

Universität Hamburg  
Fachbereich Informatik  
Arbeitsbereich Verteilte Systeme und  
Informationssysteme (VSIS)



DIPLOMARBEIT

## **Kollaborative Nutzung des World Wide Webs**

Frank Wollenweber  
im Januar 2004

Erstgutachter: Prof. Dr. Winfried Lamersdorf  
Zweitgutachter: Prof. Dr. Horst Oberquelle



Diplomarbeit:  
Kollaborative Nutzung des World Wide Webs

Autor:  
Frank Wollenweber  
Harnacksweg 44  
22417 Hamburg  
E-Mail: mail@frank-wollenweber.de  
Matrikelnummer: 4817284

Erstgutachter:  
Prof. Dr. Winfried Lamersdorf  
Arbeitsbereich Verteilte Systeme und Informationssysteme (VSIS)

Zweitgutachter:  
Prof. Dr. Horst Oberquelle  
Arbeitsbereich Angewandte und Sozialorientierte Informatik (ASI)

Universität Hamburg  
Fachbereich Informatik  
Vogt-Kölln-Straße 30  
22527 Hamburg

© Copyright Frank Wollenweber 2004

Alle Rechte vorbehalten

## **Erklärung**

Hiermit erkläre ich an Eides statt, daß ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfaßt, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt und die aus anderen Quellen entnommenen Stellen als solche gekennzeichnet habe.

Außerdem erkläre ich, daß ich mit der Einstellung dieser Diplomarbeit in den Bestand der Bibliothek des Fachbereiches Informatik der Universität Hamburg einverstanden bin.

Hamburg, am 8. Januar 2004

Frank Wollenweber

## Danksagung

Bei Herrn Prof. Dr. Winfried Lamersdorf und Herrn Prof. Dr. Horst Oberquelle bedanke ich mich für die Übernahme der Begutachtung. Mein besonderer Dank für die vielen Anregungen und Hinweise während der Erstellung dieser Arbeit gilt den Mitarbeitern unseres Fachbereichs, Herrn Tobias Baier und Herrn Harald Weinreich.

Für die Erstellung des Textes wurde  $\text{\LaTeX}$  verwandt. Als Grundlage diente dabei die Formatvorlage `hagenberg.sty`, bei deren Autor, Herrn Dr. Wilhelm Burger von der Fachhochschule Hagenberg in Österreich, ich mich ebenfalls bedanken möchte.

## Kurzfassung

Kognitive Überforderung bei der Navigation im Web, Information Overload, mangelndes Vertrauen in die Anbieter von Information und fehlende objektiv überprüfbare Qualitätsstandards sind einige der Probleme, mit denen Benutzer des World Wide Webs bei ihren oft zeitraubenden Recherchen konfrontiert werden. Einfacher als eigene Recherchen und vielfach zu besseren Ergebnissen führend ist es, einen Kollegen nach „Link-Tips“ zu einem Thema zu befragen. Derartige Verhaltensweisen werden in der Literatur unter dem Stichwort Social Navigation zusammengefaßt. Dieses ist ein Designansatz für Informationssysteme, der es Benutzern ermöglichen soll, Navigationsentscheidungen basierend auf der direkten oder indirekten Interaktion mit anderen Benutzern zu treffen. Recommender-Systeme formalisieren den Vorgang des „sich Tips geben“ und aggregieren eine Vielzahl von Bewertungen zu individuellen Empfehlungen.

Ziel der vorgestellten Diplomarbeit ist die Entwicklung eines Empfehlungs- und Annotationssystems für Web-Seiten. Da abgegebene Bewertungen auf den subjektiven Einschätzungen der Benutzer basieren, ist eine individuelle Anpassung des Systems wünschenswert, so daß die subjektiven Äußerungen derjenigen Benutzer bevorzugt Berücksichtigung finden, deren Interessen mit den eigenen übereinstimmen. Um dieses Ziel einer individuellen Selektion der Information zu erreichen, wurden Collaborative-Filtering-Techniken eingesetzt, die um die Auswertung der durch die onefC-Identitätsinfrastruktur übermittelten Benutzerprofile erweitert wurden. Ferner wurde ein Vertrauensmodell integriert, so daß jeder Benutzer steuern kann, welche Meinungsäußerungen anderer Benutzer besonderes Gewicht haben.

Das realisierte System selbst erbringt lediglich eine Vermittlerfunktion bei der Zuordnung von Empfehlungen suchenden zu gleichgesinnten bzw. vertrauenswürdigen, Bewertungen abgebenden Personen. Um einer derartigen Aufgabe gerecht zu werden, ist Wissen über die Eigenschaften der Benutzer erforderlich, welches in Benutzerprofilen zu erfassen ist. Das Empfehlungs- und Annotationssystem ist eine Plattform, über die Individuen miteinander kommunizieren, welche die Gemeinschaft der Nutzer bilden. In diesem Zusammenhang ist die Frage nach den Auswirkungen der Integration einer Identitätsinfrastruktur auf eine kollaborative Internetanwendung zu sehen, die in dieser Arbeit bearbeitet wurde.

# Inhaltsverzeichnis

<b>1. Einleitung</b>	<b>1</b>
1.1. Ziel der Arbeit . . . . .	3
1.2. Eingrenzung des Themas . . . . .	5
1.3. Struktur der Arbeit . . . . .	6
1.4. Zur Verwendung englischer Begriffe . . . . .	8
<b>2. Grundlagen</b>	<b>9</b>
2.1. Computer Supported Cooperative Work . . . . .	9
2.1.1. Begriffsbestimmung . . . . .	9
2.1.2. Formen der Interaktion . . . . .	10
2.1.3. Klassifikation von CSCW-Systemen . . . . .	10
2.1.4. Teams, Gruppen und Communities . . . . .	13
2.1.5. Community Support Systems . . . . .	14
2.1.6. Awareness . . . . .	17
2.2. Computer als soziales Medium . . . . .	20
2.2.1. Identität . . . . .	20
2.2.2. Vertrauen . . . . .	23
2.2.3. Reputation . . . . .	24
2.3. open net environment for Citizens (onefC) . . . . .	28
2.3.1. Dienste . . . . .	30
2.3.2. Architektur . . . . .	30
2.3.3. Identitätsmodell . . . . .	32
2.4. Scone . . . . .	33
2.5. HyperScout . . . . .	38
<b>3. Herausforderungen bei der Benutzung des World Wide Webs</b>	<b>41</b>
3.1. Lost in Hyperspace . . . . .	41
3.2. Information Overload . . . . .	43
3.3. Qualität im Web . . . . .	45
3.4. Vertrauen im Web . . . . .	46
3.5. Tools zur Unterstützung der Navigation im Web . . . . .	47

<b>4. Social Navigation</b>	<b>49</b>
4.1. Navigation in Informationsräumen . . . . .	50
4.2. Definitionsansätze von Social Navigation . . . . .	51
4.3. Klassifikation verschiedener Formen von Social Navigation . . . . .	54
4.4. Direkte Social Navigation . . . . .	57
4.5. Indirekte Social Navigation . . . . .	59
4.6. Voraussetzungen und Randbedingungen . . . . .	65
4.7. Wobei nützt Social Navigation? . . . . .	68
4.8. Probleme und offene Fragestellungen . . . . .	71
<b>5. Empfehlungssysteme</b>	<b>75</b>
5.1. Content-based Filtering . . . . .	79
5.1.1. Vorteile . . . . .	80
5.1.2. Nachteile . . . . .	80
5.2. Collaborative Filtering . . . . .	81
5.2.1. Pearson-r-Algorithmus . . . . .	87
5.2.2. Vorteile . . . . .	90
5.2.3. Nachteile . . . . .	90
5.2.4. Evaluationsmethoden . . . . .	92
5.3. Hybride Filtertechnologien . . . . .	93
5.4. Item-to-Item-Korrelation . . . . .	94
5.5. Social Data Mining . . . . .	95
<b>6. Verwandte Arbeiten</b>	<b>98</b>
6.1. WebPlaces . . . . .	100
6.2. GroupLens . . . . .	104
6.3. AntWorld . . . . .	108
<b>7. Anforderungen und Forschungsfragen</b>	<b>111</b>
7.1. Anforderungen . . . . .	111
7.2. Forschungsfragen . . . . .	113
<b>8. Collaborative Internet Experience</b>	<b>115</b>
8.1. Überblick . . . . .	115
8.1.1. Szenarien . . . . .	117
8.1.2. Navigation entlang sozialer Spuren . . . . .	118
8.1.3. Bewertungen und Annotationen sind Meta-Daten . . . . .	119
8.1.4. Community-Gedanke statt eines anonymen Empfehlungssystems . . . . .	120
8.2. Konzeption . . . . .	121
8.2.1. Benutzerprofile . . . . .	122
8.2.2. Abgabe von Bewertungen . . . . .	129

8.2.3.	Annotationen . . . . .	131
8.2.4.	Vertrauen . . . . .	131
8.2.5.	Vorhersage von Bewertungen . . . . .	133
8.2.6.	Kombination eines Empfehlungssystems mit Suchmaschinen . . . . .	144
8.2.7.	Benutzungsschnittstelle . . . . .	147
8.3.	Konstruktion . . . . .	149
8.3.1.	Systemarchitektur . . . . .	151
8.3.2.	Server-Komponenten . . . . .	156
8.3.3.	Client-Komponenten . . . . .	164
8.3.4.	Protokoll zwischen Client und Server . . . . .	171
8.4.	Evaluation . . . . .	173
8.4.1.	Vorgehen . . . . .	174
8.4.2.	Ergebnisse . . . . .	177
8.5.	Auswirkungen der Integration einer Identitätsinfrastruktur . . . . .	180
<b>9.</b>	<b>Fazit</b>	<b>187</b>
9.1.	Ergebnisse der Arbeit . . . . .	187
9.2.	Offene Fragen und Ausblick . . . . .	192
	<b>Abbildungsverzeichnis</b>	<b>197</b>
	<b>Tabellenverzeichnis</b>	<b>200</b>
	<b>Literaturverzeichnis</b>	<b>201</b>
	<b>A. Inhalt der CD-ROM</b>	<b>214</b>

# 1. Einleitung

The power of the network increases exponentially by the number of computers connected to it. Therefore, every computer added to the network both uses it as a resource while adding resources in a spiral of increasing value and choice. *Zitat: Robert Metcalfe, Erfinder des Ethernets und Mitbegründer der Firma 3Com*

Dieses als Metcalfe's Law bekannte Zitat mag die Entwicklung des Internets als Ganzes charakterisieren. Bezogen auf das World Wide Web (WWW) trifft es nur teilweise zu. Sicherlich ist die Attraktivität des Webs mit der Zahl an Servern, auf denen Informationen publiziert werden, gestiegen. Gleichzeitig ist es jedoch einhergehend mit dem Wachstum des WWW tendenziell schwieriger geworden, persönlich relevante, qualitativ hochwertige Inhalte aus vertrauenswürdigen Quellen aufzufinden. Die technischen und materiellen Hürden der Publikation im Web sind gering. Sicherlich ist dies einer der Gründe für die dynamische Entwicklung, die dieses Medium in den letzten Jahren genommen hat. Es mangelt dem Web jedoch an Kontrollinstanzen, die ähnlich dem Verlagswesen für die Qualität der veröffentlichten Informationen und die Seriosität der Anbieter bürgen. Ebenso mangelt es an Strukturen, die eine gezielte Navigation zu relevanten Inhalten ermöglichen.

Betrachtet man die Entwicklung der Nutzerzahlen bzw. der Zahl an Client-Rechnern, so ist Metcalfe's Law bezogen auf das heutige Web als ebenfalls nur mit Einschränkungen zutreffend zu bezeichnen. Web-Server können simultan mehrere Benutzer bedienen. Diese erfahren das Web jedoch so, als würden sie alleine mit den Servern interagieren. Sie sehen weder welche Benutzer sich noch auf der Site befinden, noch registrieren sie die aus vergangenen Handlungen der Benutzer resultierenden Spuren. Es ist beispielsweise nicht ersichtlich, welche Pfade Personen mit übereinstimmenden Interessen gegangen sind. Im Gegensatz zum Buch in der Bibliothek sieht man einem Web-Dokument auch nicht an, wie alt es ist und wie oft es in dieser Zeit gelesen wurde. Große Suchmaschinen wie Google werden jeden Tag von sehr vielen Benutzern als Ausgangspunkt für ihre Recherchen benutzt. Dabei ist es wahrscheinlich, daß viele Benutzer nach ähnlichen Themen suchen. Trotzdem fängt jeder Benutzer mit seiner Recherche wieder bei Null an, ohne daß er von den vorhergehenden Recherchen anderer Benutzer profitieren kann.

Die Möglichkeiten der Kommunikation mit anderen Nutzern beschränken sich im Web auf site-spezifische Lösungen, wie beispielsweise Web-Foren. Die Kommunikation mit

anderen Nutzern im Rahmen solcher Angebote wird dadurch erschwert, daß das Bild, das Nutzer voneinander haben, sehr eingeschränkt ist. Es fehlt an Möglichkeiten, Attribute der eigenen Persönlichkeit anderen Nutzern gegenüber auszudrücken. Folglich fällt es schwer, geeignete Kommunikationspartner zu finden sowie deren Kompetenz und Vertrauenswürdigkeit zu beurteilen. Der Gebrauchswert der im Web gebotenen Möglichkeiten der Kommunikation mit anderen Nutzern ist nicht zuletzt deshalb limitiert, da es an einem einheitlichen Identitätskonzept mangelt. Personen, die auf einer Web-Site miteinander kommuniziert haben, können sich nicht in einem anderen Umfeld im Web wiedererkennen. Der Aufbau von Beziehungen zwischen Nutzern des Webs über Grenzen von Web-Sites hinweg wird somit erschwert.

Das Oberthema dieser Arbeit ist die Unterstützung von Personen bei der Benutzung des World Wide Webs. Aus diesem Blickwinkel ist die Frage zu stellen, wie das Auffinden von persönlich relevanten Web-Seiten effizienter gestaltet werden kann, wie Benutzer dabei unterstützt werden können, in den sie interessierenden Themen auf dem Laufenden zu bleiben und wie das Vertrauen in Inhalte und Anbieter gestärkt werden kann.

Der in dieser Arbeit eingeschlagene Weg ist es, Benutzer des Webs dazu zu befähigen, Navigationsentscheidungen basierend auf der Orientierung an bzw. Kommunikation mit anderen Nutzern zu treffen. Derartige Verhaltensweisen werden unter dem Stichwort *Social Navigation* zusammengefaßt und im Verlauf der Arbeit detailliert betrachtet. Social Navigation setzt eine Wahrnehmung der Nutzer untereinander und das Vorhandensein von Einrichtungen zur direkten oder indirekten Kommunikation voraus. Auf diese Weise kann die Gestalt des Webs aus der Sichtweise der Anwender von einem menschenleeren Informationsraum zu einem bevölkerten Platz transformiert werden. Social Navigation bietet die Chance, menschliches Vermögen, Informationen auf einer semantischen Ebene zu beurteilen, in Werkzeuge zur Unterstützung der Navigation im Web einzubeziehen und somit eine unter Umständen qualitativ bessere Unterstützung zu erzielen als rein rechnergestützte Werkzeuge, etwa Suchmaschinen, zu leisten im Stande sind.

Soziales Navigieren setzt voraus, daß einige Personen Navigationshinweise anbieten, die von anderen Nutzern bei ihren Navigationsentscheidungen berücksichtigt werden können. Navigationshinweise können entweder durch explizite Aktionen, beispielsweise durch die Bewertung eines Dokumentes, gegeben werden, oder implizit aus dem Verhalten der Benutzer, im Extremfall sogar nur aus deren Gegenwart, abgeleitet werden. Benutzer kollaborieren bei der Benutzung des World Wide Webs, indem sie prinzipiell symmetrisch die Rolle der Navigationshinweise bereitstellenden und der sich an Navigationshinweisen orientierenden Person einnehmen.

In einem Informationsraum, in dem Benutzer Spuren hinterlassen, entweder durch von ihnen erstellte Artefakte, oder durch die aufgezeichnete Historie der Benutzung der Objekte der Umgebung, fügen sie dem Informationsraum Ressourcen hinzu. Diese können anderen Nutzern von Wert sein, denn sie können von der Arbeit und von den Erfahrungen

vorheriger Benutzer profitieren. Der Wert des Informationsraumes wächst somit mit der Zahl an Nutzern. Gestattet ein Informationsraum die direkte Kommunikation zwischen Nutzern, so können diese bei der Navigation kollaborativ vorgehen. Mit zunehmenden Nutzerzahlen steigt in diesem Fall auch die Zahl potentieller Kommunikationspartner. Der Wert von Systemen zum sozialen Navigieren ist abhängig von der Zahl der Nutzer. Folglich würde ein um Funktionen zum sozialen Navigieren erweitertes Web vortrefflich durch Metcalfes' Law charakterisiert werden.

Diese Arbeit ist in mancherlei Hinsicht als die Fortführung der Studienarbeit [Wol02] des Autors anzusehen. In beiden Arbeiten geht es darum, Menschen bei ihren Recherchen im World Wide Web zu unterstützen: in der Studienarbeit mittels persönlicher Web-Robots, in dieser Diplomarbeit durch Kollaboration mit anderen Nutzern.

In den restlichen Seiten dieses einleitenden Kapitels werden zunächst die Ziele und anschließend die Struktur dieser Arbeit dargestellt.

## 1.1. Ziel der Arbeit

Ziel der Arbeit ist die Entwicklung eines Empfehlungs- und Annotationssystems für Web-Seiten, welches die Kollaboration zwischen Nutzern ermöglichen und fördern soll.

Benutzer können mittels des Systems Web-Dokumente bewerten und mit textuellen Annotationen, beispielsweise Kommentaren und weitergehenden Links, versehen. Bewertungen werden unter Berücksichtigung in Benutzerprofilen modellierter Präferenzen der Benutzer zu individuellen Empfehlungen aggregiert. Die erstellte Anwendung wurde mit dem Titel *Collaborative Internet Experience* bezeichnet und wird im Text teilweise unter der Kurzform *CoInternet*-System referenziert.

Jakob Nielsen beschreibt in seiner Columnne „Alertbox“ ein von ihm als „Reputation-Manager“<sup>1</sup> bezeichnetes System, mit dem Benutzer ihre Erfahrungen mit Web-Sites kommunizieren und somit anderen Personen bei der Auffindung qualitativ hochwertiger Beiträge behilflich sein können [Nie98]. Das im praktischen Teil dieser Arbeit entwickelte System teilt diese Vision, ein Medium zur indirekten Kommunikation zwischen Web-Nutzern aufzubauen. Die Zielrichtung ist jedoch eine andere. Beim Reputation Manager geht es vornehmlich darum, die Erfahrungen mit der Qualität der Inhalte und der Vertrauenswürdigkeit einer Web-Site<sup>2</sup> an nachfolgende Benutzer zu kommunizieren, die

---

<sup>1</sup>Nielsen scheint von einem etwas anderen Reputations-Begriff auszugehen, als er im Abschnitt 2.2.3 dieser Arbeit beschrieben wird. Aus seiner Sicht ist die Reputation einer Web-Site nicht nur ein Maß für die Vertrauenswürdigkeit einer Site, sondern auch für die Kompetenz und Leistungsfähigkeit einer Site bezüglich der Erstellung qualitativ hochwertiger Inhalte.

<sup>2</sup>Es ist nur dann sinnvoll von der Reputation einer Web-Site zu sprechen, wenn sich unter dem Domain-Namen nur eine Person oder Organisation einheitlich präsentiert.

diese Angaben als ein globales Urteil zur Qualität des Gesamtangebots eines Anbieters und zur zu erwartenden Qualität der zukünftig auf der jeweiligen Site veröffentlichten Seiten verstehen können. Ziel des entwickelten Systems ist es in erster Linie, die Anwender bei der Auffindung sie persönlich interessierender, qualitativ hochwertiger Inhalte von vertrauenswürdigen Anbietern zu unterstützen.

Eine derartige Anwendung ist als eine Kommunikationsplattform zu sehen, über die Individuen asynchron und indirekt bei der Benutzung des Webs miteinander kommunizieren können und somit zu sozialem Navigieren befähigt werden. Die Einschätzung der Relevanz von Navigationshinweisen, welche vermittelt durch das System kommuniziert wurden, wird maßgeblich von der Wahrnehmung der Identität der Kommunikationspartner geprägt, wenn es beispielsweise darum geht, deren Vertrauenswürdigkeit und Kompetenz zu beurteilen.

Die Rolle des Systems ist primär eine Vermittlerfunktion zwischen Empfehlungen suchenden und Bewertungen abgebenden Personen. Der Entwurf einer derartigen Social-Navigation-Anwendung darf nicht allein aus einer technischen Perspektive heraus angegangen werden. Zu berücksichtigen sind die Implikationen technischer Entwurfsentscheidungen auf das soziale Verhalten des Nutzer.

Diese Arbeit ist Teil des Projekts *open net environment for Citizens (onefC)*<sup>3</sup>, welches sich mit der Entwicklung einer Identitätsinfrastruktur befaßt. Diese umfaßt ein universelles Konzept zur Benutzeridentifikation und ein generisches Modell zur Repräsentation von Attributen der Benutzeridentitäten. Somit wird der anwendungsübergreifende Austausch von Identitätsdaten unter Kontrolle der Benutzer durchführbar.

Im Rahmen des Projektes onefC leistet diese Arbeit den Beitrag, anhand des Empfehlungs- und Annotationssystems, exemplarisch für eine kollaborative Internetanwendung die Auswirkungen der Integration einer Identitätsinfrastruktur und der damit verbundenen weitergehenden Möglichkeiten der Benutzermodellierung auf eine kollaborative Internetanwendung aufzuzeigen. Die Untersuchung erfolgt in zwei Schritten:

Zunächst ist zu klären, wie sich der Gestaltungsspielraum bei Entwurf und Konstruktion einer derartigen Anwendung hinsichtlich des Ziels, die Unterstützung von Social Navigation zu optimieren, verändert. Wie kann die Identitätsinfrastruktur eingesetzt werden, um den Aufbau von Vertrauen zwischen den Anwendern zu begünstigen und wie kann gleichzeitig das berechnete Interesse der Benutzer nach Schutz ihrer Privatsphäre berücksichtigt werden? Wie können unter Zuhilfenahme der Identitätsinfrastruktur Navigationsunterstützung suchende Personen und passende Anbieter von Navigationshinweisen zusammengebracht werden? Des weiteren ist von Interesse, welchen Einfluß die Identitätsinfrastruktur auf die technische Realisierbarkeit eines Empfehlungssystems hat.

---

<sup>3</sup>[www.informatik.uni-hamburg.de/projects/onefc](http://www.informatik.uni-hamburg.de/projects/onefc)

Aus dieser Sichtweise stellt sich die Frage, wie sich die Identitätsinfrastruktur als Abstraktion identitätsbezogener Funktionen des Systems, wie der Persistierung und Übertragung von Identitätsdaten in die Anwendungsarchitektur integrieren läßt oder wie die Algorithmen zur Vorhersage von Bewertungen durch Einbeziehung unterschiedlicher Profilmerkmale verbessert werden können.

In einem zweiten Schritt sind dann die sozialen Implikationen der getroffenen Entwurfsentscheidungen festzustellen. Es werden Antworten auf die Frage gesucht, wie sich die Integration der Identitätsinfrastruktur und die damit verbundenen weitergehenden Systemfunktionalitäten auf das Verhalten der Anwender oder der von ihnen empfundenen Nützlichkeit der Anwendung auswirkt. Damit ist die Frage verknüpft, wie Anwender das System, die anderen Nutzer und die Gemeinschaft wahrnehmen. Diese Fragestellungen erfordern umfangreiche Evaluationen des erstellten Prototyps, welche im Rahmen dieser Arbeit nicht geleistet werden können<sup>4</sup>. Die Antworten, die diese Arbeit auf dieser Betrachtungsebene zu geben vermag, sind theoretischer Natur.

OnefC ist ein laufendes Projekt. Ob es jemals den akademischen Rahmen der Forschung am Fachbereich Informatik der Universität Hamburg verlassen wird, ist ungewiß und wird auch nicht unbedingt angestrebt. Folglich gibt es mit Ausnahme des in dieser Arbeit erstellten Systems noch keine weiteren auf der Identitätsinfrastruktur aufsetzenden Anwendungen. Prämisse der nachfolgenden Betrachtung der Auswirkungen der Integration einer Identitätsinfrastruktur ist die Annahme, daß es noch weitere Anwendungen gäbe, beispielsweise einen erweiterten Instant-Messenger, über welche die Anwender kommunizieren und Identitätsdaten austauschen können.

## 1.2. Eingrenzung des Themas

Zunächst einmal ist festzustellen, daß sich diese Arbeit mit der kollaborativen Nutzung des World Wide Webs beschäftigt, nicht jedoch mit dem Web als Medium zur Kollaboration, also mit web-basierter Groupware. In dieser Arbeit werden verschiedene Ansätze vorgestellt, wie Benutzer des World Wide Webs bei der Navigation in diesem Medium kollaborieren können. Der Schwerpunkt liegt dabei, entsprechend der Festlegung, ein Empfehlungs- und Annotationssystem für Web-Seiten entwickeln zu wollen, auf der Betrachtung von Formen asynchroner, indirekter Interaktion zwischen Web-Nutzern.

Es wurde die prototypische Umsetzung eines Empfehlungs- und Annotationsystems für Web-Seiten vorgenommen. Mit der Entwicklung dieses Systems wurden im wesentlichen zwei Ziele verfolgt. Zum einen ging es darum, einen Ansatz vorzustellen, wie das Web um Funktionen zum sozialen Navigieren erweitert werden könne, zum anderen sollte anhand

---

<sup>4</sup>Aussagekräftige Tests eines Empfehlungssystems erfordern eine große Zahl von Benutzern und einen längeren Testzeitraum, in dem Bewertungen gesammelt werden.

des Systemes gezeigt werden, wie eine Social-Navigation-Anwendung von der Integration einer Identitätsinfrastruktur profitieren könne. Das System ist als Proof-of-Concept-Implementierung zu sehen. Viele Aspekte, die beim produktiven Einsatz einer solchen Anwendung zwingend zu berücksichtigen wären, beispielsweise Fragen der Performanz, Skalierbarkeit, Sicherheit und Verfügbarkeit wurden nicht oder nur am Rande bearbeitet. Auch wenn es die realisierte Anwendung gestattet, Web-Seiten zu annotieren, so ist der realisierte Prototyp doch in erster Line ein Empfehlungssystem. Funktionen zur Annotation von Dokumenten wurden nur deshalb verwirklicht, da ein Interesse daran bestand, die aus der Kombination mit einem Empfehlungssystem resultierenden Möglichkeiten zu studieren.

### 1.3. Struktur der Arbeit

Die thematischen Grundlagen der Arbeit werden in Kapitel 2 aufgeführt. Es werden die theoretischen Hintergründe von Software zur Unterstützung kollaborativer Arbeit betrachtet. Ein Empfehlungs- und Annotationssystem ist ein Medium, über das Menschen miteinander interagieren können. Unter diesem Gesichtspunkt besteht die Notwendigkeit der Betrachtung einiger, in Bezug auf die theoretische Fundierung von Social Navigation relevanter, Aspekte rechnergestützter Interaktion zwischen Menschen. Das Projekt onefC, welches die Entwicklung einer Identitätsinfrastruktur zum Ziel hat, macht den thematischen Rahmen dieser Arbeit aus. Zudem ist der onefC-Identitätsmanager eine essentiell wichtige Komponente in der Architektur des CoInternet-Systems. Das Scone-Framework bildet die Grundlage der clientseitigen Komponenten des CoInternet-Systems. Überdies beinhaltet Scone Funktionen zur Verfolgung der Benutzeraktivitäten. Insofern ist das Scone-Framework auch unter dem für Empfehlungssysteme relevanten Aspekt des Aufbaus von Benutzerprofilen von Bedeutung. Für einen Teil der Benutzeroberfläche wurde auf das Konzept des HyperScouts zurückgegriffen.

In Kapitel 3 werden einige der Herausforderungen, mit denen Benutzer des World Wide Webs bei ihren oftmals zeitraubenden Recherchen konfrontiert werden, betrachtet. Die Navigation im WWW überfordert einige Nutzer kognitiv. Zudem ist es mit dem Wachstum des World Wide Webs in vielen Situationen schwieriger geworden, relevante Inhalte zu finden, nicht zuletzt auch wegen Unzulänglichkeiten der automatischen Selektion von Informationen durch Suchmaschinen. Naturgemäß bestehen Informationsasymmetrien zwischen Personen, die motiviert durch ein Informationsbedürfnis im Web recherchieren und den Autoren der von ihnen eingesehenen Dokumente. Folglich fällt es schwer, die Qualität der gebotenen Informationen einzuschätzen und eine Entscheidung zu treffen, ob einem bestimmten Informationsanbieter zu vertrauen sei. In diesem Kapitel werden einige Strategien zur Unterstützung von Anwendern bei der Navigation im Web vorgestellt und motiviert, warum es sinnvoll sein kann, das Web um Funktionen

zur Kollaboration zwischen Nutzern zu erweitern.

Im 4. Kapitel wird ein Überblick über verschiedene Formen der kollaborativen Navigation in Informationsräumen gegeben. Dabei wird zunächst einmal präzisiert, was im Kontext dieser Arbeit unter Navigation in Informationsräumen verstanden wird. Es werden Definitionsansätze von Social Navigation vorgestellt, die eine vereinheitlichende Sichtweise auf die verschiedenen Formen von sozialem Navigieren bieten. Zudem wird ein Klassifizierungssystem für Social-Navigation-Anwendungen präsentiert, so daß eine Grundlage für den Vergleich unterschiedlicher Ansätze geschaffen wird. Ferner wird deutlich gemacht, warum soziales Navigieren im Web nur unzureichend unterstützt wird. In diesem Kapitel wird versucht zu konkretisieren, worin der Wert von Social Navigation liegen kann. Voraussetzungen für und Randbedingungen von Social Navigation werden ebenso analysiert, wie mit dem Konzept Social Navigation einhergehende Probleme. Hinsichtlich des Ziels, am Beispiel des Empfehlungs- und Annotationssystems die Auswirkungen der Integration einer Identitätsinfrastruktur zu untersuchen, ist in diesem Zusammenhang insbesondere die Bedeutung von Identitätskonzepten in Social-Navigation-Anwendungen von Interesse.

Empfehlungssysteme sind Anwendungen, die Benutzer bei der Auswahl von Objekten aus einer Menge von Alternativen unterstützen. In Kapitel 5 werden die verschiedenen Typen von Empfehlungssystemen beschrieben. Nicht-kollaborative Techniken werden kollaborativen Ansätzen zur Realisierung von Empfehlungssystemen gegenübergestellt. Der Ansatz, durch ein Empfehlungssystem kollaboratives Verhalten bei der Auswahl von Objekten in einem Informationsraum zu ermöglichen, wird als Collaborative Filtering bezeichnet. Die Idee des Collaborative Filtering wurde bei der Realisierung des Empfehlungssystems für Web-Seiten aufgegriffen und weiterentwickelt. Die Erörterung der Schwächen von Collaborative-Filtering-Systemen ist für den Fortgang dieser Arbeit von besonderer Bedeutung, denn diese sind teilweise auf Unzulänglichkeiten der Benutzermodellierung zurückzuführen. Ein Ansatzpunkt, an dem sich durch Einbeziehung einer Identitätsinfrastruktur qualitative Verbesserungen der Anwendungen erzielen lassen könnten.

Die Leistung von Social-Navigation-Anwendungen, insbesondere von Collaborative-Filtering-Systemen, wird letztendlich aus der Gemeinschaft von Nutzern heraus erbracht, wobei die Leistung des Systems primär darin besteht, Bewertungen zu aggregieren und zu kanalisieren. Bezogen auf die Betrachtung der Auswirkungen einer Identitätsinfrastruktur ist es von Belang, welche Bedeutung die Wahrnehmung der Benutzer untereinander und die Wahrnehmung der Gemeinschaft auf das Verhalten der Anwender und deren Akzeptanz des Systems haben.

Kapitel 6 stellt einen Bezug von dieser Arbeit zu anderen Projekten her. Es werden exemplarische Anwendungen vorgestellt, welche allesamt zum Ziel haben, die Kollaboration bei der Navigation in einem Informationsraum zu fördern. Es werden bewußt

sehr unterschiedliche Systeme präsentiert, um zu verdeutlichen, welche Bandbreite von Lösungsansätzen es bei der Unterstützung kollaborativer Navigation in einem Informationsraum gibt.

Basierend auf der theoretischen Beschäftigung mit den Konzepten Social Navigation im Allgemeinen und Collaborative-Filtering-Techniken im Besonderen erfolgt in Kapitel 7 die Formulierung von Forschungsfragen. Vor allem wird eine Zuspitzung der zu Beginn dieser Arbeit gestellten Frage nach den Auswirkungen der Integration einer Identitätsinfrastruktur auf eine kollaborative Internetanwendung vorgenommen. Des Weiteren werden die Anforderungen an das zu erstellende System formuliert.

Kapitel 8 ist der Präsentation des in dieser Arbeit entwickelten Prototyps gewidmet. Es wird die Konzeption des CoInternet-Systemes vorgestellt, die von dem Leitgedanken geprägt war, die Leistung eines Empfehlungssystems durch Integration einer Identitätsinfrastruktur zu verbessern. Die Konstruktion der Anwendung wird in diesem Kapitel dokumentiert und die Ergebnisse einer ersten Evaluation werden präsentiert. Den Abschluß dieses Kapitels bildet die zusammenfassende Darstellung der Auswirkungen der Integration einer Identitätsinfrastruktur.

Kapitel 9 bildet mit der Zusammenfassung der Ergebnisse, Ausführungen über die im Rahmen der Arbeit unbeantwortet gebliebenen Fragen und einem Ausblick auf weitere Entwicklungsmöglichkeiten des CoInternet-Systems den Abschluß der Arbeit.

### 1.4. Zur Verwendung englischer Begriffe

In dieser Arbeit wurde weitestgehend darauf verzichtet, englische Fachbegriffe zu übersetzen. Einerseits ist der Großteil der zum Thema veröffentlichten Literatur englischsprachig, die Verwendung von Übersetzungen könnte folglich den Rückgriff auf vertiefende Literatur erschweren, andererseits werden in den wenigen deutschsprachigen Veröffentlichungen ebenfalls meist die englischen Originalbegriffe verwandt. Somit würden die übersetzten Begriffe, sofern eine sinnerhaltende Übersetzung überhaupt möglich ist, meist sehr ungewohnt klingen.

## 2. Grundlagen

In diesem Kapitel werden die thematischen Grundlagen der Diplomarbeit vorgestellt. Es werden zum einen theoretische Fundamente der Arbeit vorgestellt, zum anderen werden mit *Scone* und *onefC* zwei Projekte präsentiert, die den fachlichen Kontext dieser Diplomarbeit bilden. Auf eine Beschreibung der für die Implementierung des Prototyps verwendeten Techniken wird an dieser Stelle verzichtet. Die Auseinandersetzung mit den Techniken steht nicht im Zentrum dieser Arbeit. Zum besseren Verständnis wird entsprechendes Hintergrundwissen im Abschnitt über die Konstruktion des Prototyps hinzugefügt. Erfahrung im Umgang mit dem Web und Wissen über die technischen Hintergründe wird in dieser Arbeit angenommen. Auf eine Darstellung im Rahmen dieser Arbeit wurde jedoch verzichtet, da angesichts der enormen Verbreitung des Webs diesbezüglich Kenntnisse vorausgesetzt werden.

### 2.1. Computer Supported Cooperative Work

Das Teilgebiet *Computer Supported Cooperative Work (CSCW)* der Informatik beschäftigt sich mit der Unterstützung von Gruppenarbeit durch Computersysteme. Applikationen zur kollaborativen Nutzung des World Wide Webs sind in diesem fachlichen Rahmen zu betrachten. Nachfolgend werden deshalb die Grundlagen von CSCW Systemen, beginnend mit einer Definition, dargestellt. Anschließend werden verschiedene Modelle der Klassifikation von CSCW-Systemen vorgestellt, die dann als Grundlage für die Einordnung des entwickelten Prototyps und die Abgrenzung von anderen Ansätzen dienen können. Mit der Verbreitung des Internets und der damit verbundenen Möglichkeit der kostengünstigen globalen Kommunikation ist der Aufbau von sogenannten Online-Communities praktikabel geworden. Abschnitt 2.1.4 differenziert zwischen Communities, Gruppen und Teams. Nachfolgend werden dann Systeme zur Unterstützung von Communities betrachtet.

#### 2.1.1. Begriffsbestimmung

Teufel et al. [TSMB95] definieren Computer Supported Cooperative Work folgendermaßen:

Computer Supported Cooperative Work (CSCW) ist die Bezeichnung des Forschungsgebietes, welches auf interdisziplinärer Basis untersucht, wie Individuen in Arbeitsgruppen oder Teams zusammenarbeiten und wie sie dabei durch Informations- und Kommunikationstechnologie unterstützt werden können [TSMB95].

Diese Definition ist sehr weit gefaßt. Die Zusammenarbeit in Arbeitsgruppen und Teams kann sehr verschiedene Formen annehmen. Ebenso ist hinsichtlich der Betrachtung der kollaborativen Nutzung des World Wide Webs zu konkretisieren, wie Arbeitsgruppen und Teams von lose im Netz interagierenden Personengruppen abzugrenzen sind. Systeme, die auf den theoretischen Grundlagen des Forschungsgebiets CSCW basieren werden gemeinhin als *Groupware* bezeichnet. Johansen [Joh88] definiert diesen Begriff wie folgt:

Groupware is a generic term for specialised computer aids that are designed for the use of collaborative work groups. Typically, these groups are small project-oriented teams that have important tasks and tight deadlines. Groupware can involve software, hardware, services and/or group process support [Joh88].

### 2.1.2. Formen der Interaktion

Baier unterscheidet, wie in Abbildung 2.1 dargestellt ist, vier Formen der Zusammenarbeit [Bai89], die sich hinsichtlich der Intensität der Kommunikationsprozesse innerhalb der Gruppe unterscheiden. Charakteristisch für die *Kooperation* zwischen Personen ist die starke Übereinstimmung in den Zielen und die gemeinsame Verantwortung für das Arbeitsergebnis. Hiervon abzugrenzen ist, obwohl die Übergänge fließend sind, die *Kollaboration*. Zwischen kollaborierenden Personen besteht nur eine teilweise Übereinstimmung in den Zielen. Ferner sind Einzelbeiträge im Ergebnis erkennbar. Haben die beteiligten Personen kein gemeinsames Ziel, aber übereinstimmende Interessen so spricht man von *Koordination*, sofern organisatorische Zusammengehörigkeit die Abstimmung der Tätigkeiten erfordert. Die hinsichtlich der Intensität der Zusammenarbeit schwächste Form der gemeinsamen Arbeit ist die *Information*. Informationsanbieter müssen nicht notwendigerweise wissen, wer die Empfänger der Information sind. Die Kommunikation zwischen den Gruppenmitgliedern verläuft meist unidirektional.

### 2.1.3. Klassifikation von CSCW-Systemen

CSCW-Systeme lassen sich nach sehr unterschiedlichen Kriterien klassifizieren. In diesem Abschnitt werden drei der bekanntesten Klassifizierungsmodelle, die Einordnung in die

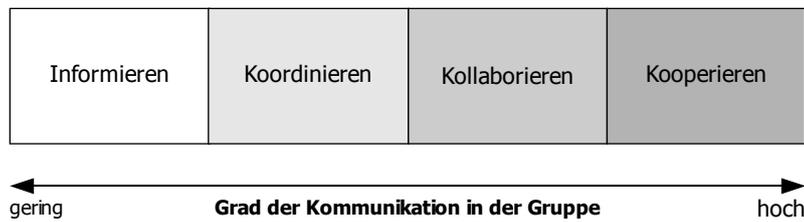


Abbildung 2.1.: Intensität des Informationsflusses in Arbeitsgruppen (Quelle: [BS98])

Raum-Zeit-Matrix, die Klassifizierung nach funktionellen Merkmalen und die Klassifizierung hinsichtlich Kommunikations-, Koordinations- und Kooperationsunterstützung, vorgestellt.

### Klassifizierung nach Raum und Zeit

Eine Taxonomie zur Klassifikation von CSCW-Systemen ist die zweidimensionale *Raume-Zeit-Matrix* [BS98]. Systeme werden nach der räumlichen und zeitlichen Verteilung der beteiligten Gruppenmitglieder differenziert. Bezüglich der räumlichen Verteilung der Gruppenmitglieder wird unterschieden, ob sich die Benutzer am gleichen Ort befinden oder räumliche verteilte Gruppenmitglieder miteinander interagieren. Die Bandbreite der Verteilung reicht dabei von in verschiedenen Räumen, aber im gleichen Flur arbeitenden Arbeitsgruppen bis hin zur Verteilung über mehrere, möglicherweise in unterschiedlichen Zeitzonen gelegenen Städten. Bezüglich der zeitlichen Verteilung ist zwischen der Kommunikation in Realzeit, in diesem Fall spricht man von synchroner Gruppenarbeit, und asynchroner Gruppenarbeit zu unterscheiden.

### Funktionelle Klassifizierung

Ein weiterer Ansatz zur Systematisierung von CSCW-Applikationen ist die Einordnung in anwendungsorientierte Funktionsklassen. Ellis et al. [EGR91] unterscheiden beispielsweise Nachrichtensysteme, Gruppeneeditoren, elektronische Sitzungsräume, Rechnerkonferenzen, Agentensysteme und Koordinationssysteme. Die Einordnung konkreter Anwendungen ist dabei nicht eindeutig, da Anwendungen Funktionalitäten mehrerer Klassen integrieren können.

*Nachrichtensysteme* bieten den asynchronen Austausch textueller oder multimedialer Nachrichten zwischen Gruppenmitgliedern. Durch Regeln und Skripte läßt sich die Verarbeitung von Nachrichten teilweise automatisieren.

## 2. Grundlagen

---

*Gruppeneditoren* unterstützen Benutzer beim gemeinsamen Verfassen eines Dokuments oder bei der gemeinsamen Programmierung von Softwaresystemen. Derartige Editoren verdeutlichen dem Benutzer die Aktivitäten der Kooperationspartner. Dabei ist zwischen Gruppeneditoren mit asynchroner Benachrichtigungsfunktion und Realzeiteditoren, welche die vorgenommenen Änderungen aller Benutzer in Echtzeit darstellen, zu unterscheiden.

*Elektronische Sitzungsräume* bieten Rechnerunterstützung bei der Durchführung von face-to-face-Sitzungen. Derartige Systeme dienen oftmals zur Findung von Gruppenentscheidungen bzw. der gemeinsamen Strukturierung von Problemen. Sie werden auch als *Group Support Systems (GSS)* bezeichnet.

*Konferenzsysteme* gestatten im Gegensatz zu den elektronischen Sitzungsräumen die Durchführung von Konferenzen räumlich oder zeitlich verteilter Teammitglieder. Die Bandbreite der Anwendungen dieser Kategorie reicht von Nicht-Realzeitkonferenzsystemen, deren Funktionsweise mit Newsgroups vergleichbar ist, bis hin zu Desktopkonferenzsysteme, die es Benutzern gestatten, von ihren mit Kamera und Mikrofon ausgestatteten Arbeitsplatzrechnern in Realzeit Videokonferenzen abzuhalten.

*Gemeinsame Informationsräume* dienen der Bearbeitung und konsistenten Verwaltung gemeinsamer Informationen. Sie erlauben die persistente Speicherung der Daten und stellen geeignete Zugriffsmechanismen bereit. Informationsräume sind vielfach als Hypertext-Systeme organisiert, bei denen mittels Hypertext-Links Zugriffspfade zwischen den Dokumenten definiert werden können.

Übernehmen Computerprogramme die Funktionen von Sitzungsteilnehmern oder agieren als aktive Teilnehmer der Gruppenarbeit, so bezeichnet man derartige CSCW-Applikationen als *Agentensysteme*. In rechnerunterstützten Konferenzen wäre es z. B. möglich, daß ein Softwareagent den Verlauf der Diskussion mitprotokolliert.

*Workflow-Management/Koordinationssysteme* unterstützen Arbeitsgruppen bei der Koordination asynchroner Tätigkeiten. Arbeitsabläufe können auf entsprechende Systeme abgebildet werden, so daß das System nach Abschluß einer Teilaufgabe automatisch die für den folgenden Bearbeitungsschritt zuständigen Gruppenmitglieder benachrichtigen kann.

### **Klassifizierung nach dem 3K-Modell**

Je nach Grad der Zusammenarbeit der Gruppenmitglieder kann zwischen Kommunikation, Koordination und Kooperation unterschieden werden. Bei der Kommunikation liegt der Schwerpunkt der Zusammenarbeit auf dem Informationsaustausch. Koordination dient der Abstimmung voneinander abhängiger Aufgaben bzw. der gemeinsamen Nutzung von Ressourcen. Dient die Zusammenarbeit der Erreichung gemeinsamer Ziele,

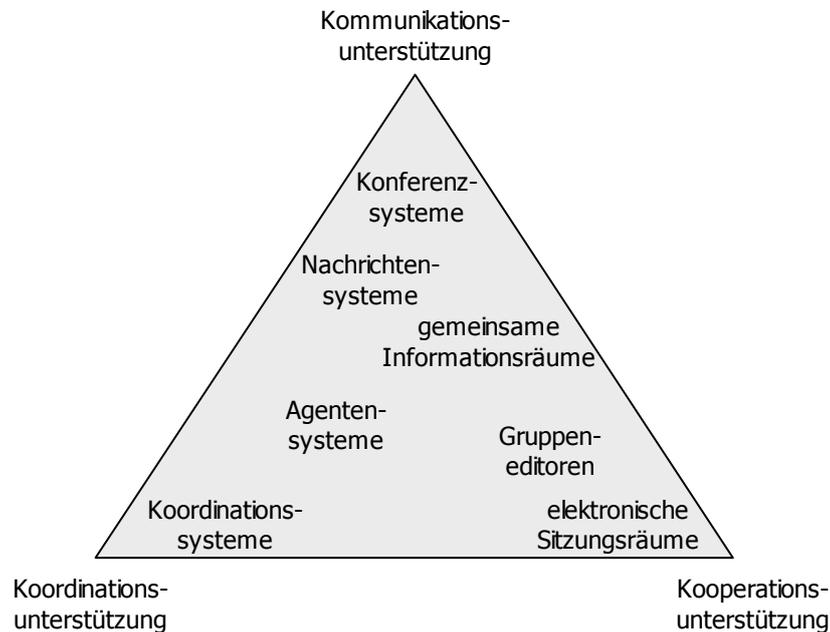


Abbildung 2.2.: Klassifizierung nach dem 3K-Modell (Quelle: [BS98])

so spricht man von Kooperation. Teufel et al. [TSMB95] differenzieren CSCW- Systeme bezüglich des Grades der Kommunikations-, Kollaborations- und Kooperationsunterstützung. Im sogenannten *3K-Modell* werden die Anwendungstypen der funktionellen Klassifizierung auf diese Weise in einem Dreieck angeordnet (siehe Abbildung 2.2).

#### 2.1.4. Teams, Gruppen und Communities

Die in Abschnitt 2.1.1 zitierte Definition von CSCW bezieht sich auf die Zusammenarbeit innerhalb von Arbeitsgruppen und Teams. Durch die starke Verbreitung des Internets wurde die Rechnerunterstützung von losen, organisationsübergreifenden Gruppen, sogenannten *Online-Communities*, möglich. Die Charakteristika dieser Communities unterscheiden sich von Arbeitsgruppen und Teams. Schlichter et al. [SKX98] differenzieren zwischen *Gruppen*, *Teams* und *Communities*. Unterscheidungskriterien sind die Größe der Personenmenge, der Interaktionsgrad, die Orientierung der beteiligten Personen, das Vorhandensein von gemeinsamen Arbeitsobjekten und die Existenz persönlicher Beziehungen.

Die Mitglieder einer *Community* teilen ein gemeinsames Interesse oder eine sonstige Gemeinsamkeit, wie z.B. den gleichen Wohnort oder die Beschäftigung in der gleichen

Branche. Sie kennen sich jedoch nicht notwendigerweise untereinander und interagieren nicht auf persönlicher Basis. Mitglieder einer Community haben in der Regel keine übereinstimmenden Arbeitsziele.

Eine sehr allgemeine Definition von Communities liefert Mynatt [MAIO97]:

...social grouping which exhibit in varying degrees: shared spatial relations, social conventions, a sense of membership and boundaries, and an ongoing rhythm of social interaction. (Zitat: [MAIO97])

Koch betrachtet zwei Typen von Communities, die sich hinsichtlich der Art der Gemeinsamkeit unterscheiden [Koc01a]:

- *Communities of Interest* Mengen von Personen mit einem gemeinsamen Interesse, z.B. Programmierer einer bestimmten Programmiersprache oder Fans eines Fußballvereins.
- *Communities of Practice* Personen, die als Gemeinsamkeit die Beteiligung an einem Prozess oder die Anwesenheit an einem gemeinsamen Ort haben. Als Beispiel wäre die Zuhörerschaft einer Vorlesung zu nennen.

Gruppen, die in der Regel deutlich weniger Personen als Communities aufweisen, sind durch die persönliche Bekanntschaft der Mitglieder gekennzeichnet. In Abgrenzung zu den Teams weisen Gruppen jedoch kein gemeinsames Ziel auf und die Gruppenmitglieder haben keine gemeinsamen Aufgaben. Die Interaktion innerhalb von Gruppen ist demnach schwächer ausgeprägt als innerhalb von Teams. Die Mitglieder von Teams beschäftigen sich mit einer gemeinsamen Aufgabe und teilen dabei die gleichen Zielvorstellungen. Teammitglieder arbeiten an gemeinsamen Artefakten, demgegenüber kommt es innerhalb von Communities nur zum gelegentlichen Austausch von persönlichen Informationen. Teams sind meist in einen organisatorischen Kontext eingebettet. Charakteristisch für sie ist ein festgelegtes Innenverhältnis und eine administrative Festschreibung.

Die Übergänge zwischen den verschiedenen Formen von Gruppierungen sind fließend. Ferner können sich innerhalb von Communities Teams herausbilden.

### 2.1.5. Community Support Systems

Wie im vorhergehenden Abschnitt aufgeführt wurde, ist durch den Fortschritt des Internets der Aufbau sogenannter Online-Communities realisierbar geworden. Communities

unterscheiden sich grundlegend von Teams. Während in Communities gemeinsame Interessen die Motivation für eine lose, keinem Gruppenprozeß untergeordnete Interaktion sind, steht bei den Teams die auf ein gemeinsames Ziel ausgerichtete Kooperation im Vordergrund. In Anlehnung an den Begriff *Groupware*, der Software zur Unterstützung von Arbeitsgruppen und Teams bezeichnet (siehe Abschnitt 2.1.1), stehen die Ausdrücke *Communityware* oder *Community Support Systems* für Programme zur Unterstützung von Communities.

Koch [Koc01a] definiert *Communityware* folgendermaßen:

Systeme und Infrastruktur, welche sich der Unterstützung von losen Gruppen und dem Aufbau und der Nutzung von in diesen Gruppen vorhandenem Synergiepotential widmen [Koc01a].

Auch im betrieblichen Kontext werden Community-Support-Systeme eingesetzt, um die lose Kommunikation über Abteilungs- und Projektgrenzen hinweg zu fördern. Eine wichtige Anwendung ist dabei das *Knowledge Management*. Wissen wird zunehmend als Produkt der (indirekten) Kommunikation gesehen [Dav98].

Bezüglich der Funktionalität von Community-Support-Systemen ist zwischen zwei Konzepten zu unterscheiden. Zum einen werden mittels derartiger Software virtuelle Treffpunkte bzw. Kommunikations- und Interaktionsplattformen realisiert, zum anderen wird versucht, softwaregestützt die Beziehungen innerhalb der Community aufzudecken und zu visualisieren.

Nachfolgend werden die grundlegenden Funktionalitäten von Communityware dargestellt [Koc01a]:

### **Matchmaking**

Die Unterstützung der Kontaktaufnahme zwischen den Mitgliedern der Community ist eine der zentralen Funktionen von Community Support Systems. Community-Support-Systeme versuchen, z. B. basierend auf dem Vergleich von Benutzerprofilen, Benutzer mit ähnlichen Interessen oder Experten auf einem bestimmten Gebiet zu ermitteln. Ziel ist es dabei, das Auffinden möglicher Kooperationspartner zu vereinfachen.

Die Kenntnis der Beziehungen innerhalb einer Community läßt sich zur (semi-) automatischen Informationsfilterung einsetzen. Grundidee der *Collaborative-Filtering*-Algorithmen ist es, automatisch Community-Mitglieder mit ähnlichen Präferenzen zu ermitteln und so aus der Vielzahl subjektiver Bewertungen für jeden Benutzer eine individuelle Selektion von Inhalten eines Suchraumes vorzunehmen. Derartige Verfahren stellen ein Instrument dar, um mit der „Informationsflut“ besser zurechtzukommen. Durch

die gemeinsame Bewertung von Inhalten wird innerhalb der Community ein gemeinsames Wissen aufgebaut. Folglich wird dieser Aspekt von Community-Support-Systemen auch als „Community-basiertes Wissensmanagement“ bezeichnet.

### **Informationsaustausch**

Matchmaking-Funktionen ermöglichen es, innerhalb der Community Kooperationspartner zu finden, mit denen dann, vermittelt durch das System, direkt interagiert werden kann. Ergänzend beinhalten typische Community-Support-Systeme Funktionen zur indirekten Kommunikation. Auf diesem Wege können sich Personen helfen, ohne sich direkt zu kennen. Von Nutzern der Community angebotene Informationen können anderen Benutzern als nützliche Informationsquelle bei ihren individuellen Aufgaben dienen. Ein typisches Beispiel ist der Austausch von Informationen zwischen den Benutzern web-basierter Diskussionsforen zu technischen Themen, welche teilweise von Produktanbietern zur Kundenbindung und zur Entlastung der eigenen Support-Abteilung eingesetzt werden.

### **Awareness**

Unter Awareness wird in diesem Zusammenhang die Information über die Präsenz und Aktivität anderer Benutzer verstanden. Awareness ist ein zentraler Aspekt bei der Gestaltung von Group- und Communityware. Gerade hinsichtlich der kollaborativen Nutzung des World Wide Webs ist dieser Gesichtspunkt zu berücksichtigen. Eine ausführlichere Darstellung wird in Abschnitt 2.1.6 gegeben.

Verschiedene Typen von Community-Support-Systemen lassen sich unterscheiden [Koc01a]. Dabei werden die vorstehend beschriebenen Funktionsmerkmale mit unterschiedlichen Schwerpunkten unterstützt.

*News- und Chatsysteme* fördern den synchronen und asynchronen Informationsaustausch zwischen den Mitgliedern einer Community.

*Buddy-Systeme* liefern Awareness-Informationen über die Mitglieder einer Community oder eines Teams. So kann ein Benutzer beispielsweise sehen, ob Freunde oder Kollegen online sind und ob diese gerade kontaktiert werden möchten.

*Recommender-Systeme* versuchen, basierend auf der Analyse der Interessen eines Benutzers, individuelle Vorschläge zu interessierenden Objekten zu unterbreiten. Auf *Collaborative-Filtering*-Algorithmen basierende Systeme beinhalten Matchmaking-Funktionen, um Benutzer mit ähnlichen Präferenzen zu ermitteln, und generieren

dann, basierend auf der Meinung der „ähnlichen“ Benutzer, individuelle Empfehlungen. Recommender-Systeme stehen im Zentrum dieser Arbeit und werden ausführlich in Kapitel 5 behandelt.

*Matchmaking-Systeme* erstellen, basierend auf der Kenntnis der Interessen eines Benutzers und der Beziehungen zwischen den Mitgliedern einer Community, Vorschläge, welche Benutzer als Kooperationspartner in Frage kommen.

### 2.1.6. Awareness

Einer der zentralen Aspekte beim Design von Groupware und Community Support Systems ist *Awareness*<sup>1</sup>. In räumlich verteilten Arbeitsgruppen sind für den Einzelnen viele Aspekte des Gruppenprozesses ohne Softwareunterstützung nicht wahrnehmbar. CSCW-Systeme, die sich auf den Austausch von Nachrichten und die gemeinsame Arbeit an Artefakten beschränken, sind nur bedingt zur Unterstützung räumlich verteilter Arbeitsgruppen geeignet, da sich in derartigen Arbeitsgruppen nur dann ein Gruppenbewußtsein einstellt, wenn Meta-Informationen über Anwesenheit und Aktivität von Gruppenmitgliedern ausgetauscht werden. Bly et al. weisen darauf hin, daß Rechnerunterstützung kooperativer Arbeit alle Aspekte des Gruppenprozesses berücksichtigen sollte:

Most tools in computer-supported cooperative work (CSCW) are devoted to the computational support of task-specific activities, but support for cooperative work is not complete without considering all aspects of the work group process. When groups are geographically distributed, it is particularly important not to neglect the need for informal interactions, spontaneous conversations, and even general awareness of people and events at other sites [BHI93].

Unter dem Begriff Awareness werden unterschiedliche Ansätze zur Erweiterung von Group- oder Communityware um Meta-Informationen verstanden, welche die Nutzer des Systems über Zusammensetzung und Zustand der Gruppe in Kenntnis setzen bzw. ihnen einen Eindruck von der Entwicklung des Gruppenprozesses oder den Veränderungen in der gemeinsamen Arbeitsumgebung verschaffen. Die Bedeutung des Begriffs Awareness weist viele Facetten auf. Liechti differenziert zwischen vier nicht-disjunkten Formen von Awareness [Lie00]:

---

<sup>1</sup>Es wird im Folgenden der englischsprachige Begriff verwendet, da es keinen äquivalenten deutschen Begriff gibt.

**Group Awareness** Dourish und Bly definieren Group Awareness wie folgt:

Group awareness is an understanding of the activities of other which provides a context for your own activity [DB92].

CSCW-Applikationen mit Group Awareness Unterstützung vermitteln den Benutzern einen Eindruck von Präsenz und Aktivität anderer Team-Mitglieder und fördern somit das Gruppenbewußtsein. Basierend auf diesen Informationen können die Benutzer ihre eigenen Aktivitäten planen.

**Workspace Awareness** Feingranulare Informationen über den Zustand einer gemeinsam genutzten Arbeitsumgebung werden mit Workspace Awareness bezeichnet. In synchronen CSCW-Architekturen, die ein gemeinsames Editieren von Dokumenten in Multi-User-Editoren erlauben, wird z. B. mittels Telepointer oder Multi-User Scrollbar die Position übermittelt, an der andere Nutzer simultan Veränderungen durchführen. In asynchronen CSCW-Architekturen ist es sinnvoll, Veränderungen an gemeinsam genutzten Dokumenten zu dokumentieren.

**Contextual Awareness** Awareness-Informationen zum über die unmittelbare Arbeitsumgebung hinausgehenden Kontext der eigenen Arbeit werden mit dem Ausdruck Contextual Awareness bezeichnet. Dem Benutzer wird versucht, einen Eindruck von der Situation zu vermitteln, in der er sich befindet, so daß eine Anpassung der eigenen Handlungen möglich wird. Insbesondere zur spontanen Kontaktaufnahme ist Contextual Awareness von Bedeutung.

**Peripheral Awareness** Der Ausdruck Peripheral Awareness bezieht sich auf die Art, wie Awareness-Informationen dargestellt werden. Kann der Nutzer Awareness-Informationen aufnehmen, ohne den Aufmerksamkeitsfokus von der eigenen Arbeit abzuwenden, so spricht man von Peripheral Awareness.

Eine weitere Einteilung von Awareness-Informationen wird in [SJP99] dargestellt. Die vorgeschlagenen Kategorien überschneiden sich teilweise mit den vorstehend genannten, beinhalten jedoch einige ergänzende Aspekte:

**Activity Awareness** bezeichnet die Vermittlung von Informationen über die projektbezogenen Aktivitäten anderer Gruppenmitglieder. Systeme zur synchronen Kollaboration versuchen, die aktuelle Tätigkeit der beteiligten Personen darzustellen. So ist beispielsweise ersichtlich, welcher Benutzer gerade welches Dokument editiert. Auch Systeme,

die eine asynchrone Zusammenarbeit fördern, präsentieren teilweise Activity Awareness-Informationen. Dies kann z. B. in Form der Historie der an einem Dokument durchgeführten Änderungen geschehen.

**Availability Awareness** Viele Groupware Applikationen stellen Informationen zur Verfügbarkeit von Benutzern dar, so daß spontane Kommunikation innerhalb der Gruppe möglich wird. Dabei werden teilweise über Präsenzinformationen hinausgehende Angaben gemacht, welche in [TSS96] mit *Social Awareness* beschrieben werden. So wird u. a. dargestellt, ob Benutzer gerade beschäftigt sind oder nicht gestört werden möchten.

**Process Awareness** bezeichnet den Versuch, den Mitgliedern einer Arbeitsgruppe einen Überblick über den Arbeitsprozess zu geben. Dadurch wird für den einzelnen ersichtlich, wie sich der eigene Arbeitsbeitrag in den Gesamtkontext einfügt, in welchem Status ein Projekt ist und welche Arbeitsschritte im Folgenden durchzuführen sind.

**Perspective Awareness** Um die Handlungen anderer Team-Mitglieder antizipieren zu können, ist es nicht nur erforderlich, über deren vergangene Aktivitäten informiert zu sein, sondern auch wünschenswert, Informationen zu den Einschätzungen der Team-Mitglieder und deren Hintergrundwissen zu haben.

**Environmental Awareness** Auch Ereignisse außerhalb der unmittelbaren Arbeitsumgebung eines Teams können Implikationen auf die gemeinsame Arbeit haben und sollten den Team-Mitgliedern vermittelt werden.

Awareness ist eine Voraussetzung für synchrone Interaktion in Arbeitsgruppen und Communities. Insbesondere die spontane, informelle Kommunikation wird durch die Integration von Awareness-Informationen in Group- oder Communityware erst praktikabel. Innerhalb von Teams kommt Awareness eine besondere Bedeutung zu, da auf diesem Wege eine implizite Koordination von Arbeitsschritten möglich wird und so die zur Koordination der Tätigkeiten notwendige Kommunikation reduziert werden kann. Durch Awareness wird es letztendlich erst möglich, einem Kreis von hauptsächlich über eine gemeinsame Softwareplattform interagierenden Personen wirklich das Bewußtsein zu vermitteln, in einer Gruppe gemeinsam zu arbeiten oder Teil einer Community zu sein.

Beim Entwurf von Awareness-Systemen sind eine Reihe von Fragestellungen zu beachten. Um eine Informationsüberlastung zu vermeiden, ist zu klären, welche Teilmenge der Benutzer über welche Ereignisse oder Zustände zu unterrichten ist. Des weiteren muß beachtet werden, daß der Betrieb von Awareness-Systemen das Risiko einer Verletzung der Privatsphäre einzelner Benutzer mit sich bringt. Die Software sollte also so gestaltet

werden, daß eine unerwünschte Überwachung anderer Gruppenmitglieder erschwert wird und daß jeder Benutzer den Umfang der über seine Aktivitäten übermittelten Awareness-Informationen steuern kann.

### 2.2. Computer als soziales Medium

Mit zunehmender Verbreitung hat sich das Internet zu einem „Ort“ entwickelt, in dem Personen sozial miteinander interagieren. „Computer und Netze beginnen als soziale Umgebung genutzt zu werden, in der Menschen sich selbst darstellen, sich mit anderen treffen, Informationen austauschen, Spiele spielen, Geschäfte machen, gemeinsam auf Informationssuche gehen und vieles andere mehr [HPPB01].“ In diesem Abschnitt sollen einige Aspekte der Betrachtung vernetzter Computers als soziales Medium erläutert werden, die im Hinblick auf die kollaborative Nutzung des World Wide Webs von besonderer Bedeutung sind. Die Benutzer nehmen in der virtuellen Welt eine Identität an und machen Erfahrungen im Umgang miteinander, welche ihren Ausdruck in der Reputation der „Bewohner“ des Netzes finden. Gerade im Kontext elektronisch abgewickelter Geschäfte sehen sich die Anwender immer wieder mit der Frage konfrontiert, ob sie einem anderen, möglicherweise gänzlich unbekanntem Benutzer vertrauen können. Die Begriffe Identität, Reputation und Vertrauen sind Gegenstand der nachfolgenden Betrachtung.

#### 2.2.1. Identität

Der Begriff der Identität ist sehr vielschichtig und läßt sich nicht in einer umfassend gültigen Definition fassen. In [Kun03] wird versucht, sich diesem Begriff aus philosophischer, psychologischer und technischer Sichtweise zu nähern. Die Definition der Identität als feinste denkbare Äquivalenzrelation, bei der jedes Objekt ausschließlich zu sich selbst in Relation steht, ist vom Standpunkt der Philosophie her gebräuchlich und auch im Kontext technischer Anwendungen, welche die Identifizierung eines Objektes aus einer Menge von Objekten erfordern, angemessen. Eine klassische psychologische Definition der Identität liefert Döring:

Das Bewußtsein, sich von anderen Menschen zu unterscheiden (Individualität) sowie über die Zeit (Kontinuität) und verschiedene Situationen (Konsistenz) hinweg - auch für die Umwelt erkennbar - dieselbe Person zu bleiben. [Dör99]

Dieses traditionelle Verständnis von Identität geht von einer einheitlichen Identität aus, die sich im Laufe der Zeit entwickelt und verfestigt. Moderne Ansätze betonen hingegen, daß die Identität einer Person Gegenstand einer fortwährenden Entwicklung ist und

bei der Übernahme unterschiedlicher Rollen verschiedene Ausprägungen erfahren kann. Identität wird dabei nicht als ein konsistentes Ganzes gesehen, sondern vielmehr als ein Patchwork unterschiedlicher Teilidentitäten, welches einem dynamischen Entwicklungsprozeß unterliegt [Kun03].

Neben der individuellen Identität von Personen können auch Gruppen eine Identität ausbilden. Man spricht in diesem Fall von *sozialen Identitäten*, welche eine Menge von für die Gruppenmitglieder charakteristischen Eigenschaften umfassen. Dabei ist das Gefühl der Zugehörigkeit zu einer Gruppe Teil der Identität der Mitglieder. Dies impliziert jedoch keine vollständige Übereinstimmung mit den Attributen der sozialen Identität.

Die öffentliche Dimension der Identität einer Person ist zu einem großen Teil das Produkt der Selbstdarstellung. Rollen- und situationsbedingt wird ein Wechsel der offenbarten Teilidentität vorgenommen und gegebenenfalls auch eine Veränderung von für den Kommunikationspartner nicht nachprüfbaren Attributen durchgeführt. Ziel ist es, den im jeweiligen Kontext gültigen sozialen Normen zu entsprechen bzw. eigenen Idealvorstellungen zu genügen. Das Image einer Person, also das beim Kommunikationspartner entstehende Abbild der Identität, ist demnach nicht einheitlich [Kun03].

Nachdem in den vorangegangenen Absätzen der Begriff der Identität aus Sicht der Philosophie und Psychologie vorgestellt wurde, widmet sich der restliche Teil dieses Abschnittes der Bedeutung von Identität in Online-Anwendungen. Technisch wird die Identität eines Benutzers vielfach auf ein Benutzerprofil reduziert, welches einen Benutzer durch eine Menge von Attributen charakterisiert, von denen einige eine eindeutige Zuordnung des Profils zu einem Benutzer ermöglichen müssen. Dabei ist zu unterscheiden, ob ein Benutzer anonym auftritt, eine gewählte Identität verwendet oder seine reale, eventuell sogar authentifizierbare Identität preisgibt.

Aus Sicht eines Dienstanbieters mag die Repräsentation eines Benutzers durch ein Benutzerprofil, welches im Minimalfall lediglich einen Login-Namen und ein Paßwort beinhaltet, ausreichend sein. Die durch Kommunikation innerhalb von Communities entstehenden *virtuellen Identitäten* sind jedoch um einiges komplexer strukturiert.

Bereits in Online-Anwendungen, wie dem Usenet, die nur minimale Hinweise zur Identität des Benutzers liefern, können virtuelle Identitäten aufgebaut werden, welche einen Teil der Persönlichkeit des Benutzers widerspiegeln und als Kontext anzusehen sind, in dem die Kommunikation innerhalb der Community zu beurteilen ist.

Identity plays a key role in virtual communities. In communication, which is the primary activity, knowing the identity with whom you communicate is essential for understanding and evaluating an interaction [Don97].

Die virtuelle Identität einer Person basiert, sofern diese keine verlässlichen Hinweise auf Attribute der realen Identität liefert, ausschließlich auf den Erfahrungen, Beziehungen

und Interaktionen in Computer-Welten. Sie ist dabei in erster Linie Produkt der Selbstdarstellung. Ein Anwender kann beliebig viele virtuelle Identitäten kreieren, welche ein von der realen Identität losgelöstes Eigenleben haben können. Die Schaffung willkürlicher Identitäten mit fiktiven Attributen wird durch die medienspezifische Kanalbegrenzung des Internets begünstigt [Kun03].

Die Bedeutung von Identitäten in Communities ist immens. In Umgebungen, in denen jeder Benutzer ohne vorherige Prüfung einer Kontrollinstanz Inhalte, z. B. in Form von Diskussionsbeiträgen oder Annotationen, veröffentlichen kann, liefert die Identität des Autors einen wichtigen Hinweis auf die zu erwartende Qualität des Beitrages. Die Darstellung der Identität des Autors ist die Voraussetzung, um neue Beiträge mit vorherigen Artikeln des selben Autors assoziieren zu können. Identität ist demnach als eine Voraussetzung für den Aufbau von Reputation (siehe Abschnitt 2.2.3) anzusehen. Beinhaltet die virtuelle Identität einer Person verifizierbare Attribute der realen Identität, so lassen sich die im virtuellen Raum des Netzes durchgeführten Handlungen einer Person unter Einbeziehung von Wertmaßstäben aus der realen Welt beurteilen. Online-Kommunikation gewinnt folglich an Relevanz, wenn eine Verknüpfung zwischen realer und virtueller Identität möglich wird. Der Meinung einer Person, deren E-Mail-Adresse auf die Tätigkeit in einem Forschungsinstitut hindeutet, sollte sicherlich gegenüber der Meinung eines AOL-Nutzers eine höhere Bedeutung in einer fachlichen Diskussion beigemessen werden. Dabei ist Vorsicht geboten: Zum einen lassen sich E-Mail-Adressen leicht fälschen, zum anderen läßt eine institutionelle E-Mail-Adresse keine sicheren Schlüsse auf die ausgeübte Tätigkeit und die fachliche Qualifikation zu [Don97].

Ein weiterer Aspekt der Betrachtung von virtuellen Identitäten ist, daß in Umgebungen, die die Identität von Nutzern erkennen lassen, und sei es nur durch ein Pseudonym, der Aufbau eines positiven Images Motivation für kooperatives, soziales Verhalten in einer Community sein kann [Don97].

Internetanwender nutzen vielfach verschiedene Dienste im Netz. Dabei müssen sie sich in der Regel bei verschiedenen Anbietern registrieren, geben jeweils Informationen über sich preis, die anbieterseitig in Benutzerprofilen gespeichert werden und bringen in der Kommunikation mit anderen Nutzern einen Teil ihrer Persönlichkeit zum Ausdruck. Die virtuelle Identität eines aktiven Internetnutzers ist also in der Regel fragmentiert. Selbstverständlich kann es der Wille eines Nutzers sein, in verschiedenen Umgebungen mit unterschiedlichen Identitäten aufzutreten. Vielfach wäre den Benutzern jedoch daran gelegen, an unterschiedlichen „Orten“ im Netz mit der selben virtuellen Identität aufzutreten. Die derzeitigen technischen Rahmenbedingungen im Internet erschweren dies oder machen es gänzlich unmöglich. Ein Ansatz, die Fragmentierung virtueller Identitäten zu verhindern und dem Benutzer weitreichende Kontrolle über seine Identitätsinformationen zu gewähren, stellen Identitätsmanagement-Systeme dar. Diese werden im nachfolgenden Abschnitt erläutert.

## Identitätsmanagement

Aufgabe von Identitätsmanagement-Systemen ist es, dem Benutzer ein Instrument zur Verwaltung seiner Identität zu geben und ihn zur feingranularen Kontrolle der Kommunikation von Identitätsinformationen in verschiedenen Situationen zu befähigen. Derartige Systeme orientieren sich an der gezielten Darstellung von Attributen der eigenen Persönlichkeit in verschiedenen Rollen und Kontexten. Identity-Management kann somit als digitales Äquivalent der Selbstdarstellung bezeichnet werden [Kun03].

Grundidee ist es, Repräsentation und Speicherung von Benutzerprofilen als eigenständigen Dienst zu etablieren und somit von auf diesen Profilverinformationen basierenden Anwendungen zu trennen. Der einheitliche Zugriff auf Identitätsdaten verhindert eine Fragmentierung von Identitäten und fördert die anwendungsübergreifende Wiederverwendung von Informationen. Dies senkt die Eintrittsbarrieren bei der Benutzung personalisierter Dienste und dem Zugang zu neuen Communities [Koc02, KW01]. Des Weiteren wird die Aktualisierung von Profilverinformationen vereinfacht, da eine redundante Speicherung bei verschiedenen Anbietern entfällt.

Benutzerzentrierte Ansätze gewähren dabei dem Anwender vollständige Kontrolle über seine Identitätsinformationen. Entsprechende Systemarchitekturen beinhalten unter Umständen als Alternative zu einer serverseitigen Speicherung, die Ablage der Identitäten auf dem Rechner des jeweiligen Nutzers.

Identitätsmanagement-Systeme versetzen den Benutzer in die Lage, verschiedene seiner Rollen entsprechende Identitäten zu definieren und zu entscheiden, wer in welchem Kontext auf welche Informationen zugreifen darf. Dabei kann der Benutzer in feinen Abstufungen zwischen einem anonymen Auftreten, Pseudonymisierung und Preisgabe der realen Identität wählen. Derartige Systeme machen den Zugriff auf Attribute seiner Identität für den Benutzer transparent und fördern das Bewußtsein für das bei den verschiedenen Anbietern über die Zeit hinweg entstehende Abbild der eigenen Identität. Resumierend kann festgestellt werden, daß benutzerzentrierte Identitätsmanagement-Systeme einen vielversprechenden Ansatz zur Stärkung des Schutzes der Privatsphäre von Internetanwendern darstellen [Koc02].

### 2.2.2. Vertrauen

Das Internet bietet mannigfaltige Möglichkeiten der Interaktion mit anderen Nutzern bzw. Diensteanbietern. Im Gegensatz zur realen Welt existieren bei der Auswahl der Interaktionspartner nur wenige auf räumliche oder kulturelle Differenzen zurückzuführende Restriktionen. Bestehen zwischen Interaktionspartnern Informationsasymmetrien, so beinhaltet dieses das Risiko opportunistischen Verhaltens einer Person zu Lasten der anderen.

Informationsasymmetrien können z. B. auf nicht überprüfbareren Angaben oder nicht nachvollziehbaren Handlungsabsichten der beteiligten Personen beruhen. Wird derjenige Online-Shop, welcher den besten Preis für ein Produkt bietet, auch wirklich die bestellte und per Vorkasse bezahlte Ware liefern? Hat eine Person, die in einem Diskussionsforum zu Medizinthemen einen Kommentar schreibt, wirklich das Expertenwissen, das sie vorgibt zu haben? Ist ein Rezensent eines Buches in einer Verbraucher-Community wirklich unabhängig oder ein Mitarbeiter der PR-Abteilung des Verlages?

Setzen sich Interaktionspartner wissentlich über ihren Informationsmangel hinweg und erbringen riskante Vorleistungen oder vertrauen nicht verifizierbaren Angaben, so basiert die Risikobereitschaft vielfach auf positiven Erfahrungen in vorherigen Interaktionen mit der betreffenden Person. Handelt es sich nicht um die eigenen Erfahrungen, sondern um die öffentlich kommunizierten Erfahrungen anderer Personen, so spricht man von *Reputation* (siehe Abschnitt 2.2.3) [KMW00].

Eine Präzisierung des Begriffs *Vertrauen* nimmt Gambetta vor:

...trust (or, symmetrically, distrust) is a particular level of the subjective probability with which an agent will perform a particular action, both before [we] can monitor such action (or independently of his capacity of ever to be able to monitor it) and in a context in which it affects [our] own action [Gam90].

Diese Definition betont, daß Vertrauen keine objektive Eigenschaft einer Person oder, allgemeiner ausgedrückt, eines Agenten ist. Vertrauen zwischen einem Vertrauensgeber und einem Vertrauensnehmer liegt nach dieser Definition vor, wenn der Vertrauensgeber es gemäß seiner subjektiven Einschätzung für wahrscheinlich hält, daß der Vertrauensnehmer eine für ihn relevante, aber nicht unmittelbar überprüfbare Handlung ausführt. Ergänzend ist festzustellen, daß Vertrauen zwischen Personen kontextspezifisch sein kann. Personen können einander bezüglich einer bestimmten zu verrichtenden Aufgabe vertrauen, ohne daß dieses Vertrauen auch in anderen Kontexten Gültigkeit besitzt [MC96].

### 2.2.3. Reputation

Vertrauensentscheidungen müssen in vielen Situationen getroffen werden, in denen entweder keine hinreichenden persönlichen Erfahrungen im Umgang mit der zu vertrauenden Person vorliegen oder die Evaluation aller Aspekte der gegebenen Situation mit den eigenen Ressourcen nicht durchführbar ist [ARH00]. Wie im vorhergehenden Abschnitt erwähnt wurde, stellt der Rückgriff auf öffentlich kommunizierte Erfahrungen

anderer Personen, ein Mittel dar, um das Risiko von Vertrauenshandlungen zu minimieren. Die *Reputation* einer Person konstituiert sich aus der Vielzahl von Bewertungen früher Interaktionspartner. Reputationsinformationen bilden die Grundlage für effektive und informierte Vertrauensentscheidungen [ARH00]. Eine Präzisierung des Begriffes *Reputation* liefern Rahman und Hailes:

A reputation is an expectation about an agent's behaviour based on information about or observations of its past behaviour [ARH00].

Die Erwartung zukünftiger Interaktionen mit derselben Person, stellt eine Motivation zu korrektem Verhalten dar. Das Bewußtsein dafür, daß die Beurteilung der gemeinsamen Vergangenheit maßgeblichen Einfluß auf zukünftige Interaktionen hat, übt eine disziplinierende Wirkung in langfristigen Interaktionsbeziehungen aus. Axelrod bezeichnet dies als „shadow of the future“ ( [Axe84] nach [RKZF00]). Reputation als transferierbares Vertrauenskapital setzt verschiedene Interaktionen einer Person zueinander in Bezug [KMW00]. Die disziplinierende Wirkung des „shadow of the future“ wird somit auf unabhängige, einmalige Interaktionen zwischen fremden Personen übertragen [RKZF00]. Gleichzeitig verbindet sich mit der Sichtweise von Reputation als *soziales Kapital* ein Anreiz zu korrektem Verhalten, denn eine positive Reputation begünstigt die Anbahnung neuer, gewinnbringender Interaktionen [ARH00].

Die Bedeutung von Reputation basiert auf der Annahme, daß die Erfahrungen Dritter im Umgang mit einer Person als Indikator für zukünftiges Verhalten dieser Person zu sehen sind. Die Validität dieser Annahme ist nach Wilson fraglich, da Reputation streng genommen nur Aussagekraft über die Gültigkeit vergangener Handlungen besitzt ( [Wil86] nach [KMW00]). Dabei ist zu berücksichtigen, daß die Relevanz von Reputation nur dann gegeben ist, wenn das Risiko eines Reputationsverlustes für die betroffene Person mit Konsequenzen, wie z. B. einer verschlechterten Marktposition, behaftet ist [KMW00].

Vorraussetzung für den Aufbau von Reputation sind lang bestehende Identitäten, die im Verlauf der Zeit in eine Vielzahl von Interaktionen involviert sind [RKZF00]. Die Reputation einer Person ist mit ihrer (virtuellen) Identität assoziiert. In Umgebungen, die ein pseudonymisiertes Auftreten gestatten, ist es von Bedeutung, daß die Verknüpfung von Pseudonym und realer Identität über die Zeit stabil bleibt. Denn das Vertrauen, welches die Benutzer eines Systems einer durch ein Pseudonym repräsentierten virtuellen Identität entgegenbringen, basiert auf der Annahme, daß sich hinter dem Pseudonym nach wie vor dieselbe Person verbirgt, welche durch kooperatives Verhalten eine positive Reputation erworben hat [DFM01]. Ferner sollte zu einem Zeitpunkt zu einem Individuum nur ein Pseudonym gehören, denn sonst könnten Personen mit mehreren Pseudonymen zum eigenen Vorteil agieren. Des weiteren sollte das Abstreifen einer negativen Reputation durch Erstellung eines neuen Pseudonyms erschwert werden [RKZF00].

Die Herkunft von Reputationsinformationen ist bei der Beurteilung ihrer Aussagekraft nicht außer Acht zu lassen. Bei der Einschätzung der Reputation einer Person wird den Meinungen vertrauter Personen naturgemäß ein höheres Gewicht beigemessen ([PRW00] nach [KMW00]).

### Reputationssysteme

In der realen Welt werden Reputationsinformationen per „Mund-zu-Mund-Propaganda“ ausgetauscht, was den Kreis der zu konsultierenden Personen auf das persönliche Netzwerk beschränkt und folglich die Menge der Personen bzw. Dienstanbieter, über deren Reputation auf diesem Wege etwas zu erfahren ist, stark begrenzt. In der virtuellen Welt des Internets sind die Unsicherheiten im Umgang mit den teilweise unter einem Pseudonym auftretenden Interaktionspartnern ungleich größer. Es bieten sich jedoch auch weitergehende Möglichkeiten des softwaregestützten Austausches von Reputationsinformationen. Reputationssysteme stellen die Historie der Bewertungen vergangener Interaktionen einer Person dar. Resnick et al. definieren Reputationssysteme folgendermaßen:

A reputation systems collects, aggregates and distributes feedback about past interactions [RKZF00].

Mittels der Hilfe von Reputationssystemen können innerhalb großer Communities Erfahrungen im Umgang mit Interaktionspartnern ausgetauscht werden. Persönliche Empfehlungen haben naturgemäß eine andere Relevanz als das über ein Reputationssystem ausgetauschte Feedback. Die Stärke von Reputationssystemen liegt jedoch in der Quantität der Bewertungen, die Einschätzungen über eine große Menge potentieller Interaktionspartner abfragbar macht. Die Vermutung, daß Reputationsinformationen von anderen potentiellen Interaktionspartnern eingesehen werden, wirkt sich disziplinierend und motivierend auf die beteiligten Personen aus (siehe Abschnitt 2.2.3). Durch Reputationssysteme erhalten unabhängige, kurzfristige Interaktionen, sofern regelmäßig Bewertungen abgegeben werden und Bewertungen Vertrauensentscheidungen maßgeblich beeinflussen, den Charakter langfristiger Beziehungen [RKZF00].

Das eBay<sup>2</sup> Feedback-Forum (siehe Abbildung 2.3) ist ein Beispiel für ein sehr einfaches Reputationssystem, das dennoch einen Teil der mit der Abwicklung von Transaktionen über diese Auktionsplattform verbundenen Risiken minimiert bzw. kalkulierbar macht. Nach Abschluß einer Transaktion können sich die beteiligten Nutzer gegenseitig auf einer Skala mit den Werten „positiv“, „neutral“ und „negativ“ bewerten und einen Freitextkommentar hinterlassen. Die akkumulierte Anzahl positiver Bewertungen wird jeweils in

---

<sup>2</sup>[www.ebay.de](http://www.ebay.de)

**Gesamtprofil**  
 2230 positive Bewertungen. 2118 stammen von unterschiedlichen Mitgliedern und gehen in die endgültige Bewertung ein.  
 46 neutrale Bewertungen.  
 22 negative Bewertungen. 22 stammen von unterschiedlichen Mitgliedern und gehen in die endgültige Bewertung ein.

**Übersicht über die jüngsten Bewertungen**

	Letzte 7 Tage	Letzter Monat	Letzte 6 Monate
Lob	36	261	1650
neutrale Bewertungen	0	8	39
Negativ	1	3	17
<b>Gesamt</b>	<b>37</b>	<b>272</b>	<b>1706</b>
Zurückgezogene Gebote	0	0	0

**alta-discount\*de 's Bewertung** Bewertung [Hilfe](#) | [Fragen & Antworten](#)

Bewertung 1 - 25 von 2298  
 [ 1 ] [2](#) [3](#) [4](#) [5](#) [6](#) ... [20](#) ... [40](#) ... [60](#) ... [80](#) ... [92](#) (nächste Seite)

[Bewertung abgeben](#) für alta-discount\*de Wenn Sie alta-discount\*de sind: [Auf Kommentare antworten](#) alta-discount\*de war der **Verkäufer = V**  
alta-discount\*de war der **Käufer = K**

	Datum	Artikelnummer	V/K
<a href="#">einweidreivierfuenf (11)</a> ★	12.01.03 19:22:34 MEZ	<a href="#">1949599357</a>	V
<b>Lob</b> : ++++++			
<a href="#">stoppelivi (49)</a> ★	12.01.03 18:06:35 MEZ	<a href="#">1942256828</a>	V
<b>Lob</b> : Alles o.k.			
<a href="#">kerlchen0 (1)</a>	12.01.03 18:02:29 MEZ	<a href="#">1943636925</a>	V
<b>Lob</b> : schnelle Lieferung, gute Ware, guter Preis, 1. Ebay-Erfahrung also nur positiv.			
<a href="#">blauwusel (485)</a> ★	12.01.03 05:21:29 MEZ	<a href="#">1949445011</a>	V

Abbildung 2.3.: Ebay Feedback-Forum

Verbindung mit dem Benutzernamen angezeigt. Eine Detailansicht listet die einzelnen Bewertungen auf, so daß ersichtlich ist, wie sich die Bewertungen über die Zeit verteilen und welche Reputation die Nutzer haben, welche Bewertungen abgegeben haben. Ein Benutzer kann somit Reputationsinformationen einbeziehen, wenn er entscheidet, ob er für einen Artikel eines bestimmten Anbieters ein Gebot abgeben möchte und welcher Betrag dabei sein Höchstgebot ist.

Ein etwas anderes Beispiel eines Reputationssystems ist die Verbraucher-Community [ciao.com](http://www.ciao.com)<sup>3</sup>. Auf dieser Plattform können Konsumenten ihre Meinungen zu Produkten und Anbietern einem breiten Publikum zugänglich machen. Die Qualität der von den Benutzern veröffentlichten Beiträge variiert stark. Um den Nutzern des Systems das Auffinden qualitativ hochwertiger Testberichte zu erleichtern, wurde ein Reputationssystem für Produktkritiker in die Site integriert. Die Beiträge der Autoren können nach Nützlichkeit bewertet werden. Im Gegensatz zu dem Beispiel des Online-Auktionshauses Ebay wird folglich nicht direkt eine Bewertung anderer Benutzer vorgenommen. Die Bewertung der Autoren erfolgt indirekt über die Bewertung der von ihnen dargestellten Meinun-

<sup>3</sup>[www.ciao.com](http://www.ciao.com)

gen. In diesem Fall könnte man von indirekter Reputation sprechen. Testberichte von Autoren mit hoher Reputation lassen sich bevorzugt darstellen, was für die Teilnehmer der Community das Risiko minimiert, sich von einer zweifelhaften Produktempfehlung zu einem Fehlkauf verleiten zu lassen.

Der Betrieb von Reputationssystemen ist mit einer Reihe von Problemen behaftet. Grundvoraussetzung für ein nützliches Reputationssystem ist es, daß möglichst viele Benutzer Bewertungen abgeben. Es ist also bei der Konzeption eines derartigen Systems zu überlegen, welche Anreize zur Bewertungsabgabe zu schaffen sind [RKZF00]. Die Bewertungsabgabe unterliegt den subjektiven Kriterien der Teilnehmer, welche bei der Beurteilung ihrer Interaktionspartner höchst unterschiedliche Wertmaßstäbe anlegen können. Eine weitere Schwierigkeit ist darin zu sehen, daß negative Bewertungen nur höchst selten abgegeben werden, da vor Abgabe einer negativen Bewertung vielfach erst verhandelt wird bzw. von einer negativen Bewertung aus Angst, ebenfalls eine negative Bewertung zu erhalten, abgesehen wird [RKZF00]. Reputationssysteme unterliegen der Gefahr, manipuliert zu werden. Gruppen von Personen könnten kollaborieren und sich gegenseitig hoch bewerten oder die Konkurrenz mit ungerechtfertigten negativen Bewertungen diskreditieren [RKZF00]. Weitere Möglichkeiten zur Manipulationen ergeben sich im Zusammenhang mit der virtuellen Identität der Benutzer, wenn z. B. Benutzer über mehrere Pseudonyme verfügen oder beliebig das Pseudonym wechseln und damit eine eventuell schlechte Reputation ablegen können.

Reputationssysteme im heutigen Internet sind Insellösungen, zwischen denen kein Austausch stattfindet [RKZF00]. Ein Nutzer, der sich z. B. bei Slashdot<sup>4</sup> durch hervorragende Beiträge profilieren konnte, muß sich in einer anderen Umgebung wieder aktiv um den Aufbau einer neuen Reputation bemühen. Die Portabilität von Reputationen würde den Eintritt in neue Communities erleichtern und spontane Kooperationen kalkulierbarer machen.

### 2.3. open net environment for Citizens (onefC)

Es gibt sicherlich aus Sicht eines Internetnutzers viele Gründe, sich möglichst anonym im Netz zu bewegen. Jedoch gibt es auch Situationen, in denen es im Interesse des Benutzers ist, „jemand im Netz zu sein“, also erkannt zu werden und Attribute seiner Persönlichkeit ausdrücken zu können. So wäre es sicherlich für einen Nutzer, der sich in einem Web-Forum durch besonders hochwertige Beiträge hervorgetan hat, wünschenswert, wenn er mit der in diesem Kontext angenommenen Identität und der damit assoziierten Reputation auch in anderen Umgebungen auftreten könnte und von Nutzern des Web-Forums an anderen Orten im Netz wiedererkannt werden würde.

---

<sup>4</sup>[www.slashdot.org](http://www.slashdot.org)

Zur Nutzung personalisierter Dienste oder zur Teilnahme an Communities ist vielfach eine Registrierung erforderlich. Ein aktiver Nutzer des heutigen Webs hat typischerweise Accounts bei den verschiedensten Anbietern eingerichtet und dort jeweils redundant einen Teil seiner persönlichen Daten, ergänzt um dienstspezifische Informationen, eingegeben. Dabei besteht die Schwierigkeit, sich jeweils Benutzernamen und Paßwort zu merken, und der Überblick, welchem Anbieter welche persönlichen Daten mitgeteilt wurden, geht schnell verloren. Die Nutzung personalisierter Dienste ist mit hohen Eintrittsbarrieren verbunden, die in erster Linie daraus resultieren, daß sinnvolle Personalisierung ein hohes Maß an Wissen über die Präferenzen des Benutzers seitens des Dienstanbieters verlangt. Benutzerprofile werden entweder explizit durch Eingaben oder implizit durch Auswertung von Benutzeraktivitäten aufgebaut. Die erstgenannte Möglichkeit ist für den Benutzer mit einem erheblichen Aufwand verbunden, die Verwendung der zweiten Alternative hat zur Folge, daß ein derartiger Dienst erst nach einer Anlaufphase nutzbringend einzusetzen ist. Eine Identitätsinfrastruktur, welche ein universelles Konzept zur Benutzeridentifikation bietet und den anwendungsübergreifenden Austausch von Identitätsinformationen unter der Kontrolle des Benutzers ermöglicht, könnte den Beitritt zu Communities, die Nutzung personalisierter Dienste, die Abwicklung von Geschäften über das Netz und die spontane Kollaboration mit anderen Nutzern deutlich vereinfachen.

Ziel des Projekts onefC ist die Entwicklung einer generischen Identitätsinfrastruktur, die auf einem vom psychologischen Identitätsbegriff (siehe Abschnitt 2.2.1) geprägten Identitätsmodell basiert. Dabei sollen zum einen auf der Systemebene liegende Anforderungen an eine Identitätsinfrastruktur, wie die Identifizierung, Authentifizierung und Autorisierung von Anwendern berücksichtigt werden. Zum anderen sollen aber auch Anforderungen der Anwendungsebene, wie die Bereitstellung und Abfrage beliebiger Identitätsattribute beachtet werden. Gerade in diesem Punkt unterscheidet sich der Ansatz des Projekts onefC von anderen Ansätzen, wie Microsoft .NET Passport<sup>5</sup> und der Liberty Alliance<sup>6</sup>, welche sich sehr spezifisch an den Anforderungen Web-basierter E-Commerce-Anwendungen orientieren. Kernkomponente der onefC-Infrastruktur ist ein Identitätsmanager, welcher den flexiblen, anwendungsübergreifenden Austausch von Identitätsinformationen gestattet und durch diverse Dienste, z. B. zur Autorisierung und Protokollierung des Zugriffs auf Attribute einer Identität, ergänzt wird.

Die Idee des Projekts onefC wird in [BZL03] beschrieben. Des weiteren sind Details über die technischen Realisierung der Kernkomponenten in [Kun03] nachzulesen.

Nachfolgend werden die aus Sicht dieser Arbeit relevanten Aspekte des Projekts onefC vorgestellt. Die von der onefC-Identitätsinfrastruktur bereitgestellten Dienste, insbesondere die vom realisierten Prototyp genutzten, werden in Abschnitt 2.3.1 beschrieben. Anschließend wird ein Überblick über die Systemarchitektur (2.3.2) gegeben und das

---

<sup>5</sup>[www.passport.net](http://www.passport.net)

<sup>6</sup>[www.projectliberty.org](http://www.projectliberty.org)

onefC-Identitätsmodell (2.3.3) präsentiert.

### 2.3.1. Dienste

Die onefC-Infrastruktur kapselt über eine einheitliche Schnittstelle den Zugriff auf lokale Identitäten, sowie die zur Abfrage fremder Identitäten notwendige Identitätskommunikation.

Um der Anforderung einer generischen Identitätsinfrastruktur Rechnung zu tragen, wurde ein universelles Identitätsmodell (siehe 2.3.3) entwickelt, daß Ausdrucksmittel für Identitäten, Subidentitäten und Attribute beinhaltet.

Die durch den Identitätsmanager realisierte Funktionalität zur Verwaltung der eigenen und fremder Identitäten schließt die Fähigkeit ein, neue Identitäten anzulegen und mit global eindeutigen Bezeichnern zu versehen. Eine Persistenzschicht ermöglicht die dauerhafte Speicherung von Identitäten auf verschiedenen Medien und in verschiedenen Formaten.

Vermittelt durch die Identitätsmanager wird der Austausch von Identitätsinformationen zwischen Anwendungen möglich, ohne daß diese explizit Kenntnis voneinander haben müssen. Diese Interoperabilität setzt natürlich ein gemeinsames semantisches Verständnis der ausgetauschten Daten voraus. Der Schutz der Privatsphäre der Benutzer war ein wichtiges Anliegen bei der Systemkonzeption. Die Beantwortung von Anfragen nach Identitätsinformationen wird durch einen Sicherheitsdienst gesteuert und durch einen Monitordienst protokolliert.

### 2.3.2. Architektur

In diesem Abschnitt wird ein Überblick über die Systemarchitektur, mit der die im vorstehenden Abschnitt beschriebenen Dienste realisiert werden, geben. Eine detailliertere Darstellung ist [Kun03] zu entnehmen.

Wie in Abbildung 2.4 ersichtlich ist, wird die Funktionalität der Identitätsinfrastruktur durch den Identitätsmanager gegenüber den Client-Anwendungen gekapselt. Um eine hohe Flexibilität und Adaptionfähigkeit zu gewährleisten, wurden Teilfunktionen des Systems aus der zentralen Manager-Komponente ausgelagert und als eigenständige Dienste realisiert, die dann über wohldefinierte Schnittstellen vom Identitätsmanager angesprochen werden.

Nachfolgend werden die Teilkomponenten, welche zusammen mit dem Identitätsmanager die Leistung der Identitätsinfrastruktur erbringen, kurz vorgestellt:

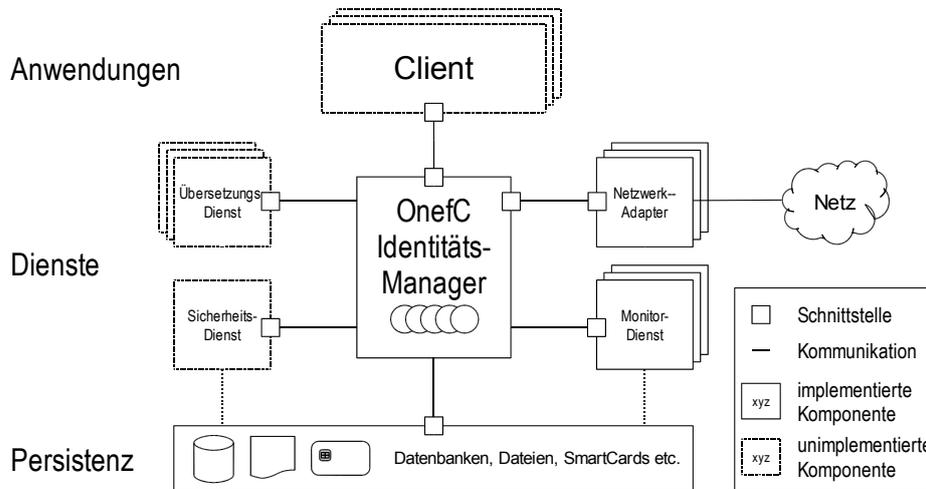


Abbildung 2.4.: Übersicht über die onefC-Architektur (Quelle: [Kun03])

**Sicherheitsdienst** Anfragen nach Attributen einer Identität autorisiert der Identitätsmanager durch Rückgriff auf den Sicherheitsdienst. Konkrete Implementierungen des Sicherheitsdienstes könnten beim Benutzer nachfragen, ob er die Herausgabe eines bestimmten Attributes seiner Identität wünscht oder regelbasiert entscheiden.

**Monitordienst** Monitordienste zur Protokollierung der externen Kommunikation können sich beim Identitätsmanager registrieren. Sie werden dann nach dem Observer-Muster über eingehende und versandte Nachrichten informiert. Anhand der erstellten Protokolle können Benutzer nachvollziehen, welche Attribute ihrer Identität kommuniziert wurden.

**Übersetzungsdienst** Als Attribute einer Identität können beliebige in XML repräsentierbare Inhalte gespeichert werden. Um den anwendungsübergreifenden Austausch von Daten zu vereinfachen, wurde das Konzept eines Übersetzungsdienstes in die Architektur integriert. Aufgabe einer derartige Komponente ist es, unterschiedliche XML-Repräsentationen ineinander zu überführen.

**Netzwerk-Adapter** Aufgabe der Netzwerk-Adapter ist es, die Kommunikationsschnittstelle des Identitätsmanagers an das verwendete Transportprotokoll anzupassen.

**Persistenzschicht** Die dauerhafte Speicherung der vom Identitätsmanager verwalteten Identitäten ist Aufgabe dieser Teilkomponente. Der aktuelle Prototyp beinhaltet eine Ablage der Identitäten im Dateisystem. Eine Datenbankanbindung oder die Speicherung auf Smartcards wären mögliche Alternativen.

### 2.3.3. Identitätsmodell

Das onefC-Identitätsmodell orientiert sich an neueren psychologischen Sichtweisen, die davon ausgehen, daß die Identität eines Individuums nicht als einheitliches Ganzes, sondern als „Identitäts-Patchwork“ zu sehen ist. Dabei wird rollen- und situationsbedingt ein Wechsel der offenbarten Teilidentität vorgenommen (siehe Abschnitt 2.2.1). Den Rahmen für die Abbildung von Identitäten bildet eine Baumdarstellung. Eine Teil-Identität, welche in einem oder mehreren Kontexten aktiv ist, wird in dieser Darstellung durch einen Identitätsknoten repräsentiert. Ausgehend von einer oder mehreren *Root-Identitäten* und damit verknüpften *Sub-Identitäten* bildet sich so eine Hierarchie von Teil-Identitäten, in der Werte vererbt bzw. überschrieben werden können. Durch die Verwendung einer Baumstruktur und die Realisierung eines Vererbungskonzeptes wurde die redundante Speicherung von Attributen minimiert und gleichzeitig die feingranulare Kontrolle über die offenbarte Teil-Identität ermöglicht.

Um inhaltlichen Einschränkungen bezüglich der Menge der darstellbaren Attribute zu vermeiden, wurde lediglich ein Container-Modell für Attribute definiert. Dieses beinhaltet eindeutige Bezeichner für Attribute und Meta-Daten zur semantischen Beschreibung der Inhalte. Beliebige, in XML repräsentierbare Daten können in den Attribut-Containern aufgenommen werden. Bezüglich der Semantik der Daten wird davon ausgegangen, daß Daten, welche dem selben XML-Schema entsprechen, eine übereinstimmende Semantik aufweisen. Alternativ wird im Rahmen des Projektes onefC daran geforscht, Attribute von Identitäten durch Ontologien zu beschreiben.

Abbildung 2.5 zeigt das UML-Klassendiagramm des onefC-Identitätsmodells. Zentrales Element ist der Identitäts-Knoten (*Identity*). Die reflexive Aggregations-Kante an diesem Modellelement drückt aus, daß zu einer Identität mehrere Sub-Identitäten gehören können. Jede Identität ist in einem oder mehreren Kontexten (*Context*) gültig und beinhaltet eine Menge von Attribute (*IdentityAttribute*). Identitäten, Kontexte und Attribute sind jeweils über *Identifier* adressierbar. Dabei wurde die Einschränkung vorgenommen, daß die Werte der Bezeichner aller Sub-Identitäten eines Identitätsbaumes mit dem Bezeichner der Root-Identität übereinstimmen. Hieraus resultiert eine kontextübergreifende Verknüpfbarkeit von Teil-Identitäten und damit die Fähigkeit, Personen in unterschiedlichen Anwendungen wiederzuerkennen.

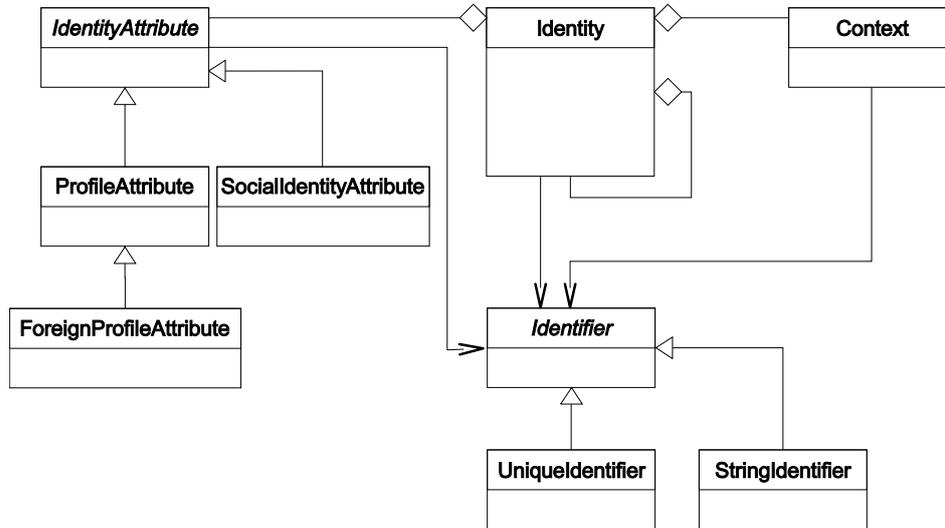


Abbildung 2.5.: UML-Klassendiagramm des onefC-Identitätsmodells (Quelle: [Kun03])

## 2.4. Scone

Die clientseitige Komponente des in dieser Diplomarbeit entwickelten Systems basiert zu großen Teilen auf dem in Java implementierten Framework Scone<sup>7</sup>, das am Fachbereich Informatik der Universität Hamburg als technische Grundlage der Dissertationen von Matthias Mayer<sup>8</sup> und Harald Weinreich<sup>9</sup> entwickelt wurde. Scone bietet eine Umgebung, die es ermöglicht, prototypisch Programme zur Unterstützung der Navigation im Web zu implementieren und zu evaluieren. Ein Plugin-Architektur gestattet es, flexibel neue Komponenten in das Framework einzubinden. Scone-Plugins können so die Funktionalität des Web-Browsers erweitern und vollkommen neue Sichten auf das Web generieren. Scone besteht aus den vier Kernkomponenten *NetObjects*, *Proxy*, *Robot* und *UserTracking*, die durch eine Reihe von Hilfskomponenten ergänzt werden. Einen Überblick über die Scone-Architektur zeigt Grafik 2.6. Die folgende Beschreibung der Komponenten ist an die Darstellung in [WBL03] angelehnt.

**NetObjects** Das Framework stellt für die wichtigsten Basiskonzepte des Internets wie URLs, Dokumente und Links ein objektorientiertes Datenmodell bereit. Zusätzlich lassen sich auch Benutzer und ihre Aktionen auf das Datenmodell abbilden. Die den Proxy

<sup>7</sup> [www.scone.de](http://www.scone.de)

<sup>8</sup> [asi-www.informatik.uni-hamburg.de/personen/mayer](http://asi-www.informatik.uni-hamburg.de/personen/mayer)

<sup>9</sup> [vsiis-www.informatik.uni-hamburg.de/members/info.phtml/20](http://vsiis-www.informatik.uni-hamburg.de/members/info.phtml/20)

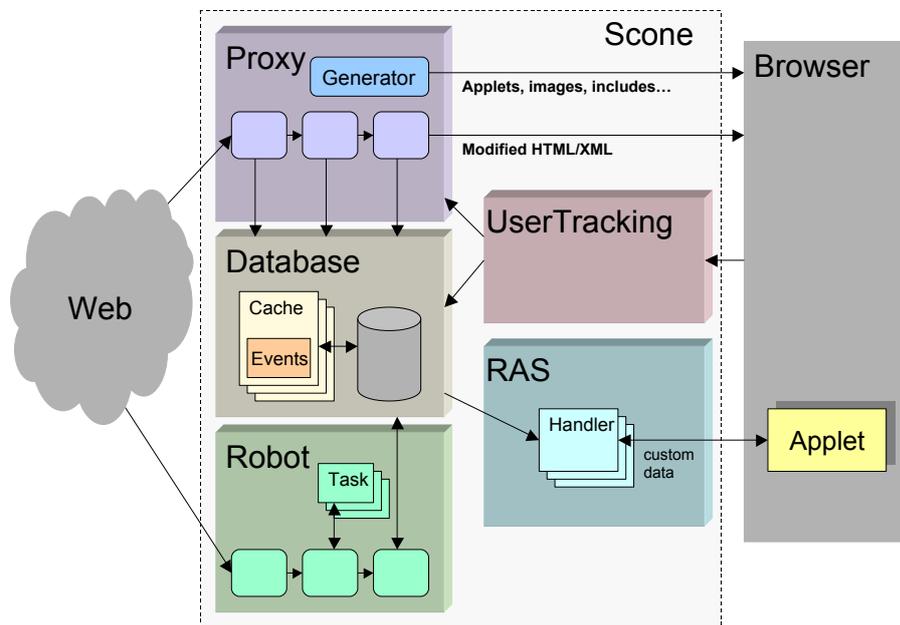


Abbildung 2.6.: Architektur des Scone-Frameworks

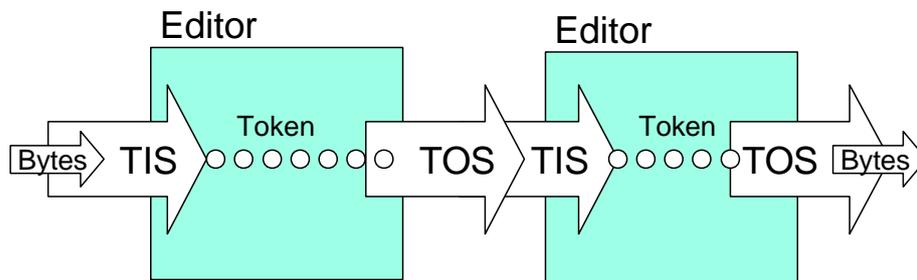
durchlaufenden bzw. vom Robot geladenen Dokumente werden von Scone persistent abgelegt. Dabei werden die gefundenen Dokumente und Links auf Java-Objekte abgebildet, die dann in Objekt-Caches abgelegt werden. Diese wiederum kapseln den Zugriff auf eine Datenbank, wobei Anbindungen für die relationale Datenbank „MySQL“<sup>10</sup> und die objektorientierte Datenbank „Poet OSS“<sup>11</sup> existieren. Für jedes neue Objekt erzeugen die Caches Events, auf die die Scone-Plugins dann reagieren können.

**Proxy** Ein Proxy-Server ist Bestandteil von Scone. Über diesen wird die Kommunikation zwischen Browser und Server abgewickelt. Dabei wird der Datenstrom analysiert und gegebenenfalls manipuliert. Diese erweiterten Proxy-Funktionalitäten basieren auf dem WBI (Web Based Intermediary) Development Kit<sup>12</sup> vom IBM Almaden Research Centre [BM99,MB00], das zur Realisierung des Scone-Proxy verwendet und erweitert wurde. Die geladenen Web-Seiten werden im Proxy automatisch in verschiedene Token zerlegt, die beispielsweise HTML-Tags oder Text repräsentieren. Scone-Plugins können sich mit

<sup>10</sup> [www.mysql.com](http://www.mysql.com)

<sup>11</sup> [www.fastobjects.de](http://www.fastobjects.de)

<sup>12</sup> [www.almaden.ibm.com/cs/wbi](http://www.almaden.ibm.com/cs/wbi)



**Abbildung 2.7.:** Bearbeitung eines TokenStreams durch zwei Editor-Plugins.  
TIS: TokenInputStream, TOS: TokenOutputStream. (Quelle: [Buc02])

Hilfe des Proxys in den Datenstrom zwischen Client und Server einklinken und haben dadurch Zugriff auf die übertragenen Dokumente. Die Token-basierte Verarbeitung erlaubt die effiziente Analyse und Manipulation von Dokumenten, da durch dieses Konzept ein mehrfaches Parsen des Bytestroms hinfällig wird. Abbildung 2.7 veranschaulicht, wie ein in Dokument als *TokenStream* mehrere Plugins durchläuft.

**Robot** Im Verlauf der Studienarbeit des Autors wurde ein sehr flexibel konfigurierbarer, für den Einsatz auf der Client-Seite optimierter Web-Robot entwickelt. Scone-Plugins können mittels dieser Komponente des Frameworks automatisch Dokumente aus dem Web laden, die dann in den Scone-Caches abgelegt werden. Das Verhalten des Robots ist durch ein *Classifier-Filter*-Konzept steuerbar. Sowohl Links als auch Dokumente werden zunächst aufgrund ihrer Eigenschaften klassifiziert und anschließend entsprechend den Filterbedingungen verarbeitet. Scone-Plugins können so spezifizieren, welche Links weiterverfolgt und wie mit den Zieldokumenten verfahren werden soll. Der Robot wurde für den clientseitigen Einsatz optimiert. Hier galt es einen Kompromiß zwischen möglichst geringer Netzlast und größtmöglicher Performanz zu finden. So kann der Robot je nach Anforderung nebenläufig mehrere Seiten gleichzeitig übertragen oder auch lastmindernd nur in Phasen geringer Systemaktivität Dokumente laden. Der Robot kann sowohl komplette Dokumente laden als auch selektiv Informationen zu Dokumententyp, Dateigröße, Änderungsdatum, Dokumententitel und Antwortzeit des Servers ermitteln.

Ausführliche Informationen zum Scone-Robot können in [Wol02] nachgelesen werden.

**UserTracking** Alle Benutzeraktionen können durch Scone verfolgt und überwacht werden. Durch Java-Events werden Scone-Plugins über Benutzeraktivitäten, z. B. das Laden eines Dokuments, informiert. Die Benachrichtigung geschieht mittels des Observer-

Patterns [GHJ95]. Das UserTracking ist von zentraler Bedeutung für den realisierten Prototyp. Denn um eine Bewertung zum gerade im Browser angezeigten Dokument vorherzusagen bzw. eine durch den Benutzer abgegebene Bewertungen korrekt einem Dokument zuzuordnen, ist es notwendig, daß der Client über Informationen zu den geöffneten Browserfenstern und den URLs der aktuell geladenen Dokumente verfügt. In alle den Proxy durchlaufende Dokumente wird ein unsichtbares Applet eingefügt, welches durch in die Seite eingebundenen JavaScript Code bei verschiedenen Browser-Events, z. B. dem Anklicken eines Links, aufgerufen wird. Die Aktionen werden dann durch das Applet über eine Socket-Verbindung, wie in Grafik 2.8 ersichtlich ist, zur Auswertung an das Framework übergeben.

Die Möglichkeiten der UserTracking-Komponente werden durch das dynamische Einblenden eines unsichtbaren Applets in jede geladene Seite gegenüber herkömmlichen, den Traffic zwischen Client und Server überwachenden Proxy-Servern stark erweitert. Browserfenster werden durch das System eindeutig benannt, so daß zwischen den Benutzeraktivitäten in verschiedenen Fenstern differenziert werden kann. Diese Funktionalität wird im Rahmen der Implementierung des Prototyps verwendet, um die Auswahl von Identität und Benutzerprofil einzelnen Browserfenstern zuzuordnen, so daß in mehreren Fenstern gleichzeitig mit verschiedenen Benutzerprofilen „gesurft“ werden kann. Als eine Schwäche des UserTrackings hat es sich erwiesen, daß Scone-Plugins zwar über das Verlassen eines Dokuments informiert werden, dabei jedoch nicht unmittelbar ersichtlich ist, ob ein neues Dokument im gleichen Fenster geladen wird, oder ob das Browserfenster geschlossen worden ist. Um das UserTracking an die spezifischen Erfordernisse der realisierten Anwendung anzupassen, wurden einige Modifikationen durchgeführt, die in Abschnitt 8.3.3 dargestellt werden.

Weitere Details zum UserTracking werden in der Studienarbeit [Haß02] beschrieben.

**Hilfskomponenten** Einige Hilfskomponenten ergänzen die Funktionalität des Frameworks.

- *Konfigurations-GUI*

Über eine grafische Benutzungsoberfläche können Plugins registriert und konfiguriert werden. Ein XML-Schema definiert ein generisches Format für die Konfigurationsdateien der Plugins. Aus diesen XML-Dokumenten wird dynamisch eine grafische Benutzungsoberfläche generiert, so daß Benutzer bequem ohne direktes Editieren der Konfigurationsdateien Änderungen vornehmen können.

- *Dokumenten-Parser*

Der Dokumenten-Parser analysiert die über den Proxy-Server geladenen Dokumente und sammelt Meta-Informationen über Inhalt, Topologie und Sprache. Mittels

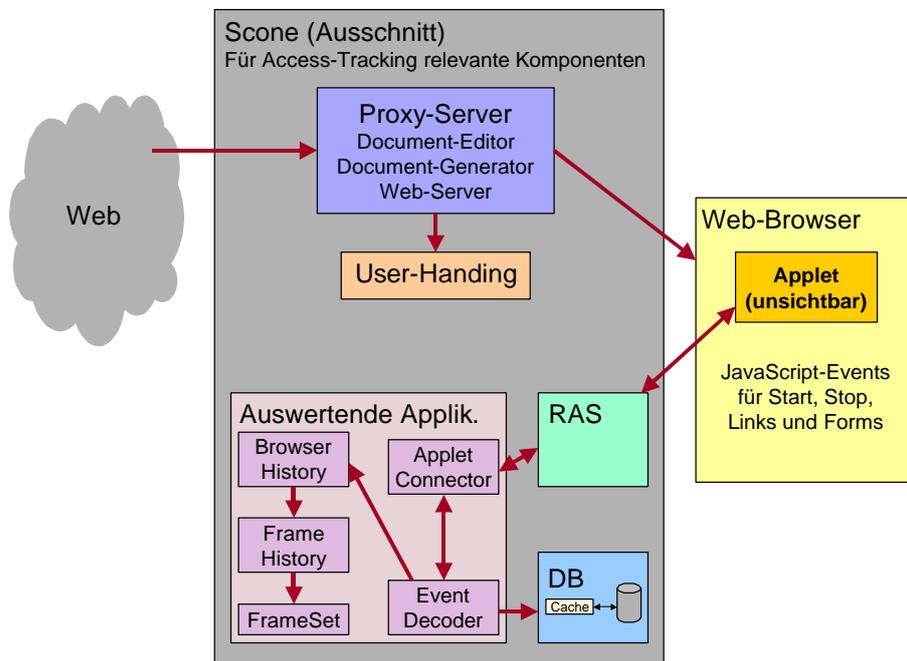


Abbildung 2.8.: Schematischer Ablauf des *UserTrackings* (Quelle: [Haß02])

einer Heuristik wird eine Kurzfassung des Dokuments erzeugt. Des Weiteren wird eine Checksumme des Dokuments, der sogenannte *Document Fingerprint* berechnet. Diese gestattet die effiziente Überprüfung von Dokumenten auf Gleichheit.

- *Browser-Steuerung*  
Das als Teil des *UserTrackings* in Dokumente eingebundene Applet kann zur Steuerung des Browserfensters verwendet werden. Die Steuerungsbefehle werden von einer Teilkomponente des Frameworks entgegengenommen und über eine Socket-Verbindung an das Applet weitergeleitet. Über die Browser-Steuerung lassen sich neue Dokumente laden, weitere Browserfenster öffnen und Fensterposition bzw. -fokus ändern. Durch direkte Betriebssystemaufrufe können Instanzen des Browsers gestartet werden.
- *Remote Access Server*  
Ein Remote Access Server bietet den Zugriff auf alle Scone-Klassen und Plugins über eine Socket-Schnittstelle. Dieses Feature läßt sich z. B. für die Interaktion von Applets mit Scone einsetzen.

- *Thumbnail-Generator*  
Verkleinerte Darstellungen von Web-Dokumenten, sogenannte Thumbnails, sind eine Möglichkeit der Repräsentation von Dokumenten in Visualisierungen. Eine in Java implementierte Komponente zum Rendern von HTML-Dokumenten wurde in das Framework integriert. Diese stellt Dokumente als Grafik in einer spezifizierbaren Größe zur Verfügung.
- *Paket zur Evaluation*  
Das Scone-Framework unterstützt Entwickler bei der prototypischen Realisierung von Erweiterungen des Webs. Um anhand dieser Prototypen Erfahrungen für die weitere Entwicklung zu gewinnen, ist eine Evaluation sinnvoll. Das Framework vereinfacht die Durchführung von Benutzertests. In die Szenarios einführende Texte und interaktive Fragebögen werden in einer Benutzungsoberfläche neben dem Browserfenster dargestellt. Zusätzlich werden die Benutzeraktionen für spätere Auswertungen protokolliert.

Scone-Anwendungen werden mittels eines Plugin-Konzepts realisiert. Die Plugins haben dabei über eine Java-API Zugriff auf alle Funktionen von Scone. Ein Plugin kann den geparteten Strom der HTML-Dokumente manipulieren und als Server für dynamische und statische Dokumente agieren. Es hat vollen Zugriff auf die Scone-Caches und kann auf die verschiedenen von Scone erzeugten Events reagieren, um z. B. beim Laden einer neuen Seite durch den Benutzer eine Visualisierung zu aktualisieren.

Die Client-Komponente des „Collaborative Internet Experience“ Systems wurde als Scone-Plugin realisiert. Andere Plugins können über eine in 8.3.3 beschriebene API auf den Rating-Server zugreifen, was exemplarisch in der Erweiterung des HyperScouts durchgeführt wurde.

### 2.5. HyperScout

Der Funktionsumfang der Web-Browser hat seit der Einführung des Webs ständig zugenommen. Dabei sind insbesondere die Gestaltungsmöglichkeiten von HTML-Dokumenten erweitert worden. Über die bekannten Features „Back-Button“ und „Bookmarks“ hinaus bieten jedoch auch aktuelle Browser kaum weitergehende Funktionen zur Vereinfachung der Navigation im Web. Das Link-Interface wurde im Verlauf der Evolution des Webs nicht grundlegend modifiziert. Hypertext-Links werden nach wie vor durch typographische Merkmale, wie Unterstreichung und Einfärbung, gekennzeichnet, dabei werden auf bereits besuchte URLs verweisende Links üblicherweise farblich gekennzeichnet und die URL des Zielobjekts in der Status-Zeile des Browsers dargestellt. Angesichts der Tatsache, daß die Benutzung von Hypertext-Links einen Großteil [CP95, TG97] der

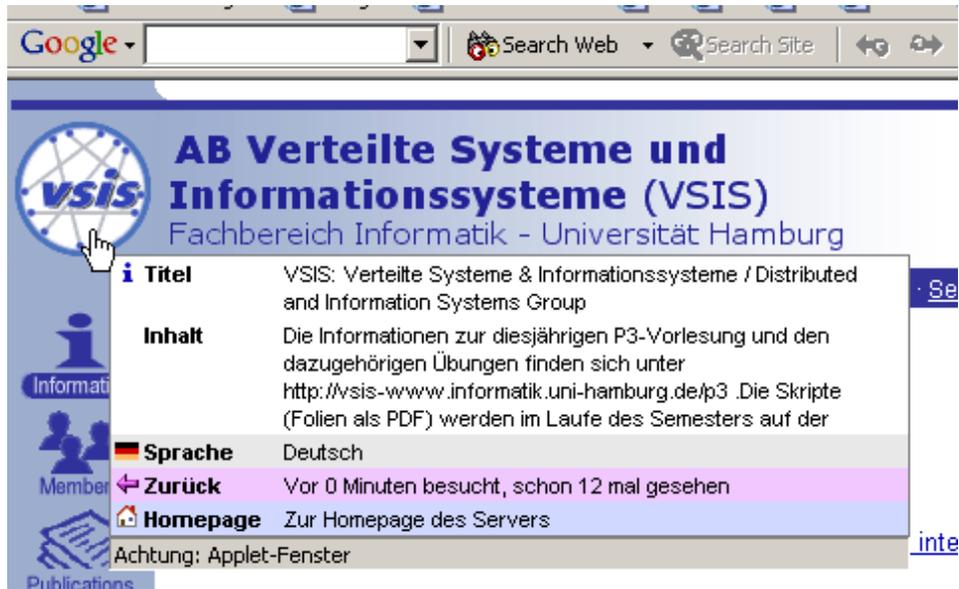


Abbildung 2.9.: Beispiel einer HyperScout-Einblendung

Navigationsaktivitäten bei der Web-Benutzung ausmachen, liegt in der Verbesserung des Link-Interfaces sicher einiges an Potential zur Vereinfachung der Navigation im Web.

Das Scone-Plugin *HyperScout* [WL00, WOL03] erweitert die Darstellung von Links in Web-Browsern um ein Popup-Fenster mit zusätzlichen Informationen zum verlinkten Dokument. Dabei wird das Konzept der typisierten Links aus früheren Hypertextsystemen aufgegriffen. Angezeigt werden eine Reihe von semantischen Informationen, wie Titel, Autor, Inhalt und Sprache des Dokuments. Überdies wurden Benutzungsinformationen, z.B. das Datum des letzten Besuchs, Hinweise auf durch den Link ausgelöste Aktionen, Informationen zum Link-Status und Infos zur Topologie des Hypertextes in die Darstellung einbezogen. Der Benutzer kann dann unter Einbeziehung der angezeigten Informationen die Entscheidung treffen, einen Link zu besuchen und so, das ist die hinter dem Projekt HyperScout stehende Hypothese, schneller und zielgerichteter relevante Informationen finden. Das Beispiel einer solchen Visualisierung ist in Abbildung 2.9 zu sehen.

Das Scone-Plugin HyperScout integriert in alle den Proxy-Server durchlaufenden HTML-Dokumente ein unsichtbares Applet, welches für die Einblendung der Popup-Fenster zuständig ist. Sämtliche Links werden um JavaScript Code ergänzt, welcher den Aufruf einer Funktion des Applets beinhaltet, die ein Popup-Fenster über dem Link darstellt, sobald der Mauszeiger über dem Link plaziert wird. Die im Popup dargestellten In-

formationen werden durch eine ins Scone-Framework integrierte Komponente ermittelt, welche mit dem Applet über eine Socket-Verbindung kommuniziert. Diesbezüglich ist die Architektur des HyperScouts mit der *UserTracking*-Komponente des Scone-Frameworks vergleichbar. Die HyperScout-Architektur wurde mit dem Ziel entworfen, verschiedene Erweiterungen der Link-Schnittstelle des Browsers einfach implementieren und evaluieren zu können.

Viele durch den HyperScout dargestellte Informationen setzen voraus, daß das verlinkte Dokument bereits geladen wurde. Frühe Versionen dieses Scone-Plugins benutzen dazu ausschließlich die in den Scone-Caches vorhandenen Daten, die auf vorhergehende Besuche der jeweiligen Sites zurückzuführen sind. Durch die Integration des Robots in das Scone-Framework kann nun der HyperScout auch aktiv im Hintergrund Dokumente aus dem Web laden. Anhand einer Heuristik wird entschieden, welche der verlinkten Dokumente durch den Robot geladen werden sollen. Einige Dokumente werden per `Http`-Befehl `GET` angefordert, wobei nur der Anfang der Datei geladen wird. Von anderen URLs wird nur durch das Kommando `HEAD` der `HTTP-Request-Header` angefordert. Dadurch kann sehr schnell die Validität eines Links überprüft und Informationen zu Dateigröße und Mime-Type gewonnen werden.

Als Teil des in dieser Diplomarbeit entwickelten Prototyps wurde das HyperScout-System um die Einblendung der individuell berechneten Vorhersage einer Bewertung des Zielobjekts eines Links ergänzt. Die flexible Architektur des HyperScouts machte eine derartige Erweiterung des Konzepts relativ problemlos möglich. Die Kommunikation zwischen dem HyperScout-Applet und der im Scone-Framework angesiedelten Komponente erfolgt über ein einfaches Protokoll, über welches sich nahezu beliebige im Popup-Fenster darzustellende Attribute übermitteln lassen. Lediglich die Übermittlung des Namens des den Link enthaltenden Fensters mußte bei den vom Applet an die Scone-Komponente gerichteten Anfragen ergänzt werden. Der *CoInternet-Client* verwaltet eine Liste der geöffneten Fenster zusammen mit den jeweils gültigen Benutzerprofilen. Aufbauend auf diesen Angaben wird dann eine für den Benutzer individuelle Vorhersage der erwarteten Bewertung einer Seite bestimmt.

## 3. Herausforderungen bei der Benutzung des World Wide Webs

Wie aus der repräsentativen ARD/ZDF-Online-Studie 2003 [EGF03] hervorgeht, nutzt mittlerweile die Hälfte der deutschen Bevölkerung ab 14 Jahren das Internet. Die aktuelle Zahl von 34,4 Millionen Nutzern im Jahr 2003 entspricht einem Zuwachs von 22 Prozent gegenüber dem Vorjahr. Zum Vergleich: 1997 waren nur 4,1 Millionen Deutsche online. Das Internet, genauer gesagt der Dienst World Wide Web, hat sich somit zu einem etablierten Kommunikations- und Informationsmedium entwickelt. Die beeindruckende Entwicklung dieses Mediums darf jedoch nicht darüber hinwegtäuschen, daß Benutzer des Webs mit einer Reihe von Schwierigkeiten konfrontiert werden, welche eine angenehme, effiziente und sichere Nutzung erschweren.

Probleme bei der Nutzung des Webs sind nur teilweise auf Unzulänglichkeiten der Browseranwendungen oder Web-Sites zurückzuführen. Vielmehr handelt es sich auch um Probleme, die dem Web als verteilter Hypertextanwendung ohne zentrale Kontrollinstanz immanent sind. Benutzer haben Schwierigkeiten sich in Hypertexten zu orientieren (siehe Abschnitt 3.1). Die niedrigen Zugangsbarrieren der Publikation im Web haben zu einer enormen Menge abrufbarer Informationen geführt, in der es immer schwerer wird, relevante Inhalte aufzufinden (3.2). Die online angebotenen Inhalte unterscheiden sich gravierend in ihrer Qualität (3.3). Dabei obliegt es dem Benutzer aus den oft umfangreichen Trefferlisten der Suchmaschinen, qualitativ hochwertige Inhalte manuell zu selektieren. Nicht jeder Anbieter von Informationen, Waren oder Dienstleistungen im Web ist vertrauenswürdig (3.4). Im abschließenden Abschnitt dieses Kapitels werden verschiedene Strategien vorgestellt, wie den genannten Problemfeldern begegnet werden kann (3.5).

### 3.1. Lost in Hyperspace

Viele Benutzer haben bei der Suche nach Informationen im World Wide Web die frustrierende Erfahrung gemacht, in komplexen Hypertexten den Überblick verloren zu haben. Hat ein Benutzer ein interessantes Dokument gefunden und „surft“ dann ein paar Seiten weiter, dann kann es leicht passieren, daß er diese Seite nicht mehr wiederfindet. Das Phänomen, daß Anwender dazu neigen, in Hypertexten die Orientierung zu verlieren,

### 3. Herausforderungen bei der Benutzung des World Wide Webs

---

wurde bereits in frühen Hypertext-Systemen beobachtet und wird in der der Literatur mit der Phrase „Lost in Hyperspace“ bezeichnet. Eine Herausforderung der Benutzung von Hypertext-Systemen liegt darin, neben der Aufgabe, den Inhalt der Dokumente zu erfassen und orientiert an den eigenen Aufgaben zu verarbeiten, auch noch die zur weiteren Navigation notwendigen Informationen zur erkundeten Struktur des Hypertextes im Kurzzeitgedächtnis zu behalten. Viele Anwender überfordert dies, ein Umstand, der als „Cognitive Overhead“ bezeichnet wird ([Con87] nach [WBL03]).

Andere Informationsquellen, wie die Printmedien, sind aus Benutzersicht um einiges einfacher zu handhaben, da die Informationen in einer linearen Folge von Seiten angeordnet sind, Seitenzahlen und Kopfzeilen einen Überblick über die Position im Dokument geben und sich der Gesamtumfang des vor einem liegenden Werks abschätzen läßt.

Kennzeichnendes Merkmal des Webs ist die Verknüpfung von Dokumenten mittels Hyperlinks. Durch diese stehen die Hypertextdokumente zueinander in einer n:m-Relation, denn eine Seite enthält Links zu m anderen Dokumenten und wird wiederum von n Seiten verlinkt. Der Benutzer sieht jedoch immer nur die von einem Dokument ausgehenden Links. Er nimmt somit nur einen kleinen Teil der Struktur wahr, in die ein Dokument eingebettet ist. Dem Benutzer fehlt es also an Überblick, zumal er immer nur die direkte Linkumgebung einer Seite wahrnimmt, ohne z. B. zu sehen, ob ein Link in eine Sackgasse führt oder ob er den Startpunkt einer linear angeordneten Folge mehrerer thematisch zusammenhängender Dokumente markiert. Die grafische Darstellung von Links beschränkt sich auch bei den heutigen Browsern in der Regel auf die Kennzeichnung durch Unterstreichung und die Markierung von bereits besuchten Links. Es gibt keine Informationen darüber, ob ein verlinktes Dokument überhaupt erreichbar ist, ob es auf einen externen Server verweist oder ob eine große Datei verlinkt wurde. Zur Unterstützung der Navigation bieten die Browser meist nur einen History-Mechanismus und eine Bookmark-Funktion, die es ermöglicht, Links zu bestimmten Seiten zu hinterlegen. Ein gängiges Benutzerverhalten bei der Suche im Netz ist es, von einer Seite ausgehend jeweils ein paar Links tief in die Hypertextstruktur hinabzusteigen, dann mittels des Back-Buttons wieder zur Ursprungsseite zurückzukehren und die nächsten Links zu besuchen. Da die History gemeinhin als Stack implementiert wird, verschwinden alle vorher besuchten Seite aus der History, wenn der Benutzer wieder zur Ursprungsseite gelangt.

Die Hyperlinks zwischen Web-Dokumente können als Ausdruck einer semantischen Relation zwischen den verknüpften Dokumenten interpretiert werden. Im Gegensatz zur Navigation in der räumlich strukturierten realen Welt, kann man sich im Web nicht Wege durch Richtungsangaben einprägen, sich an Landmarken orientieren oder mittels Karten den Weg suchen.

Erschwerend bei der Navigation im Web ist es ferner, daß das Medium durch ein hohes Maß an Inhomogenität gekennzeichnet ist. Dies gilt zum einen bezüglich der verschie-

denen Typen von Medien, die durch Hyperlinks verknüpft werden können, und zum anderen bezüglich der Darstellung der Web-Dokumente. Die Sprache HTML läßt bei der Strukturierung und Gestaltung der Dokumente viele Freiheiten. Die konkrete Darstellung der Seiten ist plattform- bzw. browserabhängig. Verbindliche Style Guides, die beispielsweise die Integration bestimmter Navigationshilfen vorschreiben, gibt es nicht.

## 3.2. Information Overload

Die Möglichkeit, schnell an Informationen zu gelangen, wird vielfach als primäres Motiv für die Anschaffung eines Internet-Zugangs genannt [C<sup>+</sup>03]. Eine zu großes Angebot an Informationen kann jedoch auch zum Problem werden, insbesondere dann, wenn die Werkzeuge, die Anwender bei den Recherchen unterstützen, mit dieser Entwicklung nicht Schritt halten können. Teilweise wirkt die Informationsflut des Internet auch abschreckend. 56 Prozent der im Rahmen der ARD/ZDF-Online-Studie befragten Personen ohne Netzzugang stimmten der Aussage zu, daß die Informationsflut des Internets nicht mehr bewältigt werden könne [GM03].

Fünf Millionen Terrabyte an Daten wurden im Jahr 2002 weltweit gespeichert. Diese Schätzung wurde in der von der University of Berkeley durchgeführten Studie „How much Information“ veröffentlicht [LV03]. Die Datenmenge der im Web abrufbaren Dokumente wird in dieser Studie mit 170 Terrabyte beziffert. Das entspricht der siebzehnfachen Größe der in der amerikanischen Library of Congress gesammelten Dokumente. Zusätzlich ist noch die Informationsmenge zu berücksichtigen, welche in Datenbanken gespeichert ist und über dynamisch generierte Web-Seiten abgerufen werden kann. Knapp 45 Millionen Web-Sites sind nach Angaben der Firma Netcraft<sup>1</sup> im November 2003 online. Als dieses Unternehmen vor acht Jahren die erste Statistik zur Zahl der Web-Server veröffentlichte, waren es lediglich 19.000 Sites.

Diese Statistiken zur Größe des Webs sollen verdeutlichen, in welchem Mißverhältnis die Menge der Informationen im Web zu den Fähigkeiten der Benutzer steht, Dokumente zu sichten. Suchmaschinen wie Google<sup>2</sup> oder Alltheweb<sup>3</sup> sind neben redaktionell gepflegten Katalogen zu einem unverzichtbaren Ausgangspunkt von Recherchen im Web geworden. Die Leistung der Suchmaschinen ist aus Anwendersicht vielfach als nicht zufriedenstellend zu beurteilen. Suchmaschinen, welche die Informationen des Webs mittels Robots indexieren und auf Grundlage des so erstellten Indexes Abfragen mit einer Liste von Verweisen auf relevante Quellen im Web beantworten, sind den Information-Retrieval-Systemen zuzuordnen. Anhand der Kenngrößen Recall und Precision [KT00] sollen nachfolgend einige der Schwächen heutiger Suchmaschinen betrachtet werden.

---

<sup>1</sup>[www.netcraft.com](http://www.netcraft.com)

<sup>2</sup>[www.google.de](http://www.google.de)

<sup>3</sup>[www.alltheweb.com](http://www.alltheweb.com)

### 3. Herausforderungen bei der Benutzung des World Wide Webs

---

Die Größe des Webs macht eine vollständige Indexierung nahezu undurchführbar. Somit enthält die Datenbasis der Suchmaschinen nur einen Ausschnitt der im Web vorhandenen Dokumente, was sich nachteilig auf die Recall-Rate der Trefferlisten auswirkt. Die hohe Dynamik des Webs führt dazu, daß die Indizes sehr schnell veralten. Recall und Precision der Suchergebnisse werden beeinträchtigt, wenn z. B. neu erstellte Seiten in den Ergebnissen nicht berücksichtigt werden oder auf geänderte bzw. gelöschte Dokumente verwiesen wird.

Zu manchen Themen ist die Zahl der Dokumente im Web so groß, daß Suchmaschinen den Anwender mit Trefferlisten konfrontieren, deren vollständige Durchsicht nicht mehr durchführbar ist. Der Qualität der Ranking-Algorithmen kommt somit eine besondere Bedeutung zu. Eine vordere Platzierung in den Ergebnislisten der Suchmaschinen ist für Inhalteanbieter, die sich in einem harten Wettbewerb um die Aufmerksamkeit der „Surfer“ befinden, äußerst erstrebenswert. Die Dienstleistung, mittels Techniken wie „Cloaking“<sup>4</sup> und „Link-Farmen“<sup>5</sup> Suchmaschinen zu manipulieren und den Sites der Auftraggeber Nutzer zuzuleiten, hat sich zu einem florierenden Geschäft entwickelt [Dwo03, Kar03]. Die Qualität der Suchergebnisse nimmt gerade bei Anfragen zu kommerziell relevanten Themen tendenziell ab.

Ein weiterer Aspekt, weshalb die Suchmaschinen bei der Selektion bezüglich einer Anfrage relevanter Inhalte unzureichend funktionieren, liegt darin begründet, daß eine Informationsverarbeitung auf semantischer Ebene nicht ausgeführt wird. Faktoren, wie die Qualität der Dokumente können bei der Bestimmung der Trefferrelevanz nicht einbezogen werden. Ebenso wenig werden Suchanfragen bei den aktuellen Suchmaschinen unter Berücksichtigung der individuellen, langfristigen Interessen der Benutzer ausgewertet. Sucht jemand bei Google nach dem Wort „apple“, so listet die Suchmaschinen 17.200.000 Treffer<sup>6</sup>, wobei in den ersten Seiten der Trefferliste nahezu ausschließlich auf das IT-Unternehmen bezogene Seiten aufgelistet werden. Für Personen, welche an Landwirtschaft und gesunder Ernährung interessiert sind, wäre die Suchmaschine bei einer solchen Anfrage nur dann nützlich, wenn es möglich wäre, mit vertretbarem Aufwand persönliche Präferenzen zum Ausdruck zu bringen, welche dann zur Personalisierung der Suchergebnisse verwendet werden würden (siehe auch [Joh03]).

---

<sup>4</sup>Anhand des `UserAgent`-Feldes der HTTP-Requests erkennt die Web-Site Robots und präsentiert eine Aufzählung von Stichwörtern. Menschliche Besucher werden jedoch zu den Sites der Auftraggeber weitergeleitet.

<sup>5</sup>Ein Netz von verlinkten Sites wird geschaffen um den PageRank-Algorithmus zu manipulieren.

<sup>6</sup>Ausprobiert am 9. November 2003

### 3.3. Qualität im Web

Die technischen und ökonomischen Hürden der Publikation im Web sind äußerst gering. Bei der ungleich aufwendigeren Verbreitung von Print-Publikationen übernimmt das Verlagswesen eine Art Mittlerrolle zwischen Autoren und Konsumenten. Es garantiert auch vor dem Hintergrund der Refinanzierbarkeit der Publikationen die Einhaltung von Qualitätsstandards. Im Web fehlt eine derartige Kontrollinstanz. Die Qualität der Angebote variiert folglich sehr stark. Dieses ist insbesondere deshalb problematisch, da die Kompetenzen der Nutzer vielfach nicht ausreichend sind, um die Qualität der Informationen, z. B. deren Richtigkeit, zu beurteilen.

Es existieren verschiedene Ansätze, die Evaluation der Qualität von Web-Sites anhand von Kriterienkatalogen zu systematisieren. Für die weitere Betrachtung erscheint es sinnvoll, Qualitätskriterien in die Klassen Technik, Inhalt und Präsentation zu gruppieren. Technische Kriterien wären beispielsweise die zuverlässige Erreichbarkeit der Site, die Performanz der Server, die Konformität mit den Standards des W3C<sup>7</sup> oder die Validität der Hyperlinks. Bei der Beurteilung der Präsentation können ergonomische und ästhetische Kriterien einbezogen werden. Inhaltliche Anforderungen an Web-Dokumente sind u. a. Vollständigkeit und Richtigkeit der Informationen, Nachvollziehbarkeit der Quellen, Aktualität und Zielgruppenaffinität. Verallgemeinernd ist die inhaltliche Qualität von online-Informationen positiv zu beurteilen, wenn gemäß den professionellen Maßstäben der jeweiligen Domäne gearbeitet wurde.

Gerade die Qualitätskriterien der Klassen Inhalt und Präsentation sind teilweise sehr subjektiv und setzen weitergehende Kenntnisse der jeweiligen Domäne voraus, beispielsweise bei der Überprüfung der Richtigkeit von Informationen. Eine vollautomatische Beurteilung der Qualität von Web-Angeboten ist folglich nicht realisierbar.

Since there is no way for computers to automatically measure quality, we have to rely on human judgment for web quality rating [Nie98].

Redaktionell betreute Verzeichnisse wie das „Open Directory Project“<sup>8</sup> haben gegenüber den Suchmaschinen den Nachteil einer deutlich kleineren Datenbasis. Ihre Stärke ist jedoch, daß durch die Einbeziehung des Urteilsvermögens der Redakteure Qualitätskriterien berücksichtigt werden können.

---

<sup>7</sup> [www.w3.org](http://www.w3.org)

<sup>8</sup> [www.dmoz.org](http://www.dmoz.org)

### 3.4. Vertrauen im Web

Nur 50,6 Prozent der in der 2003 veröffentlichten Studie „Surveying the Digital Future“ [C<sup>+</sup>03] befragten Personen sagten, daß die Informationen im Web exakt und glaubwürdig seien. In der Untersuchung des Vorjahres waren es noch 56 Prozent.

While nearly three-quarters of users consider the Internet to be a very important or extremely important source of information – a ranking higher than for books, television, radio, newspapers, or magazines – only half of users believe that most or all of the information online is reliable and accurate. Even worse, more than one-third of users say that only about half of the information they find online is reliable and accurate [C<sup>+</sup>03].

Dieses Mißtrauen erscheint, insbesondere wenn es darum geht, kritische Fragen durch Internet-Recherchen zu beantworten, durchaus angebracht. Piero Impicciatore et al. untersuchten die Korrektheit online veröffentlichter medizinischer Informationen [IPCB97]. Nur wenige der untersuchten Sites boten vollständige und akkurate Informationen, so ihr erschreckendes Fazit.

Naturgemäß besteht zwischen den Lesern von Web-Seiten und deren Autoren eine Wissensdifferenz, denn vielfach ist es gerade die Motivation für die Benutzung des Webs, Wissenslücken zu schließen oder Neuigkeiten zu erfahren. Somit fällt es vielen Anwendern schwer, die Qualität von Informationen zu beurteilen, sowie Glaubwürdigkeit und Kompetenz der Autoren einzuschätzen. Ferner können Nutzer nicht nachvollziehen, ob neutral oder vor dem Hintergrund einer bestimmten Gesinnung berichtet wird. Die Entscheidung einer online-Information zu glauben, ist eine Vertrauensentscheidung. In Ermangelung der Möglichkeiten mit anderen Nutzern zu kommunizieren und Reputationsinformationen auszutauschen, ist den Autoren von Web-Dokumenten vielfach „blindes Vertrauen“ entgegenzubringen.

Den Vorteilen, die das Netz für die Abwicklung von Geschäften bietet, wie die Überwindung räumlicher und zeitlicher Restriktionen bei der Wahl von Geschäftspartnern und die hohe Transparenz der Märkte, stehen gravierende Nachteile bezüglich der Kalkulierbarkeit der Risiken entgegen. Die Vertrauenswürdigkeit und Zuverlässigkeit von Transaktionspartnern, welche meist nur durch die von ihnen gestaltete Darstellung ihrer Web-Site bekannt sind, läßt sich nur schwerlich einschätzen. Insbesondere grenzüberschreitende Geschäfte sind mit den Unsicherheiten einer vermeintlich unklaren Rechtslage belastet. Zusätzlich sind viele online Käufer um die Wahrung des Datenschutzes im Umgang mit ihren persönlichen Daten und die Sicherheit ihrer Kreditkarteninformationen besorgt [C<sup>+</sup>03]. Unabhängige Reputationssysteme (siehe Abschnitt 2.2.3) könnten

einen wesentlichen Beitrag dazu leisten, Vertrauen zwischen Geschäftspartnern im Internet herzustellen.

Internetnutzer scheinen sich vorwiegend den Web-Sites bekannter Marken zuzuwenden. Das Markenimage etablierter Anbieter wird auf deren Internet-Auftritte übertragen [EGF03]. Beispielsweise informieren sich Nutzer in der Erwartung, die Qualität des gedruckten SPIEGELS lasse auf die Güte des online-Angebot schließen, bevorzugt bei SPIEGEL online<sup>9</sup>, statt Quellen aus dem Bereich Peer-to-Peer-Journalismus Beachtung zu schenken. Konsumenten kaufen lieber online beim vertrauten Otto-Versand<sup>10</sup>, als einen kleinen, unbekanntem, aber möglicherweise preisgünstigeren, Online-Shop zu berücksichtigen. Der PageRank-Algorithmus [PBMW98] der Suchmaschine Google verstärkt diesen Trend zu großen Anbietern tendenziell noch. Bei der Suche nach Produkten werden meist zuerst die Seiten der Hersteller und großer Händler gelistet.

Die Vielfalt der Informationen im Web resultiert insbesondere aus den vielen sehr kleinen Sites mit hoch spezialisierten Angeboten. Eventuell würden diese mehr Beachtung erfahren, wenn durch Systeme zur direkten und indirekten Kommunikation der Web-User Schwierigkeiten in der Einschätzung der Qualität der Inhalte und der Vertrauenswürdigkeit der Anbieter gemindert werden würden.

### 3.5. Tools zur Unterstützung der Navigation im Web

Ein Ansatz das Web benutzbarer zu machen ist es, die Benutzungsschnittelle der Web-Browser weiterzuentwickeln oder den Funktionsumfang der Browser zu erhöhen bzw. um weitere clientseitig ablaufende Anwendungen zu ergänzen. Visualisierungen helfen den Benutzern bei der Orientierung und Navigation in Hypertexten. Persönliche Agenten sind als ein Beispiel zu nennen, wie Anwender bei der Aufgabe, die Informationsflut des Webs zu bewältigen, unterstützt werden können [Wol02].

Benutzbarkeit ist ein wichtiges Differenzierungsmerkmal zwischen Web-Sites. Personalisierung oder der Einsatz von räumlichen Metaphern bei der Strukturierung von Informationsräumen sind exemplarisch für Bestrebungen zu nennen, serverseitig die Benutzung des Webs zu erleichtern.

Suchmaschinen sind als die zentralen Ausgangspunkte von Recherchen im Web von besonderer kommerzieller Bedeutung. Das Besucherinteresse und damit der wirtschaftliche Erfolg einer Suchmaschine wird maßgeblich durch die Qualität der Suchergebnisse bestimmt. Demgemäß ist das Thema der Optimierung von Suchmaschinentechnologien Gegenstand intensiver Forschungsaktivitäten.

---

<sup>9</sup> [www.spiegel.de](http://www.spiegel.de)

<sup>10</sup> [www.otto.de](http://www.otto.de)

### 3. Herausforderungen bei der Benutzung des World Wide Webs

---

Während die vorstehend aufgeführten Ansätze darauf abzielen, die Anwender durch Software zu unterstützen bzw. die Schnittstelle zwischen Mensch und Computer zu optimieren, wird in dieser Diplomarbeit ein anderer Weg verfolgt. Ziel ist es, die Kollaboration zwischen den Nutzern des Webs, vermittelt durch ein Softwaresystem, zu ermöglichen und zu fördern. Die Navigation durch direkte oder indirekte Interaktion mit anderen Anwendern bzw. die Orientierung an anderen Nutzern wird als *Social Navigation* (siehe Kapitel 4) bezeichnet. In diesem Kontext ist das Empfehlungs- und Annotationssystem für Web-Seiten zu sehen, welches im praktischen Teil der Arbeit realisiert und evaluiert wurde.

Trotz mangelnder technischer Unterstützung durch die heutigen Web-Browser sind den Benutzern kollaborative Vorgehensweisen bei der Navigation im Web nicht ganz fremd. In der ARD/ZDF-Online-Studie 2003 wurde die Frage nach Wegen, auf Internetseiten aufmerksam zu werden, gestellt. 67 Prozent der Teilnehmer nannten persönliche Empfehlungen aus dem Bekannten- und Freundeskreis als einen Zugangsweg [EGF03].

Social Navigation kann, z. B. durch Guided Tours, individuelle Probleme bei der Orientierung im Web mindern, wenn die Aufgabe der Orientierung auf andere Nutzer übertragen wird oder die Zahl der Navigationsalternativen basierend auf den Wegen anderer Nutzer mit ähnlichen Interessen eingeschränkt wird.

Das Potential von Social Navigation ist in Bezug auf die genannten Schwierigkeiten bei der Benutzung des Webs insbesondere darin zu sehen, daß Anwender bei der Beurteilung der Qualität von Inhalten und der Vertrauenswürdigkeit von Anbietern auf die Erfahrungen anderer Nutzer zurückgreifen können. Das in dieser Arbeit entwickelte Empfehlungs- und Annotationssystem hilft den Anwendern dabei, mit den Informationsmengen des Webs umzugehen. Durch den Einsatz von Social-Navigation-Techniken werden dabei die Kriterien persönliches Interesse, Qualität und Vertrauen für die (automatische) Informationssuche erschlossen.

## 4. Social Navigation

Im Kartenlesen ungeübte Autofahrer, die in einer fremden Stadt eine bestimmte Adresse aufsuchen müssen, neigen dazu, lieber ortskundige Passanten nach dem Weg zu befragen, als einen Stadtplan zu konsultieren ([SVW85] nach [HBM03]). Bei einer Wanderung im verschneiten Wald können wir uns möglicherweise an den Spuren im Schnee orientieren. Das abgegriffene Theorie-Buch in der Bibliothek ist sicherlich gegenüber einem alten, aber offensichtlich nur selten ausgeliehenen Buch eine gute Wahl.

Unter dem Stichwort *Social Navigation* [DC94, DDH<sup>+</sup>00] werden in der Literatur Verhaltensweisen bei der Navigation zusammengefaßt, die dadurch gekennzeichnet sind, daß Benutzer Navigationsentscheidungen basierend auf der direkten oder indirekten Interaktion mit anderen Anwendern treffen. Des weiteren bezeichnet der Ausdruck Social Navigation Designansätze für Informationssysteme, die es jeweils auf unterschiedliche Art und Weise zum Ziel haben, den Benutzern soziales Navigieren als eine Option der Erschließung der angebotenen Informationen, zur Verfügung zu stellen.

Menschen sind soziale Wesen. Das Potential, welches in der Fähigkeit gemeinschaftlich zu agieren liegt, wird bislang in Informationssystemen, wie dem World Wide Web, kaum erschlossen. Das Ziel von Social Navigation ist es, das Navigationsverhalten anderer Benutzer derselben Umgebung für die eigene Navigation und Orientierung zu nutzen [Die03].

Beim Navigieren mit anderen Personen zu kommunizieren bzw. sich an ihnen zu orientieren, ist in der realen Welt, wie vorstehende Beispiele veranschaulicht haben, eine natürliche Vorgehensweise und es erscheint auf den ersten Blick als sehr lohnend, sich mit der Abbildung dieses Phänomens von der realen Welt auf die virtuelle zu beschäftigen. Virtuelle Welten unterscheiden sich jedoch in vielerlei Hinsicht grundlegend von der realen Umgebung. So weist z. B. das Web nicht die Struktur eines euklidischen Raumes auf, in dem sinnvoll von Entfernung und Richtungsangaben gesprochen werden kann und die non-verbale Dimension geht in textbasierter online-Kommunikation verloren. Es ist folglich bei der Übernahme von Metaphern aus der realen Welt Vorsicht geboten, gewählte Designansätze sollten sorgfältig evaluiert werden.

Wie in Kapitel 3 ausgeführt wurde, ist die Benutzung des World Wide Webs mit einer Reihe von Problemen behaftet. In diesem Kapitel soll *Social Navigation* als ein Ansatz vorgestellt werden, mit dem einige der aufgeführten Probleme gemindert werden können.

In Abschnitt 4.1 wird zunächst versucht, zu präzisieren, was im Kontext dieser Diplomarbeit unter *Navigation in Informationsräumen* zu verstehen ist. In den nachfolgenden Abschnitten wird dann Social Navigation definiert und Modelle zu Klassifikation verschiedener Formen von Social Navigation werden vorgestellt. Die Abschnitte 4.4 und 4.5 beschäftigen sich weitergehend mit Beispielen direkter und indirekter Social Navigation, insbesondere sollen dabei Bezüge zum sozialen Navigieren im World Wide Web hergestellt werden. Theoretische Überlegungen zum sozialen Navigieren bilden den Abschluß dieses Kapitels. In Abschnitt 4.6 werden Voraussetzung für und Randbedingungen von Social Navigation betrachtet. Anschließend werden die Vorteile (4.7) der Integration von Social Navigation in Online-Anwendungen erörtert sowie Probleme und offene Fragestellungen (4.8) beim Entwurf von Social Navigation unterstützenden Informationssystemen aufgezeigt.

### 4.1. Navigation in Informationsräumen

Ziel dieses Abschnittes ist es, als Voraussetzung für die Beschäftigung mit dem Konzept Social Navigation, zunächst einmal zu vermitteln, was im Rahmen dieser Arbeit unter einen Informationsraum und der Navigation in diesem verstanden wird.

Der Begriff Informationsraum legt die Assoziation einer räumlichen Strukturierung von Informationen nahe. Das Konzept des Informationsraumes umfaßt jedoch auch semantisch struktuierte Sammlungen von Informationen, wie die durch Hyperlinks verknüpften Dokumente des World Wide Webs. Informationsräume müssen nicht notwendigerweise virtuell sein. Reale Räume, wie beispielsweise ein Flughafengebäude, in dem Hinweisschilder und Karten Fluggästen den Weg weisen sollen, können ebenfalls als Informationsräume betrachtet werden. Die Bedeutung des Begriffs Informationsraum umfaßt beliebig strukturierte Räume, welche als Träger von Informationen dienen können [BH97].

In einer ersten Definition bezeichnet Svensson [Sve00] Navigation als die Aktivität, von einem Ort zum anderen zu gelangen. Diese offensichtliche Bedeutung von Navigation soll im folgenden als Wegfindung bezeichnet werden, um eine Differenzierung zu anderen Typen von Navigation vornehmen zu können. Charakteristisch für das Wegfinden ist die klare Zielvorstellung der navigierenden Person. Die Tätigkeit des Wegfindes besteht aus einer Reihe von Teilaufgaben [DS73]: Zunächst einmal ist es erforderlich, sich in der Umgebung zu orientieren, um die eigene Position zu bestimmen. Nachfolgend ist die korrekte Route zum Ziel zu wählen. Während der Bewegung ist fortwährend die aktuelle Position mit der geplanten Route abzugleichen und zu überprüfen, ob das Ziel schon erreicht wurde. Diese vier Teilaufgaben werden in Anlehnung an [Sve00] um die Aufgabe, eine Zielvorgabe zu formulieren, ergänzt. Navigierende Personen stützen ihre Handlungen auf das kognitive Modell ihrer Umwelt [DS73]. Dieses kann auf der eigenen

Wahrnehmung basieren, mit Hilfe von Navigationshilfen wie Karten entstanden sein oder aber ein Resultat der Kommunikation mit anderen Personen sein.

Die Navigation in einem Raum ohne eindeutige Zielvorstellungen, jedoch mit der Motivation den Raum näher kennenzulernen, also die Art und Position der vorhandenen Objekte zu ergründen, wird als Erkundung oder in der Hypertext-Terminologie als *Browsing* bezeichnet. Ist bei der Navigation weder die Position noch die Route zu den bei der Bewegung gefundenen Objekte relevant, sondern steht die Betrachtung der Objekte des Raumes im Vordergrund, dann wird dieser Typ von Navigation als *Identifikation und Klassifikation von Objekten* bezeichnet.

Das Navigieren in Informationsräumen umfaßt gemäß den vorstehenden Begriffsbestimmungen eine große Bandbreite von Tätigkeiten in sehr unterschiedlichen Umgebungen. Insbesondere die Recherche im World Wide Web und alle damit verbundenen Teilaufgaben, wie die Formulierung von Zielvorgaben und die Orientierung im Hypertext, werden der Tätigkeit des Navigierens zugeordnet und sollen durch Social Navigation vereinfacht werden.

## 4.2. Definitionsansätze von Social Navigation

Verschiedene Autoren haben in der Vergangenheit unterschiedliche Aspekte des Konzepts Social Navigation betont, so daß nicht ohne weiteres eine allgemeingültige Definition gegeben werden kann. Zunächst einmal werden die unterschiedlichen Ausprägungen des Begriffs Social Navigation dargestellt. Anschließend wird der Vorgang des sozialen Navigierens näher betrachtet und eine Definition gegeben. Die charakteristischen Merkmale von Social Navigation werden im Anschluß anhand von Beispielen erarbeitet.

Um zu einer Definition von Social Navigation zu gelangen, wird zunächst ein Überblick über die verschiedenen Bedeutungen dieses Begriffs gegeben. In unterschiedlichen Definitionen wurden jeweils verschiedene Aspekte von Social Navigation betont. Dennoch sind grundlegende Übereinstimmungen in den unterschiedlichen Definitionsversuchen zu erkennen, so daß eine Definition formuliert werden kann, welche die verschiedenen Beispiele von Social Navigation beinhaltet und eine Abgrenzung von anderen Formen von Navigation möglich macht.

Der Begriff *Social Navigation* wurde von Dourish und Chalmers geprägt [DC94]. Sie differenzieren zwischen räumlicher bzw. semantischer Navigation einerseits und sozialer Navigation andererseits. In räumlicher Navigation bewegen sich Personen von einem Objekt zu einem anderen, weil diese in einer räumlichen Beziehung zueinander stehen. Semantische Navigation wird davon gesteuert, daß zwischen den Objekten eine semantische Beziehung besteht, die jedoch auf räumliche Beziehungen abgebildet werden kann.

Demgegenüber kann in Räumen, die kollaborative Aktivitäten unterstützen, von einer dritten Form der Navigation gesprochen werden. Navigationshandlungen können aufgrund der Wahrnehmung von Aktivitäten anderer Personen geschehen.

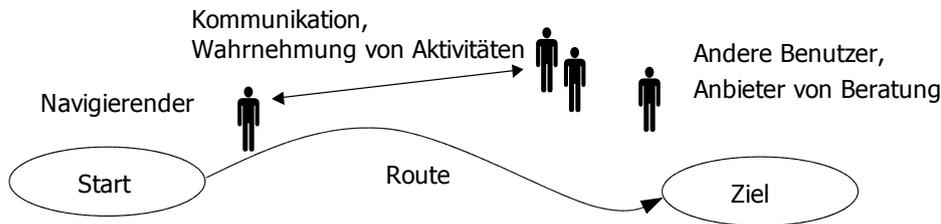
In social navigation, movement from one item to another is provoked as an artefact of the activity of another or a group of others. So, moving „towards“ a cluster of other people, or selecting objects because others have been examining them would both be examples of social navigation [DC94].

Erickson [Eri96] betont, daß viele Benutzer des Webs auch zugleich Autoren sind, die auf ihren persönlichen Homepages Listen mit Verweisen zu von ihnen, gemäß ihren individuellen Interessen und Qualitätsmaßstäben, ausgewählten Seiten veröffentlichen. Aus dieser Sichtweise interpretiert Erickson das Web als einen sozialen Hypertext, in dem persönliche Web-Sites als eine Art Selbstportrait ihre Autoren anzusehen sind und Links einen Ausschnitt der soziale Netzwerke der Autoren repräsentieren. Die Link-Listen persönlicher Seiten ermöglichen ein Vorgehen bei der Recherche im Web, daß sich an Personen orientiert, die über Kenntnisse auf dem entsprechendem Gebiet verfügen, oder in deren Umfeld sich wahrscheinlich Personen mit dem passenden Wissen befinden. Vermittelt durch persönliche Web-Sites kann auf die Empfehlungen von Experten zurückgegriffen werden, ohne diese direkt kontaktieren zu müssen. Die Strategie, bei der Lokalisierung von Informationen im Web das Wissen über Interessen, Fähigkeiten und Beziehungen von Personen einzubeziehen und entlang der Verweise persönlicher Homepages zu navigieren, stellt nach Erickson eine Form von Social Navigation dar.

Das Versenden von E-Mails mit Verweisen auf interessante, beim Browsen im Web gefundene Dokumente an Freunde und Kollegen ist ein typisches Beispiel dafür, was Dieberger [Die97] unter Social Navigation versteht. Er sieht in Social Navigation eine Form von Navigation, die auf Kommunikation und Interaktion mit anderen Nutzern basiert.

Um zu einer für diese Arbeit gültigen Definition von Social Navigation zu gelangen, soll zunächst einmal der Vorgang des sozialen Navigierens näher betrachtet werden.

In Social-Navigation-Handlungen sind grundsätzlich zwei Parteien involviert: Zum einen die im Informationsraum navigierende und nach Navigationsunterstützung suchende Person, zum anderen ein oder mehrere Anbieter von Beratung, welche sich ebenfalls in dem Informationsraum befinden. Dabei ist anzumerken, daß die Benutzer des Informationsraumes natürlich situationsabhängig die Rollen wechseln können, d. h. diejenigen Personen, die in der einen Situation Navigationsunterstützung anbieten, können in einer anderen Situation selbst Navigationsunterstützung verlangen. Charakteristisch für Social Navigation ist nun die Beeinflußung der Navigationshandlungen durch die Kommunikation zwischen der navigierenden Person und den Anbietern von Beratung bzw. durch



**Abbildung 4.1.:** Schematischer Ablauf von Social Navigation (Quelle: [For98])

die Wahrnehmung deren Aktivitäten (siehe Abbildung 4.1). Dabei können die beteiligten Parteien entweder direkt kommunizieren oder indirekt über die von den Anbietern von Navigationsunterstützung als Resultat ihrer Handlungen hinterlassenen Artefakte. Anbieter von Beratung können auch Agenten sein. In diesem Fall wird versucht dem Informationsraum eine Benutzungsoberfläche zu geben, die Social Navigation als eine Möglichkeit der Navigation in diesem Raum vorsieht.

Als Definition von Social Navigation soll nun die Definition von Svensson [Sve00] genutzt werden, welche die zu Beginn dieses Abschnitts aufgeführten Sichtweisen von Dourish [DC94], Erickson [Eri96] und Dienberger [Die97] umfaßt:

Navigation that is conceptually understood as driven by the actions from one or more advice providers [Sve00].

Zwei, an die Darstellung in [Sve00] angelehnte Beispiele sollen die charakteristischen Merkmale von Social Navigation veranschaulichen.

Eine Person, die einen Trampelpfad begeht, wählt eine bestimmte Streckenführung genau deshalb, weil viele andere Personen vorher ebenfalls diese Strecke gegangen sind. Dieses ist folglich gemäß vorstehender Definition ein Beispiel für soziales Navigieren. Nun könnte man argumentieren, daß auch das Verfolgen der Streckenführung einer Straße Social Navigation sei, denn die Straße ist schließlich das Produkt menschlicher Planung und Arbeit. Die Straße ist jedoch als ein immanenter Bestandteil des Raumes anzusehen. Die Streckenführung der Straße ist im Gegensatz zu der des Trampelpfades nicht das Produkt der vielfachen Begehung genau dieser Strecke, sondern das Resultat der Überlegungen von Städteplanern. Der Trampelpfad hat sich dynamisch durch die vielfache Begehung der selben Strecke herausgebildet. Wenn er lange Zeit nicht mehr begangen wird, verschwindet er langsam wieder. Bedenkt man jedoch, daß Trampelpfade zu Straßen ausgebaut werden können, dann wird deutlich, daß die Grenzen zwischen Social Navigation und anderen, an den gegebenen Strukturen und Navigationshilfsmitteln orientierten Formen von Navigation fließend sein können. In Umgebungen, die Social Navi-

gation ermöglichen, können die Benutzer durch ihr Handeln die Strukturen des Raumes verändern bzw. erweitern. Dieberger [DDH<sup>+</sup>00] schreibt dazu sehr pointiert:

In this way, social navigation is a closer reflection of what people actually do than it is a reflection of what designers think people should actually do [DDH<sup>+</sup>00].

Einem ankommenden Fluggast weisen Hinweisschilder den Weg zur Gepäckausgabe. Alternativ könnte er auch ein Mitglied des Bodenpersonals nach dem Weg fragen. Letztgenannte Situation ist eindeutig als Social Navigation zu interpretieren, der erstgenannte Fall nicht. Die Auskunft der Bodenpersonals ist individuell auf den ratsuchenden Fluggast zugeschnitten, während die Hinweisschilder als ein Bestandteil des Raumes anzusehen sind, welcher der navigierenden Personen keine individuellen Hinweise bietet.

Zusammenfassend läßt sich feststellen, daß die Navigation unter Zuhilfenahme der Produkte oder Spuren der Aktivitäten anderer, nur dann als Social Navigation zu bezeichnen ist, wenn entweder eine individuelle Ausrichtung der Navigationsunterstützung an die hilfesuchende Person vorliegt oder sich die Navigationsunterstützung dynamisch aus den Handlungen früherer Benutzer entwickelt hat.

### 4.3. Klassifikation verschiedener Formen von Social Navigation

Kennzeichnend für Social Navigation ist die Beeinflußung der Navigationshandlungen durch Kommunikation oder allgemeiner ausgedrückt Interaktion mit anderen Personen. Verschiedene Formen von Social Navigation lassen sich hinsichtlich des Kommunikationsmodus, den Randbedingungen der Kommunikation und den Eigenschaften der ausgetauschten Nachrichten differenzieren [For98].

Kennzeichnend für *direkte Social Navigation* ist bidirektionale, gemeinsame Kommunikation zwischen der navigierenden Person und dem Anbieter von Navigationsinformationen. In diesem Fall befinden sich der Navigierende und der Berater in einem Dialog. Der Berater geht direkt auf die Fragestellung des Navigierenden ein, stellt gegebenenfalls Rückfragen oder verweist auf andere Berater. Davon abzugrenzen ist die *indirekte Social Navigation*. Charakteristisch ist hierbei die unidirektionale Kommunikation zwischen den beteiligten Personen. Der Navigierende und der Anbieter von Beratung kommunizieren in diesem Fall nicht gemeinsam, sondern indirekt über Spuren und Artefakte, die der Anbieter von Beratung bewußt oder unbewußt im Raum hinterläßt und die von anderen Benutzern bei der Navigation wahrgenommen werden [Sve98].

Sowohl direkte als auch indirekte Social Navigation können *synchrone* und *asynchrone* Kommunikation beinhalten. Hinsichtlich des oben aufgeführten Merkmals der direkten Kommunikation macht es keinen Unterschied, ob man synchron in einem Chat

Link-Tips austauscht oder asynchron per E-Mail Hinweise zu interessanten Web-Seiten kommuniziert. Typische Beispiele indirekter Social Navigation, wie die Einbeziehung von Annotation von Dokumenten in Navigationsentscheidungen, basieren auf asynchroner Kommunikation. Vom Ergebnis her ist es gleichgültig, ob man sich an den Spuren, die andere Benutzer in der Vergangenheit hinterlassen haben orientiert, oder ob man einer oder einer Menge von Personen folgt, ohne sich über die gemeinsame Route und das Ziel zu verständigen. In beiden Situationen ist von indirekter Social Navigation zu sprechen. Der erstgenannte Fall basiert auf asynchroner Kommunikation, mediatisiert über die im Raum hinterlassenen Spuren. Im zweitgenannten Fall werden die Positionsänderungen synchron kommuniziert oder wahrgenommen.

Die Unterscheidung verschiedener Formen von Social Navigation bezüglich des Kommunikationsmodus mag ausreichend sein, um grundlegend verschiedenen Ausprägungen abzugrenzen. Das Phänomen Social Navigation wird jedoch noch von einer Reihe anderer Faktoren beeinflusst, die insbesondere hinsichtlich der Untersuchung der Auswirkungen der Integration einer Identitätsinfrastruktur in eine kollaborative Internetanwendung von Interesse sind. Nachfolgend werden in Übereinstimmung mit [For98] weitere Klassifizierungsmerkmale aufgeführt.

Eine Person, die einer anderen den Weg weist, tut dieses mit der *Intention*, das Navigieren der anderen Person zu determinieren. Es sind aber auch Situationen denkbar, in denen die eigenen Handlungen die Navigation anderer Personen beeinflussen, ohne daß dieses der entsprechenden Person bewußt ist. Ein Benutzer des Online-Shops Amazon<sup>1</sup>, der ein Buch erwirbt, gibt damit anderen Nutzern indirekt eine Kaufempfehlung. Dennoch darf bezweifelt werden, daß Leute Bücher mit der Intention kaufen, die von ihnen erstandenen Werke anderen zu empfehlen.

Tabelle 4.1 zeigt, wie einige exemplarische Anwendungen bezüglich der Dimensionen Kommunikationsmodus und Intention bei der Kommunikation von Navigationshinweisen einzuordnen sind. Benutzer, die direkt mit anderen interagieren, indem sie beispielsweise auf Fragen eingehen, tun dieses in der Intention, daß ihre Äußerungen vom Kommunikationspartner beachtet werden und dessen Handeln beeinflussen. Folglich gibt es kein Beispiel direkter Social Navigation, bei der die Navigationsunterstützung kommunizierende Person unbeabsichtigt handelt. ActiveWorlds ist eine dreidimensionale Virtual Reality Umgebung, in der Personen durch Avatare repräsentiert werden. In dieser virtuellen Welt können Anwender kommunizieren und gemeinsam navigieren. Ähnliches leisten Collaborative-Browsing-Anwendungen für das Web. Beide Beispiele beinhalten direkte, beabsichtigte Kommunikation zwischen Benutzern. Die Annotation von Dokumenten können von nachfolgenden Besuchern gelesen werden und deren Weg durch den Informationraum beeinflussen. Dieses entspricht der Intention des Verfassers von Anmerkungen. Collaborative Filtering (siehe Abschnitt 5.2), als ein Verfahren die Bewertungen

---

<sup>1</sup>[www.amazon.com](http://www.amazon.com)

#### 4. Social Navigation

	DIREKT	INDIREKT
INTENTIONAL	Active Worlds, Collaborative Browsing, Link-Tipper E-Mail	Annotationen, Collaborative Filtering basierend auf expliziten Bewertungen
UNBEABSICHTIG	-	History-Enriched Objects, Footprints, Collaborative Filtering basierend auf impliziten Bewertungen

**Tabelle 4.1.:** Klassifikation von Social Navigation angelehnt an die Darstellung in [For98]

vieler Benutzer bei der automatischen Selektion von Informationen einzubeziehen, kann entweder auf expliziten Bewertungen von Benutzern basieren oder deren Verhalten im System auswerten und daraus implizite Bewertungen generieren. In beiden Ausprägungen werden die Bewertungen nicht direkt zwischen den Anwendern kommuniziert. Im erstgenannten Fall ist die Abgabe von Bewertungen für den Anwender mit einem Aufwand verbunden, welchen er nur mit einer bestimmten Intention auf sich nehmen wird. Im anderen Fall ist die Generierung von Bewertungen ein Randprodukt der Handlungen der Benutzer. Diese sind sich möglicherweise nicht einmal der Tatsache bewußt, daß ihre Aktionen ausgewertet werden. Das Handeln der Benutzer dürfte also im seltensten Fall von dem Bestreben geleitet sein, anderen Benutzern Navigationshinweise zu geben.

Ein weiteres Klassifizierungsmerkmal von Social-Navigation-Anwendungen ist, ob die Navigationsunterstützung einzelnen Benutzer zuzuordnen ist, oder *aggregierte Information* einer Menge von Benutzern kommuniziert wird. So ist es beispielsweise ein Unterschied, ob lediglich der Prozentsatz von Benutzern dargestellt wird, die zum Kauf eines Produktes raten, oder ob sich die Zusammensetzung dieses Votums aus den einzelnen Meinungen nachvollziehen läßt. Die Wahrnehmbarkeit von einzelnen Benutzern in der Menge ist mit Risiken für die Privatsphäre der Anwender verbunden.

Die Auflistung der meistgelesenen Meldungen einer Web-Site kann Benutzern bei der Auswahl der Meldungen dienlich sein und ist als Social Navigation anzusehen, da sich diese Hinweise dynamisch durch die Aktivitäten anderer Benutzer entwickeln. Dennoch haben diese Hinweise gegenüber einer auf dem Vergleich der individuellen Lesengewohnheiten mit denen anderer Benutzer basierenden Nachrichtenauswahl eine andere Qualität. Social Navigation kann auch danach klassifiziert werden, ob die gegebenen Navigationshinweise *personalisiert* wurden, oder ob für alle Benutzer einheitliche Hinweise vorliegen.

Hinsichtlich der *Identität* der Benutzer läßt sich zwischen Informationsräumen unterscheiden, in denen die Benutzer anonym, mit einem gewählten Pseudonym oder mit

ihrer realen Identität auftreten (siehe Abschnitt 2.2.1). Der Grad der Identifizierbarkeit eines Interaktionspartners ist ein Kriterium bei der Bewertung seiner Beiträge. Des Weiteren ist der Aufbau einer Reputation als eine Voraussetzung für das Schaffen von Vertrauen nur dann möglich, wenn Benutzer zumindest unter einem Pseudonym auftreten, das in dem Kontext des spezifischen Informationsraumes oder kontextübergreifend Gültigkeit besitzt.

Eine Dimension der Betrachtung von Umgebungen, in denen Benutzer miteinander interagieren können, sind die kommunizierten *Benutzerprofile* und deren Authentifizierbarkeit. Im einfachsten Fall ist eine Einschätzung von Interaktionspartnern nur aufgrund deren früherer Beiträge in einer Umgebung, eventuell ergänzt um die Meinung anderer Benutzer zu diesen, möglich. In komplizierteren Umgebungen beinhalten Benutzerprofile weitergehende, eventuell sogar authentifizierbare Angaben zu Interessen und Qualifikationen der Benutzer. Hinsichtlich der Verwendung von Benutzerprofilen ist zu differenzieren, ob diese automatisch zur Erbringung eines Dienstes ausgewertet werden oder ob Nutzer gegenseitig ihre Profile einsehen können.

#### 4.4. Direkte Social Navigation

Kennzeichnend für direkte Social Navigation ist die gemeinsame, bidirektionale Kommunikation zwischen der navigierenden Personen und anderen Nutzern, die Unterstützung bei der Navigation gewähren [Die03]. Zentrale Fragestellung ist dabei das Auffinden von geeigneten Kommunikationspartnern. Dabei kann es von Bedeutung sein, an welchem Ort sich potentielle Kommunikationspartner befinden, oder ob diese bestimmte Eigenschaften, wie z. B. Expertenwissen auf einem bestimmten Gebiet, aufweisen. Nachfolgend wird Collaborative Browsing als ein Beispiel direkter Social Navigation und als ein Ansatz zur kollaborativen Nutzung des World Wide Webs vorgestellt.

Web-Server können simultan mehrere Benutzer bedienen. Diese erfahren das Web jedoch so, als würden sie alleine mit den Servern kommunizieren.

On most areas of the Web, for example, users are given the illusion of being the only person present [DDH<sup>+</sup>00].

Anfragende Benutzer sind im Web, vorausgesetzt es wurde kein anwendungsspezifisches Autorisierungskonzept implementiert, nur durch eine IP-Adresse und einen Port gekennzeichnet. Nicht einmal einzelne Rechner lassen sich verlässlich identifizieren. Anfragende Rechner können über dynamisch vergebene IP-Adressen verfügen. Ebenso ist es möglich, daß sich hinter einer IP-Adresse mehrere Rechner und damit auch mehrere Benutzer verbergen. Dies ist der Fall, wenn der Zugriff auf das World Wide Web vermittelt

durch einen Proxy-Server erfolgt oder die Anbindung des lokalen Netzes an das Internet über einen Router mit Network Address Translation erfolgt. Das zur Anfrage und Übertragung der Dokumente verwendete HTTP-Protokoll [FGM<sup>+</sup>99] ist zustandslos. Seitens des Web-Servers liegt keine Information darüber vor, ob ein Benutzer die geladene Seite nach wie vor in seinem Browserfenster geöffnet hat oder ob ein neues Dokument geladen oder die Applikation beendet wurde. Allenfalls kann anhand einer Heuristik entschieden werden, welche Dokumente gerade von welchen Anwendern betrachtet werden. Sofern nicht durch Cookies auf Anwendungsebene ein Session-Konzept realisiert wurde, können keine sicheren Aussagen zu Bewegungen von Nutzern auf einer Site getroffen werden, und Benutzer lassen sich bei wiederholten Besuchen nicht ohne weiteres wiedererkennen. HTTP ist ein einfaches Request-Response-Protokoll, bei dem die Kommunikation durch den Client initiiert wird. Einmal geladene Seiten können nicht durch den Server aktualisiert oder um Meta-Informationen, wie beispielsweise eine aktuelle Liste der eine Seite anfragenden Benutzer, ergänzt werden. Das Web in der jetzigen Form erfüllt folglich aus genannten technischen Gründen nicht die Voraussetzungen, um ein Bewußtsein für die Gegenwart anderer Nutzer (siehe Abschnitt 2.1.6) und damit die Voraussetzung für die Interaktion zu schaffen. Die Spezifikationen des World Wide Webs beinhalten keine Protokolle zur Synchronisation der Bewegungen im Web, zur synchronen Kommunikation zwischen Anwendern oder zum asynchronen Austausch von Verweisen auf interessante Quellen.

Aufgrund der mangelnden Unterstützung von direkter Social Navigation durch die Technologie des WWW weichen Anwender auf andere Internetanwendungen aus, um bei der Navigation im Web zu interagieren. Typische Beispiele einer derartigen Verhaltensweise sind die asynchrone Kommunikation von Link-Tips per E-Mail, Newsgroups oder Web-Foren oder der synchrone Austausch von Verweisen auf Ressourcen im Web per Instant-Messaging oder Chat. Instant-Messaging-Dienste wie ICQ<sup>2</sup> beinhalten Awareness-Funktionalität. So können gezielt persönlich bekannte Personen kontaktiert werden, welche ebenfalls gerade online sind. Die unzureichende Integration von Funktionalität zur Kommunikation zwischen Benutzern im Web führt zu Medienbrüchen beim Austausch von Navigationsinformationen, wenn beispielsweise URLs per Kopieren und Einfügen zwischen Anwendungen transferiert werden. Das kommunizierte Wissen um Ressourcen im Web steht, sofern es per E-Mail oder Instant Messaging kommuniziert wurde, nur einem eingeschränkten Personenkreis zur Verfügung und liegt zudem unstrukturiert vor.

Collaborative Browsing [HRCV01, SSW97] bezeichnet verschiedene Anwendungen, die versuchen, integrierte Lösungen für die Erweiterung des Webs um Awareness, Kommunikation zwischen Nutzern und Koordination von Browsing-Aktivitäten zu bieten. Denkbare Anwendungsgebiete sind dabei virtuelle Lernumgebungen, in denen Tutoren ihre Schüler durch die Lektionen leiten, Informationsschalter auf Unternehmenswebsites

---

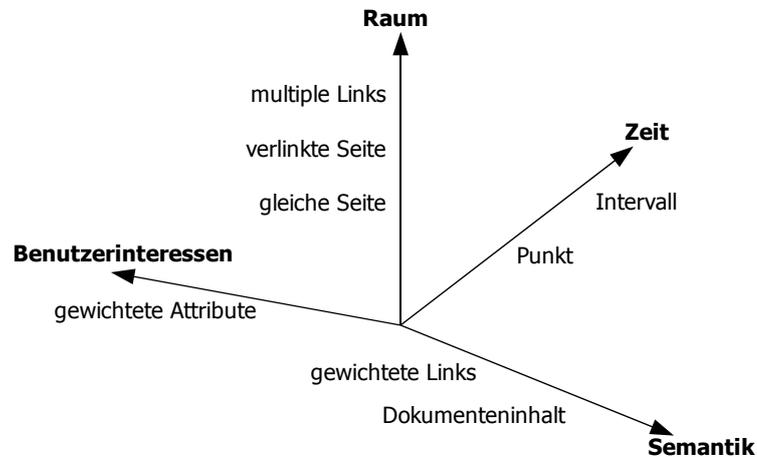
<sup>2</sup>[www.icq.com](http://www.icq.com)

oder einfach nur die Recherche im Web, bei der Anwender mit ähnlichen Interessen bzw. Aufgaben zusammengebracht werden sollen, so daß Synergien nutzbar werden.

Das Web wird gleichzeitig von einer unüberschaubaren Menge von Personen genutzt. Eine Herausforderung beim Entwurf einer Collaborative-Browsing-Anwendung besteht darin, Individuen zusammenzubringen, welche voneinander profitieren können. Im Rahmen des Projektes CoBrow [SSW97] wurde zu diesem Zweck eine Metrik definiert, welche die Nähe von Benutzern über den Abstand in einem von den Dimensionen Raum, Zeit, Semantik und Benutzerinteressen aufgespannten, vierdimensionalen Raum (siehe Abbildung 4.2) ausdrückt. Die Betrachtung des Webs als einen gerichteten Graphen legt die Definition der räumlichen Entfernung von Web-Seiten über die minimale Zahl von Links, mit der man von einer Seite zur anderen gelangt, nahe. Des weiteren ist die Definition einer semantischen Distanz von Dokumenten denkbar. Knoten im Hypertext, welche eine ähnliche Semantik aufweisen, werden dabei als nahe beieinander liegend definiert. Im Extremfall ist der gleiche Text unter mehreren URLs abrufbar, so daß der semantische Abstand Null beträgt. Die Sichtbarkeit von Benutzern läßt sich über die Betrachtung der räumlichen und semantischen Distanz der von ihnen geladenen Dokumente einschränken. Ergänzend kann die zeitliche Dimension hinzugezogen werden, bei der der Abstand zwischen Benutzern über die Überlappung der Zeiträume, in denen die Dokumente jeweils im Browser dargestellt werden, definiert wird. Die vorstehenden Metriken definieren Nähe zwischen Benutzern über die von ihnen geladenen Dokumente. Betrachtet man die Benutzer selbst, so kann über den Vergleich der sie beschreibenden Attribute eine weitere Dimension der Betrachtung von Nähe definiert werden. Benutzer, welche die gleiche Sprache sprechen oder übereinstimmende Interessen aufweisen, werden als nahe beieinander stehend aufgefaßt.

## 4.5. Indirekte Social Navigation

Wie in Abschnitt 4.3 dargestellt wurde, unterscheidet sich indirekte von direkter Social Navigation durch das Fehlen gemeinsamer Kommunikation zwischen den beteiligten Personen. Stattdessen werden von den Benutzern eines Informationsraumes bewußt oder unbewußt generierte Navigationsinformationen unidirektional zu anderen Besuchern des Raumes kommuniziert. Indirekte Social Navigation in Informationsräumen kann sehr unterschiedliche Ausprägungen haben. Sie teilen das gemeinsame Merkmal, daß Handlungen im Informationsraum für andere Benutzer sichtbar gemacht werden. Dieses kann durch Visualisierung von Gegenwart und Aktivität der aktuell im Informationsraum befindlichen Individuen geschehen, so daß ähnlich dem Verhalten, sich an Menschenansammlungen zu orientieren oder einfach einer Menge anderer Menschen nachzulaufen, auch in Informationsräumen eine Orientierung an anderen Personen möglich wird, ohne mit diesen direkt in Kontakt zu treten. Ein gutes Beispiel für dieses Paradigma ist der

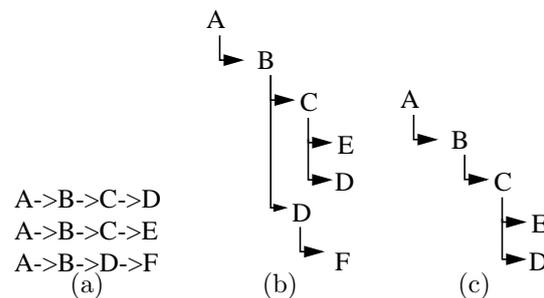


**Abbildung 4.2.:** Verschiedene Metriken für die Definition von Nähe zwischen Web-Nutzern formen einen vierdimensionalen Raum. (Quelle: [SSW97])

WebPlaces-Prototyp, welcher in Abschnitt 6.1 präsentiert wird. Sollen die vergangenen Handlungen von Nutzern eines Informationsraumes für nachfolgende Personen nutzbar gemacht werden, so kann dieses entweder über sichtbare, von Aktivitäten hinterlassene Spuren geschehen oder über von den Benutzern erstellte Artefakte, wie Annotationen und Bewertungen.

Das Web bietet keinerlei Unterstützung für indirekte Social Navigation. Der Abruf von Web-Seiten hinterläßt in der Regel auf Web-Servern Spuren in Form von Log-Einträgen. Log-Dateien werden jedoch typischer Weise nur seitens der Site-Betreiber ausgewertet. In der ursprünglichen Architektur des Webs, die auf zustandslosen HTTP-Servern basiert, ist eine Darstellung der Spuren der Benutzung von Web-Seiten nicht vorgesehen. Das Web beinhaltet zudem keine Funktionen, um Dokumente zu kommentieren, Links zu ergänzen und Seiten zu bewerten. Zu diesen Zwecken wird, wie bereits in Bezug auf indirekte Social Navigation erwähnt wurde, auf andere Formen der Internet-Kommunikation ausgewichen.

In History-Enriched Environments wird in Analogie zu den Eigenschaften realer Objekte versucht, digitale Objete mit Spuren ihrer Benutzung anzureichern. Ein erster Ansatz war das Projekt [HHWM92]. Hier wurden Scrollbars um die Visualisierung der Häufigkeit von Editiervorgängen in den jeweiligen Textabschnitten ergänzt. Bezogen auf das Web ist, abgesehen von Web-Site spezifischen Lösungen, wie z. B. der Darstellung von Zugriffszählern, insbesondere der von Wechselblat entwickelte Prototyp Footprints [WM97, WM99, Wex99], aber auch das System CoWeb [DL00] zu nennen.



**Abbildung 4.3.:** Schema der Visualisierung von Pfaden im Web durch den Footprints-Prototyp. (a) Seiten wurden in dieser Reihenfolge abgerufen, (b) Visualisierung der von Dokument A ausgehenden Pfade, (c) Einem Benutzer, der von A über B nach C navigiert ist, wird diese Grafik präsentiert. (Quelle: [Wex99])

Idee des Projektes Footprints ist es, Pfade im Web zu visualisieren und Benutzern in Abhängigkeit von der Abfolge der zuvor von ihnen besuchten Dokumente Navigationsempfehlungen zu geben. Dabei wird eine Idee aufgegriffen, welche auf eine der ersten Beschreibungen von Hypertext-Systemen zurückgeht. Bereits im Artikel „As we may think“ [Bus96]<sup>3</sup>, der 1939 von Vannevar Bush veröffentlicht wurde, wird die Idee beschrieben, Pfade in Informationsräumen sichtbar zu machen und mit anderen Benutzern zu teilen. Anhand des in Abbildung 4.3 dargestellten Beispiels [Wex99] soll die Visualisierung von Pfaden durch das Footprints-System veranschaulicht werden. Dabei werden als Datenbasis für die Visualisierung die in Abbildung 4.3(a) aufgelisteten Pfade angenommen. Einem Benutzer, der sich im Hypertext von Dokument A zum Dokument B bewegt hat, wird die in Abbildung 4.3(b) dargestellte Grafik präsentiert. Die Repräsentationen einzelner Dokumente können in der Visualisierung mehrfach dargestellt werden. Dabei wird der unterschiedlichen Position des Abrufs des Dokuments in verschiedene Pfaden Rechnung getragen und berücksichtigt, daß bestimmte Folgen von Seitenabrufen von vorherigen Besuchern nicht beschriftet wurden. So wurden die verschiedenen Darstellungen von Dokument D nicht zusammengelegt, da sie unterschiedliche Pfade repräsentieren. Keine früheren Besucher haben, nachdem sie den Pfad A,B,C gegangen sind, ihren Weg über D nach F fortgesetzt. Setzt der Besucher mit dem Aufruf des Dokuments C seinen Weg fort, dann verändert sich die Visualisierung dynamisch, wie in Abbildung 4.3(c) ersichtlich ist. Der mit der Entscheidung des Benutzers, den Link von Dokument B zum Dokument C aufzurufen, bedeutungslos gewordenen Pfad von B,D,F wurde ausgeblendet.

Wenn eine Auswahl aus einer unüberschaubar großen Menge von Alternativen getroffen

<sup>3</sup>Es handelt sich bei der referenzierten Quelle um einen Nachdruck

werden muß oder wenn die eigenen Informationen zum Treffen einer qualifizierten Entscheidung nicht ausreichend sind, ist die Befragung einer vertrauten, kompetenten Person eine mögliche Vorgehensweise bei der Entscheidungsfindung. Empfehlungssysteme (siehe Kapitel 5) sollen Benutzern die Auswahl von Objekten vereinfachen. Derartige Funktionalität kann durch Vergleich der Objekteigenschaften mit den in Profilen dokumentierten Präferenzen der Benutzer geschehen (siehe Abschnitt 5.1). Eine Alternative ist es, den Vorgang des „sich Tips geben“ zu formalisieren und eine Vielzahl von individuellen, explizit oder implizit abgegebenen Bewertungen zu Empfehlungen zu aggregieren. Im Gegensatz zu der vorstehend aufgeführten Verhaltensweise der Entscheidungsfindung durch direkte Kommunikation sind auf den Handlungen der Benutzer eines Informationsraumes basierende Empfehlungssysteme ein typisches Beispiel für indirekte Social Navigation.

Einfache Empfehlungssysteme bieten den Benutzern keine individuellen Empfehlungen, sondern versuchen, anhand von Verfahren, wie der Berechnung von durchschnittlichen Bewertungen oder der Zählung von Zugriffen, Objekte gemäß ihrer Popularität in eine Reihenfolge zu bringen und diese Information anderen Nutzern zur Entscheidungsunterstützung zur Verfügung zu stellen. Abbildung 4.4 zeigt ein in die Web-Site des Zweiten Deutschen Fernsehens integriertes Empfehlungssystem. Unter den Artikeln befindet sich jeweils ein Bedienelement (Abbildung 4.4(a)), welches der Abgabe von Bewertungen dient. Aus der Summe der Bewertungen wird dann die nach durchschnittlicher Bewertung sortierte Liste der Meldungen (Abbildung 4.4(b)) erzeugt.

Individuelle Empfehlungen werden beispielsweise mittels Collaborative-Filtering-Techniken generiert. Dabei wird die individuelle Übereinstimmung mit der Meinung anderer Benutzer bei der Berechnung von Empfehlungen berücksichtigt. Collaborative Filtering wird ausführlich in Abschnitt 5.2 behandelt. An dieser Stelle soll jedoch exemplarisch das Projekt Sitieseer [RP97] vorgestellt werden, welches individuelle Empfehlungen von Web-Dokumenten bietet. Prämisse für die Funktion dieses Systems ist, daß das Setzen eines Bookmarks einer impliziten Deklaration von Interesse an den unter der hinterlegten URL abrufbaren Inhalten entspricht und daß die Organisation von Bookmarks in einer Verzeichnisstruktur semantische Beziehungen zwischen den Dokumenten ausdrückt. Die Verzeichnisse werden als ein persönliches Klassifikationssystem betrachtet. Sie bilden den Kontext der vom System hervorgebrachten Empfehlungen. Um für einen Benutzer im Kontext eines Ordners seiner Bookmarksammlung Empfehlungen zu berechnen, wird die Überlappung der Menge der in diesem Verzeichnis abgelegten URLs mit den Ordnern der Bookmarks anderer Systemnutzer ermittelt. Übereinstimmung in URLs, welche bei der globalen Betrachtung der von den Nutzern hinterlegten Adressen selten vorkommen, werden bei der Berechnung der Überlappung besonders berücksichtigt. Basierend auf diesem Verfahren wird für jedes Bookmark-Verzeichnis jedes Anwenders dynamisch eine Community of Interest (siehe Abschnitt 2.1.4) gebildet, wobei vorausgesetzt wird, daß Anwender mit hohen Übereinstimmungen in einem Ver-

zeichnis ihrer Bookmarksammlung im Kontext dieses Verzeichnisses übereinstimmende Interessen haben. Der entsprechende Bookmark-Ordner der Empfehlungen suchenden Person, welcher zur dynamischen Bildung einer Community verwandt wurde, wird im letzten Schritt des Empfehlungsprozess um von den Mitglieder der Community in ihren korrespondierenden Verzeichnissen abgelegten URLs ergänzt. Abbildung 4.5 veranschaulicht diesen Algorithmus anhand des Verzeichnisses „Location Spots“ des Benutzers John. Dieses Verzeichnis stimmt in drei URLs mit Mary’s „Tropical Getaways“ Verzeichnis überein, was dieses Verzeichnis zur bevorzugten Quelle von Empfehlungen für John macht. Die von Mary abgelegte URL [www.vtourist.com](http://www.vtourist.com) wird deshalb John im Kontext seines „Vacation Spots“ Ordners empfohlen.

Die Annotationen anderer Besucher können eine wertvolle Ergänzung der von den Autoren der Objekte eines Informationsraumes offerierten Informationen darstellen und die Navigation vereinfachen. Annotationssysteme bieten Nutzern eines Informationsraumes die Möglichkeit, Annotationen, wie Kommentare, weitergehende Information oder Fehlerkorrekturen, zu verfassen und die Strukturen eines Informationsraumes um neue Verknüpfungen zu ergänzen. Annotationen werden vom System als Meta-Informationen zu den Objekten des Informationsraumes verwaltet und zusammen mit diesen ausgegeben.

Eine Erweiterung der online-Dokumentation der Programmiersprache PHP um ein Annotationssystem wird in Abbildung 4.6 gezeigt. Die von den mit PHP programmierenden Web-Entwicklern verfaßten Kommentare beinhalten vielfach nützliche Algorithmen, weisen auf Bugs hin oder enthalten Links zu weitergehenden Informationen. Eine Qualitätssicherung findet nicht statt, so daß es durchaus möglich ist, daß ein als Teil einer Annotation veröffentlichter Algorithmus nicht der dokumentierten Funktionalität entspricht.

Neben Web-Site spezifischen Ansätzen, gibt es eine Reihe von Projekten, welche es zum Ziel haben, beliebige Web-Dokumente annotierbar zu machen. CritLink [Yee02] basiert auf einem Proxy-Server, welcher Web-Seiten um eine Toolbar mit Schaltflächen zum Erstellen von Kommentaren ergänzt und annotierte Stellen eines Dokuments mit kleinen verlinkten Markern kennzeichnet. Annotationen werden als normale HTML-Dokumente unter einer von Annotations-Server vergebenen URL abgelegt und sind somit ebenfalls annotierbar. Auch das W3C<sup>4</sup> hat sich dem Thema der Annotation von Web-Dokumenten angenommen. Annotea [KK01] ist eine vom W3C entwickelte Infrastruktur für Web-Annotationen, welche auf Web-Standards basiert. Annotationen werden ganz allgemein als Meta-Daten betrachtet, welche sich in RDF<sup>5</sup> ausdrücken lassen und sich auf durch XPointer<sup>6</sup> markierte Stellen eines Dokuments beziehen.

---

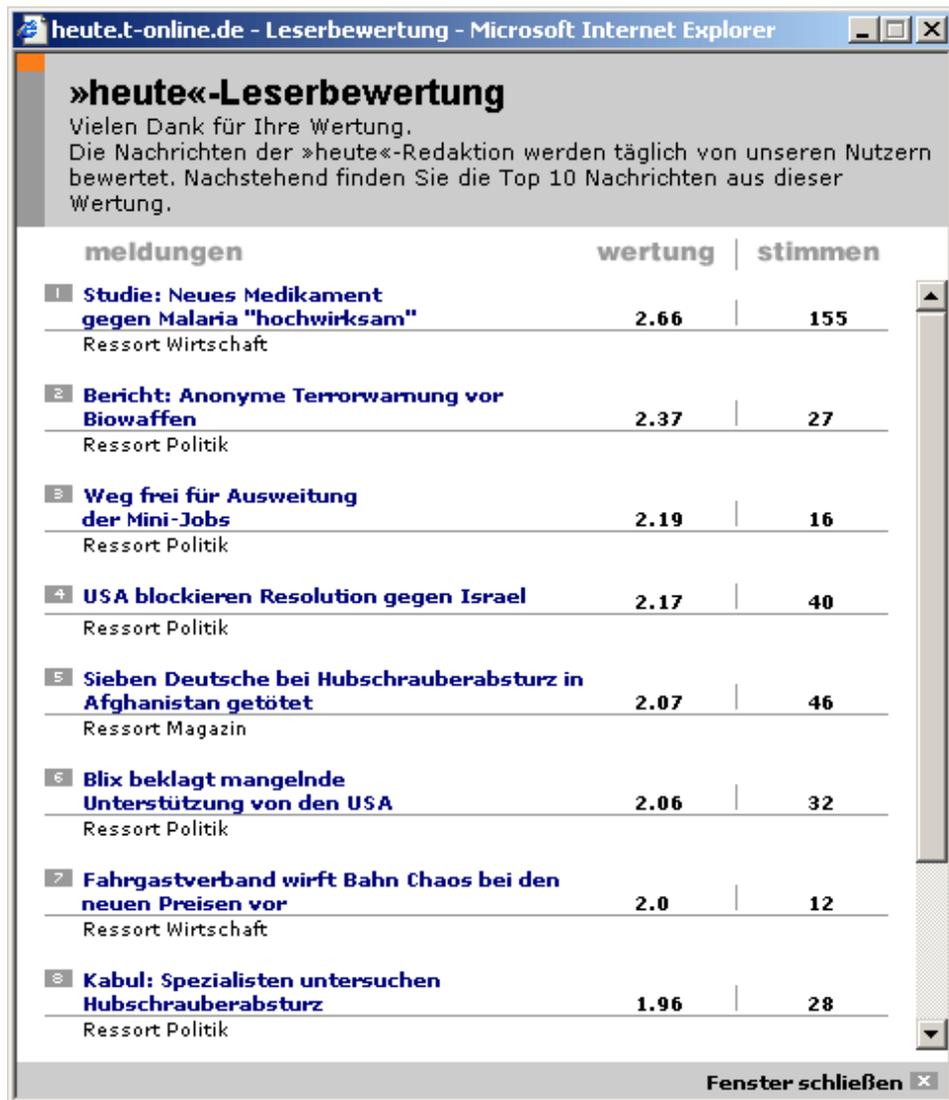
<sup>4</sup>[www.w3c.org](http://www.w3c.org)

<sup>5</sup>[www.w3.org/RDF](http://www.w3.org/RDF)

<sup>6</sup>[www.w3.org/TR/WD-xptr](http://www.w3.org/TR/WD-xptr)



(a)



(b)

Abbildung 4.4.: Bewertungssystem für Meldungen auf der Web-Site des ZDF (www.zdf.de). (a) Abgabe von Bewertungen, (b) Nach durchschnittlichen Bewertungen sortierte Liste der Meldungen

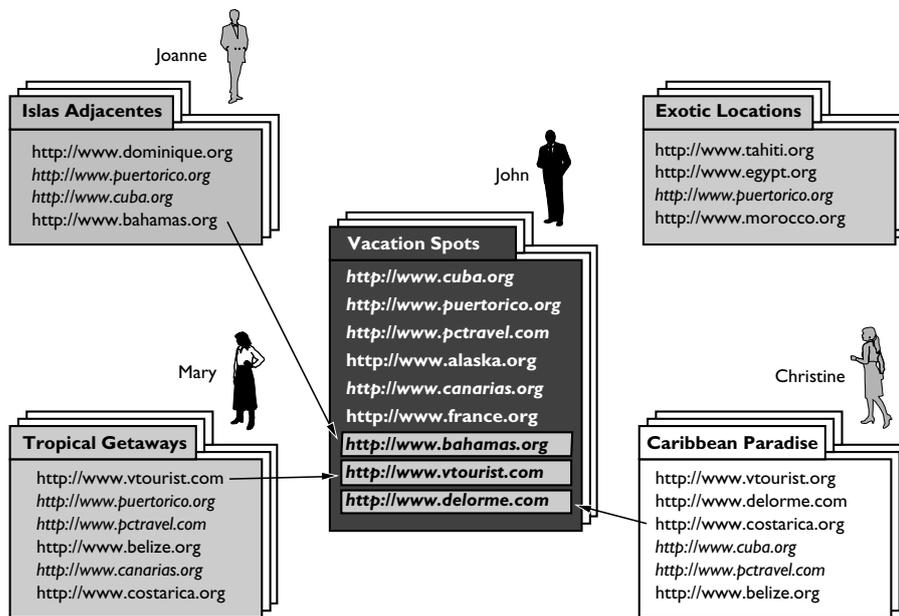


Abbildung 4.5.: Generierung von Empfehlungen basierend auf der Auswertung von Bookmark-Listen (Quelle: [RP97])

## 4.6. Voraussetzungen und Randbedingungen

Das Phänomen Social Navigation wurde in den vorhergehenden Abschnitten dieses Kapitels zum einen aus einer abstrakten und von Online-Anwendungen losgelösten Sicht beschrieben und zum anderen anhand von Beispielen illustriert. In diesem Abschnitt sollen nun einige Voraussetzungen erläutert werden, die eine Online-Anwendung erfüllen sollte, um soziales Navigieren zu ermöglichen. Es werden Randbedingungen betrachtet, welche die Handlung des sozialen Navigierens beeinflussen und beim Entwurf von Social Navigation unterstützenden Systemen zu beachten sind. Abbildung 4.7 zeigt schematisch, mit welchen anderen Faktoren rechnergestützter Kommunikation und Interaktion Social Navigation in Wechselwirkung steht.

Bei der Navigation Unterstützung suchende Personen haben gegenüber den Navigationsunterstützung anbietenden Personen einen Informationsnachteil. Die Validität von Navigationshinweisen kann meist erst zu einem späteren Zeitpunkt, z. B. nach dem Erwerb des empfohlenen Artikels, festgestellt werden. Soziales Navigieren ist folglich nur dann möglich, wenn Vertrauen (siehe Abschnitt 2.2.2) in die Qualität der Navigationsinformationen und in die positiven Absichten von deren Urheber besteht [FHS98]. Vertrauen

## 4. Social Navigation

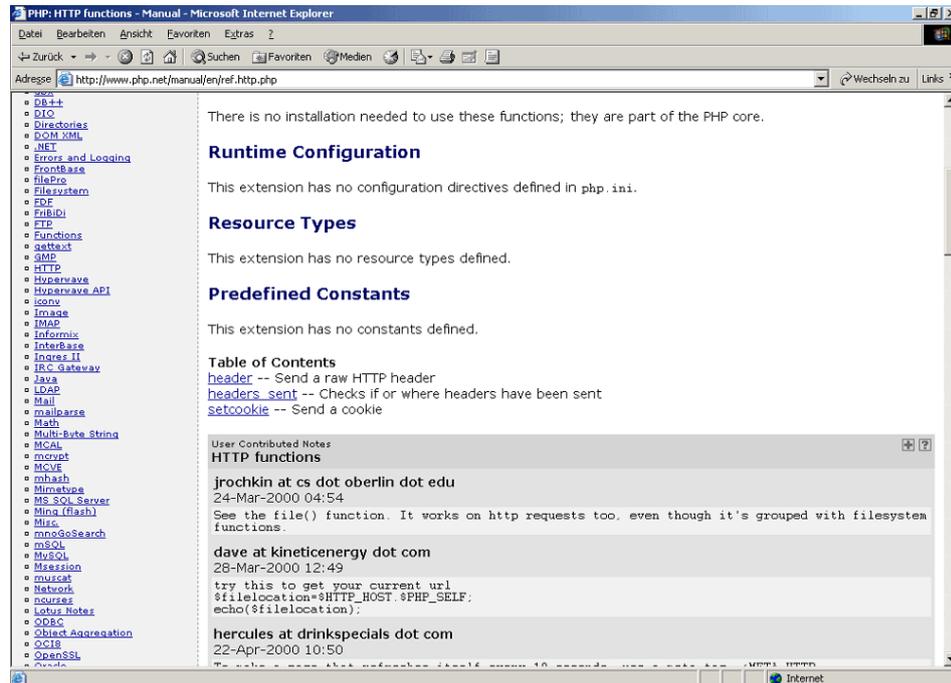


Abbildung 4.6.: PHP Annotated Manual (www.php.net/manual)

kann zum einen über die Offenbarung von Attributen der realen Identität aufgebaut werden, welche die Kompetenzen einer Person dokumentieren, zum anderen kann Vertrauen über Reputation (siehe Abschnitt 2.2.3), also die kommunizierten Erfahrungen anderer Personen, geschaffen werden. Voraussetzung für den Aufbau von Vertrauen ist die Identifizierbarkeit von Personen. Werden Navigationshinweise von anonymen Nutzern hervorgebracht oder in Form von aggregierter Information präsentiert, so kann kaum von Vertrauen in die Anbieter von Navigationshinweisen gesprochen werden. In diesem Fall ist ein Mindestmaß an Vertrauen in das System, z. B. in dessen Schutz vor Manipulationen, erforderlich, um sich von den Navigationsinformationen leiten zu lassen.

Social Navigation ist Navigation durch Kommunikation mit anderen Menschen bzw. Agenten. Dabei ist es entscheidend, welcher Teil der Identität (siehe Abschnitt 2.2.1) der Interaktionspartner wahrnehmbar ist. Wenn Personen in einer virtuellen Umgebung mit ihrer realen Identität auftreten, also ihren wirklich Namen nennen und eventuell weitere sie identifizierende Attribute offenbaren, dann werden damit auch Regeln der realen Welt auf den virtuellen Raum übertragen. Forsberg vertritt die These, daß sich Menschen in Informationsräumen unterschiedlich verhalten, je nachdem, wie sie in dieser Umgebung identifiziert werden können [For98]. Die Regeln in einem Informations-

raum und damit die Relevanz von Navigationsinformationen werden folglich durch den Grad an Identifizierbarkeit der beteiligten Personen beeinflusst. Umgebungen, die ein vollständig anonymes Auftreten gestatten, machen es Personen leicht, sich opportunistisch zu verhalten, da sie für ihre Handlungen nicht verantwortlich<sup>7</sup> gemacht werden können. Online publizierte medizinische Informationen sind vielfach von mangelhafter Qualität [IPCB97]. Insbesondere in dieser Domäne ist Wissen um die Identität derjenigen Besucher eines Informationsraumes, mit denen man Informationen austauscht, deren Ratschläge man befolgt oder an denen man sich orientiert, notwendig, um die Qualität der von ihnen gegebenen Navigationsunterstützung einschätzen zu können [FHS98]. Benutzer eines medizinischen Diskussionsforums sollten den kommunizierten Informationen mit größter Skepsis begegnen, wenn diese von anonymen Personen stammen. Kommen jedoch Hinweise von einer Person, die mit realer Identität auftritt und nachweisen kann, daß sie als Arzt in einem Krankenhaus arbeitet, dann kann von einer höheren Qualität deren Beiträge ausgegangen werden. Zum einen kann antizipiert werden, daß ein Krankenhaus nur hinreichend qualifizierte Mediziner einstellt und zum anderen beinhaltet das Auftreten unter realer Identität das Risiko, für die eigenen Handlungen zur Verantwortung gezogen zu werden. Es ist also im Spannungsfeld zwischen dem Schutz der Privatsphäre und einer verbindlichen Kommunikation, bei der die Herkunft von Informationen nachvollziehbar ist und die Verantwortlichkeit für Beiträge sichtbar wird, beim Entwurf von Systemen abzuwägen.

Awareness (siehe Abschnitt 2.1.6) ist eine Voraussetzung für Social Navigation. Nur wenn die Gegenwart anderer Benutzer wahrnehmbar ist, ist es möglich mit ihnen direkt in Kontakt zu treten bzw. sich an ihnen zu orientieren [FHS98]. Im Kontext indirekter Social Navigation ist es überdies von Interesse, daß Spuren vergangener Handlungen wahrnehmbar sind. Wenn Awareness der Anbahnung von Interaktionen dient, dann ist es erforderlich, daß Awarenessinformationen Angaben zur Identität der Benutzer beinhalten, die eine Unterscheidung von Personen gestatten. Direkte Social Navigation setzt das Auffinden eines kompetenten Interaktionspartners voraus, welcher zu helfen bereit ist. Danis argumentiert, daß das Auffinden relevanter Interaktionspartner vereinfacht werden könne, wenn Awarenessinformationen um Attribute der Persönlichkeit eines Benutzers ergänzt werden [Dan00]. Überdies kann die Ergänzung von Aufzeichnungen vergangener Aktivitäten der Nutzer eines Informationsraumes um Identitätsinformationen sinnvoll sein, um eine persönliche Relevanzeinschätzung dieser Navigationsinformationen durchführbar zu machen. Wexelbat vertritt die Auffassung, daß Menschen ein natürliches Bestreben haben, sich an den Handlungen der ihnen nahe stehenden Personen zu orientieren [Wex99].

Zu viele Awarenessinformationen können einen Nutzer kognitiv überfordern, und sie verlieren damit ihre Nützlichkeit. Des weiteren steht die Kommunikation von Awarenessin-

---

<sup>7</sup>In der englischsprachigen Literatur wird in diesem Zusammenhang von *Accountability* gesprochen

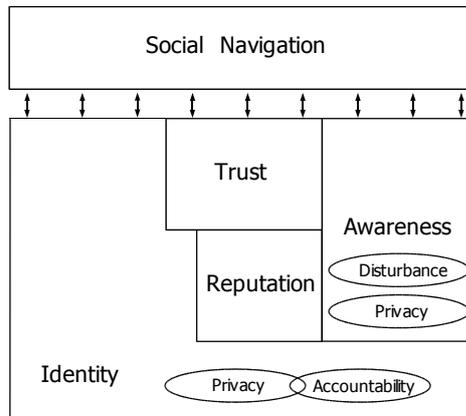


Abbildung 4.7.: Einflussfaktoren sozialen Navigierens

formationen insbesondere dann, wenn diese zusammen mit Attributen der Identität kommuniziert werden, in einem Spannungsfeld mit dem Schutz der Privatsphäre der Benutzer. Bezüglich Awareness ist beim Entwurf von Systemen zum einen zwischen Awareness und Disturbance, zum anderen zwischen Awareness und Privacy abzuwägen [HS96].

Identität und Vertrauen sind nicht nur als Randbedingungen von Social Navigation zu sehen, sondern werden auch umgekehrt von den Handlungen des sozialen Navigierens beeinflusst. Wenn eine Person einer anderen kooperativ den Weg weist oder sich durch kompetente Annotationen auszeichnet, begünstigt dies den Aufbau einer positiven Reputation und schafft somit Vertrauen. Soziales Navigieren hat Rückwirkungen auf die virtuelle Identität einer Person, da soziales Navigieren, wie andere kooperative Handlungen auch, Einfluß auf die wahrgenommene Persönlichkeit einer Person hat (siehe Abschnitt 2.2.1). Awareness, im Sinne von Wahrnehmung anderer Personen bzw. deren aktueller oder vergangener Aktivitäten, ist ebenfalls von Social Navigation beeinflusst, wenn man z. B. an einen Chat denkt, der mitprotokolliert wird, und in dem sich Personen über eine bestimmte Web-Site austauschen. Nachfolgende Personen können wahrnehmen, daß im Kontext dieses Chats ein Interesse an dieser Site bestand. Sie sehen möglicherweise auch, wer sich über die Site geäußert hat, und sie können sich von den Meinungen leiten lassen.

### 4.7. Wobei nützt Social Navigation?

In diesem Abschnitt sollen einige Gründe dafür aufgeführt werden, warum es sich lohnen kann, Online-Anwendungen um Social-Navigation-Funktionen zu erweitern. Die vorge-

stellten Argumente wurden nur teilweise anhand von an Prototypen durchgeführten Studien belegt. Es sind vielmehr Arbeitshypothesen, welche aus den Erfahrungen verschiedener Projekte resultieren und in [DDH<sup>+</sup>00] formuliert wurden. Einschränkend ist anzumerken, daß die Aussagen bezüglich der Wirkung von Social Navigation keinen allgemeingültigen Charakter besitzen. Social-Navigation-Techniken, die sich im Kontext einer Anwendung als vorteilhaft erweisen, können im Kontext einer anderen Anwendung von den Benutzern eher als hinderlich oder als Eingriff in die Privatsphäre empfunden werden [FHS98]. Zusätzliche Komplexität erhält die Frage nach dem sinnvollen Einsatz von Social Navigation in verschiedenen Anwendungskontexten durch individuelle Unterschiede zwischen Anwendern bezüglich der von ihnen empfundenen Nützlichkeit von Social Navigation [HLSW00].

**Filtern von Informationen** History-Enriched Environments (siehe Abschnitt 4.5) und Recommender-Systeme (siehe Kapitel 5) scheinen das Auffinden relevanter Informationen in großen Informationsräumen zu vereinfachen. Erfahrungen mit den Prototypen Footprints [Wex03] und GroupLens [MRK97] unterstützen diese Annahme.

Einfache Information-Retrieval-Techniken bestimmen die Relevanz von Information indem sie überprüfen, ob die gesuchten Begriffe enthalten sind. Für die Bestimmung der individuellen Relevanz von Dokumenten bei der Informationssuche ist dieses Kriterium jedoch unzureichend. Als weitere Auswahlkriterien sind u. a. Qualität, Neuigkeitswert und Wirkung im jeweiligen Kontext zu berücksichtigen.

**Beurteilung der Qualität und Wirkung von Informationen** Unter der Qualität einer Information wird hier verstanden, daß sie wahr ist und daß der Autor, gemessen an den Maßstäben der jeweiligen Domäne, professionell gearbeitet hat. Derartige Qualitätskriterien lassen sich nicht automatisch überprüfen. Den Einschätzungen und Reaktionen anderer Personen kommt somit eine besondere Bedeutung zu. Für die Beurteilung der Qualität eines Dokuments kann Wissen darüber, wer den Text gelesen oder in eigenen Dokumenten zitiert hat, von Interesse sein. Ein renommierter Herausgeber birgt für die Einhaltung von Qualitätsstandards bei den in einer Publikation veröffentlichten Beiträgen. Ein angesehenen Wissenschaftler wird nur Dokumente referenzieren, welche seinen eigenen professionellen Ansprüchen genügen.

Die Frage, welche Wirkung eine Information im jeweiligen Zusammenhang hat, ist unmittelbar mit der Frage verknüpft, von welchen Personen die Information zur Kenntnis genommen wurde und wie sie das Handeln der Personen beeinflußt hat [Har03]. Der Einfluß einer wissenschaftlichen Publikation auf die Entwicklung eines Faches läßt sich beispielsweise anhand der diese Quelle referenzierenden Dokumente ermesen.

Wie eine Information in das persönliche „Framework of Relevance“ [Har03] einzuordnen ist, ob sie als interessant einzustufen oder zu ignorieren ist, kann nicht ausschließlich

aufgrund inhärenter Merkmale der Information entschieden werden. Es ist die „soziale Textur“ von Information, also die Meta-Information darüber, wie andere Personen mit der jeweiligen Information umgegangen sind, die in der Praxis maßgeblichen Einfluß auf die Beurteilung von Informationen hat [SH03].

**Social Affordance** <sup>8</sup> Die sichtbaren Aktionen anderer Benutzer vermitteln einen Eindruck davon, was in der entsprechenden Umgebung angemessenes Verhalten ist, was getan werden kann und was unterlassen werden sollte. So wird ein erstmaliger Besucher einer Uni-Bibliothek bereits nach sehr kurzer Zeit durch Beobachtung des Verhaltens der Anwesenden erkennen, daß laute Unterhaltung kein angemessenes Verhalten im Lesesaal einer Bibliothek ist.

Durch die Visualisierung der Benutzeraktionen wirkt ein Informationsraum auf Besucher lebendig und einladend. Die Analyse des Nutzens von Social Navigation sollte sich nicht ausschließlich auf die Betrachtung der Effizienz, mit der Personen in einem Informationsraum navigieren, beschränken. Vielmehr ist zu berücksichtigen, daß Social Navigation die Nutzung des Systems zu einer angenehmeren Erfahrung für die Benutzer machen kann. Diese werden von den Aktionen anderer Personen dazu inspiriert, neue Funktionen auszuprobieren und den Informationsraum über das ursprüngliche Ziel hinausgehend zu erkunden.

Die Möglichkeit der Interaktion mit und der Wahrnehmung von anderen Nutzern formt aus Räumen Plätze, in denen bestimmte soziale Normen gelten [HD96].

Physically, a place is a space which is invested with understandings of behavioral appropriateness, cultural expectations, and so forth. We are located in „space“, but we act in „place“ [HD96].

**Formung des Informationsraumes durch die Benutzer** Social Navigation eröffnet die Perspektive der Einbeziehung von Benutzern in den Prozess der Strukturierung eines Informationsraumes. Informationsräume können sich durch Auswertung des Benutzerverhaltens dynamisch restrukturieren und somit an die Nutzergewohnheiten anpassen. Benutzern bieten sich Alternativen zur Navigation entlang der von den Architekten eines Informationsraumes geplanten Strukturen.

---

<sup>8</sup>Auch hier wurde auf eine Übersetzung verzichtet, da dem Autor kein bedeutungsgleiches deutsches Wort bekannt ist.

## 4.8. Probleme und offene Fragestellungen

Nachdem im vorhergehenden Abschnitt, vielleicht aus einer etwas zu optimistischen Perspektive, der Nutzen von Social Navigation aufgeführt wurde, sollen in diesem abschließenden Abschnitt des Kapitels Probleme von Social Navigation und offene Fragestellungen bei der Konzeption von Social Navigation unterstützenden Systemen dargelegt werden.

Social Navigation steht in vielerlei Hinsicht in einem Spannungsfeld zu dem berechtigten Interesse der Benutzer nach dem Schutz ihrer Privatsphäre. In Abschnitt 4.6 wurde Awareness als eine Voraussetzung für Social Navigation bezeichnet. Die Darstellung der Gegenwart und Aktivitäten einer Person oder die Protokollierung von Aktivitäten birgt das Risiko der Verletzung der Privatsphäre der Anwender [Die03]. Um personalisierte Navigationshinweise zu erhalten, müssen sich Benutzer bei einem System anmelden und akzeptieren, daß Daten über sie in Benutzerprofilen erfaßt werden. Der Einsatz von Pseudonymisierungstechniken relativiert jedoch das Risiko für die Privatsphäre der Anwender.

Systeme, in denen Benutzer Navigationsentscheidungen durch Orientierung an anderen Personen treffen, bieten Angriffsflächen für Personen, welche versuchen das Navigationsverhalten der Benutzer zu ihren Gunsten zu beeinflussen. Entsprechende Systeme unterliegen schnell dem Verdacht, Navigationshinweise zu manipulieren. Die Akzeptanz von Social Navigation dürfte im konkreten Fall eng damit verbunden sein, ob es einerseits gelingt, Vertrauen in das System, Vertrauen in dessen Unabhängigkeit, Vertrauen in den Schutz vor Manipulationen und Vertrauen in die Beachtung von Datenschutzrichtlinien zu schaffen und ob es andererseits gelingt, die Schaffung von Vertrauen in die Urheber von Navigationsinformationen, z. B. durch Integration von Reputationsmechanismen (Abschnitt 2.2.3), zu ermöglichen.

Social-Navigation-Anwendungen setzen, sofern sie explizite Handlungen der Benutzer erfordern, ein hohes Maß an Kooperationsbereitschaft voraus. Dabei darf nicht außer acht gelassen werden, daß die Anwender eines Systems durchaus auch in einer Konkurrenzsituation stehen können oder Personen unter Zeitdruck agieren. Es stellt sich die Frage, warum sich Menschen bei der Navigation helfen sollten [Die03].

Beispiele wie Slashdot<sup>9</sup> zeigen, daß Benutzer teilweise ohne erkennbaren Nutzen bereit sind, Aufwand auf sich zu nehmen, wenn zum einen das ausgeprägte Gefühl, einer Community anzugehören, vorliegt und zum anderen der Aufbau einer positiven Reputation innerhalb der Gemeinschaft motivierend wirkt (Abschnitte 2.2.1 und 2.2.3). Ob sich beispielsweise Besucher einer Web-Site dazu motivieren lassen, anderen Nutzern, denen sie sich nicht in einer online-Gemeinschaft verbunden fühlen, den Weg zu weisen, erscheint

---

<sup>9</sup>[www.slashdot.org](http://www.slashdot.org)

fraglich. Direkte Social Navigation wird in vielen Fällen nur dann funktionieren, wenn sich die beteiligten Personen, also die navigierende Person und die Anbieter von Navigationsunterstützung, in einem gemeinsamen organisatorischen Kontext befinden, sich persönlich kennen oder sich der Anbieter von Navigationsunterstützung als Dienstleister sieht und seine Tätigkeit in irgendeiner Form materiell kompensiert wird. Man denke nur an den Online-Helpdesk einer Unternehmenswebsite oder Knowledge-Management-Anwendungen in Organisationen, wie das System AnswerGarden [AM96].

Das Maß an Kooperationsbereitschaft der Benutzer als Erfolgsfaktor von Social Navigation unterstützenden Anwendungen ist hinsichtlich indirekter Social Navigation nur dann von Bedeutung, wenn die Erstellung von Navigationshinweisen mit einem Aufwand für die Benutzer verbunden ist. Stellt die Gewinnung von Navigationshinweisen ein Randprodukt der Tätigkeit der Anwender dar, so kann dieser Faktor vernachlässigt werden. Problematisch ist, bezogen auf indirekte Social Navigation, daß das Anbieten von Navigationsinformationen meist nicht mit einem unmittelbar wahrnehmbaren Vorteil verbunden ist, sondern der Nutzen, sofern vorhanden, erst zu einem späteren Zeitpunkt, z. B. in Form von personalisierten Informationen, deutlich wird.

Eine weitere offene Fragestellung ist, wie sichergestellt werden kann, daß sich im Verhalten der Anwender die passive Nutzung des Systems, bei der von der Arbeit anderer Anwender profitiert wird, und die aktive Beteiligung in einem Gleichgewicht befinden. Ökonomische Modelle, wie die in dem von Sun betriebenen Java Diskussions-Forum erprobten Duke Dollars<sup>10</sup>, stellen einen möglichen Lösungsansatz dar [Die03].

Social Navigation wird nur dann von den Anwendern akzeptiert werden, wenn aus Sicht der einzelnen Benutzer das Verhältnis von Aufwand und Nutzen positiv ausfällt. Der bezogen auf die Social-Navigation-Funktionalität eines Systems anfallende Aufwand umfaßt dabei nicht nur die Arbeit der aktiven Bereitstellung von Navigationsinformationen. Ferner ist zu berücksichtigen, daß die Evaluation der präsentierten Navigationsinformationen mit Aufwand verbunden sein kann. Was nützt z. B. eine hoch entwickelte Visualisierung von Pfaden im Web, wenn die Interpretation der Grafik die Anwender zusätzlich belastet.

Social Navigation setzt voraus, daß ein System von einer „kritischen Masse“ von Anwendern genutzt wird. Direkte Social Navigation ist nur dann möglich, wenn eine hinreichende Zahl potentieller Kooperationspartner zur Verfügung steht. Indirekte Social Navigation ist nur dann dienlich, wenn resultierend aus der vorhergehenden Benutzung des Systems ausreichend Navigationsinformationen gesammelt werden konnten. Es ist folglich schwer, Nutzer für einen frühzeitigen Eintritt in ein neu gestartetes System zu begeistern, da in dieser Phase vom System nur wenig Nutzen zu erwarten ist. Diese Schwierigkeiten bei der Etablierung einer neuen Social-Navigation-Anwendung werden in

---

<sup>10</sup><http://forum.java.sun.com/rewardFaq.jsp>

der Literatur auch als Kaltstart oder Bootstrap-Problem bezeichnet. Eine weitergehende Betrachtung erfolgt im Rahmen der Darstellung von Collaborative-Filtering-Techniken in Abschnitt 5.2.

Im Falle indirekter Social Navigation kommunizieren die Nutzer über im Informationsraum hinterlassene Artefakte, die als Meta-Informationen in einer engen Verbindung zu den Objekten des Raumes stehen. Unterliegt der Informationsraum einer hohen Dynamik, so besteht das Risiko, daß Navigationsinformationen veralten, wenn beispielsweise Objekte entfernt werden, sich deren Inhalt ändert oder die Strukturen des Raumes verändert werden [Die03]. Zusätzlich ist zu beachten, daß Navigationshinweise unter bestimmten Randbedingungen erstellt worden sein können. Ändern sich diese, so verlieren auch die Navigationshinweise ihre Relevanz. Ein Shop-System, welches basierend auf der automatischen Auswertung des Einkaufsverhaltens der Kunden Kaufempfehlungen berechnet, wird in der Vorweihnachtszeit die in diesem Zeitraum vielgefragten Weihnachtsartikel empfehlen. Ein über die Feiertage hinausgehendes Interesse der Kunden an weihnachtlichen Waren ist jedoch als unwahrscheinlich anzusehen.

Social-Navigation-Systeme, welche aus den aggregierten Spuren vergangener Besucher Navigationshinweise generieren, bergen das Risiko sogenannter Schneeballeffekte. Eine Person geht einen „falschen“ Weg, nachfolgende Personen folgen ihr, was dazu führt, daß die Navigationshinweise für den falschen Weg verstärkt und weitere Personen fehlgeleitet werden [SHLW01]. Die in [MRK97] dokumentierten Erfahrungen aus dem Grouplens-Projekt zeigen, daß bereits hoch bewertete Artikel doppelt so oft wie schlecht bewertete Artikel beurteilt werden. Insbesondere dann, wenn Empfehlungen implizit, z.B. aus Kaufentscheidungen, abgeleitet werden, können Feedbackmechanismen Schneeballeffekte begrenzen [DDH<sup>+</sup>00].

Wie im vorhergehenden Abschnitt als ein Vorteil von Social Navigation herausgestellt wurde, können Nutzer auf diesem Wege an der Gestaltung von Informationsräumen teilhaben. Sie können die Strukturen eines Informationsraumes verändern und die Objekte mit Meta-Informationen, wie beispielsweise Annotationen oder Bewertungen, ergänzen. Dieses steht in Konflikt zu den Interessen von Informationsanbietern, möglichst vollständige Kontrolle über die Präsentation der Inhalte zu behalten. Ein Web-Shop, in dem Besucher Kommentare zu Artikeln hinterlassen können, läuft Gefahr, daß sich auch Leute kritisch zu den angebotenen Waren äußern und dadurch Kaufinteressenten von einem Kauf absehen. Werden kritische Kommentare von den Systembetreibern gelöscht, so kann dieses Akzeptanzprobleme nach sich ziehen.

Wie das Beispiel des Web-Annotationsdienstes „Thirdvoice.com“<sup>11</sup> zeigte, ist mit dem Widerstand von Informationsanbietern zu rechnen, wenn der Versuch unternommen

---

<sup>11</sup>Mit der Einstellung von Thirdvoice.com sind auch Sites wie [saynotothirdvoice.com](http://saynotothirdvoice.com), die es sich zum Ziel gemacht haben, den Kritikern von Thirdvoice.com eine Plattform zu bieten, von der Bildfläche verschwunden.

#### 4. *Social Navigation*

---

wird, einen Informationsraum mittels eines eigenständigen Dienstes um die Möglichkeit des sozialen Navigierens zu ergänzen.

## 5. Empfehlungssysteme

Im täglichen Leben gibt es viele Situationen, in denen Personen mit der Aufgabe, Objekte aus einer großen Menge zu selektieren, konfrontiert werden. Beispielsweise können zeitliche Restriktionen dazu zwingen, sich bei der Zeitungslektüre auf eine Auswahl von Artikeln zu beschränken. Während eines Städteurlaubs können nur einige der lokalen Restaurants ausprobiert werden. Bei der Recherche im Web können nur wenige der von Suchmaschinen als potentiell relevant eingeschätzten Seiten begutachtet werden.

Malone et al. untersuchten in einer Studie die Vorgehensweise von Personen bei der Selektion von Informationen [MGT<sup>+</sup>87]. Sie identifizierten dabei eine Reihe von Regeln, nach denen die Testpersonen bei der Entscheidung, einer Information aus einer Vielzahl von Quellen weitergehende Beachtung zu schenken, vorgingen. Diese Entscheidungsregeln lassen sich in die folgenden drei Kategorien einteilen, welche die unterschiedlichen Vorgehensweisen bei der Informationsfilterung charakterisieren:

**Kognitives Filtern** Diese Methode der Informationsfilterung basiert auf der inhaltlichen Analyse der Objekte eines Suchraumes. Gängige Herangehensweise beim Lesen einer Zeitung ist es, die Artikel kurz anzulesen und dann anhand von Überschrift und Vorspann zu entscheiden, ob es sich lohnt, den gesamten Artikel zu lesen.

**Soziales Filtern** Die Herkunft von Informationen war ein wichtiges Kriterium, nach dem die Testpersonen der von Malone et al. durchgeführten Studie Informationen auswählten. Insbesondere die Beziehung im organisationalen Kontext zwischen dem Autor einer Nachricht und dem Empfänger beeinflusste die Selektionsentscheidung.

Im Kontext dieser Arbeit werden darüber hinaus jegliche Formen von Informationsfilterung, bei der eine Orientierung an der sozialen Textur (siehe Abschnitt 4.7) der Informationen erfolgt, dem sozialen Filtern zugeordnet. Empfehlungen verbreiten sich oftmals per „Mund-zu-Mund-Propaganda“, wobei die individuelle Relevanz einer Empfehlung maßgeblichen von deren Herkunft abhängig ist. Die Filterung von Objekten basierend auf der Beachtung von Empfehlungen ist ein typisches Beispiel für soziales Filtern. Insbesondere dann, wenn die eigenen Kenntnisse und Ressourcen nicht zur Evaluation aller Alternativen eines Suchraumes hinreichend sind oder wenn nicht einmal alle Alternativen bekannt sind, verlassen wir uns auf den Rat von „Experten“.

Soziales Filtern erfordert im Gegensatz zum kognitiven Filtern keine inhaltliche Betrachtung der auszuwählenden Objekte.

**Ökonomisches Filtern** Die Auswahl erfolgt in diesem Fall anhand von Kosten-Nutzen-Kriterien. Dabei werden entweder Objekte ausgewählt, die bei gegebenen Kosten einen hohen Nutzen versprechen, oder solche, welche lediglich geringe Kosten verursachen. Kosten können auch nicht-monetären Größen, wie beispielsweise der Zeit, die das Lesen eines Artikels in Anspruch nimmt, entsprechen. Ein langer Artikel wird also bei Anwendung dieses Filterprinzips nur dann gelesen werden, wenn der Nutzen so hoch erscheint, daß dieser den Zeitaufwand rechtfertigt.

Mit der wachsenden Vielzahl der über elektronische Medien abrufbaren Informationen, wird die Eignung der vorstehend beschreibenden manuellen Filterverfahren zunehmend geringer. Bestrebungen, den Prozess der Informationsfilterung durch Software zu (teil-) automatisieren, werden in diesem Kapitel unter Abwägung der Vor- und Nachteile der jeweiligen Ansätze vorgestellt.

*Automatisches Information Filtering* ist eng verwandt mit *Information-Retrieval-Techniken* [BC92]. Beide Applikationstypen haben das Ziel, Benutzer beim Auffinden relevanter Informationen zu unterstützen und gleichzeitig die Zahl der präsentierten nicht-relevanten Dokumente zu minimieren. Information Retrieval bezeichnet die aktive Suche nach Informationen, motiviert von kurzfristigen Interessen und mit dem Ziel, Wissenslücken zu schließen. Das Informationsbedürfnis wird in Abfragen formuliert, welche auf relativ statische Datenbestände angewandt werden. In Abgrenzung davon ist Information Filtering von grundsätzlichen, langfristigen Benutzerinteressen bestimmt, welche durch Benutzerprofile modelliert werden. Benutzer haben das Ziel, in einem Themengebiet auf dem Laufenden zu bleiben. Information-Filtering-Techniken unterstützen sie bei der passiven Informationssuche in einem dynamischen Datenstrom.

Die Begriffe „automatisches Information Filtering“ und „*Empfehlungssysteme*“ werden im Kontext dieser Arbeit als Synonyme verwendet. Letztendlich sind beides lediglich unterschiedliche Sichtweisen auf Systeme, welche Anwender bei der Informationsselektion oder der Auswahl von Produkten unterstützen. Information-Filtering-Techniken zielen darauf ab, nicht-relevante Informationen auszusortieren. Hingegen ist es das Ziel von Empfehlungssystemen, relevante Informationen zu selektieren bzw. Objekte gemäß ihrer Relevanz in eine Reihenfolge zu bringen.

In Analogie zu den von Malone aufgezeigten Formen manueller Informationsfilterung, existieren zwei grundlegend unterschiedliche Paradigmen der Realisierung von Empfehlungssystemen.

*Eigenschaftsbasierte (Feature-based) Filtertechniken*, in Bezug auf die Auswahl von Medieninhalten ist vornehmlich von *Content-based Filtering* (Abschnitt 5.1) zu sprechen,

---

orientieren sich an der Vorgehensweise des manuellen kognitiven Filterns. Sie selektieren Informationen aufgrund des Abgleichs der Inhalte mit den im Benutzerprofil modellierten Präferenzen.

*Empfehlen-Systeme*, gebräuchlicher ist der englische Begriff *Recommender-Systeme*, aggregieren und kanalisieren die von Nutzern des Systems abgegebenen Empfehlungen. Der hier gewählte Ansatz besteht darin, von einer Analyse der zu empfehlenden Objekte abzusehen und stattdessen ein Medium zu schaffen, über das Objekte bewertende und Empfehlungen suchende Personen kommunizieren können.

In a typical recommender system people provide recommendations as inputs, which the system then aggregates and directs to appropriate recipients. In some cases the primary transformation is the aggregation; in others the system's value lies in its ability to make good matches between the recommenders and those seeking recommendations [RV97].

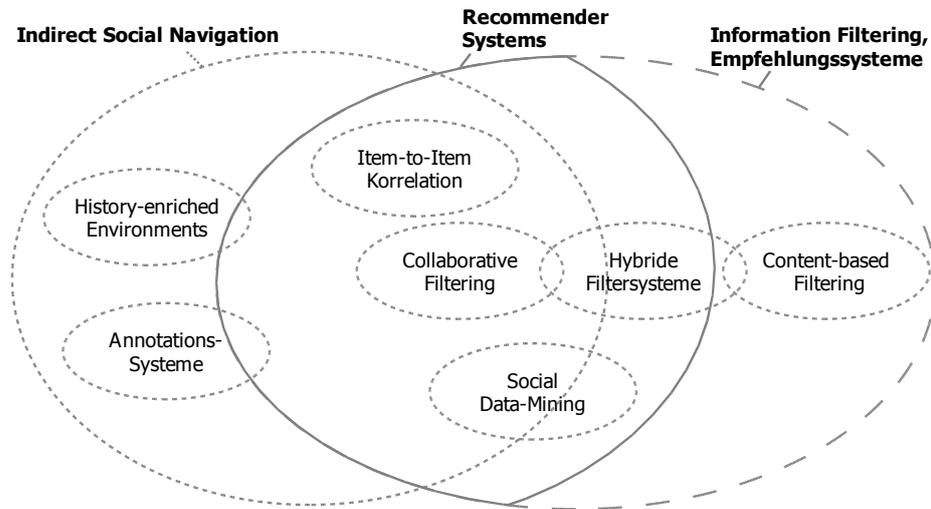
Durch *CollaborativeFiltering*-Verfahren (Abschnitt 5.2) formalisieren Recommender-Systeme den Vorgang des sozialen Filterns. Objekte werden nach dieser Methode nicht vorgeschlagen, indem nach Objekten gesucht wird, deren Eigenschaften dem Benutzerprofil entsprechen, sondern durch Auffinden von Personen mit ähnlichen Benutzerprofilen. Diese fungieren dann als Empfehler und von ihnen hoch bewertete Objekte werden vorgeschlagen, sofern sie vom Rat suchenden Benutzer noch nicht bewertet wurden. Mit dieser Methode wird der Verbreitungsmechanismus der „Mund-zu-Mund-Propaganda“ großen Gemeinschaften von unter Umständen anonymen Benutzern zugänglich gemacht. *Hybride Verfahren* (Abschnitt 5.3) versuchen, die Vorteile von Content-based und Collaborative Filtering zu vereinen. Große Ähnlichkeit mit den Collaborative-Filtering-Algorithmen weist das Verfahren der *Item-to-Item-Korrelation* (Abschnitt 5.4) auf. Hier werden nicht die Profile der Benutzer verglichen, sondern verwandte Artikel basierend auf der Berechnung von Korrelationen zwischen den für die einzelnen Artikel abgegebenen Bewertungen ermittelt. Überdies sind *Social-Data-Mining-Techniken* (Abschnitt 5.5) als eine Möglichkeit der Realisierung von Recommender-Systemen zu nennen. Diese werten diverse Nutzungs- und Strukturinformationen eines Informationsraumes zur Bestimmung von Vorschlägen aus. Die beiden letztgenannten Techniken nehmen eine Sonderstellung ein, da sie im Gegensatz zu den anderen Alternativen keine individualisierten Empfehlungen generieren.

Die Methode des ökonomischen Filterns ist einer softwaregestützten Ausführung nur schwer zugänglich, da sich zwar der mit der Auswahl eines Objektes verbundene Aufwand, sofern einfache Kriterien wie die Textlänge berücksichtigt werden, automatisch feststellen läßt, der Nutzen eines Objektes jedoch von den höchst subjektiven Anforderungen des Benutzers abhängig ist.

Zwischen Recommender-Systemen, welche als eine Teilmenge der Empfehlungssysteme anzusehen sind, und indirekter Social Navigation bestehen starke Bezüge. Die Konzepte sind jedoch weder identisch, noch sind Recommender-Systeme als eine Teilmenge der Social Navigation unterstützenden Anwendungen anzusehen. Abbildung 5.1 soll die Beziehung zwischen den Konzepten Social Navigation und Empfehlungssysteme verdeutlichen und zeigen, wie die wichtigsten Formen von Empfehlungssystemen und exemplarische Beispiele von indirekte Social Navigation unterstützenden Anwendungen einzuordnen sind.

Aus den Bewertungen der Anwender Vorschläge ableitende Recommender-Systeme, insbesondere Collaborative-Filtering-Anwendungen und Systeme, die basierend auf Item-to-Item-Korrelationsanalysen Empfehlungen generieren, sind dem Konzept indirekter Social Navigation zuzuordnen (siehe Abschnitt 4.5). Jedoch ist zu betonen, daß die Generierung von Empfehlungen durch hybride Ansätze, welche Collaborative- und Content-based-Filtering-Techniken kombinieren, nur teilweise auf einer sozialen Komponente beruht. Bisher noch nicht bewertete Objekte werden durch diese Systeme einer inhaltlichen Analyse unterzogen. Folglich weist die Orientierung an den Empfehlungen in diesem Fall nicht die Charakteristik des sozialen Navigierens auf. Social-Data-Mining-Methoden zur Empfehlungsberechnung sind nur dann dem Konzept indirekter Social Navigation zuzuordnen, wenn die ausgewerteten Informationen ausschließlich auf Handlungen der Besucher eines Informationsraumes zurückzuführen sind, und die von den Urhebern des Raumes, z. B. durch das Setzen von Links, gemachten Vorgaben unberücksichtigt bleiben.

Recommender-Systeme und indirekte Social Navigation sind nicht identisch. Wie vorstehend beschrieben wurde, ist die Funktion von Recommender-Systemen teilweise nicht mit den Kriterien von Social Navigation zu vereinbaren. Überdies ist indirekte Social Navigation als ein allgemeineres Konzept anzusehen, das weitere Anwendungsklassen umfaßt. History-Enriched Environments, wie die von Wechselblatt realisierte Visualisierung von Pfaden im Web (siehe Abschnitt 4.5), veranschaulichen die Aktivitäten vorheriger Benutzer ohne daraus konkrete Empfehlungen abzuleiten. Letztendlich ist es eine Frage der Präsentation von Informationen, ob History-Enriched Environments den Recommender-Systemen zuzuordnen sind. Die Annotation von Links mit der Prozentzahl derjenigen Besucher, welche die aktuelle Seite über den jeweiligen Link verlassen haben, kann als Information-Filtering-System betrachtet werden. History Enriched Environments und Social-Data-Mining-Techniken weisen große Übereinstimmungen auf. Die Auswertung der Zugriffsprotokolle eines Web-Servers, um daraus Navigationsempfehlungen abzuleiten, läßt sich einerseits vom Standpunkt der History Enriched Environments betrachten und stellt andererseits ein auf Social-Data-Mining-Techniken basierendes Recommender-System dar. Anwender können durch Annotations-Systeme Objekte eines Informationsraumes mit Anmerkungen versehen. Dabei hängt es stark vom Einsatzkontext, der Struktur und den Inhalten der konkreten Annotationen ab, ob ein Bezug zu



**Abbildung 5.1.:** Zusammenhang zwischen Social Navigation und Recommender-Systemen

den Recommender-Systemen zu sehen ist. Die Kombination von in Freitextkommentaren abgefaßten Produkttests mit Bewertungen auf einer numerischen Skala, wie sie in der Verbraucher-Community *ciao.com* anzutreffen ist, kann als Grenzfall von Annotations- und Recommender-Systemen gesehen werden.

## 5.1. Content-based Filtering

*Content-based-Filtering-Systeme* analysieren Eigenschaften und Inhalte von Objekten eines Informationsraumes und vergleichen diese mit den in einem Benutzerprofil modellierten Präferenzen. Dokumente, deren Inhalt bzw. Metadaten dem Profil des eine Empfehlung suchenden Benutzers entsprechen, werden vorgeschlagen.

Einfache Benutzerprofile bestehen aus Vektoren gewichteter Keywords. Diese lassen sich dann entsprechend dem Vektorraummodell des Information Retrieval [SM83] mit Dokumenten abgleichen. Dokumente, die vom Benutzer als relevant eingestuft wurden, werden in das Benutzerprofil integriert. Auf diese Weise wird das Wissen des Systems über den Benutzer kontinuierlich ausgebaut. Dieser Vorgang wird in der Literatur als *Relevance Feedback* bezeichnet.

Content-based-Filtering-Verfahren suchen Objekte, die eine Ähnlichkeit zu bereits durch den jeweiligen Benutzer positiv beurteilten Objekten aufweisen. Einem Benutzer, der

bereits mehrere DVDs mit dem Darsteller Chevy Chase erworben hat, würden beispielsweise weitere Filme mit diesem Hauptdarsteller oder Filme des Genres Komödie vorgeschlagen werden.

### 5.1.1. Vorteile von Content-based Filtering

Vorteilhaft an der Vorgehensweise des Content-based Filtering ist es, daß auch Objekte, die noch von keinem anderen Nutzer bewertet wurden, vorgeschlagen werden können.

Die Objektivität, mit der beim Content-based Filtering Objekte selektiert werden, kann als Vorteil dieser Technik gewertet werden, wenn die Nachvollziehbarkeit von Empfehlungen bedeutsam ist.

### 5.1.2. Nachteile von Content-based Filtering

Shardanand und Maes benennen Limitationen von Content-based-Filtering-Techniken, mit denen sie die Entwicklung von Collaborative-Filtering-Techniken motivieren [SM95]:

Die Notwendigkeit der inhaltlichen Betrachtung der potentiell zu empfehlenden Objekte bedingt die maschinelle Verarbeitbarkeit der Objekte oder die Zuweisung von Meta-Daten, anhand derer die Empfehlungsgenerierung erfolgen kann. Beim gegenwärtigen Stand der Technik ist die automatische inhaltliche Analyse multimedialer Daten nicht durchführbar. Content-based-Filtering-Techniken sind folglich kaum für die Auswahl von Musik, Bildern und Filmen geeignet, sofern diese nicht hinreichend manuell durch Meta-Daten beschrieben werden, was in vielen Situationen durch Ressourcenbeschränkungen nicht praktikabel sein dürfte.

Content-based-Filtering-Techniken neigen zur Überspezialisierung. Tendenziell werden Objekte empfohlen, die eine Ähnlichkeit zu den in der Vergangenheit positiv gewerteten Objekten aufweisen. Der Suchraum wird durch die strikte Orientierung an den durch Benutzerprofile spezifizierten Filterregeln stark eingeschränkt. Inhaltlich verwandte Objekte, die den Kriterien nicht entsprechen, werden den Anwendern unter Umständen vorenthalten. Durch den Einsatz von Content-based Filtering geht auch die Chance verloren, bei der Recherche durch einen glücklichen Zufall interessante Dokumente zu entdecken, die nicht unmittelbar mit den ursprünglichen Intentionen der Recherche übereinstimmen<sup>1</sup>.

Die objektiven Kriterien, nach denen Content-based-Filtering-Techniken Dokumente auswählen, mögen als Vorteil erscheinen, denn sie machen die Entstehung von Empfehlungen zu einem gewissen Grad transparent. Kriterien wie Qualität, Stil oder Sichtweise

---

<sup>1</sup>In der englischsprachigen Literatur wird diese Art Objekte zu finden als Serendipity Effect bezeichnet

des Autors bleiben jedoch bei dieser Vorgehensweise der automatischen Informationsfilterung unberücksichtigt. Dokumente können beispielsweise das gleiche Thema behandeln, was sich aus Sicht des Filteralgorithmus im Vorhandensein der gleichen Schlüsselwörter manifestiert, aber gravierende Qualitätsunterschiede aufweisen, die dem Filtersystem bei der Klassifikation von Dokumenten verborgen bleiben.

Resümierend läßt sich feststellen, daß einige der Probleme von Content-based-Filtering-Techniken auf Defizite in der automatischen Verarbeitung von Dokumenten auf einer semantischen Ebene zurückführen sind. Texte, die nicht die gesuchten Schlüsselwörter, wohl aber deren Synonyme enthalten, würden ausgefiltert werden. Enthält umgekehrt das Benutzerprofil Homonyme, so werden Dokumente selektiert, die nicht den Interessen des Anwenders entsprechen.

## 5.2. Collaborative Filtering

Von Menschen, die wir gut kennen und deren Geschmack mit unserem übereinstimmt, lassen wir uns gerne einen Tip geben, welchen Film wir sehen oder welche Musik wir hören sollten. Eine Auswahlentscheidungen zu treffen, indem kompetente Personen um Rat gefragt werden, ist nur dann praktikabel, wenn zum einen die Zahl der Alternativen überschaubar ist und andererseits potentielle Ratgeber persönlich bekannt sind. *Collaborative-Filtering-Techniken*<sup>2</sup> haben es zum Ziel, die Verbreitung von Empfehlungen per „Mund-zu-Mund-Propaganda“ zu automatisieren. Somit entstehen Empfehlungssysteme, die einerseits bezüglich der Zahl der Benutzer und der Größe des Suchraumes skalierbar sind und andererseits die Charakteristik persönlicher Empfehlungen, nämlich die Einbeziehung menschlichen Urteilsvermögens, beibehalten.

Der Ausdruck *Collaborative Filtering* geht auf eine Veröffentlichung über das „Information Tapestry System“ [GNOT92] zurück. Es wurde am Xerox PARC (Palo Alto Research Center) als Reaktion auf den mit der Einführung eines E-Mail-Systems einhergehenden Information Overload zur Filterung von Nachrichten unternehmensinterner Mailing-Listen entwickelt.

Collaborative filtering simply means that people collaborate to help one another perform filtering by recording their reactions to documents they read [GNOT92].

Ähnlich, wie der Moderator einer Newsgroup<sup>3</sup> entscheidet, welche Nachrichten an die Abonnenten versandt werden, konnte jeder Nutzer des Tapestry-Systems individuell

---

<sup>2</sup>In der Literatur wird auch der Begriff *Social Information Filtering* verwendet, teilweise als Synonym [SM95], teilweise mit in Nuancen abweichender Bedeutung [Ols98].

<sup>3</sup>Natürlich werden nicht alle Newsgroups moderiert

die interessantesten, lesenswerten eingehenden Nachrichten selektieren. Diese Bewertungen wurden, ergänzt um die Protokollierung von Reaktionen auf Nachrichten, beispielsweise deren Beantwortung, abgespeichert. Sie waren durch an das System gestellte Abfragen, wie „zeige mir alle neuen Nachrichten, die Benutzer Meier oder Müller als interessant bewertet haben“, nutzbar. Filterregeln mussten explizit formuliert werden, was die persönliche Kenntnis der anderen Nutzer und deren Präferenzen bedingt. Dieser Ansatz mag zum Einsatz in Arbeitsgruppen geeignet sein, zur Filterung von Informationen durch Communities ist er es nur bedingt, denn hier ist die Voraussetzung der persönlichen Bekanntschaft der beteiligten Personen nur in Ausnahmefällen gegeben (siehe Abschnitt 2.1.4).

Nachfolgende Collaborative-Filtering-Systeme, zu nennen sind hier insbesondere die Projekte „Grouplens“ [RIS<sup>+</sup>94,KMM<sup>+</sup>97] und „Ringo“ [SM95], zeichneten sich dadurch aus, daß sie automatisch für jeden Benutzer eine Auswahl der relevanten Empfehler trafen. Infolgedessen konnte der Einsatzbereich von Collaborative Filtering auf große Communities und somit auf einen erweiterten Kreis von Objekte bewertenden Personen ausgedehnt werden.

Das Usenet hat mit zunehmender Zahl publizierter Nachrichten einen Teil seiner Nützlichkeit eingebüßt, da es zunehmend schwerer wurde, die relevanten Beiträge zu selektieren. Das Grouplens-System bietet Anwendern die Möglichkeit, über modifizierte Newsreader Beiträge zu bewerten und prognostizierte Bewertungen abzufragen. Mittels Korrelationsanalysen der von den Benutzern abgegebenen Bewertungen werden Ähnlichkeiten zwischen Anwendern bestimmt, die dann zu einer gewichteten Mittelwertbildung der zu einem Beitrag abgegebenen Bewertungen verwendet werden. Die durch Grouplens errechneten Vorhersagen, der für einen Anwender zu erwartenden Relevanz einer Nachricht, basieren auf dem Urteil anderer Nutzer des Systems, ohne daß deren Identität gegenüber der Empfehlungen abfragenden Person aufgedeckt wird. Details zur Architektur des Grouplens-Prototyps werden in Abschnitt 6.2 dargestellt.

Ein ähnlicher Ansatz wurde im Projekt Ringo, einem Empfehlungssystem für Musik, verfolgt. Diese Domäne erscheint für die Evaluation von Collaborative-Filtering-Techniken prädestiniert, denn es kann davon ausgegangen werden, daß die Meinungen der Benutzer zu einzelnen Musiktiteln nicht zufällig verteilt sind, sondern sich an Genres und Interpreten orientieren. Überdies ist der Musikgeschmack meistens langfristiger Natur. Anders als Grouplens, das primär auf die Filterung der eingehenden Meldungen abzielt, ist es das Ziel von Ringo, aktiv Empfehlungen zu generieren, nachdem die Präferenzen eines neuen Nutzers durch eine Auswahl von zu bewertenden Titeln kennengelernt wurden. Durch Ringo können Musikliebhaber neue, ihrem Geschmack entsprechende, Musiktitel entdecken, welche von Interpreten stammen, die sie vorher möglicherweise gar nicht kannten.

Die Unterschiede zwischen dem Tapestry-System auf der einen Seite und Grouplens

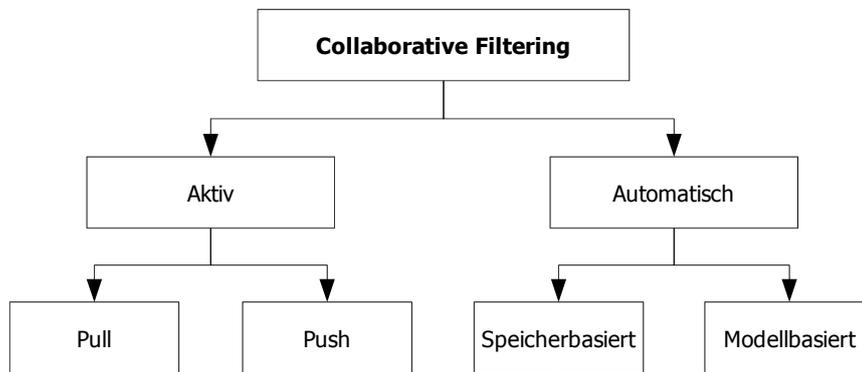


Abbildung 5.2.: Klassifikation von Collaborative-Filtering-Verfahren

bzw. Ringo auf der anderen Seite liegen im Automatisierungsgrad der Zuordnung von Bewertungen abgebenden und Empfehlungen suchenden Personen. Collaborative-Filtering-Systeme lassen sich gemäß diesem Klassifizierungsmerkmal, wie in Abbildung 5.2 dargestellt wird, in *aktive* und *automatische* Systeme unterteilen.

Für die aktiven Filtersysteme ist es charakteristisch, daß sich die Systemfunktionalität darauf beschränkt, mit Personen assoziierte Bewertungen von Objekten abzuspeichern und abfragbar zu machen bzw. als Kommunikationsmedium den Austausch von Empfehlungen zu fördern. Weitergehend ist zwischen *pull-active* und *push-active* Filtering zu unterscheiden [KR03] (siehe Abbildung 5.2).

Müssen die Benutzer manuell Filterregeln formulieren, welche Angaben zu den Personen enthalten, deren Urteil in die persönliche Informationsfilterung einbezogen werden soll, so spricht man von *pull-active* Collaborative Filtering. Der der Zusatz „pull“ bezieht sich darauf, daß die Anwender Beurteilungen bzw. vorgeschlagene Dokumente vom Server abfragen müssen. Das Tapestry-System läßt sich in diese Kategorie einordnen, da dort ebenfalls die Auswahl von Informationen die Initiative des jeweiligen Anwenders erfordert.

Push-active Collaborative Filtering basiert darauf, daß Benutzer anderen Personen, welche vermutlich von den Informationen profitieren, Verweise auf Dokumente zusenden [ME95]. Hier geht die Initiative von der eine Information bewertenden und potentielle Interessenten auswählenden Person aus. Der Empfänger der Hinweise bekommt diese automatisch zugestellt und befindet sich in der Rolle eines passiven Konsumenten. Diese Technik mag beim Auffinden relevanter Informationen hilfreich sein, sofern die Objekte auswählenden und zusendenden Personen den Informationsbedarf der die Informationen erhaltenden Personen gut einschätzen können. Kontraproduktiv könnte sich jedoch ein

entsprechendes System auswirken, wenn beispielsweise ein Entscheidungsträger in einem Unternehmen mit sehr vielen Zusendungen überhäuft wird.

Automatisches Collaborative Filtering – in diese Kategorie sind die vorstehend aufgeführten Systeme GroupLens und Ringo einzuordnen – besteht darin, daß die zur Generierung von Empfehlungen notwendige Auffindung gleichgesinnter Personen algorithmisiert wird. Das genaue Verfahren wird nachfolgend erklärt. Sofern es nicht explizit angegeben wird, bezieht sich im weiteren Verlauf des Dokuments der Ausdruck Collaborative Filtering auf das automatisierte Verfahren.

Typische Content-based-Filtering-Systeme empfehlen einem Benutzer Gegenstände, die ähnlich zu den Objekten sind, die er in der Vergangenheit mochte. Hingegen werden von auf Collaborative Filtering basierenden Empfehlungssystemen Objekte empfohlen, die Benutzer mit ähnlichem Geschmack hoch bewertet haben. Prämisse beim kollaborativen Filtern ist, daß Personen, deren Meinung in der Vergangenheit übereinstimmte, auch neue Beiträge ähnlich beurteilen werden [RIS<sup>+</sup>94]. Ob dieser Schluß von vergangener Übereinstimmung auf zukünftige Ähnlichkeit in den Beurteilungen zulässig ist, hängt u. a. davon ab, aus welchen Domänen die zu bewertenden Objekte stammen. Enthält das Benutzerprofil Bewertungen aus verschiedenen Gebieten, so kann die Qualität der Vorhersagen zweifelhaft sein. Eine übereinstimmende Meinung bezüglich interessanter und qualitativ hochwertiger Seiten zum Thema Programmierung läßt kaum Schlüsse auf einen gleichartigen Geschmack bei der Auswahl von Kochrezepten zu.

Nachstehend soll anhand des in Grafik 5.3 schematisch dargestellten Beispiels die Funktionsweise von Collaborative-Filtering-Techniken veranschaulicht werden. Einem Benutzer „XY“, dessen Benutzerprofil seine z.B. aus vorherigem Kauf- oder Navigationsverhalten abgeleitete Präferenz für „Buchtitel 2“, „Buchtitel 4“, „Link 2“ und „Link 3“ dokumentiert, sollen Vorschläge für weitere ihn potentiell interessierende Bücher und Web-Seiten vorgelegt werden. Der Vergleich des Profils von „XY“ mit den weiteren in der Datenbasis des Systems gespeicherten Profilen ergibt eine hohe Übereinstimmung mit „Benutzer A“. Dieser teilt mit „XY“ das Interesse an „Buchtitel 2“, „Buchtitel 4“ und „Link 2“, was ihn für diesen das System abfragenden Benutzer zum bevorzugten Empfehler macht. Auch „Benutzer A“ hat „Buchtitel 1“ positiv bewertet. Folglich wird dieses Buch vorge schlagen.

Für die formale Betrachtung des Collaborative Filtering Prozesses gelten folgende Vereinbarungen:

$U = \{1, 2, \dots, a, \dots, m\}$	Menge des Benutzer-Identifikatoren
$a \in U$	Identifikator des aktiven Benutzers
$I = \{1, 2, \dots, c, \dots, n\}$	Menge der Objekt-Identifikatoren
$I_i \subseteq I$	Menge der vom Benutzer $i$ bewerteten Objekte
$v_{i,j}$	Wertung des Benutzers $i$ zum Objekt $j$

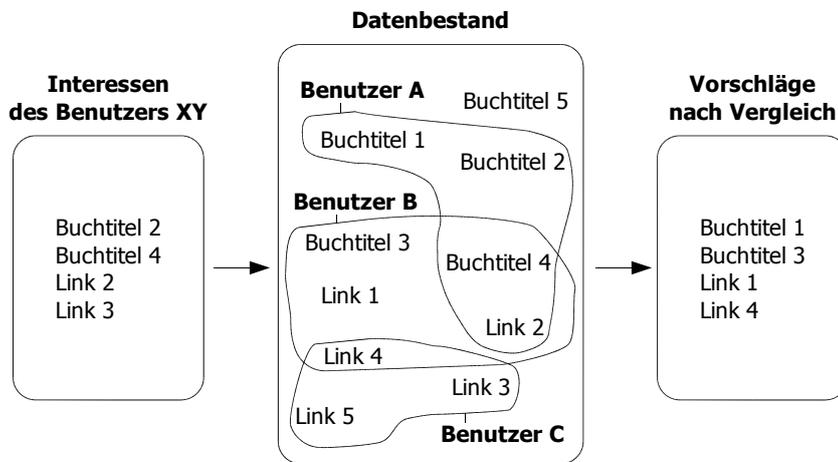


Abbildung 5.3.: Anschauliche Erklärung von Collaborative Filtering (Quelle: [JZ00])

Die Aufgabe des Algorithmus besteht darin, für den aktiven Benutzer  $a$  eine Vorhersage  $p_{a,j}$  der zu erwartenden Relevanz des Objektes  $j$  zu treffen oder eine sortierte Liste  $(T_1, T_2, \dots, T_N)$  von Empfehlungen zu generieren (siehe Abbildung 5.4) [SKKR01].

Da beim Collaborative Filtering keine inhaltliche Betrachtung der Objekte durchgeführt wird, können ausschließlich Prognosen zu Objekten errechnet werden, welche bereits bewertet wurden. Voraussetzung für den Collaborative-Filtering-Algorithmus ist folglich,

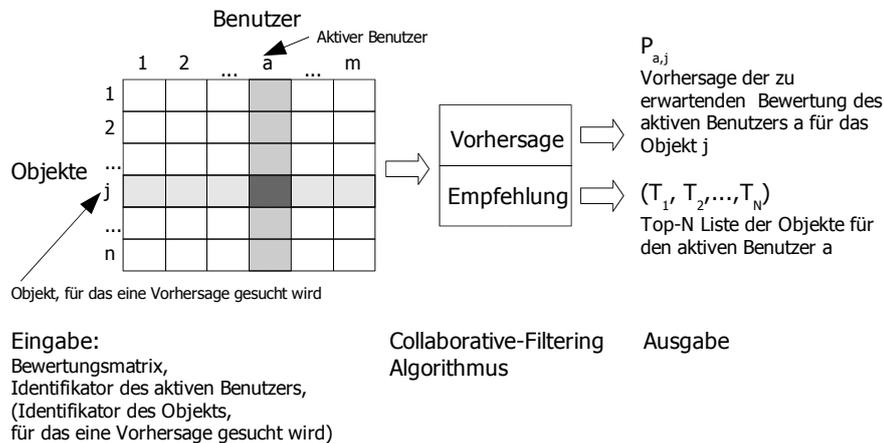


Abbildung 5.4.: Collaborative-Filtering-Prozeß (In Anlehnung an [SKKR01])

daß das System bereits über einige Benutzer verfügt, welche Bewertungen abgegeben haben. Die Bewertungen entsprechen Werten aus einer vorgegebenen diskreten oder kontinuierlichen Skala. Sie können expliziten Angaben entstammen oder implizit durch Beobachtung des Benutzerverhaltens erstellt werden. Der Kauf eines Artikels könnte beispielsweise als ein positives Votum interpretiert werden. Bewertungen werden in einer Bewertungsmatrix zusammengefaßt. Die Bewertung  $v_{i,j}$  des Benutzers  $i$  zum Objekt  $j$  wird dabei in der  $i$ . Spalte und  $j$ . Zeile der Matrix erfaßt. Die Bewertungsmatrix repräsentiert das gesamte Wissen des Systems über die Präferenzen der Benutzer und ist die Datenbasis für den Collaborative-Filtering-Algorithmus. Einzelne Benutzerprofile lassen sich als Spaltenvektoren der Bewertungsmatrix interpretieren.

Um persönliche Empfehlungen zu erhalten, muß der aktive Benutzer in einer Initialisierungsphase durch Abgabe von Bewertungen ein Benutzerprofil aufgebaut haben. Dieses wird fortwährend mit jeder abgegebenen Bewertung weiter verfeinert. Das System lernt somit mit fortschreitender Dauer der Benutzung die Interessen des Benutzers kontinuierlich besser kennen und die Personalisierung der Empfehlungen kann mit zunehmender Detaillierung des Benutzerprofils verbessert werden.

Der Collaborative-Filtering-Algorithmus läßt sich in zwei Phasen unterteilen. In der ersten Phase müssen die im System vorliegenden Benutzerprofile mit dem Profil des aktiven Benutzers verglichen werden. Ziel ist es, die *Nearest Neighbours* zu ermitteln, welche die größte Übereinstimmung in den vergangenen Bewertungen mit dem aktuellen Benutzer aufweisen und somit als bevorzugte Empfehler verwendet werden [RIS<sup>+</sup>94]. Je mehr Nutzer das System besitzt, je größer also die Datenbasis ist, desto höher ist die Wahrscheinlichkeit, andere Personen mit hoher Übereinstimmung in den Meinungen zu finden.

In der zweiten Phase wird dann die Vorhersage durch gewichtete Summation der von den anderen Systemnutzern abgegebenen Bewertungen ermittelt [RIS<sup>+</sup>94]. Eine sortierte Liste von Vorschlägen läßt sich beispielsweise dadurch ermitteln, daß die Benutzer gemäß ihrer Ähnlichkeit zum Profil des aktiven Benutzers sortiert werden und Objekte empfohlen werden, die von ihnen hoch bewertet wurden und sich nicht im Profil des eine Empfehlung suchenden Benutzers befinden.

Konkrete Instanzen von Collaborative-Filtering-Algorithmen lassen sich in *speicherbasierte* und *modellbasierte* Techniken einteilen [BHK98] (siehe Abbildung 5.2). Speicherbasierte Verfahren nutzen bei jeder Anfrage die kompletten zur Verfügung stehenden Daten um Ähnlichkeiten zwischen den Nutzern zu berechnen. Exemplarisch für diese Kategorie wird in Abschnitt 5.2 der Pearson-r-Algorithmus vorgestellt.

Modellbasierte Verfahren transformieren die Datenbasis in regelmäßigen Zeitabständen in ein Modell, welches die Beziehungen zwischen den Benutzern ausdrückt und zur Generierung von Empfehlungen verwendet werden kann. Die aufwendige Operation, ähnliche

		Benutzer			
		Ken (1)	Lee (2)	Meg (3)	Nan (4)
Meldung Nr.	1	1	4	2	2
	2	5	2	4	4
	3			3	
	4	2	5		5
	5	4	1		1
	6	?	2	5	

**Bewertungsmatrix**

**Abbildung 5.5.:** Wie wird Benutzer Ken Artikel 6 beurteilen? (Quelle: [RIS<sup>+</sup>94])

Nutzer zu identifizieren, kann auf diesem Wege von der zeitkritischen Bearbeitung von Anfragen nach Empfehlungen losgelöst werden. Beispielfür modell-basierte Verfahren sind Clustering-Techniken und Bayesianische Netzwerke zu nennen [UF98].

### 5.2.1. Pearson-r-Algorithmus

Im folgenden soll anhand eines [RIS<sup>+</sup>94] entnommenen Beispiels der Pearson-r-Algorithmus vorgestellt werden. Die Details des Algorithmus sind an die Darstellung in [BHK98] angelehnt. Abbildung 5.5 zeigt die Bewertungsmatrix eines Recommender-Systems. Die Benutzer haben gemäß subjektiven Kriterien die Artikel auf einer Skala von 1 (nicht relevant) bis 5 (sehr relevant) bewertet. Benutzer „Ken“ stellt an das System die Anfrage, wie relevant Meldung 6 für ihn vermutlich sei. Gesucht ist also die vorhergesagte Bewertung  $p_{a,m}$  für den Benutzer  $a = 1$  und die Meldung  $m = 6$ . Die Berechnung wird in drei Schritten durchgeführt:

1. Bestimmung des Durchschnittsvotums  $\bar{v}_i$  für den aktiven Benutzer  $a$  und alle Benutzer  $b \in U$  mit  $m \in I_b$ .

$$\bar{v}_i = \frac{1}{|I_i|} \sum_{j \in I_i} v_{i,j} \quad (5.1)$$

Wie in Abbildung 5.6 ersichtlich ist, ergeben sich für die Benutzer des Beispiels die Durchschnittsvoten  $\bar{v}_{Ken} = 3$ ,  $\bar{v}_{Lee} = 2,8$  und  $\bar{v}_{Meg} = 3,5$ .

2. Berechnung der Korrelationskoeffizienten  $w_{a,j}$  zwischen dem Benutzerprofil  $a$  und dem Vergleichsprofil  $i$  für alle Benutzer  $i$  mit  $m \in I_i$  und  $I_a \cap I_i \neq \{\}$ .

$$w_{a,i} = \frac{\sum_{k \in I_a \cap I_i} (v_{a,k} - \bar{v}_a) (v_{i,k} - \bar{v}_i)}{\sqrt{\sum_{k \in I_a \cap I_i} (v_{a,k} - \bar{v}_a)^2 \sum_{k \in I_a \cap I_i} (v_{i,k} - \bar{v}_i)^2}} \quad (5.2)$$

Die Korrelationskoeffizienten sind ein Maß dafür, wie „ähnlich“ andere Benutzer, welche die Meldung  $m$  bereits bewertet haben, zum aktiven Benutzer  $a$  sind. Mögliche Werte liegen im Intervall von -1 bis 1. Liegt der Korrelationskoeffizient nahe bei 0, so gleichen sich die Profile nicht, denn einige Meldungen wurden in der Vergangenheit übereinstimmend und andere entgegengesetzt bewertet. Hatten die Benutzer in der Vergangenheit Meldungen in der Tendenz, also der Abweichung von der jeweiligen Durchschnittsbewertung, übereinstimmend beurteilt, so wird der Wert bei 1 liegen. Demgegenüber signalisiert ein Korrelationskoeffizient von -1 durchgängig gegenteilige Voten bei allen gemeinsam bewerteten Meldungen. Zu beachten ist, daß bei dieser Art, die Übereinstimmung in den Bewertungen zu ermitteln, nicht die absoluten Werte, sondern die Abweichungen vom mittleren Votum berücksichtigt werden. Der ermittelte Korrelationskoeffizient wird also maßgeblich von den Bewertungen bestimmt, die deutlich vom Durchschnittsvotum des Benutzers abweichen und von denen folglich eine höhere Aussagekraft angenommen werden kann.

Ken und Meg haben jeweils die Artikel 1 und 2 bewertet. Den ersten Artikel beurteilten sie übereinstimmend schlechter, den zweiten besser als ihr Durchschnittsvotum. Die Korrelation zwischen Ken und Lee beträgt somit 1. Demgegenüber waren Ken und Lee in der Vergangenheit gegenteiliger Auffassung. Der Korrelationswert  $w_{Ken, Lee}$  beträgt -0,8 (siehe Abbildung 5.6).

3. Abschließend wird die Kalkulation der Vorhersage durch gewichtete Summation der Abweichungen der Bewertungen der Benutzer von ihrem Durchschnittsvotum ausgeführt.

$$p_{a,m} = \bar{v}_a + \frac{\sum_{i \in \{b | b \in U \wedge m \in I_b\}} w_{a,i} (v_{i,m} - \bar{v}_i)}{\sum_{i \in \{b | b \in U \wedge m \in I_b\}} |w_{a,i}|} \quad (5.3)$$

Es wird unterstellt, daß der aktive Benutzer auch für die fragliche Meldung mit seinem Grundvotum stimmen wird. Zu diesem Wert werden dann mit dem jeweiligen Korrelationsfaktor gewichtet alle von anderen Nutzern des Systems zu dieser Meldung abgegebenen Wertungen aufsummiert. Eine denkbare Modifikation des Algorithmus besteht darin, nur die Stimmen derjenigen Benutzer zu berücksichtigen, zu denen die durch den Korrelationskoeffizienten ermittelte Ähnlichkeit einen

		Benutzer				
		Ken (1)	Lee (2)	Meg (3)	Nan (4)	
Meldung Nr.	1	1	4	2	2	$\bar{v}_{Ken} = 3$
	2	5	2	4	4	$\bar{v}_{Lee} = 2,8$
	3			3		$\bar{v}_{Meg} = 3,5$
	4	2	5		5	
	5	4	1		1	
	6	<b>4,2</b>	2	5		

$w_{Ken, Lee} = -0,8$   
 $w_{Ken, Meg} = 1$

**Abbildung 5.6.:** Berechnung der vorhergesagten Bewertung des Artikels 6 für den Benutzer Ken (Quelle: [RIS<sup>+</sup>94])

definierten Schwellenwert übersteigt. Der vorhergesagte Wert kann die Intervallgrenzen der Bewertungsskala überschreiten, in diesem Fall ist der minimale bzw. maximale Wert anzunehmen. Man beachte, daß auch negative Korrelationen in die Berechnung eingehen. Der vorhergesagte Wert wird folglich nicht nur von den Meinungen gleichgesinnter Anwender dominiert, sondern auch von Personen beeinflusst, welche überwiegend konträrer Meinung waren.

Für die Berechnung der vorhergesagten Bewertung des Artikels 6 für den Benutzer Ken ergibt sich, wie in Abbildung 5.6 gezeigt wird, folgende Rechnung:

$$p_{Ken,6} = 3 + \frac{-0,8 \cdot (2 - 2,8) + 1 \cdot (5 - 3,5)}{0,8 + 1} \approx 4,2$$

Meldung 6 wurde durchschnittlich mit 3,5 bewertet. Demgegenüber fällt die errechnete Vorhersage mit einem Wert von 4,2 positiver aus, da Benutzer Ken, der die Meldung mit 5 bewertet hat, die Summenbildung dominiert<sup>4</sup>.

<sup>4</sup>Die Abweichung gegenüber dem Rechenbeispiel in [RIS<sup>+</sup>94] ist zum darauf zurückzuführen, daß bei der Ermittlung des Durchschnittsvotums der einzelnen Benutzer über alle Bewertungen im jeweiligen Profil und nicht nur über die mit dem aktiven Benutzer gemeinsam bewerteten Artikel aufsummiert wurde.

### 5.2.2. Vorteile von Collaborative Filtering

Collaborative Filtering macht Empfehlungen basierend auf subjektiven Kriterien, wie der Beurteilung von Qualität, möglich [Koc01b]. Nicht der Computer bewertet Objekte, sondern die Empfindungen anderer Benutzer werden berücksichtigt. Durch Software läßt sich leicht eine Menge von Artikeln nach bestimmten Stichwörtern durchsuchen, aber es wird beispielsweise problematisch sein, algorithmisch diejenigen Beiträge zu selektieren, welche gemessen an journalistischen Qualitätsstandards als besonders hochwertig anzusehen sind. Menschen haben weit weniger Probleme mit der semantischen Beurteilung von Informationen. Dieses Potential wird durch Collaborative Filtering nutzbar gemacht.

Durch Collaborative Filtering werden domain-übergreifende Empfehlungen durchführbar. Basierend auf der Analyse von Bewertungen können aus der Präferenz für ein Objekt auf Vorlieben für weitere Objekte geschlossen werden, auch wenn die Objekte in keiner an formalen Attributen festzumachenden Beziehung zueinander stehen.

Automatisches Collaborative Filtering hat den Vorzug, daß es nicht erforderlich ist, die individuell relevanten Empfehler persönlich zu kennen [SM95]. Die Herkunft von Empfehlungen kann anonym bleiben. Die Person, welche Empfehlungen erhält, muß nicht notwendigerweise wissen, zu welchen Anwendern sie große Interessensähnlichkeiten aufweist. Dies ist sicherlich unter dem Aspekt des Schutzes der Privatsphäre der Anwender vorteilhaft. Persönliche Beziehungen bleiben jedoch in automatischen Collaborative-Filtering-Systemen bei der Generierung von Empfehlungen unberücksichtigt.

### 5.2.3. Nachteile von Collaborative Filtering

Das Potential von Collaborative-Filtering-Techniken basiert auf der Einbeziehung menschlichen Urteilsvermögens. Aus diesem Ansatz resultieren jedoch auch einige Probleme. Dabei sind die im Abschnitt über Social Navigation gemachten Angaben übertragbar (siehe Abschnitt 4.8). Darüber hinaus werden in diesem Abschnitt spezifische Herausforderungen des Entwurfs von Collaborative-Filtering-Anwendungen aufgeführt.

Ein Collaborative-Filtering-System zu etablieren gestaltet sich diffizil. Einerseits kann ein leeres System keine Objekte empfehlen, da, wie vorstehend erwähnt wurde, der Empfehlungsalgorithmus auf der Auswertung der von den Benutzern abgegebenen Voten basiert, andererseits fällt es schwer, in der Startphase des Systems neue Anwender zu gewinnen. Diese haben von Beginn an die hohe Erwartungshaltung, personalisierte Empfehlungen zu erhalten – ein Anspruch, dem das System wiederum ohne eine hinreichende Zahl aktiver Benutzer nicht entsprechen kann. Ein Aspekt dieses sogenannten *Bootstrap-Problems*<sup>5</sup> ist es, daß Personen, welche in der Etablierungsphase des Systems

---

<sup>5</sup>Alternative Bezeichnungen sind Kaltstart- bzw. Cold-Start-Problem

Bewertungen eingeben gegenüber späteren Benutzern, die von ihrer Arbeit profitieren, tendenziell benachteiligt werden. Diese Personen nehmen Aufwand auf sich, ohne daß dieser durch personalisierte Empfehlungen kompensiert wird. Ein Sachverhalt, der auch als *Early-Rater-Problem* bezeichnet wird [DDH<sup>+</sup>00]. Festzuhalten ist, daß auf reinen Collaborative-Filtering-Algorithmen basierende Empfehlungssysteme eine lange Initialisierungsphase benötigen, wobei der Nutzen des Systems mit der Zahl aktiver Anwender steigt.

Eine weitere Ausprägung des Kaltstartproblems ist mit dem Neueinstieg in das laufende System verbunden. Personalisierte Empfehlungen lassen sich nur dann berechnen, wenn ein hinreichend detailliertes Benutzerprofil vorliegt. Infolgedessen wird ein neuer Anwender zunächst eine Phase durchlaufen müssen, in der er eine Vielzahl von Beurteilungen erfaßt, ohne jedoch unmittelbar von der Nutzung des Systems zu profitieren.

Für Objekte, die noch nicht bewertet wurden, können keine Vorhersagen der zu erwartenden Relevanz generiert werden. Dies bedeutet ein schlechtes Verhältnis von Aufwand und Nutzen für diejenigen Anwender, welche den Aufwand, neue Objekte zu beurteilen, auf sich nehmen, statt abzuwarten bis andere Anwender diese Aufgabe erledigt haben. Ein Umstand der als *First-Rater-Problem* bezeichnet wird [KMM<sup>+</sup>97].

Die Funktion von Collaborative-Filtering-Systemen, welche auf expliziten Bewertungen basieren, ist maßgeblich davon abhängig, ob es gelingt, dauerhaft Anwender zur Bewertungsabgabe zu motivieren. So lange die Abgabe von Bewertungen und damit die zunehmende Detaillierung des Benutzerprofils zu persönlicheren Vorhersagen führt und dieser Zusammenhang im Bewußtsein der Anwender verankert ist, mag die Motivation der Anwender unproblematisch erscheinen. Sind die Benutzer jedoch erst einmal mit der Qualität der Empfehlungen zufrieden, so könnte es zu einem Verhalten kommen, bei dem die passive Nutzung des Systems überwiegt. Diese Problematik wird mit dem Begriff *Free-Rider-Problem* benannt [KR03].

Werden Collaborative-Filtering-Techniken zur Selektion von Objekten aus sehr großen, dynamisch wachsenden Mengen eingesetzt, so wird jeder einzelne Benutzer nur einen Bruchteil der Objekte bewerten können. Insbesondere dann, wenn zudem noch die Menge der Benutzer verhältnismäßig klein ist, werden einige Objekte ganz ohne Wertung bleiben. Eine in diesem Fall sehr dünn besetzte Bewertungsmatrix hat zwei Konsequenzen, die unter der Bezeichnung *Sparsity-Problem* zusammengefaßt werden. Zum einen wird die Menge der empfehlbaren Objekte eingeschränkt. Zum anderen wird die für die Personalisierung der Vorhersagen essentiell wichtige Auffindung ähnlicher Personen erschwert, denn die Korrelationsanalysen setzen voraus, daß zwischen den Anwendern eine Überlappung in der Menge der bewerteten Objekte vorliegt [BS97, KMM<sup>+</sup>97, SKKR01].

Die Repräsentation der Benutzerprofile als Vektor der Bewertungsmatrix ist nur mit Einschränkungen geeignet, um zuverlässig Übereinstimmungen in den Interessen der Benutzer zu ermitteln. Personen können die gleichen Interessen aufweisen, aber verschiedene,

jedoch thematisch ähnliche Objekte bewertet haben. Algorithmen, wie der vorgestellte Pearson-r-Algorithmus (Abschnitt 5.2.1), können nur dann gleichgesinnte Anwender identifizieren, wenn diese exakt die gleichen Gegenstände mit hoher Übereinstimmung bewertet haben. Beispielsweise könnte ein Benutzer Wetternachrichten bei CNN bewerten und ein anderer bei Yahoo. Die Präferenz beider Benutzer für Wetternachrichten würde nicht zum Ausdruck kommen [BS97].

Collaborative Filtering funktioniert nur dann zufriedenstellend, wenn es gelingt, für eine Empfehlungen abfragende Person eine Menge von Empfehlern mit ähnlichen Präferenzen zu finden. „Exotische“ Benutzer bekommen eventuell keine spezifischen Bewertungsvorhersagen, insbesondere dann, wenn die Zahl der Systemnutzer sehr gering ist [BS97].

Collaborative-Filtering-Systeme lernen die Interessen der Benutzer kontinuierlich besser kennen. Veränderungen in den Benutzerinteressen werden jedoch nur sehr langsam nachvollzogen. Insbesondere abrupte oder temporäre Veränderungen der Interessen sind als problematisch anzusehen [KR03].

Ein Problem, welches mit der expliziten Abgabe von Bewertungen zusammenhängt, ist, daß Benutzer vom Wertebereich der Bewertungsskala höchst unterschiedlich Gebrauch machen können, wenn die Bewertung von Objekten nach subjektiven Kriterien erfolgt. So werden manche Benutzer die maximale Bewertung nur in Ausnahmefällen verwenden, während andere inflationär oft mit dem maximalen Wert votieren werden.

Empfehlungen durch Collaborative-Filtering-Algorithmen basieren auf statistischen Verfahren. Eine Schwäche dieses Ansatzes ist, daß für Anwender der Empfehlungsprozess nicht unbedingt transparent ist. Recommender-Systeme, wie beispielsweise GroupLens, präsentieren sich als „Black-Box“. Anwender können aufgrund der ihnen präsentierten Informationen nicht nachvollziehen, warum ein bestimmter Artikel empfohlen wird und das System von einem anderen abrät [DDH<sup>+</sup>00].

Collaborative-Filtering-Systeme liefern tendenziell bessere Resultate, wenn sie über eine große Benutzerbasis verfügen. Ein Recommender-System sollte also bezüglich der Zahl der Anwender skalierbar gestaltet werden. Der Aufwand der Generierung von Empfehlungen ist abhängig von der Zahl der Objekte in der Datenbasis des Systems und der Anzahl der Nutzer [SKKR01]. Bei speicherbasierten Verfahren hat die Suche nach den „nächsten Nachbarn“ eine Komplexität von  $O(n^2)$ , wenn das System  $n$  Nutzer aufweist.

### 5.2.4. Evaluationsmethoden

Collaborative-Filtering-Algorithmen basieren auf statistischen Verfahren. Demgemäß können statistische Kenngrößen als Metrik zur objektiven Beurteilung der Qualität verschiedener Vorhersagealgorithmen herangezogen werden.

Eine in [SM95] vorgeschlagene Evaluationsmethode basiert auf der Idee, 20 Prozent der Bewertungen der Datenbasis zu löschen, als Target-Set zu speichern und mittels der verbleibenden Bewertungen, dem Source-Set, die gelöschten Werte zu extrapolieren. Betrachtet werden anschließend die Differenzen der berechneten Vorhersagen zu den Werten des Target-Sets. Erwartungswert und Standardabweichung der Differenzen sollten möglichst nahe bei Null liegen. Des weiteren ist es ein Vergleichskriterium, welcher Anteil der Werte des Source-Sets durch den gewählten Vorhersagealgorithmus rekonstruiert werden kann.

Als Vergleichsmaßstab können der triviale Fall, nämlich die Vorhersage des mittleren Elements der Bewertungsskala, unabhängig von den vorhandenen Bewertungen und der sogenannte Base-Case Algorithmus, also die Extrapolation von Werten durch ungewichtete Mittelwertbildung, herangezogen werden. Beide Algorithmen sollten bezüglich der obengenannten Kriterien deutlich übertroffen werden. Die Signifikanz der so gewonnenen Aussagen ist stark vom Umfang der zugrunde liegenden Testdaten, also der Zahl der Benutzer und der von ihnen abgegebenen Wertungen, abhängig. Bei der Evaluation von Collaborative-Filtering-Algorithmen kann neben der Präzision der Vorhersagen auch die Performanz der Berechnungen ein Kriterium sein, denn beim Entwurf von Empfehlungssystemen ist zwischen der Qualität der personalisierten Vorschläge und der Skalierbarkeit des Systems abzuwägen.

### 5.3. Hybride Filtertechnologien

Hybride Systeme versuchen, die Vorteile von Collaborative-Filtering-Techniken zu nutzen und gleichzeitig die Nachteile dieses Paradigmas durch Kombination mit Methoden des Content-based Filterings zu relativieren [BS97, SKB<sup>+</sup>98, CGM<sup>+</sup>99]. Content-based Filtering kommt bei hybriden Systemen vorzugsweise dann zum Einsatz, wenn für ein neues Objekt noch keine Bewertungen vorliegen. Das „First-Rater-Problem“ (siehe Abschnitt 5.2.3) wird somit abgeschwächt.

Ein Beispiel für die Kombination von Content-based und Collaborative Filtering sind die im Grouplens-Prototyp verwendeten „FilterBots“ [SKB<sup>+</sup>98]. Mehrere Agenten geben basierend auf inhaltlichen Analysen für jeden neuen Newsgroup-Beitrag Bewertungen ab. Jeder Agent entspricht dabei einem hypothetischen Benutzer, der alle Beiträge mag, die einem bestimmten formalen Kriterium entsprechen. Dabei wurden Agenten entwickelt, die z. B. die Menge an zitiertem Text oder die Zahl der Rechtschreibfehler als Kriterium verwendeten. Wie in Abbildung 5.7 dargestellt, werden die Bewertungen der Agenten zusammen mit den Einträgen der Benutzer in der Bewertungsmatrix gespeichert. Somit kann die Korrelation eines Benutzerprofils mit der Menge der Bewertungen eines Agenten zur individuellen Bestimmung der Gewichtung des jeweiligen Agenten bei der Berechnung von Relevanzvorhersagen herangezogen werden.

	Benutzer 1	Benutzer 2	...	Agent 1	Agent 2	...
Objekt 1	$V_{b1,1}$	$V_{b2,1}$	...	$V_{a1,1}$	$V_{a2,1}$	...
Objekt 2	$V_{b1,2}$	$V_{b2,2}$	...	$V_{a1,2}$	$V_{a2,2}$	...
...	...	...	...	...	...	...

Abbildung 5.7.: GroupLens FilterBots bewerten Artikel wie menschliche Benutzer.

## 5.4. Item-to-Item-Korrelation

In typischen E-Commerce-Szenarios ist die Zahl der Artikel im Vergleich zur Entwicklung der Menge der Benutzer als relativ statisch anzusehen. Folglich erscheint es bezüglich der Performanz vorteilhaft, den Collaborative-Filtering-Algorithmus dahingehend zu verändern, daß nicht nach ähnlichen wertenden Benutzern, sondern nach ähnlich bewerteten Artikeln gesucht wird. Es wird bei diesem Ansatz davon ausgegangen, daß ein Benutzer vorzugsweise Artikel erwerben wird, die ähnlich zu den Artikeln sind, die er in der Vergangenheit mochte und umgekehrt Artikel meidet, die denen entsprechen, die er nicht mochte [SKKR01].

Bei dieser Herangehensweise wird die Bewertungsmatrix (siehe Abschnitt 5.2) nach Paaren von Artikeln durchsucht, deren Bewertungen korrelieren. So werden basierend auf den Handlungen vorhergehender Benutzer Beziehungen zwischen Artikeln identifiziert, die dann als Grundlage von Empfehlungen dienen, sobald ein Benutzer für mindestens einen Artikel, z. B. durch Aufruf einer Beschreibungsseite oder Plazierung im Warenkorb, sein Interesse bekundet hat. Abbildung 5.8 zeigt, wie der Web-Shop Amazon<sup>6</sup> einem Benutzer, der ein Buch ausgewählt hat, basierend auf dieser Technik weitere Titel vorschlägt.

Beläßt man es bei der Auffindung korrelierender Artikel, so sind zwar keine personalisierten Empfehlungen möglich. Dafür können anonymen Benutzern in Abhängigkeit von den gerade ausgewählten Objekten weitere Alternativen vorgeschlagen werden [SKR99].

Analog zur Vorgehensweise des kollaborativen Filterns können jedoch auch einem am System angemeldeten Besucher individuelle Vorhersagen der zu erwartenden Relevanz von ihm bisher unbewerteter Artikel präsentiert werden. Dabei wird zunächst ermittelt, wie ähnlich die im Profil des aktiven Benutzers vorhandenen Artikel zu dem fraglichen

---

<sup>6</sup>[www.amazon.de](http://www.amazon.de)

Kunden, die Design Patterns. Elements of Reusable Object-Oriented Software. gekauft haben, bestellten auch:

The screenshot shows a grid of recommended books. The top row features three books: 'Refactoring. Improving the Design of Existing Code' (EUR 37,20), 'A System of Patterns. Pattern-Oriented Software Architecture' (EUR 57,17), and 'UML Distilled. A Brief Guide to the Standard Object Modeling Language' (EUR 36,95). The bottom row features three more books: 'Effective C++. 50 Specific Ways to Improve your Programs and Designs' (EUR 36,42), 'Applying UML and Patterns' (EUR 47,41), and 'Modern C++ Design: Generic Programming and Design Patterns Applied (C++ in Depth)' (EUR 41,02). Each book listing includes a small cover image, the title, author, and price, along with a 'Wagen' (shopping cart) icon. To the right, a blue sidebar titled 'IHR EINKAUFSWAGEN' (Your Shopping Cart) shows the selected item 'Design Patterns. Elements of Reusable Object-Oriented Software' for EUR 39,40, with a 'Zur Kasse gehen' (Go to checkout) button and a note to log in for 1-click ordering.

Abbildung 5.8.: Ausgabe von Empfehlungen basierend auf dem Inhalt des Warenkorbes im Web-Shop Amazon

Objekt sind. Anschließend werden die Bewertungen des Benutzerprofils entsprechend gewichtet aufsummiert [SKKR01].

## 5.5. Social Data Mining

Durch *Social Data Mining* werden Datenbestände, beispielsweise Newsgroup-Beiträge, Webserver-Zugriffsprotokolle oder Web-Dokumente mit den darin enthaltenen Hyperlinks, mit dem Ziel ausgewertet, die Informationen zu Navigationshinweisen zu aggregieren [TH01]. Kennzeichnend für Social Data Mining ist, daß die ausgewerteten Informationen nicht auf Handlungen zurückzuführen sind, welche in der Intention ausgeführt wurden, vermittelt über ein Empfehlungssystem, anderen Nutzern Navigationshinweise zu geben [TH01]. Die erfaßten Daten sind als Randprodukte anderer Tätigkeiten anzusehen. Möglicherweise wurden auch die Daten zu einem ganz anderen Zweck als der Generierung von Navigationshinweise erhoben. Insofern stimmen Social-Data-Mining-Anwendungen mit implizite Bewertungen einbeziehenden Collaborative-Filtering-Anwendungen überein. Als unterschiedlich ist jedoch die Rollenverteilung anzusehen. Der Anwender einer Collaborative-Filtering-Applikation befindet sich symmetrisch in der Rolle des Anbie-

ters und Konsumenten von Empfehlungen. Demgegenüber kann bei auf Social Data Mining basierenden Systemen von einer asymmetrischen Rollenverteilung ausgegangen werden [TH01, THA<sup>+</sup>97].

Der Vorteil dieses Verfahrens liegt darin begründet, daß die Anwender keinen Aufwand mit der Bewertung von Objekten haben und meistens auf sehr umfangreiche Datenbestände zurückgegriffen werden kann. Nachteiligerweise werden jedoch keine personalisierten Empfehlungen ausgegeben.

Exemplarisch für Social-Data-Mining-Techniken sind das System PHOAKS [THA<sup>+</sup>97] und der Google<sup>7</sup> PageRank [PBMW98] zu nennen.

Der PHOAKS-Prototyp wertet regelmäßig Newsgroups des Usenet bezüglich der Nennung von URLs aus. Dabei wird die Erwähnung einer URL in einem Newsgroup-Artikel als eine im Bezugsrahmen der jeweiligen Newsgroup gültige Empfehlung der referenzierten Web-Ressource interpretiert. Auf diesem Wege werden Empfehlungslisten von Web-Dokumenten erstellt (siehe Abbildung 5.9), wobei die Reihenfolge der Einträge durch die Häufigkeit, mit der die jeweilige URL in der Newsgroup erwähnt wurde, bestimmt wird.

Zur Bewertung der Relevanz von Dokumenten bezüglich einer Suchanfrage benutzt Google die Link-Struktur des WWW, welche durch den Google-Robot partiell erfaßt und auf Datenstrukturen abgebildet wird. Der *PageRank* genannte Algorithmus basiert auf der Annahme, daß ein vielfach verlinktes Dokument von größerer Relevanz ist als ein Dokument auf das nur wenige Seiten verweisen. Der PageRank-Algorithmus gewichtet nun zusätzlich die einzelnen Links, so daß ein Link von einer Seite, die ihrerseits einen hohen PageRank-Wert besitzt, mehr als der Link von einer unbedeutenden Seite zählt.

---

<sup>7</sup>[www.google.com](http://www.google.com)

People Helping One Another Know Stuff

**Phoaks**  
Together, we know it all.

bootleg CDs

Navigate Up: [PHOAKS Entry Page](#) : [Newsgroup Areas](#): [rec](#) . [music](#) . [dylan](#)

### Frequently Mentioned Resources

Resource Title	Distinct Posters	Click on Bars for Message Context(s) *
01) <a href="#">Bob Links</a>	<u>9</u>	■■■■■■■■
02) <a href="#">Image at 128.39.161.105</a>	<u>7</u>	■■■■■■
03) <a href="#">The Book of Bob</a>	<u>5</u>	■■■■
04) <a href="#">Tour Update</a>	<u>5</u>	■■■■
05) <a href="#">Telegraph Home Page</a>	<u>4</u>	■■■■
06) <a href="#">Mailing List WWW Gateway</a>	<u>3</u>	■■■
07) <a href="#">Bob Dylan Chords</a>	<u>3</u>	■■■
08) <a href="#">Expecting Rain</a>	<u>3</u>	■■■
09) <a href="#">Jan Heiberg's WWW Page</a>	<u>3</u>	■■■
10) <a href="#">"_" same title, different location</a>	<u>3</u>	■■■
11) <a href="#">Bringing It All Back Home Page</a>	<u>3</u>	■■■
12) <a href="#">CDnow : welcome</a>	<u>2</u>	■■

Abbildung 5.9.: Durch das PHOAKS-System ermittelte Liste empfohlener Web-Dokumente zu einer Newsgroup (Quelle: [THA<sup>+</sup>97])

## 6. Verwandte Arbeiten

In diesem Kapitel werden einige Projekte präsentiert, welche die Konzeption des Empfehlungs- und Annotationssystems beeinflusst haben. Eine Sichtweise auf das System ist das in Kapitel 4 vorgestellte Konzept „Social Navigation“. Navigation durch Orientierung an Empfehlungen, welche auf der Aggregation von Bewertungen anderer Nutzer basieren oder durch Beachtung von Annotationen, ist der Vorgehensweise des sozialen Navigierens zuzuordnen. Abstrahiert man von den kollaborativen Elementen des Empfehlungsprozesses, dann läßt sich das System ganz allgemein als ein Anwendung betrachten, die Benutzern bei der Selektion von Informationen assistiert. Aus diesem Blickwinkel betrachtet bestehen Bezüge zu allen in Kapitel 5 vorgestellten Paradigmen der Entwicklung von Empfehlungssystemen bzw. Information-Filtering-Anwendungen. Die eingesetzten Collaborative-Filtering-Techniken stehen in der Schnittmenge der Konzepte indirekter Social Navigation und Empfehlungssysteme.

Auch das Projekt WebPlaces (Abschnitt 6.1) hat es zum Ziel, das Web um die Möglichkeit des sozialen Navigierens zu erweitern. Das Konzept sieht vornehmlich die Unterstützung direkter Social Navigation basierend auf synchroner Interaktion vor. Es steht damit im Kontrast zu dem CoInternet-System. Konkrete Bezüge sind im Bereich der Implementierung zu sehen, wo ein ähnlicher Ansatz gewählt wurde, um das System in die Web-Architektur zu integrieren.

Das System Grouplens – es wird in Abschnitt 6.2 beschrieben – kann als Muster eines auf Collaborative-Filtering-Techniken basierenden Empfehlungssystems betrachtet werden. Gegenstand der Empfehlungen des Grouplens-Systems sind über einen Internet-Dienst publizierte Informationen. Auch in diesem Punkt bestehen Parallelen zu dem im Zentrum dieser Arbeit stehenden System.

Der in Abschnitt 6.3 vorgestellte Dienst AntWorld bietet eine Plattform zur asynchronen Kollaboration bei der Recherche im Web. Suche im Web ist ebenfalls ein Szenario, welches bei der Konzeption des CoInternet-Systems berücksichtigt wurde.

Die Suchmaschine Alexa<sup>1</sup> gestattet es Benutzern, Web-Sites zu bewerten und „Rezensionen“ zu verfassen. Somit fördert dieser Dienst die asynchrone Kollaboration der Nutzer. Über die in den Browser integrierbare Alexa-Toolbar können viele Funktionen der Such-

---

<sup>1</sup>[www.alex.com](http://www.alex.com)

---

maschine direkt aufgerufen und zusätzliche Informationen zum aktuellen Dokument, beispielsweise eine Liste verwandter Seiten, eingesehen werden. Mit jeder geladenen Seite werden durch die Toolbar Informationen zur aktuellen URL vom Alexa-Server abgefragt. Diese Anfragen wertet Alexa aus, um ein Traffic Ranking zu erstellen und verwandte Dokumente zu identifizieren. Inwiefern die explizit von den Benutzern abgegebenen Bewertungen und die Traffic Analysen einen Einfluß auf des Ranking der Suchergebnisse haben, bleibt offen.

Als ein Beispiel indirekter Social Navigation wurde das Empfehlungssystem SiteSeer in Abschnitt 4.5 präsentiert. SiteSeer generiert Empfehlungen basierend auf der Auswertung der Bookmark-Verzeichnisse der Nutzer. Bookmarks werden dabei als impliziter Ausdruck von besonderem Interesse an der hinterlegten Seite angesehen. Bei der Konzeption der CoInternet-Anwendung wurde diese Annahme geteilt und Bookmarks als implizite Voten bzw. zur Initialisierung von Benutzerprofilen verwendet.

Das System Knowledge Pump [GAD97] schlägt Personen durch Aggregation von Bewertungen anderer Nutzer für sie lesenswerte Web-Seiten vor. Hierbei kommt eine Technik zum Einsatz, die von den Urhebern des Systems als *Community-centred Collaborative Filtering* bezeichnet wird. Objekte, die von den Benutzern des Knowledge-Pump-Systems bewertet werden, müssen auch gleichzeitig in ein vom System vorgegebenes Klassifikationsschema eingeordnet werden. Dadurch wird die Menge der Objekte partitioniert und die Suche nach ähnlichen Personen durch den Collaborative-Filtering-Algorithmus auf Personen beschränkt, die weitere Objekte der Domäne bewertet haben, in der sich das Objekt befindet, für welches eine Relevanzvorhersage bestimmt werden soll. Automatische Collaborative-Filtering-Systeme erschweren es neuen Benutzern, in das System einzutreten, denn die zur Berechnung individueller Empfehlungen notwendige Suche nach ähnlichen Benutzern setzt voraus, daß der jeweilige Benutzer selbst einige Bewertungen abgegeben hat (siehe Abschnitt 5.2.3). Knowledge Pump gestattet es den Anwendern, eine Liste von Ratgebern anzugeben, denen sie vertrauen und deren Meinung ihnen wichtig ist. Diese Auswahl wird durch den Collaborative-Filtering-Algorithmus insbesondere in der Phase einbezogen, in der die Menge der durch den Empfehlungen abfragenden Benutzer bewerteten Objekte noch nicht ausreichend ist, um mittels der statistischen Verfahren automatisch Ratgeber zu bestimmen. Bei der Konzeption des CoInternet-Prototyps wurde der Ansatz des Community-centred Collaborative Filtering aufgegriffen und weiterentwickelt (siehe Abschnitt 8.2.1).

Bei der Entwicklung des CoInternet-Servers galt es, die Funktionseinheit zur Vorhersage von Bewertungen so modular wie möglich zu konstruieren, um einfach unterschiedliche Implementierungen des Vorhersagealgorithmus erproben zu können. Diese Anforderung teilt das in dieser Arbeit entwickelte System mit dem Java-Framework SWAMI [FHH<sup>+</sup>00]. Dieses Framework dient primär der Entwicklung und Evaluation von Collaborative-Filtering-Algorithmen. Bestandteile des Frameworks sind eine *Predicti-*

*on Engine*, die eine Programmierschnittstelle zur Implementierung von Vorhersagealgorithmen bietet, sowie Module zur Evaluation und Visualisierung der Leistung der untersuchten Algorithmen. Eine Implementierung des Pearson-r-Algorithmus (siehe Abschnitt 5.2.1), welche Bestandteil des Frameworks ist, konnte in modifizierter Form bei der Entwicklung des Prototyps eingesetzt werden.

Eine Funktion des CoInternet-Systems ist die Annotation von Web-Seiten mit beliebigen Freitextkommentaren, welche auch Links zu weiteren Ressourcen im Web beinhalten können. Somit ist diese Anwendung in einer Reihe mit früheren Versuchen, die Annotation von Web-Dokumenten zu ermöglichen, wie den Projekten CritLink (Abschnitt 4.5) und Thirdvoice<sup>2</sup>, zu sehen.

### 6.1. WebPlaces

Das Web ist ein Raum, in dem sich viele Personen gleichzeitig aufhalten. Aber es ist kein Platz<sup>3</sup>, auf dem sich die Leute gegenseitig sehen und miteinander interagieren können. Ziel des Systems WebPlaces [MB99, Fis00, MBF03] ist es, aus dem Web einen Platz zu machen, auf dem die Gegenwart und Aktivität anderer Nutzer sichtbar und synchrone Interaktion möglich wird. WebPlaces steht mit dieser Betonung der synchronen Interaktion im Kontrast zu den nachstehend genannten Systemen GroupLens (Abschnitt 6.2) und AntWorld (Abschnitt 6.3), welche ebenso wie der in dieser Arbeit entwickelte Prototyp (Kapitel 8) die asynchrone Interaktion bei der Navigation in einem Informationsraum ermöglichen sollen.

Awareness (siehe Abschnitt 2.1.6) zwischen Benutzern des Webs ist nur dann sinnvoll, wenn die Menge der wahrnehmbaren Personen überschaubar bleibt. Dabei ist die Frage zu klären, wie die Benutzer des Webs dynamisch Communities zugeordnet werden können und wie die Communities sinnvoll visualisiert werden können.

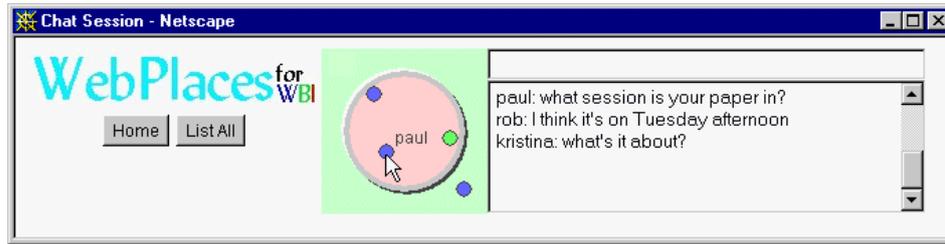
Im WebPlaces-Prototyp sind alle eingeloggtten Benutzer Teil derselben Community. Eine einfache Vorgehensweise wäre es, aus allen Besuchern einer Web-Site oder aus allen Personen die gerade die selbe Seite anschauen, dynamisch eine Community zu bilden. Alternativ könnten Benutzer, die sich in der gleichen Region des Web-Graphen befinden, gerade Dokumente mit ähnlichem Inhalt geladen haben oder überlappende Pfade im Web gegangen sind, einer Community zugeordnet werden (siehe auch Abbildung 4.2 auf Seite 60).

Die Gegenwart und Aktivitäten anderer Community-Mitglieder werden durch das WebPlaces-System in einem Fenster neben dem Browser visualisiert. Dabei wurde die

---

<sup>2</sup>Der unter der URL [www.thirdvoice.com](http://www.thirdvoice.com) erreichbare Annotationsdienst wurde mittlerweile eingestellt.

<sup>3</sup>Der begriffliche Unterschied zwischen Raum und Platz wurde in Abschnitt 4.7 erörtert.



**Abbildung 6.1.:** Das WebPlaces-Fenster. In der Mitte befindet sich die Visualisierung des *Social Proxy*, die Chat-Funktionalität wird durch die Elemente auf der rechten Seite realisiert. (Quelle: [MB99])

Idee des *Social Proxy* von Erickson et al. aufgegriffen [EK03]. Die Community wird in der gewählten Visualisierung durch einen Kreis repräsentiert, einzelne Benutzer werden durch farbige Punkte dargestellt (siehe Abbildung 6.1). Durch die Bewegung eines Punktes in Richtung des Zentrums des Kreises wird eine Interaktion mit der Gruppe, beispielsweise eine Äußerung im Chat, ausgedrückt. Interaktionen zwischen zwei Personen werden durch die Annäherung der zugeordneten Punkte dargestellt. Rotiert ein Punkt um das Zentrum, so symbolisiert dieses individuelle Aktionen, wie beispielsweise das Browsen. Ist eine Person eine längere Zeit nicht aktiv, so wandert der zugeordnete Punkt an den Rand des Kreises. Durch diese Art der Visualisierung sind sowohl der Zustand als auch die Aktivität der Community-Mitglieder sichtbar. Bereits durch ein beiläufigen Blick auf die Grafik läßt sich die aktuelle Situation in der Community erfassen.

WebPlaces unterstützt mit der Chat-Funktion die synchrone Kommunikation innerhalb der Community ebenso wie die von den anderen Gruppenmitgliedern isolierte Eins-zu-Eins-Konversation. Die Wahrnehmung von anderen Nutzern verbunden mit der Möglichkeit der Kontaktaufnahme macht WebPlaces zu einem System, welches das Web um Funktionen direkter Social Navigation erweitert.

Auch Formen indirekter Social Navigation werden durch WebPlaces unterstützt. Bereits die aus der Visualisierung ersichtlichen Informationen über die Aktivitäten auf einem Platz im Web können das Navigationsverhalten beeinflussen. Des weiteren bietet das System die Funktion, anderen Benutzern bei ihren Bewegungen im Web zu folgen, ohne mit diesen in Kontakt treten zu müssen.

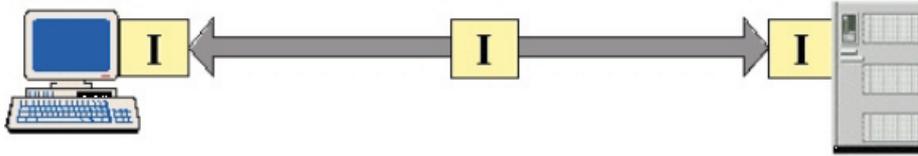
Benutzer werden im WebPlaces-System durch Pseudonyme repräsentiert, deren Gültigkeit sich auf dieses System beschränkt. Intention des WebPlaces-Systems ist es, Awareness und soziale Interaktion im Web allgegenwärtig zu machen und somit den Aufbau von Beziehungen zwischen Anwendern zu fördern. In diesem Sinne würde WebPlaces entscheidend von der Integration einer Identitätsinfrastruktur profitieren. Durch ein ein-

heitliches, über Grenzen von Anwendungen und Web-Sites hinweg gültiges Identitätsmodell und global eindeutige Identifikatoren wäre es zum einen möglich, Benutzer in verschiedenen Kontexten wiederzuerkennen und zum anderen könnte nahtlos zwischen verschiedenen Formen der Internet-Kommunikation gewechselt werden.

Durch WebPlaces soll das Web, über den Rahmen einzelner Web-Sites hinausgehend und prinzipiell für alle Benutzer zugänglich, um Funktionen zur Schaffung von Awareness und zur Unterstützung sozialer Interaktion erweitert werden. Technische Realisierungsformen der Integration des Systems in die Web-Architektur, die Modifikationen der Web-Server oder -Browser bedingen, erscheinen bei dieser Vorgabe als wenig zielführend. Der Lösungsansatz des WebPlaces-Projektes ist die Technik der *Intermediaries*.

*Intermediaries* sind Computerprogramme, die zwischen dem Produzenten von Informationen und dem Verbraucher im Informationsstrom plziert sind und dabei einen Mehrwert generieren [BM99, MB00]. Intermediaries können an verschiedene Stellen in den Informationsfluß integriert werden. Alternativen sind der Betrieb auf dem Client-Rechner unter Kontrolle des Anwenders, die Platzierung auf Seiten des Servers, als Teil des Systems eines Anbieters von Informationen und der Betrieb von Intermediaries durch zwischengeschaltete Dienstleister, welche diese einer Menge von Benutzern zugänglich machen (siehe Abbildung 6.2). Es lassen sich folgende Funktionen von Intermediaries unterscheiden:

- *Anpassen*  
Intermediaries können basierend auf Informationen über den Benutzer und seine Umgebung Informationsströme anpassen bzw. personalisieren. Beispielhaft ist die Anpassung der Reihenfolge der Punkte eines Menüs gemäß den Präferenzen des Benutzers.
- *Filtern*  
Entfernen von Informationen aus dem Strom.
- *Annotieren*  
Hinzufügen zusätzlicher Informationen.
- *Aggregieren*  
Kombinieren von Informationsströmen aus verschiedenen Quellen.
- *Transkodieren*  
Umwandlung des Informationsstroms in ein anderes Format. Ein Beispiel wäre die Modifikation der Darstellung von Dokumenten gemäß den Eigenschaften des verwendeten Endgerätes.
- *Zwischenspeichern*  
Speichern übertragener Informationen für spätere Verwendung.



**Abbildung 6.2.:** Intermediaries können an beliebiger Stelle im Informationsfluß zwischen dem Produzenten und Konsumenten von Informationen plaziert werden. (Quelle: [MB00])

Die Zuordnung von Intermediaries zu diesen Funktionklassen muß nicht eindeutig sein. Ferner sind Systeme möglich, bei denen Intermediaries zur Realisierung ihrer Funktionalität weitere Dienste einbeziehen.

Das WBI (Web Based Intermediary) Development Kit<sup>4</sup> vom IBM Almaden Research Centre ist ein programmierbarer Proxy-Server, der eine Programmierschnittstelle für Intermediaries bietet. Kombinationen der drei grundlegenden Bausteine *Monitors*, *Editors* und *Generators* können zur Realisierung beliebiger Intermediaries über eine Plugin-Architektur in den den Proxy-Server durchlaufenden Datenstrom integriert werden.

Monitore überwachen den Datenstrom, eine Anwendung wäre der Aufbau von Benutzermodellen basierend auf der Auswertung des Navigationsverhaltens im Web. Mittels Editoren können Intermediaries Dokumente modifizieren, etwa um zusätzliche Informationen hinzuzufügen. Ein Generator, der dynamisch Daten generiert, kann als der Sonderfall eines Editors angesehen werden.

WebPlaces wurde basierend auf dem WBI implementiert. In diesem Punkt bestehen Parallelen zwischen WebPlaces und dem Empfehlungs- und Annotationssystem, welches im Zentrum dieser Arbeit steht und mittelbar über das Framework Scone ebenfalls auf dem WBI basiert (siehe Abschnitt 2.4). Beide Systeme haben es zum Ziel, Social Navigation in einer universellen Art und Weise im Web zu unterstützen, ohne dabei Veränderungen an der Architektur des Webs vornehmen zu müssen. Auch wenn sich die konkreten Formen von Social Navigation zwischen beiden Systemen grundlegend unterscheiden, die Realisierung als Intermediary ist beiden Anwendungen gemein.

Konkret nutzt WebPlaces den WBI-Proxy-Server um mittels eines Editors JavaScript Code in die Dokumente einzubetten, der bewirkt, daß das WebPlaces-Applet in einem separaten Fenster angezeigt wird. Ein Monitor überwacht die Bewegungen der Benutzer im Web, um diese dynamisch Communities zuzuordnen. Die Information der Zugehörigkeit zu einer Community wird über eingebetteten JavaScript Code an das WebPlaces Applet kommuniziert. Zusätzlich wurde ein Generator realisiert, der beispielsweise die Login-Seite erstellt.

<sup>4</sup>[www.almaden.ibm.com/cs/wbi](http://www.almaden.ibm.com/cs/wbi)

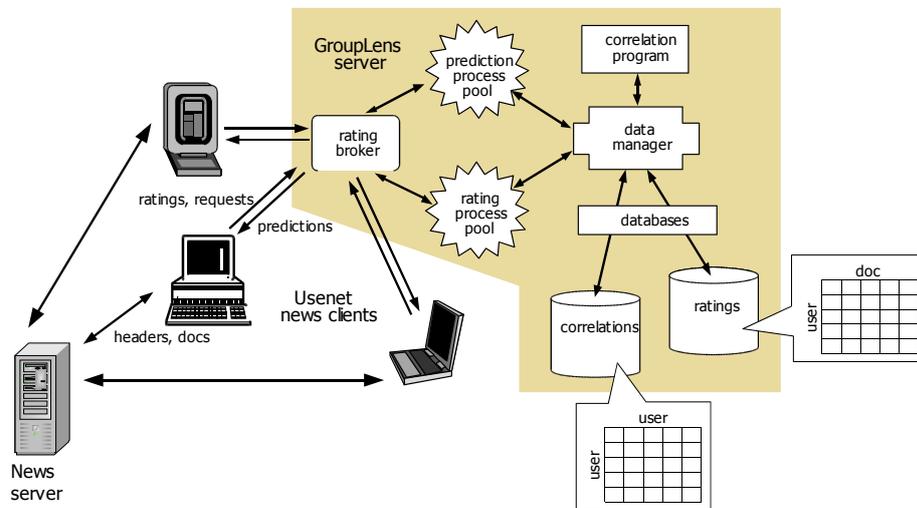


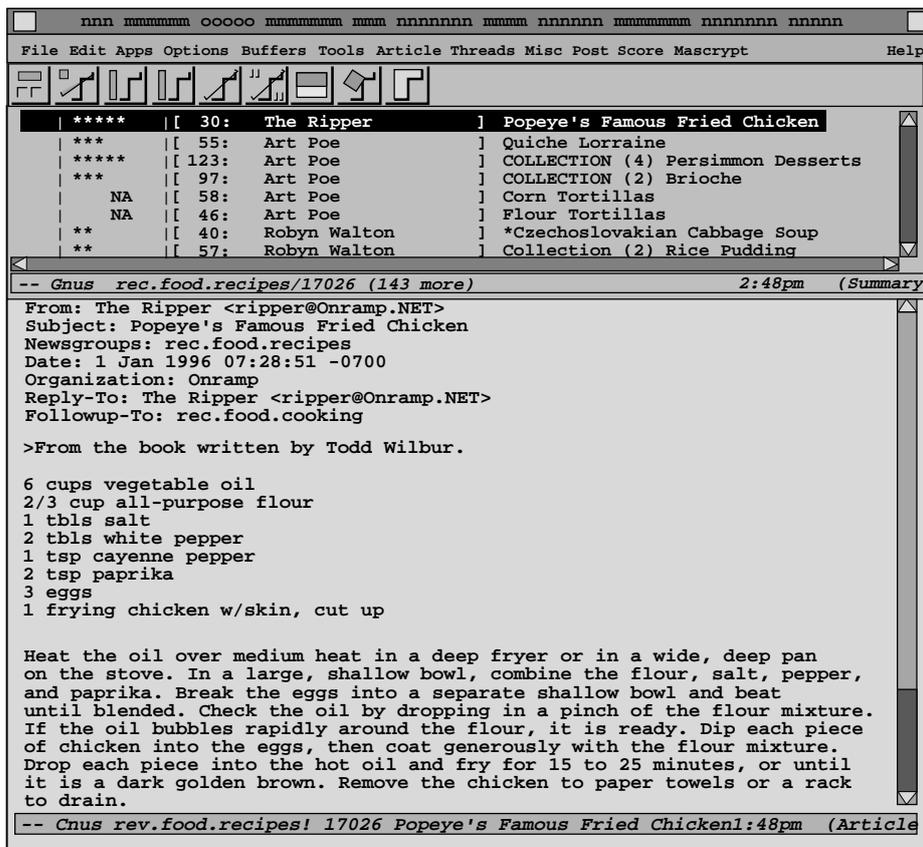
Abbildung 6.3.: Architektur des Grouplens-Systems (in Anlehnung an [KMM<sup>+</sup>97])

## 6.2. Grouplens

Grouplens [RIS<sup>+</sup>94, KMM<sup>+</sup>97, MRK97, SKB<sup>+</sup>98] ist, wie bereits im Rahmen der Ausführungen über Collaborative Filtering (Abschnitt 5.2) erwähnt wurde, ein Empfehlungssystem für Usenet-News. Zur Vorhersage der individuell zu erwartenden Relevanz einer Nachricht nützt Grouplens den Pearson-r-Algorithmus (siehe Abschnitt 5.2.1). Die eindeutigen Identifizierung der Benutzer erfolgt über Pseudonyme. Dadurch können Bewertungen mit Benutzern assoziiert und individuelle Empfehlungen gegeben werden, ohne daß die Benutzer gegenüber dem System und anderen Nutzern ihre reale Identität offenbaren müssen. Gegenstand der nachfolgenden Darstellung ist es, die Grouplens-Architektur exemplarisch für einen Realisierungsweg bei der Erweiterung eines Internet-Dienstes um ein auf Collaborative-Filtering-Techniken basierendes Empfehlungssystem vorzustellen. Überdies sollen die Charakteristika des Usenets denen des World Wide Webs gegenübergestellt werden, um daraus Implikationen für den Entwurf eines Empfehlungssystems für Web-Seiten abzuleiten.

Grouplens basiert auf einer Client-Server-Architektur (siehe Abbildung 6.3).

Als Clients kommen verschiedene modifizierte News-Reader zum Einsatz. Diese kommunizieren per NNTP-Protokoll [KL86] mit einem News-Server, um Nachrichten zu laden. Sie weisen eine Anbindung an den Grouplens-Server auf, über die einerseits Bewertungen von Nachrichten zum Server gesandt und andererseits Empfehlungen abgefragt werden.



**Abbildung 6.4.:** Benutzungsoberfläche eines zur Darstellung von Bewertungsvorhersagen angepaßten News-Readers. Die Anzahl der Sterne in der linken Spalte symbolisiert das prognostizierte Interesse an der jeweiligen Meldung. (Quelle: [KMM<sup>+</sup>97])

Die Benutzungsoberfläche der Clients wurde dahingehend erweitert, daß eine einfache Erfassung von Bewertungen möglich wird und die Auflistung der Nachrichten um die Darstellung von Relevanzvorhersagen ergänzt wird (siehe Abbildung 6.4).

Die individuellen Vorhersage der zu erwartenden Relevanz einer Bewertung durch einen Collaborative-Filtering-Algorithmus basiert auf einer Bewertungsmatrix, welche die von den Benutzern zu den Artikeln abgegebenen Bewertungen enthält. Die Datenbank *ratings* des Grouplens-Servers bietet die persistente Speicherung und den effizienten Zugriff auf die Daten der Bewertungsmatrix. Die Ermittlung der Korrelationen zwischen den vergangenen Bewertungen der Benutzer ist ein aufwendiger Vorgang (siehe Formel 5.2 auf Seite 88). Der Zugriff auf vorberechnete Korrelationswerte bei der Beantwortung von

Anfragen nach Relevanzvorhersagen verspricht folglich einen Performanzgewinn. Zudem sind die Korrelationen als ein Ausdruck historischer Übereinstimmung keinen sprunghaften Veränderungen unterworfen, so daß die Vorhersagegenauigkeit nur wenig abnimmt, wenn nicht mit jeder Anfrage aktuelle Korrelationen bestimmt werden. Der Datenbank *ratings* wurde eine zweite Datenbank *correlations* zur Seite gestellt, in welcher die Korrelationen des Bewertungsverhaltens der Nutzer abgelegt sind. Das *correlation program* liest aus der Datenbank *ratings* und aktualisiert die Einträge von *correlations*. Dabei wird die Ausführung dieses Prozesses so terminiert, daß die Korrelation jedes Paares von Benutzern ungefähr alle 24 Stunden neu berechnet wird.

Die Kommunikation mit den Clients wird über den *rating broker* abgewickelt. Ankommende Anfragen nach Relevanzvorhersagen werden an einen freien Prozeß des *prediction process pool* weitergeleitet. Ein *prediction process* liest aus der Datenbank *ratings* die für den angefragten Artikel abgegebenen Bewertungen und summiert dieses gewichtet mit den in der Datenbank *correlations* abgelegten Korrelationen. Abgegebene Bewertungen werden durch einen freien *rating process* in die Datenbank geschrieben, dabei werden gegebenenfalls neue Artikel erfaßt.

Die Kommunikation zwischen den News-Readern und dem Grouplens-Server wird über ein im Rahmen des Projektes entwickeltes offenes Protokoll abgewickelt. Sofern das spezialisierte Übertragungsprotokoll eingehalten wird, sind Client- und Server-Komponenten gegen alternative Implementierungen austauschbar. Das Protokoll beinhaltet folgende Kommandos:

- *Register*  
Registriert das Pseudonym eines neuen Benutzers.
- *Login*  
Vor dem Zugriff auf Grouplens ist ein Login erforderlich. Der Server antwortet mit einem *session token*, der zur Authorisierung des weiteren Zugriffs auf das System verwendet wird.
- *Logout*  
Abmeldung vom Grouplens-Server.
- *GetPredictions*  
Abfrage von Relevanzvorhersagen für den durch das *session token* identifizierten aktiven Benutzer und eine Liste von *message-ids*, welche Nachrichten aus der News-group benennen, deren Namen ebenfalls als Parameter zu übertragen ist.
- *PutRatings*  
Sendet eine Liste von Tupeln aus *message-id* und *rating* an den Server. Weitere Parameter sind das *session token* und der Name der Newsgroup.

Ein interessanter Aspekt der GroupLens-Architektur ist es, daß es vorgesehen ist, zur Verbesserung der Skalierbarkeit mehrere Rating-Server zu betreiben, die jeweils einen Teil der Benutzer bedienen. Um auf den Servern jeweils identische Datenbestände vorzuhalten, wurde eine Replikation der Bewertungen über die Verteilungsmechanismen des Usenets vorgesehen.

Das Usenet ist durch die hierarchische Benennung der Newsgroups inhaltlich strukturiert. Die einzelnen Newsgroups bilden einen sinnvollen Gültigkeitsrahmen für die Modellierung der Benutzerinteressen durch die abgegebenen Bewertungen. Durch Partitionierung der Menge der Bewertungen entsprechend der Zuordnung der bewerteten Meldungen zu verschiedenen Newsgroups kann verhindert werden, daß die Übereinstimmung zwischen Benutzern bezüglich der Nachrichten einer Newsgroup Einfluß auf die Vorhersage von Bewertungen für Nachrichten einer thematisch andersartigen Newsgroup hat. Newsgroups sind ein sehr homogenes Medium; die einzige Form von Dokumenten sind Nachrichten. Somit ist es möglich, in einem Bewertungssystem Kriterien vorzugeben, welche sich im Kontext einer Newsgroups auf alle Nachrichten anwenden lassen.

Die Architektur des Webs sieht keine Einordnung von Inhalten in ein globales Klassifizierungssystem<sup>5</sup> vor. Eine Strukturierung der Inhalte findet auf Ebene der Web-Sites statt. Dabei ist zu bedenken, daß einerseits die verschiedenen Dokumente einer Site nicht notwendigerweise einen inhaltlichen Bezug aufweisen müssen und andererseits Dokumente verschiedener Sites in einem inhaltlichen Kontext stehen können. Durch die Strukturierung von Web-Sites werden den einzelnen Dokumenten implizit Typen zugewiesen. So ist, wie in Abbildung 6.5 gezeigt wird, zwischen der Homepage, Inhaltsseiten und eventuell zwischengeschalteten Navigationsseiten zu differenzieren.

Beim Entwurf eines Bewertungssystems für Web-Seiten ist die Herausforderung zu überwinden, ein sinnvolles Verfahren zur Partitionierung der Menge der Bewertungen zu finden. Bewertet eine Person beispielsweise Seite mit Kochrezepten und fragt anschließend Bewertungsvorhersagen von Dokumenten zum Thema Programmiersprachen ab, dann sollte vermieden werden, daß der Meinung von Benutzern, die übereinstimmende Vorlieben bezüglich der Wahl von Kochrezepten aufweisen, ein besonderes Gewicht zugewiesen wird. Ein weiteres Problem ist die Angabe von Bewertungskriterien. Hier ist die große inhaltliche Bandbreite der Informationen im Web und das Vorhandensein verschiedener Dokumententypen zu berücksichtigen. Sollten für die Bewertung der Seiten eines Online-Shops die gleichen Kriterien Gültigkeit wie für eine News-Site besitzen?

Die Nützlichkeit eines Empfehlungssystems läßt sich am Verhältnis des Wertes korrekter Relevanzvorhersagen zu den Kosten falscher Vorhersagen ermessen. Vorhersagen sind als korrekt anzusehen, wenn übereinstimmend mit der Beurteilung des Benutzers Objekte

---

<sup>5</sup>Web-Kataloge wie Yahoo könnten jedoch als ein Versuch der inhaltlichen Klassifizierung von Web-Dokumenten verstanden werden.

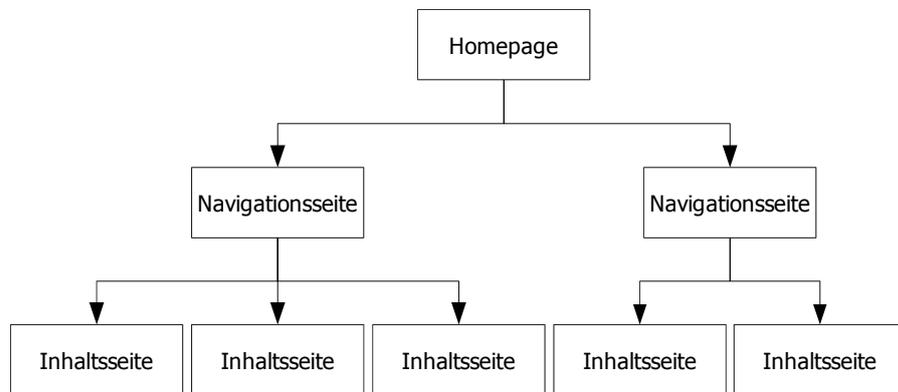


Abbildung 6.5.: Seitentypen einer hierarchisch strukturierten Web-Site

empfohlen werden und die Objekte, von denen abgeraten wird, auch vom Benutzer negativ beurteilt werden würden. Empfiehlt das System Objekte, die dem Benutzer nicht gefallen und rät von Objekten ab, die eigentlich relevant wären, so ist von falschen Empfehlungen zu sprechen. Die Urheber von Grouplens sehen in ihrem System einen hohen Nutzen für die Anwender, da nur ein geringer Anteil der Nachrichten einer Newsgroup relevant sei. Folglich sei es vom hohen Wert, auf die wenigen interessanten Nachrichten hingewiesen zu werden. Gleichzeitig sei es zu verschmerzen, wenn eine interessante Nachricht versäumt wird, oder eine nicht relevante Nachricht gelesen wird, denn dieses nimmt nur sehr wenig Zeit in Anspruch. Bezogen auf das Web kann ebenfalls davon ausgegangen werden, daß nur ein geringer Anteil der zu einem Thema publizierten Dokumente von individueller Relevanz für einen Anwender ist und somit korrekte Empfehlungen einen hohen Wert besitzen. Die Kosten falscher Vorhersagen müssen jedoch differenzierter betrachtet werden. Während das Usenet in erster Linie ein Kommunikationsmedium ist, ist das Web vorrangig ein Informationsmedium. Für jemanden, der im Web ernsthaft recherchiert, etwa nach wissenschaftlichen Veröffentlichungen, kann es durchaus Konsequenzen haben, wenn das System von einem eigentlich relevanten Dokument abrät. Bezogen auf die Seiten eines Web-Shops könnten Empfehlungen auch Einfluß auf das Kaufverhalten der Anwender haben. Eine falsche Empfehlung wäre hier mit direkten monetären Kosten verbunden.

### 6.3. AntWorld

Recherchen im Web erfordern in vielerlei Hinsicht menschliche Intelligenz und Kreativität. So müssen Anfragen formuliert und verfeinert, geeignete Suchmaschinen und

Kataloge ausgewählt und gefundene Dokumente hinsichtlich der Rechercheaufgabe bewertet werden. Durch das AntWorld-System<sup>6</sup> [KBM<sup>+</sup>99, KBM<sup>+</sup>00, MNS00] sollen Team-Mitglieder in Organisationen zu kollaborativem Verhalten bei Recherchen im Web befähigt werden. Die Benutzer von AntWorld kollaborieren asynchron über hinterlassene Bewertungen, die dann zu Empfehlungen für nachfolgende Personen mit ähnlichen Rechercheaufgaben aggregiert werden. Der Fokus des Systems liegt weniger bei der Unterstützung ungerichteten Browsens im Web, sondern vielmehr bei der Hilfestellung bei zielgerichteten Recherchen. Demgemäß läßt sich das AntWorld als kollaboratives Information-Retrieval-System charakterisieren.

Die Ziele eines Benutzers bei der Recherche werden durch einen sogenannten *Quest* repräsentiert. Dieser ist zu Beginn einer neuen Recherche durch die Formulierung einer kurzen Beschreibung – diese ist mit einer Suchmaschinenanfrage vergleichbar – und einer optionalen langen Beschreibung der Zielvorgaben zu initialisieren. Eine durch das System unterstützte Vorgehensweise ist es, Quests an Suchmaschinen zu übermitteln.

Gefundene Dokumente können in einem Fenster neben dem Browser hinsichtlich ihrer Nützlichkeit bezogen auf die Beantwortung der aktuellen Fragestellung bewertet werden. Beurteilungen und Aufzeichnung der bei der Suche im Web gewählten Pfade werden vom *Organizational Ant World Server* assoziiert mit dem jeweiligen Quest gespeichert und analysiert. Basierend auf diesen Daten werden nachfolgenden Benutzern, die nach ähnlichen Themen suchen, Listen von Empfehlungen präsentiert und außerdem Links zu relevanten Dokumenten durch Modifikation des Hypertextes mit einem Icon als *AntMark* gekennzeichnet.

Die Generierung einer Empfehlungsliste erfolgt in einem zweistufigen Prozeß. Zunächst wird das *Quest Matching* durchgeführt. Die Aufgabe, „ähnliche“ Quests zu finden, ist nicht trivial, da Anwender verschiedene Begriffe verwenden können, um übereinstimmende Ziele zu formulieren. Ein Quest umfaßt neben der von dem Anwendern verfaßten Beschreibung den Inhalt der als nützlich bewerteten Seiten. Dadurch wächst die Beschreibung des Quests dynamisch während des Suchprozesses. Mit jeder abgegebenen Bewertung können präzisere Empfehlungen ausgegeben werden. Die Bewertungsabgabe ist demnach mit den Relevance-Feedback-Mechanismen von Information-Retrieval-Systemen zu vergleichen. Zur Durchführung der Ähnlichkeitsberechnungen finden verschiedene Methoden aus dem Information-Retrieval-Vektorraummodell Verwendung. Die als ähnlich identifizierten Quests bilden dann die Grundlage für die im abschließenden Schritt des *Link Scoring* erstellte Liste von Empfehlungen relevanter Dokumente. Dabei werden URLs, die in mehreren als ähnlich identifizierten Quests übereinstimmend hoch gewertet wurden, bevorzugt empfohlen.

---

<sup>6</sup>Der Name soll auf die Analogie der durch das AntWorld unterstützten Suche im Web zur Futtersuche von Ameisen hinweisen. Findet eine Ameise Nahrung, dann hinterläßt sie eine chemische Spur, an der sich ihre Artgenossen orientieren können.

AntWorld wurde als Client-Server-Anwendung realisiert. Dabei umfaßt der zentrale Server eine Datenbank, in der die Quests und Bewertungen verwaltet werden, und einen durch Java Servlets realisierten Proxy-Server. Dieser überwacht Bewegungen im Netz und integriert JavaScript bzw. JavaApplets für die clientseitigen Komponenten der Benutzungsoberfläche in die übertragenen HTML-Dokumente. Zusammen mit dem eingebetteten Code werden auch die Daten der serverseitig berechneten Empfehlungslisten übertragen. Ein optionaler clientseitiger Proxy fügt die AntMark Icons zur Kennzeichnung potentiell relevanter Links ein.

In Kapitel 5 wurden die Unterschiede zwischen Information Retrieval und Information Filtering aufgezeigt. Das AntWorld-System, bei dem die kurzfristige Interessen ausdrückenden Quests die Basis der kollaborativen Elemente des Suchvorgangs sind, läßt sich der Kategorie Information Retrieval zuordnen. Hingegen ist das in dieser Arbeit entwickelte System (Kapitel 8) dem Bereich Information Filtering zuzuordnen, weil die Empfehlung von Dokumenten hier unter Einbeziehung der in den Benutzerprofilen modellierten langfristigen Interessen erfolgt.

Quests beinhalten explizite Formulierungen der aktuellen Rechercheziele der Anwender. Diese werden nicht nur zur Abfrage von Suchmaschinen verwendet, sondern insbesondere auch, um Kriterien für die Suche nach ähnlichen Quests zu definieren und damit das Wissen vorheriger Benutzer zu erschließen. Das Potential von Suchmaschinenanfragen wird auch im CoInternet-System genutzt. Die im Laufe der Zeit gestellten Anfragen werden in den Profilen der Benutzer gesammelt und verwendet, um nach ähnlichen Benutzern und damit nach Empfehlern für Web-Seiten zu suchen. Die Methoden des Vergleichs von Profilen sind dabei wie im AntWorld-System den Techniken des Information Retrieval entlehnt.

## 7. Anforderungen und Forschungsfragen

Aus der Beschäftigung mit den Konzepten Social Navigation (Kapitel 4) und Recommender-Systemen (Kapitel 5) heraus sowie unter Einbeziehung der aus der Betrachtung verwandter Arbeiten gewonnenen Erkenntnisse (Kapitel 6) sollen in diesem Kapitel einige Anforderungen an das zu erstellende Empfehlungs- und Annotationssystem formuliert werden (Abschnitt 7.1). Des Weiteren werden Forschungsfragen benannt (Abschnitt 7.2). Diese haben ihren Ursprung vorwiegend in der Darstellung der Probleme von Social Navigation (Abschnitt 4.8) im Allgemeinen und Collaborative Filtering (Abschnitt 5.2.3) im Besonderen. Die Forschungsfragen konkretisieren die eingangs gestellte Frage nach den Auswirkungen der Integration einer Identitätsinfrastruktur auf eine kollaborative Internetanwendung (Abschnitt 1.1) und sind bei der Konzeption des Prototyps zumindest ansatzweise zu beantworten.

### 7.1. Anforderungen an ein Empfehlungs- und Annotationssystem für Web-Seiten

Die Anforderungen an eine Social-Navigation-Anwendung sollten sich an den Bedürfnissen der Anwender orientieren und sind zunächst nur unvollständig spezifizierbar. Das Ziel, die Benutzung des Webs durch die Bereitstellung von Funktionen zum sozialen Navigieren zu erleichtern, wird demzufolge nur durch ein iteratives Vorgehen bei der Systementwicklung, bei dem sich Phasen der Konzeption, Implementation und Evaluation abwechseln, zu erreichen sein. Die in der Einleitung als ein Ziel dieser Arbeit formulierte Frage nach den Auswirkungen der Integration einer Identitätsinfrastruktur auf eine kollaborative Internetanwendung, in diesem Fall ein Empfehlungs- und Annotationssystem, läßt sich nur mittels umfangreicher Evaluationen erkunden, bei denen insbesondere auch das Zusammenspiel mit anderen die Identitätsinfrastruktur benutzenden Anwendungen zu untersuchen ist. Eine Anforderung an den zu erstellenden Prototypen ist folglich eine flexible Architektur, die als Grundlage für weitergehende Forschungsaktivitäten dienen kann. Im Kontext dieser Arbeit liegt der Schwerpunkt auf der Unterstützung indirekter Social Navigation. Formen direkter Social Navigation sollten jedoch mit geringen Modifikationen des Systems realisiert werden können, um die Perspektive offen zu halten, Kombinationen aus direkter und indirekter Social Navigation studieren zu können.

Alle Komponenten des Systems sollten über offene Protokolle kommunizieren, so daß ein einfacher Austausch gegen alternative Implementierungen möglich wird. Bezüglich der Funktionseinheit, welche aus abgegebenen Bewertungen unter Berücksichtigung der Benutzerprofile Bewertungsvorhersagen berechnet, ist ein hoher Grad an Modularisierung anzustreben, um verschiedene Vorhersagealgorithmen und die Auswertung unterschiedlicher Profilmerekmale erproben zu können. Bezüglich der empfehlbaren bzw. annotierbaren Objekte sollen keine unnötigen Einschränkungen vorgenommen werden. Es wird angestrebt, beliebige Objekte empfehlbar zu machen, so lange diese über *Uniform Resource Identifier (URIs)* [BL94] identifizierbar sind. Insbesondere werden beliebige Dokumentensammlungen mit Web-Frontend unterstützt.

Das System soll sich nahtlos in die Web-Architektur integrieren. Notwendige Veränderungen an Web-Servern würden den Kreis der Sites, an denen das entwickelte System ausprobiert werden kann, drastisch einschränken. Ebenfalls sind weitreichende Änderungen an den Browsern, die über die Installation eines Plugins hinausgehen, auszuschließen, da diese in der Praxis die Etablierung eines Empfehlungs- und Annotationssystems gravierend erschweren würden.

Social-Navigation-Funktionalität sollte möglichst gut in die zugrunde liegende Anwendung integriert werden [FHS98]. Bezogen auf das zu realisierende System bedeutet dieses, daß eine nahtlose Integration in das Bedienkonzept des Web-Browsers gefordert wird. Die Abgabe von Bewertungen sollte mit so wenig Aufwand wie möglich verbunden sein und Vorhersagen sollten sich ohne größeren kognitiven Aufwand in Navigationsentscheidungen einbeziehen lassen. Der Einstieg in das System ist einfach zu gestalten. Aufwendige Anmeldeprozeduren sind zu vermeiden, diese würden in der Praxis die Gewinnung neuer Nutzer erschweren, und das System sollte sehr schnell nützliche Empfehlungen liefern.

Ein Empfehlungs- und Annotationssystem ist mit Risiken für die Privatsphäre der Benutzer verbunden. Personen, die Bewertungen abgeben, sind dazu unter Umständen nur dann bereit, wenn sichergestellt ist, daß die Bewertungen nicht oder nur unter ihrer Kontrolle von anderen Nutzer mit ihrer realen Identität assoziiert werden können. Weitere Risiken für die Privatsphäre der Anwender liegen im Aufbau der Benutzerprofile. Es sollte dem Anwender bewußt gemacht werden, welche Daten gesammelt und wie diese verwendet werden. Dieser Punkt ist insbesondere dann wichtig, wenn Profile implizit durch Auswertung von Benutzeraktionen aufgebaut werden. Anwender sollten die vollständige Kontrolle über ihre Benutzerprofile erhalten und gegebenenfalls auch Daten vollständig löschen können. Eine Weitergabe oder Verwendung in einem anderen Kontext verbietet sich selbstverständlich. Ein zusätzliches Risiko für die Privatsphäre der Anwender ist, daß die automatische Abfrage von Bewertungen bei der Navigation im Web das Risiko der Überwachung mit sich bringt, da fortwährend die aktuell geladenen URLs an den Dienst übermittelt werden.

## 7.2. Forschungsfragen

In dieser Arbeit sollen Collaborative-Filtering-Techniken (Abschnitt 5.2) verwendet werden, um aus einer Vielzahl von Bewertungen individuelle Empfehlungen zu berechnen. Dieser Ansatz ist jedoch mit einigen Problemen verbunden, die auf unzureichendes Wissen des Systems über die Eigenschaften der Nutzer bzw. unzureichende Kenntnis der Nutzer untereinander zurückzuführen sind (Abschnitt 5.2.3). Weitere Unzulänglichkeiten ergeben sich aus der Verwendung eines auf Bewertungsmuster reduzierten Benutzermodells, wie es in typischen Collaborative-Filtering-Systemen, exemplarisch ist GroupLens zu nennen (Abschnitt 6.2), verwendet wird. Ein zentraler Aspekt der in der Einleitung (Abschnitt 1.1) gestellten Fragen nach den Auswirkungen der Integration einer Identitätsinfrastruktur auf eine kollaborative Internetanwendung ist die Frage, wie Collaborative-Filtering-Anwendungen durch Einbeziehung der Identitätsinfrastruktur verbessert werden können.

Ein Gesichtspunkt des sogenannten Kaltstartproblems ist, daß Benutzer, die neu in das System eintreten, noch über kein für die Berechnung persönlicher Empfehlungen hinreichendes Benutzerprofil verfügen. Sofern die Suche nach ähnlichen Benutzern einzig auf der statistischen Auswertung der Bewertungsmatrix basiert, sind Überschneidungen in der Menge der bewerteten Objekte Voraussetzung für den Vergleich zweier Personen. Gerade bezogen auf ein Empfehlungssystem für Web-Seiten ist diese Voraussetzung als überaus problematisch anzusehen, da bei der Vielzahl abrufbarer Seiten und der im Verhältnis dazu immer recht kleinen Menge an Nutzern des Empfehlungssystems mit einer spärlich besetzten Bewertungsmatrix zu rechnen ist. Dieses Problem wird dadurch verstärkt, daß Anwender ähnliche Interessen haben können, sich aber bei unterschiedlichen Anbietern informieren und demzufolge ihre Bewertungsmuster keine Übereinstimmungen erkennen lassen. Eine Fragestellung ist, wie das Kaltstartproblem und die aus dem Sparsity-Problem resultierenden Schwierigkeiten beim Auffinden ähnlicher Benutzer durch Einsatz der Identitätsinfrastruktur gemindert werden können. Dabei ist zu ergründen, wie Benutzerinteressen in Profilen modelliert werden können und wie Profilinformationen zu gewinnen sind. Des weiteren wird nach Algorithmen zum Vergleich von Profilen gesucht.

Auf der Technik des automatischen Collaborative Filtering basierende Systeme lassen persönliche Beziehungen unberücksichtigt. Prinzipiell wird bei diesem Verfahren davon ausgegangen, daß alle Benutzer in gleichem Maße als Empfehler geeignet sind. Dieses erscheint unangebracht, denn möglicherweise wird eine Empfehlungen abfragenden Person einigen Benutzern mehr Vertrauen entgegenbringen als anderen oder der Meinung von Experten, Freunden oder Kollegen besonderes Interesse beimessen. Zu untersuchen ist, wie sich persönliche Netzwerke, welche unter Umständen in anderen Kontexten aufgebaut wurden, auf das System abbilden und in den Empfehlungsprozeß einbeziehen lassen?

Collaborative-Filtering-Systeme wie GroupLens präsentieren sich den Anwendern als „Black Box“. Dieses erscheint für ein Empfehlungssystem für Web-Seiten unangemessen, denn die Aufgaben, welche Personen im Web verfolgen, können höchst unterschiedlich sein, so daß bezüglich der Risiken der Orientierung an Empfehlungen keine allgemeine Einschätzung getroffen werden kann (siehe Abschnitt 6.2). Auf Verlangen sollte ein Empfehlungssystem Details zum Empfehlungsprozeß präsentieren. Die Frage, ob einer bestimmten Empfehlung Vertrauen entgegenzubringen ist, ist eng mit der Fragen nach den relevanten Empfehlern, deren Kompetenzen und Vertrauenswürdigkeit verknüpft. Wie kann durch Einbeziehung der Identitätsinfrastruktur ein Empfehlungssystem dahingehend gestaltet werden, daß jeder Anwender eine informierte Entscheidung treffen kann, inwiefern er einer bestimmten Empfehlung bzw. einem Empfehler vertraut?

Der Nutzen eines Empfehlungssystems, welches auf der expliziten Abgabe von Bewertungen basiert, ist maßgeblich vom Engagement der Benutzer abhängig (siehe Abschnitte 4.8 und 5.2.3). Es ist zu untersuchen, wie ein Empfehlungs- und Annotationssystem zu gestalten ist, so daß das Zugehörigkeitsgefühl zur Community der Nutzer gefördert wird.

Die Berechnung von Empfehlungen bzw. Bewertungsvorhersagen aus einer Menge von Bewertungen ist nur dann als sinnvoll zu erachten, wenn zwischen den Nutzern ein Konsens bezüglich der bei der Bewertung anzuwendenden Kriterien vorliegt. Wie in Abschnitt 6.2 ausgeführt wurde, ist es schwierig, allgemeingültige Kriterien für die Bewertung von Web-Dokumenten anzugeben. Einerseits ist bezüglich der Funktion der Seiten innerhalb der Site-Struktur zwischen verschiedenen Seitentypen zu unterscheiden, andererseits ist die Bandbreite der Themen im Web sehr groß. Eine Aufgabe beim Entwurf des Empfehlungssystems ist es zunächst einmal Erkenntnisse darüber zu gewinnen, nach welchen Kriterien Anwender Web-Dokumente in verschiedenen Kontexten bewerten.

Anhand der Gegenüberstellung des Usenets mit dem Web in Abschnitt 6.2 wurde deutlich, daß es dem Web an Strukturen fehlt, die eine automatische Partitionierung der Menge der Bewertungen bezüglich der Domäne der bewerteten Objekte möglich machen. Bei der Konzeption eines Empfehlungssystems für Web-Seiten müssen alternative Ansätze gefunden werden, wie verhindert werden kann, daß aus übereinstimmenden Bewertungen von Seiten zu einem Thema, auf eine einhellige Meinung zu thematisch entfernten Dokumenten geschlossen wird.

In Abschnitt 3.2 wurden einige der Unzulänglichkeiten von Suchmaschinen beschrieben. Eine der Fragen dieser Arbeit ist, wie durch die Kombination eines Empfehlungssystems mit Suchmaschinen das Potential menschlichen Urteilsvermögens für die automatische Informationssuche erschlossen werden kann.

## 8. Collaborative Internet Experience – Ein Empfehlungs- und Annotationssystem für Web-Seiten

In diesem Kapitel wird der im Rahmen dieser Diplomarbeit entwickelte Prototyp vorgestellt. Abschnitt 8.1 gibt einen ersten Überblick über die Anwendung. Einsatzszenarien und hinter dem Entwurf stehende Ideen werden an dieser Stelle erläutert. Die Konzeption des Systems, unter Berücksichtigung der in Kapitel 7 formulierten Anforderungen und Forschungsfragen, wird in Abschnitt 8.2 präsentiert. Im Anschluß erfolgt die Darstellung der Konstruktion des Systems (Abschnitt 8.3). Dabei wird die gewählte Architektur vorgestellt und Details der softwaretechnischen Realisierung werden erörtert. Die Durchführung und die Ergebnisse einer ersten Evaluation sind Gegenstand von Abschnitt 8.4. Zum Abschluß dieses Kapitels werden Antworten auf die Frage gegeben, welche Auswirkungen die Integration der onefC-Identitätsinfrastruktur auf das Empfehlungs- und Annotationssystem habe.

### 8.1. Überblick

Mittels des Empfehlungs- und Annotationssystems *Collaborative Internet Experience* soll ein unabhängiger Dienst geschaffen werden, der Benutzer bei der Auffindung sie interessierender Inhalte unterstützt und bei der Beurteilung der Qualität und Glaubwürdigkeit von Inhalten im World Wide Web hinzugezogen werden kann.

Betrachtet man im Gegensatz dazu ein Empfehlungssystem für Musiktitel, so kann angenommen werden, daß die Motivation der Benutzer das Kennenlernen neuer, ihrem Geschmack entsprechender, Musik ist. Hingegen muß bei einem Empfehlungssystem für Web-Seiten davon ausgegangen werden, daß die Motivation ein Informationsbedürfnis ist, welches sich durch langfristige Interessen ausdrückende Benutzerprofile nur unvollständig spezifizieren läßt. Folglich wird nicht der aktive Vorschlag von Inhalten durch das System vorgenommen, sondern dem Benutzer durch ein in den Browser integriertes System bei der Benutzung des Webs assistiert.

Mittels des Konzepts der Intermediaries (Abschnitt 6.1) wird das Web durch diesen Dienst dahingehend erweitert, daß soziales Navigieren allgegenwärtig möglich wird.

Die Nutzer des CoInternet-Systems können kollaborieren, indem sie Web-Seiten bewerten oder mit Annotationen, welche insbesondere auch weiterführende Links enthalten können, versehen. Die Bedeutung der Meinung anderer Benutzer unterscheidet sich subjektiv, wenn beispielsweise Personen mit ähnlichen Interessen oder besonderen Kompetenzen ein besonderes Gewicht beigemessen wird. Somit erscheint eine einfache Mittelwertbildung der abgegebenen Bewertungen wenig zweckmäßig. Ferner ist anzunehmen, daß die Qualität der Annotationen stark variiert und deren Nützlichkeit abnimmt, wenn vielbesuchte Seiten um viele Anmerkungen ergänzt werden und somit eine Auswahl persönlich interessanter Annotationen erschwert wird<sup>1</sup>. Durch den Einsatz im Rahmen dieser Arbeit erweiterter Collaborative-Filtering-Algorithmen (Abschnitt 5.2) präsentiert das System individuelle Vorhersagen der zu erwartenden Relevanz von Web-Dokumenten. Zudem werden diese Filteralgorithmen analog auch auf Annotationen angewandt. Benutzer, welche die vom System berechneten Relevanzvorhersagen oder Annotationen zur Kenntnis nehmen und diese Informationen in ihre Navigationsentscheidungen einbeziehen, profitieren damit von den Erfahrungen vorheriger Benutzer, welche ihnen, vermittelt durch das CoInternet-System, kommuniziert wurden. Diese Form der Navigation durch Orientierung an von anderen Nutzern hinterlassenen Artefakten ist in Übereinstimmung mit den in Abschnitt 4.2 aufgeführten Definitionen und der Klassifikation in Abschnitt 4.3 als indirekte Social Navigation zu bezeichnen.

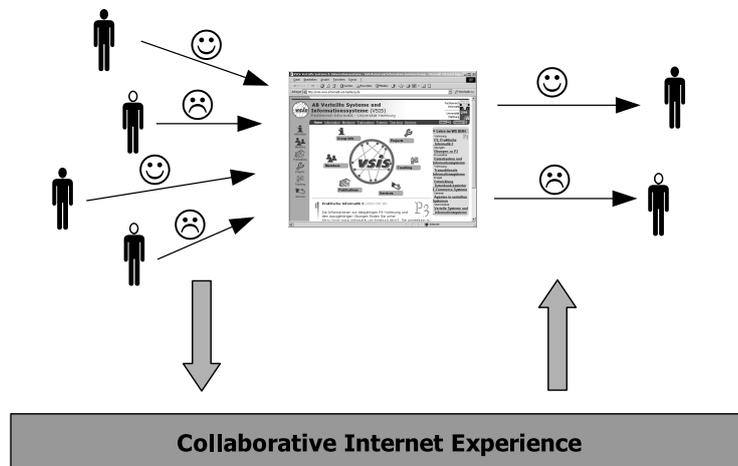
Die Bewertung der Objekte wird ausschließlich durch die Benutzer durchgeführt. Die Rolle des Systems ist es dabei, als eine Kommunikationsplattform zu dienen und eine Vermittlerfunktion zwischen Empfehlungen suchenden und gleichgesinnten bzw. vertrauten Personen, die Bewertungen abgegeben haben, zu erfüllen. Abbildung 8.1 zeigt, wie Benutzer indirekt und unidirektional über das CoInternet-System kommunizieren. Die abgegebenen Bewertungen werden durch das System zu Relevanzvorhersagen aggregiert, wobei die Aufgabe, „ähnliche“ Personen zu finden, zu bewältigen ist.

Der bei der Konzeption der Anwendung geplante Einsatzbereich ist die Unterstützung von Communities (Abschnitt 2.1.4), welche durch die Nutzer des Systems gebildet werden. Basierend auf der Systemfunktionalität, durch Vergleich von Benutzerprofilen gleichgesinnte Benutzer zu finden, können sich innerhalb der großen Community der Systemnutzer dynamisch kleinere „Communities of Interest“ herausbilden.

Die Mitglieder der Community mögen individuell sehr unterschiedliche Ziele haben und kollaborieren nur sehr lose. Die Anwendung ist folglich als Community Support System von einer Teams unterstützenden Groupware abzugrenzen, wenngleich nicht ausgeschlossen werden soll, daß es sinnvoll sein kann, Teile eines derartigen Empfehlungs- und Annotationsystem in eine Groupware-Lösung zu integrieren. Betrachtet man die Anwendung allgemein als ein CSCW-System, so kann entsprechend den in Abschnitt 2.1.3

---

<sup>1</sup>Man schaue sich nur einmal beliebige News-Sites zu IT-Themenen wie [www.heise.de](http://www.heise.de) und [www.golem.de](http://www.golem.de) an, die es gestatten, Artikel in Foren zu diskutieren.



**Abbildung 8.1.:** Das System Collaborative Internet Experience bietet eine Plattform zur Kommunikation von Bewertungen und Annotationen. „Ähnliche“ Personen werden durch gleichfarbige Köpfe dargeseilt.

aufgeführten Klassifizierungssystemen eine Einordnung vorgenommen werden. Gemäß der Klassifizierung nach Raum und Zeit handelt es sich um eine Anwendung zur asynchronen Zusammenarbeit räumlich entfernter Personen. Entsprechend der funktionellen Klassifizierung ist das System zwischen den Nachrichtensystemen und den gemeinsamen Informationsräumen anzusiedeln.

Ein Empfehlungs- und Bewertungssystem, wie das in dieser Arbeit entwickelte, könnte von großen Organisationen oder Unternehmen als ein auf den Kreis der Mitarbeiter beschränkter Dienst betrieben werden. Alternativen wären der Betrieb durch einen Internet Service Provider als ein den Kunden offerierter Mehrwertdienst. Zudem wäre ein Betrieb als ein prinzipiell allen Nutzern offenstehender Dienst im Internet möglich, welcher idealerweise an eine bereits etablierte Online-Community gekoppelt werden könnte.

Nachfolgend soll zunächst anhand zweier Szenarien die mögliche Verwendung des Systems illustriert werden. Anschließend werden in den Abschnitten 8.1.2-8.1.4 die grundlegenden hinter dem Systementwurf stehenden Ideen dargestellt.

### 8.1.1. Szenarien

Den Nutzungsgewohnheiten vieler deutscher Internet-Nutzer scheint es zu entsprechen, nur eine geringe Zahl von Web-Sites regelmäßig aufzusuchen [EGF03]. Eine Motivation

für eine solchen habitualisierte Nutzung des Webs mag es sein, in interessierenden Themen auf dem Laufenden zu bleiben. Folgende Szenariobeschreibung soll einen Eindruck davon vermitteln, wie sich das gemeinsame Sichten von Meldungen unterstützt durch das in dieser Arbeit realisierte System aus Sicht eines Anwenders darstellen könnte.

Eine Person möchte sich täglich bei Spiegel online über die aktuellen Geschehnisse informieren, hat aber leider nur Zeit, um einen kleinen Teil der Artikel zu lesen. Glücklicherweise lesen auch viele ihrer Kollegen und andere Benutzer des Collaborative-Internet-Servers täglich Spiegel online und bewerten gemäß ihren subjektiven Interessen Artikel. Bei ihren Besuchen auf der Spiegel-Seite erscheinen, wenn sie den Mauszeiger über einem Link platziert, Einblendungen individuell errechneter Relevanzvorhersagen. Benutzer, denen sie vertraut, die in der Vergangenheit ähnlich bewertet haben oder die ein ähnliches Interessensprofil aufweisen, sich beispielsweise ebenfalls für Politik interessieren, werden bei der Berechnung der Vorhersagen besonders berücksichtigt. Artikel, bei denen eine niedrige Vorhersage angezeigt wird, liest sie gar nicht erst. Manchmal findet sie einen Artikel, dessen Überschrift gar nicht so interessant aussieht, für den aber eine hohe Relevanzvorhersage ausgegeben wird. Deshalb liest sie diesen ebenfalls.

Ein weiteres Szenario, in welchem die Anwendung nutzbringend eingesetzt werden kann, ist die gemeinsame Suche nach Informationen. Dieser Aspekt ist insbesondere deshalb von Interesse, da er einen Weg aufzeigt, wie innerhalb einer Community gemeinsames Wissen gebildet werden kann.

Eine Person benötigt für ein Projekt Informationen zur Peer-to-Peer Technologie. Sie sucht nach „JXTA“, eine Suchabfrage, die auch befreundete Programmierer, die ebenfalls Teilnehmer des Empfehlungssystems sind und als vertraute Personen gekennzeichnet wurden, schon oft gestellt haben. Da sie die Abfrage über das Suchen-Feld des CoInternet-Clients gestellt hat, werden die Ergebnisse von Google durch den CoInternet-Server nach Relevanzvorhersagen sortiert. Die von den Freunden als relevant eingestuften Seiten erscheinen in der Liste an vorderster Position.

### 8.1.2. Navigation entlang sozialer Spuren

Durch Annotationen können Benutzer die Strukturen des Webs dynamisch ergänzen. Nachfolgende Benutzer bewegen sich in diesem erweiterten Informationsraum und hinterlassen Spuren durch Bewertungen, welche explizit abgegeben oder implizit aus dem

Navigationsverhalten abgeleitet wurden. Diese Art von Spuren haben jedoch eine weitgehende Bedeutung als beispielweise der Eintrag im Log-File eines Web-Servers oder die über eine Abstimmungsfunktion auf einer Web-Seite abgegebene Stimme, die anschließend in einer aufsummierten Zahl von Stimmen aufgeht. Die Spuren der Anwender des CoInternet-Systems stehen über universelle, anwendungsübergreifende Bezeichner in Verbindung mit umfangreichen Benutzerprofilen. Die Privatsphäre der Anwender wird dennoch geachtet, denn sämtliche Informationen werden pseudonymisiert gespeichert.

Die sozialen Spuren vieler Benutzer werden durch das System unter Berücksichtigung des Profils der navigierenden Person und der Profile der Bewertungen abgebenden Benutzer aggregiert. Die Links eines Hypertextes bekommen somit eine Gewichtung, welche bei der Wahl eines Weges durch den Hypertext dienlich sein kann.

Das Web wird durch das als Intermediary realisierte System zu einem Informationsraum, der teilweise durch die Benutzer geformt wird und sich bis zu einem gewissen Grad individuell an die Benutzer anpaßt, welche ihren Interessen, als einem Ausschnitt ihrer Persönlichkeit, durch die Identitätsinfrastruktur gegenüber dem Mittlersystem Ausdruck verschaffen.

### 8.1.3. Bewertungen und Annotationen sind Meta-Daten

Aus einer verallgemeinernden Sichtweise heraus sind Bewertungen und Annotationen mit Inhalten des Web verknüpfte Meta-Daten, welche über eine im weiteren Verlauf dieser Arbeit spezifizierte Schnittstelle abfragbar sind und in einem definierten Format vorliegen. Die Bedeutung der Meta-Daten wird durch die Assoziation mit Benutzerprofilen angereichert. Führt man diesen Gedanken fort, so lassen sich Benutzerprofile als Meta-Meta-Daten interpretieren, die einen Bedeutungskontext der Meta-Daten definieren. So ist beispielsweise eine positive Bewertung einer Person im Kontext der anderen von dieser Person positiv bewerteten Seiten zu betrachten.

Bewertungen und Annotation können einerseits Benutzern präsentiert werden, die ihr Verhalten entsprechend ausrichten können, andererseits sind derartige Meta-Daten, insbesondere Bewertungen und aus Freitextkommentaren extrahierte Links einer automatischen Verarbeitung zur Realisierung weitergehender Dienste zugänglich. Eine Anwendung wäre die Optimierung der Trefferlisten von Suchmaschinen. Dieser Ansatz wird in Abschnitt 8.2.6 weiterverfolgt werden. In [Wol02] wurde das Konzept beschrieben, wie persönliche Web-Robots basierend auf *Classifizieren* und *Filtern* an spezifische Aufgaben angepaßt werden können. Durch Einbeziehung des Bewertungsdienstes wäre ein Shopping-Bot realisierbar, der nicht nur nach dem günstigsten Preis für einen Artikel sucht, sondern überdies die Reputation des Händlers berücksichtigt. Ferner könnten persönliche Agenten Web-Sites nach Inhalten durchsuchen, welche nicht nur eine Men-

ge gesuchter Begriffe enthalten, sondern zusätzlich vermutlich den Interessen des den Agenten kontrollierenden Benutzers entsprechen.

#### 8.1.4. Community-Gedanke statt eines anonymen Empfehlungssystems

Collaborative-Filtering-Systeme verlangen den Benutzern ein dauerhaft hohes Engagement ab. Insbesondere bei der Abgabe der ersten Bewertung für ein Objekt ist ein gewisses Maß an Altruismus erforderlich. Bei der Konzeption des CoInternet-Systems stand der Gedanke im Vordergrund, eine Online-Community zu schaffen. Damit verbunden ist die Annahme, die Wahrnehmung des Systems als eine soziale Gemeinschaft, in Verbindung mit dem Bewußtsein dieser anzugehören, habe positiven Einfluß auf die Motivation der Benutzer, was das Free-Rider- und First-Rater-Problem (siehe Abschnitt 5.2.3) mindern würde. Das Zugehörigkeitsgefühl zu einer Online-Community und das Streben nach einer positiven Reputation innerhalb der Gemeinschaft scheint sich positiv auf die Bereitschaft der Nutzer, aktiv an einer kollaborativen Anwendung zu partizipieren, auszuwirken. Ein ermutigendes diesbezügliches Beispiel ist das eBay Feedback-Forum (siehe Abschnitt 2.2.3). Transaktionspartner bewerten sich hier gegenseitig, obwohl diese Informationen ausschließlich anderen Nutzern dienlich ist. Der Erfolg einer Verbraucher-Community wie ciao.com ist eng damit verknüpft, daß es gelingt, hinreichend viele Testberichte für eine große Zahl aktueller Produkte bereitzustellen. Es ist schon etwas verwunderlich, daß anscheinend eine für den Betrieb dieses Dienstes ausreichende Zahl von Personen bereit ist, nach dem Erwerb eines Produktes teils umfangreiche Testberichte zu verfassen und somit nachfolgenden Interessenten eine informierte Kaufentscheidung zu ermöglichen.

GroupLens ist, wie das in dieser Arbeit erstellte System, eine kollaborative Anwendung. Die Benutzungsschnittstelle macht diesen Umstand jedoch nicht deutlich und läßt das System wie eine Einzelbenutzer-Anwendung erscheinen (siehe Abbildung 6.4 auf Seite 105). Das Empfehlungssystem wird als eine Black-Box präsentiert. Das Zustandekommen einer Relevanzvorhersage ist für den Benutzer nicht nachvollziehbar.

In Kontrast dazu ist es das Ziel der CoInternet-Anwendung, den Empfehlungsprozeß transparent zu gestalten. Die Zusammensetzung einer Empfehlung läßt sich bis zu den einzelnen abgegebenen Bewertungen nachvollziehen. Außer einfachen Bewertungen auf einer festgelegten Skala können auch mit einem kurzen Text begründete Bewertungen und beliebige Annotationen an das System übermittelt werden. Diese Form der indirekten Kommunikation mit anderen Nutzern bietet den Nutzern die Gelegenheit, individuell als Persönlichkeit wahrgenommen zu werden.

Wie in Abschnitt 2.2.3 ausgeführt wurde, ist die Identifizierbarkeit von Nutzern eine Grundvoraussetzung für den Aufbau von Reputation. Durch die mit der oneC-Identitätsinfrastruktur einhergehende Möglichkeit, universelle Identifikatoren zu verwen-

den, wird die Übertragung von Reputation aus anderen Kontexten auf das CoInternet-System möglich. Umgekehrt könnte es sich aber auch motivierend auf die Anwender auswirken, daß eine im Kontext dieses Systems z. B. durch hochwertige Annotationen gewonnene Reputation auf andere kollaborative Anwendungen übertragbar ist.

„To be someone on the net“ lautet das Motto des Projektes onefC [BZL03]. Durch die Identitätsinfrastruktur werden Nutzer dazu befähigt, Attribute ihrer Persönlichkeit auszudrücken, aber auch von anderen Nutzern als Individuum kontextübergreifend wahrgenommen werden und mit ihnen kommunizieren zu können. In dieser Hinsicht scheint die Integration einer Identitätsinfrastruktur ein sinnvoller Schritt zu sein, um aus einem Empfehlungs- und Annotationssystem eine lebendige Community zu machen.

## 8.2. Konzeption eines Empfehlungs- und Annotationssystems

Gegenstand dieses Abschnittes ist die Konzeption des Empfehlungs- und Annotationssystems. Dabei soll zunächst das konzeptionelle Modell der Anwendung präsentiert werden. Ausführungen zu den Details erfolgen in Unterabschnitten.

Abbildung 8.2 zeigt das konzeptionelle Modell der Anwendung als UML-Klassendiagramm. Ausgangspunkt sind die Benutzer, welche bei der Nutzung des Dienstes unter einer Identität auftreten. Teil der Identität ist ein vom Anwender gewählter Benutzername und ein *Universal Unique Identifier (UUID)* [Kun03]. Eine Identität kann in verschiedenen Diensten Gültigkeit besitzen. Der durch die onefC-Identitätsinfrastruktur vergebene eindeutige Bezeichner gestattet die Wiedererkennung von Identitäten über Grenzen von Anwendungen hinweg. Ein Benutzer kann mehrere Identitäten annehmen, um beispielsweise berufliches und privates Surfen zu trennen. Korrespondierend zu den unterschiedlichen Interessenschwerpunkten können zu einer Identität mehrere Profile angelegt werden (Abschnitt 8.2.1). In diesen sind die Interessen der Benutzer modelliert. Sie bilden die Grundlage der Berechnung von Relevanzvorhersagen (Abschnitt 8.2.5).

Benutzer können beliebige über URIs identifizierte Objekte bewerten. Insbesondere gestattet der implementierte Client die Bewertung von Web-Dokumenten, welche über *Uniform Resource Locators (URLs)* [BLMM94] adressiert werden (Abschnitt 8.2.2). Bewertungen sind Teil des Benutzerprofils. Zusätzlich erlaubt das System, Dokumente mit textuellen Annotationen zu versehen (Abschnitt 8.2.3). Annotationen werden durch das CoInternet-System einer URL zugewiesen, unter der sie als normale HTML-Dokumente [BLC95] abrufbar sind. Diese Art der Modellierung impliziert, daß auch Annotationen bewertet bzw. annotiert werden können. Bewertungen, die neben einem Wert aus der Bewertungsskala zudem noch eine Begründung enthalten, wurden über eine optionale Assoziation zwischen einer Annotation und einer Bewertung modelliert, welche sich

beide auf die gleiche URL beziehen müssen. Als für die anderen Nutzer sichtbare Information zur Urheberschaft von Bewertungen und Annotationen wird der Benutzername der jeweiligen Identität zusammen mit der UUID ausgegeben.

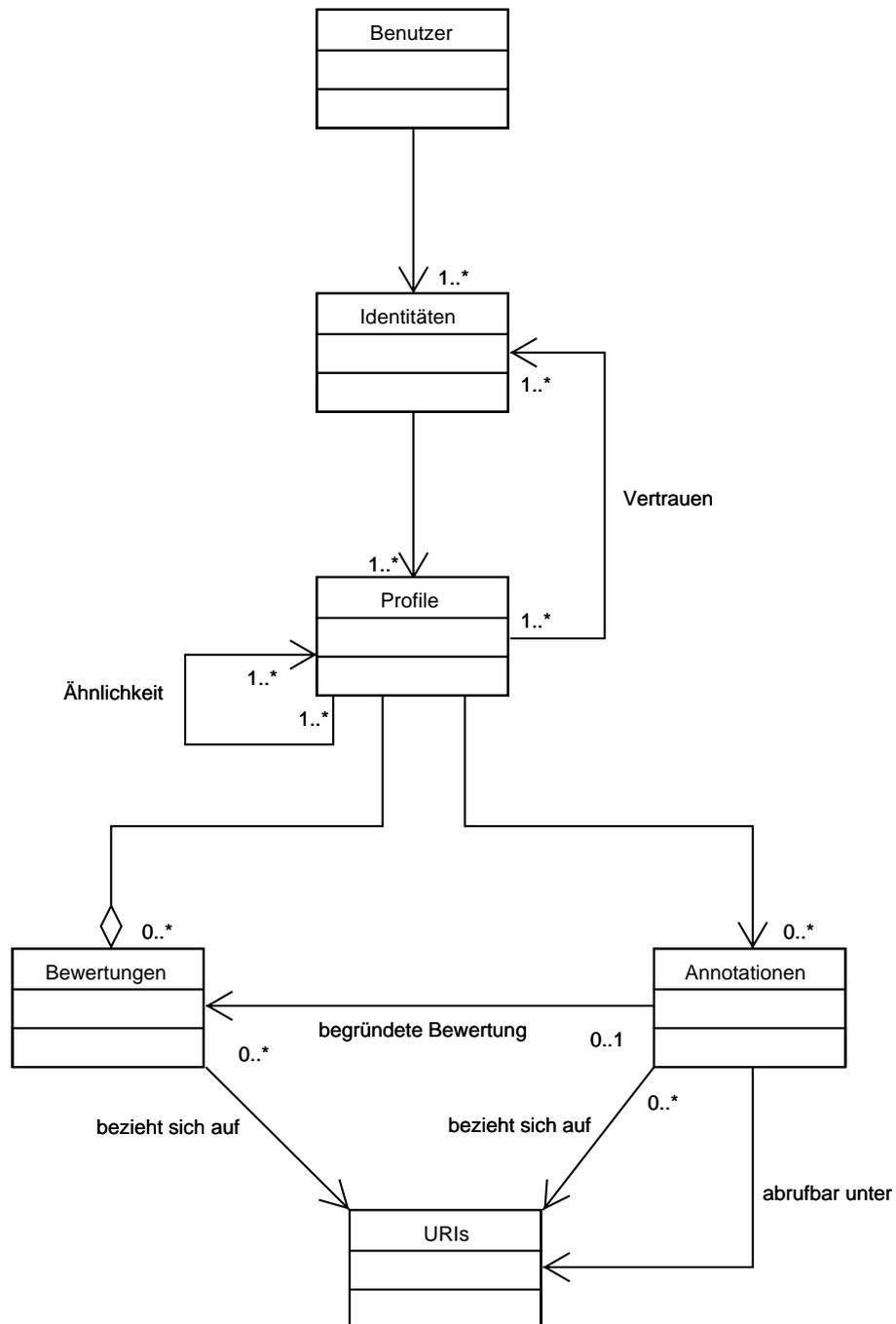
Die Verwendung von URLs zur Adressierung der Dokumente ist nicht unproblematisch. Zu einem Dokument kann es mehrere URLs geben, wenn beispielsweise eine Session-ID Bestandteil der URL ist oder die Datei auf verschiedenen Servern gespiegelt wurde. Zusätzlich können die unter einer URL abrufbaren Dokumente verändert werden, so daß nicht gewährleistet ist, daß sich die im Laufe der Zeit zu einer URL erfaßten Bewertungen und Annotationen auf die gleiche Version eines Dokuments beziehen. In der prototypischen Umsetzung wurden diesbezügliche Probleme vernachlässigt. Die Kombination von URLs mit aus den Dokumenten berechneten Checksummen, sogenannten *Document Fingerprints*, wie sie u. a. durch den *Dokumenten-Parser* des Frameworks Scone (Abschnitt 2.4) berechnet werden, erscheint jedoch sinnvoll.

Die Berechnung von Bewertungsvorhersagen erfordert die Bestimmung der Ähnlichkeit zwischen Profilen (Abschnitt 8.2.5). Zusätzlich können Benutzer für jedes ihrer Profile eine Reihe von Identitäten spezifizieren, denen sie vertrauen und deren Bewertungen bei der Berechnung von Relevanzvorhersagen besonderes Gewicht haben soll (Abschnitt 8.2.4).

Ein wesentlicher Aspekt des CoInternet-Systems ist die Unterstützung der Suche im Web durch die Kombination der Vorhersage von Bewertungen mit den Ergebnislisten einer Suchmaschine. Ein entsprechendes Konzept ist Gegenstand von Abschnitt 8.2.6. Die entworfene Benutzungsschnittstelle wird zum Abschluß der Präsentation der Konzeption in Abschnitt 8.2.7 vorgestellt.

### 8.2.1. Benutzerprofile

Die langfristigen Interessen der Benutzer werden in Benutzerprofilen modelliert. Diese dienen dazu, gemäß dem Ansatz des Collaborative Filtering, gleichgesinnte Empfehler aufzufinden und mittelbar über deren Wertungen, Relevanzvorhersagen für Dokumente zu berechnen. Die durch das System wahrgenommene und modellierte Persönlichkeit eines Benutzers ist das Produkt von dessen Selbstdarstellung bzw. von dessen Beobachtung. Naturgemäß sind derartige Benutzerprofile als unvollständig anzusehen und Gegenstand eines kontinuierlichen Lernprozesses. Die Herausforderung besteht darin, mit den gewählten Modellen und den Methoden zur Erhebung von Profilinformatoren eine für die Berechnung von Relevanzvorhersagen hinreichende Kenntnis über den jeweiligen Benutzer zu erlangen. Die Qualität der Vorhersagen läßt sich letztendlich nur durch den jeweiligen Anwender beurteilen. Nur wenn Empfehlungen mit einer für den Benutzer akzeptablen Häufigkeit mit seiner Einschätzung nach Begutachtung der jeweiligen Objekte übereinstimmen, wird der Dienst als hilfreich empfunden werden.



**Abbildung 8.2.:** Konzeptionelles Modell des Empfehlungs- und Annotationssystems als UML-Klassendiagramm

Ein auf Bewertungsmuster reduziertes Benutzermodell erscheint für ein System zur Empfehlung von Web-Seiten unzureichend, da erst nach einer langen Initialisierungsphase sinnvolle Bewertungsvorhersagen errechnet werden können und der Vergleich von Profilen Überschneidungen in der Menge der bewerteten URLs erfordert (siehe Abschnitte 5.2.3 und 7.2).

Der in dieser Arbeit gewählte Ansatz ist es, dem Kaltstart- und Sparsity-Problem dadurch zu begegnen, daß einerseits zusätzlich zu den Bewertungen weitere Attribute in das Benutzermodell einbezogen werden, um die Datenbasis für den Vergleich von Benutzerprofilen zu vergrößern, und andererseits durch Verwendung der Identitätsinfrastruktur die Möglichkeiten, Profilinformatoren zu erheben und zu kommunizieren, verbessert werden.

Personalisierte Dienste im Internet sind ähnlich dem CoInternet-System darauf angewiesen, die Präferenzen der Anwender auf Benutzerprofile abzubilden. Durch die oneC-Identitätsinfrastruktur können Profile auf dem Rechner des Benutzers abgelegt und unter seiner Kontrolle an verschiedene Dienste übertragen werden. Sofern mehrere Dienste die gleichen Daten interpretieren können, etwa weil man sich auf ein gemeinsames XML-Schema verständigt hat, wäre der anwendungsübergreifende Austausch von Benutzerprofilen praktikabel. Dem Benutzer bliebe die redundante manuelle Erfassung von Daten erspart und er könnte von Beginn an ein an seinen Bedürfnissen orientiertes System nutzen.

Ein Empfehlungssystem würde vom Austausch von Profilinformatoren über eine Identitätsinfrastruktur profitieren, denn es könnten ohne Initialisierungsphase Bewertungsvorhersagen errechnet werden. Die Bandbreite der Daten, die in ein zum kollaborativen Filtern verwendetes Benutzerprofil integriert werden könnte, ist groß. Beispiele wären die Historie der bei einem online Händler gekauften Bücher oder die Liste der interessierenden Themen, die zur Personalisierung eines News-Portals angegeben wurde.

Benutzerprofile des CoInternet-Systems werden aus verschiedenen Attributen komponiert. Sie können aus verschiedenen Quellen gespeist werden, explizit durch die Anwender spezifiziert oder implizit aus ihren Aktionen abgeleitet werden. Ziel bei der Konstruktion des CoInternet-Systems ist es, die Verarbeitung von Attributen der Benutzerprofile flexibel zu gestalten, so daß die Hinzunahme weiterer Attribute nur wenige Änderungen bedingt.

Die Benutzerprofile werden ausschließlich zur Berechnung von Empfehlungen benutzt und können mit Ausnahme der Auflistung der Bewertungen nicht von anderen Teilnehmern der CoInternet-Community eingesehen werden.

ATTRIBUT	ERFASSTE WERTE
Bewertungen	{{(URL, Bewertung, Datum)}}
Interessen	{{(Bezeichnung, Gewicht)}}
Gruppenzugehörigkeiten	{Gruppenbezeichnung}

**Tabelle 8.1.:** Für die Berechnung von Ähnlichkeiten herangezogene *explizit* anzugebende Attribute eines Benutzerprofils

### Expliziter Angabe von Profilinformatoren

Eine Vorgehensweise bei der Erhebung von Benutzerprofilen ist es, die benötigten Informationen mittels Formularen oder anderen Interaktionselementen abzufragen. Dieser explizite Ausdruck der Präferenzen eines Benutzers führt zu akkuraten Daten, ist jedoch mit Aufwand für den Anwender verbunden, der folglich ein Mindestmaß an Motivation aufbringen muß. Die für die prototypische Umsetzung des Empfehlungs- und Annotationsystems verwendeten Attribute sind aus Tabelle 8.1 ersichtlich.

Explizit abzugebende Bewertungen sind ein Bestandteil des Benutzerprofils. Für Details zur Bewertungsabgabe sei auf Abschnitt 8.2.2 verwiesen.

Ein sehr einfacher Ansatz, die Interessen eines Benutzers zu modellieren, ist eine durch den Benutzer zu erfassende Liste mit entsprechend ihrer Bedeutung gewichteten Stichwörtern. Die Verarbeitung von Formulierungen eines Benutzers enthaltenden Attributen eines Benutzerprofils beinhaltet Schwierigkeiten auf einer semantischen Betrachtungsebene. Das Ergebnis des Vergleichs der Attribute zweier Profile kann verfälscht werden, wenn diese Synonyme oder Homonyme enthalten. Möglicherweise relativieren sich diese Probleme jedoch mit zunehmender Anzahl der in einem Profilattribut gesammelten Begriffen.

Ein weiteres Attribut der modellierten Benutzerprofile ist die Auflistung der Gruppen, denen sich ein Benutzer zugehörig fühlt. An dieser Stelle könnten beispielsweise Mitgliedschaften in Verbänden, Vereinen und Fan-Clubs genannt werden. Benutzer könnten jedoch unterschiedliche Bezeichnungen für die gleiche Gruppen nennen. Insofern ist diese Art der Modellierung als nicht optimal anzusehen. Eine Alternative wäre die Auswahl aus einer kollaborativ verwalteten oder von den Administratoren vorgegeben Liste von Gruppen.

Die Erfassung von Benutzerprofilen und die spätere Anpassung und Erweiterung bestehender Profile geschieht über den Profil-Editor. Ein Screenshot ist in Abbildung 8.3 dargestellt.



**Abbildung 8.3.:** Mittels des Profil-Editors können alle Attribute eines Benutzerprofils editiert werden.

### Impliziter Aufbau von Benutzerprofilen

Aus dem Verhalten des Benutzers lassen sich verschiedene Attribute eines Benutzerprofils ableiten. Die Genauigkeit der so gewonnenen Informationen ist tendenziell schlechter als bei der expliziten Erfassung. Kompensiert wird dieser Nachteil jedoch dadurch, daß der Aufwand für die Anwender minimiert wird und mit zunehmender Dauer der Benutzung des Systems die Menge der gewonnenen Profilinformatoren automatisch zunimmt.

In Tabelle 8.2 sind die implizit ermittelten Profilattribute aufgeführt. Auf implizite Bewertungen und Bookmarks wird in Abschnitt 8.2.2 näher eingegangen.

Die an Suchmaschinen übermittelten Anfragen werden zum Aufbau von Benutzerprofilen herangezogen. Zur Vereinfachung werden nur Anfragen berücksichtigt, die aus einer Abfolge mit „und“ verknüpfter Begriffe bestehen. Die Suchbegriffe werden zusammen mit dem Datum der letzten Suche nach diesem Begriff und der Anzahl der Suchanfragen, in denen nach diesem Begriff recherchiert wurde, erfaßt.

Auch die implizit aufgebauten Attribute eines Profils können jederzeit eingesehen und verändert werden. So ist es möglich, im Profil-Editor Begriffe aus der Liste der Suchbegriffe zu entfernen.

Teil des Scone-Frameworks ist eine Komponente zur Benutzerverfolgung (Abschnitt 2.4). So gewonnene Daten zum Navigationsverhalten im Web wären für die Verwendung in Benutzerprofilen verwendbar. Die im Web gegangenen Pfade sind exemplarisch für ein auf diesem Wege zu erfassendes Profilattribut zu nennen.

ATTRIBUT	ERFASSTE WERTE
Bewertungen	{(URL, Bewertung, Datum)}
Bookmarks	{URL}
Suchanfragen	{(Wort, Datum der letzten Suche, Anzahl Suchen)}

**Tabelle 8.2.:** Für die Berechnung von Ähnlichkeiten herangezogene *implizit* erhobene Attribute eines Benutzerprofils

### Verhinderung fragwürdiger Cross-Domain-Empfehlungen

Die dem Collaborative Filtering zugrundeliegende Annahme, daß übereinstimmende Meinungen in der Vergangenheit auf eine ähnliche Bewertung neuer Objekte schließen lassen, ist nur dann als zulässig zu erachten, wenn die beurteilten Objekte einer Domäne zuzuordnen sind (siehe Abschnitt 5.2). Ein Ansatz, um zu verhindern, daß unzulässige Schlüsse aus übereinstimmenden Wertungen geschlossen werden, ist die Partitionierung des Objektraumes gemäß eines Klassifizierungsschematas (siehe Abbildung 8.4). Diese Technik wird in den Projekten Grouplens und Knowledge-Pump angewendet, erscheint aber bezogen auf die Empfehlung von Web-Seiten als ungeeignet (Kapitel 6).

Der im CoInternet-System gewählte Ansatz ist die Verwendung mehrerer, einer Identität zugeordneter Benutzerprofile. Es empfiehlt sich, für grundlegend unterschiedliche Themengebiete jeweils ein Profil anzulegen. Die Einteilung der Benutzerprofile ist für andere Nutzer unsichtbar. Für jeden der in einem Profil modellierten Interessensschwerpunkte eines Benutzers wird separat eine Menge von Empfehlern ermittelt. Jeder Benutzer legt durch die Erstellung von Profilen und die Zuordnung von Attributen und Bewertungen zu den Profilen sein individuelles Klassifikationsschema fest und definiert durch die erfassten Profilm Merkmale Kriterien für die Auswahl von Empfehlern. Legt ein Benutzer beispielsweise ein Profil mit der Bezeichnung „Java“ an, dann kann er durch die Abgabe von Bewertungen selbst einige Seiten diesem Thema zuordnen und damit die Menge der in diesem Kontext zum Vergleich von Profilen herangezogenen Dokumente einschränken. Eine Aufteilung des Benutzermodells auf mehrere Profile hat zur Folge, daß abhängig vom gewählten Profil die Menge der zum Vergleich mit anderen Benutzern verwendeten Informationen insbesondere auch die Bewertungsmatrix eingeschränkt wird (siehe Abbildung 8.5). Für den Anwender ist die Vorgehensweise mit einem geringen Mehraufwand verbunden. Er muß selbstständig eine Aufteilung seiner Interessensschwerpunkte auf verschiedene Profile vornehmen und beim „Surfen“ im Web gegebenenfalls zwischen Profilen umschalten.

		Benutzer 1	Benutzer 2	Benutzer 3	Benutzer 4
Domäne 1	Objekt 1	1	...	...	...
	Objekt 2	...	...	...	...
	Objekt 3	2	...	...	...
Domäne 2	Objekt 4	-1	...	...	...
	Objekt 5	2	...	...	...
	Objekt 6	?	...	...	...

**Abbildung 8.4.:** Durch die Partitionierung des Objektraumes wird der Teil der Bewertungsmatrix, welcher bei der Berechnung einer Bewertungsvorhersage ausgewertet wird, eingegrenzt.

	Benutzer 1			Benutzer 2	Benutzer 3	
	Identität 1		Identität 2	Identität 3	Identität 4	
	Profil 1	Profil 2	Profil 3	Profil 4	Profil 5	Profil 6
Objekt 1		-2	...	...	...	...
Objekt 2	-2		...	...	...	...
Objekt 3	2		...	...	...	...
Objekt 4		2				
Objekt 5		0	...	...	...	...
Objekt 6	2		...	...	...	...
Objekt 7	?		...	...	...	...

**Abbildung 8.5.:** Der bei der Berechnung einer Bewertungsvorhersage berücksichtigte Teil der Bewertungsmatrix ist abhängig vom gewählten Profil und den darin enthaltenen Wertungen.

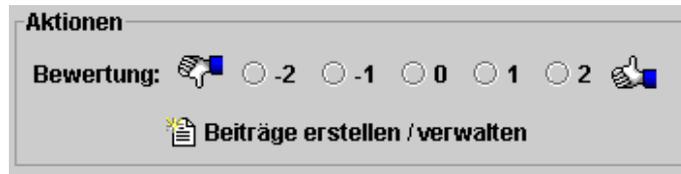


Abbildung 8.6.: Explizite Abgabe von Bewertungen

### 8.2.2. Abgabe von Bewertungen

Jedes Mitglieder der CoInternet-Community kann und soll Web-Seiten bewerten. Diese sind Teil seines Benutzerprofils, nehmen aber eine Sonderstellung ein, da sie als Profilattribut nicht nur der Auffindung ähnlicher Benutzer dienen, sondern die Berechnung von Empfehlungen direkt auf den gesammelten Werten basiert.

Bewertungen können explizit durch den Benutzer erfaßt oder implizit aus seinen Aktionen abgeleitet werden.

#### Explizite Abgabe von Bewertungen

Die Verwendung expliziter Bewertungen ist ein Weg, um an präzise Beurteilungen der Objekte eines Informationsraumes zu gelangen. Nachteilig ist jedoch der Mehraufwand, den diese Art der Datenerhebung für die Benutzer mit sich bringt. Dieser besteht nicht nur darin, durch einen Mausklick oder Tastendruck eine Wertung abzuschicken. Vielmehr bedeutet es eine zusätzliche Belastung für den Anwender, über die persönliche Relevanz der aktuellen Seite zu reflektieren und eine Entscheidung bei der Auswahl eines Wertes aus der Bewertungsskala zu treffen. Jemand, der aktiv eine Bewertung abgibt, handelt in der Intention, Navigationshinweise an nachfolgende Besucher zu kommunizieren und dabei das eigene Benutzerprofil zu verbessern. Insofern kann die Verwendung expliziter Bewertungen als „fair“ gegenüber der Bewertungen abgebenden Person bezeichnet werden, da diese selbstständig entscheiden kann, wann sie welche Dokumente bewertet.

Die explizite Abgabe von Bewertungen erfolgt über eine Gruppe von Radio-Buttons in einem Fenster neben dem Browser (siehe Abbildung 8.6). Eine Bewertung im CoInternet-System umfaßt neben der URL, auf die sie sich bezieht, und dem Datum, an dem sie abgegeben wurde, eine Benotung auf einer Skala mit den Werten -2,-1,0,1,2. Mit der Höchstnote 2 sind Web-Seiten zu bewerten, die als besonders interessant angesehen werden, von hoher Qualität sind und von vertrauenswürdigen Urhebern stammen. Diese Bewertungskriterien sind als vorläufig anzusehen und müssen weiter konkretisiert werden, wobei die Vielfalt der Inhalte im Web zu berücksichtigen ist.

Sofern nicht die erste Bewertung für ein Objekt abgegeben wird, wird dem Benutzer vor Abgabe seiner Beurteilung bereits eine Bewertungsvorhersage präsentiert. Diese könnte dessen Meinung beeinflussen [CLA<sup>+</sup>03]. Im schlimmsten Fall kann es zu Schneeballeffekten kommen (siehe Abschnitt 4.8).

Die obligatorische Angabe von Bewertungen zu vom System vorgeschlagenen Objekten zur Initialisierung eines Profils beim Eintritt in ein Empfehlungssystem ist eine Option, um das Kaltstartproblem zu mindern [SM95]. In diesem Fall kommt der Auswahl der Objekte eine besondere Bedeutung zu, was diesen Ansatz für ein Web-Seiten-Empfehlungssystem ungünstig erscheinen läßt.

### **Implizite Abgabe von Bewertungen**

Implizit aus der Beobachtung des Verhaltens eines Benutzers abgeleitete Bewertungen haben bei der Feststellung der Interessen eines Benutzers eine geringere Aussagekraft als explizite Bewertungen. Es ist einzukalkulieren, daß ein Teil der impliziten Bewertungen nicht mit der tatsächlichen Beurteilung des Anwenders in Einklang steht. Hingegen liegt der Vorteil darin begründet, daß die Abgabe von Bewertungen mit keinem zusätzlichen Aufwand für den Benutzer verbunden ist. Jeder Teilnehmer eines Empfehlungssystems leistet somit bereits durch seine gewöhnlichen Aktivitäten einen Beitrag, der anderen Nutzern zu Gute kommt. Es ist anzunehmen, daß ein Benutzer eines auf impliziten Bewertungen basierenden Empfehlungssystems nicht in der Absicht agiert, durch sein Handeln Navigationshinweise zu kommunizieren. Eventuell ist ihm dieser Zusammenhang auch gar nicht bewußt. Die Verwendung impliziter Bewertungen ist folglich mit Risiken für den Schutz der Privatsphäre der Anwender verbunden. Der Nutzen impliziter Bewertungen liegt insbesondere darin, daß durch ihre Einbeziehung die Menge der erfaßten Bewertungen maximiert werden kann und sich somit das Sparsity-Problem relativiert.

Implizite Bewertungen können zum einen aus der Analyse des Navigationsverhaltens geschlossen werden und zum anderen durch Überwachung von Aktionen des Benutzers im Browserprogramm erstellt werden.

Das Scone-Framework bietet optimale Möglichkeiten zur Beobachtung der Navigation im Web. Basierend auf Scone wurde eine Schnittstelle zur Programmierung von Modulen zur Durchführung impliziter Bewertungen entwickelt.

Im einfachsten Fall könnte bereits die Tatsache, daß eine Seite aufgerufen wurde, als eine, wenngleich wenig aussagekräftige, Wertung angesehen werden. Eine geringfügige Verbesserung wäre die Hinzunahme der Anzahl der Zugriffe, so daß nur mehrmalige Zugriffe als positives Votum gewertet werden würden. Im Rahmen des Projekts GroupLens wurde ermittelt, daß eine positive Korrelation zwischen der mit der Lektüre ei-

nes Usenet-Artikels verbrachten Zeit und der anschließend abgegebenen Bewertung besteht [KMM<sup>+</sup>97]. Diesem Beispiel folgend wurde der Versuch unternommen, implizite Wertungen in das CoInternet-System zu integrieren. Liest ein Benutzer einen langen Artikel, nach seiner Verweilzeit auf der Seite zu urteilen, vollständig, so wird das als positive Bewertung interpretiert. Ob diese Annahme Gültigkeit besitzt, muß in weiterführenden Untersuchungen ermittelt werden.

Das Setzen eines Bookmarks ist ein relativ sicherer Hinweis auf Interesse an einem Dokument [RP97]. Bookmarks liegen unter Umständen bereits vor. Sie sind somit vorzüglich geeignet, um Benutzerprofile zu initialisieren. Bookmarks werden im CoInternet-System zur Berechnung von Empfehlungen und zum Vergleich von Profilen verwendet. Sie werden dennoch separat von den aus dem Navigationsverhalten abgeleiteten Bewertungen in das Benutzerprofil integriert. Ihnen wird gegenüber den beispielsweise aus der Überwachung der Lesezeiten abgeleiteten Bewertungen eine höhere Aussagekraft zugesprochen.

Weitere Indikatoren für das Interesse an einem Dokument sind das Abspeichern und Ausdrucken. Die aufgeführten Beispiele impliziter Bewertungen verdeutlichen, daß es viele Hinweise auf das Interesse an einer Web-Seite gibt, es hingegen diffizil ist, Desinteresse festzustellen [CLWB01].

### 8.2.3. Annotationen

Web-Seiten können mittels des CoInternet-Systems mit Annotationen versehen werden. Diese bestehen aus einer Überschrift und einem unstrukturierten Text. In den Text eingegebene URLs werden automatisch in Hyperlinks umgewandelt. Annotationen beziehen sich immer auf das gesamte Dokument. Dadurch wurde dem Problem der Verankerung von Annotationen in sich ändernden Dokumenten ausgewichen.

### 8.2.4. Vertrauen

Vertrauen in die Anbieter von Navigationsunterstützung ist eine wichtige Randbedingung des sozialen Navigierens (siehe Abschnitt 4.6). Folglich war es beim Entwurf des CoInternet-Systems ein Ziel, die individuelle Berechnung von Bewertungsvorhersagen nicht nur auf der Ähnlichkeit zwischen Benutzerprofilen zu basieren, sondern auch den Grad an Vertrauen zwischen den Mitgliedern der CoInternet-Community zu berücksichtigen. Mit dieser Anforderung sind zwei Fragen verbunden. Zum einen ist der Benutzer dabei zu unterstützen, zu entscheiden, wem er vertrauen kann. Zum anderen ist ein Vertrauensmodell zu wählen, aus dem abgeleitet werden kann, inwieweit ein Nutzer einem anderen vertraut.

Eine Voraussetzung für eine informierte Entscheidung, einem anderen Nutzer Vertrauen entgegenzubringen, ist die Sichtbarkeit seiner zuvor in die Gemeinschaft eingebrachten

Beiträge. Im CoInternet-System läßt sich die Zusammensetzung von Relevanzvorhersagen bis hin zu einzelnen Stimmen nachvollziehen. Die Urheberschaft von Annotationen ist nachvollziehbar und zu jeder am System registrierten Identität kann die Liste abgegebener Bewertungen abgefragt werden (siehe Abschnitt 8.2.5).

Bei der Angabe vertrauter Personen kann die Tatsache hilfreich sein, daß onfC-Identitäten anwendungsübergreifend Gültigkeit besitzen können. Aus anderen Kontexten eventuell sogar unter der realen Identität bekannte Personen, zu denen eine Beziehung besteht oder die eine gute Reputation besitzen, können innerhalb der CoInternet-Community wiedererkannt und als vertraute Person angegeben werden. Somit ist es möglich, bereits beim Eintritt in das System, zu einem Zeitpunkt also, in dem noch keine auf der Benutzung des Empfehlungs- und Annotationssystems beruhenden Erfahrungen zu den Eigenschaften anderer Nutzer vorliegen, einige vertraute Personen zu benennen. Unter Verwendung dieser Informationen können ohne Initialisierungsphase persönliche Relevanzvorhersagen berechnet werden.

Der onefC-Identitätsmanager speichert nicht nur eigene Identitäten, sondern auch sogenannte *Foreign Identities*. Dies sind Abbilder der Identitäten anderer Personen, mit denen unter Einbeziehung des Identitätsmanagers interagiert wurde [Kun03]. Ein Assistent könnte bei der Angabe vertrauter Personen hilfreich sein, indem er auf Personen hinweist, deren Foreign Identity im lokalen Identitätsmanager gespeichert ist und die ebenfalls Teil der CoInternet-Community sind.

Da in einer virtuellen Umgebung viele Eigenschaften eines Menschen unsichtbar sind, prägt die Kommunikation maßgeblich das Bild von einer Person (siehe Abschnitt 2.2.1). Durch ein einheitliches Identitätsmodell und global eindeutige Identifikatoren wird die Interoperabilität zwischen verschiedenen Online-Anwendungen erleichtert. Personen, die durch den CoInternet-Dienst aufeinander aufmerksam geworden sind, könnten über andere Anwendungen miteinander direkt kommunizieren. Der Aufbau vertrauensvoller Beziehungen zwischen den Nutzern des Systems wird somit begünstigt.

Eine Liste der vertrauten Personen ist Bestandteil des Benutzerprofils. Dieses Attribut nimmt eine Sonderstellung ein, da es nicht der Auffindung ähnlicher Personen dient, sondern explizit Ausdruck von bei der Berechnung von Vorhersagen besonders zu berücksichtigenden Personen ist. Die Anwender erhalten durch manuelle Angabe von Personen, denen sie vertrauen und deren Meinung als relevant erachtet wird, partielle Kontrolle über den Filterprozeß. Damit wird dem Umstand Rechnung getragen, daß die Meinung von Experten, Freunden oder Kollegen, unabhängig von der Ähnlichkeit der Profile, bedeutsam sein kann.

Das Vertrauen zu einer anderen Person bezieht sich meistens auf einen bestimmten Kontext (Abschnitt 2.2.2). In den verschiedenen Profilen können deshalb unterschiedliche Identitäten angegeben werden, denen man vertraut.

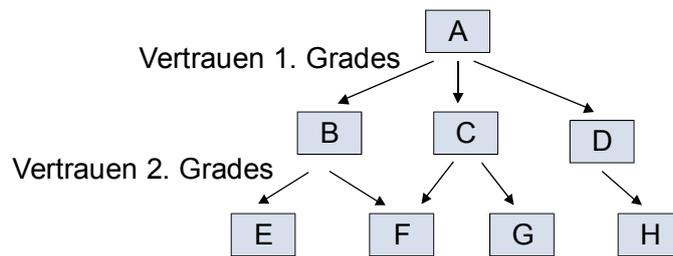


Abbildung 8.7.: Web of Trust des Benutzers A

Das Vertrauensmodell des CoInternet-Systems berücksichtigt nicht nur die von einem Benutzer ausgewählten vertrauenswürdigen Mitglieder, sondern auch die wiederum von diesen Personen angegebene Liste vertrauter Personen. Anschaulich läßt sich dieses Modell als *Web-Of-Trust* beschreiben [KMW00]. Abbildung 8.7 zeigt das Web-of-Trust des Benutzers A. Er vertraut B,C und D und damit indirekt auch E,F,G und H.

Durch die Benennung vertrauter Personen wird eine Relation zwischen den Benutzern des Systems definiert, die sich auf einen gerichteten Graphen abbilden läßt. Um zu ermitteln, ob eine Person einer anderen indirekt vertraut, ist eine Breitensuche in diesem Graphen durchzuführen. Das Vertrauen zwischen zwei Nutzern wurde mit dem Reziprokwert der Länge des kürzesten Pfades quantifiziert. Im Beispiel aus Abbildung 8.7 vertrauen sowohl B als auch C dem Benutzer G. Eine Weiterentwicklung des Berechnungsmodells wäre es, nicht nur die Länge der Pfade, sondern auch die Anzahl der Pfade zwischen zwei Benutzern zu berücksichtigen.

Das soziale Netzwerke der Anwender wird durch das Web-of-Trust-Vertrauensmodell und die Integration einer Identitätsinfrastruktur für den Filteringprozeß nutzbar gemacht.

### 8.2.5. Vorhersage von Bewertungen

In diesem Abschnitt wird die Funktionalität zur Berechnung von Bewertungsvorhersagen mittels Collaborative-Filtering-Techniken vorgestellt. In Unterabschnitten wird der Vergleich von Profilen, die Aggregation der Bewertungen zu Vorhersagen und die Präsentation der Ergebnisse beschrieben. Ein Exkurs zeigt Wege auf, wie die Berechnung von Empfehlungen verteilt in einem Netzwerk ausgeführt werden kann. Zum Abschluß dieses Abschnittes wird gezeigt, wie bei der Gestaltung der Benutzungsoberfläche des CoInternet-Systems versucht wurde, den Bewertungsprozeß für den Anwender transparent zu gestalten.

### Vergleich von Benutzerprofilen

Um eine individuelle Vorhersage der zu erwartenden Relevanz eines Objektes berechnen zu können, müssen die Profile der Benutzer, die für das fragliche Objekte eine Bewertung abgegeben haben, mit dem Profil des aktiven Benutzers verglichen werden. Ein Benutzerprofil besteht aus verschiedenen Attributen, die separat verglichen werden (siehe Tabellen 8.1 und 8.2 auf den Seiten 125 und 127.) Als Ergebnis dieses Schrittes wird die Ähnlichkeit zweier Benutzer durch einen Vektor ausgedrückt, dessen Komponenten die durch Vergleich der einzelnen Attribute errechneten Ähnlichkeiten sind.

Die Ähnlichkeit zwischen zwei Profilen wird anhand der expliziten Bewertungen wie in Abschnitt 5.2.1 beschrieben durch Korrelationsanalysen ermittelt. Für implizite Bewertungen, bei denen ausschließlich positive Wertungen erfaßt wurden, ist die Berechnung von Korrelationen nicht geeignet. In diesem Fall ist nicht die Frage, ob Benutzer bei den gemeinsam bewerteten Meldungen einer Meinung waren, relevant, sondern primär die Überschneidung in der Menge der implizit bewerteten Dokumente zu überprüfen.

Profilattribute, wie die Menge der interessierenden Themen, lassen sich als Dokumente betrachten, die mit Methoden des Information-Retrieval-Vektorraummodells [SM83] verglichen werden. Profilattribute eines Profils  $i$  werden als ein Eigenschaftsvektor  $A_i = (e_{i,1}, e_{i,2}, \dots, e_{i,d})$  repräsentiert. Dabei sind Eigenschaften Begriffe oder URLs und die Komponenten  $e_{i,k}$  das Gewicht der jeweiligen Eigenschaft in den Werten eines Profilattributes eines Benutzerprofils. Im einfachsten Fall wird nur das Vorhandensein einer Eigenschaft beachtet. Die Gewichtung nimmt somit nur die Werte 0 (Eigenschaft vorhanden) oder 1 (Eigenschaft nicht vorhanden) an. Diese Art der Gewichtung ist für aus einer Menge von Eigenschaften bestehenden Attributen, wie den Bookmarks, den impliziten Bewertungen<sup>2</sup> oder den Gruppenzugehörigkeiten angemessen. Bei den Attributen Interesse und Suchbegriffe wird mit dem explizit vom Benutzer angegebenen Gewicht bzw. der Anzahl der Suchen gewichtet. Die Dimension  $d$  der Vektoren entspricht der Zahl unterschiedlicher Eigenschaften eines Attributes in der Menge der im System vorliegenden Profile. Abbildung 8.8 zeigt exemplarisch die Attribute Bookmarks und Interessen zweier Profile in Vektordarstellung.

Zwei Attribute  $A_i$  und  $A_j$  der Profile  $i$  und  $j$  werden durch folgende Vektoren beschrieben:

$$A_i = (e_{i,1}, e_{i,2}, \dots, e_{i,d})$$
$$A_j = (e_{j,1}, e_{j,2}, \dots, e_{j,d})$$

Die Ähnlichkeit  $sim_{A,i,j}$  der Profile  $i$  und  $j$  auf Basis des Vergleichs von Attribut  $A$  läßt sich nun mit im Vektorraummodell des Information Retrieval gebräuchlichen Methoden bestimmen:

---

<sup>2</sup>Sofern nur positive Bewertungen implizit ermittelt werden

	Profil 1	Profil 2
http://www.dvz.de	1	1
http://www.eurailpress.com	1	0
http://www.maritime-trade-press.com	0	1
http://www.marilog.de	0	1
http://www.tfi-publications.com	0	0
http://www.cim-publications.de	0	0
http://www.spiegel.de	1	1

(a)

	Profil 1	Profil 2
HTML	1	5
Java	2	1
DHTML	0	3
Zope	0	0
Phyton	0	0
CGI	0	2
Web Services	5	2

(b)

**Abbildung 8.8.:** Attribute der Benutzerprofile werden als Vektoren repräsentiert.  
 (a) Bookmarks, (b) Interessen (gewichtet)

$$sim_{A,i,j} = \frac{2 \left[ \sum_{k=1}^d (e_{i,k} \cdot e_{j,k}) \right]}{\sum_{k=1}^d e_{i,k} + \sum_{k=1}^d e_{j,k}} \quad (8.1)$$

$$sim_{A,i,j} = \frac{\sum_{k=1}^d (e_{i,k} \cdot e_{j,k})}{\sqrt{\sum_{k=1}^d e_{i,k}^2 \cdot \sum_{k=1}^d e_{j,k}^2}} \quad (8.2)$$

Formel 8.1 wird als *Dice-Koeffizient* bezeichnet [SM83]. Sofern nur das Vorhandensein bzw. Abhandensein von Eigenschaften berücksichtigt wird, entspricht der Zähler des Koeffizienten der doppelten Anzahl gemeinsamer Eigenschaften, während der Nenner anschaulich als die Summe der Eigenschaften der beiden zu vergleichenden Profilattribute zu verstehen ist. Der Dice-Koeffizient wird zum Vergleich der impliziten Bewertungen, Bookmarks und Gruppenzugehörigkeiten verwendet.

Formel 8.2 stellt den Cosinus-Koeffizient<sup>3</sup> dar [SM83]. Er entspricht anschaulich dem Winkel zwischen den Vektoren in einem mehrdimensionalen Raum. Die Interessen und gesuchten Begriffe werden durch den Cosinus-Koeffizient verglichen.

Übereinstimmungen in insgesamt über die Menge aller Profile hinweg betrachtet seltenen Eigenschaften sollten in besonderem Maße beim Vergleich von Profilattributen berücksichtig

<sup>3</sup>Man beachte die große Ähnlichkeit mit Formel 5.2 auf Seite 88 zur Berechnung der Korrelation von Bewertungen. Bei der Berechnung der Korrelationen werden jedoch nur die beiden Vergleichsprofilen gemeinen Eigenschaften (Bewertungen) betrachtet.

sichtigt werden. Eine Gewichtung der Eigenschaften mit der *Inverse Document Frequency* [SM83] könnte zu aussagekräftigeren Ergebnissen der Vergleichsoperation führen.

Die gewählten Verfahren zum Vergleich von Profilen sind exemplarisch für mögliche Implementationen der Ermittlung ähnlicher Profile zu sehen. Die Leistung eines gewählten Algorithmus zum Vergleich von Profilen ist letztendlich an der damit erreichten Vorhersagegenauigkeit zu messen, welche sich empirisch mit dem in Abschnitt 5.2.4 skizzierten Verfahren bestimmen läßt.

Festzuhalten bleibt an dieser Stelle, daß das CoInternet-System Benutzer durch aus mehreren Attributen bestehenden Profilen modelliert und das Resultat des Vergleichs von Profilen nicht eine skalare Größe, wie bei ausschließlich auf Bewertungsmustern basierenden Collaborative-Filtering-Algorithmen ist, sondern ein Vektor, dessen Dimension durch die Zahl der Attribute der Profile bestimmt wird.

### **Berechnung von Vorhersagen**

Die Berechnung von Prognosen zur zu erwartenden Relevanz eines Dokuments basiert im CoInternet-System auf zwei Säulen: Auf der einen Seite werden die Benutzerprofile in mehreren Dimensionen verglichen um auf diesem Wege automatisch die *Nearest-Neighbours* zu einem Benutzerprofil zu ermitteln. Diese Herangehensweise entspricht dem automatischen Collaborative Filtering. Auf der anderen Seite werden explizit von den Benutzern gemachte Angaben zum Vertrauen zu bzw. zur Bedeutung von anderen Nutzern zum kollaborativen Filtern herangezogen. Jeder Nutzer kann über das Modell des Web-of-Trusts sein persönliches Netzwerk relevanter Empfehler aufbauen. Dieser Aspekt des Empfehlungsprozesses ist als aktives Collaborative Filtering zu charakterisieren.

Korrespondierend zu den verschiedenen Attributen eines Profils existieren als Resultat des Vergleichs der Benutzerprofile mehrere Ähnlichkeitswerte. Diese werden zusammen mit den das Vertrauen zwischen Benutzern quantifizierenden Werten zur Berechnung von Vorhersagen durch gewichtete Summation der Bewertungen verwendet. Aus den einzelnen Vorhersagen wird gemäß eines vorgegebenen Verhältnisses ein gewichteter Durchschnitt gebildet. So läßt sich berücksichtigen, daß beispielsweise implizite Bewertungen in der einem Benutzer präsentierten Relevanzvorhersage eine geringere Bedeutung als explizite Wertungen haben sollten.

Die Berechnung von Vorhersagen soll zunächst anhand des in Abbildung 8.9 dargestellten Beispiels erläutert werden. Ziel ist es, eine Prognose  $pred_{a,u}$  für die vom aktiven Benutzer A zu erwartenden Bewertung der von den Benutzern B,C und D mit den Werten -2, 1 und 2 beurteilten Web-Seite, welche unter der URL  $U_u$  abrufbar ist, zu berechnen. Im Beispiel wird zur Vereinfachung lediglich ein Ähnlichkeitswert zusammen mit dem Vertrauenswert in die Berechnung einbezogen.

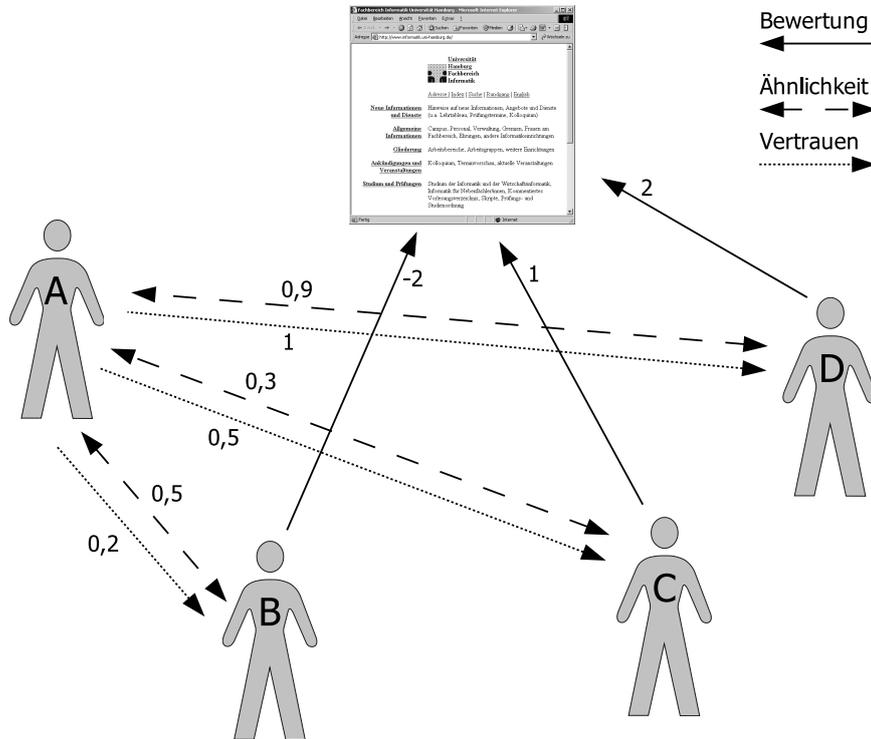


Abbildung 8.9.: Schema der Berechnung von Bewertungsvorhersagen

$$\begin{aligned}
 pred_{a,u} &= \left( 6 \cdot \frac{0,5 \cdot (-2) + 0,3 \cdot 1 + 0,9 \cdot 2}{0,5 + 0,3 + 0,9} + 4 \cdot \frac{0,2 \cdot (-2) + 0,5 \cdot 1 + 1 \cdot 2}{0,2 + 0,5 + 1} \right) \cdot \frac{1}{10} \\
 &\approx 0,89
 \end{aligned}$$

Die errechneten Prognosen werden im Verhältnis 6 zu 4 aufsummiert. Der Meinung des Benutzers D kommt aufgrund der Profilähnlichkeit und des von A ausgesprochenen Vertrauens eine besondere Bedeutung zu. Der Vorhersagewert liegt deshalb mit 0,89 höher als der Durchschnittswert von 0,33.

Für die formale Betrachtung der Berechnung der Vorhersagen gelten folgende Vereinbarungen:

$$\begin{aligned}
 Profiles &= \{P_1, P_2, \dots, P_a, \dots, P_m\} && \text{Menge der Profile} \\
 P_i &= (A, B, C) && \text{Profil mit den Attribut-Vektoren } A, B \text{ und } C
 \end{aligned}$$

$P_a \in Profiles$	Profil des aktiven Benutzers
$URIs = \{U_1, U_2, \dots, U_u, \dots, U_n\}$	Menge der URIs
$I_i \subseteq URIs$	Menge der im Profil $P_i$ bewerteten Objekte
$R_u = \{p   P_p \in Profiles \wedge U_u \in I_p\}$	Menge der Identifikatoren der Profile, in denen eine Bewertung für die URI $U_u$ enthalten ist
$v_{i,j}$	Wertung des Benutzers des Profils $P_i$ zur URL $U_j$
$sim_{A,i,j}$	Ähnlichkeit der Profile $P_i$ und $P_j$ durch Vergleich des Attributs $A$ (analog für die Attribute $B$ und $C$ )
$trust_{i,j}$	Vertrauen des Benutzers des Profils $P_i$ zum Benutzer des Profils $P_j$
$weights = (w_A, w_B, w_C, w_{Trust})$	Gewichtung der Attribute $A, B, C$ und des Vertrauens zwischen den Benutzern bei der Berechnung von Vorhersagen

Gesucht ist also die vorhergesagte Bewertung  $pred_{a,u}$  für das Profil des aktiven Benutzers  $P_a$  und die URI  $U_u$ . Zu diesem Zweck wird zunächst basierend auf dem Vergleich des Attributs  $A$  und analog für die Attribute  $B$  und  $C$  eine gesonderte Vorhersage bestimmt:

$$pred_{A,a,u} = \frac{\sum_{i \in R_u} sim_{A,a,i} \cdot v_{i,u}}{\sum_{i \in R_u} sim_{A,a,i}} \quad (8.3)$$

Ein weiterer Vorhersagewert ergibt sich aus der Einbeziehung des Web-of-Trusts:

$$pred_{Trust,a,u} = \frac{\sum_{i \in R_u} trust_{a,i} \cdot v_{i,u}}{\sum_{i \in R_u} trust_{a,i}} \quad (8.4)$$

Kann basierend auf einem Attribut keine Vorhersage berechnet werden, da im aktuellen Profil oder in allen Vergleichsprofilen keine Werte für dieses Attribut vorliegen, so ist das Gewicht des Attributs im Vektor  $weights$  auf 0 zu setzen und so sind die anderen Gewichte proportional zu erhöhen. Im letzten Schritt wird gemäß dem durch  $weights$  spezifizierten Verhältnis eine gewichtete Durchschnittsbildung vorgenommen, um die einzelnen Vorhersagen zu einem Wert zu aggregieren:

$$pred_{a,u} = \frac{w_A \cdot pred_{A,a,u} + w_B \cdot pred_{B,a,u} + w_C \cdot pred_{C,a,u} + w_{Trust} \cdot pred_{Trust,a,u}}{w_A + w_B + w_C + w_{Trust}} \quad (8.5)$$

Um die Relevanz einer Vorhersage einschätzen zu können, ist ein deren Genauigkeit beziffernder Wert wünschenswert. Zu diesem Zweck wird für das Attribut A und in gleicher Weise für die anderen Attribute der Durchschnitt der Ähnlichkeitswerte unter Berücksichtigung der Bedeutung des jeweiligen Profils bei der Berechnung des Vorhersagewertes bestimmt:

$$\overline{sim_{A,a,u}} = \frac{\sum_{i \in R_u} sim_{A,a,i}^2}{\sum_{i \in R_u} sim_{A,a,i}} \quad (8.6)$$

Die durchschnittlichen Ähnlichkeitswerte der verschiedenen Attribute werden zu einem Wert zusammengezogen:

$$\overline{sim_{a,u}} = w_A \cdot \overline{sim_{A,a,u}} + w_B \cdot \overline{sim_{B,a,u}} + w_C \cdot \overline{sim_{C,a,u}} \quad (8.7)$$

Analog zur Berechnung der mittleren Ähnlichkeit wird ein durchschnittlicher Vertrauenswert bestimmt:

$$\overline{trust_{a,u}} = \frac{\sum_{i \in R_u} trust_{a,i}^2}{\sum_{i \in R_u} trust_{a,i}} \quad (8.8)$$

Der Wert  $rel_{a,u}$  ist ein Maß dafür, mit welcher Sicherheit die für das Profil  $P_a$  des aktiven Benutzers prognostizierte Bewertung des unter der URL  $U_u$  abrufbaren Dokuments zutreffend ist.

$$rel_{a,u} = \frac{(w_A + w_B + w_C) \cdot \overline{sim_{a,u}} + w_{Trust} \cdot \overline{trust_{a,u}}}{w_A + w_B + w_C + w_{Trust}} \quad (8.9)$$

### Exkurs: Verteilungsaspekte der Berechnung von Vorhersagen

Der vorstehend beschriebene Prozeß zur Berechnung von Vorhersagen setzt zentral vorliegende Daten voraus. Es sind jedoch auch ähnlich dem in [Tve01] beschriebenen Ansatz verteilte Collaborative-Filtering-Algorithmen bis hin zur Ausführung von Collaborative Filtering in einem Peer-to-Peer-Netz denkbar, bei dem jeder Netzknote lediglich die eigenen Bewertungen und das eigene Profil vorhält. Eine von einem Netzknote ausgehende Anfrage nach der Relevanz eines Objektes wird einschließlich dem Identifikator des Objektes und dem eigenen Benutzerprofil, dieses besteht eventuell nur aus dem Vektor der eigenen Bewertungen, an alle anderen Netzknote geschickt. Diese antworten, sofern sie das fragliche Objekt bewertet haben, mit ihrer Bewertung und der Ähnlichkeit ihres Profils zu dem Profil des anfragenden Netzknotes. Sind alle Antworten eingetroffen,



Abbildung 8.10.: Präsentation einer Bewertungsprognose

dann ist die Aggregation der Bewertungen durchzuführen. Analog läßt sich auf Vertrauen basiertes aktives Collaborative Filtering realisieren [Ols98]. In diesem Fall werden Anfragen ausschließlich an vertraute Netzknotten gesandt. Nachteilig bei hochgradig verteilten Collaborative-Filtering-Algorithmen ist die hohe Zahl an zu versendenden Nachrichten und die durch Nicht-Erreichbarkeit von Netzknotten bedingte Unvollständigkeit der Datenbasis.

### Präsentation von Bewertungsvorhersagen

Bewertungsvorhersagen werden auf zweierlei Art und Weise präsentiert. Eine Vorhersage zur aktuelle Seite wird zusammen mit Informationen zum Prozeß der Aggregation von Bewertungen als Teil der Benutzungsschnittstelle des CoInternet-Clients in einem Fenster neben dem Browser präsentiert (siehe Abbildung 8.10). Die Vorhersage wird als Zahl in einem farbig hinterlegten Quadrat dargestellt. Die Färbung wird je nach Wert der Vorhersage von Grüntönen für positive Vorhersagen über Gelbtöne für neutrale Vorhersagen bis hin zu Rot für negative Werte variiert. Durch die Benutzung unterschiedlicher Farben sollen die Werte von Vorhersagen peripher wahrnehmbar werden. Die kleinere eingeklammerte Zahl entspricht dem arithmetischen Mittel der Bewertungen. Die anderen Elemente dieser Ansicht werden im nachfolgenden Unterabschnitt erläutert.

Eine Herausforderung bei der Navigation in Hypertexten ist es, eine Entscheidung zu treffen, welcher der in ein Dokument eingebetteten Hyperlinks hinsichtlich der persönlichen Ziele sinnvollerweise verfolgt werden sollte. Das Scone-Plugin HyperScout unterstützt Anwender bei der Auswahl von Links durch Einblendung von Popup-Fenstern mit Informationen zum Zieldokument (Abschnitt 2.5). HyperScout wurde dahingehend modifiziert, daß nun als zusätzlicher Hinweis die prognostizierte Bewertung des Zieldokuments eines Hyperlinks ausgegeben wird. Abbildung 8.11 zeigt ein um die Darstellung von Relevanzvorhersagen ergänztes HyperScout-Popup. Die Vorhersage wird in der letzten Zeile ausgegeben. Die farbliche Hinterlegung ist übereinstimmend zu der Präsentati-



**Abbildung 8.11.:** Einblendung der prognostizierten Bewertung des Ziel-Objekts eines Hyperlinks mittels eines modifizierten HyperScout-Popups

on von Bewertungsvorhersagen im Fenster des CoInternet-Clients. Durch die Beachtung der individuellen Relevanzprognosen können Benutzer unter Umständen den Aufruf von Dokumenten vermeiden, von denen sie nach kurzem Studium der Inhalte feststellen, daß diese entgegen den Erwartungen bei der Verfolgung des Links uninteressant sind. Die für das Web typische „Hub-and-Spoke“-Navigation [TG97] könnte somit einem zielgerichteten Vorgehen weichen.

Eine Alternative zu den Popups wäre die Modifikation der Seiten, so daß die Vorhersagen, visualisiert beispielsweise durch Icons, direkt neben den Links im Dokument gezeigt werden würden. Nachteilig bei dieser Art der Darstellung ist jedoch, daß das Layout der Seite beeinträchtigt wird. Desweiteren müssten bereits beim Laden des Dokuments Vorhersagen für alle verlinkten Dokumente abgefragt werden. Hingegen wird durch das modifizierte HyperScout-Plugin immer nur die Vorhersage zu dem Link abgefragt, über dem der Mauszeiger plazierte wurde.

### Transparenz des Empfehlungsprozesses

Die Qualität der vom CoInternet-System präsentierten Empfehlungen unterliegt großen Schwankungen. Einflußfaktoren sind die Zahl der für ein Dokument abgegebenen Bewertungen, deren Streuung, die Übereinstimmung des Profils eines Empfehlungen abfragenden Benutzers zu den Profilen der Bewerter und die Menge an in den Benutzerprofilen enthaltenen Daten. Wenn bei polarisierenden Themen auch die „Nearest Neighbours“ nicht einheitlich votiert haben, ist eine daraus abgeleitete neutrale Vorhersage sicherlich nicht optimal. Die eine Bewertungsvorhersage beachtende Person wird entweder dem

einen oder dem anderen Lager zuzuordnen sein – eine Entscheidung, die das Empfehlungssystem in derartigen Situationen nicht treffen kann.

Das aus der Orientierung an Empfehlungen resultierende Risiko kann je nach den Zielen, die ein Benutzer bei der Navigation im Web verfolgt, variieren. Der Prozeß der Berechnung von Empfehlungen sollte für die Anwender nachvollziehbar sein. Die Frage, warum eine Web-Seite als empfehlenswert und eine andere als uninteressant eingestuft wird, sollte durch die dem Benutzer präsentierten Angaben beantwortbar sein. Darüber hinaus sollen die gezeigten Informationen Aufschluß darüber geben, wie akkurat eine Vorhersage ist.

Das Entwurfsziel eines für die Anwender transparenten Empfehlungsprozesses verbindet sich mit dem Ziel, die Empfehlungen derart zu präsentieren, daß ein Bewußtsein dafür geschaffen wird, daß die Berechnung von Vorhersagen ein kollaborativer Prozeß ist, an dem die Mitglieder der CoInternet-Community durch die Abgabe von Bewertungen maßgeblichen Anteil haben. Die Herkunft von Empfehlungen soll bis hin zu einzelnen Bewertungen nachvollziehbar sein, so daß das Engagement einzelner Benutzer für die Gemeinschaft deutlich und damit der Aufbau von Reputation begünstigt wird.

Das zusammen mit der prognostizierten Wertung gezeigte Histogramm der Stimmenverteilung (siehe Abbildung 8.10) gibt einen ersten Überblick über die Verteilung der Meinungen vorheriger Besucher einer Seite. Betrachtet man zudem noch den Mittelwert der Bewertungen und dessen Abweichung von der individuellen Vorhersage, so sind Rückschlüsse auf die Verteilung der Stimmen der besonders ähnlichen bzw. vertrauten Benutzer möglich. Weitere Angaben sind die Zahl der Stimmen und die gewichteten Mittelwerte der Ähnlichkeit und des Vertrauens zu den Mitgliedern der CoInternet-Community, welche das aktuelle Dokument bewertet haben. Aus diesen Zahlen wurde der Wert für die Relevanz einer Vorhersage abgeleitet. Diese Kenngrößen werden gemäß den vorstehend in diesem Abschnitt aufgeführten Formeln 8.7-8.9 bestimmt.

Das Vertrauen in eine Bewertungsvorhersage ist eng mit dem Vertrauen in die Empfehler verbunden. Um die Herkunft von Empfehlungen nachvollziehbar zu machen, lassen sich alle zu einem Dokument abgegebenen Bewertungen wie in Abbildung 8.12 dargestellt einsehen. Die Bewertungen werden zusammen mit den Pseudonymen der Benutzer aufgelistet. Zusätzlich werden Angaben zur Ähnlichkeit zum Profil des jeweiligen Nutzers und zu dem dieser Person ausgesprochenen Vertrauen gemacht. Die Aktualität der Bewertungen kann anhand des ebenfalls ausgegebenen Datums eingeschätzt werden. Eine derartige Ansicht ist eine Form von Awareness-Information. Nicht die aktuelle Gegenwart anderer Mitglieder der Community wird präsentiert, sondern historische Informationen über deren Aktivitäten im Informationsraum. Die Schaffung von Gruppenbewußtsein könnte auf diesem Wege gefördert werden, wenn beispielsweise für einen Benutzer wahrnehmbar wird, daß es einen kleinen Kreis von Personen gibt, welche die Seiten, die er besonders mag, ebenfalls positiv bewertet und eventuell kommentiert haben. In diesem

**Beiträge zum Dokument:**

<http://testserver/news/meldung1.htm> ☹️ -1,2

**Bewertungen:**

Bewertung	Autor	Datum
0	Robert Ähnlichkeit: <input type="checkbox"/> Vertrauen: <input type="checkbox"/>	24.06.2003
-1	Boerni Ähnlichkeit: <input type="checkbox"/> Vertrauen: <input type="checkbox"/>	19.06.2003
-2	Stefan Ähnlichkeit: <input type="checkbox"/> Vertrauen: <input type="checkbox"/>	18.06.2003
0	thoralf Ähnlichkeit: <input type="checkbox"/> Vertrauen: <input type="checkbox"/>	18.06.2003
-2	Björn Ähnlichkeit: <input type="checkbox"/> Vertrauen: <input type="checkbox"/>	18.06.2003
-1	Lars Ähnlichkeit: <input type="checkbox"/> Vertrauen: <input type="checkbox"/>	17.06.2003
2	nika Ähnlichkeit: <input type="checkbox"/> Vertrauen: <input type="checkbox"/>	12.06.2003

Abbildung 8.12.: Auflistung aller zu einem Dokument abgegebenen Bewertungen

Zusammenhang wäre die Integration von Anwendungen zur direkten Kommunikation unter Umständen sinnvoll. Eine Erweiterung dieser Ansicht wäre die Kennzeichnung aus anderen Kontexten bekannter Identitäten. Zu diesem Zweck könnte ein auf dem Rechner des Benutzers laufendes Intermediary die UUIDs der im lokalen Identitätsmanager gespeicherten Foreign Identities mit denen der gelisteten Benutzer vergleichen und die Liste entsprechend anpassen.

Die Pseudonyme der Benutzer sind Hyperlinks zu einer Seite, auf der alle Beiträge der Person aufgelistet sind (siehe Abbildung 8.13). Anhand dieser Informationen kann ein Benutzer entscheiden, ob er eine Person in das Web-of-Trust seines aktuellen Profils aufnehmen möchte. Die neben den URLs eingefügten Icons symbolisieren die Relevanzvorhersage für das jeweilige Dokument.

**Beiträge des Benutzers Frank**

Ähnlichkeit:

Vertrauen:

[▶ Zu den vertrauten Personen hinzufügen](#)

**Bewertungen:**

URL	Bewertung	Datum
http://testserver/news/meldung17.htm ☹️ 1,2	-2	06.12.2003
http://testserver/news/meldung24.htm ☹️ 0,6	-2	06.12.2003
http://testserver/news/meldung20.htm ☹️ 0,3	1	06.12.2003
http://testserver/news/meldung19.htm ☹️ -0,6	-2	06.12.2003
http://testserver/news/meldung11.htm ☹️ -0,0	2	06.12.2003
http://testserver/news/meldung8.htm ☹️ 1,3	2	06.12.2003
http://testserver/news/meldung7.htm ☹️ 0,2	-2	06.12.2003

**Annotationen:**

URL	Überschrift	Datum
http://www.scone.de/ 😊 2,0	Tolle Anwendung des WBI	06.12.2003

Abbildung 8.13.: Ausgabe aller Bewertungen und Annotationen eines Benutzers

**8.2.6. Kombination eines Empfehlungssystems mit Suchmaschinen**

Die oftmals als unzureichend empfundene Leistung von Suchmaschinen bei der Recherche im Web ist u. a. auf fehlendes semantisches Verständnis der Dokumente zurückzuführen (Abschnitt 3.2). Doch selbst, wenn eine Suchmaschine in der Lage wäre, die Bedeutung indexierter Dokumente zu erfassen, so könnte sie aufgrund einer Suchanfrage nicht entscheiden, ob ein Benutzer bei der Suche nach beispielsweise dem Begriff „apple“ an Computern oder Äpfeln interessiert ist. Ferner könnte nicht festgestellt werden, ob jemand, der nach dem Namen eines Produktes sucht, an Händlern interessiert ist oder Test- und Erfahrungsberichte lesen möchte. Die Formulierung eines Informationsbedürfnisses als Suchanfrage ist folglich in manchen Situationen als unzureichend anzusehen, da der Kontext einer Suche unberücksichtigt bleibt. Weitere Defizite von Suchmaschinen sind darin zu sehen, daß bei der Zusammenstellung der Suchergebnisse die Qualität der Dokumente ebenso unberücksichtigt bleibt wie die Vertrauenswürdigkeit der Informationsanbieter. Beides sind Kriterien, die naturgemäß nicht einer automatischen Beurteilung zugänglich sind.

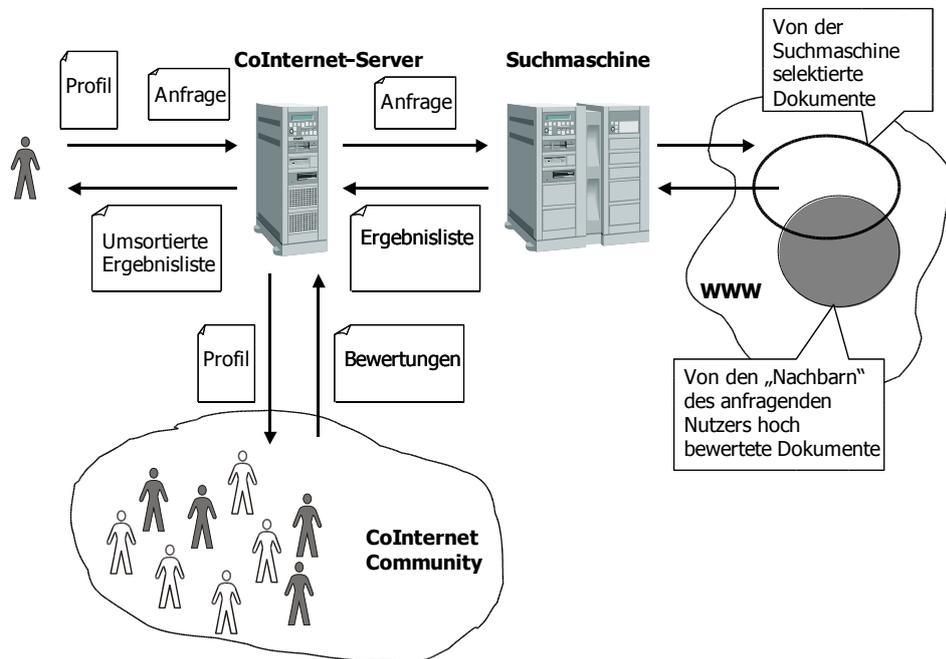
Der in dieser Arbeit gewählte Ansatz zur Verbesserung der Präzision von Suchergebnis-

sen ist es, durch Kombination des Empfehlungssystems mit einer oder mehreren Suchmaschinen einen leistungsfähigeren Suchdienst zu realisieren. Die Suchanfragen der Benutzer werden an eine Suchmaschine weitergeleitet. Die von der Suchmaschine erstellten Ergebnislisten werden anschließend gemäß der vorhergesagten Bewertung der einzelnen Einträge umsortiert. Das CoInternet-System bietet somit den Mehrwert, Social-Navigation-Informationen direkt in die dem Anwender präsentierte Benutzungsoberfläche eines Suchdienstes zu integrieren.

Abbildung 8.14 zeigt schematisch den Ablauf der Verarbeitung einer Suchanfrage durch das CoInternet-System unter Rückgriff auf die Funktionalität einer Suchmaschine. Suchmaschinen beinhalten einen Index der durch Robots aus dem Web eingelesenen Dokumente. Dieser Index ist über Suchanfragen abfragbar, deren Resultat eine Teilmenge der indexierten Web-Dokumente ist. In Benutzerprofilen sind die langfristigen Interessen der Benutzer modelliert. Zusammen mit den durch Angabe vertrauter Personen explizit gemachten Beziehungsnetzwerken dienen sie der Auffindung im Kontext eines Profils relevanter Empfehler. Durch Aggregation der von diesen abgegebenen Bewertungen selektiert das CoInternet-System ebenfalls eine Menge von Web-Dokumenten, die für den anfragenden Benutzer als potentiell relevant eingestuft werden. Eine entsprechend den Bewertungsvorhersagen umsortierte Ergebnisliste einer Suchmaschinenabfrage enthält somit an den vordersten Positionen Verweise zu Dokumenten, die einerseits den als Suchanfrage formulierten kurzfristigen Interessen entsprechen und andererseits von den basierend auf dem Benutzerprofil ausgewählten Empfehlern positiv beurteilt wurden.

Die über die Identitätsinfrastruktur übermittelten Profile bestimmen die Zugehörigkeit zu einer Gruppe von Empfehlern innerhalb der CoInternet-Community, die durch ihre Bewertungen einen Kontext für die Auswertung von Suchanfragen aufbaut. Durch Einbeziehung des Urteils anderer Benutzer in den Suchprozeß finden subjektive Kriterien wie Qualität, Interesse und Vertrauen bei der Auswahl von Informationen durch Suchmaschinen Beachtung. Die Fähigkeit menschlicher Benutzer, Dokumente auf einer semantischen Ebene zu verarbeiten, wird für die Suchmaschinenteknologie erschlossen.

Abbildung 8.15 zeigt den Ausschnitt eines Screenshots der Ergebnisseite einer an das CoInternet-System abgesetzten Suchanfrage nach dem Begriff „Robots“. An den ersten beiden Positionen der Liste werden Dokumente zum Thema Web-Robots gelistet, da aufgrund des Profils des aktiven Benutzers eine Affinität zu an Internettechnologien interessierten Benutzern festgestellt werden konnte. Diese haben zuvor Seiten zum Thema Web-Robots positiv bewertet und bei früheren Suchanfragen präsentierte Dokumente zum Thema Robotik negativ beurteilt. Das Ergebnis einer weiteren Suchanfrage, bei der ebenfalls nach „Robots“ gesucht wurde, ist aus Abbildung 8.16 ersichtlich. In diesem Fall wird an erster Stelle auf eine Seite zum Thema Robotik verwiesen, denn die Suchanfrage wurde im Kontext eines Benutzerprofils ausgewertet, das den Kreis der Empfehler auf an diesem Thema interessierte Personen einschränkt.



**Abbildung 8.14.:** Schema der Kombination des CoInternet-Systems mit einer Suchmaschine

Der durch Kombination des Empfehlungssystems mit einer Suchmaschine realisierte Suchdienst ist lernfähig. Bewertungen von Web-Seiten können aus dieser Sichtweise mit den *Relevance-Feedback-Techniken* des Information Retrieval verglichen werden, denn sie präzisieren die Suche nach ähnlichen Benutzern und damit mittelbar das Ranking zukünftiger Suchergebnisse. Ebenso lernt das System mit jeder gestellten Suchanfrage dazu, denn Suchbegriffe sind Teil der Benutzerprofile. Mit jeder Suchanfrage werden indirekt über den Vergleich der Profile auch vorherige Suchanfragen berücksichtigt.

Nachteilig am beschriebenen Verfahren ist, daß mit jeder Suchanfrage eine Vielzahl von Relevanzvorhersagen bestimmt werden müssen, was sich negativ auf die Performanz der Suche mit einem derartigen System auswirken kann. Die in einem Empfehlungssystem für Web-Seiten vorliegenden Bewertungen werden immer nur einen kleinen Teil der Dokumente des Webs abdecken können. Folglich ist es nicht möglich, bei jeder Suchanfrage Empfehlungen anderer Benutzer zu berücksichtigen.

Suche nach: **Robots** Ergebnisse **1-10** von 2320000

The Web **Robots** Pages 😊 **1,6**

The Web **Robots** Pages The Web **Robots** Pages. Web **Robots** are programs that traverse the Web automatically. Some ... **Robots**. The Web **Robots** Pages.  
<http://www.robotstxt.org/wc/robots.html>

**Robots** Exclusion 😊 **1,1**

The Web **Robots** Pages **Robots** Exclusion. Sometimes people find they have ... The **Robots** Exclusion Protocol. The **Robots** Exclusion Protocol is ...  
<http://www.robotstxt.org/wc/exclusion.html>

Real **Robots** On The Web 😞 **-0,7**

Real **Robots** on the Web: ARITI- Augmented Reality Interface for Telerobotic applications via Internet - control a 4-DOF robot using a remote computer ...  
[http://ranier.hq.nasa.gov/telerobotics\\_page/realrobots.html](http://ranier.hq.nasa.gov/telerobotics_page/realrobots.html)

**robots.net** - Robot news and Robotics Info 😞 **-1,3**

Welcome to **robots.net**, the place to read the latest news on personal and industrial robotics, robot competitions, and other cool stuff. ... Recently added **robots** ...  
 Beschreibung: A news and discussion site for those interested in **robots** and robotics. Incl  
<http://robots.net/>

We Are **Robots**

Beschreibung: An animated series documenting the lives of **robots**.  
<http://www.wearerobots.com/>

**Abbildung 8.15.:** Ergebnissseite der Suche nach „Robots“ unter Einbeziehung des Profils eines an Internet-Themen interessierten Benutzers

### 8.2.7. Benutzungsschnittstelle

Zentrales Element der Benutzungsschnittstelle des CoInternet-Systems ist das in Abbildung 8.17 gezeigte Fenster, welches neben dem Web-Browser auf dem Desktop platziert wird.

Um die Praxistauglichkeit des Prototyps zu erhöhen, wurde berücksichtigt, daß mehrere Browserfenster gleichzeitig geöffnet werden können. Im oberen Bereich des Fensters wird der Titel und die URL des Dokuments angezeigt, auf das sich die Darstellung im Client-Fenster bezieht. Sind mehrere Browserfenster geöffnet, so kann mittels der Combo-Box, in der die Titel der angezeigten Dokumente eingetragen sind, zwischen den Fenstern umgeschaltet werden. Das ausgewählte Fenster wird automatisch in den Vordergrund

Suche nach: **Robots** Ergebnisse **1-10** von 2320000

**robots.net** - Robot news and Robotics Info 😊 **1,3**

Welcome to **robots.net**, the place to read the latest news on personal and industrial robotics, robot competitions, and other cool stuff. ... Recently added **robots** ...  
Beschreibung: A news and discussion site for those interested in **robots** and robotics. Inclu  
<http://robots.net/>

Real **Robots** On The Web 😊 **0,7**

Real **Robots** on the Web: ARITI- Augmented Reality Interface for Telerobotic applications via Internet - control a 4-DOF robot using a remote computer ...  
[http://ranier.hq.nasa.gov/telerobotics\\_page/realrobots.html](http://ranier.hq.nasa.gov/telerobotics_page/realrobots.html)

**Robots** Exclusion 😊 **0,3**

The Web **Robots** Pages **Robots** Exclusion. Sometimes people find they have ... The **Robots** Exclusion Protocol. The **Robots** Exclusion Protocol is ...  
<http://www.robotstxt.org/wc/exclusion.html>

The Web **Robots** Pages 😞 **-1,3**

The Web **Robots** Pages The Web **Robots** Pages. Web **Robots** are programs that traverse the Web automatically. Some ... **Robots**. The Web **Robots** Pages.  
<http://www.robotstxt.org/wc/robots.html>

We Are **Robots**

Beschreibung: An animated series documenting the lives of **robots**.  
<http://www.wearerobots.com/>

**Abbildung 8.16.:** In diesem Fall werden die Ergebnisse der Suche nach „Robots“ in einer anderen Reihenfolge als in Abbildung 8.15 präsentiert, da das Profil, welches den Kontext der Suchanfrage bildet, eine Affinität zu an Robotik interessierten Personen dokumentiert.

der Anzeige gestellt. Die Schwächen dieser Lösung werden bei vielen Browserfenstern offenkundig. Anzustreben wäre eine Integration des Clients in den Browser als Side- oder Toolbar. Aus Gründen der technischen Realisierbarkeit wurde jedoch einem separaten Fenster der Vorzug gegeben. Die Bewertung einzelner Frames ist ebenfalls möglich. Jedes Frame wird vom System als ein eigenständiges Fenster verwaltet. Jedoch sind mittels Framesets realisierte Sites für ein Web-Seiten-Empfehlungssystem ungünstig, denn die Zuordnung der Relevanzvorhersagen zu einzelnen Frames ist nur schwer darstellbar.

Die im Block „Einstellungen“ gruppierten Elemente gestatten die Auswahl der aktuellen Identität und eines dieser zugeordneten Profils. Über die neben den Combo-Boxen befindlichen Icons können Identitäten bzw. Profile angelegt, geändert und gelöscht werden.

Durch Deaktivierung der mit einer Glühlampe gekennzeichneten Schaltfläche wird der CoInternet-Dienst deaktiviert, so daß sich der Anwender durch das Web bewegen kann, ohne daß im Hintergrund der CoInternet-Server kontaktiert wird.

Suchanfragen, die dann unter Einbeziehung der Empfehlungen beantwortet werden, können über das Eingabefeld „Suchen“ abgesetzt werden. Die Ergebnisse werden im gerade aktiven Browserfenster präsentiert.

In der Gruppe „Aktionen“ können wie in Abschnitt 8.2.2 beschrieben explizite Bewertungen abgesendet werden. Wird eine Seite geladen, die bereits bewertet wurde, so wird der entsprechende Radio-Button selektiert. Durch einen Mausklick auf „Beiträge erstellen/verwalten“ gelangt man zu einer Ansicht, in der alle bisherigen Beiträge eingesehen, verändert und gelöscht werden können und darüber hinaus Annotationen und kommentierte Bewertungen zum aktuellen Dokument verfasst werden können.

Bewertungsvorhersagen werden in der Gruppe „Bewertungen“ wie in Abschnitt 8.2.5 beschrieben dargestellt. Die zu einer Seite vorliegenden Annotationen werden im unteren Bereich des Client-Fensters ausgegeben.

Weitere Elemente der Benutzungsoberfläche sind über den Links in den Seiten eingebettete Popups (Abbildung 8.11) und einige als dynamische Web-Seiten realisierte Ansichten, darunter die Liste aller Bewertungen einer Seite (Abbildung 8.12), die Auflistung aller Beiträge eines Mitglieds der CoInternet-Community (Abbildung 8.13) und die bei der Beantwortung von Suchanfragen generierten Trefferlisten (Abbildungen 8.15 und 8.16).

## 8.3. Konstruktion des CoInternet-Systems

In diesem Abschnitt wird die Konstruktion des CoInternet-Systems, dessen Konzeption im vorherigen Abschnitt skizziert wurde, beschrieben. Die Systemarchitektur wird in Abschnitt 8.3.1 dargestellt. Das System wurde vollständig in Java implementiert, was einerseits durch die technischen Randbedingungen des Scone-Frameworks und des oneC-Identitätsmanagers vorgegeben war und andererseits durch eine angestrebte Plattformunabhängigkeit zu begründen ist. Details zur Implementation der einzelnen Systemkomponenten werden in den Abschnitten 8.3.2 und 8.3.3 gegeben. Die Darstellung der Implementation erfolgt auf einer relativ abstrakten Ebene. Gegebenenfalls kann der auf der beiliegenden CD-ROM befindliche Source-Code eingesehen werden. Das Protokoll zur Kommunikation mit dem CoInternet-Server ist Gegenstand von Abschnitt 8.3.4.

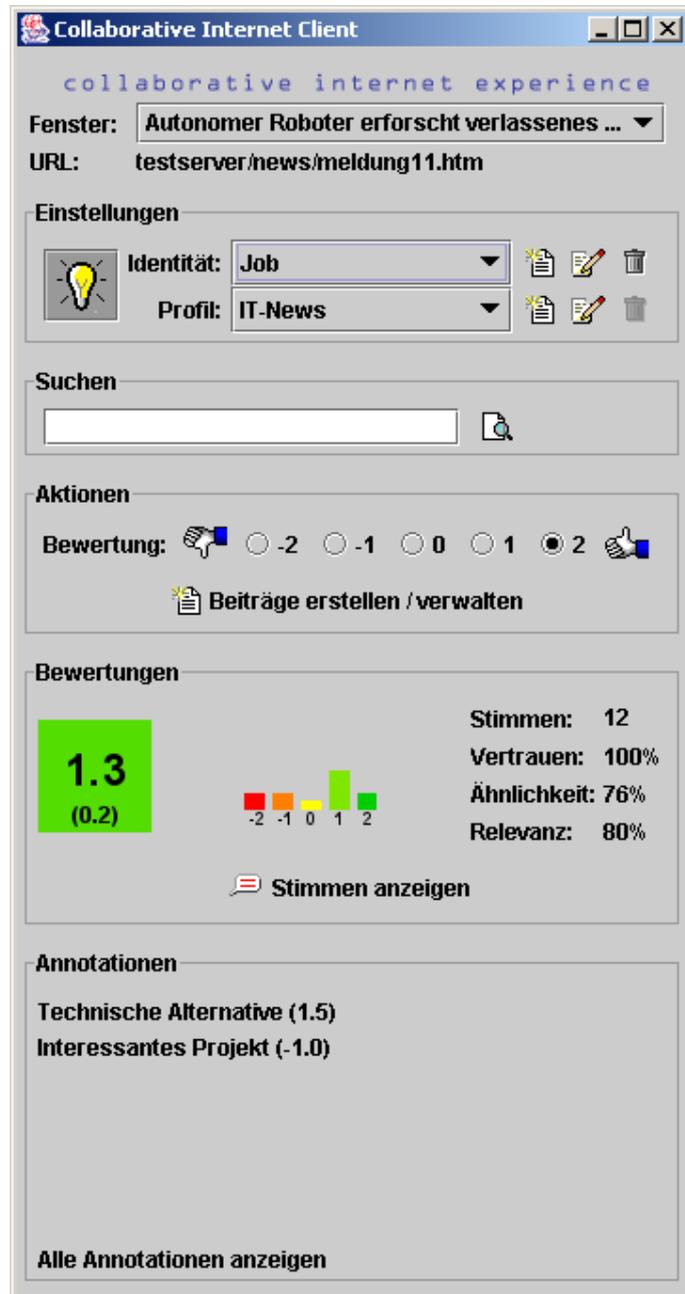


Abbildung 8.17.: Screenshot des CoInternet-Clients

### 8.3.1. Systemarchitektur

Aus dem beschriebenen Konzept eines Empfehlungs- und Annotationssystem lassen sich vier grundlegende logische Funktionseinheiten bestimmen, die zur Realisierung des Systems benötigt werden, und bei deren Realisierung zwischen mehreren Alternativen abzuwägen ist. Diese sind jedoch nicht unabhängig voneinander zu betrachten, da Entscheidungen bezüglich der Umsetzung einer Funktionseinheit Implikationen auf die mögliche Realisierung anderer Einheiten haben kann. Zusätzlich sind Entscheidungen bezüglich der räumlichen Verteilung der Einheiten zu treffen.

**Datenhaltung** Die wesentlichen Daten des CoInternet-Systems sind Bewertungen, Annotationen und Benutzerprofile. Sie müssen persistent abgelegt und über Schnittstellen zugreifbar gemacht werden. Ferner werden Funktionen zur Verwaltung der Datenbestände benötigt.

Eine Vorgehensweise ist es, sämtliche Daten zentral in einem Server zu halten. Eine derartige Lösung ist jedoch mit dem Risiko verbunden, daß es einen zentralen Ausfallpunkt gibt und Performanz-Engpässe auftreten können. Eine Alternative ist ein Cluster von mit verteilten oder replizierten Daten arbeitenden Servern. Aus Sicht des Benutzers ist es wünschenswert, jederzeit vollständige Kontrolle über die Profile zu behalten. Dieser Anforderung würde eine Lösung gerecht werden, bei der die Profile durch einen auf dem Rechner des Benutzers laufenden Identitätsmanager verwaltet werden und nur auf Anfrage Attribute eines Profils nach außen kommuniziert werden. Diese Variante bietet überdies den Vorteil, daß auf im Rahmen der Benutzung des CoInternet-Dienstes aufgebaute Profile auch von anderen Anwendungen zugegriffen werden könnten.

**Berechnung von Bewertungsvorhersagen** Die Berechnung von Vorhersagen benötigt den Zugriff auf Benutzerprofile und Bewertungen. Folglich ist es sinnvoll, beide Funktionseinheiten auf einem Rechnerknoten zu vereinen bzw. eine direkte Verbindung zwischen diesen Funktionseinheiten vorzusehen. Die Berechnung von Vorhersagen läßt sich zentral oder verteilt auf mehrere Rechner ausführen. Die verteilte Ausführung, bei der ein Server jeweils einen Teil der Clients bedient, bringt gegenüber einer nur auf einem zentralen Server basierenden Architektur Vorteile bezüglich der Skalierbarkeit und Ausfallsicherheit. Es ist wie in Abbildung 8.18 ersichtlich zwischen vier Varianten zu unterscheiden:

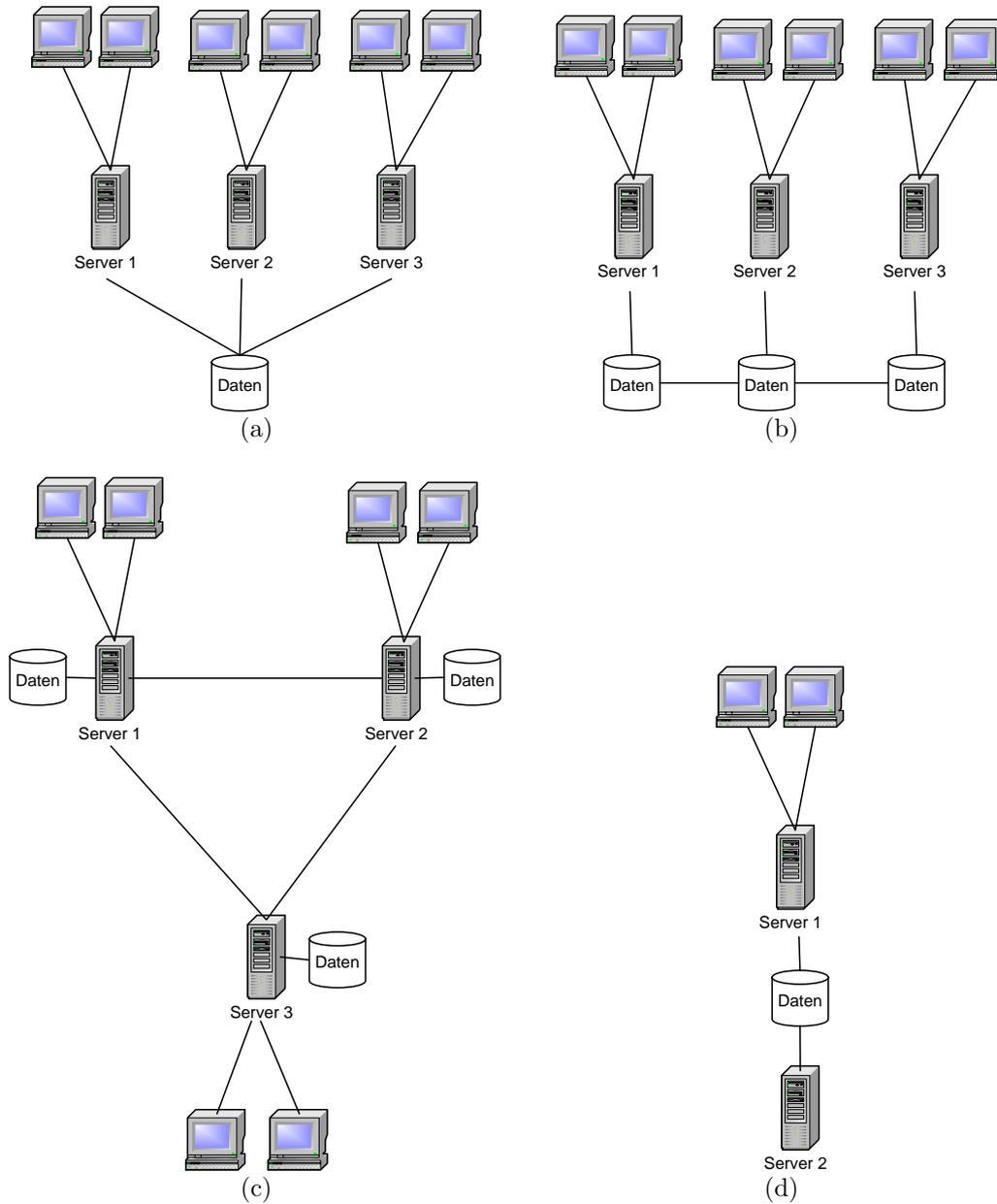
- Mehrere Rechner, die den vollständigen Prozeß der Berechnung von Bewertungsvorhersagen basierend auf einem gemeinsamen Datenbestand ausführen (Abbildung 8.18(a)).

- Vorhaltung lokaler Replikate des Datenbestandes auf allen Vorhersagen berechnenden Maschinen (Abbildung 8.18(b)).
- Berechnung von Teilergebnissen durch die einzelnen Rechner beruhend auf dem lokal gespeicherten Anteil des Datenbestandes. Diese Teilergebnisse werden zwischen den Rechnern ausgetauscht und zu Vorhersagen zusammengefaßt (Abbildung 8.18(c)). In diese Kategorie ist das in Abschnitt 8.2.5 skizzierte Konzept der Realisierung von kollaborativem Filtern auf einem Peer-to-Peer-Netz einzuordnen.
- Loslösung von asynchron zur Beantwortung von Anfragen ausführbaren Funktionen des Vorhersageprozesses und deren Ausführung auf getrennten Rechnern. Die Aktualisierung zwischengespeicherter Werte zur Ähnlichkeit der Profile ist ein Aufgabe, die sich aus der Beantwortung von Anfragen ausgliedern läßt (Abbildung 8.18(d)).

**Anbindung an die Web-Architektur** Wesentliche Aufgabe einer Funktionseinheit zur Anbindung des Empfehlungs- und Annotationssystems an die Web-Architektur ist die Synchronisation der Benutzungsschnittstelle mit den Bewegungen im Netz. Es ist notwendig, die Aktionen des Benutzers im Netz zu überwachen, so daß für das aktuelle Dokument Bewertungsvorhersagen und die zugeordneten Annotationen abgefragt werden können. Zusätzlich werden diese Informationen gebraucht, damit das aktuelle Dokument bewertet werden kann und aus der Beobachtung des Navigationsverhaltens implizite Bewertungen abgeleitet bzw. Benutzerprofile aufgebaut werden können.

Ein Option der Herstellung einer Anbindung an die Web-Architektur ist die Modifikation des Browsers bzw. die Verwendung der durch den Browser bereitgestellten APIs. Eine Alternative dazu ist die Benutzung eines Proxy-Servers. Der Dienst würde in diesem Fall als Intermediary realisiert werden, der in den Datenstrom zwischen Web-Client und Web-Server geschaltet wird. Es ist zwischen einem zentralen Proxy-Server, über den die Web-Kommunikation sämtlicher Benutzer abgewickelt wird oder mehreren, auf den Rechnern der Benutzer laufenden Proxies abzuwägen.

Eine weitere Variante der Umsetzung einer Anbindung an die Web-Architektur ist die Integration direkt in Web-Sites oder die Verwendung eines einem Web-Server vorgeschalteten Proxy-Servers. Diese Option besitzt gegenüber den vorstehend beschriebenen Alternativen den Nachteil, daß sich die gebotene Social-Navigation-Funktionalität auf die jeweilige Web-Site beschränkt. Letztendlich wäre es auch bei der Realisierung eines Web-Seiten Empfehlungs- und Annotationssystems in Erwägung zu ziehen, keine Anbindung an das Web, im Sinne der Überwachung der Navigationsaktivitäten und der fortwährenden Weiterleitung dieser Informationen an andere Systemkomponenten, vorzunehmen, sondern stattdessen das System als eine eigenständige Web-Site zu realisieren.



**Abbildung 8.18.:** Architekturvarianten bei der Realisierung einer Funktionseinheit zur Bewertungsvorhersage.

(a) Gemeinsame Daten, (b) Replizierte Daten, (c) Verteilte Daten, (d) Ausgliederung von asynchron zur Beantwortung von Anfragen ausführbaren Funktionen

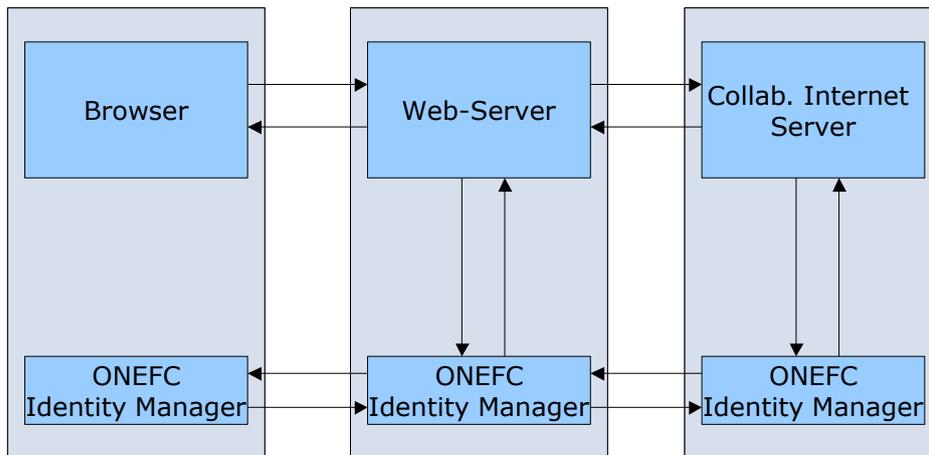
Diese könnte es beispielsweise Benutzern gestatten, die in einem Katalog gespeicherten Links zu Internet-Ressourcen zu kommentieren und zu bewerten.

**Benutzungsoberfläche** Zur Eingabe von Bewertungen und Annotationen sowie zur Anzeige von Bewertungsvorhersagen und Annotationen ist eine Benutzungsoberfläche zu entwickeln. Diese sollte darüberhinaus Funktionen zur Verwaltung der eigenen Beiträge, zum Absenden von Suchanfragen und zum Editieren der Profile bereitstellen.

Die Benutzungsoberfläche des System könnte ganz oder teilweise als web-basierte Anwendung durch dynamisch generierte HTML-Seiten realisiert werden. Eine weitere Variante ist die Modifikation der abgefragten Web-Seiten, wobei eventuell auch Java-Applets eingebettet werden könnten. Ferner kann die Benutzungsschnittstelle in den Browser integriert werden, etwa als Side- oder Toolbar oder eine separate Client-Anwendung realisiert werden.

Ein realistischer Zwischenschritt bei der Etablierung von Social Navigation im Web, ist die Integration entsprechender Erweiterungen in Web-Sites. Beispiele für Internet-Präsenzen, die Funktionen zum sozialen Navigieren anbieten, wurde in Kapitel 4 aufgeführt. Eine Systemarchitektur für das CoInternet-Systems wäre die Bereitstellung der Funktionen des Systems als ein Service, der in dynamische Web-Sites eingebunden werden kann (siehe Abbildung 8.19). Um für die Inhalte einer Web-Site ein Empfehlungs- und Annotationsystem bereitzustellen, müßte lediglich der Code zur Generierung der Web-Dokumente um Aufrufe von Funktionen des CoInternet-Servers erweitert werden. Ferner sind Elemente zur Abgabe von Bewertungen, eventuell zum Verfassen von Annotationen und zur Darstellung von Empfehlungen in das Layout der Web-Site zu integrieren. Interessante Synergieeffekte können sich ergeben, wenn mehrere Sites auf den CoInternet-Dienst zugreifen. Beispielsweise könnten Bewertungen von Meldungen zu neuen Produkten auf dem IT-Sektor zur Berechnung von Empfehlungen auf einer Shopping-Site herangezogen werden. Kooperationen zwischen auf den CoInternet-Dienst zugreifenden Sites werden durch die Möglichkeit, mittels der oneC-Identitätsmanager Single-Sign-On-Anmeldeprozeduren zu realisieren, begünstigt.

Der Ablauf des Ladevorgangs einer Seite wird nachstehend anhand von Abbildung 8.19 beschrieben. Der Web-Browser sendet einen Request an den Web-Server. Durch den Web-Server werden, vermittelt durch den Identitätsmanager, Daten zur Identifizierung des Nutzers und eventuell noch weitere für die Web-Site oder das CoInternet-System relevante Attribute der Identität des Benutzers abgefragt. Dieser Schritt ist notwendig, da die Wiedererkennung von Nutzern Grundvoraussetzung für den Aufbau von Benutzerprofilen und für personalisierte Dienste ist. Bei der Generierung der an den Client zurückzusendenden Seite wird gegebenenfalls der CoInternet-Server kontaktiert, um Vorhersagen zur Relevanz der aktuellen Seite und der verlinkten Seiten abzufragen. Zusätzlich werden die mit dem Dokument assoziierten Annotationen vom CoInternet-Server



**Abbildung 8.19.:** Das CoInternet-System könnte als ein in Web-Sites einzubindender Dienst realisiert werden.

geladen. Diese Elemente werden während der dynamischen Erstellung von Seiten in das Layout integriert. Ein Austausch von Benutzerprofilen zwischen dem Web-Server und dem CoInternet-Server findet auf Ebene der Identitätsmanager statt.

In dieser Arbeit wurde für die Verwirklichung des Empfehlungs- und Annotationssystems eine Client-Server-Architektur gewählt, wie sie schematisch in Abbildung 8.20 abgebildet ist.

Sämtliche Bewertungen, Annotationen und Benutzerprofile werden in der Datenbank eines zentralen Servers abgelegt. Basierend auf diesen Daten berechnet der Server Empfehlungen und generiert Ansichten als HTML-Dokumente, wie beispielsweise die Liste aller zu einer URL vorliegender Beiträge (siehe Abschnitt 8.2.5). Es ist fraglich, ob es möglich ist, mit einer derartigen Lösung eine für den produktiven Betrieb des Systems hinreichende Zahl von Nutzern zu bedienen. Gegebenenfalls wäre für den praktischen Einsatz des Systems eine Verteilung der Server-Funktionalität auf mehrere Rechner vorzunehmen.

Der Client wurde als Java-Anwendung mit einer grafischen Benutzungsoberfläche realisiert. Die Anbindung an die Web-Architektur erfolgt über den auf dem Rechner des Benutzers laufenden Proxy-Server des Scone-Frameworks. Die Identitäten des Benutzers und die Benutzerprofile werden durch einen Identitätsmanager verwaltet und persistent abgelegt.

Anzustreben wäre sicherlich eine Integration des Clients in den Browser, denn auf diesem Wege ließe sich eine Benutzungsoberfläche realisieren, die insbesondere bei mehreren

geöffneten Browserfenstern ergonomische Vorteile hätte. Technisch hätte eine Integration in den Browser den Vorteil, daß es zuverlässigere und weitreichendere Möglichkeiten zur Überwachung der Aktivitäten des Nutzes als bei Verwendung eines Proxy-Servers in Verbindung mit der UserTracking-Komponente des Scone-Frameworks gäbe. Die Kombination aus einer Java-Applikation und einem lokalen Proxy-Server wurde vorgezogen, da das Scone-Framework die derartige Umsetzung einer Erweiterung des Webs umfassend unterstützt. Auf diesem Wege konnte mit weitaus weniger Aufwand, als es bei der Modifikation eines Browsers oder der Implementation eines Browser-Plugins der Fall gewesen wäre, eine prototypische Implementierung vorgenommen werden. Die Realisierung als Browser-Plugin hätte die Festlegung auf eine browser-spezifische API bedingt und somit den Kreis der mit dem CoInternet-Dienst verwendbaren Browser stark eingeschränkt.

Die vollständige Realisierung des Clients als Applet oder durch in die Seiten eingebettete Bedienelemente erscheint nicht sinnvoll, da das Layout der Web-Seiten beeinträchtigt werden würde und bei nicht standardkonformen Seiten keine fehlerfreie Einbettung zu gewährleisten ist. Java-Applets laufen in einer *Sandbox* ab, die den Zugriff auf lokale Daten unterbindet und somit den Betrieb eines lokalen Identitätsmanagers verhindert.

Ein lokaler Proxy wurde gewählt, da dieser auf der gleichen virtuellen Maschine wie der Client ablaufen kann und somit eine direkte Kommunikation zwischen diesen Komponenten möglich wird.

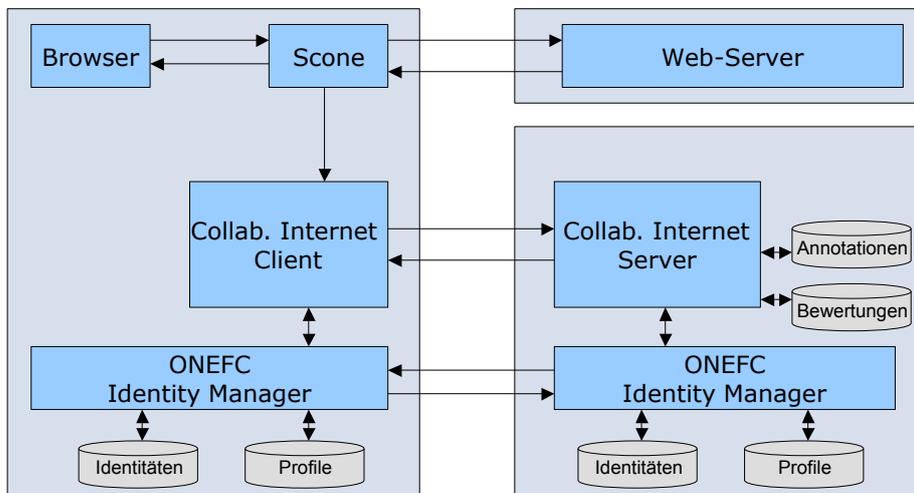
Fordert ein Benutzer mittels des Web-Browsers ein Dokument eines Web-Servers an, so wird die Kommunikation zwischen Browser und Web-Server über den Scone-Proxy-Server geleitet (siehe Abbildung 8.20). Jeder Ladevorgang wird an den CoInternet-Client kommuniziert. Dieser fragt vom lokalen Identitätsmanager die Identifikatoren der aktuell gewählten Identität und des zugeordneten Profils ab. Anschließend wird eine Anfrage nach einer Relevanzvorhersage zum geladenen Dokument und nach den mit dem Dokument assoziierten Annotationen an den CoInternet-Server zusammen mit den Identifikatoren von Identität und Profil übermittelt. Sofern serverseitig aktuelle Kopien des Benutzerprofils vorliegen, wird eine Vorhersage berechnet und an den Client zurückgesendet. Anderenfalls wird zunächst durch Kommunikation der Identitätsmanager das Benutzerprofil aktualisiert.

### 8.3.2. Server-Komponenten

Die Kernfunktionalität des Servers wird als Web-Service zur Nutzung durch beliebige Clients bereitgestellt. Konkret bedeutet dies, die öffentlichen Methoden des Servers lassen sich über auf dem *Simple Object Access Protocol (SOAP)*<sup>4</sup> aufsetzende Remote

---

<sup>4</sup>[www.w3.org/TR/SOAP](http://www.w3.org/TR/SOAP)



**Abbildung 8.20.:** Verwirklicht wurde das CoInternet-System als eine Client-Server-Architektur.

Procedure Calls aufrufen. SOAP-Nachrichten werden über HTTP als Transportprotokoll ausgetauscht. Realisiert wurde der Web Service unter Zuhilfenahme von Apache Axis<sup>5</sup>. Axis bietet einerseits eine Umgebung zum Hosting von Web-Services und andererseits ein SOAP-Toolkit zum Erstellen von und zum Zugriff auf Web-Services. Teile der Funktionalität des Servers sind zudem über Web-Seiten zugreifbar, welche durch Servlets dynamisch generiert werden. Sowohl Axis als auch die erstellten Servlets benötigen zum Betrieb einen Servlet Container. Hier kam die Referenzimplementierung eines Servlet Containers, der Tomcat Application Server<sup>6</sup>, zum Einsatz. Bei der Realisierung des CoInternet-Servers konnte somit auf einen leistungsfähigen, multi-threaded arbeitenden HTTP-Server aufgesetzt werden.

In der nachfolgenden Darstellung werden jeweils die zentralen Klassen der in Java realisierten Implementation aufgeführt. Diese realisieren mit verschiedenen Hilfsklassen jeweils eine Systemkomponente.

<sup>5</sup>[ws.apache.org/axis](http://ws.apache.org/axis)

<sup>6</sup>[jakarta.apache.org/tomcat](http://jakarta.apache.org/tomcat)

## **CoInternetServer**

Im Mittelpunkt der Implementation des CoInternet-Servers steht eine Instanz der Klasse `CoInternetServer`. Aufgabe des `CoInternetServer` ist es, die Funktionalität des Servers gegenüber den Clients zu kapseln und über eine Schnittstelle verfügbar zu machen. Von Clients empfangene Anfragen werden unter Rückgriff auf die anderen Komponenten des Servers beantwortet. Angaben zu den vom `CoInternetServer` für den entfernten Zugriff bereitgestellten Operationen werden im Rahmen der Ausführungen über das Protokoll der Kommunikation zwischen Client und Server in Abschnitt 8.3.4 gemacht.

## **Web-Services Schnittstelle**

Apache Axis bietet eine auf einem beliebigen Servlet Container aufsetzende Umgebung zum Hosting von Web Services. Prinzipiell können die öffentlichen Methoden jeder Klasse mittels Axis für den entfernten Aufruf über das SOAP-Protokoll verfügbar gemacht werden. Lediglich die Parameter und Rückgabewerte müssen, sofern es sich dabei um Objekte handelt, den Java-Bean-Konventionen entsprechen. Dies ist notwendig, um eine automatische Serialisierung in eine XML-Repräsentation durchführbar zu machen. Das Axis-Servlet interpretiert die per HTTP übertragenen SOAP Nachrichten. Gegebenenfalls werden, übereinstimmend mit der XML-Darstellung der Parameter, Objekte instanziiert. Axis führt den Aufruf der in der SOAP-Nachricht benannten Methode durch. Das Resultat wird in einer SOAP-Nachricht verpackt an den Client übermittelt.

Mittels des Axis SOAP-Toolkits wurde eine in der *Web Services Description Language (WSDL)*<sup>7</sup> verfasste Beschreibung der Schnittstellen der Klasse `CoInternetServer` generiert. Dieses Dokument dient der automatischen Generierung von Code zum Zugriff auf diesen Dienst. Darüber hinaus wurde durch das Toolkit ein Deployment-Deskriptor erstellt, der benötigt wird, um den `CoInternetServer` als einen über Axis zugreifbaren Web-Service zu registrieren.

Für den Betrieb des CoInternet-Servers wird Axis zusammen mit einem Tomcat Application Server betrieben. Während der Initialisierung des Servers erzeugt Axis genau eine Instanz des `CoInternetServer`, deren Methoden bei der Verarbeitung ankommender Requests aufgerufen werden.

## **ProfileLoader**

Identitäten und Profile der das CoInternet-System benutzenden Personen werden in der Datenbank des Servers abgelegt. Bestandteil der an den Dienst gesendeten Requests

---

<sup>7</sup><http://www.w3c.org/TR/wsdl.html>

sind UUIDs, welche die Identität und das gewählte Profil des anfragenden Benutzers referenzieren. Aufgabe des `ProfileLoader` ist es, bei der Beantwortung von Anfragen nach Relevanzprognosen zunächst zu überprüfen, ob mit der Identität und dem Profil korrespondierende Einträge in der Datenbank existieren. Ist dies der Fall, so ist darüber hinaus die Aktualität der Attribute des Profils zu kontrollieren. Dazu werden die im Request übermittelten Zeitstempel der letzten Aktualisierung der jeweiligen Attribute mit den in der Datenbank abgelegten Zeitstempeln abgeglichen. Sollte es notwendig sein, so wird über den `OnefcAdapter` eine Anfrage an den serverseitig laufenden Identitätsmanager formuliert, der dann vom clientseitigen Identitätsmanager die benötigten Profilinformationen anfordert, so daß die Datenbank des Servers aktualisiert werden kann.

#### **OnefcAdapter**

Ähnlich dem `OnefcAdapter` des CoInternet-Clients kapselt der serverseitige `OnefcAdapter` den Zugriff auf die Identitätsinfrastruktur. Der `OnefcAdapter` beinhaltet Funktionen, um über die Identitätsinfrastruktur Identitätsdaten der Clients zu laden. Die vom Identitätsmanager übermittelte Identität eines Benutzers ist wie in Abschnitt 8.3.3 beschrieben, auf das interne Modell des CoInternet-Systems abzubilden. Insbesondere sind die als XML-Dokumente serialisierten Attribute zu deserialisieren.

#### **DBAdapter**

Alle persistenten Daten des Servers werden in einem relationalen Datenbankschema abgelegt. Gerade hinsichtlich der Berechnung von Bewertungsvorhersagen ist die Verwendung einer Datenbank gegenüber anderen Persistenztechniken zu bevorzugen, da ein effizienter Zugriff auf die Datenbestände erforderlich ist. Es wird das MySQL-Datenbanksystem<sup>8</sup> verwendet, wobei InnoDB-Tabellen<sup>9</sup> zum Einsatz kommen. MySQL zeichnet sich durch eine gute Leistungsfähigkeit insbesondere bei lesenden Zugriffen aus. Durch den Einsatz des Tabellentyps InnoDB konnte ein Datenbankschema implementiert werden, das die referentielle Integrität der Daten gewährleistet.

Der `DBAdapter` stellt die Anbindung des Servers an die Datenbank her. Er bietet eine an den Erfordernissen des CoInternet-Servers ausgerichtete Funktionsbibliothek zum Zugriff auf die Datenbank.

---

<sup>8</sup>[www.mysql.com](http://www.mysql.com)

<sup>9</sup>[www.innodb.com](http://www.innodb.com)

## PredictionEngine

Die **PredictionEngine** berechnet wie in Abschnitt 8.2.5 beschrieben individuelle Bewertungsvorhersagen. Der verfolgte Ansatz ist es, Collaborative-Filtering-Techniken anzuwenden, jedoch ein über einfache Bewertungsmuster hinausgehendes Benutzermodell zu verwenden. Die der Berechnung zugrunde liegenden Benutzerprofile bestehen aus verschiedenen Attributen (Abschnitt 8.2.1), welche für den Vergleich von Benutzerprofilen durch Vektoren von Eigenschaften repräsentiert werden. Eine Festlegung auf die bei der Berechnung einzubeziehenden Attribute sollte ebenso vermieden werden, wie eine Festlegung der zum Vergleich von Profilattributen verwendeten Algorithmen. Vielmehr galt es, eine Umgebung für die Entwicklung und Erprobung von Vorhersagealgorithmen zu schaffen.

Das Modulkonzept der **PredictionEngine** basiert auf der Verwendung einer Objekt-Fabrik [GHJ95]. Die **PredictionEngineFactory** instanziiert beim Startvorgang des CoInternet-Servers die **PredictionEngine** und alle in der Konfigurationsdatei des Servers benannten Module, welche das Interface **PredictionModule** implementieren müssen (siehe Abbildung 8.21). Diese werden gemäß der Konfigurationsdatei konfiguriert und bei der **PredictionEngine** registriert. Durch diese Vorgehensweise können Module unabhängig entwickelt und in das compilierte System integriert werden. Für verschiedene Testszenarien läßt sich die Zusammenstellung der in den Vorhersageprozeß einbezogenen Module durch einfaches Editieren der Konfigurationsdatei verändern.

Die Klasse **PredictionEngine** realisiert den in Abschnitt 8.2.5 dargestellten Ablauf der Berechnung von Bewertungsvorhersagen. Dabei werden durch die registrierten Module jeweils getrennt Vorhersagen errechnet und entsprechend der in der Konfigurationsdatei vorgegebenen Gewichtung der Module aufsummiert. Sollte es einem Modul aufgrund für einen Vergleich der Profile unzureichender Daten im ausgewerteten Profilattribut nicht möglich sein, eine Prognose zu bestimmen, so wird das Modul bei der Bestimmung des Endergebnisses übergangen. Der Anteil der verbliebenen Module erhöht sich folglich proportional. Abbildung 8.22 zeigt schematisch den Ablauf der Bestimmung individueller Bewertungsprognosen durch die **PredictionEngine**.

Das Interface **PredictionModule** definiert eine Schnittstelle, welche Module, die in den Prozeß der Vorhersage von Bewertungen eingeklinkt werden sollen, zu implementieren haben (Abbildung 8.21). Bei der Berechnung einer Vorhersage wird jeweils die Methode **predict()** der registrierten Module aufgerufen. Parameter sind der in der Datenbank verwendete Identifikator des Profils des aktiven Benutzers, der Identifikator der URL des Dokuments, für das eine Vorhersage ermittelt werden soll, und die Liste der für die fragliche URL abgegebenen Bewertungen. Die Einträge dieser Liste sind Tupel aus Profilidentifikatoren und dem Wert der Bewertung.

Die Klassenhierarchie der realisierten Module ist aus Abbildung 8.21 ersichtlich. Die Vor-

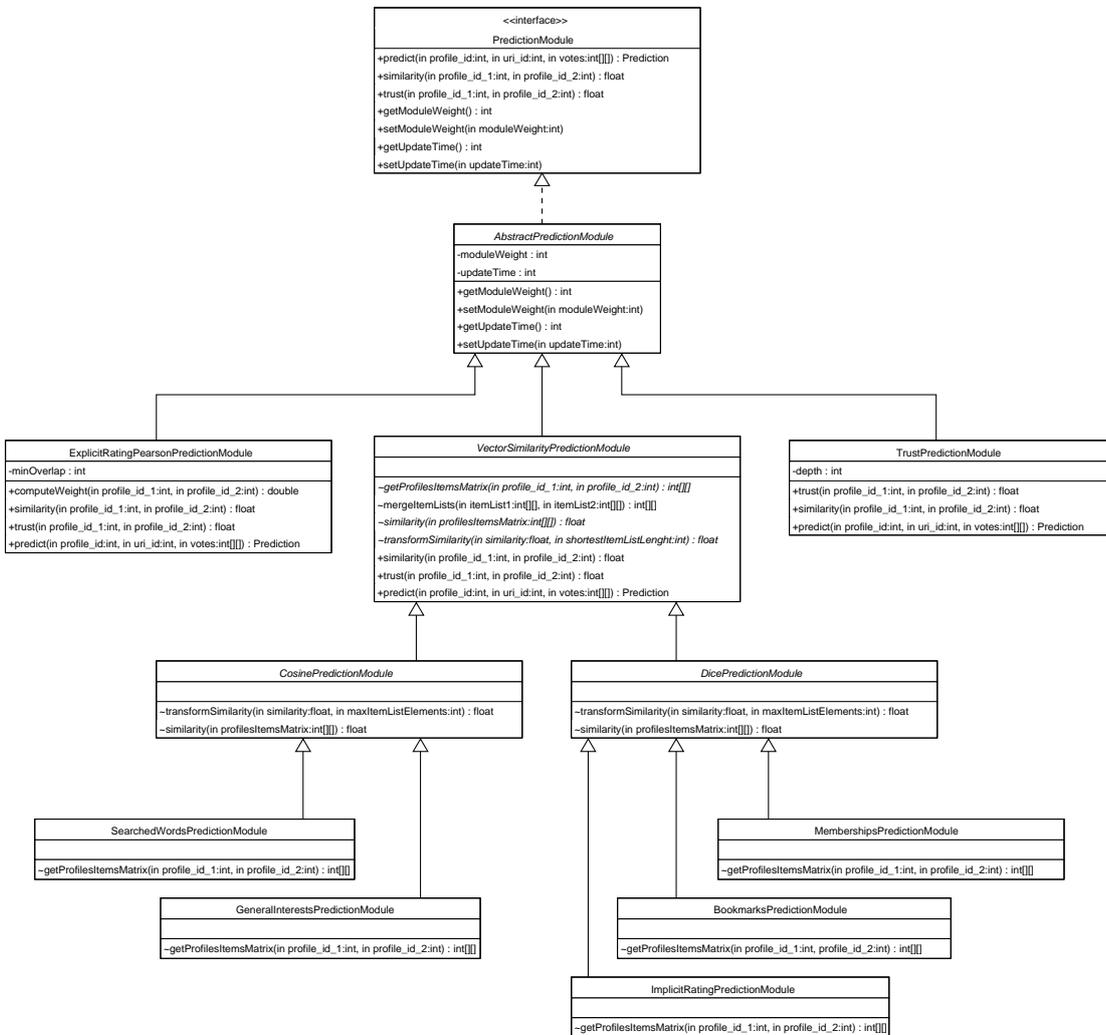
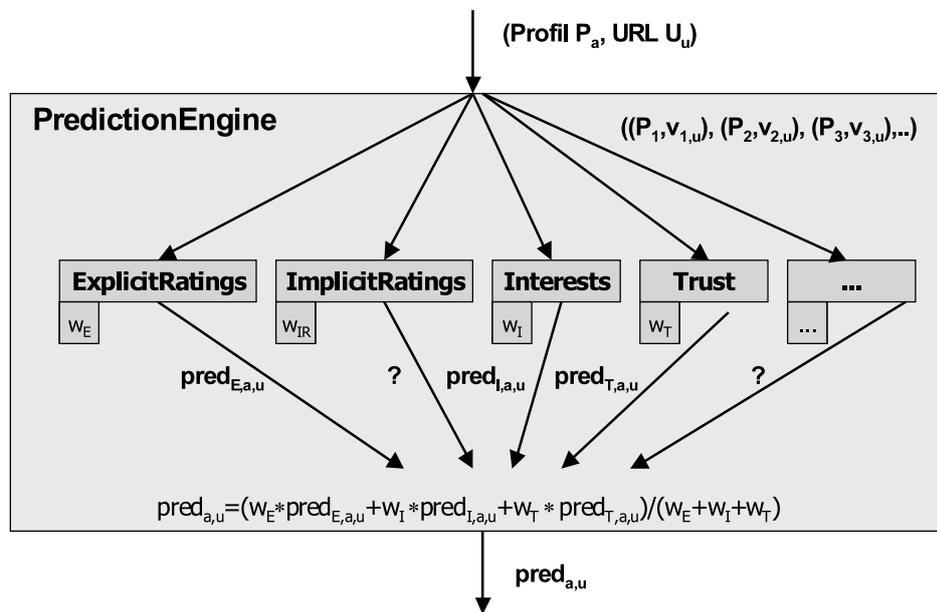


Abbildung 8.21.: Klassendiagramm der entwickelten Module zur Bestimmung von Bewertungsprognosen



**Abbildung 8.22.:** Schema der Berechnung von Bewertungsvorhersagen unter Einbeziehung mehrerer Module durch die PredictionEngine

hersage von Bewertungen basierend auf der Auffindung ähnlicher Benutzer durch Vergleich der von ihnen abgegebene Bewertungen wird durch das **ExplicitRatingModule** realisiert. Implementiert wurde der in Abschnitt 5.2.1 erläuterte Pearson-r-Algorithmus. Collaborative Filtering auf Grundlage der von den Benutzern durch Angabe vertrauter Personen ausgedrückten Beziehungen wird durch das **TrustPredictionModule** ausgeführt.

Den übrigen Attributen der Benutzerprofile ist gemein, daß sie sich wie in Abschnitt 8.2.5 beschrieben als Eigenschaftsvektoren repräsentieren lassen. Bei der Realisierung der Module zum kollaborativen Filtern basierend auf den Profilattributen Interesse, Suchbegriffe, implizite Bewertungen, Bookmarks und Mitgliedschaften wurde auf das Template-Pattern [GHJ95] zurückgegriffen. Die Methode `predict()` der abstrakten Klasse `VectorSimilarityPredictionModule` realisiert den Collaborative-Filtering-Algorithmus, wobei die Methode `similarity()` zum Vergleich der Benutzerprofile durch Unterklassen zu implementieren ist. Der Methode `similarity()` ist als Parameter ein zweidimensionales Array zu übergeben, welches der aus den beiden zu vergleichenden Eigenschaftsvektoren gebildeten Matrix entspricht (siehe Abbildung 8.8 auf Seite 135). Der Aufbau dieser Matrix ist ebenfalls durch eine Unterklasse zu implementieren, so daß in der Klasse `VectorSimilarityPredictionModule`

keine Festlegung auf das auszuwertende Attribut erfolgen mußte. Der Algorithmus zum Vergleich der Profile wird in den von `VectorSimilarityPredictionModule` ererbenden abstrakten Klassen `DicePredictionModule` und `CosinePredictionModule` implementiert. Diese verwenden zum Vergleich der Vektoren den Dice-<sup>10</sup> bzw. Cosinus-Koeffizient<sup>11</sup>. Die zum kollaborativen Filtern basierend auf bestimmten Attributen eines Benutzerprofils eingesetzten Module, beispielsweise das `GeneralInterestsPredictionModule` und das `ImplicitRatingPredictionModule`, implementieren die Methode `getProfilesItemMatrix()` und legen somit das auszuwertende Profilattribut fest. Die Festlegung auf einen zum Vergleich der Eigenschaftsvektoren anzuwendenden Algorithmus erfolgt durch die Wahl der Oberklasse. In der prototypischen Implementierung wird zum Vergleich impliziter Bewertungen der Dice-Koeffizient verwendet. Soll stattdessen der Cosinus-Koeffizient angewandt werden, so ist lediglich die Klasse `ImplicitRatingPredictionModule` von der Klasse `CosinePredictionModule` abzuleiten. Die Realisierung von Modulen zur Bewertungsvorhersage auf Grundlage der beschriebenen Anwendung des Template-Pattern bietet den Vorteil, daß einerseits leicht neue Module für weitere Profilattribute verwirklicht werden können und andererseits der Vergleichsalgorithmus problemlos änderbar ist.

Die Berechnung der Ähnlichkeiten zwischen Benutzerprofilen durch die bei der `PredictionEngine` registrierten Module erfolgt, sobald den Server eine Anfrage nach einer Vorhersage erreicht und die `PredictionEngine` aufgerufen wird. Die Beantwortung von Anfragen ist zeitkritisch, denn das Empfehlungssystem wird nur dann als nützlich wahrgenommen werden, wenn die Vorhersagen ohne größere Verzögerung bei der Navigation im Web angezeigt werden. Gleichzeitig ist die Bestimmung der Vorhersagen durch separate Betrachtung verschiedener Profilattribute ein aufwendiger Vorgang. Um die Performanz des Systems zu steigern, wurde durch die Klasse `PredictionEngineCache` ein zweistufiges Caching-Konzept für die Ähnlichkeit von Attributen der Benutzerprofile und für das aus dem Web-of-Trust abgeleitete Vertrauen zwischen den Benutzern realisiert. Eine begrenzte Zahl von Einträgen wird in einem Hashtable im Arbeitsspeicher gehalten. Die aus dem Hashtable verdrängten Einträge werden mit einem Zeitstempel versehen in die Datenbank geschrieben. Jedes Modul der `PredictionEngine` kann einen Zeitraum festlegen, in dem die von dem Modul berechneten Werte Gültigkeit besitzen. Der Ansatz, Ähnlichkeiten unmittelbar bei der Beantwortung von Anfragen zu bestimmen, aber diese Werte zur späteren Verwendung in einem Cache abzulegen, wurde gegenüber einer zu einem definierten Zeitpunkt ausgeführten Berechnung der Ähnlichkeiten zwischen allen Profilen bevorzugt. Der Aufwand für die Aktualisierung der Werte würde in diesem Fall quadratisch zur Zahl der Profile steigen, wobei viele Werte berechnet werden würden, die möglicherweise niemals benötigt werden.

---

<sup>10</sup>Formel 8.1 auf Seite 135

<sup>11</sup>Formel 8.2 auf Seite 135

### Web-Interface

Teile der Benutzungsschnittstelle des CoInternet-Systems wurden als dynamische Web-Seiten realisiert (Abschnitt 8.2.7). Zu diesem Zweck wurden verschiedene *Servlets* entwickelt, welche die zur Generierung der Seiten notwendige Programmlogik enthalten und die Erstellung der HTML-Ausgabe an *Java Server Pages (JSPs)* delegieren.

### GoogleSearcher

In Abschnitt 8.2.6 wurde das Konzept beschrieben, wie durch Kombination des Empfehlungssystems mit einer Suchmaschine ein Suchdienst bereitgestellt werden kann, der eine individuelle Anpassung der Suchergebnisse entsprechend dem Profil des suchenden Benutzers vornimmt. Als Teil des CoInternet-Systems wurde aufbauend auf der Suchmaschine „Google“<sup>12</sup> ein derartiger Suchdienst verwirklicht. Die Suchanfragen werden von einem Servlet entgegengenommen, welches die Darstellung der Trefferlisten übernimmt. Die UUIDs von Identität und Profil des Benutzers werden in der URL übertragen. Die Bearbeitung der Suchanfragen erfolgt durch die Klasse *GoogleSearcher*. Über die Google SOAP-API<sup>13</sup> werden die Resultate der Suchanfrage ermittelt. Anschließend wird für jeden Treffer durch die *PredictionEngine* eine Bewertungsvorhersage bestimmt und eine Umsortierung der Ergebnisliste entsprechend den prognostizierten Bewertungen vorgenommen. Abbildung 8.23 zeigt den Ablauf der Beantwortung einer Suchanfrage durch den CoInternet-Server.

### 8.3.3. Client-Komponenten

In diesem Abschnitt werden die wichtigsten Komponenten aufgeführt, welche die in Java implementierte Client-Anwendung des CoInternet-Systems ausmachen.

### CoInternetPlugin

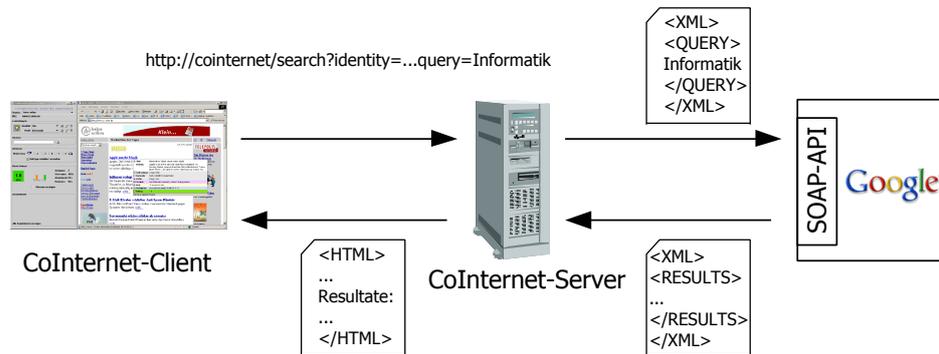
Die Verfolgung der Aktivitäten der Benutzer im Web erfolgt über das Scone-Plugin *CoInternetPlugin*. Gemäß der in [BM99] vorgestellten Terminologie für Komponenten von als Intermediary realisierten Anwendungen handelt es sich bei dem *CoInternetPlugin* um einen *Monitor*.

Plugins, die sich über ein Observer-Muster [GHJ95] beim *AccessCache* der *AccessTracking*-Komponente des Scone-Frameworks registriert haben, werden über den Beginn und Ende des Ladevorgangs eines Dokuments sowie über das Verlassen einer Seite

---

<sup>12</sup>[www.google.com](http://www.google.com)

<sup>13</sup>[www.google.com/apis](http://www.google.com/apis)



**Abbildung 8.23.:** Realisierung der Kombination eines Empfehlungssystems mit einer Suchmaschine unter Einbeziehung der über die Google SOAP-API verfügbar gemachten Dienste der Suchmaschine Google

benachrichtigt. Derartige Ereignisse werden durch Objekte der Klasse `AccessEvent` repräsentiert. Diese geben Aufschluß über die URL des Dokuments, den Fensternamen und beim Verlassen einer Seite durch Navigation zu einer weiteren Seite oder Schließen des Browserfensters auch über die Verweildauer. `AccessEvents` werden an den `CoInternetClient`, der Kernkomponente der Client-Installation, und den zur Absendung impliziter Bewertungen eingesetzten `ImplicitRatingAgent` weitergeleitet.

`AccessEvents` informieren über das Verlassen eines Dokuments, jedoch ist dabei nicht ersichtlich, ob das Fenster geschlossen wurde oder lediglich ein neues Dokument im gleichen Fenster geladen wird. Da im CoInternet-System die Wahl der Identität und des Profils an das jeweilige Fenster gebunden ist, es also möglich ist, parallel in verschiedenen Browserfenstern mit verschiedenen Identitäten aufzutreten, ist es nicht ausreichend, bei jedem Verlassen einer Seite zunächst das zugeordnete `Frame`-Objekt aus der Liste der geöffneten Fenster zu entfernen und gegebenenfalls mit der nächsten Notifikation über ein `AccessEvent` wieder einzufügen. Das `CoInternetPlugin` stellt bei jedem durch das Verlassen eines Dokuments ausgelösten `AccessEvent` das zugeordnete `Frame` in die Liste der Fenster, von denen angenommen wird, daß sie geschlossen wurden. Wird innerhalb eines einstellbaren Zeitraums, beispielsweise nach drei Sekunden, kein weiteres `AccessEvent` für ein Fenster dieser Liste empfangen, so wird der `CoInternetClient` über das Schließen des Fensters benachrichtigt, anderenfalls wird das Fenster aus dieser Liste entfernt.

## HyperScoutPlugin

Um die Prognose der zu erwartenden Relevanz des Zielobjektes eines Hyperlinks als ein Popup-Fenster darzustellen, welches angezeigt wird, sobald der Mauszeiger über dem Link plaziert wird, wurde das Scone-Plugin HyperScout (siehe Abschnitt 2.5) modifiziert. Die auf Seite 141 befindliche Abbildung 8.11 zeigt den Screenshot eines HyperScout-Popups. Die eingeblendeten Vorhersagen werden individuell für das im jeweiligen Fenster gültige Profil errechnet.

Der in die Web-Seiten integrierte JavaScript Code und das HyperScout-Applet wurden dahingehend erweitert, daß der Fenstername über die Socket-Verbindung zum `LinkDataGenerator` übermittelt wird. Der `LinkDataGenerator` ist die in das Scone-Framework eingebundene Komponente des HyperScout-Systems, welche die in den Popups anzuzeigenden Informationen zusammenstellt. Der `LinkDataGenerator` kontaktiert, vermittelt durch den `RatingServerProxy`, den CoInternet-Server um eine Vorhersage abzufragen. Beim Aufruf des `RatingServerProxy` werden als Parameter der Fenstername und die URL des verlinkten Objektes übergeben. Die gewählten Profile und Annotationen werden durch den `CoInternetClient` jeweils mit dem durch Scone vergebenen eindeutigen Fenstername assoziiert. Somit ist es möglich, bei Nennung des Fenstername die zugeordnete Identität und das Profil zu bestimmen. Der vom CoInternet-Server berechnete Wert wird an das Applet weitergeleitet und im Popup-Fenster angezeigt.

## CoInternetClient

Die Klasse `CoInternetClient` repräsentiert die zentrale clientseitige Komponente des CoInternet-Systems. Sie wurde als Singleton [GHJ95] realisiert, so daß sichergestellt ist, daß auf einer Java Virtual Maschine immer nur eine Instanz existiert, die durch Aufruf der statischen Methode `getInstance()` referenziert werden kann. Der `CoInternetClient` verwaltet Identitäten, Profile und eine Liste der Browserfenster. Als zentrale Komponente beinhaltet der `CoInternetClient` Assoziationen zu allen anderen Systemkomponenten. Um eine Koordination der Systemkomponenten zu gewährleisten und dennoch eine möglichst lose Koppelung zwischen den Komponenten beizubehalten, wurde ein auf dem Observer-Pattern [GHJ95] basierender Benachrichtigungsmechanismus implementiert. Objekte können sich für die Benachrichtigung über Ereignisse, wie das Laden eines Dokuments, das Öffnen oder Schließen eines neuen Browserfensters oder das Eintreffen einer vom CoInternet-Server abgefragten Bewertungsvorhersage registrieren. Auf diesem Wege wird u. a. die mit den Navigationsaktivitäten des Benutzers synchronisierte Aktualisierung der Benutzungsoberfläche vorgenommen.

### **CoInternetClientGui**

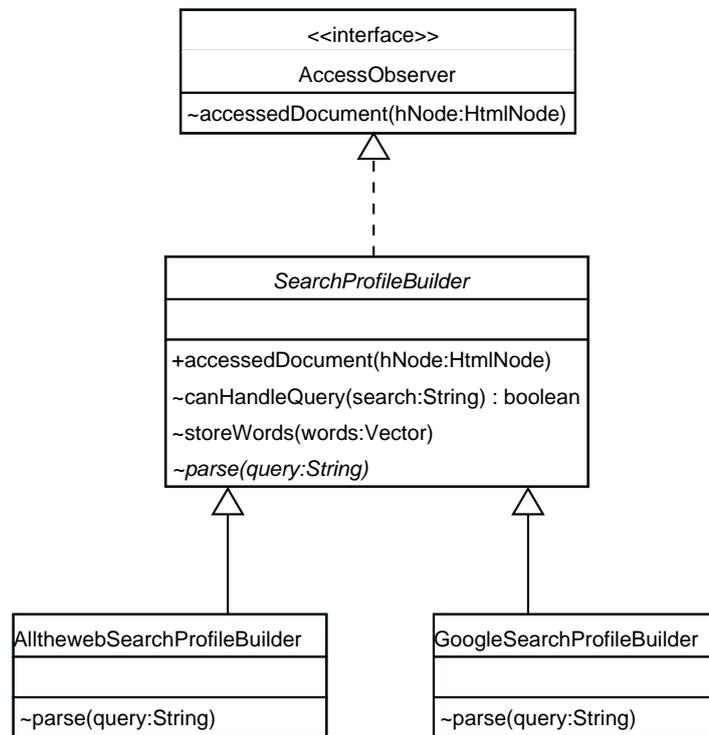
Mittels der Klasse `CoInternetClientGui` wurde das Hauptelement der Benutzungsschnittstelle des CoInternet-Systems, ein neben dem Browser auf dem Desktop anzuordnendes Fenster, realisiert (siehe Abbildung 8.17 auf Seite 150). Zur Gestaltung der Benutzungsoberfläche der Anwendung wurden ausschließlich Swing-Komponenten verwendet. Die Elemente der Benutzungsoberfläche wurden auf Panels gruppiert, die in separaten Klassen implementiert wurden. Die einzelnen Panels registrieren sich jeweils für die relevanten Ereignisse beim `CoInternetClient` und rufen als Reaktion auf Benutzereingaben Methoden des `CoInternetClients` oder des `RatingServerProxy` auf, die den Systemzustand entsprechend ändern bzw. Aufrufe des Servers vornehmen. Wesentlich für die Funktion der Benutzungsoberfläche ist die Fähigkeit, über Scone den Browser steuern zu können. So kann als Reaktion auf eine Benutzeraktion eine vom CoInternet-Server generierte Ansicht im Browserfenster dargestellt werden.

### **ProfileBuilder**

Aufgabe des `ProfileBuilder` ist der Aufbau und die Modifikation von Benutzerprofilen. Profile lassen sich interaktiv durch die mittels der Klasse `ProfileBuilderGui` realisierte Bedienoberfläche editieren. Die an Suchmaschinen gestellten Anfragen werden in die Benutzerprofile aufgenommen (Abschnitt 8.2.1). `SearchProfileBuilder` werden beim `CoInternetClient` für die Benachrichtigung über Zugriffe auf Dokumente registriert. Sie analysieren die URL des geladenen Dokuments und tragen, sofern es sich um eine Suchmaschinenabfrage handelt, die gesuchten Begriffe in das Profil ein. Wie das in Abbildung 8.24 abgebildete Klassendiagramm zeigt, wird die Auswertung der an unterschiedliche Suchmaschinen gestellten Anfragen jeweils von spezifischen Unterklassen von `SearchProfileBuilder` übernommen. Die Klasse `SearchProfilebuilder` fungiert dabei als Template [GHJ95]. Der komplette Vorgang der Verarbeitung von Suchmaschinenabfragen wird von dieser Klasse implementiert, wobei das suchmaschinen-spezifische Parsen des Query-Strings der URL einer Trefferliste als abstrakte Methode ausgegliedert wurde und von Unterklassen zu implementieren ist.

### **ImplicitRatingAgent**

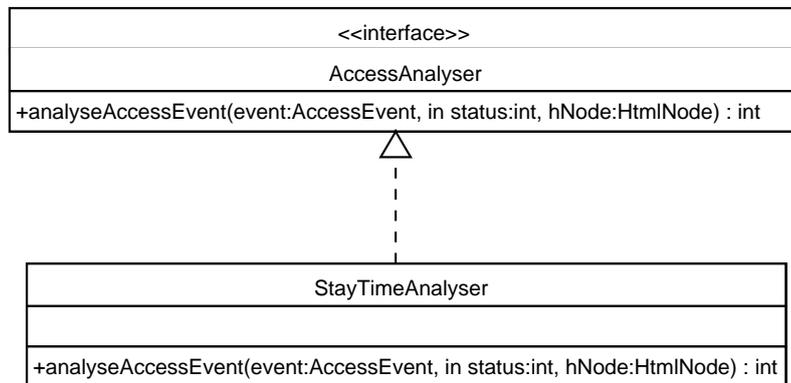
Der `ImplicitRatingAgent` analysiert das Navigationsverhalten des Benutzers und leitet daraus automatisch implizite Bewertungen ab, die über den `RatingServerProxy` an den CoInternet-Server geschickt werden. Um eine Festlegung auf einen konkreten Algorithmus zur Bestimmung impliziter Bewertungen zu vermeiden und anhand des Prototyps derartige Algorithmen entwickeln und evaluieren zu können, wurde für diesbezügliche



**Abbildung 8.24.:** Klassendiagramm der in die Auswertung von Suchmaschinenabfragen involvierten Klassen

Funktionen ein Modulkonzept verwirklicht. Ähnlich einer Objekt-Factory [GHJ95] instanziiert der `ImplicitRatingAgent` die in der Konfigurationsdatei des Clients benannten Klassen, welche das Interface `AccessAnalyser` (siehe Abbildung 8.25) implementieren müssen. Somit können neue Funktionen zur Bestimmung impliziter Bewertungen in das kompilierte System integriert werden. Der `ImplicitRatingAgent` wird direkt vom `CoInternetPlugin` über Zugriffe auf Dokumente benachrichtigt. Im Falle einer Benachrichtigung wird sequentiell die Methode `analyseAccessEvent()` aller `AccessAnalyser` aufgerufen. Als Parameter werden das `AccessEvent`, eine Konstante, die angibt, ob das Event durch den Beginn eines Ladevorgangs, das Ende eines Ladevorgangs oder das Verlassen einer Seite ausgelöst wurde, und das die geladene oder verlassene Web-Seite repräsentierende `HtmlNode`-Objekt übergeben. Die `AccessAnalyser` antworten entweder mit einer Bewertung oder einer Konstante, die ausdrückt, daß keine Bewertung abgeleitet werden konnte.

Die Kombination des `CoInternet`-Systems mit dem `Scone`-Framework eröffnet vielver-



**Abbildung 8.25.:** Klassen zum Ableiten impliziter Bewertungen müssen das Interface `AccessAnalyser` implementieren. Zur Erprobung wurde die Klasse `StayTimeAnalyser` realisiert, welche die Aufenthaltsdauer auf Web-Seiten auswertet.

sprechende Perspektiven der Entwicklung von Modulen zum Ableiten impliziter Bewertungen. Über die Scone-Caches lassen sich diverse Informationen zur Historie des Zugriffs auf Web-Seiten abfragen, die in die Implementation eines `AccessAnalyser`s einbezogen werden können.

### OnefcAdapter

Die Klasse `OnefcAdapter` kapselt den Zugriff auf den `onefc`-Identitätsmanager durch die Komponenten des `CoInternet`-Clients. Beim Systemstart werden die durch den Identitätsmanager verwalteten Identitäten eingelesen und in das interne Modell des `CoInternet`-Systems überführt. Umgekehrt sind Änderungen an Identitäten und Benutzerprofilen über Aufrufe der API des Identitätsmanagers auf dessen Identitätsdaten zu übertragen.

In Abschnitt 2.3 wurde das `onefc`-Identitätsmodell beschrieben (siehe Grafik 2.5 auf Seite 33). Um die Dienste des `onefc`-Identitätsmanager in Anspruch nehmen zu können, galt es, das basierend auf dem konzeptionellen Modell des `CoInternet`-Systems implementierte Modell auf das `onefc`-Identitätsmodell abzubilden. `CoInternet`-Identitäten werden dabei auf `Root`-Identitäten abgebildet und mit einem Kontext versehen, so daß ersichtlich ist, daß die jeweilige Identität durch das `CoInternet`-System verwendet wird. Das Pseudonym des Benutzers wird als Attribut der `RootIdentität` hinterlegt. Für jedes Benutzerprofil wird eine `Sub-Identität` angelegt. Um gezielt Attribute eines Profils abfragen zu können, wird an den Bezeichner des zugeordneten Kontextes eine `UUID` angehängt, um somit eine eindeutige Benennung sicherzustellen. Attribute eines Benutzerprofils

werden als Attribute der mit dem Profil korrespondierenden Sub-Identität dargestellt. Dabei ist das objektorientierte Attributmodell des CoInternet-Systems in eine XML-Darstellung zu überführen. Zu diesem Zweck wurde jede Klasse, die ein Attribut der Benutzerprofile des CoInternet-Systems modelliert, mit Funktionen zur Serialisierung in eine XML-Darstellung und zur Deserialisierung aus einem entsprechenden XML-String versehen.

### **Web-Services Schnittstelle**

Aus der WSDL-Datei, welche die über entfernte Funktionsaufrufe nutzbare Schnittstelle des Servers beschreibt, wurde unter Zuhilfenahme des Axis SOAP-Toolkits ein Client-Stub generiert. Dieser fungiert als lokaler Stellvertreter des Servers. Er kapselt die über das SOAP-Protokoll abgewickelten Funktionsaufrufe. Insbesondere übernimmt er das Kodieren der beim Methodenaufruf zu übertragenden Parameter in eine dem SOAP-Standard entsprechende XML-Darstellung sowie das Dekodieren der Rückgabewerte.

Der Client-Stub ermöglicht den Aufruf von Methoden des Servers analog zu Methodenaufrufen lokaler Objekte. Dennoch ist die direkte Verwendung des Client-Stubs für die realisierte Anwendung unvorteilhaft, da der Client-Stub nur Funktionen zur synchronen Kommunikation mit dem Server bietet. Der Programmablauf blockiert dadurch vom Absenden des Requests bis zum Empfang der Response. Ferner ist zur Minimierung der Netzlast ein Caching-Konzept für die Kommunikation mit dem Server wünschenswert.

### **RatingServerProxy**

Mit dem `RatingServerProxy` wurde auf Grundlage des Client-Stubs eine den Anforderungen des CoInternet-Systems entsprechende Schnittstelle zum Aufruf der Funktionen des Servers geschaffen. Der `RatingServerProxy` bietet Methoden zu synchronen und asynchronen Kommunikation mit dem Server. Asynchrone Kommunikation wurde dadurch realisiert, daß für die Abwicklung der Kommunikation jeweils ein neuer Thread gestartet wird. Nach Empfang der Antwort des Servers oder im Fehlerfall wird eine Callback-Methode des Objekts aufgerufen, das vermittelt durch den `RatingServerProxy` auf eine Funktion des Servers zugegriffen hatte.

Um zu vermeiden, daß beim Surfen im Netz innerhalb kurzer Zeit wiederholt Anfragen nach einer Prognose der Relevanz des selben Dokuments gestellt werden, werden Abfragen von Bewertungsvorhersagen in einem Cache abgelegt. Der Cache wurde als Hashtable realisiert, bei dem Requests als Schlüssel zu den gespeicherten Responses dienen. Ein im Hintergrund laufender Thread löscht regelmäßig obsoleete Einträge aus dem Cache.

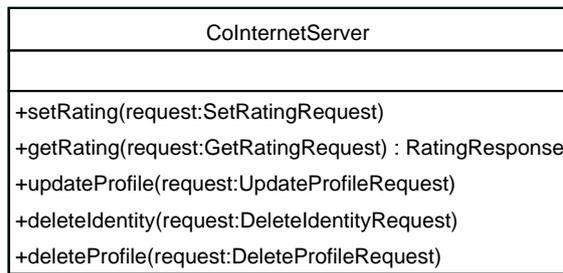


Abbildung 8.26.: Schnittstelle des CoInternetServer

### 8.3.4. Protokoll zwischen Client und Server

Grundmuster der Kommunikation zwischen dem CoInternet-Client und dem Server ist der entfernte Funktionsaufruf von Methoden des Servers durch den Client. Abstrahierend von dem Format der ausgetauschten SOAP-Nachrichten läßt sich das Protokoll anhand der Menge der von der Klasse `CoInternetServer` zum entfernten Aufruf bereitgestellten Operationen beschreiben (Abbildung 8.26).

Die Abgabe von Bewertungen erfolgt über den Aufruf der Methode `setRating()`. Bewertungsvorhersagen sowie die Liste der einem Dokument zugeordneten Annotationen lassen sich über die Methode `getRating()` abfragen. Wurden lokale Änderungen an einem Benutzerprofil durchgeführt, so kann durch Aufruf der Methode `updateProfile()` der Server explizit dazu aufgefordert werden, durch Kommunikation der Identitätsmanager das in der Datenbank des Servers abgelegte Profil zu aktualisieren. Identitäten und alle damit assoziierten Profile sind durch Aufruf der Operation `deleteIdentity()` löscherbar. Entsprechend können einzelne Profile durch `deleteProfile()` vom Server entfernt werden. Die Parameter der Methoden und die Rückgabewerte wurden jeweils in Klassen zusammengefaßt. Diese entsprechen den Java-Bean-Konventionen und können automatisch durch Axis in eine XML-Darstellung überführt werden.

Abbildung 8.27 zeigt das Klassendiagramm der Requests. Jeder Request beinhaltet die UUID der gewählten Identität und des Profils. Requests von Operationen, die den Abgleich von Benutzerprofilen erfordern, enthalten zudem noch die für die Kommunikation der Identitätsmanager benötigten Angaben und Timestamps, welche den Zeitpunkt der letzten Modifikation der verschiedenen Profilattribute markieren.

Im jetzigen Entwicklungsstand des Systems liefert lediglich die Methode `getRating()` einen Rückgabewert. In einem Objekt der Klasse `RatingResponse` werden u. a. die individuell berechnete Bewertungsprognose (`individualRating`) zu der im Request benannten URI, die durchschnittliche Bewertung (`averageRating`) und die Verteilung der

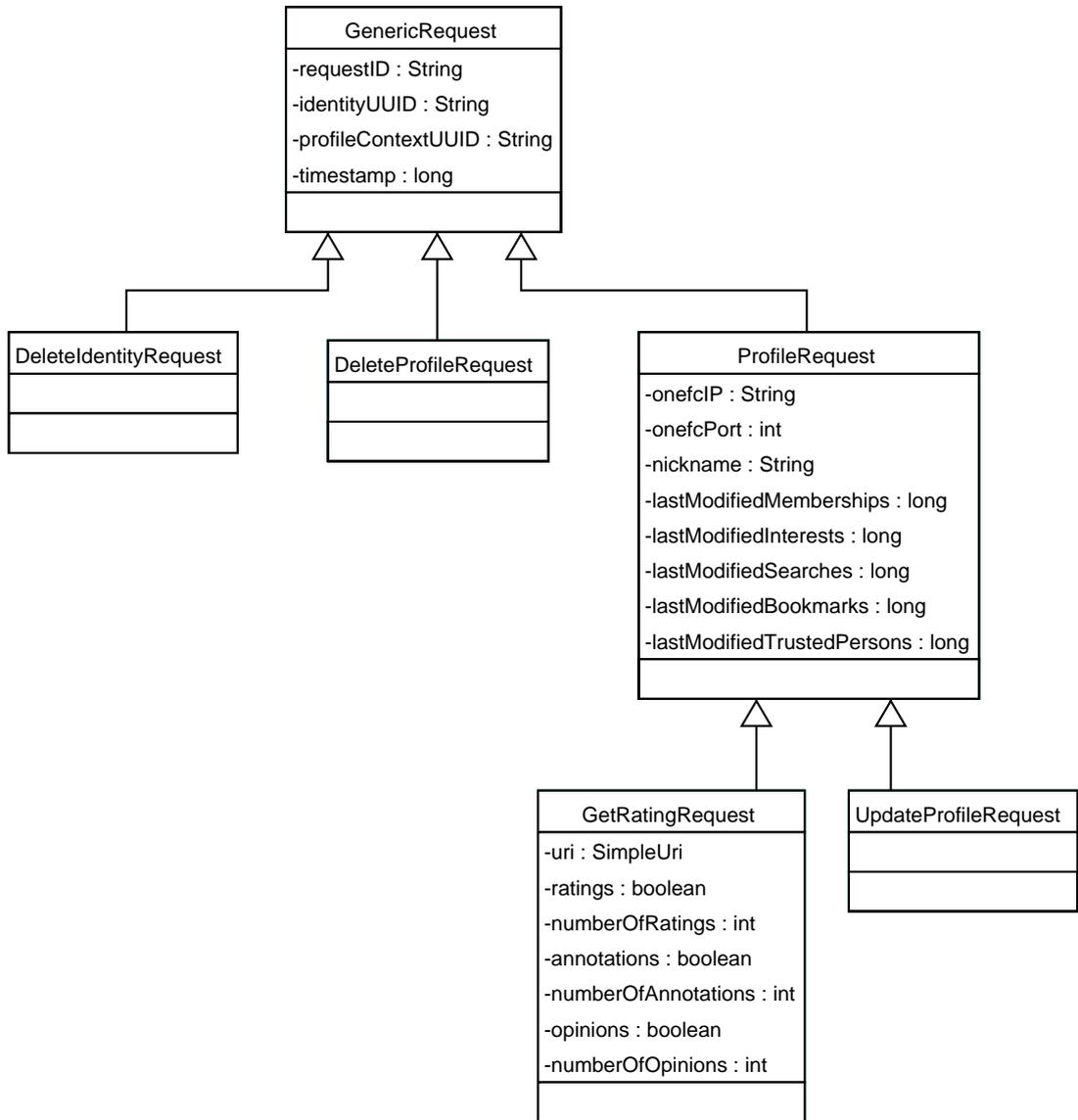


Abbildung 8.27.: Klassendiagramm der an den CoInternet-Server zu sendenden Requests

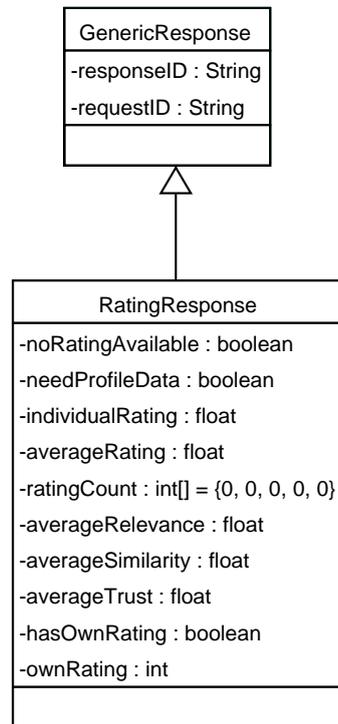


Abbildung 8.28.: Klassendiagramm der an den Client übermittelten Responses

Stimmen (`ratingCount`) übermittelt. In Abbildung 8.28 wird das Klassendiagramm der Responses dargestellt.

## 8.4. Evaluation des CoInternet-Systems

Die während der Entwicklung des CoInternet-Systems getroffenen Entwurfsentscheidungen basieren auf theoretischen Vorüberlegungen und technischen Randbedingungen. Offen bleibt die Frage, wie Benutzer ein derartiges System benutzen würden und mit welchen Problemen sie sich dabei konfrontiert sehen. In diesem Abschnitt werden die Durchführung und die Ergebnisse einer ersten Evaluation des Systems vorgestellt. Ziel war es, durch Benutzertests qualitative Aussagen zur Gestaltung der Benutzungsoberfläche und zu den realisierten Konzepten zu gewinnen. Auf diesem Wege sollten Anregungen gewonnen werden, in welchen Aspekten das Konzept eines Empfehlungs- und Annotationsystems und die gewählte Realisierung hinsichtlich eines (hypothetischen) produktiven Einsatzes zu verbessern seien.

Anhand eines konkreten Szenarios, in diesem Fall der Aufgabe, eine Menge aktueller Meldungen zu sichten, sollte überprüft werden, ob das Konzept des Empfehlungs- und Annotationssystem mit den Nutzungsgewohnheiten der Anwender bei der Navigation im Web in Einklang steht. Dabei war zu beobachten, wie die Testpersonen bei der Auswahl von Meldungen vorgehen und welche Kriterien hinter ihren Auswahlentscheidungen stehen.

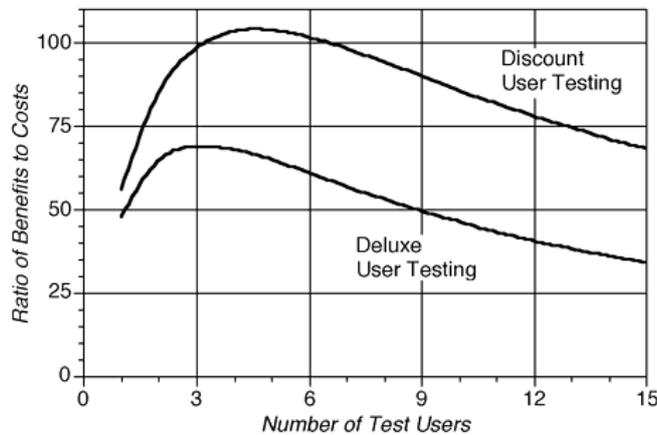
Der Erfolg eines Empfehlungssystems, welches wie das im jetzigen Entwicklungsstand befindliche CoInternet-System vorwiegend von explizit abzugebenden Bewertungen Gebrauch macht, wird entscheidend dadurch bestimmt, ob es gelingt, Anwender zur regelmäßigen Abgabe von Bewertungen zu motivieren. Die Abgabe von Bewertungen bedeutet für den Anwender einen Mehraufwand bei der Benutzung des Webs. Zusätzlich erfordert die Einbeziehung von Bewertungsvorhersagen in die Navigationsentscheidungen einen mentalen Aufwand. Hinsichtlich der zu erwartenden Akzeptanz eines Empfehlungssystems ist die Frage von besonderem Interesse, wie Anwender den aus der Nutzung des Systems resultierenden Aufwand empfinden.

Der Wert von Empfehlungen ist gering, wenn die zugrunde liegenden Bewertungen nach stark abweichenden Kriterien abgegeben wurden. Durch die Evaluation sollte ein Eindruck davon gewonnen werden, nach welchen Kriterien Anwender Web-Seiten bewerten, in dem Wissen, damit unter Umständen das Verhalten späterer Benutzer zu beeinflussen. In diesem Kontext ist auch die Frage zu sehen, welche Maßstäbe die Testpersonen bei der Bewertung von Seiten anlegen und wie sie gemäß ihrem Urteil eine Wert der Bewertungsskala auswählen.

Eine weiterer Gesichtspunkt der Evaluation des Systems ist die Frage, wie Anwender die Darstellung der prognostizierten Bewertungen interpretieren und wie sie sich das Zustandekommen einer bestimmten Empfehlung erklären.

### 8.4.1. Vorgehen

Den zeitlichen Restriktionen einer Diplomarbeit folgend wurde versucht, bei der Evaluation des Systems einen Kompromiß zwischen dem Aufwand und dem Nutzen der mit der gewählten Methode zu erzielenden Resultate zu finden. Eine in diesem Zusammenhang vorgenommene Einschränkung ist die Beschränkung auf die Kernfunktionalität des Systems. Lediglich die mit der Abgabe und Vorhersage von Bewertungen in Beziehung stehenden Teile der Anwendung wurden untersucht. Ferner wurde eine Beschränkung auf genau ein typisches Anwendungsszenario vorgenommen. Nielsen betont aus seiner langjährigen Projekterfahrung, daß man bei der Evaluation der Benutzbarkeit einer Anwendung am meisten von den ersten Testbenutzern lernt [Nie94] (siehe Abbildung 8.29). Der Test wurde mit einer relativ kleinen Gruppe von nur elf Personen durchgeführt. Diese haben an zwei jeweils einstündigen Terminen mit dem System gearbeitet. Die



**Abbildung 8.29.:** Das beste Verhältnis von Aufwand und Nutzen bei Benutzertests ist bei relativ kleinen Testgruppen zu erreichen. (Quelle: [Nie94])

Testbenutzer waren allesamt sehr erfahren im Umgang mit dem WWW. Somit konnte ausgeschlossen werden, daß auftretende Schwierigkeiten auf unzureichende Kompetenz bei der Benutzung des Webs zurückzuführen sind. Retrospektiv läßt sich sagen, daß die Größe der Testgruppe für die mit der Evaluation angestrebten Ziele hinreichend war. Bereits nach nur wenigen Sitzungen wurde eine hohe Redundanz in den Kommentaren der Benutzer festgestellt. Selbstverständlich sind die gemachten Einschränkungen bei der Interpretation der Resultate zu berücksichtigen. Was die Evaluation unter den gewählten Randbedingungen nicht leisten konnte, war die Gewinnung von Aussagen zur Genauigkeit der berechneten Vorhersagen. Hierzu wären langfristige Untersuchungen mit einer deutlich größeren Testgruppe notwendig gewesen, die den Rahmen einer Diplomarbeit überstiegen hätten (Abschnitt 5.2.4). Zudem lassen sich aus der durchgeführten Evaluation keine Aussagen darüber ableiten, wie die Leistung und die Wahrnehmung des Systems durch die Integration der onefC-Identitätsinfrastruktur beeinflusst wird.

Methode der Evaluation sind die aus der kognitiven Psychologie stammenden *Thinking Aloud Tests* [Nie93]. Benutzer werden dazu aufgefordert, bei der Bedienung der Anwendung fortwährend ihre Gedanken zu verbalisieren. Auf diese Weise kann ein insbesondere bei auftretenden Problemen aufschlußreiches Bild des mentalen Modells eines Anwenders vom System gewonnen werden.

Die Durchführung der Evaluation ist anschaulich in Abbildung 8.30 dargestellt. Das nachfolgend präsentierte Konzept der Evaluation machte es erforderlich, die Testbenutzer in zwei Gruppen einzuteilen und mit jeder Person zwei Termine durchzuführen. Vor jedem Test wurden bewußt nur sehr sparsame Hinweise zur Bedienung des Systems gege-

						
	1	2	3	4	5	6
Meldung 1	2	-2	1	0	-2	2
Meldung 2	1	1	-1	-2	1	2
Meldung 3	-2	0	2	1	0	1
Meldung 4	1	1	-2			
Meldung 5	-1	0	1			
Meldung 6	-2	2	1			
Meldung 7				2	1	0
Meldung 8				0	1	2
Meldung 9				1	0	2

**Abbildung 8.30.:** Zur Durchführung der Evaluation wurden die Testbenutzer in zwei Gruppen eingeteilt. Im ersten Termin stand die Abgabe von Bewertungen im Vordergrund. Basierend auf den im ersten Termin erfaßten Meldungen wurden im zweiten Termin jeweils für die korrespondierende Gruppe Bewertungsvorhersagen generiert.

ben, um einen Eindruck davon gewinnen zu können, ob der Umgang mit der Anwendung intuitiv erlernbar ist.

In Anlehnung an das erste der in Abschnitt 8.1.1 beschriebenen Szenarien wurde für den Test jenes Szenario gewählt, daß sich der Anwender mit dem Ziel „auf dem Laufenden zu bleiben“ auf einer Web-Site zu aktuellen IT-Themen informieren möchte. Es wurden zum Zeitpunkt des Tests aktuelle Meldungen von verschiedenen Sites in die den Testpersonen präsentierten Listen von Meldungen übernommen.

Die Testpersonen wurden gebeten, zum ersten Termin eine Auswahl von zehn favorisierten Bookmarks mit Bezug zu IT-Themen mitzubringen. Diese wurden zur Initialisierung in das Benutzerprofil eingetragen. Im Vordergrund des ersten Termins stand die Abgabe von Bewertungen. Die Personen wurden aufgefordert, während des Testablaufs Erwartungen an angezeigte Meldungen zu formulieren und Begründungen zu ihren Wertungen zu geben. Eine Auswahl von acht Meldungen war von allen Benutzern zu bewerten. Diese Bewertungen dienten zusammen mit den Bookmarks als Grundlage für den Vergleich von Benutzern der beiden Testgruppen. Zusätzlich waren aus einer Liste von dreizehn Artikeln einige zu lesen und zu bewerten. In diesem Schritt wurde jeder der beiden Testgruppen eine unterschiedliche Liste von Meldungen präsentiert.

Im zweiten Termin galt es, die Liste der von der korrespondierenden Gruppe im ersten

Termin bewerteten Meldungen unter Berücksichtigung der errechneten Vorhersagen zu sichten. Die Personen wurden aufgefordert, bevor sie auf einen Link zu einer Meldung klicken, das angezeigte HyperScout-Popup zu beachten. Sie sollten mitteilen, wie die angezeigte Bewertungsvorhersage mit der eigenen Beurteilung auf Grundlage der in der Liste verlinkten Überschrift der Artikels übereinstimmt. Nach anschließendem Durchlesen des Artikels sollten sie ihre Bewertungen nennen. Bei großen Differenzen zwischen der errechneten Vorhersage und der eigenen Einschätzung war anhand der vom System gebotenen Möglichkeiten nachzuvollziehen, wie sich die Vorhersage aus den Stimmen der anderen Testpersonen zusammensetzt.

### 8.4.2. Ergebnisse

Aufgrund der gewählten Vorgehensweise bei der Evaluation haben die an dieser Stelle präsentierten Ergebnisse nicht den Anspruch, repräsentativ zu sein. Vielmehr handelt es sich um ein Meinungsbild, das die Aussagen einer verhältnismäßig kleinen Testgruppe widerspiegelt, die in einem klar umrissenen Szenario mit dem System gearbeitet hat.

Die Abgabe von Bewertungen mittels der in Abbildung 8.6 auf Seite 129 gezeigten Gruppe von Radio-Buttons wurde von allen Testpersonen sofort verstanden, jedoch vermißten viele ein Feedback über das erfolgte Absenden der Bewertung an den Server. Einige Testpersonen gaben sogar an, daß sie, einen zusätzlichen Mausclick in Kauf nehmend, einen separaten Button zum Abschicken der Bewertung bevorzugt hätten.

Bei Anzeige einer bereits bewerteten Seite wird der entsprechende Button der Gruppe von Radio-Buttons selektiert. Eine Änderung kann einfach durch Anklicken eines anderen Buttons der Gruppe erfolgen. Das Löschen von Bewertungen war bislang nur über die web-basierten Elemente der Benutzungsschnittstelle möglich. In Situationen, in denen versehentlich eine Seite bewertet wurde, hat sich dieser Weg als unnötig kompliziert erwiesen. Sinnvoll wäre es, die Gruppe der Radio-Buttons um eine weitere Schaltfläche zu ergänzen, die standardmäßig bei nicht bewerteten Seiten selektiert wird und auszuwählen ist, wenn eine Bewertung zurückgenommen werden soll.

Es besteht die Tendenz Meldungen, die schon aufgrund der Überschrift als uninteressant eingestuft wurden, nicht negativ zu bewerten. Es kann davon ausgegangen werden, daß negative Wertungen eine höhere Aussagekraft besitzen als positive. Dies ist beim Vergleich von Bewertungsmustern und bei der Aggregation der Bewertungen in Anrechnung zu bringen. Die Benutzungsoberfläche des CoInternet-Clients gestattet es nur, das aktuell im Browserfenster angezeigte Dokument zu bewerten. Offenkundig uninteressante Meldungen werden jedoch gar nicht erst geladen und folglich auch nicht bewertet. Eine zu erprobende Verbesserung wäre es, die direkte Bewertung verlinkter Dokumente vorzusehen, beispielsweise durch ein Kontextmenü, welches bei einem rechten Mausclick auf einen Link erscheint.

Es scheint den Nutzungsgewohnheiten einiger Anwender zu entsprechen, Übersichtsseiten in einem Browserfenster zu belassen und für einzelne Meldungen jeweils neue Fenster zu öffnen. Bei einer derartigen Vorgehensweise schien es den Testpersonen nicht unbedingt bewußt zu sein, auf welches Browserfenster sich die Darstellung in dem Fenster des CoInternet-Clients bezieht. Andere Personen haben nur ein Browserfenster verwendet, aber aus Gewohnheit sehr schnell nach dem Durchlesen der Meldung auf den „Zurück-Button“ geklickt, um zurück zur Liste der Meldungen zu gelangen, und erst dann die Bewertung abgegeben, die sich eigentlich auf die zuvor angezeigte Meldung beziehen sollte. Wenn das Client-Fenster betrachtet wird, ist das Dokument, welches bewertet werden kann oder auf das sich eine Vorhersage bezieht, außerhalb des Aufmerksamkeitsfokus, was es erschwert, den Bezug zwischen den Elementen des Client-Fensters und dem im Browser angezeigten Dokument herzustellen. Eine Integration der Benutzungsoberfläche in den Browser ist aus diesem Grund anzustreben.

Die Testpersonen hatten bei der Beurteilung der Meldungen sehr unterschiedliche Bewertungskriterien und -maßstäbe verwendet. Vorherrschende Kriterien waren das persönliche Interesse an einer Meldung in Verbindung mit deren Erwartungskonformität. Artikel, deren Überschrift interessant klang, die jedoch inhaltlich enttäuschten, wurden negativ bewertet. Auffallend ist jedoch, daß einzelne Personen sich bemühten, bei der Bewertung möglichst objektiv vorzugehen und statt dem subjektiven Interesse schwerpunktmäßig ihrer Meinung nach objektive Qualitätskriterien zu berücksichtigen.

Teilweise wurde die Bewertung nicht auf das Dokument selbst bezogen, sondern die eigene Meinung zum Gegenstand der Meldung bzw. die Haltung zur Meinung des Autors zum Ausdruck gebracht. So wurde etwa ein Artikel, der über einen MP3-Player berichtete, negativ bewertet, obwohl großes Interesse an diesem Thema vorlag, da die Qualitäten des genannten Players schlecht beurteilt wurden. Eine Meldung, die über den Betrug mit 0190er-Nummern berichtete, erhielt ein negatives Votum, da die Testperson zum Ausdruck bringen wollte, daß sie gegen derartiges betrügerisches Verhalten sei.

Interessant ist, daß in einigen Fällen unter strategischen Gesichtspunkten abgestimmt wurde. So wurde mit der Intention, andere Nutzer auf ein Thema aufmerksam zu machen, positiv für eine eigentlich als uninteressant angesehene Meldung votiert. Zudem haben Benutzer zum Ausdruck gebracht, daß sie sich bei der Auswahl eines Wertes der Skala an der ihnen präsentierten Vorhersage und dem Durchschnittswert orientieren um die nachfolgenden Personen präsentierten Werte zu beeinflussen.

Hinsichtlich der Ausschöpfung des durch die Bewertungsskala vorgegebenen Wertebereichs gab es zwischen den Benutzern Unterschiede. Einige von ihnen haben fast ausschließlich die Extremwerte gewählt, während andere Testpersonen sich bemüht haben, weiter zu differenzieren. Ein Teil der Anwender wäre sicherlich mit einer nur zwei Werte umfassende Skala zufrieden gewesen. Hingegen wurde von einigen Personen sogar der Wunsch geäußert, die Bewertungen in deutlich feineren Abstufungen über einen Schie-

bereglert erfassen zu können. Der mittlere Wert der Bewertungsskala wurde kaum verwendet und scheint verzichtbar zu sein. Es ist zweifelhaft, ob Anwender die Motivation aufbringen, eine Seite zu bewerten, die sie weder besonders gut noch schlecht finden.

Deutlich wurde, daß der Aufwand eine Seite zu bewerten, nur zweitrangig in der motorischen Aufgabe, auf einen der Radio-Buttons zu klicken, besteht, sondern primär in der Reflexion über das persönliche Interesse an einer Seite und in der Beurteilung deren Qualität zu sehen ist. Gemäß dem eigenen Urteil einen Wert der Skala auszuwählen, also eine Entscheidung zu treffen, scheint überdies eine Belastung darzustellen.

Eine Beschriftung der Bewertungsskala, die Richtlinien für die Verwendung der einzelnen Werte vorgibt, wäre wünschenswert, erscheint jedoch aufgrund der thematischen Vielfalt im Web äußerst problematisch.

Typische Verhaltensweise beim Sichten einer Liste von Meldungen ist das Scannen der Überschriften. Für eine derartige Vorgehensweise ist die Anzeige von Bewertungsvorhersagen in Popup-Fenstern, die erst eingeblendet werden, sobald der Mauszeiger über einem Link plaziert wird, unzureichend. Bereits beim Überblicken der aufgelisteten Überschriften, erkannten die Testbenutzer sehr schnell sie potentiell interessierende Artikel. Der Wert eines Empfehlungssystem ist bei der Aufgabe, eine Menge von Meldungen zu sichten, insbesondere bei den Meldungen gegeben, bei denen die errechnete Vorhersage von der Einschätzung des Benutzers nach Lesen der verlinkten Überschrift abweicht. Wenn also, nach der Überschrift zu urteilen, ein Artikel uninteressant zu sein scheint, er aber in Wirklichkeit für den Anwender äußerst lesenswert ist. Anwender sollten folglich diejenigen Links, für die hohe Relevanzvorhersagen errechnet wurden, auf einen Blick erkennen können. Die durch Modifikation der Dokumente realisierte Präsentation von Bewertungsvorgaben unmittelbar neben den Links scheint jedoch angesichts dessen, daß somit das Layout der Seiten in Mitleidenschaft gezogen werden würde, keine zu erwägende Alternative zu den Popup-Fenstern zu sein. Eine Option wäre es hingegen, entsprechend dem Wert der Bewertungsvorhersage eingefärbte Flächen, halbtransparent über den Links einzublenden. Diese *Overlays* oder Kennzeichnungen innerhalb der Seite könnten zudem durch den Benutzer per Tastendruck zu aktivieren sein. Mittels dieser *Link-On-Demand-Technik* könnte zwischen einer unverfälschten Darstellung und der sofort sichtbaren Kennzeichnung interessanter Links umgeschaltet werden [WOL01].

Bezüglich der Präsentation der Informationen zur Erklärung des Bewertungsprozesses sind Unzulänglichkeiten des Prototypen offenkundig geworden. Die Werte „Ähnlichkeit“, „Vertrauen“ und „Relevanz“ (Abschnitt 8.2.5) lassen sich mathematisch begründen, sind aber für den Benutzer anscheinend nicht sehr eingängig. Die Darstellung der Verteilung der Stimmen als einfaches Histogramm ist für das Verständnis der Entstehung von Vorhersagen unzulänglich, da diese Visualisierung unberücksichtigt läßt, daß die Meinung der Anwender abhängig von Ähnlichkeit und Vertrauen zum aktiven Benutzer unterschiedliches Gewicht besitzt. Optional könnten die Balken des Histogrammes gemäß des

Gewichts der Stimmen in mehrere Klassen unterteilt werden. Ersatzweise wäre auch die in [HKR00] vorgeschlagene Visualisierung zu erproben, bei der jede Stimme durch einen sehr schmalen Balken repräsentiert wird, dessen Länge die abgegebene Wertung darstellt. Durch die Sortierung der Balken nach dem Gewicht der Stimme, ist nicht nur sichtbar, wie die Bewertungen insgesamt streuen, sondern auch wie einheitlich die ähnlichen Benutzer abgestimmt haben.

Durch die Evaluation wurde deutlich, daß Benutzerinteressen nicht nur evolutionären Veränderungen unterliegen, sondern sich auch sehr abrupt ändern können. Sichtbar wurde dies vor allem bei den Kommentaren zu Meldungen, die einen Bezug zu Produkten haben. Ist die Anschaffung eines Produktes der entsprechenden Kategorie geplant, so werden diesbezügliche Web-Seiten als äußerst relevant angesehen. Mit der erfolgten Anschaffung endet jedoch in vielen Fällen das Interesse an Texten mit Bezügen zur entsprechenden Produktkategorie. Eine Herausforderung bei der Gestaltung eines Empfehlungssystems für Web-Seiten ist es, Werkzeuge zur komfortablen Anpassung von Benutzerprofilen zu entwerfen.

### 8.5. Auswirkungen der Integration einer Identitätsinfrastruktur

In diesem Abschnitt wird anhand des realisierten Empfehlungs- und Annotationssystems eine zusammenfassende Beantwortung der eingangs in dieser Arbeit gestellten Frage nach den Auswirkungen der Integration einer Identitätsinfrastruktur auf eine kollaborative Internetanwendung vorgenommen. Wie bereits in der Einleitung erwähnt wurde, erfordert die Beantwortung dieser Fragestellung eine Betrachtung in zwei Schritten (siehe Abschnitt 1.1).

Zunächst ist festzustellen, welche Bedeutung Identitätskonzepten in einer Social-Navigation-Anwendung, insbesondere einem auf Collaborative-Filtering-Techniken aufsetzenden Empfehlungssystem, zukommt – in welchen Aspekten also durch die Integration einer Identitätsinfrastruktur Veränderungen zu erwarten sind. Diese Analyse liefert Ansatzpunkte, an denen die Anwendung durch Integration einer Identitätsinfrastruktur optimiert werden kann. Resultat der Betrachtung auf dieser Ebene sind Hinweise darauf, wie ein Empfehlungs- und Annotationssystem zu gestalten sei, um aus der Integration einer Identitätsinfrastruktur Nutzen zu ziehen.

In einem zweiten Schritt der Betrachtung sind die sozialen Implikationen eines an die Identitätsinfrastruktur angepassten Empfehlungs- und Annotationssystems zu betrachten. Es ist die Frage nach den Veränderungen des Verhaltens des Benutzer und der von ihnen empfundenen Nützlichkeit des Systems von Bedeutung. Damit verbunden ist die Frage, wie sich die Wahrnehmung, die Anwender von den anderen Nutzern und der

Gemeinschaft haben, wandelt. Letztendlich wären in diesem Schritt die Entscheidungen zu verifizieren, die beim von der Integration einer Identitätsinfrastruktur geprägten Entwurf eines Empfehlungssystems getroffen wurden. Dazu wären langfristige Evaluationen mit einer großen Zahl von Benutzern notwendig, die im Rahmen dieser Arbeit nicht durchführbar sind. Dennoch werden Tendenzen aufgezeigt, welche sozialen Implikationen die Integration der Identitätsinfrastruktur haben könnte.

Charakteristisch für die onefC-Identitätsinfrastruktur (Abschnitt 2.3) ist die Verwendung global eindeutiger Identifikatoren für die Benutzeridentitäten und ein generisches Identitätsmodell, das den anwendungsübergreifenden Austausch von Identitätsdaten praktikabel macht, sofern ein gemeinsames Verständnis der Semantik der Daten gegeben ist. Zudem wird diese Identitätsinfrastruktur durch eine dezentrale Datenhaltung gekennzeichnet. Sämtliche Identitätsdaten werden durch den lokal auf dem Rechner des Benutzers laufenden Identitätsmanager verwaltet und können unter seiner Kontrolle an verschiedene Dienste und andere Kommunikationspartner übermittelt werden. Diese prinzipiell dezentrale Datenhaltung schließt jedoch nicht aus, daß auf der Identitätsinfrastruktur aufsetzende Applikationen, wie das CoInternet-System, an zentraler Stelle Replikate der Identitätsdaten vorhalten. In Abgrenzung davon wird angenommen, daß eine herkömmliche kollaborative Internetanwendung, welche nicht auf einer Identitätsinfrastruktur aufsetzt, nur im Rahmen dieser Anwendung gültige Benutzeridentifikatoren, beispielsweise Paare aus einem Benutzernamen und einem Passwort, verwendet. Ferner wird für eine derartige Anwendung angenommen, daß sie ein propriäteres Identitätsmodell aufweist und Identitätsdaten zentral auf dem Server des Diensteanbieters abgelegt werden. Sie sind somit für eine weitergehende Nutzung unter der Kontrolle des jeweiligen Anwenders unzugänglich.

Ein Gesichtspunkt der Betrachtung von Auswirkungen der Integration einer Identitätsinfrastruktur ist die Vereinfachung der Implementation kollaborativer Anwendungen. Zur Realisierung identitätsbezogener Programmteile kann auf die Dienste der onfC-Identitätsinfrastruktur zurückgegriffen werden. Anwendungsspezifische Benutzerprofile lassen sich auf Grundlage des generischen Identitätsmodells realisieren. Überdies wird die Persistierung von Identitäten und die Kommunikation von Identitätsattributen über die Identitätsmanager abgewickelt.

Die Leistung eines Collaborative-Filtering-Systems wird maßgeblich von der Kenntnis des Systems über die Benutzer bestimmt. Nur bei hinreichend genauen Benutzerprofilen wird ein Empfehlungssystem automatisch zwischen Bewertungen abgebenden und gleichgesinnten, Empfehlungen suchenden Personen vermitteln können. Neu in das System eintretende Benutzer verfügen noch über kein aussagekräftiges Benutzerprofil und können somit keine individuellen Empfehlungen erhalten. Dieser Sachverhalt wurde in Abschnitt 5.2.3 als Kaltstartproblem bezeichnet.

Die onefC-Identitätsinfrastruktur eröffnet den Benutzern weitergehende Möglichkeiten,

die eigene Persönlichkeit gegenüber dem Empfehlungssystem zum Ausdruck zu bringen. Durch den lokalen Identitätsmanager verwaltete Attribute eines Benutzerprofils, die aus der Benutzung verschiedenster Anwendungen resultieren können, lassen sich an ein Empfehlungssystem übertragen und können in den Prozeß der Berechnung persönlicher Empfehlungen einbezogen werden. Ein Empfehlungssystem, welches flexibel mit verschiedenen Profilattributen umgehen kann, das also auf Ebene der Benutzermodellierung interoperabel mit einigen anderen Diensten ist, könnte neu in das System eintretenden Benutzern ohne Anlaufphase persönliche Empfehlungen präsentieren. Das Kaltstartproblem würde somit abgeschwächt oder sogar beseitigt werden.

Durch das CoInternet-System wurde demonstriert, daß Collaborative Filtering auf Grundlage verschiedenster Attribute erfolgen kann, sofern diese einen Bezug zu den Präferenzen des jeweiligen Benutzers aufweisen (Abschnitte 8.2.1 und 8.2.5). In Ermangelung anderer, auf der Identitätsinfrastruktur aufbauender Anwendungen wurden zur Generierung von Empfehlungen ausschließlich durch den CoInternet-Client bzw. den Profil-Editor erfaßte Profilattribute ausgewertet. Jedoch sind Bookmarks und Suchbegriffe exemplarisch für Profilattribute zu nennen, die, auch wenn sie durch das CoInternet-System in das durch den Identitätsmanager verwaltete Benutzerprofil eingetragen wurden, nicht in unmittelbarem Zusammenhang mit der Benutzung des CoInternet-Systems erfaßt wurden. Analog könnten Profilattribute anderer Anwendungen, beispielsweise die Historie erworbener Bücher oder zur Personalisierung einer News-Site angegebene Themenbezeichnungen, durch das CoInternet-System ausgewertet werden.

Einen Dienst im Internet an die eigenen Bedürfnisse anzupassen, ist entweder mit einigem Aufwand verbunden, wenn beispielsweise umfangreiche Formulare auszufüllen sind, oder erfordert, sofern Benutzerprofile implizit aufgebaut werden, einen längeren Benutzungszeitraum. Hat ein Benutzer einen Dienst erst einmal so weit konfiguriert bzw. hat sich der Dienst im Laufe der Zeit so weit an den Benutzer angepaßt, daß bei der Personalisierung zufriedenstellende Ergebnisse erzielt werden, so wird sich ein Anwender schwer tun, zu einem konkurrierenden Angebot zu wechseln. Dieser Umstand kommt insbesondere Dienst Anbietern aus dem E-Commerce-Umfeld nicht ganz ungelegen, denn sie haben ein hohes Interesse daran, die Kunden an sich zu binden. Die Möglichkeit, vermittelt durch die Identitätsinfrastruktur Identitätsdaten auszutauschen oder gemeinsam zu nutzen, fördert die Mobilität der Anwender bei der Wahl der von ihnen genutzten Dienste. Hingegen haben konkurrierende Dienste kein Interesse daran, die von ihnen erhobenen Profildaten mit anderen Anbietern zu teilen. Sie werden folglich entweder nicht davon Gebrauch machen, die Daten mittels des Identitätsmanagers auf dem Rechner des Benutzers abzulegen oder regelmäßig „neue Features hinzufügen“, um durch die entstehende Inkompatibilität des Formats der Profilattribute zu verhindern, daß auch andere Anwendungen die Daten nutzen. Festzuhalten ist, durch Einbeziehung der Identitätsinfrastruktur werden Empfehlungssysteme realisierbar, die aus verschiedenen Quellen gespeiste Benutzerprofile verwenden. Offen bleibt die Frage, ob im angestrebten Maße

Profilattribute anderer Anwendungen zur Verfügung stehen und ob Anwender bereit sind, den Zugriff eines Empfehlungssystems auf diese Daten zu billigen.

Ein ausschließlich auf Collaborative-Filtering-Techniken basierendes Empfehlungssystem wird nur bereits bewertete Objekte empfehlen können (siehe Abschnitt 5.2.3). Angesichts der Menge der im Web abrufbaren Dokumente, wird es einem Empfehlungssystem für Web-Seiten selbst bei einer sehr großen Nutzerbasis nur möglich sein, für einen kleinen Anteil der Dokumente Relevanzprognosen zu präsentieren. Demgemäß wäre die Einbeziehung von Techniken des Content-based Filtering (Abschnitt 5.1) eine sinnvolle Weiterentwicklung des CoInternet-Systems. Einige der durch die onefC-Identitätsinfrastruktur übertragenen und vom CoInternet-System zum Auffinden ähnlicher Benutzer genutzten Profilattribute wären auch durch Content-based-Filtering-Algorithmen zur direkten Abschätzung der Relevanz eines Dokuments nutzbar. Somit könnten, begünstigt durch den Einsatz der onefC-Identitätsinfrastruktur, hybride Empfehlungssysteme (Abschnitt 5.3) realisiert werden.

Ebenso, wie ein Empfehlungssystem auf Identitätsdaten anderer Anwendungen zugreifen kann, können umgekehrt die im Rahmen der Benutzung eines Empfehlungssystems aufgebaute Benutzerprofile zur Personalisierung anderer Dienste herangezogen werden. Infolgedessen würde sich der Aufwand, beispielsweise durch Abgabe expliziter Bewertungen ein Benutzerprofil aufzubauen, relativieren.

Durch die Verwendung der Identitätsinfrastruktur wird die Ad-hoc-Nutzung des CoInternet-Dienstes möglich. Anwender können den Dienst ohne vorherige Anmeldung verwenden und müssen sich nicht bei jedem Aufruf des Systems durch Nennung eines Benutzernamens identifizieren und durch die Eingabe eines Passworts autorisieren, denn durch die Identitätsinfrastruktur werden Single-Sign-On-Anmeldeprozeduren realisierbar. Der Beitritt zum und der Zugriff auf den Dienst wird folglich vereinfacht.

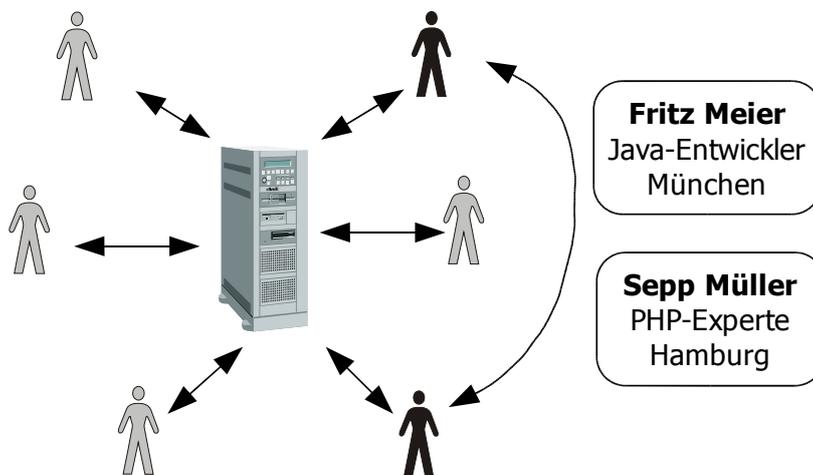
Die im CoInternet-System hinterlassenen Bewertungen und Annotationen sind Navigationshinweise, die nachfolgende Benutzer in ihre Navigationsentscheidungen einbeziehen können. Das CoInternet-System befähigt die Benutzer zum sozialen Navigieren, wobei eine Herausforderung darin besteht zu beurteilen, welche Bedeutung einzelnen Navigationshinweisen beizumessen ist. In diesem Zusammenhang ist das Wissen über die Urheber von Navigationshinweisen maßgeblich. Die Wahrnehmung der Identität eines Kommunikationspartners ist bestimmend für die Einschätzung von dessen Vertrauenswürdigkeit und Kompetenz (siehe Abschnitte 2.2 und 4.6).

Der Erfolg einer Social-Navigation-Anwendung wie dem CoInternet-System, ist abhängig von dem Engagement der Benutzer. Die Nutzer eines auf Collaborative-Filtering-Techniken basierenden Empfehlungssystems bilden eine soziale Gemeinschaft. Nur wenn sich im Verhalten eines Großteils der Nutzerschaft die aktive Beteiligung und die passive Benutzung in einem Gleichgewicht befindet, wird das Empfehlungssystem seiner Aufgabe gerecht werden können (siehe Abschnitte 4.8 und 5.2.3).

Einer der Leitgedanken bei der Gestaltung des CoInternet-Systemes war es, den Empfehlungs-Prozeß möglichst transparent zu gestalten, bis hin zum Aufzeigen der Beiträge einzelner Benutzer. Somit sollte die Einschätzung der durch das System generierten Empfehlungen vereinfacht und das Bewußtsein dafür geschärft werden, daß die Empfehlungen in einem kollaborativen Prozeß entstehen. Ziel war es, eine Online-Community zu schaffen und das Zugehörigkeitsgefühl der Anwender zu dieser Gemeinschaft zu stärken. Nachstehend werden einige Aspekte aufgeführt, welche Konsequenzen die Integration der onefC-Identitätsinfrastruktur in Hinblick auf dieses Entwurfsziel hat.

Die Identitätsinfrastruktur fördert die Interoperabilität zwischen Online-Anwendungen. Global eindeutige Identifikatoren und ein einheitliches Identitätsmodell machen es möglich, unter der selben durch den Identitätsmanager verwalteten Identität in verschiedenen kollaborativen Anwendungen aufzutreten. Die Nutzer des CoInternet-Systems können somit nahtlos über Grenzen von Anwendungen hinweg miteinander kommunizieren und kollaborieren. Personen, die eine Seite, welche einem selbst außerordentlich gut gefällt, hoch bewertet haben und zudem noch eine große Ähnlichkeit des Benutzerprofils aufweisen, könnten außerhalb des CoInternet-Dienstes direkt kontaktiert werden. Die Leistung des CoInternet-Dienstes ist es in diesem Fall, Anwender, basierend auf der Auswertung der durch die onefC-Identitätsinfrastruktur übertragenen Benutzerprofile, beim Auffinden geeigneter Interaktionspartner zu unterstützen. In virtuellen Umgebungen ist die Kommunikation mit einer Person prägend für das Bild dieser Person. Durch die direkte Kommunikation mit anderen Nutzern des CoInternet-Systems und aufgrund der Möglichkeit, mittels der Identitätsinfrastruktur Attribute der eigenen Persönlichkeit anderen Nutzern gegenüber zu offenbaren, können vertrauensvolle Beziehungen zu anderen Nutzern aufgebaut und das Gefühl, einer Gemeinschaft anzugehören, gestärkt werden.

Die von der Identitätsinfrastruktur vergebenen global eindeutigen Benutzeridentifikatoren gestatten ein kontextübergreifendes Wiedererkennen von Benutzern. Das Bild der Benutzer voneinander wird somit auch von deren Handlungen außerhalb des CoInternet-Systems und den in anderen Kontexten offenbarten Attributen ihrer Persönlichkeiten geformt. Durch das Vorhandensein anwendungsübergreifend gültiger Identifikatoren ist eine Voraussetzung gegeben, um eine außerhalb des CoInternet-Systems erworbene Reputation auf diese Anwendung zu übertragen. Umgekehrt könnte der Aufbau einer in andere Kontexte transferierbare Reputation die Motivation der Mitglieder der CoInternet-Community zu aktiver Beteiligung erhöhen. Der Grad an Identifizierbarkeit der Kommunikationspartner ist ein bestimmender Einflußfaktor bei der Beurteilung von Kommunikation (Abschnitt 4.6). Resultierend aus der Eigenschaft der Identitätsinfrastruktur, kontextübergreifendes Wiedererkennen von Personen zu gestatten, können Benutzer innerhalb der CoInternet-Community anonym bleiben und dennoch von ausgewählten Personen, die beispielsweise persönlich bekannt sind oder mit denen in anderen Umgebungen Identitätsinformationen ausgetauscht wurden, unter ihrer realen Identität wahrgenom-



**Abbildung 8.31.:** Anwender können außerhalb des CoInternet-Systems Identitätsdaten austauschen und sich innerhalb der CoInternet-Community wiedererkennen. Der Grad an Identifizierbarkeit der Nutzer ist folglich nicht für alle einheitlich festgelegt, sondern variiert aus Sicht jedes Teilnehmers.

men werden (siehe Abbildung 8.31). Durch die Integration der onefC-Identitätsinfrastruktur kann eine pauschale Festlegung vermieden werden, wie sich Nutzer innerhalb der Community identifizieren müssen. Statt dessen kann jeder Nutzer selbst differenzieren, welchen Teil seiner Identität er welchen anderen Nutzern gegenüber offenbart.

Aktives Collaborative Filtering (Abschnitt 5.2) hat gegenüber automatischem Collaborative Filtering den Vorteil, daß kein Kaltstartproblem besteht, sofern Anwender bereits beim Eintritt in das System einige für sie relevante Empfehler benennen. Anwender erlangen mehr Kontrolle über den Prozeß der Generierung von Empfehlungen. Unabhängig von der Ähnlichkeit der Benutzerprofile kann bestimmten Personen, etwa Experten auf einem Gebiet, eine besondere Bedeutung zugesprochen werden. Aktives Collaborative Filtering setzt jedoch voraus, daß Anwender in der Lage sind, eine Liste von Personen anzugeben, deren Meinung ihnen wichtig ist und denen sie vertrauen. Die Kombination eines Empfehlungssystems mit einer Identitätsinfrastruktur läßt den Einsatz von aktivem Collaborative Filtering auch in großen Communities sinnvoll erscheinen, denn einerseits können vertraute bzw. kompetente Personen aus anderen Kontexten wiedererkannt werden, und andererseits ist begünstigt durch die Interoperabilität mit Anwendungen zur direkten Kommunikation der Aufbau von Beziehungen innerhalb der CoInternet-Community möglich.

Die Verwendung einer Identitätsinfrastruktur bedeutet nicht automatisch, daß Benutzer

mehr über sich preisgeben müssen. Der Unterschied liegt darin, daß es für sie nachvollziehbar wird, wann welche Daten an einen anderen Rechner im Internet übermittelt wurden und sie mehr Kontrolle über die Herausgabe von Daten erlangen. Zudem können Redundanzen bei der Eingabe von Daten durch die anwendungsübergreifende Verwendung von Identitätsdaten vermieden werden. Dennoch resultieren aus der Integration einer Identitätsinfrastruktur in ein Empfehlungs- und Annotationssystem ernstzunehmende Risiken für den Schutz der Privatsphäre der Anwender. Gelingt es einer Person, unberechtigt eine Verbindung zwischen der virtuellen und realen Identität eines Benutzers herzustellen, so vergrößert sich in diesem Fall durch die Verwendung von über Grenzen von Anwendungen hinweg nutzbaren Identitäten der Schaden für den Anwender. Profilattribute, die im Glauben eingegeben wurden, daß diese pseudonymisiert gespeichert werden würden, ließen sich plötzlich einer bestimmten Person zuordnen.

Es wurde aufgezeigt, wie durch die Integration einer Identitätsinfrastruktur die Fähigkeit eines Collaborative-Filtering-Systems, neuen Benutzern individuelle Empfehlungen zu präsentieren, verbessert werden kann. Ferner wurde dargelegt, welche neuen Möglichkeiten für die Anwender aus der Einbeziehung einer Identitätsinfrastruktur resultieren. Offen bleibt jedoch die Frage, welche Konsequenzen die Integration einer Identitätsinfrastruktur in ein Empfehlungs- und Annotationssystem, wie dem in dieser Arbeit entwickelten CoInternet-System, aus Sicht eines Anwenders hat, der verschiedene auf der Identitätsinfrastruktur aufsetzende Dienste nutzt. Aus diesem Blickwinkel ist die Frage von Bedeutung, ob einhergehend mit der Verwendung über die Identitätsinfrastruktur übertragener Profilattribute, die von den Anwendern erfahrene Vorhersagegenauigkeit eines Empfehlungssystems steigt, und ob dieses letztendlich zu einer höheren Akzeptanz des Systems führt. Nur einem von den Anwendern als nützlich wahrgenommenen Empfehlungssystem wird es gelingen, die für den erfolgreichen Betrieb notwendige Zahl von Benutzern an sich zu binden. Die Frage ist auch, inwiefern Benutzer die vermeintlichen Risiken auf sich nehmen werden, Daten von sich preiszugeben. Sicherlich sind sie dazu eher bereit, wenn sie die Leistungen des Empfehlungssystems positiv beurteilen. Ungewiß ist auch, ob es wirklich begünstigt durch den Einsatz einer Identitätsinfrastruktur gelingen kann, Vertrauen zwischen den Anwendern und damit auch das Vertrauen in vom System generierte Empfehlungen herzustellen. Durch die Identitätsinfrastruktur wird die Wahrnehmung der Benutzer untereinander und die Möglichkeit der Kommunikation miteinander verbessert. Ob auf diesem Wege unter den Nutzern eines Empfehlungssystems tatsächlich das Gefühl, einer Gemeinschaft anzugehören, entsteht, ob das Streben nach Reputation innerhalb der Gemeinschaft das Engagement der Anwender vergrößert und somit das Free-Rider-Problem minimiert wird, bleibt an dieser Stelle unbeantwortet.

## 9. Fazit

In diesem abschließenden Kapitel der Diplomarbeit werden zunächst resümierend die Ergebnisse der Arbeit zusammengefaßt. Anschließend wird ein Ausblick auf Entwicklungsmöglichkeiten des CoInternet-Systems gegeben und es werden offene Fragen aufgeführt.

### 9.1. Ergebnisse der Arbeit

Das Wachstum des World Wide Webs hat es teilweise sehr schwierig gemacht, persönlich relevante Informationen zu finden. Ein Grund dafür ist, daß es den vielfach zum Auffinden von Informationen verwendeten Suchmaschinen nicht möglich ist, Dokumente auf einer semantischen Ebene zu analysieren. Zudem sind Kriterien wie die Qualität von Dokumenten und die Vertrauenswürdigkeit der Anbieter kaum einer automatischen Beurteilung zugänglich (Kapitel 3). Der in dieser Arbeit beschrittene Weg ist es, die Kollaboration zwischen Benutzern des Webs zu ermöglichen und zu fördern. Dadurch soll das Potential menschlichen Urteilsvermögens für Werkzeuge zur Unterstützung bei der Navigation im Web erschlossen werden.

Realisiert wurde in dieser Arbeit das Konzept eines Empfehlungs- und Annotations-systemes für Web-Seiten, welches Benutzern beim Auffinden interessanter, qualitativ hochwertiger Inhalte aus vertrauenswürdigen Quellen assistieren soll. Benutzer des als CoInternet-System bezeichneten Dienstes können bei der Benutzung des Webs kollaborieren, indem sie Web-Seiten bewerten oder mit Annotationen versehen, die beispielsweise Links zu weiterführenden Dokumenten enthalten können. Mittels Collaborative-Filtering-Techniken aggregiert das System abgegebene Wertungen zu individuellen Vorhersagen der zu erwartenden Relevanz einer Seite. Das System selbst erbringt lediglich eine Vermittlerfunktion bei der Zuordnung von Empfehlungen suchenden zu gleichgesinnten bzw. vertrauenswürdigen, Bewertungen abgebenden Personen. Um einer derartigen Aufgabe gerecht zu werden, ist Wissen über die Eigenschaften der Benutzer erforderlich, welches in Benutzerprofilen zu erfassen ist. Das Empfehlungs- und Annotationssystem ist eine Plattform, über die Individuen miteinander kommunizieren, welche die Gemeinschaft der Nutzer bilden. In diesem Zusammenhang ist die Frage nach den Auswirkungen der Integration einer Identitätsinfrastruktur auf eine kollaborative Internetanwendung zu sehen, die in dieser Arbeit bearbeitet wurde.

Die Navigation im Web unter Beachtung von Empfehlungen und Annotationen ist als Social Navigation zu bezeichnen (Kapitel 4). Social Navigation ist Navigation in einem Informationsraum durch Kommunikation mit und Orientierung an anderen Personen. Der Wahrnehmung der Identität der Navigationshinweise anbietenden Personen wurde als ein bestimmender Einflußfaktor sozialen Navigierens identifiziert. Die Entscheidung, ob einem Navigationshinweis Beachtung zu schenken ist, wird davon beeinflusst, ob man deren Urheber vertraut, ob man ihn für kompetent hält und ob Beziehungen zu ihm bestehen. Zudem werden die Regeln in einem Informationsraum und damit die Relevanz von Navigationsinformationen durch den Grad an Identifizierbarkeit der beteiligten Personen geprägt.

Auf Collaborative-Filtering-Techniken (Abschnitt 5.2) basierende Empfehlungssysteme erstellen individuelle Empfehlungen, indem sie die von Benutzern des Systems abgegebenen Bewertungen aggregieren. Sie ermöglichen somit soziales Navigieren. Bei der Aggregation der Bewertungen wird eine individuelle Gewichtung der Stimmen vorgenommen. Aktives Collaborative Filtering ist dadurch gekennzeichnet, daß die bevorzugten Empfehler manuell anzugeben sind. Für die Technik des automatischen Collaborative Filtering ist es kennzeichnend, daß automatisch ähnliche Empfehler ermittelt werden. Eine gängige Vorgehensweise ist es dabei, die Korrelation früherer Bewertungen identischer Objekte zu bestimmen. Automatisches Collaborative Filtering ist mit dem Problem verbunden, daß für einen neu in das System eintretenden Benutzer, der noch keine Bewertungen abgegeben hat, auch keine Empfehlungen berechnet werden können. Ein Umstand, der als Kaltstartproblem bezeichnet wird. Ein auf dem Vergleich abgegebener Bewertungen basierendes Vorgehen bei der automatischen Auffindung ähnlicher Benutzer ist überdies mit der Schwierigkeit verbunden, daß zwei Nutzer nur dann vergleichbar sind, wenn sie eine Menge von Objekten gemeinsam bewertet haben. In einem Empfehlungssystem für Web-Seiten kann davon ausgegangen werden, daß die Bewertungsmatrix immer sehr dünn besetzt sein wird. Ein Aspekt dieses Sparsity-Problems ist, daß der Vergleich von Benutzern schwer fällt, zumal zwei Personen ja auch ähnliche Interessen haben können, sich aber bei unterschiedlichen Anbietern informieren. Automatisches Collaborative Filtering läßt Beziehungen zwischen Benutzern beim Prozeß der Berechnung von Empfehlungen unberücksichtigt, dies ist hingegen die Stärke von aktiven Collaborative-Filtering-Techniken. Derartige Empfehlungssysteme können nach Angabe relevanter Empfehler ohne Initialisierungsphase Empfehlungen berechnen, haben aber den Nachteil, daß sie die Kenntnis der Nutzer untereinander voraussetzen.

Collaborative-Filtering-Systeme nehmen selbst keine Analyse der empfehlbaren Objekte vor. Die Informationsverarbeitung findet in den Köpfen der Anwender statt. Anwender sind somit nicht nur passive Nutzer eines solchen Systemes, sondern tragen auch einen Teil zu seiner Funktionalität bei. Daraus resultieren die Chancen von derartigen Empfehlungssystemen, aber auch einige der Probleme. Kriterien wie subjektives Interesse, Qualität und Vertrauen können auf diesem Weg bei der Berechnung von Empfehlungen

Berücksichtigung finden. Hinderlich ist jedoch, daß für Objekte, für die noch keine Bewertungen vorliegen, auch keine Relevanzvorhersagen gegeben werden können. Der erste Bewerter eines Objektes nimmt folglich einen Aufwand auf sich, ohne daß er selbst bei der Auswahl des Objektes unterstützt wurde. Die Leistung eines derartigen Empfehlungssystems ist abhängig vom Engagement der Benutzer. Wenn zu viele Personen das System lediglich passiv nutzen, so wird es keine zufriedenstellende Leistung erbringen können. In diesem Zusammenhang spricht man auch von Free-Rider-Problem. Soziales Verhalten der Nutzer ist eine Voraussetzung für den Betrieb eines auf Collaborative-Filtering-Techniken basierenden Empfehlungssystems. Annahme bei der Konzeption des CoInternet-Systems war es, daß Benutzer am ehesten zur aktiven Beteiligung zu bewegen sind, wenn das Bestehen einer Gemeinschaft der Nutzer verdeutlicht wird, ein Zugehörigkeitsgefühl zu dieser Gemeinschaft gefördert wird und das Streben nach Reputation innerhalb dieser Gemeinschaft motivierend wirken kann.

Resümierend kann aus der Beschäftigung mit dem Konzept Social Navigation und dessen Ausprägung Collaborative Filtering festgehalten werden, daß einige der Probleme dieser Techniken auf unzureichendes Wissen des Systems über die Eigenschaften der Nutzer bzw. ungenügende Kenntnis der Nutzer untereinander zurückzuführen sind. Unter diesem Gesichtspunkt ergeben sich durch die Integration einer Identitätsinfrastruktur Verbesserungspotentiale.

Die wesentliche Verbesserung gegenüber herkömmlichen Social-Navigation-Anwendungen besteht darin, daß die im Informationsraum hinterlassenen Spuren über eine Identitätsinfrastruktur in Verbindung mit Benutzerprofilen stehen, welche über den Kontext des Empfehlungs- und Annotationssystems hinaus Gültigkeit besitzen. Um den Unzulänglichkeiten eines auf Bewertungsmuster reduzierten Benutzermodells auszuweichen, wurden im CoInternet-System Benutzerprofile vorgesehen, die flexibel aus verschiedenen Attributen zusammengesetzt werden können (Abschnitt 8.2.1). Benutzerprofile können somit in mehreren Dimensionen verglichen werden, dadurch wird die Möglichkeiten der Auffindung ähnlicher Benutzer vergrößert (Abschnitt 8.2.5). Durch die Einsatz der Identitätsinfrastruktur ist es möglich, Benutzerprofile aus verschiedenen Quellen zu speisen, was das Kaltstartproblem relativiert.

Typische Collaborative-Filtering-Systeme präsentieren sich als Black-Box bei der die Entstehung der Empfehlungen nicht nachvollziehbar ist. Dieser Ansatz scheint für ein Empfehlungssystem für Web-Seiten in zweierlei Hinsicht unangemessen zu sein. Zum einen können Benutzer mit höchst unterschiedlichen Aufgaben im Web surfen und das Risiko der Orientierung an Empfehlungen kann dabei stark variieren. Den Benutzern müssen folglich Informationen präsentiert werden, anhand derer sie entscheiden können, welcher Bedeutung sie einer Empfehlung beimessen. Zum anderen macht ein Empfehlungssystem, das sich als Black-Box präsentiert, nicht deutlich, daß Empfehlungen in einem kollaborativen Prozeß entstehen, an dem die Benutzer maßgeblichen Anteil ha-

ben. Wenn einzelne Nutzer nicht sichtbar werden, so ist es auch schwer, ein Gefühl von Gemeinschaft entstehen zu lassen. Im CoInternet-System ist der Empfehlungsprozeß bis hin zur Stimme einzelner Benutzer nachvollziehbar. Durch die Integration der Identitätsinfrastruktur wurde die Interoperabilität mit weiteren Anwendungen ermöglicht. Personen können somit auch nahtlos über andere Anwendungen miteinander interagieren. Die universellen Benutzeridentifikatoren haben zur Folge, daß sich Personen aus anderen Kontexten wiedererkennen können bzw. im CoInternet-System erworbene Reputation auf andere Kontexte übertragbar ist.

Aktives Collaborative Filtering trägt dem Umstand Rechnung, daß die Bedeutung der Bewerter subjektiv unterschiedlich sein kann. Unabhängig von der Ähnlichkeit der Bewertungen in der Vergangenheit kann die Meinung von Experten, Kollegen und Freunden individuell von besonderer Wichtigkeit sein. Aktives Collaborative Filtering setzt die Kenntnis der Benutzer untereinander voraus. Diese Voraussetzung ließ aktives Collaborative Filtering für den Einsatz in Communities ungeeignet erscheinen. Durch die Gestaltung des Empfehlungssystems, welche die Wahrnehmung einzelner Benutzer ermöglicht sowie durch die mit der Integration einer Identitätsinfrastruktur einhergehende Möglichkeit der Interaktion mit anderen Nutzern mittels interoperabler Anwendungen, wurde der Aufbau vertrauensvoller Beziehungen zwischen den Nutzern des Systems vereinfacht. Zudem ist es möglich, daß sich Personen, die sich aus einem anderen Kontext kennen, innerhalb des Systems wiedererkennen. Auf diesem Wege werden die Vorteile von aktivem Collaborative Filtering auch in großen Benutzergruppen nutzbar. In dieser Arbeit wurde der Ansatz gewählt, aktive und automatische Collaborative-Filtering-Techniken zu kombinieren. Vorhersagen werden einmal unter Berücksichtigung automatisch ermittelter Stimmengewichtung berechnet und einmal unter Einbeziehung der durch Bestimmung einer Liste vertrauter Personen explizit gemachten Beziehungen (Abschnitte 8.2.4 und 8.2.5).

Ein System, das es jedem Nutzer erlaubt, Web-Seiten mit für alle anderen Nutzer einsehbaren Annotationen zu versehen, wird damit umgehen müssen, daß es bei viel frequentierten Seiten zu einem unüberschaubaren Übermaß an Annotationen kommen kann, eine sinnvolle Filterung der Annotationen also notwendig erscheint. In dieser Arbeit wurden Annotationen als Web-Seiten abgelegt, die über das CoInternet-System mit den annotierten Dokumenten verknüpft werden. Annotationen können wie gewöhnliche Web-Seiten bewertet und ebenso auch annotiert werden. Die gleichen Verfahren, die zur Berechnung von Relevanzvorhersagen für Web-Seiten verwendet werden, werden auch auf Annotationen angewendet.

Die Bedeutung von Annotationen ist bezogen auf das Ziel, unter den Benutzern des CoInternet-Systems eine Community zu schaffen auch darin zu sehen, daß Annotationen eine Gelegenheit bieten, innerhalb der Gemeinschaft als Individuum wahrgenommen zu werden.

Suchmaschinen basieren auf Information-Retrieval-Techniken und liefern als Antwort auf eine Suchanfrage eine Liste potentiell relevanter Web-Seiten. Angesichts der vielfach unüberschaubar großen Zahl, der zu einer Anfrage aufgelisteten Seiten, kommt den Algorithmen, welche die Treffer in eine Reihenfolge bringen, eine besondere Bedeutung zu. Die Leistung von Suchmaschinen ist bei einigen Anfragen als unzureichend anzusehen, wobei Defizite in der semantischen Informationsverarbeitung offenkundig werden. Doch selbst wenn Suchmaschinen erkennen würden, daß ein Begriff mehrere Bedeutungen hat und selbst wenn sie die Dokumente dahingehend analysieren könnten, daß sie im Stande wären zu entscheiden, welcher der alternativen Bedeutungen das Vorkommen des Suchbegriffes in einem Dokument entspricht, so wären sie doch nicht in der Lage allein aufgrund der Suchanfrage zu beurteilen, an welcher der verschiedenen Bedeutungen der Nutzer interessiert ist. In dieser Arbeit wurde der Ansatz gewählt, das Empfehlungssystem mit einer Suchmaschine zu kombinieren (Abschnitt 8.2.6). Benutzerprofile definieren einen Kontext, der bei der Auswertung von Suchanfragen berücksichtigt wird. Durch die Suchanfragen selektiert die Suchmaschine eine Menge passender Dokumente aus dem Web. Zusätzlich identifiziert das Empfehlungssystem auf Grundlage des Benutzerprofils eine Gruppe relevanter Empfehler, durch deren Bewertungen mittelbar ebenfalls eine Menge von Dokumenten ausgewählt wird. Die Umsortierung der Trefferlisten gemäß den Bewertungsvorhersagen bewirkt, daß an vorderster Position Dokumente aufgelistet werden, die sowohl der Suchanfrage als auch den grundsätzlichen Interessen des Anfragers entsprechen. Auf diesem Weg wurde menschliches Urteilsvermögen für die (automatische) Informationssuche erschlossen.

Die Realisierung des CoInternet-Systems erfolgte basierend auf dem onefC-Identitätsmanager, dem Scone-Framework, dem Tomcat Application Server und dem Axis SOAP Toolkit. Durch die Verwendung fertiger Software-Komponenten konnte eine relativ kurze Entwicklungszeit erreicht werden und die Entwicklungstätigkeit auf die fachlichen Aspekte der Anwendung beschränkt werden. So konnte beispielsweise eine Beschäftigung mit dem Thema der Netzwerkkommunikation unterbleiben.

Die Komponenten zur Generierung impliziter Bewertungen und zur Vorhersage von Bewertungen wurden mittels eines Modul-Konzepts verwirklicht. Der entstandene Prototyp kann somit als Grundlage für weitergehende Forschungsaktivitäten dienen.

Der CoInternet-Client stellt seine Funktionen als Web-Service zur Verfügung. Infolgedessen ist eine Integration der Funktionen des CoInternet-Servers in neue Client-Anwendungen problemlos möglich.

Im Rahmen dieser Arbeit wurde das CoInternet-System einer ersten Evaluation unterzogen (Abschnitt 8.4). Eines der Ergebnisse dabei war, daß Benutzer nur sehr selten negative Bewertungen abzugeben scheinen, da Objekte, die bereits aufgrund des Link-Textes als uninteressant eingestuft werden, gar nicht erst aufgerufen und bewertet werden. Die zur Anzeige der Relevanzvorhersage von Link-Zielen verwendeten Popup-Fenster sind für

das Verhalten, eine Auflistung verlinkter Überschriften nur zu scannen, schlecht geeignet, denn jeder Link muß mit der Maus angefahren werden. Die Bewertung von Web-Seiten besteht nur sekundär in einem motorischen Aufwand. Gravierender scheint die damit verbundene mentale Anstrengung zu sein. Es ist fraglich, ob viele Personen bereit sind, derartige Mühen auf sich zu nehmen. Infolgedessen erscheint die Beschäftigung mit impliziten Bewertungen sehr lohnend.

Der Beitrag, den diese Arbeit im Rahmen des Projektes onefC leistet, ist es, die Bedeutung einer Identitätsinfrastruktur hinsichtlich der Förderung von Kollaboration und Personalisierung aufzuzeigen. Konkret wurde deutlich gemacht, welches Potential eine Identitätsinfrastruktur in Hinblick auf die Verbesserung von Social Navigation bietet. Es wurden Bezüge aufgezeigt, die zwischen dem Konzept Social Navigation und Identitätskonzepten bestehen und wo sich Möglichkeiten für Verbesserungen ergeben. Exemplarisch anhand des realisierten Empfehlungs- und Annotationssystem wurden einige Vorschläge erarbeitet, wie eine Social-Navigation-Anwendung durch die Integration einer Identitätsinfrastruktur zu verbessern sei. In erster Linie sind Benutzerprofile vorzusehen, die auch aus anderen Anwendungen gespeist werden können und die Wahrnehmbarkeit einzelner Benutzer in der Gemeinschaft ist zu fördern, um die Grundlage für Interaktionen zwischen den Nutzern und den Aufbau von Vertrauen zu schaffen.

### 9.2. Offene Fragen und Ausblick

Das in dieser Arbeit entstandene System weist einen Funktionsumfang und eine Stabilität auf, die es für die Durchführung einer Evaluation tauglich machte. Dennoch ist es in einigen Punkten weit davon entfernt, für einen praktischen Einsatz geeignet zu sein.

Eine mögliche Weiterentwicklung des Systems kann in zwei Richtungen erfolgen. Zum einen bietet es sich an, eine Reduzierung auf eine Kernfunktionalität vorzunehmen, die technische Komplexität der Client-Anwendung zu reduzieren und eine derartiges System als Plugin in den Web-Browser zu integrieren. Alternativ könnten jedoch auch mit dem Ziel, die Schaffung von Gruppenbewußtsein und vertrauensvollen Beziehungen zwischen den Nutzern zu fördern, Funktionen in das System integriert werden, welche die durch Analyse der Benutzerprofile und Angabe vertrauter Personen dynamisch gebildete Community der für eine Person relevanten Empfehler veranschaulichen. Zudem könnten Funktionen zur Unterstützung direkter Social Navigation in das System aufgenommen werden.

Der Wert eines Empfehlungssystemes für Web-Seiten ist abhängig von der Anzahl der Benutzer. Eine für den realen Betrieb zu erfüllende Anforderung ist die Skalierbarkeit sowohl bezüglich der Zahl bedienbarer Anwender als auch hinsichtlich der Menge der Dokumente, für die Bewertungsvorhersagen bestimmt werden können. Steigerungen der

Performanz ließen sich durch Optimierung der Vorhersagealgorithmen, aber auch durch Verteilung der Server-Funktionalität auf mehrere Maschinen erzielen. Dieser Aspekt ist auch in Bezug auf die zu erreichende Verfügbarkeit eines derartigen Dienstes von Interesse. Die Anforderungen an die Sicherheit eines Empfehlungssystems sind hoch. Auch wenn Benutzerprofile lediglich pseudonymisiert abgelegt werden, so ist der unberechtigte Zugriff auf diese Daten unbedingt zu unterbinden, denn letztendlich wird es nie ganz ausgeschlossen werden können, daß Rückschlüsse auf die reale Identität eines Benutzers gezogen werden können. Ein Empfehlungssystem bietet ein Angriffsziel für Manipulationen. Durch technische Maßnahmen sollte ein solches böses Vorgehen, beispielsweise das Überfluten des Systems mit Bewertungen, erschwert werden.

Ein auf Collaborative-Filtering-Techniken basierendes System wird nur für diejenigen Seiten Relevanzvorhersagen berechnen können, die bereits bewertet wurden. Die Leistungsfähigkeit eines Empfehlungssystems mit einer dünn besetzten Bewertungsmatrix ist folglich stark eingeschränkt. Wie gravierend das aus diesem Sachverhalt resultierende Sparsity-Problem in der Realität ist, ist abhängig von den Überschneidungen in der Menge der von den Nutzern des Systems besuchten Seiten. Es wäre hinsichtlich der Frage, wie nützlich ein Empfehlungssystem für Web-Seiten in der Praxis sein kann zu untersuchen, wie groß die Überschneidung der Web-Seiten in verschiedenen für den Betrieb eines solchen Systems möglichen Szenarien ist. Dabei wären Arbeitsgruppen und Organisationen ebenso zu untersuchen, wie Communities of Interest im Internet oder Gruppen beliebiger Nutzer, die sich bei einem für alle Web-Nutzer offenstehenden Empfehlungs- und Annotationssystem registriert haben.

Um die Fähigkeit zu erlangen, auch für noch nicht bewertete Seiten Relevanzvorhersagen bestimmen zu können, wäre eine Kombination mit Techniken des Content-based Filtering wünschenswert. Dabei könnte ebenfalls von der Einbeziehung über die Identitätsinfrastruktur übermittelter Profilattribute profitiert werden. Nimmt man an, daß Hyperlinks eine semantische Beziehung von Dokumenten ausdrücken, so wäre es überdies in Erwägung zu ziehen, Bewertungen der verlinkten Dokumente in die Bestimmung von Bewertungsvorhersagen eines noch nicht bewerteten Dokuments einzubeziehen.

Die Visualisierung des Zustandekommens von Bewertungsvorhersagen hat sich während der Evaluation als unzureichend erwiesen. Hier wären weitere Forschungsaktivitäten notwendig, um eine auf einen Blick erkennbare Darstellung zu finden, die zur Erklärung prognostizierter Bewertungen hinzugezogen werden kann. Es hat sich ebenfalls gezeigt, daß die Darstellung der Bewertungsvorhersagen verlinkter Dokumente in Popup-Fenstern verbesserungsbedürftig ist. Zu evaluieren wäre, ob alternative Techniken, wie Overlays und Modifikationen der Dokumente kombiniert mit Link-on-Demand-Techniken, dem Nutzungsverhalten zuträglicher sind.

Die zur Modellierung der Benutzerinteressen gewählten Attribute entstammt im Prinzip einer willkürlichen Festlegung. Ziel war es zu verdeutlichen, daß Collaborative Filtering

basierend auf der Kombination mehrerer Profilattribute möglich ist. Hier wäre zu untersuchen, mit welcher Art der Benutzermodellierung sich die besten Vorhersageergebnisse erzielen lassen. Zusätzlich müßte, sofern es denn andere auf der Identitätsinfrastruktur basierende Anwendungen gebe, eine Festlegung getroffen werden, welche anderen Identitätsdaten in die zum Collaborative-Filtering herangezogenen Benutzerprofile zu integrieren sind.

Um den Aufwand der Benutzung des Systems zu minimieren, ist der implizite Aufbau von Benutzerprofilen, insbesondere die implizite Abgabe von Bewertungen sinnvoll. Die in diesem Bereich realisierten Lösungen sind lediglich als Beispiele zu sehen. Jedoch wurden die Voraussetzungen geschaffen, um derartige Techniken entwickeln und erproben zu können.

Der Vergleich von Benutzerprofilen mit einfachen Methode des Information Retrieval ist als Proof-of-Concept ausreichend, läßt aber noch einigen Spielraum für Optimierungen. Einerseits kann durch Verbesserungen an dieser Stelle die Qualität der Empfehlungen gesteigert werden, andererseits bieten sich hier Ansatzmöglichkeiten, um die Performanz des Systems zu verbessern, beispielsweise durch Gruppierung der Benutzer zu Clustern.

Die Evaluation hat verdeutlicht, daß die Notwendigkeit besteht, Interessensänderungen zu erfassen und den Benutzern eine einfache Anpassung ihrer Benutzerprofile zu gestatten. In diesem Zusammenhang wäre es notwendig, den Benutzern transparent zu machen, welche Daten ihres Benutzerprofils entscheidend für die Auffindung der für eine konkrete Bewertungsvorhersage relevanten Empfehler ist, so daß gezielt Änderungen vorgenommen werden können.

Die Entscheidung, ein Empfehlungs- und Annotationssystem für Web-Seiten erstellen zu wollen, ist eine zu Beginn der Diplomarbeit getroffene Festlegung. Nicht auszuschließen ist, daß andere Formen von Social Navigation in Abhängigkeit von den Aufgaben und Eigenschaften der Benutzer besser geeignet sind.

Die Frage nach den Auswirkungen der Integration einer Identitätsinfrastruktur auf eine kollaborative Internetanwendung wird sich über rein theoretische Argumentationen hinaus nicht in dieser Allgemeinheit fundiert beantworten lassen. Der Versuch, Antworten auf diese Frage zu finden, setzt die Einschränkung auf die Betrachtung konkreter Systeme voraus. Doch selbst bei einer derartigen Herangehensweise sind lediglich Aussagen zu Teilaspekten möglich, die zudem noch von begrenzter Gültigkeit sind.

Die Leistung eines Empfehlungssystems zu beurteilen, gestaltet sich schwierig, denn die Korrektheit einer berechneten Empfehlung ist letztendlich nur durch den jeweiligen Benutzer zu beurteilen. Durchführbar ist die Überprüfung der Vorhersagegenauigkeit, basierend auf statistischen Verfahren. Neben der Abweichung berechneter Werte zu den Bewertungen der Benutzer ist dabei eine Kenngröße der Anteil der angefragten Dokumente, für den eine Vorhersage bestimmt werden konnte. Jedoch ist eine derartige

Untersuchung immer eine Momentaufnahme. Dabei wird nicht berücksichtigt, daß die Leistungsfähigkeit eines Empfehlungssystems einem dynamischen Prozeß unterliegt. Die von den Benutzern wahrgenommene Vorhersagegenauigkeit beeinflusst als ein Faktor die Akzeptanz des Systems und damit deren Verbleiben im System sowie deren Motivation zur Abgabe von Bewertungen. Beispiele für andere Faktoren, die die Akzeptanz des Systems bestimmen können, sind der mit der Benutzung verbundene Aufwand, die Freude an der Benutzung des Systems und die Verbundenheit zur Gemeinschaft der Anwender. Gleichzeitig ist die Leistungsfähigkeit des Systems abhängig von der Mitwirkung der Benutzer. Es besteht folglich eine zyklische Abhängigkeit, die bewirkt wird, daß die Leistungsfähigkeit eines ohnehin schon zufriedenstellend funktionierenden Systems über die Zeit tendenziell ansteigt, währenddessen ein unzureichend arbeitendes System die Nutzer nur schwer dauerhaft an sich binden wird, und somit dessen Fähigkeit im Laufe der Zeit weiter abnimmt.

Die Evaluation eines auf einer Identitätsinfrastruktur basierenden Empfehlungssystems ist methodisch sehr schwierig. Die Stärke eines solchen Systems liegt im Zusammenspiel mit anderen Anwendungen. Unter „Laborbedingungen“ wird man kaum realistische Test-szenarien schaffen können, in denen eine Vielzahl von Benutzern über einen längeren Zeitraum mit verschiedenen Anwendungen arbeitet, zwischen denen durch das Empfehlungssystem eine Interoperabilität besteht. Die Fähigkeit, neuen Benutzern präzise Empfehlungen zu präsentieren, wird durch die übermittelten Profilattribute beeinflusst. In der Realität kann jedoch nicht davon ausgegangen werden, daß alle Anbieter bereit sind, von ihnen aufgebaute Benutzerprofile zu teilen. Ebenso ist es unwahrscheinlich, daß der überwiegende Teil der Benutzer den vorbehaltlosen Zugriff auf die von einem Empfehlungssystem auswertbaren Attribute seiner Identität gestattet. Die Frage, ob innerhalb eines Empfehlungssystems Vertrauen zu anderen Nutzern aufgebaut wird, ob ein Zugehörigkeitsgefühl zur Gemeinschaft besteht und ob sich dieses positiv auf die Motivation der Benutzer auswirkt, wird sich ebenso wie die Frage nach dem Aufbau von Beziehungen zu anderen Nutzern und deren Verwendung zum aktiven Collaborative Filtering empirisch untersuchen lassen. Die Ergebnisse werden jedoch stark von der Zusammensetzung der Testgruppe geprägt, denn nicht nur die Eigenschaften jedes einzelnen Nutzers sind relevant, sondern auch das soziale Gefüge innerhalb der Nutzerschaft.

Um nicht nur zu Aussagen bezüglich einer auf der Identitätsinfrastruktur aufsetzenden Anwendung zu gelangen, sondern isoliert die Auswirkungen der Integration betrachten zu können, wäre eine vergleichende Studie von Anwendungen mit einem propriärem Identitätskonzept und Anwendungen, die eine Identitätsinfrastruktur verwenden, notwendig. Abgesehen davon, daß es, wie vorstehend aufgeführt wurde, diffizil ist, Social-Navigation-Anwendungen zu evaluieren, so ist es kaum möglich, den von der Integration der Identitätsinfrastruktur ausgelösten Effekt isoliert zu betrachten, da zu viele Einflußfaktoren bestehen. Zwei bis auf die unterschiedliche Realisierung der Identitätskonzepte identische Anwendungen zu betrachten und isoliert von anderen Anwendungen zu evalu-

ieren, erscheint nicht sinnvoll, da in diesem Fall die aus der Interoperabilität mit anderen Applikationen resultierende Stärke einer Identitätsinfrastruktur unberücksichtigt bleiben würde. Nimmt man jedoch als Vergleichsapplikation ein System, das mit der Maßgabe entworfen wurde, aus der Identitätsinfrastruktur Nutzen zu ziehen, wie es beim Entwurf des CoInternet-Systems der Fall war, so wird man nicht abschätzen können, welcher Einfluß letztendlich aus der Integration einer Identitätsinfrastruktur herrührt und welcher auf die Spezifika des Entwurfes zurückzuführen ist. Aus vorstehend genannten Gründen erscheint zudem die Evaluation eines derartigen Systems kaum möglich, bei der eine Kombination aus mehreren Anwendungen betrachtet wird und Fähigkeiten, wie das Wiedererkennen von Benutzern, zur Geltung kommen.

Durch das CoInternet-System wurde eine Grundlage geschaffen, um die Kollaboration bei der Benutzung des World Wide Webs zu ermöglichen und damit den Nutzern zu helfen, die Vielfalt der in diesem Medium gebotenen Informationen für sich zu erschließen. Aus Sicht des Autors bietet sich jedoch ein weiteres Problem, welches es den Anwendern schwer macht, die immer größer werdenden Möglichkeiten der Informationsbeschaffung auch wirklich gewinnbringend einzusetzen. Viele der für einen Benutzer relevanten Informationen sind auf dessen Rechner in diversen Ablagen verteilt. Eventuell verfügt eine Person sogar über verschiedene Rechner und mobile Geräte, was dieses Problem noch verstärkt. Anwender haben interessante Web-Seiten in Bookmarks hinterlegt, haben News-Groups und News-Letter abonniert, wichtige E-Mails abgelegt und Dateien in einer hierarchischen Verzeichnisstruktur organisiert. Mit der wachsenden Zahl persönlich bedeutsamer Informationen wird es dabei immer schwieriger, diese dem eigenen Arbeitsstil entsprechend abzulegen, so daß ein effizienter, an Aufgaben orientierter Zugriff gewährleistet ist.

# Abbildungsverzeichnis

2.1. Intensität des Informationsflusses in Arbeitsgruppen . . . . .	11
2.2. Klassifizierung nach dem 3K-Modell . . . . .	13
2.3. Ebay Feedback-Forum . . . . .	27
2.4. Übersicht über die onefC-Architektur . . . . .	31
2.5. UML-Klassendiagramm des onefC-Identitätsmodells . . . . .	33
2.6. Architektur des Scone-Frameworks . . . . .	34
2.7. Bearbeitung eines TokenStreams durch zwei Editor-Plugins . . . . .	35
2.8. Schematischer Ablauf des <i>UserTrackings</i> . . . . .	37
2.9. Beispiel einer HyperScout-Einblendung . . . . .	39
4.1. Schematischer Ablauf von Social Navigation . . . . .	53
4.2. Metriken für die Definition von Nähe zwischen Web-Nutzern . . . . .	60
4.3. Schema der Visualisierung von Pfaden im Web durch den Footprints- Prototyp . . . . .	61
4.4. Bewertungssystem für Meldungen auf der Web-Site des ZDF . . . . .	64
4.5. Generierung von Empfehlungen basierend auf der Auswertung von Bookmark-Listen . . . . .	65
4.6. PHP Annotated Manual . . . . .	66
4.7. Einflußfaktoren sozialen Navigierens . . . . .	68
5.1. Zusammenhang zwischen Social Navigation und Recommender-Systemen	79
5.2. Klassifikation von Collaborative-Filtering-Verfahren . . . . .	83
5.3. Anschauliche Erklärung von Collaborative Filtering . . . . .	85
5.4. Collaborative-Filtering-Prozeß . . . . .	85
5.5. Wie wird Benutzer Ken Artikel 6 beurteilen? . . . . .	87
5.6. Berechnung der vorhergesagten Bewertung des Artikels 6 für den Benutzer Ken . . . . .	89
5.7. GroupLens FilterBots . . . . .	94
5.8. Ausgabe von Empfehlungen basierend auf dem Inhalt des Warenkorbes im Web-Shop Amazon . . . . .	95
5.9. Durch das PHOAKS-System ermittelte Liste empfohlener Web- Dokumente zu einer Newsgroup . . . . .	97

6.1. Das WebPlaces-Fenster . . . . .	101
6.2. Platzierung von Intermediaries im Informationsfluß zwischen dem Produzenten und Konsumenten von Informationen . . . . .	103
6.3. Architektur des Grouplens-Systems . . . . .	104
6.4. Benutzungsoberfläche eines zur Darstellung von Bewertungsvorhersagen angepaßten News-Readers . . . . .	105
6.5. Seitentypen einer hierarchisch strukturierten Web-Site . . . . .	108
8.1. Das System Collaborative Internet Experience bietet eine Plattform zur Kommunikation von Bewertungen und Annotationen. . . . .	117
8.2. Konzeptionelles Modell des Empfehlungs- und Annotationssystems als UML-Klassendiagramm . . . . .	123
8.3. Profil-Editor . . . . .	126
8.4. Aufteilung der Bewertungsmatrix durch Partitionierung des Objektraumes	128
8.5. Aufteilung der Bewertungsmatrix durch Verwendung mehrerer Profile . .	128
8.6. Explizite Abgabe von Bewertungen . . . . .	129
8.7. Web of Trust . . . . .	133
8.8. Repräsentation von Attributen der Benutzerprofile als Vektoren . . . . .	135
8.9. Schema der Berechnung von Bewertungsvorhersagen . . . . .	137
8.10. Präsentation einer Bewertungsprognose . . . . .	140
8.11. Einblendung der prognostizierten Bewertung des Ziel-Objekts eines Hyperlinks mittels eines modifizierten HyperScout-Popups . . . . .	141
8.12. Auflistung aller zu einem Dokument abgegebenen Bewertungen . . . . .	143
8.13. Ausgabe aller Bewertungen und Annotationen eines Benutzers . . . . .	144
8.14. Schema der Kombination des CoInternet-Systems mit einer Suchmaschine	146
8.15. Ergebnisseite der Suche nach „Robots“ unter Einbeziehung des Profils eines an Internet-Themen interessierten Benutzers . . . . .	147
8.16. Ergebnisseite der Suche nach „Robots“ unter Einbeziehung des Profils eines an Robotik interessierten Benutzers . . . . .	148
8.17. Screenshot des CoInternet-Clients . . . . .	150
8.18. Architekturvarianten bei der Realisierung einer Funktionseinheit zur Bewertungsvorhersage . . . . .	153
8.19. Das CoInternet-System könnte als ein in Web-Sites einzubindender Dienst realisiert werden. . . . .	155
8.20. Client-Server-Architektur des CoInternet-Systems . . . . .	157
8.21. Klassendiagramm der entwickelten Module zur Bestimmung von Bewertungsprognosen . . . . .	161
8.22. Schema der Berechnung von Bewertungsvorhersagen unter Einbeziehung mehrerer Module durch die PredictionEngine . . . . .	162

8.23. Realisierung der Kombination eines Empfehlungssystems mit einer Suchmaschine unter Einbeziehung der über die Google SOAP-API verfügbar gemachten Dienste der Suchmaschine Google . . . . .	165
8.24. Klassendiagramm der in die Auswertung von Suchmaschinenabfragen involvierten Klassen . . . . .	168
8.25. Interface <code>AccessAnalyser</code> . . . . .	169
8.26. Schnittstelle des <code>CoInternetServer</code> . . . . .	171
8.27. Klassendiagramm der an den CoInternet-Server zu sendenden Requests .	172
8.28. Klassendiagramm der an den Client übermittelten Responses . . . . .	173
8.29. Verhältnis von Aufwand und Nutzen bei Benutzertests in Abhängigkeit von der Größe der Testgruppe . . . . .	175
8.30. Durchführung der Evaluation . . . . .	176
8.31. Austausch von Identitätsdaten außerhalb des CoInternet-Systems . . . . .	185

## Tabellenverzeichnis

4.1. Klassifikation von Social Navigation . . . . .	56
8.1. Für die Berechnung von Ähnlichkeiten herangezogene <i>explizit</i> anzugeben- de Attribute eines Benutzerprofils . . . . .	125
8.2. Für die Berechnung von Ähnlichkeiten herangezogene <i>implizit</i> erhobene Attribute eines Benutzerprofils . . . . .	127

## Literaturverzeichnis

- [AM96] ACKERMAN, MARK S. und DAVID W. McDONALD: *Answer Garden 2: merging organizational memory with collaborative help*. In: *Proceedings of the 1996 ACM conference on Computer supported cooperative work*, Seiten 97–105, Boston, Massachusetts, United States, 1996. ACM Press.
- [ARH00] ABDUL-RAHMAN, ALFAREZ und STEPHEN HAILES: *Supporting Trust in Virtual Communities*. In: *Proceedings of the Hawaii International Conference on System Sciences*, Maui, Hawaii, 2000. IEEE Computer Society.
- [Axe84] AXELROD, R.: *The Evolution of Cooperation*. Basic Books, New York, 1984.
- [Bai89] BAIER, J. H.: *Supporting Cooperative Work with Computers: Addressing Meeting Mania*. In: *IEEE Intellectual Digest of Papers, COMPCON '89*, 1989.
- [BC92] BELKIN, NICHOLAS J. und W. BRUCE CROFT: *Information filtering and information retrieval: two sides of the same coin?* *Communications of the ACM*, 35(12):29–38, 1992.
- [BH97] BENYON, DAVID und KRISTINA HÖÖK: *Navigation in Information Spaces: supporting the individual*. In: *Proceedings of the IFIP TC13 Interantional Conference on Human-Computer Interaction*, Seiten 39–46. Chapman & Hall, Ltd., 1997.
- [BHI93] BLY, SARA A., STEVE R. HARRISON und SUSAN IRWIN: *Media spaces: bringing people together in a video, audio, and computing environment*. *Communications of the ACM*, 36(1):28–46, 1993.
- [BHK98] BREESE, JACK, DAVID HECKERMAN und CARL KADIE: *Empirical Analysis of Predictive Algorithms for Collaborative Filtering*. In: *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, Madison, WI, Juli 1998. Morgan Kaufmann Publisher.

- [BL94] BERNERS-LEE, T.: *RFC 1630: Universal Resource Identifiers in WWW: A Unifying Syntax for the Expression of Names and Addresses of Objects on the Network as used in the World-Wide Web*, Juni 1994.
- [BLC95] BERNERS-LEE, T. und D. CONNOLLY: *RFC 1866: Hypertext Markup Language — 2.0*, November 1995.
- [BLMM94] BERNERS-LEE, T., L. MASINTER und M. MCCAHILL: *RFC 1738: Uniform Resource Locators (URL)*, Dezember 1994.
- [BM99] BARRETT, R. und P. P. MAGLIO: *Intermediaries: An approach to manipulating information streams*. IBM Systems Journal, 38(4):629–641, 1999.
- [BS97] BALABANOVIC, MARKO und YOAV SHOHAM: *Fab: content-based, collaborative recommendation*. Communications of the ACM, 40(3):66–72, 1997.
- [BS98] BORGHOFF, UWE und JOHANN SCHLICHTER: *Rechnergestützte Gruppenarbeit – Eine Einführung in Verteilte Anwendungen*. Springer Verlag, Berlin, 2. Auflage, 1998.
- [Buc02] BUCHMANN, VOLKERT: *Multiple Links im World-wide web: Entwicklung eines Client/ Server-Navigationswerkzeuges mit dem Framework Scone*. Studienarbeit, Universität Hamburg, Fachbereich Informatik, Arbeitsbereich Verteilte Systeme und Informationssysteme (VSIS), Hamburg, November 2002.
- [Bus96] BUSH, VANNEVAR: *As we may think*. interactions, 3(2):35–46, 1996.
- [BZL03] BAIER, TOBY, CHRISTIAN ZIRPINS und WINFRIED LAMERSDORF: *Digital Identity: How To Be Someone On The Net*. In: REIS, PEDRO ISAIAS ANTONIO PALMA DOS (Herausgeber): *e-Society 2003*, Band II, Seiten 815–820, Lisbon, Portugal, Juni 2003. IADIS International association for development of the information society, IADIS Press.
- [C<sup>+</sup>03] COLE, JEFFREY I. et al.: *THE UCLA INTERNET REPORT – Surveying the Digital Future*. Technischer Bericht, UCLA Center for Communication Policy, Los Angeles, Januar 2003.
- [CGM<sup>+</sup>99] CLAYPOOL, M., A. GOKHALE, T. MIRANDA, P. MURNIKOV, D. NETES und M. SARTIN: *Combining Content-Based and Collaborative Filters in an Online Newspaper*. In: *Proceedings of ACM SIGIR Workshop on Recommender Systems*, Berkeley, CA, August 1999. ACM Press.
- [CLA<sup>+</sup>03] COSLEY, DAN, SHYONG K. LAM, ISTVAN ALBERT, JOSEPH A. KONSTAN und JOHN RIEDL: *Is seeing believing?: how recommender system interfaces*

- affect users' opinions.* In: *Proceedings of the conference on Human factors in computing systems*, Seiten 585–592, Ft. Lauderdale, Florida, USA, 2003. ACM Press.
- [CLWB01] CLAYPOOL, MARK, PHONG LE, MAKOTO WASED und DAVID BROWN: *Implicit interest indicators.* In: *Proceedings of the 6th international conference on Intelligent user interfaces*, Seiten 33–40, Santa Fe, New Mexico, United States, 2001. ACM Press.
- [Con87] CONKLIN, JEFF: *Hypertext: An Introduction and Survey.* IEEE Computer, 20(9):17–40, September 1987.
- [CP95] CATLEDGE, LARA D. und JAMES E. PITKOW: *Characterizing browsing strategies in the World-Wide Web.* In: *Proceedings of the Third International World-Wide Web conference on Technology, tools and applications*, Seiten 1065–1073, Darmstadt, Germany, 1995. Elsevier North-Holland, Inc.
- [Dan00] DANIS, CATALINA M.: *Extending the concept of awareness to include static and dynamic person information.* ACM SIGGROUP Bulletin, 21(3):59–62, 2000.
- [Dav98] DAVENPORT, T.: *Working knowledge.* Harvard Business School Press, Boston, 1998.
- [DB92] DOURISH, PAUL und VICTORIA BELLOTTI: *Awareness and coordination in shared workspaces.* In: TURNER, J. und R.KRAUT (Herausgeber): *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, Seiten 107–114, Toronto, Ontario, Canada, 1992. ACM Press.
- [DC94] DOURISH, PAUL und MATTHEW CHALMERS: *Running Out of Space: Models of Information Navigation.* In: *Proceedings of Human Computer Interaction (HCI'94)*, Glasgow, UK, August 1994.
- [DDH<sup>+</sup>00] DIEBERGER, A., P. DOURISH, K. HÖÖK, P. RESNICK und A. WEXELBLAT: *Social navigation: techniques for building more usable systems.* interactions, 7(6):36–45, 2000.
- [DFM01] DINGLEDINE, ROGER, MICHAEL J. FREEDMAN und DAVID MOLNAR: *Accountability.* In: ORAM, ANDY (Herausgeber): *Peer-to-Peer: Harnessing the Power of Disruptive Technologies*, Kapitel 16, Seiten 271–340. O'Reilly & Associates, Inc., Sebastopol, CA, 1. Auflage, März 2001.
- [Die97] DIEBERGER, ANDREAS: *Supporting Social Navigation on the World Wide Web.* International Journal of Human-Computer Studies, 46(6):805–825, 1997.

- [Die03] DIEBERGER, ANDREAS: *Social connotations of space in the design for virtual communities and social navigation*. In: HÖÖK, KRISTINA, DAVID BENYON und ALAN J. MUNRO (Herausgeber): *Designing information spaces: the social navigation approach*, Seiten 293–313. Springer-Verlag, London, 2003.
- [DL00] DIEBERGER, ANDREAS und PETER LÖNNQVIST: *Visualizing interaction history on a collaborative web server*. In: *Proceedings of the eleventh ACM on Hypertext and hypermedia*, Seiten 220–221, San Antonio, Texas, United States, 2000. ACM Press.
- [Don97] DONATH, J.: *Identity and Deception in the Virtual Community*. In: KOLLOCK, P. und M. SMITH (Herausgeber): *Communities in Cyberspace: Perspectives on New Forms of Social Organization*. University of California Press, Berkeley, 1997.
- [Dör99] DÖRING, NICOLA: *Identität und Internet*, Seiten 255–313. Hogrefe, Göttingen, 1999.
- [DS73] DOWNS, ROGER und DAVID STEA: *Cognitive Maps and Spatial Behaviour: Process and Products*. In: DOWNS, ROGER und DAVID STEA (Herausgeber): *Image and Environment*, Seiten 8–26. Aldine, Chicago, 1973.
- [Dwo03] DWORSCHAK, MANFRED: *Sumpfhühner im Anflug*. SPIEGEL online, November 2003. URL, [www.spiegel.de/spiegel/0,1518,272384,00.html](http://www.spiegel.de/spiegel/0,1518,272384,00.html).
- [EGF03] EIMEREN, BIRGIT VAN, HEINZ GERHARD und BEATE FREES: *ARD/ZDF-Online-Studie 2003 – Internetverbreitung in Deutschland: Unerwartet hoher Zuwachs*. Media Perspektiven, August 2003.
- [EGR91] ELLIS, CLARENCE A., SIMON J. GIBBS und GAIL REIN: *Groupware: some issues and experiences*. Communications of the ACM, 34(1):39–58, 1991.
- [EK03] ERICKSON, THOMAS und WENDY A. KELLOGG: *Social translucence: using minimalist visualisations of social activity to support collective interaction*. In: HÖÖK, KRISTINA, DAVID BENYON und ALAN J. MUNRO (Herausgeber): *Designing information spaces: the social navigation approach*, Seiten 17–41. Springer-Verlag, London, 2003.
- [Eri96] ERICKSON, THOMAS: *The World-Wide-Web as social hypertext*. Communications of the ACM, 39(1):15–17, 1996.
- [FGM<sup>+</sup>99] FIELDING, R., J. GETTYS, J. MOGUL, H. FRYSTYK, L. MASINTER, P. LEACH und T. BERNERS-LEE: *RFC 2616: Hypertext transfer protocol – HTTP*, 1999.

- [FHH<sup>+</sup>00] FISHER, DANYEL, KRIS HILDRUM, JASON HONG, MARK NEWMAN, MEGAN THOMAS und RICH VUDUC: *SWAMI (poster session): a framework for collaborative filtering algorithm development and evaluation*. In: *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, Seiten 366–368, Athens, Greece, 2000. ACM Press.
- [FHS98] FORSBERG, MATTIAS, KRISTINA HÖÖK und MARTIN SVENSSON: *Design Principals of Social Navigation*. In: *Proceedings of the 4th ERCIM Workshop on User Interfaces for All (UI4All'98)*, Stockholm, Schweden, Oktober 1998.
- [Fis00] FISHER, BY DANYEL: *Implementation of Web Places*. Diplomarbeit, UC Berkeley, Mai 2000.
- [For98] FORSBERG, J. MATTIAS: *Social Navigation: An Extended Definition*. URL, [www.nada.kth.se/~forsberg/Documents/SocNav.pdf](http://www.nada.kth.se/~forsberg/Documents/SocNav.pdf), Juli 1998.
- [GAD97] GLANCE, N., D. ARREGUI und M. DARDENNE: *Knowledge Pump: Community-centred Collaborative Filtering*. In: *Proceedings of the 5th DELOS Workshop on Filtering and Collaborative Filtering*, Budapest, November 1997.
- [Gam90] GAMBETTA, D.: *Can We Trust Trust?*, Kapitel 13, Seiten 213–237. Basil Blackwell, Oxford, 1990.
- [GHJ95] GAMMA, ERICH, RICHARD HELM und RALPH JOHNSON: *Design Patterns. Elements of Reusable Object-Oriented Software*. Addison-Wesley Professional Computing Series. Addison-Wesley, Reading, Massachusetts, 1995.
- [GM03] GERHARDS, MARIA und ANNETTE MENDE: *ARD/ZDF-Offline-Studie 2003 – Offliner 2003: Stabile Vorbehalte gegenüber dem Internet*. Media-Perspektiven, August 2003.
- [GNOT92] GOLDBERG, DAVID, DAVID NICHOLS, BRIAN M. OKI und DOUGLAS TERRY: *Using collaborative filtering to weave an information tapestry*. *Communications of the ACM*, 35(12):61–70, 1992.
- [Har03] HARPER, R. H. R.: *Information that counts: a sociological view of information navigation*. In: HÖÖK, KRISTINA, DAVID BENYON und ALAN J. MUNRO (Herausgeber): *Designing information spaces: the social navigation approach*, Seiten 343–353. Springer-Verlag, London, 2003.
- [Haß02] HASS, TORSTEN: *Java in heterogenen Umgebungen: Konzepte der Interaktion zwischen Javaprogrammen und Web-Browsern zur Erfassung von Benutzeraktionen*. Studienarbeit, Universität Hamburg, Fachbereich Informatik,

- Arbeitsbereich Verteilte Systeme und Informationssysteme (VSIS), Hamburg, September 2002.
- [HBM03] HÖÖK, KRISTINA, DAVID BENYON und ALAN J. MUNRO: *Editors' introduction: footprints in the snow*. In: HÖÖK, KRISTINA, DAVID BENYON und ALAN J. MUNRO (Herausgeber): *Designing information spaces: the social navigation approach*, Seiten 1–13. Springer-Verlag, London, 2003.
- [HD96] HARRISON, STEVE und PAUL DOURISH: *Re-place-ing space: the roles of place and space in collaborative systems*. In: *Proceedings of the 1996 ACM conference on Computer supported cooperative work*, Seiten 67–76, Boston, Massachusetts, United States, 1996. ACM Press.
- [HHWM92] HILL, WILLIAM C., JAMES D. HOLLAN, DAVE WROBLEWSKI und TIM MCCANDLESS: *Edit wear and read wear*. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, Seiten 3–9, Monterey, California, United States, 1992. ACM Press.
- [HKR00] HERLOCKER, JONATHAN L., JOSEPH A. KONSTAN und JOHN RIEDL: *Explaining collaborative filtering recommendations*. In: *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, Seiten 241–250, Philadelphia, Pennsylvania, United States, 2000. ACM Press.
- [HLSW00] HÖÖK, KRISTINA, JARMO LAAKSOLAHTI, MARTIN SVENSSON und ANNIKA WAERN: *Individual Differences in Social Navigation*. In: *Workshop on Social Navigation: a design approach? (CHI'2000)*, Hague, Niederlande, 2000.
- [HPPB01] HOSCHKA, PETER, WOLFGANG PRINZ und UTA PANKOKE-BABATZ: *Der Computer als soziales Medium*. In: SCHWABE, GERHARD, NORBERT STREITZ und RAINER UNLAND (Herausgeber): *CSCW-Kompendium*. Springer Verlag, Berlin, 2001.
- [HRCV01] HOYOS-RIVERA, GUILLERMO DE JESÚS, JEAN-PIERRE COURTIAT und THIERRY VILLEMUR: *A Design Framework for Collaborative Browsing*. In: *Proceedings of the 10th IEEE International Workshops on Enabling Technologies*, Seiten 362–367, Cambridge, MA, 2001. IEEE Computer Society.
- [HS96] HUDSON, SCOTT E. und IAN SMITH: *Techniques for addressing fundamental privacy and disruption tradeoffs in awareness support systems*. In: *Proceedings of the 1996 ACM conference on Computer supported cooperative work*, Seiten 248–257, Boston, Massachusetts, United States, 1996. ACM Press.
- [IPCB97] IMPICCIATORE, PIERO, CHIARA PANDOLFINI, NICOLA CASELLA und MAURIZIO BONATI: *Reliability of health information for the public on the world*

- wide web: systematic survey of advice on managing fever in children at home.* British Medical Journal, 314(7098), 1997.
- [Joh88] JOHANSEN, R.: *Groupware: Computer Support for Business Teams.* Free Press, New York, 1988.
- [Joh03] JOHNSON, STEVEN: *Digging for Googleholes.* URL, [slate.msn.com/id/2085668](http://slate.msn.com/id/2085668), Juli 2003.
- [JZ00] JANETZKO, DR. DIETMAR und DIRK ZUGENMAIER: *Viele Gesichter - Personalisierte Websites stellen sich auf Besucher ein.* c't, (18):88, 2000.
- [Kar03] KARZAUNINKAT, STEFAN: *Google zugemüllt - Spam überschwemmt die Suchergebnisse.* c't, (20):88-91, 2003.
- [KBM<sup>+</sup>99] KANTOR, PAUL, ENDRE BOROS, BEN MELAMED, DAVID J. NEU, VLADIMIR MEŇKOV, QIN SHI und MYUNG-HO KIM: *Ant World. (A brief one-page description of Ant World.).* In: *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, Seite 323, Berkeley, California, United States, 1999. ACM Press.
- [KBM<sup>+</sup>00] KANTOR, PAUL B., ENDRE BOROS, BENJAMIN MELAMED, VLADIMIR MEŇKOV, BRACHA SHAPIRA und DAVID J. NEU: *Capturing human intelligence in the net.* Commun. ACM, 43(8):112-115, 2000.
- [KK01] KAHAN, JOSÉ und MARJA-RITTA KOIVUNEN: *Annotea: an open RDF infrastructure for shared Web annotations.* In: *Proceedings of the tenth international conference on World Wide Web*, Seiten 623-632, Hong Kong, Hong Kong, 2001. ACM Press.
- [KL86] KANTOR, B. und P. LAPSLEY: *RFC 977: Network News Transfer Protocol: A Proposed Standard for the Stream-Based Transmission of News*, Februar 1986. Status: PROPOSED STANDARD.
- [KMM<sup>+</sup>97] KONSTAN, JOSEPH A., BRADLEY N. MILLER, DAVID MALTZ, JONATHAN L. HERLOCKER, LEE R. GORDON und JOHN RIEDL: *GroupLens: applying collaborative filtering to Usenet news.* Communications of the ACM, 40(3):77-87, 1997.
- [KMW00] KOCH, MICHAEL, KATHRIN MÖSLEIN und MICHAEL WAGNER: *Vertrauen und Reputation in Online-Anwendungen und virtuellen Gemeinschaften.* In: ENGELIEN, MARTIN und DETLEF NEUMANN (Herausgeber): *Virtuelle Organisation und Neue Medien - Workshop Gemeinschaften in Neuen Medien (GeNeMe2000)*, Seiten 69-84, Dresden, Oktober 2000. Eul Verlag, Lohmar.

- [Koc01a] KOCH, MICHAEL: *Community-Support-Systeme*. In: SCHWABE, GERHARD, NORBERT STREITZ und RAINER UNLAND (Herausgeber): *CSCW-Kompodium*, Seiten 286–296. Springer Verlag, Berlin, 2001.
- [Koc01b] KOCH, MICHAEL: *Kollaboratives Filtern*. In: SCHWABE, GERHARD, NORBERT STREITZ und RAINER UNLAND (Herausgeber): *CSCW-Kompodium*, Seiten 351–357. Springer Verlag, Berlin, 2001.
- [Koc02] KOCH, MICHAEL: *Global Identity Management to Boost Personalization*. In: SCHUBERT, PETRA und PETRA UND UWE LEIMSTOLL (Herausgeber): *Proceedings of the Ninth Research Symposium on Emerging Electronic Markets*, Seiten 137–147, Basel, Switzerland, September 2002. Institute for Business Economics, University of Applied Sciences Basel.
- [KR03] KONSTAN, JOSEPH A. und JOHN RIEDL: *Collaborative filtering: supporting social navigation in large, crowded infospaces*. In: HÖÖK, KRISTINA, DAVID BENYON und ALAN J. MUNRO (Herausgeber): *Designing information spaces: the social navigation approach*, Seiten 43–82. Springer-Verlag, London, 2003.
- [KT00] KOBAYASHI, MEI und KOICHI TAKEDA: *Information retrieval on the web*. ACM Computing Surveys, 32(2):144–173, 2000.
- [Kun03] KUNZE, CHRISTIAN P.: *Digitale Identität und Identitäts-Management*. Diplomarbeit, Universität Hamburg, Fachbereich Informatik, Arbeitsbereich Verteilte Systeme und Informationssysteme (VSIS), Hamburg, 2003.
- [KW01] KOCH, MICHAEL und WOLFGANG WÖRNDL: *Community Support and Identity Management*. In: *Proc. Europ. Conference on Computer-Supported Cooperative Work (ECSCW2001)*, Seiten 319–338, Bonn, Deutschland, September 2001.
- [Lie00] LIECHTI, OLIVIER: *Awareness and the WWW: an overview*. ACM SIGGROUP Bulletin, 21(3):3–12, 2000.
- [LV03] LYMAN, PETER und HAL R. VARIAN: *How Much Information*. Technischer Bericht, School of Information Management and Systems, University of California at Berkeley, Berkeley, 2003.
- [MAIO97] MYNATT, ELIZABETH D., ANNETTE ADLER, MIZUKO ITO und VICKI L. O'DAY: *Design for network communities*. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, Seiten 210–217, Atlanta, Georgia, United States, 1997. ACM Press.

- 
- [MB99] MAGLIO, PAUL P. und ROB BARRETT: *WebPlaces: Adding People to the Web*. In: *Proceedings of the Eighth International World Wide Web Conference*, Toronto, Kanada, 1999.
- [MB00] MAGLIO, PAUL und ROB BARRETT: *Intermediaries personalize information streams*. *Communications of the ACM*, 43(8):96–101, 2000.
- [MBF03] MAGLIO, PAUL P., ROB BARRETT und STEPHEN FARRELL: *Webplaces: using intermediaries to add people to the Web*. In: HÖÖK, KRISTINA, DAVID BENYON und ALAN J. MUNRO (Herausgeber): *Designing information spaces: the social navigation approach*, Seiten 249–269. Springer-Verlag, London, 2003.
- [MC96] MCKNIGHT, D. H. und N. L. CHERVANY: *The Meaning of Trust*. Technischer Bericht 94-04, Carlson School of Management, University of Minnesota, 1996.
- [ME95] MALTZ, DAVID und KATE EHRlich: *Pointing the way: active collaborative filtering*. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, Seiten 202–209, Denver, Colorado, United States, 1995. ACM Press/Addison-Wesley Publishing Co.
- [MGT<sup>+</sup>87] MALONE, THOMAS W, KENNETH R GRANT, FRANKLYN A TURBAK, STEPHEN A BROBST und MICHAEL D COHEN: *Intelligent information-sharing systems*. *Communications of the ACM*, 30(5):390–402, 1987.
- [MNS00] MENKOV, VLADIMIR, DAVE NEU und QIN SHI: *AntWorld: A Collaborative Web Search Tool*. In: *Proceedings of the Third International Workshop on Distributed Communities on the Web*, Seiten 13–22, Quebec City, Canada, 2000. Springer-Verlag.
- [MRK97] MILLER, BRADLEY N., JOHN T. RIEDL und JOSEPH A. KONSTAN: *Experience with GroupLens: Making Usenet Useful Again*. In: USENIX (Herausgeber): *1997 Annual Technical Conference, January 6–10, 1997. Anaheim, CA, USA*, Seiten 219–233, Berkeley, CA, USA, 1997. USENIX.
- [Nie93] NIELSEN, JAKOB: *Usability Engineering*. Academic Press, Cambridge, MA, 1993.
- [Nie94] NIELSEN, JAKOB: *Guerilla HCI: Using Discount Usability Engineering to Penetrate the Intimidation Barrier*. URL, [www.useit.com/papers/guerrilla\\_hci.html](http://www.useit.com/papers/guerrilla_hci.html), 1994.
- [Nie98] NIELSEN, JAKOB: *Jakob Nielsen's Alertbox for February 8, 1998: The Reputation Manager*. URL, [www.useit.com/alertbox/980208.html](http://www.useit.com/alertbox/980208.html), Februar 1998.

- [Ols98] OLSSON, TOMAS: *Decentralised Social Filtering based on Trust*. In: *The Working Notes of the AAAI-98 Recommender Systems Workshop*, Madison, Wisconsin, 1998.
- [PBMW98] PAGE, LAWRENCE, SERGEY BRIN, RAJEEV MOTWANI und TERRY WINOGRAD: *The PageRank Citation Ranking: Bringing Order to the Web*. Technischer Bericht, Stanford Digital Library Technologies Project, 1998.
- [PRW00] PICOT, A., R. REICHWALD und R. WIGAND: *Die grenzenlose Unternehmung. Information, Organisation und Management*. Gabler Verlag, Wiesbaden, 4. Auflage, 2000.
- [RIS<sup>+</sup>94] RESNICK, PAUL, NEOPHYTOS IACOVOU, MITESH SUCHAK, PETER BERGSTROM und JOHN RIEDL: *GroupLens: an open architecture for collaborative filtering of netnews*. In: *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, Seiten 175–186, Chapel Hill, North Carolina, United States, 1994. ACM Press.
- [RKZF00] RESNICK, PAUL, KO KUWABARA, RICHARD ZECKHAUSER und ERIC FRIEDMAN: *Reputation systems*. *Communications of the ACM*, 43(12):45–48, 2000.
- [RP97] RUCKER, JAMES und MARCOS J. POLANCO: *Siteseer: personalized navigation for the Web*. *Communications of the ACM*, 40(3):73–76, 1997.
- [RV97] RESNICK, PAUL und HAL R. VARIAN: *Recommender systems*. *Communications of the ACM*, 40(3):56–58, 1997.
- [SH03] SVENSSON, MARTIN und KRISTINA HÖÖK: *Social navigation of food recipes: designing Kalas*. In: HÖÖK, KRISTINA, DAVID BENYON und ALAN J. MUNRO (Herausgeber): *Designing information spaces: the social navigation approach*, Seiten 201–222. Springer-Verlag, London, 2003.
- [SHLW01] SVENSSON, MARTIN, KRISTINA HÖÖK, JARMO LAAKSOLAHTI und ANNIKA WAERN: *Social navigation of food recipes*. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, Seiten 341–348, Seattle, Washington, United States, 2001. ACM Press.
- [SJP99] STEINFELD, CHARLES, CHYNG-YANG JANG und BEN PFAFF: *Supporting virtual team collaboration: the TeamSCOPE system*. In: *Proceedings of the international ACM SIGGROUP conference on Supporting group work*, Seiten 81–90, Phoenix, Arizona, United States, 1999. ACM Press.

- [SKB<sup>+</sup>98] SARWAR, BADRUL M., JOSEPH A. KONSTAN, AL BORCHERS, JON HERLOCKER, BRAD MILLER und JOHN RIEDL: *Using filtering agents to improve prediction quality in the GroupLens research collaborative filtering system*. In: *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, Seiten 345–354, Seattle, Washington, United States, 1998. ACM Press.
- [SKKR01] SARWAR, BADRUL, GEORGE KARYPIS, JOSEPH KONSTAN und JOHN REIDL: *Item-based collaborative filtering recommendation algorithms*. In: *Proceedings of the tenth international conference on World Wide Web*, Seiten 285–295, Hong Kong, Hong Kong, 2001. ACM Press.
- [SKR99] SCHAFER, J. BEN, JOSEPH KONSTAN und JOHN RIEDI: *Recommender systems in e-commerce*. In: *Proceedings of the 1st ACM conference on Electronic commerce*, Seiten 158–166, Denver, Colorado, United States, 1999. ACM Press.
- [SKX98] SCHLICHTER, JOHANN H., MICHAEL KOCH und CHENGMAO XU: *Awareness - The Common Link Between Groupware and Community Support Systems*. In: ISHIDA, TORU (Herausgeber): *Community Computing and Support Systems*, Seiten 77–93. Springer Verlag, Juni 1998.
- [SM83] SALTON, GERARD und MICHAEL J. MCGILL: *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc., New York, 1983.
- [SM95] SHARDANAND, UPENDRA und PATTIE MAES: *Social information filtering: algorithms for automating „word of mouth“*. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, Seiten 210–217, Denver, Colorado, United States, 1995. ACM Press/Addison-Wesley Publishing Co.
- [SSW97] SIDLER, G., A. SCOTT und H. WOLF: *Collaborative Browsing in the World Wide Web*. In: *Proceedings of the 8th Joint European Networking Conference*, Edinburgh, Schottland, May 1997.
- [Sve98] SVENSSON, MARTIN: *Social Navigation*. In: DAHLBÄCK, NILS (Herausgeber): *Exploring Navigation; Towards a Framework for Design and Evaluation of Navigation in Electronic Spaces*, Kapitel 6, Seiten 73–88. Swedish Institute of Computer Science, Kista, Schweden, 1998.
- [Sve00] SVENSSON, MARTIN: *Defining and Designing Social Navigation*. Licenciate Thesis, University of Stockholm, Stockholm, Schweden, 2000.
- [SVW85] STREETER, LYNN A., DIANE VITELLO und SUSAN A. WONSIEWICZ: *How to tell people where to go: comparing navigational aids*. *International Journal of Man-Machine Studies*, 22(5):549–562, 1985.

- [TG97] TAUSCHER, LINDA und SAUL GREENBERG: *How people revisit web pages: empirical findings and implications for the design of history systems*. Int. J. Hum.-Comput. Stud., 47(1):97–137, 1997.
- [TH01] TERVEEN, LOREN und WILL HILL: *Beyond Recommender Systems: Helping People Help Each Other*. In: CARROLL, JOHN M. (Herausgeber): *Human-Computer Interaction in the New Millenium*. Addison Wesley, New York, 2001.
- [THA+97] TERVEEN, LOREN, WILL HILL, BRIAN AMENTO, DAVID McDONALD und JOSH CRETER: *PHOAKS: a system for sharing recommendations*. Communications of the ACM, 40(3):59–62, 1997.
- [TSMB95] TEUFEL, S., C. SAUTER, T. MÜHLHERR und K. BAUKNECHT: *Computer-unterstützte Gruppenarbeit*. Addison-Wesley, Bonn, 1995.
- [TSS96] TOLLMAR, KONRAD, OVIDIU SANDOR und ANNA SCHÖMER: *Supporting social awareness @ work design and experience*. In: *Proceedings of the 1996 ACM conference on Computer supported cooperative work*, Seiten 298–307, Boston, Massachusetts, United States, 1996. ACM Press.
- [Tve01] TVEIT, AMUND: *Peer-to-peer based recommendations for mobile commerce*. In: *Proceedings of the 1st international workshop on Mobile commerce*, Seiten 26–29, Rome, Italy, 2001. ACM Press.
- [UF98] UNGAR, L. und D. FOSTER: *Clustering Methods For Collaborative Filtering*. In: *Proceedings of the Workshop on Recommendation Systems*, Menlo Park California, 1998. AAAI Press.
- [WBL03] WEINREICH, HARALD, VOLKERT BUCHMANN und WINFRIED LAMERSDORF: *Scone: Ein Framework zur evaluativen Realisierung von Erweiterungen des Webs*. In: K. IRMSCHER, K.-P. FÄHNRIK (Herausgeber): *Ta-gungsband Kommunikation in Verteilten Systemen - KiVS 2003*, Informatik Aktuell, Seiten 31–42, Leipzig, Februar 2003. Springer-Verlag.
- [Wex99] WEXELBLAT, ALAN: *History-Based Tools for Navigation*. In: *Proceedings of the Thirty-Second Annual Hawaii International Conference on System Sciences-Volume 2*, Seite 2033. IEEE Computer Society, 1999.
- [Wex03] WEXELBLAT, ALAN: *Results from the footprints project*. In: HÖÖK, KRISTINA, DAVID BENYON und ALAN J. MUNRO (Herausgeber): *Designing information spaces: the social navigation approach*, Seiten 223–248. Springer-Verlag, London, 2003.

- [Wil86] WILSON, R.: *Reputations in games and markets*. In: ROTH, ALVIN E. (Herausgeber): *Game-theoretic models of bargaining*. Cambridge University Press, Cambridge, 1986.
- [WL00] WEINREICH, HARALD und WINFRIED LAMERSDORF: *Concepts for improved visualization of Web link attributes*. In: *Proceedings of the 9th International WWW Conference.*, Amsterdam, Niederlande, 2000. Elsevier.
- [WM97] WEXELBLAT, ALAN und PATTI MAES: *Footprints: History-rich web browsing*. In: *Proceedings of the Conference on Computer-Assisted Information Retrieval (RIA0)*, Seiten 75–84, Montreal, Quebec, 1997.
- [WM99] WEXELBLAT, ALAN und PATTIE MAES: *Footprints: history-rich tools for information foraging*. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, Seiten 270–277, Pittsburgh, Pennsylvania, United States, 1999. ACM Press.
- [WOL01] WEINREICH, H., H. OBENDORF und W. LAMERSDORF: *The Look of the Link - Concepts for the User Interface of Extended Hyperlinks*. In: *Proceedings of the 12th ACM Conference on Hypertext and Hypermedia*, Seiten 19–28, Aarhus, Denmark, August 2001. The Association for Computing Machinery, ACM.
- [Wol02] WOLLENWEBER, FRANK: *Entwicklung eines generischen Robots für das Scone-Framework*. Studienarbeit, Universität Hamburg, Fachbereich Informatik, Arbeitsbereich Verteilte Systeme und Informationssysteme (VSIS), Hamburg, Juni 2002.
- [WOL03] WEINREICH, HARALD, HARTMUT OBENDORF und WINFRIED LAMERSDORF: *HyperScout: Darstellung erweiterter Typinformationen im World Wide Web – Konzepte und Auswirkungen*. In: JÜRGEN ZIEGLER, GERD SZWILLUS (Herausgeber): *Mensch und Computer 2003*, Berichte des German Chapter of the ACM, Seiten 155–164, Stuttgart, September 2003. B.G. Teubner Verlag.
- [Yee02] YEE, KA-PING: *CritLink: Advanced Hyperlinks Enable Public Annotation on the Web*. In: *Demonstration abstract submitted to CSCW 2002*, New Orleans, Louisiana, USA, November 2002.

## A. Inhalt der CD-ROM

Die beiliegende CD-ROM enthält die das komplette Collaborative Internet Experience System einschließlich aller Quelltexte und zum Betrieb notwendiger Komponenten.

### **Pfad:**

/readme.txt . . . . .	Anmerkungen zur CD-ROM
/cointernet . . . . .	Quelltexte und Binärdateien des CoInternet-Systems
/onefc . . . . .	Quelltexte und Binärdateien der onefC-Identitätsinfrastruktur
/scone . . . . .	Quelltexte und Binärdateien des Scone-Frameworks einschließlich des modifizierten HyperScout-Plugins
/software . . . . .	Zum Betrieb des CoInternet-Systems notwendige Zusatzsoftware