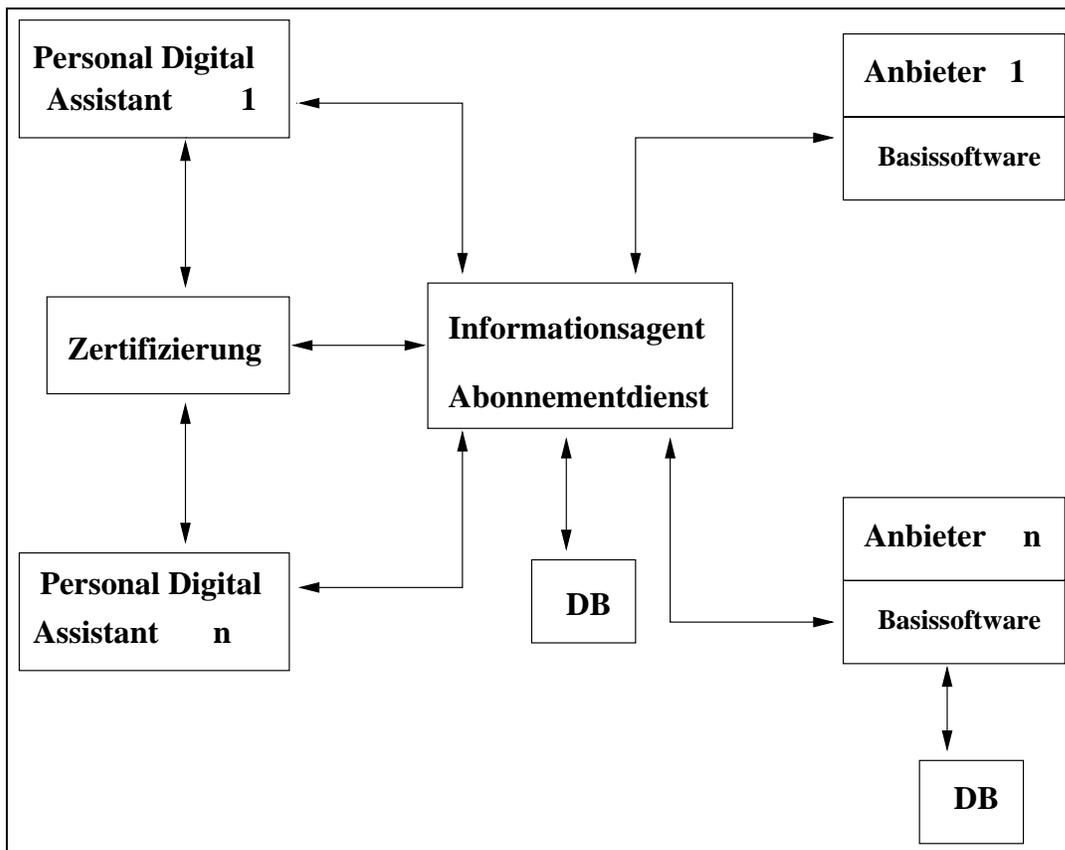


Abbildung 4.2: Architekturmodell eines Abonnementdienstes

Mit dem in Abbildung 4.2 vorgestellten Architekturmodell soll die Möglichkeit eines globalen Agentensystems geschaffen werden, in dem Anbieter ihre Informationen über einheitliche Schnittstellen offerieren, verschiedenste Dienste auf diese Informationen zugreifen können und Nutzer ihre Profilinformatoren den Diensten in einer genormten Art und Weise zur Verfügung stellen.

Die starre Client-Server-Architektur bestehender Abonnementdienste wird von diesem Modell aufgebrochen. Alle Komponenten (PDA, Agent, Anbieter, Zertifizierungsinstanz) müssen sowohl in die Rolle des Client als auch in die Rolle des Servers schlüpfen können. Der Personal Digital Assistant (PDA) ist beim Nutzer beziehungsweise Provider angesiedelt, da dieser das Profil des Anwenders und seine technische Basis verwaltet. Die Verwaltung des Informationsagenten unterliegt dem Anbieter des Abonnementdienstes. Da alle Informationsagenten auf dasselbe Profil des Nutzers zugreifen, muß über eine unabhängige Zertifizierungsinstanz die Berechtigung für den Zugriff eingeholt werden. Möchte ein Nutzer die Dienste

eines Abonnementanbieters in Anspruch nehmen, fordert der Informationsagent bei der Zertifizierungsstelle einen Zugriffsschlüssel an. Die Zertifizierungsstelle fragt beim PDA nach, ob der Informationsagent berechtigt ist, auf das Profil zuzugreifen. Bestätigt dies der PDA, erhalten sowohl der PDA als auch der Informationsagent einen Schlüssel. Mit diesem Schlüssel erkennt zum einen der PDA, daß der Informationsagent berechtigt ist, auf das Nutzerprofil zuzugreifen. Zum anderen gilt dieser Schlüssel als Authorisation des Nutzer beim Informationsagenten, so daß eine explizite Authentifikation durch den Nutzer entfällt. Möchte ein Nutzer Abstand vom Service nehmen, so fordert er den PDA auf, den Schlüssel zu löschen. Dazu muß sich dieser an die Zertifizierungsstelle wenden. Diese informiert den Informationsagenten, daß ein Löschbegehren eines Nutzers vorliegt. Existieren durch den Informationsagenten berechnete Forderungen an den Nutzer, so kann die Zertifizierungsstelle den Löschwunsch verweigern, andernfalls kann der PDA den Schlüssel jetzt löschen.

Die Informationsbeschaffung und Filterung erfolgt durch einen mobilen Agenten des Informationsagenten. Dieser mobile Agent hat die Fähigkeit, sich zu diversen Anbietern zu bewegen, Information des Anbieters zu sammeln und dort bereits zu filtern. Damit der Informationsagent nicht ziellos nach Anbietern suchen muß, kann er Auskunftsagenten zu Rate ziehen, die die Adressinformationen der Informationsanbieter enthalten. Bei der Migration des Agenten von einem Knoten des Netzes zu einem anderen ist er in der Lage, Störungen des Netzes auf seinem Verbindungsweg zu einem Knoten zu erkennen und durch die Wahl eines alternativen Weges zu umgehen. Sowohl der Informationsagent als auch der Auskunftsagent sollten über eine Ontologie verfügen, die Synonyme, Akronyme und Mehrfachbedeutungen von Begriffen enthalten. Sollen beispielsweise Informationen zum Begriff Golf gefunden werden, kann es sich um ein Auto, eine Meeresströmung beziehungsweise eine Sportart handeln. Falls also der Agent nicht im vorhinein die Bedeutung festgelegt hat, sollte der Auskunftsagent ihn darüber informieren.

Um eine hohe Ergebnisqualität zu erreichen, ist es wichtig, daß von den Anbietern neben semi- und unstrukturierten Daten auch strukturierte Informationen aus Datenbanken bereitgestellt werden. Im Gegensatz zu Teilen eines aus dem Kontext gerissenen Textdokumentes besitzen Daten einer Datenbank ein hohes Informationsniveau. Dazu ist es jedoch erforderlich, daß Anbieter die Schemainformationen ihrer Datenbanken über eine Funktion des Agenteninterfaces den mobilen Agenten zur Verfügung stellen. Da nicht in jedem Fall zu erwarten ist, daß das Metaschema die Suchanfrage enthält, könnte der Agent mit Hilfe seiner Ontologie prüfen, ob die Metainformationen in einer Beziehung zum Suchstring stehen. Sucht der Agent nach Informationen über den Sport Golf, das Schema enthält aber nur das Attribut Sportart, sollte der Agent erkennen, daß Golf eine bestimmte Sportart darstellt. Daraufhin bittet er den Anbieter Informationen zur Sportart Golf aus der Datenbank zu selektieren. Das Anfrageergebnis kann daraufhin vom Agenten nach weiteren Kriterien gefiltert beziehungsweise direkt

in die Ergebnismenge aufgenommen werden.

Aufgrund der standardisierten Schnittstellen sowohl auf der Nutzer- als auch auf der Anbieterseite, ist das Agentensystem in einem hohen Maße skalierbar. Für den Informationsanbieter genügt die Installation einer Basissoftware, um durch mobile Agenten eines Abonnementdienstes erreichbar zu sein. Ähnliches gilt für Benutzer der Dienste, die durch die Bereitstellung des PDA Teilnehmer des Agentensystems werden.

4.3.1 Der Personal Digital Assistant

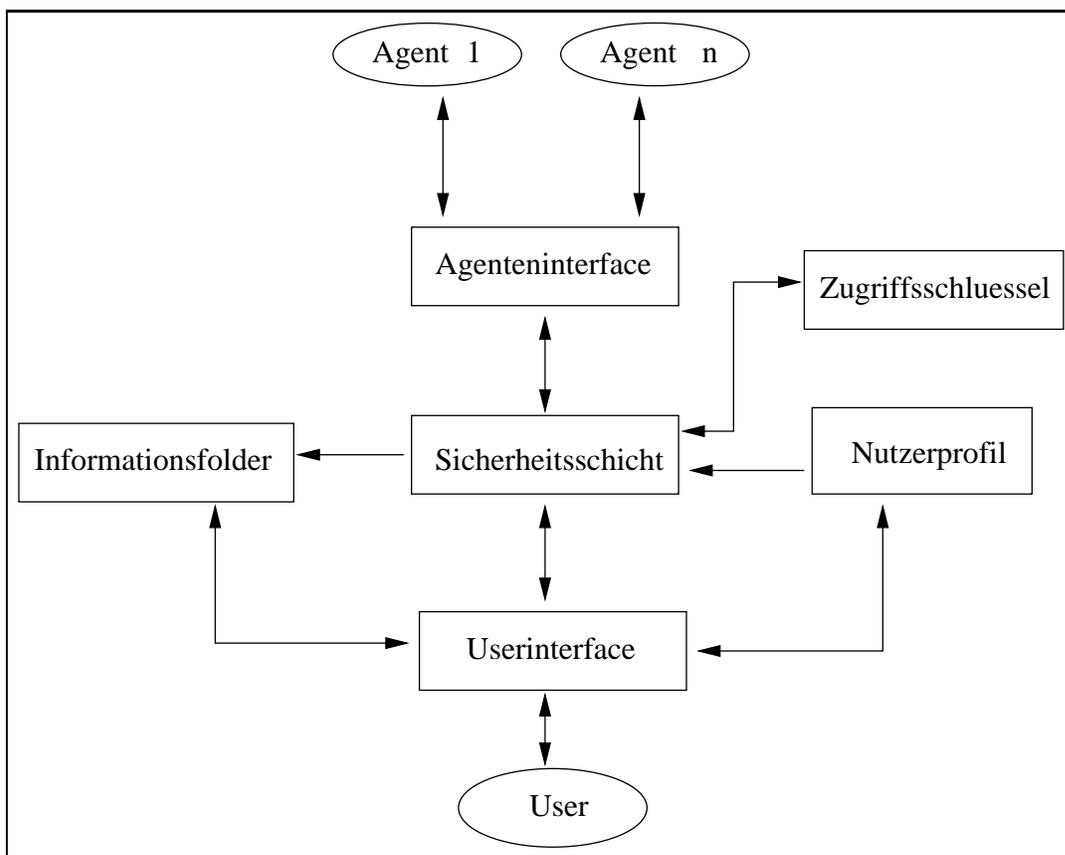


Abbildung 4.3: Aufbau des Personal Digital Assistant

Abbildung 4.3 zeigt den schematischen Aufbau des Personal Digital Assistant¹ (PDA). Im letzten Abschnitt wurde bereits über die Funktion des PDA im allgemeinen gesprochen. In diesem Abschnitt soll detaillierter auf den PDA und seine Komponenten eingegangen werden.

¹Die Bezeichnung Personal Digital Assistant wird u.a. für mobile Organizer verwendet. Dies ist in diesem Zusammenhang aber nicht gemeint.

Durch die Verwendung des PDA als eigenständigen Teil soll eine Entkoppelung zwischen Informationsagent und Nutzerprofil und zwischen Informationsbeschaffung/Bereitstellung und der Inanspruchnahme erreicht werden. Da die Informationsversorgung durch den Agenten initiiert wird, senken sich auch die Kommunikationskosten des Benutzers.

Die Kontrolle und Ausführung aller aktiven Agenten auf dem PDA obliegen dem Agenteninterface. Sowohl das Agenteninterface als auch die Attribute des Nutzerprofils besitzen eine einheitliche, öffentlich bekannte Form, so daß allen Agenten, die Informationen aus dem Nutzerprofil benötigen, eine gemeinsame Basis geschaffen wird.

Das Agenteninterface übernimmt die Prüfung der Zutrittsberechtigung eines Agenten, die Verwaltung der Agenten sowie deren Codeausführung. Zur Durchführung greift das Agenteninterface als auch das User-Interface auf die Sicherheitsschicht zurück. Diese nimmt eine Ver- und Entschlüsselung der Daten vor. Die Codierung der Informationen muß in einer Form erfolgen, die es jedem autorisierten Agenten aber auch der Sicherheitsschicht des PDA ermöglicht, diesen Prozeß umzukehren. Eine Möglichkeit wäre den Zugriffsschlüssel des Agenten zu verwenden.

Die Zutrittsberechtigung zu dem PDA erhalten Agenten, deren Zugriffsschlüssel zu einem Schlüssel aus der Schlüsselmenge des PDA paßt. Diese Agenten dürfen sowohl Informationen aus dem Nutzerprofil abrufen, als auch Informationen in den Folder schreiben. Desweiteren zugriffsberechtigt sind spezielle Agenten wie zum Beispiel Agenten der Zertifizierungsstelle. Diese Agenten sind aber in ihrem Aktionsradius auf den PDA eingeschränkt. Sie sind nur berechtigt, einen bestimmten Status abzufragen, beziehungsweise den neuen Zugriffsschlüssel für einen Agenten zur Verfügung zu stellen.

Möchte ein Nutzer sein Profil ändern beziehungsweise bereitgestellte Informationen aus den Folders lesen, ändern oder löschen, kann er dies mit Hilfe des User-Interface vornehmen. Das User-Interface könnte ein Web-Browser sein, in den beispielsweise ein spezieller Profileditor integriert ist. Im letzten Abschnitt wurde bereits angedeutet, daß das Löschen eines Informationsagenten nicht ohne weiteres durch den Nutzer erfolgen darf. Solange es sich um einen kostenlosen Service handelt, mag dies nicht einleuted sein. Da aber auch kostenpflichtige Dienste auf den PDA zugreifen können, darf sich ein Nutzer nicht einfach durch Löschen des Zugriffsschlüssels seiner Zahlungsverpflichtung entziehen.

4.3.2 Der Informationsagent

Der Informationsagent des Abonnementdienstes kann von Abonnementdienst zu Abonnementdienst variieren. Abbildung 4.4 soll daher nur ein Beispiel für einen möglichen Aufbau geben. Da aber der Informationsagent einen Datentransfer mit sehr sensiblen Daten (geheime Schlüssel, elektronisches Geld) vornimmt,

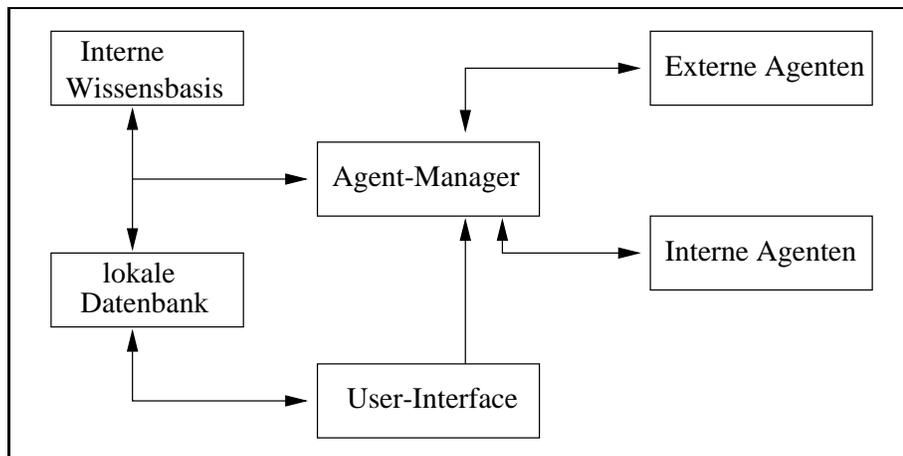


Abbildung 4.4: Beispiel eines Informationsagenten

werden an die verwendete Agentensprache in punkto Sicherheit hohe Ansprüche gesetzt. So muß die Sprache die Objekte vor Angriffen von außen abschirmen können. Weitere Kriterien der Agentensprache sind Objektorientiertheit, Multi-tasking, Verarbeitung von Signalen, die Unterstützung der Kommunikation über TCP/IP und die Möglichkeit der persistenten Datenhaltung. Die externen Agenten müssen zudem die standardisierten Schnittstellen von Anbietern, Zertifizierungsstelle und PDA unterstützen.

Das User-Interface ist die Kommunikationsschnittstelle des Anwenders zum Informationsagenten. Es repräsentiert für den Nutzer einen Assistenten, der seine Wünsche entgegennimmt und über den Agenten-Manager eine Verbindung zu den Agenten herstellt, um seine Probleme zu lösen. Daß der Informationsagent sich zur Lösung unter Umständen externer Informationsquellen bedient, muß für den Nutzer völlig transparent bleiben. Für die wesentliche Aufgabe des Informationsagenten des Abonnementdienstes, der regelmäßigen bzw. ereignisorientierten Versorgung des Nutzers mit Informationen, stehen ihm die interne Wissensbasis, die lokale Datenbank, der Agenten-Manager und die externen und internen Agenten zur Verfügung. Die interne Wissensbasis enthält ein Modell der Umwelt, aus dem Wünsche, Ziele und Intentionen abgeleitet werden können. In der lokalen Datenbasis sind zum einen Informationen der Nutzer enthalten, die nicht aus dem Nutzerprofil abgeleitet werden können (Zugriffsschlüssel, Häufigkeit der Informationsbereitstellung u.ä.), zum anderen können dort die Ontologie, eigene Informationen u.ä. abgelegt sein. Aufgrund der Komplexität der Gesamtaufgabe, wird dieses in Teilprobleme zerlegt. Der Agenten-Manager stellt dabei einen Moderator für die zu lösenden Teilprobleme dar. Er vergibt diese an die internen und externen Agenten und übernimmt die Koordination. Beispiele für Teilprobleme der externen Agenten sind die Informationsbeschaffung und Filterung bei

den Anbietern, die Schlüsselbeschaffung bei der Zertifizierungsstelle, die Selektion des Nutzerprofils und die Bereitstellung der abonnierten Information. Zu lösende Teilprobleme der internen Agenten sind die Aufbereitung der Ergebnismenge entsprechend der Systemeinstellungen des Nutzers und die Anwendung der Wissensbasis auf Informationen und Nutzerprofil.

4.3.3 Der Informationsanbieter

Der prinzipielle Aufbau der Basissoftware des Informationsagenten aus Abbildung 4.5 entspricht im wesentlichen dem Aufbau des PDA aus dem Abschnitt 4.3.1. So übernimmt das Agenten-Interface des Informationsanbieters ähnliche Aufgaben wie das Agenten-Interface des PDA. Die ankommenden Agenten werden durch die Schnittstelle identifiziert, verwaltet und ausgeführt. Allen Agenten steht dabei eine einheitliche Umgebung zur Verfügung. So bietet die Agentenschnittstelle Funktionen zum Auslesen der Schemainformationen und Selektion der Information der bereitgestellten Datenbanken an. Die Agenten der Abonnementdienste müssen daher nicht von vornherein wissen, ob ein Anbieter bestimmte Informationen zur Verfügung stellt, sondern können ihre Suchfunktionen auf die Schemainformationen der Datenbanken der Anbieter anpassen. Damit durch die Suchfunktionen der Agenten auf den Datenbanken des Informationsanbieters kein Sicherheitsrisiko entsteht, erfolgt der eigentliche Zugriff auf die Datenbank über die Sicherheitsschicht, in der eine Ver- bzw. Entschlüsselung der Daten vorgenommen wird.

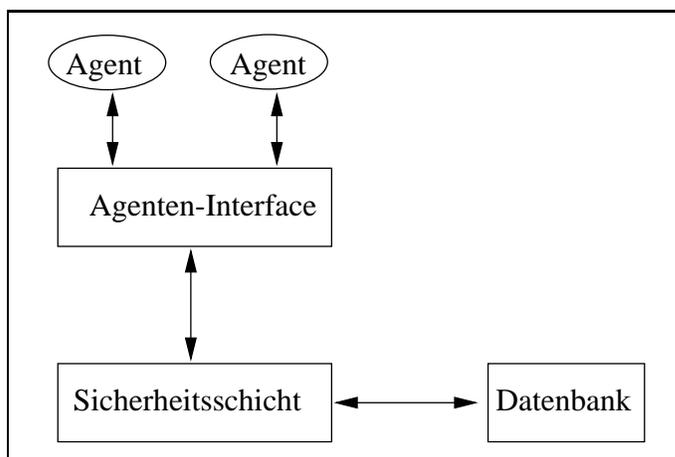


Abbildung 4.5: Aufbau der Basissoftware des Informationsanbieters

4.3.4 Klassifikation des Abonnementdienstes

Abbildung 4.6 nimmt eine Einordnung des vorgestellten Abonnementdienstes entsprechend der Klassifikationsmatrix aus Abschnitt 3.2.3 vor. So handelt es sich um ein Multiagentensystem, deren Agenten überwiegend mobil agieren und mit einfachem bis komplexem Intelligenzverhalten ausgestattet sind.

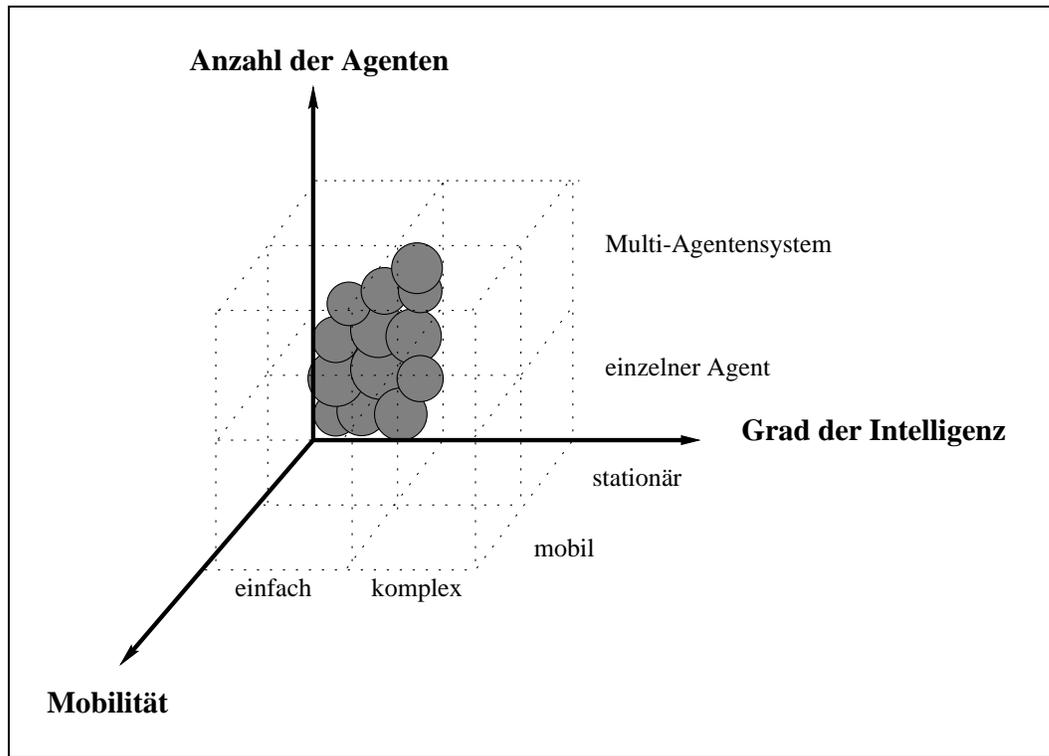


Abbildung 4.6: Klassifikation des Abonnementdienstes

4.3.5 Zusammenfassung

Abschließend soll auf mögliche Probleme bei der Realisierung dieses Modells eingegangen werden.

Die vorgestellte Architektur stellt ein globales Agentensystem dar, daß eine Dezentralisierung der bereitgestellten Informationen vorsieht. Um diesen Vorschlag jedoch in die Realität umzusetzen, müssen einige Hürden genommen werden. Das größte Hindernis bildet sicherlich die Forderung nach standardisierten Schnittstellen und der Vergabe von Zugriffsschlüsseln durch eine einheitliche Instanz. Obwohl, beispielsweise in [CC97], Ansätze einer Normung von mobilen Agenten auf

der Basis von CORBA² existieren, ist auf längere Zeit nicht mit einer Umsetzung in die Praxis zu rechnen. Eine weitere Anforderung ist ein sicheres und breitbandiges Netz, da der überwiegende Teil der Daten erst von entfernten Knoten zusammengetragen werden muß. Für dieses Modell wurde die Verwendung von mobilen Agenten vorgeschlagen, was zusätzlich eine gute Hardwareausstattung der Informationsanbieter erfordert, um eine prompte Ausführung der eintreffenden Agenten zu ermöglichen. Sicherlich wäre für dieses Modell auch die Verwendung von statischen Agenten möglich, die über einen Nachrichtenaustausch Informationen abfragen. Mobile Agenten bieten aber den Vorteil das ein großer Teil ihrer Kommunikation lokal beim Anbieter stattfindet, daß bedeutet, das Informationsfilterung nicht beim Abonnementdienst stattfindet, sondern beim Anbieter und das nur der relevante Anteil über das Netz transportiert wird. Zudem muß der mobile Agent nicht ständig Kontakt zu seinem Ausgangsort halten, er kann bei der Migration Netzstörungen erkennen und alternative Wege einschlagen. Leider gibt es bisher kaum kommerzielle Anwendungen die mobile Agenten einsetzen, so daß ein Netzwerk, in dem tausende mobiler Agenten umherwandern bisher nicht erforscht ist. Möglicherweise kehren sich auch die Vorteile der mobilen Agenten durch den mittransportieren Overhead ihrer Reaktionsfähigkeit um, so daß das obige Agentensystem nur noch zu einem gewisser Teil von mobilen Agenten ausgeführt wird, der andere Teil aber durch Nachrichtenaustausch. Beim Einsatz von statischen Agenten gelten ähnliche Anforderungen wie für die Agentensprache. Es sollte kein festgelegtes Kommunikationsprotokoll existieren, vielmehr müssen Translationsmöglichkeiten zwischen verschiedenen Protokollen bestehen. Die Forderung nach einem kohärenten Nutzerprofil ist in diesem System sicherlich das geringste Problem, da auch bereits Vorschläge für die Schaffung einer Plattform für Benutzerpräferenzen³ gemacht wurden.

²Common Object Request Broker Architecture

³Platform for Privacy Preferences - [W3Con]

Kapitel 5

Evaluierung ausgewählter Abonnementdienste

Nachdem im letzten Kapitel ein Vorschlag für zukünftige Profildienste unterbreitet wurde, sollen in diesem Kapitel existierende Profil-Abonnementdienste auf ihre Funktion und Leistungsfähigkeit untersucht werden. Wünschenswert ist eine Betrachtung von Abonnementdiensten, die innerhalb von Suchmaschinen angeboten werden. Leider bieten Suchmaschinen keine bzw. nur sehr rudimentäre Profildienste an, so daß für die Untersuchung alternative Abonnementdienste gefunden werden mußten. Die durch diese Systeme bereitgestellten Informationen sind dabei von ganz unterschiedlicher Natur, so daß ein qualitativer Vergleich der Dienste untereinander nicht sinnvoll erscheint. Es werden vielmehr Kriterien untersucht, die prinzipiell allen Abonnementdiensten zugeordnet werden können unabhängig von der bereitgestellten Information. Die aus der Evaluierung gewonnen Erkenntnisse über positive bzw. negative Eigenschaften der Systeme sollen dann in den Entwurf des SWING-Abonnementdienstes einfließen.

Der folgende Abschnitt beschreibt zunächst die Kriterien, nach denen die betrachteten Abonnementdienste bewertet werden. Anschließend werden einige Abonnementdienste vorgestellt und mit Hilfe der Kriterien genauer unter die Lupe genommen.

5.1 Kriterien für Abonnementdienste

Für die Untersuchung wurden die Dienste Ariadne, The Informant, Pointcast Business Network und Elsevier ContentDirect herangezogen. Besonderes Augenmerk wurde dabei auf die nachfolgend beschriebenen Kriterien gelegt.

- **Datenschutzmechanismen:** Hierunter wird der Schutz gespeicherter bzw. übertragener Daten vor Angriffen von außen verstanden, in deren Folge

sie manipuliert, zerstört oder unberechtigt gelesen werden. Der Aufwand für Datenschutzmechanismen hängt wesentlich von den Informationen im Nutzerprofil, den abonnierten Daten oder Semantiken ab, die durch die Kombination von Nutzerprofil und abonnierte Information entstehen. Die Datensicherheit der Abonnementdienste soll daher unter den Aspekten *notwendig* und *ausreichend* betrachtet werden.

- **Präferenzerstellung:** Hierbei handelt es sich um die initiale Erstellung des Nutzerprofils. Dies kann grob auf drei verschiedenen Wegen erfolgen.
 - **Initialisierung durch Nutzereigenschaften:** Der Nutzer wird aufgefordert, seine persönlichen Daten einzugeben, nach denen er bestimmten Nutzergruppen zugeordnet wird. Der Vorteil dieser Methode ist die Einfachheit.
 - **Initialisierung durch Beispiele:** Dies entspricht dem Ausfüllen eines Fragebogens. Das Abonnementsystem gibt dem Nutzer eine Reihe von Beispielen mit einer entsprechenden Bewertungsskala vor, die der Nutzer zu bewerten hat. Eine Analyse des Bewertungsergebnisses liefert ein bestimmtes Nutzermuster, woraus ein konkretes Nutzerprofil generiert werden kann. Vorteil dieser Methode ist die sehr einfache Handhabung für den Nutzer. Allerdings können die Beispiele so ungünstig gewählt sein, daß sie die Wünsche des Nutzers nicht widerspiegeln und dadurch ein nicht präzises Nutzerprofil liefern.
 - **Initialisierung durch Beobachtung:** Hier wird zunächst das Verhalten des Nutzer beobachtet. Ist das System der Meinung, genug Informationen gesammelt zu haben, wird aus diesem Wissen ein Nutzerprofil erstellt. Von Vorteil ist, daß der Nutzer gänzlich von Eingaben zu seinen Präferenzen befreit wird. Desweiteren kann die Beobachtung auf die gesamte Nutzungsdauer des Anwenders ausgeweitet werden, um seine Verhaltensänderungen zu registrieren. Leider kann die durchgehende Beobachtung auch als Nachteil angesehen werden, wenn es für den Nutzer nicht möglich ist, die Beobachtung seines Verhaltens zu unterdrücken. Jedes ziellose Referenzieren wird dann vom System registriert, obwohl es keinen direkten Zusammenhang mit den Präferenzen des Nutzers darstellt. Ein weiterer Nachteil dieser Form der Präferenzerstellung ist die Anlaufzeit bis das System endgültig einsatzbereit ist. Außerdem besteht die Gefahr, daß vom System ein Verhalten des Nutzer beobachtet wird, das völlig neben den Interessen des Nutzers liegt.
- **Präferenzgranularität:** Je feiner die Granularität des Benutzerprofils ist, desto präziser sind natürlich auch die Ergebnisse der Informationsfilterung. Allerdings nimmt sowohl für den Nutzer als auch für den Entwickler eines

Abonnementdienstes der Aufwand in Richtung feinere Granularität zu. Die Bewertung der zu betrachtenden Systeme soll in den Stufen fein, mittel und grob erfolgen.

- **Adaptives Verhalten:** Unter der Adaptivität wird die Fähigkeit eines Profil-Abonnementdienstes verstanden, sich in gewisser Weise dem Verhalten des Nutzers anzupassen. Dies setzt allerdings das Vorhandensein eines Zielsystems in Form von regel- bzw. wissensbasierten Systemen oder neuronalen Netzwerken voraus. Das adaptive Verhalten entspricht der Schlußfolgerungsfähigkeit von intelligenten Agenten aus dem Abschnitt 3.2.2.
- **Beachtung der Nutzersystemumgebung:** Dieses Kriterium ist erfüllt, wenn für den Anwender die Möglichkeit besteht, Übertragungszeiten bzw. Informationsausgaben auf seine systemspezifischen Vorgaben einzustellen.
- **Verfügbarkeit:** In Abschnitt 3.3 wurde eine Klassifikation der Abonnementdienste nach client- bzw. serverbasierten Systemen vorgenommen. Die Verfügbarkeit soll nun angeben, ob der Dienst jedem Nutzer unabhängig von seiner soft- und hardwaretechnischen Ausstattung zur Verfügung steht. Vorausgesetzt wird dabei allerdings ein Zugang zum WWW.
- **Ereignisnotifikation:** Die Ereignisnotifikation gibt an, in wie weit der Nutzer Einfluß auf den Zeitpunkt und die Art der Benachrichtigung nehmen kann. Handelt es sich bei dem Dienst eher um einen Alertingservice, hat der Nutzer in der Regel kaum Möglichkeiten, die Benachrichtigung zu steuern. Handelt es sich dagegen um einen Profil-Abonnementdienst, solltes es Möglichkeiten geben eine periodische Festlegung der Ereignisbenachrichtigung zu treffen.
- **Zeitliche Entkoppelung** Eine zeitliche Entkopplung liegt vor, wenn die Informationsübertragung asynchron zur Betrachtung beim Dienstanwender erfolgen kann. Diese Form wird auch Push-Technologie genannt. Wird dagegen die Auslieferung der Inhalte vom Nutzer gesteuert, spricht man von Pull-Technologie.

5.2 Ariadne

Ariadne [Ariad] ist ein Projekt der Freien Universität Berlin. Es ist ein Partnerprojekt der elektronischen Informatik-Bibliothek MeDoc¹. Innerhalb von Ariadne können Informationen aus dem Bereich der Informatik angeboten und vermittelt werden. Da das Anbieten von Informationen auf keine Nutzerklasse beschränkt

¹MeDoc - Multimedia electronic Documents ist die erste digitale online zugreifbare Volltext-Bibliothek Deutschlands

ist, aber dennoch der Datenbestand auf Informationen der Informatik begrenzt bleiben soll, bietet Ariadne die Registration als Qualitätskontrolleur an. Die Aufgabe des Qualitätskontrolleurs entspricht der eines Moderators einer News-Gruppe, wobei das Thema einer CR-Klassifikationsnummer entspricht. Ariadne ist ein serverbasiertes System, das jedem Internetnutzer zur Verfügung steht. Die Suche innerhalb des Systems kann navigierend oder über die Eingabe eines Suchstrings erfolgen. Bei der navigierenden Suche ist das Verzeichnis nach dem CR-Klassifikationsschema geordnet. Ariadne bietet Nutzern die Anlage eines Nutzerprofils an. Dieses Profil besteht aus gespeicherten Suchanfragen und archivierten URL's. Die gespeicherten Suchanfragen können vom System in einer durch den Nutzer festgelegten Periode wiederholt werden. Der festgelegte Zeitraum kann nicht individuell für jede Suchanfrage vereinbart werden, sondern gilt für das gesamte Profil. Für die Ereignisbenachrichtigung beim Nutzer gilt ähnliches wie für das Festlegen eines Zeitraums. Es gibt die Optionen, alle Meldungen die das Profil des Nutzers betreffen, per Email zuzusenden bzw. keine Nachrichten durch Ariadne zustellen zu lassen. Der zeitliche Rahmen für die Überwachung der archivierten URL's entspricht dem der Suchanfragen. Neben der direkten Eingabe einer zu überwachenden URL in das Nutzerprofil ist auch die Bekanntmachung über die Mail Document Option des WWW-Browsers möglich. Die Beobachtung ausgewählter Teile eines Dokumentes ist nicht möglich. Die Präferenzgranularität entspricht den Selektionsbedingungen einer Suchanfrage. Ariadne unterstützt den Anwender mit multiplen Eingrenzungsmöglichkeiten der Anfrage (z.B. exact match, stemm match, Suche in abstracts, keywords, title, Selektion der Objekttypen). Die Ergebnisse einer Anfrage werden dem Nutzer in einer übersichtlichen Form bereitgestellt. Neben der Referenz auf das Dokument wird der Objekttyp, eine Kurzbeschreibung des Inhaltes und der Status der URL (gültig, nicht kontrolliert) bereitgestellt. Es wird allerdings kein Ranking der Ergebnisse vorgenommen, so daß für den Nutzer nicht zu erkennen ist, ob die Ergebnismenge eine Ordnung besitzt. Datenschutzmechanismen existieren bei Ariadne nur in Form einer Authentifizierungspflicht für Nutzer, die Anfrage- und Updateoperationen auf gespeicherten Profilen vornehmen möchten. Die Datenübertragung zwischen Client und Server und die Ereignisnotifikation des Systems erfolgen dagegen unverschlüsselt.

Zusammenfassend kann Ariadne als Abonnementdienst für Informationen aus dem Gebiet der Informatik bezeichnet werden. Das Anlegen und Benutzen des persönlichen Profils ist einfach. Die Ergebnispräsentation einer Suchanfrage ist übersichtlich und homogen gestaltet. Gut gelöst wurde auch das Problem, nicht informatikbasierte Datenquellen durch externe Qualitätskontrolleure aus dem System zu entfernen. Nachteilig innerhalb eines Nutzerprofils ist, daß alle abonnierten Suchanfragen und der URL-Reminderservice dem selben Zeitintervall unterliegen. Ein weiteres Manko von Ariadne ist die beschränkte Informationsmenge, auf die innerhalb des Systems zugegriffen werden kann, da Ariadne keine Inte-

gration externer Informatik-Datenquellen vornimmt.

5.3 The Informant

The Informant [Infor] ist ein von der Computer Engineering Gruppe des Dartmouth College entwickelter Webmonitor und Notifikationsdienst. Es handelt sich um ein serverbasiertes System, das Nutzern die Anlage persönlicher Profile anbietet, in denen bis zu drei Suchanfragen und bis zu fünf zu überwachende URL's gespeichert werden können. Zur Auswertung der Suchanfragen kann der Anwender zwischen den Suchmaschinen AltaVista ([Alta]), Lycos ([Lycos]), Excite ([Excite]) und Infoseek ([Info]) wählen. Das Zeitintervall, in dem sowohl Suchanfragen als auch die URL's überwacht werden, entspricht einer globalen Präferenz. Dieser Wert kann daher nur für das gesamte Profil festgelegt werden. Der Nutzer hat die Wahl zwischen 3, 7, 14, 30 und 60 Tagen. Für jede zu überwachende Suchanfrage liefert Informant maximal 10 Dokumente als Resultat zurück. Die einzelnen Referenzen werden als updated, new bzw. unreachable gekennzeichnet. Wurde eine gültige Email-Adresse in die globalen Präferenzen eingetragen, erfolgt die Ereignisbenachrichtigung in jedem Fall auf diesem Weg. Da das Feld vom System nicht überprüft wird, können auch Internetnutzer, die über keine Email-Adresse verfügen diesen Dienst nutzen. Aufgrund mißbräuchlicher Nutzung wurde die Option, URLs in die Ergebnisbenachrichtigung mitaufzunehmen, wieder entfernt.

Informant bietet dem Nutzer eine sehr übersichtliche Darstellung der Ergebnisse einer Suchanfrage. Es ist auf einen Blick zu erkennen, ob ein neues Dokument im Ergebnis enthalten ist. Auch "tote" Links werden durch das System explizit dargestellt. Nachteilig fallen aber die vielen Beschränkungen des Systems ins Gewicht. So kann der Nutzer maximal 3 Anfragen und 5 URL's in sein Profil aufnehmen. Alle Anfragen und URL's werden im selben Zeitabschnitt wiederholt und die Ergebnismenge von 10 Dokumenten pro Anfrage ist einfach nicht ausreichend. Die Entwickler von Informant möchten aber in einer nächsten Version mehr Seiten pro Anfrage zurückliefern. Geplant ist zudem die Möglichkeit, bestimmte Ergebnisreferenzen als irrelevant kennzeichnen zu können. Diese werden bei einer erneuten Bearbeitung aus der Ergebnismenge eliminiert. Weiterhin möchten sie die Ergebnisse verschiedener Suchmaschinen kombinieren, um eine Liste relevanter Referenzen erhalten zu können.

5.4 Pointcast Business Network 2.0

Pointcast [Point] bietet mit Business Network 2.0 einen kostenfreien Internet-Nachrichtendienst an. Es handelt sich um ein client-basiertes System, das zur

Zeit für Windows95, Windows NT 4.0 und Mac OS zur Verfügung steht. Alle Informationen, die Pointcast bereitstellt, sind über Themen in Kanälen organisiert. Durch die Anlage eines persönlichen Nutzerprofils, in welchem der Anwender nicht gewünschte Kanäle bzw. Themen aus der Gesamtmenge deselektiert, werden ihm seine interessanten Nachrichten präsentiert. Für die Bereitstellung von Informationen stehen bei Pointcast u.a. CNN² und ZDNet³ unter Vertrag. Zur spezifischeren Personalisierung des Profils werden die Kanäle weiter in Bereiche unterteilt. Hier kann ebenfalls eine Selektion/Deselektion durch den Anwender vorgenommen werden. Weitere Profilinformationen innerhalb von Pointcast sind zum Beispiel das Intervall, in dem die Nachrichten aktualisiert werden sollen, ob Werbeanfragen an Pointcast an den Nutzer weitergeleitet werden sollen und die Einrichtung eines Bildschirmschoners mit den aktuellen Nachrichten. Die Präsentation der Nachrichten beim Nutzer erfolgt durch den zentralen Pointcast-Bildschirm. Dieser ist in die vier Bereiche Werbefenster, Anzeigefenster, Tool-Leiste und Registerfenster unterteilt. Die Tool-Leiste enthält Funktionen zum Auswählen der Themen, Aktualisieren der nutzerspezifischen Informationen, der Präferenzdefinition sowie zum Drucken und Organisieren der Oberfläche. Innerhalb des Registerfensters werden die Titel zu Publikationen und Dokumenten der Bereiche eines Kanals dargestellt. Durch Anwahl eines Titels wird das vollständige Dokument innerhalb des Anzeigefensters präsentiert. Aus welchem Update die gegenwärtig präsentierten Daten stammen, erkennt der Anwender aus dem Kopf des Pointcast-Hauptfensters.

Betrachtet man Pointcast Business Network unter den oben genannten Kriterien, so sind Sicherheitsmechanismen für die Authorisierung der Nutzer nicht erforderlich, da die abonnierten Informationen direkt auf den Rechner des Anwenders übertragen werden. Für den Transport der Daten zwischen dem Pointcast-Server und dem Client wird ebenfalls keine explizite Verschlüsselung vorgenommen. Da es sich um einen kostenlosen Service handelt, dessen abonnierte Information überwiegend aus dem Nachrichtenbereich stammt, würde eine Verschlüsselung auch wenig Sinn machen. Das Einstellen der Präferenzen wird im Anschluß an die Installation vorgenommen. Überdies können sie aber jederzeit geändert werden. Die Initialisierung erfolgt durch das Festlegen der Nutzereigenschaften. Die Präferenzgranularität ist recht fein, da für jeden abonnierten Kanal des Nutzers wiederum Gruppen existieren, die von ihm aktiviert bzw. deaktiviert werden können. Daß das Abonnementintervall nicht für jeden Kanal festgelegt werden kann, fällt nicht weiter ins Gewicht, da es sich im Wesentlichen um tagesaktuelle Informationen handelt, die eine tägliche Aktualisierung erfordern. Adaptive Verhalten weist Business Network nicht auf. Es werden nur Daten bereitgestellt, die einem im Profil enthaltenen Kanal zugeordnet werden können. Business Network ist ein client-basiertes System, daher hängen die Kriterien Verfügbarkeit,

²CNN - Cable News Network Interactive

³ZDNet - Web-Site der Ziff Davis Verlags GmbH

Ereignisnotifikation und Beachtung der Systemumgebung sehr eng zusammen. Bei der Verfügbarkeit schneidet Pointcast recht schlecht ab, da das System nur für die beiden Microsoft-Systeme Windows95 und Windows NT 4.0 sowie für Mac OS bereitgestellt wird. Da das Interface speziell für das darunterliegende System angepaßt ist, wird eine dementsprechend gute Beachtung der Systemumgebung vorgenommen. Pointcast verzichtet darauf, eine zusätzliche Ereignisnotifikation vorzunehmen. Die Möglichkeit zur zeitlich entkoppelten Bereitstellung der Daten durch Pointcast hängt essentiell vom Nutzer ab. Trennt er seinen Rechner stets vom Netz, kann die Bereitstellung nur während der Onlinesitzung erfolgen. Prinzipiell bietet Pointcast aber die Datenübertragen zu beliebigen Zeiten an.

5.5 Elsevier Science - ContentsDirect

Elsevier Science ist Marktführer im Publizieren und Verbreiten von wissenschaftlichen Informationen eines breiten Spektrums. Innerhalb des Systems stehen Lesern mehr als 1200 englischsprachige Journale zur Auswahl.

Mit ContentsDirect bietet Elsevier Wissenschaftlern einen frei nutzbaren Alertingdienst für Inhalte demnächst erscheinender Veröffentlichungen an. Es ist ein server-basiertes System, das uneingeschränkt zur Verfügung steht. Die Nutzerpräferenzen des ContentsDirect-Benachrichtigungsdienstes bestehen aus Journalen verschiedener wissenschaftlicher Gebiete. Zu den ausgewählten Journalen bezieht der Anwender die Inhaltsverzeichnisse der Neuveröffentlichungen. Im Gegensatz zu den drei Profildiensten der letzten Abschnitte hat der Nutzer hier keinen Einfluß auf Zeit und Form der Ereignisnotifikation. Diese wird allein durch Elsevier gesteuert. Jede innerhalb von ContentsDirect festgelegten Präferenz (Journale) entspricht einer Mailingliste, in die der Nutzer aufgenommen wird. Die Sicherheitsmechanismen bei ContentsDirect bestehen aus einer Authentifizierungskomponente, die die Nutzung eines fremden Nutzeraccounts verhindert. Eine Verschlüsselung der per Email übertragenen Informationen erfolgt nicht. Die Funktionalität von ContentsDirect ist auf die Verbreitung der Inhaltsverzeichnisse von Neuerscheinungen beschränkt, wodurch der Informationsagent des System über keine adaptiven Fähigkeiten verfügt.

5.6 Übersicht und Zusammenfassung

In Tabelle 5.1 werden die vier Abonnementdienste unter dem Aspekt der Betrachtungskriterien aus dem ersten Abschnitt dieses Kapitels zusammengefaßt.

	Ariadne	Informant	Pointcast	Elsevier	SWING
Anwendungsgebiet	IR / IF ⁴	IR / IF	Newswatcher ⁵	IR / IF	IR / IF
Dienst	Profil	Profil	Profil	Alerting	Profil
Datenschutz	auf Profilebene ausreichend	auf Profilebene ausreichend	nicht notwendig	auf Profilebene ausreichend	auf Profilebene ausreichend
Präferenz-erstellung	Nutzereigenschaften	Nutzereigenschaften	Nutzereigenschaften	Nutzereigenschaften	Nutzereigenschaften
Präferenzgranularität	mittel ⁶	grob ⁷	fein ⁸	grob ⁹	mittel-fein ¹⁰
Adaptives Verhalten	nein	nein	nein	nein	nein
Beachtung der Nutzerumgebung	nein	nein	ja ¹¹	nein	nein
Verfügbarkeit des Abonnementdienstes	uneingeschränkt über WWW	uneingeschränkt über WWW	eingeschränkt Windows95 NT 4.0 Mac OS	uneingeschränkt über WWW	uneingeschränkt über WWW
Festlegung Ereignisnotifikation	globale Präferenz im Profil	globale Präferenz im Profil	globale Präferenz im Profil	Steuerung durch Anbieter	lokale Präferenz je Anfrage/URL
Ereignisnotifikation auf / durch	Email	Email	PC des Nutzer	Email	Email / Newschannel im Abodienst
zeitliche Enkopplung	bedingt ¹² nur Referenzen	bedingt nur Referenzen	besteht	-	bedingt nur Referenzen

Tabelle 5.1: Übersicht der betrachteten Abonnementdienste

⁴Information Retrieval / Information Filtering - Anwendungen aus diesem Bereich unterstützen den Benutzer bei der gezielten Suche nach Informationen im WWW

⁵automatische Aktualisierung nutzerspezifischer Nachrichten und Informationen

⁶sehr feine Präferenzfestlegung für jede Nutzeranfrage, aber kaum zeitliche Einstellungen

⁷sehr eingeschränkte Möglichkeiten der Präferenzfestlegung

⁸Individuelle Einstellungen für jeden abonnierten Kanal

⁹Präferenzen bestehen nur aus Journalen

¹⁰diverse Einstellungsmöglichkeiten zu überwachender Anfragen/URL (Groß-/Kleinschreibung, Rechtschreibfehler, Überwachungsbereiche etc.)

¹¹client-basierter Abonnementdienst, der speziell an die Plattform angepasst wurde

¹²Eine zeitliche Enkopplung wäre möglich, wenn statt der Referenz das gesamte Dokument übertragen wird

Zusätzlich wurde in diese Tabelle der SWING-Abonnementdienst aufgenommen, um einen direkten Vergleich zu den betrachteten Systemen aufzuzeigen. Eine genauere Beschreibung des SWING-Abonnementdienstes erfolgt jedoch erst im nächsten Kapitel.

Diese Systeme stellen nur eine kleine Auswahl existierender Abonnementdienste dar. So bietet in Deutschland der Springer Verlag mit LINK [Link] einen ähnlichen Benachrichtigungsdienst an, wie Elsevier. Jedoch werden durch Elsevier wesentlich mehr Journale angeboten als durch den Springer Verlag. Eine weitere breite Palette an frei verfügbaren als auch kommerziell vertriebenen Abonnementdiensten findet sich unter den sogenannten Newswatchern. Neben dem oben beschriebenen Pointcast Business Network seien an dieser Stelle noch Backweb Infocenter [Back], Marimba Castanet [Marim] und Intermind Communicator [Inter] genannt. Es ist aber fraglich, ob diese Systeme eine große Akzeptanz beim Nutzer erlangen, da sie durch ständig wechselnde Werbeeinblendungen die Aufmerksamkeit des Anwenders von der eigentlichen Information ablenken. Zudem sind sie sehr speicher- und rechenintensiv. Ein weiterer Kritikpunkt dieser Dienste sind in der Regel fehlende adaptive Fähigkeiten. Sie stellen Informationen nur auf Grund der im Profil gespeicherten Präferenzen bereit. Rückschlüsse und darausfolgenden Empfehlungen durch eine Wissensbasis oder Ähnlichkeitsanalyse nehmen sie im Allgemeinen nicht vor. Werden solche Systeme von Nutzern als ausschließliche Nachrichten-Informationsquellen verwendet, besteht die Gefahr eines eingeschränkten Blickwinkels. Eine mögliche Lösung des Problems bietet die Firma Firefly Network Inc.[Fire] mit dem Konzept des Advanced Collaborative Filtering (ACF) an. Das Firefly-System stellt personalisierte Informationen aus dem Entertainmentbereich zur Verfügung. Die Interessen aus diesem Bereich unterliegen in einem hohem Maß subjektiven Einschätzungen. ACF liegt die Theorie zugrunde, daß die Präferenzen der Teilnehmer einer Gemeinschaft herangezogen werden, um weitere Präferenzen beziehungsweise Interessensgebiete eines einzelnen Nutzers vorhersagen zu können [Fir97].

Kapitel 6

Der SWING-Abonnementdienst

Die im letzten Kapitel vorgenommene Evaluierung verschiedener Informationsdienste hat sowohl deren funktionale Stärken als auch Schwächen für die Anwendung auf einen neuen Abonnementdienst aufgezeigt. Diese Erkenntnisse sollen nun in den Entwurf sowie die Implementierung des SWING-Abonnementdienstes einfließen.

In diesem Kapitel wird zunächst der konzeptionelle Entwurf des SWING-Abonnementdienstes vorgenommen. Dabei ist zu analysieren, inwieweit die Vorteile der verschiedenen Dienste auf den SWING-Abonnementdienst übertragbar sind. Daran anschließend erfolgt eine Beschreibung der implementierten Funktionalität.

6.1 Konzeption

6.1.1 Zielstellung

In der Einleitung der Studienarbeit wurde als primäres Ziel der Entwurf und die prototypische Implementierung eines Profildienstes für die SWING-Suchmaschine genannt. Dabei ist von Untersuchungen existierender Lösungen auszugehen. Die Ziele, die beim Entwurf verfolgt werden sollen, sind:

- den Nutzer über Veränderungen im System zu informieren
- zyklisch wiederkehrende Anfragen zu bearbeiten
- turnusmäßige Informationen zu einem speziellen Thema bereitzustellen

6.1.2 Analyse

Wie bereits erwähnt, diente die Evaluierung der Abonnementdienste dem Erkennen der Vor- bzw. Nachteile der einzelnen Systeme und dem Herausfiltern von Funktionalitäten für den Abonnementdienst. Insbesondere wurden dabei folgende positive und negative Eigenschaften erkannt.

- Positive Eigenschaften
 1. sehr gute Strukturierung der Dokumentreferenzen (The Informant)
 2. Filtermöglichkeit auf Objekttypen (Ariadne)
 3. neben Anfrage- auch URL-Überwachung (Ariadne, The Informant)
 4. Einfügen einer URL zum Profil an beliebiger Stelle über die Email-Funktion des WWW-Browsers (Ariadne)
 5. nach Abonnementzustellung Nutzung der Informationen ohne Netzverbindung (Pointcast)
 6. Advanced Collaborative Filtering (ACF) - über den Vergleich eines Nutzerprofils mit Profilen anderer Nutzer können Vorhersagen und Empfehlungen hinsichtlich der Interessensschwerpunkte getroffen werden (Firefly)
 7. Informationen über Neuigkeiten des Systems (Firefly)
- Negative Eigenschaften
 1. kein individuelles Abonnement einzelner Anfragen/URL (Ariadne, The Informant)
 2. keine Filtermöglichkeiten für die Beobachtung von Webinhalten (Ariadne, The Informant)
 3. keine individuelle Systemanpassung (Ariadne, The Informant, Elsevier)
 4. keine Relevancebewertung der Ergebnismenge (Ariadne, The Informant)
 5. geringe Ergebnismenge (The Informant)
 6. keine persönliche Informationsvermittlung, sondern Zuordnung zu Mailinglisten (Elsevier)

Das Ziel ist es nun, die Vorteile der einzelnen Systeme in den Entwurf des SWING-Abonnementdienstes zu übernehmen. Die als nachteilig bewerteten Funktionen dagegen zu vermeiden oder aber durch Verbesserungen in Vorteile umzuwandeln.

Zunächst werden die Stärken der verschiedenen Systeme auf eine Umsetzung in das SWING-System analysiert. Es ist wahrscheinlich, daß einige gute Eigenschaften nicht auf SWING übertragbar sind, da diese an ein spezielles Subsystem angepaßt sind.

- zu 1. Die Strukturierung der Dokumentenreferenzen aufgrund der Filterinformationen ist ohne größere Probleme auch in das SWING-System integrierbar. Die Dokumentenreferenzen einer gespeicherten Anfrage können bei der Wiederholung dieser Anfrage den Zuständen neu bzw. verändert zugeordnet werden. Nicht mehr vorhandene Referenzen werden aus der Ergebnismenge der Datenbank entfernt.
- zu 2. Die im Ariadne-System gegebene Möglichkeit, nur bestimmte Objekttypen in die Anfrage aufzunehmen ist auf den zu entwickelnden Service nicht übertragbar. Der SWING-Abonnementdienst nutzt eine Schnittstelle zur SWING-Suchmaschine und ist daher bei der Anfragebearbeitung auf die Gegebenheiten von SWING angewiesen. In SWING wird nur zwischen den Objekttypen Hypertextdokument und Postscript u.a. Dokumente unterschieden.
- zu 3. Die Umsetzung eines URL-Reminders und die Bereitstellung von Informationen über Neuigkeiten des Systems sind für den zu entwickelnden Dienst zu berücksichtigen. Die Neuigkeiten werden über einen integrierten Newschannel publiziert, der jedoch zunächst nur als unidirektionaler Kanal geplant ist.
- zu 4. Die Aufnahme einer URL in den Reminder via Email ist nicht vorgesehen. Prinzipiell ist diese Funktionalität nicht ausgeschlossen, da aber der SWING URL-Reminder verschiedene zusätzliche Angaben verlangt, besteht die Gefahr einer hohen Fehlerquote.
- zu 5. Die Nutzung der abonnierten Information ohne zusätzliche Netzverbindung ist auf jeden Fall wünschenswert. Leider ist dies in SWING mit einigen Schwierigkeiten verbunden. Das Subsystem des Abonnementdienstes ist eine Suchmaschine, deren Anfrageergebnis Dokumentenreferenzen sind. Eine Informationsnutzung ohne Internetverbindung würde eine Übertragung aller Dateien (HTML, Images, Icons, usw.), die innerhalb des Dokumentes verwendet werden, zur Folge haben. Je nach Größe der Dateien und Anzahl der Ergebnisse kann der Speicherbedarf beim Nutzer schnell anwachsen. Eine reduzierte Variante wäre, nur das Dokument selbst an den Nutzer zu senden. Dies hat jedoch den Nachteil, daß unter Umständen sehr viel Information verloren geht. Dies ist zum Beispiel der Fall, wenn innerhalb des Dokumentes viele externe Dateien verwendet werden. Für die periodische

Bearbeitung von Nutzeranfragen im SWING-Abonnementdienst werden daher vorläufig nur die Referenzen der Ergebnisse und eine Kurzbeschreibung bereitgestellt.

- zu 6. Das ACF des Firefly-Systems ist auch in SWING denkbar. Die innerhalb des Themenkatalogs von den Anwendern spezifizierten Interessen unterliegen den gleichen subjektiven Einschätzungen wie die Auswahl von Musik und Filmen in Firefly. Man könnte nun die Präferenzen der Nutzer untereinander vergleichen und bei Ähnlichkeiten Empfehlungen für bestimmte Interessensgebiete aussprechen. An dieser Stelle könnte man statt Empfehlungen für Interessensgebiete auch direkt Nutzer benennen, so daß Nutzer mit Nutzern über den Newschannel in Verbindung treten und ihre Erfahrungen austauschen.

Von den positiven Merkmalen konnten zwei nicht auf den SWING-Abonnementdienst übertragen werden. Als nächstes werden die Nachteile analysiert. Sie stellen in der Regel nicht fehlende Funktionen dar, sondern eingeschränkte Fähigkeiten vorhandener. Es ist nun zu überlegen, wie diese Einschränkungen für das SWING-System aufgehoben oder verringert werden können.

- zu 1. Im SWING-Abonnementdienst ist eine individuelle Anfrage- und URL-Überwachung konzipiert. Im ersten Prototypen können zunächst jedoch nur drei verschiedene Intervalle festgelegt werden. Eine Erweiterung auf beliebige Zeiträume ist aber jederzeit möglich.
- zu 2. Es wurde bereits erwähnt, daß der Service unter anderem einen URL-Reminder beinhalten wird. Wird, wie in den Systemen Ariadne und Informant, nur die gesamte Seite auf Änderungen beobachtet, führen auch irrelevante Abweichungen zur Benachrichtigung beim Nutzer. Dies ist zum Beispiel der Fall, wenn sich auf der Seite ein Zähler regelmäßig verändert. Um diese Teile aus der Beobachtung auszuschließen, können im SWING-Abonnementdienst HTML-Strukturelemente angegeben werden (Tabellen, Listen ect.). Benachrichtigungen beim Nutzer erfolgen dann nur aufgrund von Änderungen innerhalb dieser Bereiche.
- zu 3. Die individuelle Systemanpassung des Abonnementdienstes an die heterogene Umgebung der Nutzer ist eine sehr komplexe Aufgabe. Die Differenzierungen reichen von unterschiedlichen Systemvoraussetzungen (Unix, Dos, Windows) , über abweichende Notifikationswünsche (Email, Fax, Voicemail ect.) und Form und Formate der Benachrichtigung (Postscript, Worddokument, Image usw.). Objektiv betrachtet wird in SWING daher keine Systemanpassung vorgenommen. Der Nutzer kann nur als Ausgabe-medium zwischen der Email und dem internen Newschannel wählen.

- zu 4. Bei der Relevanzbewertung einer Anfrage profitiert der Service von der SWING-Suchmaschine. Die Ergebnisdokumente einer Anfrage werden bereits durch den SWING-Agenten nach einem bestimmten Algorithmus bewertet. Aufgrund der Nutzung einer Schnittstelle zum SWING-Agenten stehen dem Informationsagenten die Rankingergebnisse der Anfrage ebenfalls zur Verfügung.
- zu 5. Im SWING-Abonnementdienst wird eine maximale Ergebnismenge von 50 Dokumenten zugelassen. Die persönliche Nutzung von Suchmaschinen hat gezeigt, daß in der Regel nicht mehr als 30 oder 40 Dokumente aus einer Anfrage besichtigt werden. Wurde danach kein zufriedenstellendes Resultat gefunden, wird in der Regel die Anfrage verfeinert. Sollte sich in der Praxis herausstellen, daß 50 Dokumente nicht ausreichen, kann eine Erweiterung bis auf 1000 Dokumente problemlos vorgenommen werden. Der Grund für die erste Einschränkung ist die Speicherung der Resultate in der Datenbank. Mehr Resultate bedeuten leider nicht mehr qualitativ hochwertige Dokumente. Im Endeffekt stehen möglicherweise 1000 URL's in der Datenbank, von denen sich der Nutzer für maximal 50 interessiert. Jede Updateoperation auf dieser Anfrage besitzt aber eine höhere Zeitkomplexität.
- zu 6. Mailinglisten sind immer ein Zeichen, daß durch den Anwender nur sehr eingeschränkt Präferenzen festgelegt werden können. Die persönlichen Anfragen und URL's der Nutzer im SWING-Abonnementdienst eignen sich wenig für Mailinglisten. Hier wird eine individuelle Benachrichtigung beim Nutzer vorgenommen. Es wird auch darauf geachtet, daß einem Anwender nicht mehrere Mails an einem Tag zugestellt werden, sondern eine Zusammenfassung dieser Benachrichtigungen erfolgt.

6.1.3 Funktionalität

Im letzten Abschnitt wurden bereits einige Funktionen genannt, die in den SWING-Abonnementdienst einfließen sollen. In diesem Abschnitt werden die Konzepte konkretisiert, wobei das Hauptaugenmerk bei der Entwicklung des Konzeptes für einen Abonnementdienst neben der Einfachheit auch auf der Erweiterbarkeit des Systems liegt.

Die Unterabschnitte repräsentieren die Schwerpunkte des Abonnementdienstes. Diese Grobkonzepte stellen relativ unabhängige Teile dar. Das Hinzufügen weiterer Module bietet somit genug Flexibilität für Erweiterungen.

Die Elementarfunktionen wie Authentifizieren oder Neuanmeldung eines Nutzers werden hier nicht extra aufgeführt. Die nachfolgende Strukturierung soll die Funktionen möglichst fein granularisieren. In der ersten Hierarchiestufe werden

die Wünsche der Nutzer dargestellt. Die tiefer in der Hierarchie liegenden Punkte bezeichnen die nötigen Einzelschritte.

6.1.3.1 Anfragebearbeitung

Die Anfragebearbeitung ermöglicht es dem Nutzer Anfragen an die Suchmaschine in einem von ihm festgelegten Zeitintervall zu abonnieren. Der Unterschied zu herkömmlichen Suchmaschinen liegt in der autonomen Initiierung der Anfrageausführung durch den Abonnement-Agenten. Neben der Abonnementfunktion werden Funktionen zum Löschen, Editieren und manuellem Wiederholen von Anfragen angeboten.

1. neue Anfrage abonnieren
 - 1.1. Anfragedialog auslösen
 - 1.2. Anfragetext eingeben
 - 1.3. Anfrage auslösen
 - 1.3.1. Dokumentenreferenzen als Email
 - 1.3.2. Dokumentenreferenzen auf WWW-Browser anzeigen
 - 1.3.2.1. Anfrage speichern
 - 1.3.2.2. Anfrageparameter (Intervall, Medium) einstellen
2. Anfrage löschen
 - 2.1. Anfrage(n) anzeigen lassen
 - 2.2. Anfrage selektieren durch Auswahl einer Anfrage aus der Menge der gespeicherten Anfragen des Nutzers
 - 2.3. Anfrage löschen
3. Anfrage ändern
 - 3.1. Anfrage(n) anzeigen lassen
 - 3.2. Anfrage selektieren
 - 3.3. Informationen zur Anfrage anzeigen (Anfrageterm, Ausgabemedium, Abonnementintervall)
 - 3.4. Änderung innerhalb der zulässigen Parameter durchführen
 - 3.5. Änderungen bestätigen
4. Anfrage wiederholen
 - 4.1. Anfrage(n) anzeigen lassen

- 4.2. Anfrage selektieren
- 4.3. Anfrage der Schnittstelle zum Agenten der Suchmaschine übergeben
- 4.4. Filtern der Ergebnismenge
- 4.5. Dokumentenreferenzen gruppiert anzeigen (neue, geänderte, unveränderte und nicht erreichbare¹ Dokumente)

Zu erkennen ist, daß der Nutzer, abgesehen vom Erstellen einer neuen Anfrage, immer zuerst eine Anfrage selektieren muß, um sie dann weiter zu bearbeiten. Dies bedeutet, daß die Auswahl einer Anfrage sehr einfach und schnell möglich sein muß.

In der obigen Auflistung wurde das Speichern einer Anfrage nicht vom Erstellen getrennt, obwohl es für den Nutzer nicht zwingend erforderlich ist, einen Anfrage-text auch dauerhaft in sein Profil aufzunehmen. Prinzipiell ist der Anfragedialog transparent zum Anfragedialog der Suchmaschine und kann wie dieser verwendet werden. Die vier Hauptpunkte stellen die Wünsche des Nutzers dar und sind für sich als Einheit (Transaktion) zu betrachten. Im Ergebnis einer Transaktion wird das Profil des Nutzers in einen neuen Zustand überführt. Eine Ausnahme bildet hier das Versenden einer Email nach einer Ad hoc Anfrage (Punkt 1.3.1.). In diesem Fall wird kein neuer Zustand erreicht.

Die Funktionen arbeiten auf Anfragen. Anfragen wiederum repräsentieren eine Menge von Dokumentenreferenzen. Außer der Selektion einer Anfrage gibt es keine weitere Operation auf Referenzen. An dieser Stelle ist zu überlegen, ob es sinnvoll ist, auch direkt mit einzelnen Dokumentenreferenzen zu operieren, indem man eine Querverbindung zum URL-Reminder schafft.

6.1.3.2 URL-Reminder

Innerhalb des URL-Reminders können durch den Anwender Internetadressen spezifiziert werden, die durch den Abonnementdienst auf Änderungen beobachtet werden. Dabei kann das gesamte Dokument betrachtet werden oder aber nur ausgewählte Bereiche.

1. URL-Überwachung abonnieren
 - 1.1. Reminder-Dialogfeld selektieren
 - 1.2. neue URL eingeben
 - 1.3. URL Status ermitteln

¹Diese Dokumente können nicht überprüft werden, da entweder das Dokument nicht bereitsteht oder der Server zur Zeit nicht erreichbar ist

- 1.4. zu überwachende Strukturelemente auswählen (bspw. Listen, Tabellen, Body)
- 1.5. Intervallparameter spezifizieren (Überwachungszeitraum)
2. URL löschen
 - 2.1. URL(s) anzeigen lassen
 - 2.2. URL selektieren (durch Auswahl einer URL aus der Menge der gespeicherten URL's des Anwenders)
 - 2.3. URL löschen
3. URL ändern
 - 3.1. URL(s) anzeigen lassen
 - 3.2. URL selektieren
 - 3.3. Änderungen innerhalb der zulässigen Parameter durchführen

Äquivalent zur Anfragebearbeitung muß der Nutzer immer zuerst eine URL selektieren, um sie dann weiter zu bearbeiten. Somit muß diese Funktion möglichst schnell und einfach gestaltet sein. Die Punkte abonnieren, löschen und ändern einer URL sind jeweils als Einheit zu betrachten. Im Falle eines erfolgreichen Abschlusses einer Einheit, wird das Profil des Nutzers in einen neuen, dauerhaften Zustand überführt.

Für die regelmäßige Beobachtung einer abonnierten URL ist zu überlegen, ob bei der Statusgruppe 4xx diese URL aus der Datenbank und damit aus dem Profil aller betroffenen Nutzer entfernt werden soll.

6.1.3.3 Newschannel

Nachrichten an einen Nutzer können sowohl durch die Festlegung der Ausgabe auf den Newschannel innerhalb der Funktionsgruppen Anfragebearbeitung und Reminderservice als auch durch SWING-Systemnachrichten resultieren.

1. News lesen
 - 1.1. Newsdialogfeld selektieren
 - 1.2. News anzeigen lassen (nicht betrachtete sind als *new* gekennzeichnet)
 - 1.3. Artikel selektieren
 - 1.4. new Label deaktivieren
2. News löschen

- 2.1. News anzeigen lassen
- 2.2. Artikel selektieren
- 2.3. selektierten Artikel löschen

News können zunächst nur in eine Richtung vom SWING-System an Abonnementnutzer übertragen werden. Theoretisch wäre auch das Senden von Nachrichten zwischen Nutzern möglich. Da aber die Erkennung von Nutzern in SWING über die Email-Adresse erfolgt, ist hier zunächst noch der Datenschutz zu klären.

6.1.3.4 Themenkatalog

Der Themenkatalog ist eine Art Browsing-System, in dem Dokumentenreferenzen nach Themen sortiert sind. Der Nutzer kann durch Auswahl von Themen Neuigkeiten direkt auf seinem Profil sehen. Aufgrund der gleichen Begriffswelt des Themenkatalogs sind Interessensanalysen zwischen Profilen unterschiedlicher Nutzer möglich.

1. Themeneinstellungen bearbeiten
 - 1.1. Themendialogfeld selektieren
 - 1.2. Anzeige aller Themen als Browsingkatalog
 - 1.3. Selektion eines Themas für die Aufnahme in den persönlichen Katalog
 - 1.4. Änderung auf dem Profil durchführen
2. Information zu Themen anzeigen
 - 2.1. Themen anzeigen
 - 2.2. Thema selektieren
 - 2.3. Dokumentenreferenzen anzeigen (neu gekennzeichnet)

6.1.4 Datenbankentwurf des SWING-Abonnementdienstes

Da zu allen Funktionen des Abonnementdienstes Daten persistent gespeichert werden müssen, bietet sich die Verwendung eines Datenbank-Managementsystems (DBMS) an. Die dauerhafte Speicherung von Daten innerhalb eines DBMS bietet viele Vorteile gegenüber der Verwendung von separaten Dateien als Persistentespeicher. Die schwerwiegendsten Probleme, die ohne Nutzung eines einheitlichen DBMS auftreten würden, sind folgende (siehe auch [KE96]):

- Redundanz und Inkonsistenz

- Beschränkte Zugriffsmöglichkeit
- Verlust von Daten
- Integritätsverletzung
- Sicherheitsprobleme
- Hohe Entwicklungskosten

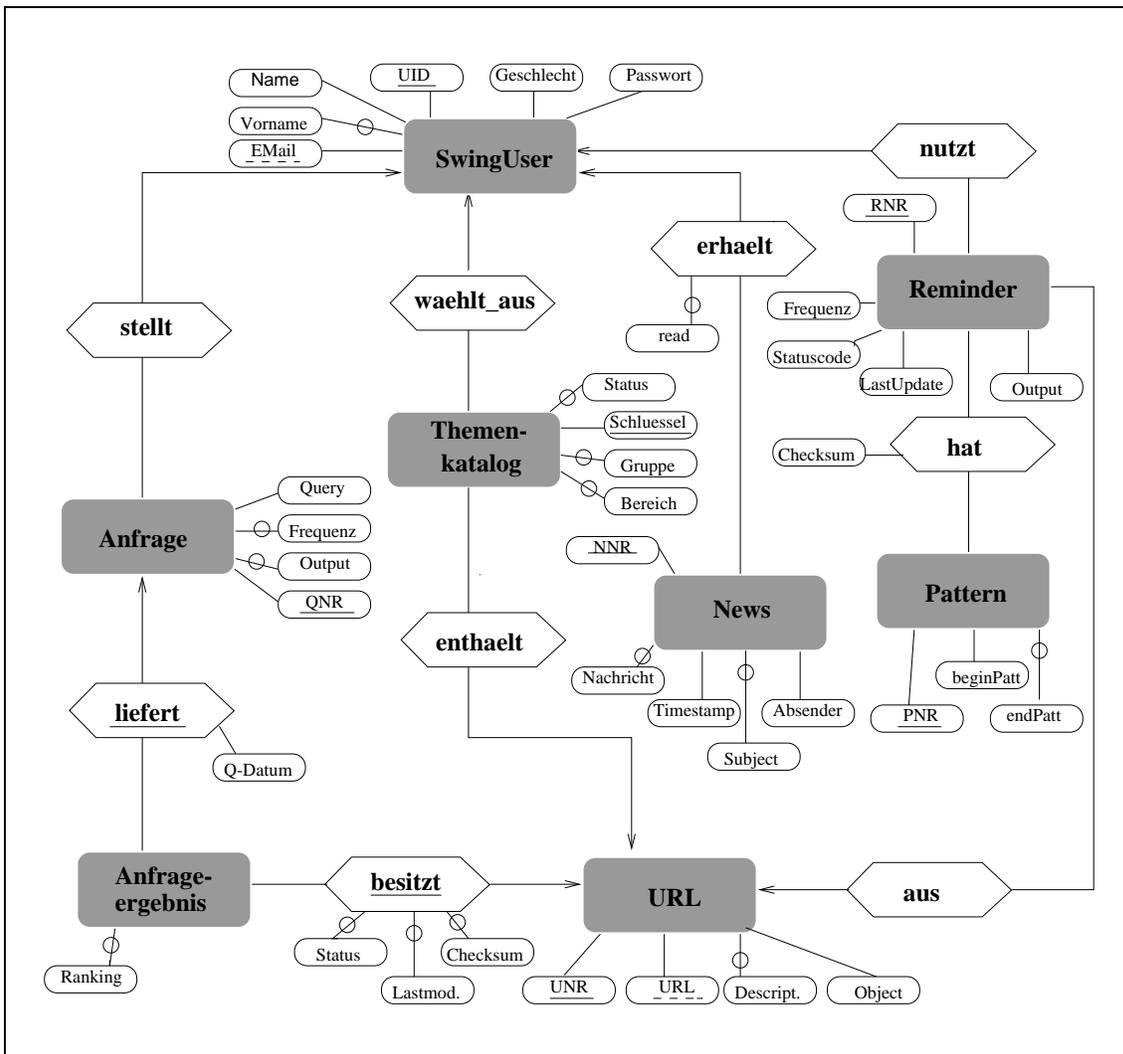


Abbildung 6.1: ER-Schema des SWING-Abonnementdienstes

Der konzeptionelle Datenbankentwurf nimmt eine Abgrenzung eines Teils der realen Welt vor, um den Ausschnitt zu bestimmen, der in der Datenbank modelliert werden soll. In der Studienarbeit ist der zu modellierende Ausschnitt der

SWING-Abonnementdienst. Das Ergebnis des konzeptionellen Entwurfs ist ein zielsystem-unabhängiges Datenbankschema.

Eines der populärsten Modelle für den konzeptionellen Entwurf ist das Entity-Relationship-Modell (ER-Modell). Wie der Name bereits ausdrückt, sind die grundlegendsten Modellierungskonstrukte dieses Modells Entities und Relationships zwischen den Entities. Weitere Konstrukte sind Attribute und Rollen. Eine tiefergehende Beschreibung des Modells findet sich u.a. in [KE96].

Abbildung 6.1 zeigt die Umsetzung der Grobkonzepte aus dem letzten Abschnitt in das ER-Modell. Rechtecke in dieser Darstellung repräsentieren Entities, Beziehungen zwischen den Entities sind durch Rauten dargestellt. Die Attribute der Entities sind durch Ovale charakterisiert. Alle Beziehungstypen des SWING-ER-Modells sind binär, das heißt, daß genau zwei Entitytypen an einer Beziehung teilnehmen. Die Beziehungen des SWING-Modells werden durch die zwei Funktionalitäten 1:n-Beziehung und n:m-Beziehung charakterisiert. Beispielsweise besteht zwischen den Entities *SwingUser* und *Anfrage* eine 1:n-Beziehung, da jedem Entity aus *SwingUser* beliebig viele Entities aus *Anfrage* zugeordnet sein können. Umgekehrt kann aber jedes Entity aus *Anfrage* mit maximal einem Entity aus *SwingUser* in Beziehung stehen. Eine n:m-Beziehung gehen die Entities *Reminder* und *Pattern* ein. Jedes Entity aus *Reminder* kann mit beliebig vielen Entities aus *Pattern* in Beziehung stehen, genauso wie jedem Entity aus *Pattern* beliebig viele Entities aus *Reminder* zugeordnet sein können.

Bei der genaueren Betrachtung des Abonnementdienstes fallen die zentralen Entities *SwingUser* und *URL* ins Auge. Anfragen (*Anfrage*) und URL-Beobachtungen (*Reminder*) werden durch Anforderungen von SwingUsern ausgelöst. Die abonnierten Informationen innerhalb von SWING (Anfragen, Themen, Reminder) sind Referenzen auf WWW-Inhalte (URL's). Das Entity *Pattern* beinhaltet die HTML-Strukturelemente, die durch den Reminderservice unterstützt werden. Jedes Entity aus *Reminder* kann mehrere Entities aus *Pattern* enthalten. Somit stellt *Pattern* ein mengenwertiges Attribut für *Reminder* dar. Für fast alle Entities wurden künstliche Schlüsselwerte eingeführt, obwohl beispielsweise in *SwingUser* und *URL* bereits einfache identifizierende Schlüsselattribute existieren. Der Grund ist der leichtere Umgang bei Anfragen und Updates. Zudem benötigen die künstlichen Schlüsselattribute für die Indizierung weniger Speicherplatz als beispielsweise die Email-Adresse oder eine URL.

Zum Abschluß soll noch verdeutlicht werden, welche Entities des SWING-Schemas mit welchen Grobkonzepten des letzten Abschnittes assoziieren.

- Unterabschnitt 6.1.3.1: *SwingUser*, *Anfrage*, *Anfrageergebnis*, *URL*
- Unterabschnitt 6.1.3.2: *SwingUser*, *Reminder*, *Pattern*, *URL*
- Unterabschnitt 6.1.3.3: *SwingUser*, *News*

- Unterabschnitt 6.1.3.4: *SwingUser, Themenkatalog, URL*

6.1.5 Architektur

Nach der Beschreibung aller Konzepte des Abonnementdienstes kann nun auch die Einordnung des Abonnementdienstes in eine Architektur vorgenommen werden (Abbildung 6.2). Deutlich zu erkennen ist die Schichtenarchitektur, wie sie in Abschnitt 4.1. beschrieben wurde.

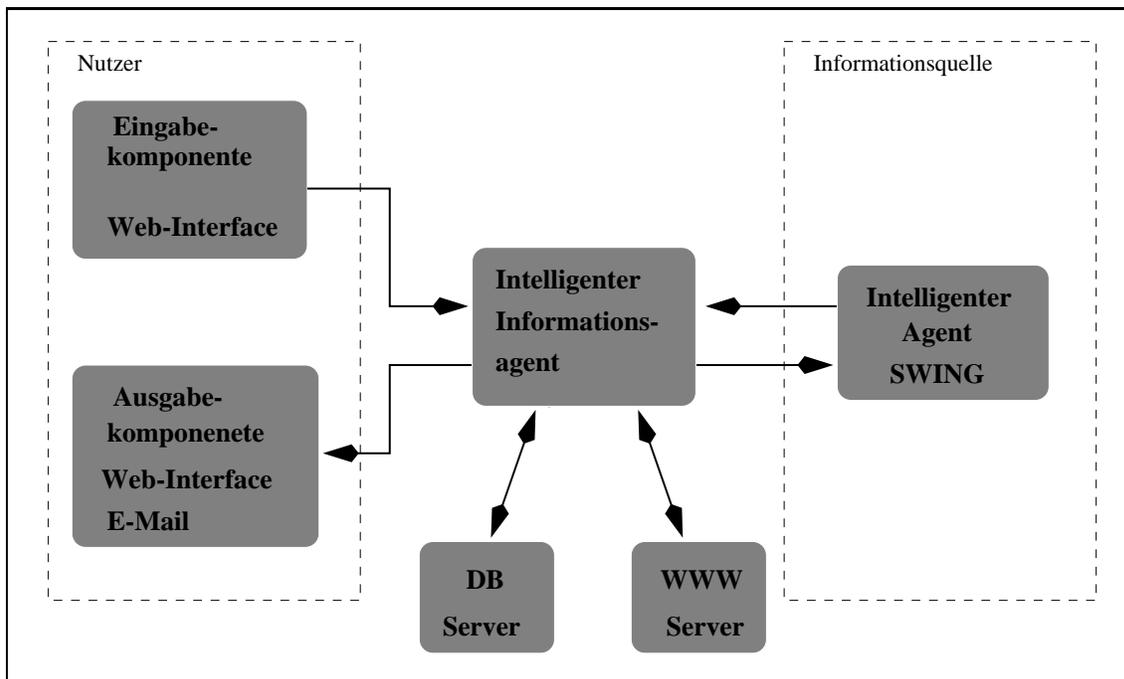


Abbildung 6.2: Architektur des SWING-Abonnementdienstes

Mit Hilfe der Eingabekomponente stellt der Nutzer des SWING-Abonnementdienstes Kontakt zum Informationsagenten her. Wichtigstes Werkzeug ist für ihn in diesem Zusammenhang der WWW-Browser. Die Ausgaben der personalisierten Informationen des Agenten können dem Nutzer sowohl in Form von Emails als auch über das WWW bereitgestellt werden.

Der Informationsagent verwaltet alle Daten der Nutzer, wie Name, Paßwort usw.. Im besonderen verwaltet er auch die vom Nutzer gestellten Anfragen, die zurückgelieferten Dokumentenreferenzen und die zu überwachenden URL's. Zusätzlich übernimmt er die Visualisierung der Daten. Dies geschieht in diesem Prototypen mittels HTML-Formularen.

Zur Informationsgewinnung bedient sich der Informationsagent des SWING-Agenten. Über eine Schnittstelle fordert der Informationsagent Informationen zu einer

Nutzeranfrage beim SWING-Agenten an. Die Antwort, in Form eines HTML-Formulars, beinhaltet neben den Dokumentenreferenzen auch Relevanzbewertungen und Kurzbeschreibungen der Referenzen.

6.1.6 Ausblick auf zusätzliche Funktionalität

Dieser Abschnitt soll mit einem Ausblick auf zusätzliche Funktionalität abgeschlossen werden. Dadurch werden sich dem Nutzer zwar keine grundlegend neuen Anwendungsgebiete erschließen, aber eine Unterstützung der vorhandenen.

Eine Funktion, die unterstützt werden könnte, ist die schrittweise Verfeinerung einer Anfrage. Hierbei werden die Dokumentenreferenzen eines Anfrageergebnisses nach einem neuen Anfragestring überprüft. Für eine periodische Anfragebearbeitung muß zur Rekonstruktion die gesamte Liste der Anfrageterme in der Datenbank vermerkt werden. Eine weitere Funktion ist die Kennzeichnung irrelevanter Referenzen. Diese Dokumente werden aus der Beobachtung der Anfrage ausgeschlossen. Denkbar für einen Abonnementdienst ist auch der Einsatz eines Terminkalenders. In diesen Kalender kann der Nutzer Referenzen eintragen. Zu dem angegebenen Termin wird der Nutzer auf die eingetragene Referenz aufmerksam gemacht.

6.2 Implementierung

Wie in Abschnitt 6.1.4 bereits angedeutet, wird für den SWING-Abonnementdienst zur persistenten Datenhaltung ein DBMS verwendet, um mögliche Probleme wie inkonsistente, redundante Daten oder den Verlust von Daten zu vermeiden.

Für diesen Prototypen wurde ein relationales Datenbanksystem (DBS) gewählt. Zum einen sind die zu speichernden Daten alle von einfachen Datentypen. Desweiteren werden im Verhältnis für wenige Objekttypen viele Objekte gespeichert. Komplexe Datentypen oder Texte (CLOB - Character Large Object) und Images (BLOB - Binary Large Object) werden für diesen Prototypen zunächst nicht benötigt. Daher ist es auch nicht erforderlich ein Objekt-Relationales DBS oder ein Object-Orientiertes DBS zu verwenden.

In diesem Abschnitt wird die Umsetzung des ER-Schemas in das relationale Modell sowie die prototypische Implementierung des Abonnementdienstes beschrieben. Aufgrund des Umfangs ist bisher nur ein Teil der in der Konzeption gestalteten und angedachten Funktionalität implementiert worden.

6.2.1 Umsetzung des ER-Schemas in das relationale Schema

Der Prototyp verwendet zur Speicherung der Nutzerdaten das relationale Datenbanksystem INGRES in der Version 6.4. Die Wahl von INGRES war u.a. durch die verwendete Programmiersprache und der hierfür vorhandenen Schnittstelle begründet. Leider hat diese Version gegenüber dem Nachfolgersystem OpenIngres noch einige Nachteile. So wird die Sicherung der referentiellen Integrität nicht implizit durch das Datenbanksystem vorgenommen, sondern muß explizit durch den Datenbankdesigner durch Regeln definiert werden. Eine Portierung der Datenbank nach OpenIngres ist aber ohne größeren Aufwand möglich.

Dem relationalen Datenbanksystem INGRES liegt das relationale Datenmodell zugrunde. Aufgrund der unterschiedlichen Strukturierungskonzepte von ER-Modell (Entitytypen und Beziehungstypen) und relationalem Modell (Relationen) ist es jedoch nicht möglich, eine einfache 1:1 Umsetzung des konzeptionellen Modells in das relationale vorzunehmen. Um eine einheitliche Basis zu haben, werden in einem initialen Entwurf für jeden Entitytyp und jeden Beziehungstyp eigene Relationen definiert. Diese Relationen werden dann in weiteren Schritten verfeinert. Dabei können Relationen eliminiert oder zusammengefaßt werden. Da die Beschreibung der Umsetzung den Rahmen dieser Arbeit sprengen würde, wird an dieser Stelle auf [KE96] verwiesen. Dort wird eine beispielhafte Umsetzung vorgenommen. Das Ergebnis der Umsetzung in das relationale Modell ist eine Strukturbeschreibung (Datenbankschema). Für den SWING-Abonnementdienst findet sich die Strukturbeschreibung in Anhang A wieder.

6.2.2 Konfliktauflösung

Konflikte in einer Datenbank können u.a. durch Update-Operationen auf Schlüsseln bzw. Fremdschlüsseln entstehen. Für einen konsistenten Datenbankzustand muß sichergestellt werden, daß für jeden Wert eines Fremdschlüssels ein identischer Wert des zugehörigen Primärschlüssels existiert. Dies wird auch referentielle Integrität genannt.

Der in Ingres 6.4. verwendete SQL89-Standard² [Ing91] sieht keine Schlüssel-Fremdschlüssel-Angaben beim Anlegen einer Relation vor. Sämtliche Konflikte die durch Operationen auf Schlüsseln/Fremdschlüsseln auftreten, müssen durch Regeln überwacht werden. Die Aktivierung von Regeln erfolgt durch ECA-Mechanismen (Trigger).

- **Event** kennzeichnet ein bestimmtes punktuellere Ereignis (z.B. Ausführung einer SQL-Anweisung).

²SQL - Structured Query Language basiert auf dem relationalen Datenmodell

- **Condition** ist ein Prädikat über der Datenbank. Die Bedingung legt fest, unter welchen Voraussetzungen man sich für ein Ereignis interessiert.
- **Action** spezifiziert, wie auf das Eintreten einer Situation reagiert werden soll (z.B. durch eine Datenbank-Prozedur).

Für die Auflösung eines Konfliktes im SWING-Abonnementdienst werden die Verfahren:

- **Kaskadierung**: Dieses Verfahren wird bei Operationen auf Primärschlüsseln angewandt. Die Kaskadierung nimmt diese Operation dann auch auf den zugehörigen Fremdschlüsseln vor. Als Beispiel sei das Löschen eines Primärschlüsselwertes genannt. Beim kaskadierenden Löschen werden auch alle betroffenen Fremdschlüsselwerte eliminiert.
- **Restriktion**: Die Operation auf dem Primärschlüssel wird nur ausgeführt, wenn es keine entsprechenden Fremdschlüsselwerte in referenzierten Relationen gibt.
- **Verboten**: Verbietet eine Operation gänzlich.

angewendet. Welches Verfahren dabei auf welchen Operationen und Attributen einer Relation angewendet wird, zeigt Anhang B.

6.2.3 Beschreibung der Implementierung

Da bereits Teile der SWING-Suchmaschine mit Hilfe der Programmiersprache Python [Python] umgesetzt wurden, wurde auch die Implementierung des Informationsagenten in dieser Programmiersprache vorgenommen. Python ist eine interpretierende objektorientierte Programmiersprache. Der Interpreter ist für die gängigsten Systeme frei verfügbar. Vom Sprachumfang ist Python mit der sehr verbreiteten Programmiersprache Perl zu vergleichen. Wie in Perl existiert auch für Python eine umfangreiche Bibliothek zusätzlicher Module, die die Programmierung vereinfachen. Ein großer Vorteil von Python gegenüber einer Programmiersprache wie C, ist das Vorhandensein von höheren Datentypen wie Liste, Tupel und Dictionary. Gerade im Zusammenhang mit Datenbanken werden diese Datentypen häufig benötigt. Die Visualisierung der Anfragen und die Kommunikation mit den Nutzern erfolgt durch HTML-Strukturelemente auf dem WWW-Client. In Abschnitt 3.1.5. wurden die verschiedenen Möglichkeiten der Datenbankbindung an das WWW beschrieben. Im Abonnementdienst kommt die Variante eins zum tragen. Das heißt, für jede Anfrage an die Datenbank wurde ein separates CGI-Programm erstellt.

Die Konzeption des Abonnementdienstes hat doch eine recht umfangreiche Funktionalität zu Tage gefördert. Daher wurde in dieser Studienarbeit zunächst nur das Grobkonzept der Anfragebearbeitung, die Neuanmeldung und Authentifizierung von Nutzern sowie Änderungen auf dem initialen Profil der Nutzer implementiert.

Der Abonnementdienst ist zwar ein System mit einer gewissen Eigendynamik, in der Form, daß der Informationsagent die periodische Anfragebearbeitung autonom vornimmt. Jedoch finden viele Funktionen des Informationsagenten im Dialog mit dem Nutzer statt. Der in Abbildung 6.3 dargestellte Graph beschreibt den Nutzerdialog.

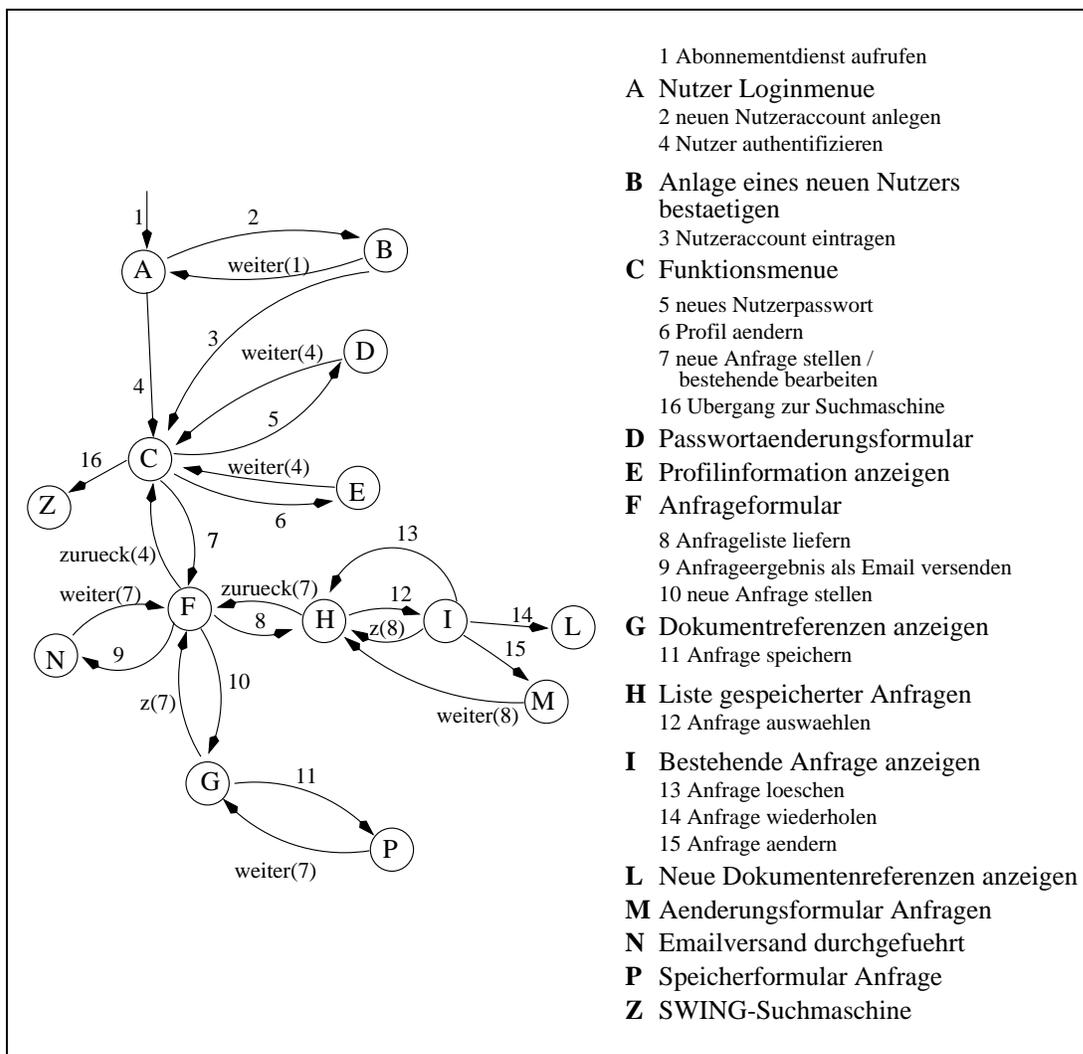


Abbildung 6.3: Nutzerdialog des SWING-Abonnementdienstes

Dabei entspricht ein Knoten des Graphen jeweils einem HTML-Formular. Die

Kanten bilden die zur Verfügung stehenden Funktionen ab. Die Knoten F - P und Kanten 8 - 14 realisieren das Grobkonzept der Anfragebearbeitung. Durch den direkten Vergleich der Abbildung mit den konzipierten Funktionen des Abschnittes 6.1.3.1. ist die Vollständigkeit leicht zu überprüfen.

Die einzige Funktion, die nicht im Nutzerdialoggraphen enthalten ist, ist die regelmäßige Wiederholung der Nutzeranfragen. Hierbei handelt es sich um eine Funktion, die einem gewissen Zeitintervall unterliegt. Vom logischen Gesichtspunkt her, müßte die Funktion durch ECA-Regeln (siehe Abschnitt 6.1) des Datenbanksystems aktiviert werden. Das aktive Verhalten des Ingres-Datenbanksystems unterstützt nur sogenannte primitive Ereignisse, die auf Datenbankoperationen beruhen. Zeitereignisse hingegen beruhen nicht auf Datenbankoperationen. Sie werden durch das Eintreten eines spezifizierten Zeitpunktes ausgelöst. Um dennoch eine regelmäßige Anfragewiederholung vorzunehmen, wurde auf die Installation einer *crontab*-Datei unter UNIX zurückgegriffen. Innerhalb dieser Datei können 5 Zeitparameter und ein String angegeben werden. Der String wird durch die Shell zum angegebenen Zeitpunkt ausgeführt.

Beispielsweise wird durch die *crontab*-Datei

```
10 10 * * * 0-6 echo "Python ist (k)eine Schlange" > $HOME/python.tex
```

jeden Tag um 10.10 Uhr der String "Python ist (k)eine Schlange" in die Datei *python.tex* im Homeverzeichnis des *crontab*-Eigentümers geschrieben.

Kapitel 7

Zusammenfassung und Ausblick

Dieses Kapitel setzt sich aus der Zusammenfassung der Schwerpunkte dieser Studienarbeit, der Bewertung der Ergebnisse und einem Ausblick auf die weitere Verwendbarkeit der Arbeit zusammen.

7.1 Zusammenfassung

Der rasante Informationszuwachs führt mittlerweile zu keiner befriedigenden Informationsversorgung der Nutzer unter zu hilfenahme simpler Suchmaschinen. Der Trend führt in Richtung personalisierter Informationssysteme, die nach individuellen Präferenzen Information Retrieval, Information Filtering und Informationsbenachrichtigung vornehmen. Trotz der verbesserten Informationsqualität haben diese Systeme sowohl beim Information Retrieval als auch bei der Bereitstellung beim Nutzer Schwachstellen. Diese Schwachstellen sind in der Heterogenität des Internets begründet.

Im ersten Teil der Studienarbeit wurde der Versuch unternommen, ein Architekturmodell aufzustellen, das durch den Einsatz mobiler Agenten und durch die Leistungserweiterung des Informationsagenten einen Teil dieser Schwachstellen beseitigt. Dies sollte aber nicht durch die Vereinheitlichung der verwendeten Agentensprache, sondern durch standardisierte Schnittstellen erreicht werden. Mobile Agenten bilden eine gute Basis beim Information Retrieval, gerade wenn eine hohe Selektivität zu erwarten ist. Solange der Nutzer aber Zweifel an den Sicherheitsmechanismen beim Umgang mit seinen sensiblen Daten haben muß, wird ein dauerhafter Einsatz in Frage gestellt sein.

Der zweite Teil der Studienarbeit befaßte sich mit der Evaluierung der Funktionalität existierender Abonnementdienste, wobei der Gesichtspunkt auf den Möglichkeiten zur Interaktion der Funktionen in den SWING-Abonnementdienst lag. Aus den Ergebnissen wurde ein Konzept entwickelt und ein Datenbankent-

wurf für dieses Konzept vorgenommen. Neben der Umsetzung des Datenbankentwurfes in ein relationales Modell wurde ein Teil des doch recht umfangreichen Konzeptes im Prototypen implementiert.

7.2 Bewertung der Ergebnisse

Der Versuch ein Architekturmodell für einen zukünftigen Informationsdienst aufzustellen, hat gezeigt, daß zwischen der reinen theoretischen Betrachtung und den praktischen Möglichkeiten einer Umsetzung eine breite Lücke klafft. Selbst unter der Annahme einheitlicher Schnittstellen weisen viele der in diesem Modell gedachten Problemlösungen große Sicherheitsrisiken auf.

Der entstandene Prototyp des SWING-Abonnementdienstes weist die volle Anwendungsfunktionalität nur für einen Teil der Konzeption auf, so daß dieser auch noch nicht als vollwertiges Anwendungsprogramm betrachtet werden kann. Trotzdem eröffnen sich für einen Nutzer des SWING-Systems bereits neue Möglichkeiten der Anfragebearbeitung.

7.3 Ausblick

Im Abschnitt 6.1. wurde ein Ausblick auf zusätzliche Funktionalität gegeben. Diese Funktionalität sollte in einem nächsten Schritt miteinbezogen werden. Dazu sollte die vorhandene Konzeption verfeinert und erweitert werden. Es empfiehlt sich auch, einen weiteren Prototypen nicht auf der HTML-Basis zu gestalten. Die direkte Rückmeldung auf Ereignisse im System wird durch den statischen Aufbau von HTML verhindert. In diesem Zusammenhang sollten sich auch durch den Einsatz von ODBC (JDBC) ¹Freiheiten bei der Wahl des Datenbanksystems ergeben.

Eine weitere Möglichkeit bessere inhaltsbezogene Aussagen über die gefundenen Dokumente machen zu können, wäre die Verwendung von XML²/RDF³. In HTML können Informationen über Informationen (Metadaten) nur über das Meta-Tag innerhalb des HTML-Dokumentes weitergereicht werden. Mit Hilfe von RDF kann nun die Beschreibung des Inhaltes eines Dokumentes von der Darstellung getrennt werden. Diese Beschreibung muß nun nicht mehr Teil des HTML-Dokumentes sein. Im Gegensatz zu den Meta-Tags können die Beschreibungselemente eine beliebig tiefe Strukturierung erreichen. Die gemeinsame Syntax ist dabei XML. XML ist eine Untermenge von SGML, so daß auch jedes wohlgeformte HTML-Dokument ein gültiges XML-Dokument ist.

¹ODBC - Open Database Connectivity, JDBC - Java Database Connectivity

²XML - Extensible Markup Language

³RDF - Resource Description Framework

Beispiel:

Es existiert eine Schema-Beschreibung in XML die Entities und Attribute von Unterkünften enthält. Diese könnte beispielsweise folgende Gestalt haben:

```
<! DOCTYPE MyUnterkunft [
<! ELEMENT hotel (Name, Adresse) >
<! ELEMENT Name (#PCDATA) >
<! ELEMENT Adresse (Strasse, Ort, PLZ) >
<! ELEMENT Strasse (#PCDATA) >
<! ELEMENT Ort (#PCDATA) >
<! ELEMENT PLZ (#PCDATA) >
]>
```

Eine Beschreibung eines Dokumentes, das Informationen über Hotels enthält könnte dann folgendermaßen aussehen:

```
<?xml:namespace ns = "http://www.w3.org/RDF/RDF" prefix="RDF" ?>
<?xml:namespace ns = "http://uri-unterkunft-schema/" prefix="MU" ?>
<RDF:RDF>
  <RDF:Description RDF:HREF = "http://uri-of-Unterkunft-Dokument">
    <MU:hotel>
      <MU:Name>Schöne Aussicht</MU:Name>
      <MU:Adresse>
        <MU:Strasse>Am Strand 20</MU:Strasse>
        <MU:Ort>Warnemünde</MU:Ort>
        <MU:PLZ>18123</MU:PLZ>
      </MU:Adresse>
    </MU:hotel>
  </RDF:Description>
</RDF:RDF>
```

Die Beschreibung des Dokumentes mit Hilfe von RDF und dem Unterkunft-Schema liefert nun gezielte Informationen zum Namen und Anschrift des Hotels. Aufgrund der Strukturierung der Daten ist ihre Aussagekraft sehr hoch, so daß diese Informationen beispielsweise direkt in entsprechende Attributfelder von Datenbanken einfließen könnten. Umgekehrt ist es nun auch möglich gezielte Anfragen an eine entsprechende Ressource zu stellen.

Literaturverzeichnis

- [BZW98] Brenner, W.; Zarnekow, R.; Wittig, H.: *Intelligente Softwareagenten*. Springer Verlag, Berlin, 1998
- [CC97] Chang, D.; Covaci, S.: The OMG Mobile Agent Facility: A Submission. In: *First International Workshop on Mobile Agents 1997 (MA'97)*, LNCS 1219, 1997
- [DG96] Dittrich, K.; Gatzju, S.: *Aktive Datenbanksysteme*. Thomson Publishing, Bonn, 1996
- [Fir97] Firefly Network Inc.: *Collaborative Filtering Technology: An Overview*. <http://www.firefly.net/products/CollaborativeFiltering/html>. 1997
- [FG96] Franklin, S.; Graesser, A.: Is It an Agent, or Just a Program ?. In: *Proceedings on ECAI Workshop on Agent Theories, Architectures and Languages (ATLA'96)*, LNAI 1193, Seite 21-35, 1996
- [FZ96] Franklin, M.; Zdonik, S.: Dissemination-Based Information Systems. In: *IEEE Bulletin of the Technical Committee on Data Engineering*, Vol. 20, Seite 20-30, 1996
- [Glo98] Global-Info: *Vorschlag der technischen Arbeitsgruppe in Schwerpunkt 4 zur Rahmenarchitektur*. 1998
- [Her96] Hermans, B.: *Intelligent Software Agents on the Internet*. Diplomarbeit an der Universität Tilburg, Niederlande, Juli 1996
- [HMDL97] Heuer, A.; Meyer, H.; Düsterhöft, A.; Langer, U.: SWING: Der Anfrage- und Suchdienst des Regionalen Informationssystems MV-Info. In: *Tagungsband IuK-Tage Mecklenburg-Vorpommern*, Schwerin, Juni 1997
- [Ing91] Ingres Corp.: *Ingres Release 6.4 SQL Reference Manual*. 1991

- [KE96] Kemper, A.; Eickler, A.: *Datenbanksystem*. Oldenbourg Verlag, München, 1996
- [Kop96] Kopka, H.: *L^AT_EX Einführung*. Addison-Wesley, Bonn, 1996
- [Klu96] Klute, R.: *Das World Wide Web*. Addison-Wesley, Bonn, 1996
- [LF97] Martin von Löwis, Nils Fischbeck: *Das Python-Buch*. Addison-Wesley, Bonn, 1997
- [Loe97] Loeser, H.: *Datenbankanbindung an das WWW*. Tagungsband BTW, 1997
- [MRK96] Magedanz, T.; Rothermel K.; Krause, S.: Intelligent Agents: An Emerging Technology for Next Generations Telecommunications?. In *INFOCOMM'96*, San Francisco, 1996
- [Ariad] Ariadne: <http://ariadne.inf.fu-berlin.de:8000/>
- [Alta] AltaVista Suchmaschine: <http://www.altavista.digital.com>
- [Back] Backweb Infocenter.: <http://www.backweb.com>
- [Else] Elsevier Science: <http://www.elsevier.nl>
- [Excite] Excite Suchmaschine: <http://www.excite.com>
- [Fire] Firefly Network Inc.: <http://www.firefly.com>
- [Info] InfoSeek Suchmaschine: <http://www.infoseek.com>
- [Infor] The Informant: <http://informant.dartmouth.edu/>
- [Inter] Intermind Communicator.: <http://www.intermind.com>
- [Link] LINK Information Service.: <http://link.springer.de>
- [Lycos] Lycos Suchmaschine: <http://www.lycos.com>
- [Marim] Marimba Castanet 3.0.: <http://www.marimba.com>
- [Point] Pointcast Business Network: <http://www.pointcast.com>
- [Python] Official Python Homepage.: <http://www.python.org>
- [W3Con] World Wide Web Consortium.: <http://www.w3.org/P3P>

Abbildungsverzeichnis

3.1	Charakteristische Eigenschaften intelligenter Agenten	12
3.2	Klassifikationsmatrix nach [BZW98]	14
3.3	Einordnung von Informationsagenten in die Klassifikationsmatrix	15
3.4	Klassifikationsmatrix eines Fernsehagentensystems	17
3.5	Mechanismen der Informationsbereitstellung	18
4.1	Aufbau eines Abonnementdienstes	21
4.2	Architekturmodell eines Abonnementdienstes	24
4.3	Aufbau des Personal Digital Assistant	26
4.4	Beispiel eines Informationsagenten	28
4.5	Aufbau der Basissoftware des Informationsanbieters	29
4.6	Klassifikation des Abonnementdienstes	30
6.1	ER-Schema des SWING-Abonnementdienstes	50
6.2	Architektur des SWING-Abonnementdienstes	52
6.3	Nutzerdialog des SWING-Abonnementdienstes	56

Anhang A

Das SWING-Datenbankschema

Attribut	Domäne	Schlüssel	Integritäts- bedingung	Status
UID	Integer	Primär	> 0	Not Null
Email	Character(8)	Sekundär		Not Null
Name	Character(30)			Not Null
Vorname	Character(30)			
Geschlecht	Character(1)		F v M	Not Null
Passwort	Character(13)			Not Null
Counter	Integer			

Tabelle A.1: Relation: SwingUser

Attribut	Domäne	Schlüssel	Integritäts- bedingunn	Status
QNR	Integer	Primär	> 0	Not Null
UID	Integer	Fremdschl. (SwingUser)		Not Null
Query	Character(128)			Not Null
Frequenz	Integer			
Output	Character(1)			
Querydatum	Date			Not Null
Querycounter	Integer			

Tabelle A.2: Relation: Anfrage

Attribut	Domäne	Schlüssel	Integritäts- bedingung	Status
UNR	Integer	Primärschl. Fremdschl. (URL)		Not Null
QNR	Integer	Primärschl. Fremdschl. (Anfrage)		Not Null
Ranking	Integer			
Statuscode	Integer			
Lastmodified	Date			
Checksum	Character(16)			
Checkdate	Date			

Tabelle A.3: Relation: Anfrageergebnis

Attribut	Domäne	Schlüssel	Integritäts- bedingung	Status
UNR	Integer	Primärschl.	> 0	Not Null
URL	Character(256)	Sekundär		Not Null
Description	Character(150)			
Object	Char(1)			

Tabelle A.4: Relation: URL

Attribut	Domäne	Schlüssel	Integritäts- bedingung	Status
RNR	Integer	Primärschl.	> 0	Not Null
UNR	Integer	Fremdschl.(URL)		Not Null
UID	Integer	Fremdschl.(SwingU.)		Not Null
Frequenz	Integer			Not Null
Checksum	Character(16)			
Nextupdate	date			Not Null

Tabelle A.5: Relation: Reminder

Attribut	Domäne	Schlüssel	Integritäts- bedingung	Status
PNR	Integer	Primärschl.	> 0	Not Null
BeginTag	Character(20)			Not Null
EndTag	Character(20)			

Tabelle A.6: Relation: Pattern

Attribut	Domäne	Schlüssel	Integritäts- bedingung	Status
PNR	Integer	Primärschl. Fremdschl. (Pattern)		Not Null
RNR	Integer	Primärschl. Fremdschl. (Reminder)		Not Null
Checksum	Character(16)			

Tabelle A.7: Relation: RePat

Attribut	Domäne	Schlüssel	Integritäts- bedingung	Status
NNR	Integer	Primärschl.	> 0	Not Null
Nachricht	Character(1800)			
Timestamp	Date			Not Null
Absender	Character(80)			Not Null
Subject	Character(80)			

Tabelle A.8: Relation: News

Attribut	Domäne	Schlüssel	Integritäts- bedingung	Status
NNR	Integer	Primärschl. Fremdschl. (News)		Not Null
UID	Integer	Primärschl. Fremdschl. (SwingUser)		Not Null
Read	Char(1)			

Tabelle A.9: Relation: erhaelt

Attribut	Domäne	Schlüssel	Integritäts- bedingung	Status
ThemenURL	Integer	Fremdschl. (URL)		Not Null
Schluessel	Integer	Primärschl		Not Null
Gruppe	Character(40)			
Bereich	Character(40)			
Status	Character(1)			

Tabelle A.10: Relation: Themenkatalog

Attribut	Domäne	Schlüssel	Integritäts- bedingung	Status
Schlüssel	Integer	Primärschl. Fremdschl. (Themenk.)		Not Null
UID	Integer	Primärschl. Fremdschl. (SwingU.)		Not Null

Tabelle A.11: Relation: waehlt-aus

Anhang B

Konfliktauflösung im SWING-Schema

Operation	Relation	Attribut	Verfahren	betroffene Relation
Delete	Anfrage	QNR	cascade	Anfrageergebnis
	News	NNR	cascade	erhaelt
	Pattern	PNR	cascade	RePatt
	Reminder	RNR	cascade	RePatt
	Themenkatalog	Schluessel	cascade	waehtl-aus
	URL	UNR / URL	restricted	Anfrageergebnis Themenkatalog Reminder
	SwingUser	UID / EMAIL	cascade	Anfrage Reminder erhaelt waehtl-aus
Insert	Anfrageergebnis	QNR / UNR	restricted	
	erhaelt	UID / NNR	restricted	
	Reminder	UID / UNR	restricted	
	RePatt	RNR / PNR	restricted	
	waehtl-aus	UID / Schluessel	restricted	
	SwingUser	Email	restricted	
	URL	URL	restricted	

Tabelle B.1: Sicherung der referentiellen Integrität

An einigen Stellen entspricht das Konfliktauflösungsverfahren nicht der Beschreibung aus Abschnitt 6.2.2. So besitzt das Attribut URL in URL keinen Fremdschlüssel. Die restriktiven Insert bzw. Update-Operationen beziehen sich auf

Operation	Relation	Attribut	Verfahren	betroffene Relation
Update	Pattern	PNR	cascade	RePatt
	Themenkatalog	Schlüssel	cascade	wahlt-aus
	SwingUser	Email	restricted	
	URL	URL	restricted	
	Anfrageergebnis	QNR	forbidden	
	Anfrageergebnis	UNR	forbidden	
	Anfrage	QNR	forbidden	
	erhaelt	UID	forbidden	
	erhaelt	NNR	forbidden	
	News	NNR	forbidden	
	Reminder	UID	forbidden	
	Reminder	UNR	forbidden	
	RePatt	RNR	forbidden	
	SwingUser	UID	forbidden	
	URL	UNR	forbidden	
	wahlt-aus	Schlüssel	forbidden	
	wahlt-aus	UID	forbidden	

die Relation selbst. Identisches gilt auch für das Attribut Email in der Relation SwingUser.