

# Personalisierung der Internetsuche – Lösungstechniken und Marktüberblick

Kai Riemer, Fabian Brüggemann  
*Westfälische Wilhelmsuniversität  
Institut für Wirtschaftsinformatik  
Leonardo-Campus 3  
D-48149 Münster  
wikari@wi.uni-muenster.de*

**Abstract.** Ziel dieses Beitrags ist es, einen State-of-the-Art-Überblick der Forschung zur Suchpersonalisierung zu geben, Techniken zur Personalisierung der Informationssuche zu systematisieren und den Markt der Suchdienste auf deren Verbreitung zu untersuchen. Hierzu werden ausgehend von typischen Problemen im Prozess der webbasierten Suche die wesentlichen Personalisierungstechniken identifiziert; diese werden genutzt, um 20 Suchdienste zu klassifizieren. Die Ergebnisse der Marktuntersuchung werden diskutiert und Implikationen für die weitere Forschung werden abgeleitet.

**Keywords:** Suchpersonalisierung, Personalisierungstechniken, Internetsuche

## Einleitung

Informationen gewinnen in der heutigen Wissensgesellschaft immer mehr an Bedeutung. Das Internet scheint die Informationsbedürfnisse aus Anwendersicht auf den ersten Blick zu lösen: Es ist jederzeit verfügbar, nicht lokal begrenzt und Informationen stehen in nahezu beliebiger Menge zur Verfügung. Suchmaschinen – browserbasierte Programme mit denen gezielt nach Begriffen gesucht werden kann – bilden das Rückgrat der Informationssuche im Internet [20; 34]. Allerdings kann das Hauptbedürfnis des Suchenden, die in der Situation notwendigen Information in der richtigen Qualität und Granularität zu finden und dabei einen möglichst geringen Aufwand erbringen zu müssen [77], häufig nicht befriedigt werden. Suchmaschinen stehen vor der Herausforderung, aus Milliarden von Dokumenten mit einer Suchanfrage von durchschnittlich 1,7 Suchworten die für den Nutzer passenden Treffer zu finden [11]. Dabei müssen sie zudem ein möglichst breites Spektrum möglicher Nutzerinteressen berücksichtigen. Die typische Ergebnisliste einer Suchanfrage führt beim Nutzer entsprechend oft zu einem Problem des „Information Overload“ [1], so dass er die für ihn relevanten Informationen zunächst mühsam selbst herausfiltern muss. Eine Möglichkeit zur Verbesserung des Problems liegt in der Personalisierung der Informationssuche. Hierzu verfolgt der vorliegende Beitrag zwei Ziele. Zum einen soll ein State-of-the-Art-Überblick der Forschung zu Techniken und Verfahren der Suchpersonalisierung gegeben werden. Zum anderen wird der Markt für

Suchmaschinen auf die Verbreitung dieser Techniken hin untersucht. Der Beitrag wird zunächst in den Kontext der Suchmaschinenforschung eingeordnet.

## 1. Abgrenzung und Einordnung des vorliegenden Beitrags

Die Suchmaschinenforschung hat ihre Wurzeln in der Bibliothekswissenschaft, die sich im Wesentlichen mit der Organisation und Klassifikation großer Datenbestände, den Techniken für das Auffinden von Informationen und dem Nutzerverhalten bei der Informationssuche beschäftigt. Mit der Popularität des Internets hat sich die Suchmaschinenforschung in den Bereich des E-Commerce und der Wirtschaftsinformatik ausgedehnt. Der relevante Gegenstandsbereich der Informationssuche (der sog. Korpus) ist in diesem Fall das WWW [37]. In Anlehnung an Hevner et al. lässt sich die Suchmaschinenforschung in gestaltungsorientierte und erkenntnistheoretische Arbeiten unterscheiden [31]; siehe Abbildung 1.

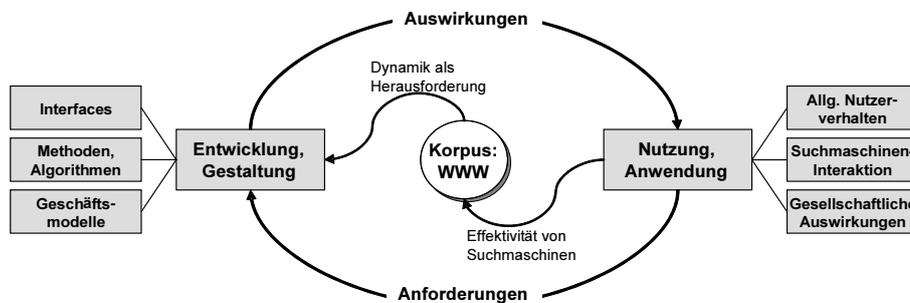


Abbildung 1. Übersicht über Bereiche in der Suchmaschinenforschung

Gestaltungsorientierte Studien beschäftigen sich mit der Entwicklung von Suchmaschinen. Hierzu gehört neben der Entwicklung von Suchalgorithmen [1] auch das Design von Nutzer-Interfaces. Letztes hat zum Ziel, die Interaktion bei der Eingabe der Suchanfrage und der Nutzung der Ergebnisse zu verbessern; diese Forschung baut auf Usability-Studien auf [29]. Schließlich beschäftigt sich ein recht neuer Zweig innerhalb des E-Commerce mit Geschäftsmodellen für Suchdienste; hier ist im Wesentlichen die Werbefinanzierung von Suchergebnissen (das sog. Paid Placement) zu nennen [9]. Insgesamt steht die Gestaltung von Web-Suchmaschinen vor besonderen Herausforderungen durch die Art des zugrunde liegenden Korpus. Dieser wurde im Gegensatz z.B. zu Bibliothekssystemen nicht mit Blick auf die Informationssuche gestaltet [64]. Das Web wächst vielmehr dezentral und unterliegt einer besonderen Dynamik in mehreren Dimensionen: der verfügbare Inhalt und die Zahl der Seiten wächst schnell (jeden Tag kommen 60 Terabytes an Daten hinzu [71]), die Inhalte ändern sich zudem teils täglich und Seiten werden wieder aus dem Netz genommen [37].

Im Bereich der erkenntnistheoretischen Studien sind zunächst solche zu nennen, die sich allgemein mit dem Nutzerverhalten bei der Informationssuche beschäftigen; einen aktuellen Überblick bietet [83]. Den Kern bilden jedoch Studien zur Suchmaschineninteraktion und Nutzerwahrnehmung [43]. Populär ist dabei die Auswertung großer Datenbestände von Suchanfragen, um hieraus Erkenntnisse zum

Nutzerverhalten zu ziehen [42; 72]; einen guten Überblick gibt [52]. Weitere Studien setzen sich mit den gesellschaftlichen Effekten von Internetsuchmaschinen auseinander, z.B. mit dem Thema Bias und möglichen Auswirkungen eines mangelnden Wettbewerbs zwischen Suchmaschinenanbietern [56]. In diesem Zusammenhang sind besonders auch Studien zur Effektivität von Suchmaschinen zu nennen, die sich mit der Abdeckung der Suchdienste in Bezug auf den Korpus Web beschäftigen, siehe z.B. [6; 25; 27; 32].

Aus den Ergebnissen zum Nutzerverhalten und zum Einsatz von Suchmaschinen lassen sich die Anforderungen für deren Entwicklung und Gestaltung ableiten [83]. Gleichzeitig bildet die Gestaltung von Suchmaschinen (z.B. von Prototypen) die Grundlage für die Erforschung des Nutzungsverhaltens, siehe z.B. [62]. Die Zusammenhänge werden in Abbildung 1 visualisiert. Der vorliegende Beitrag lässt sich wie folgt in diesen Ordnungsrahmen einordnen: Zunächst werden aus erkenntnistheoretischen Arbeiten gewonnene Ergebnisse zur Nutzung von Suchmaschinen dazu genutzt, typische Probleme der Informationssuche im Internet zu identifizieren (siehe Kap. 2). Diese Probleme können als Anforderungen an die Gestaltung von Suchmaschinen verstanden werden, die durch Personalisierungstechniken adressiert werden sollen. Den Kern der Studie bildet ein strukturierter Überblick über Techniken und Verfahren zur Personalisierung von Suchmaschinen in Kapitel 3. Die beschriebenen Techniken und Verfahren wiederum finden Eingang in die Praxis; hierzu werden in Kapitel 4 die Ergebnisse einer Marktuntersuchung vorgestellt. Kapitel 5 diskutiert abschließend Implikationen für die weitere Forschung.

## **2. Probleme bei der Informationssuche**

Um den Sinn und Zweck der Personalisierung von Suchmaschinen zu verdeutlichen, ist es zunächst sinnvoll, typische Probleme der Informationssuche aus Anwendersicht aufzuzeigen, die von den Techniken zur Personalisierung adressiert werden. Sowohl die Probleme, wie auch die Techniken in Kapitel 3 werden anhand eines typischen Suchprozesses strukturiert.

### *2.1. Ablauf eines typischen Suchprozesses*

Der im Weiteren genutzte Suchprozess geht auf Arbeiten zurück, die sich mit dem Verhalten bei der Informationssuche in elektronischen Verzeichnissen [51] und im Web befassen [15]. Dieser allgemeine Verhaltensprozess der Informationssuche kann in Anlehnung an [35] auf die Nutzung von Suchmaschinen übertragen werden, wobei im Einzelnen sechs Phasen unterschieden werden (siehe auch Abb. 2): In der Initiierungsphase (1) realisiert der Suchende sein Informationsbedürfnis, in der Auswahlphase (2) wählt er eine geeignete Suchmaschine bzw. Suchdomäne und in der Formulierungsphase (3) stellt er die konkrete Suchanfrage. In der nun folgenden Durchführungsphase (4) wird die Anfrage von der Suchmaschine verarbeitet; sie führt zu einer Liste von Suchergebnissen, die in der Präsentationsphase (5) angezeigt wird. Der Suchende kann dann die gewünschten Ergebnisse auswählen und in der Nutzungsphase (6) lesen und verarbeiten.

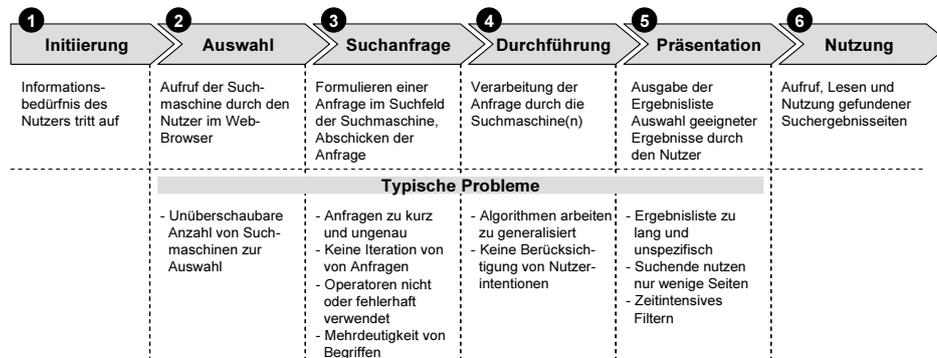


Abbildung 2. Typischer Suchprozess und Probleme des Suchenden

## 2.2. Probleme aus Nutzersicht im Prozess der Informationssuche

Im oben beschriebenen Suchprozess beziehen sich die erste und letzte Phase auf den Nutzungskontext der eigentlichen Informationssuche. Die im Weiteren erläuterten Probleme beziehen sich dagegen konkret auf die mittleren vier Phasen, in denen die eigentliche Interaktion mit der Suchmaschine stattfindet.

### 2.2.1. Probleme bei der Auswahl von Suchmaschine und Suchdomäne

Bereits bei der Auswahl einer geeigneten Suchmaschine stellt sich ein nicht unerhebliches Suchproblem. Es gibt mittlerweile weltweit mehr als 70.000 Suchdienste und Datenbanken [20], von denen sich viele auf ein bestimmtes Themengebiet konzentrieren [77]. Es kann also sinnvoll sein, eine auf eine konkrete Suchdomäne spezialisierte Suchmaschine auszuwählen. Zudem decken auch die großen Suchmaschinen jeweils nur einen Teil des Internets ab [56], so dass eine Kombination von Suchdiensten für eine Anfrage sinnvoll erscheint [20]. Eine solche Kombination ist die Aufgabe von Meta-Suchmaschinen, die ihre Anfragen parallel an mehrere (teils spezialisierte) Suchmaschinen leiten [14].

### 2.2.2. Probleme bei der Formulierung der Suchanfrage

Um gute Resultate zu erhalten, ist es notwendig, dass der Suchende die richtigen Suchwörter in der richtigen Kombination in das Suchfeld der Suchmaschine eingibt [46; 70]. Untersuchungen haben jedoch gezeigt, dass dies meist nicht der Fall ist [4], weil z.B. der Suchende zu wenig Wissen über eine Suchdomäne oder die genutzte Suchmaschine hat, um eine Suchanfrage sinnvoll zu formulieren [68; 71]. Zudem stellen zwei Drittel der User meist nur eine einzige Suchanfrage, ohne diese zu variieren [33; 72]. Dabei sind die meisten Anfragen zudem zu unspezifisch und kurz, d.h. es werden zu wenige Begriffe verwendet [4; 1; 24; 40]. So besteht eine Suchanfrage im Durchschnitt aus ca. 2 Begriffen und 62 % aller Anfragen enthalten nur ein oder zwei Wörter [19; 22; 33; 72]. Auch werden die Möglichkeiten zur Nutzung von Operatoren („AND“/„OR“) meist nicht oder nur fehlerhaft genutzt, wie Log-Analysen zeigen [72; 83]. Aufgrund der enormen Anzahl von verfügbaren Seiten im

Internet reichen solche Anfragen meist nicht, um eine präzise und handhabbare Ergebnismenge zu liefern [34]. Das Problem wird dabei durch die Mehrdeutigkeit von Begriffen noch verschärft, so dass Ergebnisse zu verschiedenen Themengebieten im Suchergebnis auftauchen können [63]. Eine Anfrage zur Raubkatze Jaguar liefert z.B. auch Ergebnisse zur entsprechenden Automarke [71; 85]. Diesen Erkenntnissen entsprechend gibt regelmäßig eine Mehrzahl von Suchenden in Umfragen an, nicht mit dem Sucherlebnis und den gefundenen Ergebnissen zufrieden zu sein [64].

### *2.2.3. Probleme durch die Art der Suchalgorithmen*

Identisch formulierte Suchanfragen liefern immer dieselben Ergebnisse, auch wenn die Suchenden unterschiedliche Absichten verfolgen [73]. Zur Determinierung der Relevanz von Suchtreffern nutzen Suchmaschinen so genannte PageRank-Algorithmen [1]. Das Ziel der Suchmaschinenbetreiber ist es im Normalfall, diese Algorithmen so einzustellen, dass eine möglichst große Bandbreite von unterschiedlichen Nutzerabsichten mit einem Suchergebnis berücksichtigt wird [79]. Entsprechend wird die gleiche Ergebnismenge von Anwendern mit unterschiedlichen Absichten bei der Informationssuche als unterschiedlich relevant empfunden [75]. Eine Einflussnahme auf den Einsatz bestimmter Algorithmen sowie auf die Kriterien, nach denen die Suchtreffer sortiert werden sollen, ist dem Nutzer i. d. R. nicht möglich [77], da die Algorithmen meist von den Suchmaschinenbetreibern als Geschäftsgeheimnis gehütet werden [1]. Die Nutzer könnten jedoch daran interessiert sein, bestimmte Teile ihrer Suchanfrage explizit stärker zu gewichten und somit die subjektive Relevanz zu erhöhen.

### *2.2.4. Probleme mit der Präsentation der Suchergebnisse*

Als Ergebnis präsentieren Suchmaschinen eine nach Relevanz geordnete Auflistung von Links zu Internetseiten, die die gesuchten Informationen enthalten sollten. Da die Produktion neuer Informationen immer kostengünstiger wird, erhöht sich das Volumen verfügbarer Informationen im Internet stetig [78]. Entsprechend werden die Nutzer bei ihrer Suche mit Informationen überschwemmt, von denen sie jedoch nur einen Bruchteil benötigen („Information Overload“). Viele Nutzer beachten dabei nur die ersten angezeigten Ergebnisse [40]. So wurde in einer Untersuchung festgestellt, dass 58 % der Nutzer nur die Ergebnisse der ersten Seite der Suchergebnisse und weitere 19% lediglich zusätzlich die der zweiten Seite betrachtet haben [33]. Auf diese Weise gehen relevante Informationen verloren, da das Angebot zu groß ist und die relevanten Informationen nicht erkennbar sind [4]. Meist fehlt dem Nutzer die Zeit, die relevanten Informationen auszufiltern, was zu Frustration führen kann [69; 73; 78].

Die Ausführungen machen deutlich, dass die klassische Informationssuche mittels Suchmaschinen mehrere Nachteile für den Nutzer hat. Einerseits hat der Nutzer Probleme bei der Formulierung der Anfrage, zudem wird er mit einer Flut von Informationen in der Ergebnismenge konfrontiert, die immer noch eine weiter gehende manuelle Suche erfordern. Darüber hinaus bleibt ein Gefühl von Unsicherheit, ob die relevanten oder besten Informationen überhaupt gefunden wurden. Einige der genannten Probleme könnten durch eine bessere Schulung der Suchenden vermieden werden. Da sich die Internetsuche jedoch zunehmend als Massenphänomen präsentiert und weite Nutzerkreise betrifft, arbeiten sowohl Forscher wie auch Hersteller an

technischen Lösungen, die die Nutzer gezielt unterstützen und entlasten sollen [11]. Hierbei stehen insbesondere die im Weiteren vorgestellten Personalisierungstechniken im Mittelpunkt.

### 3. Personalisierungstechniken

Mit Personalisierung wird allgemein das Anpassen eines Objekts an die Bedürfnisse eines Subjekts bezeichnet [65]. Personalisierung im Internetkontext kann spezifischer als Anpassung von Webseiten an die Bedürfnisse, Präferenzen und Fähigkeiten einzelner Nutzer definiert werden [54; 65; 69; 82]. Ziel der Personalisierung von Suchdiensten ist es, jedem Nutzer die für ihn relevanten Informationen so aufzubereiten, dass sie in der für ihn geeigneten Form vorliegen [10]. Dabei kann zwischen verschiedenen Ebenen der Personalisierung unterschieden werden, z.B. zwischen inhaltlicher Personalisierung sowie Personalisierung der Web-Site hinsichtlich Navigation und Layout [36; 50; 53; 65]. In Anlehnung an diese Unterteilung kann die Personalisierung der Informationssuche in die folgenden drei Ebenen unterteilt werden:

1. Personalisierung der Suchverfahren: Diese bildet den Kern der Suchpersonalisierung und deckt damit die Phasen drei und vier im Suchprozess ab. Hierzu gehören entsprechend Verfahren zur Personalisierung der Suchanfrage des Nutzers, sowie die Personalisierung auf Algorithmebene der Suchmaschine.
2. Personalisierung der Suchmaschine: Diese umfasst zusätzlich die Phasen zwei und fünf und bezieht sich auf weitere Techniken, wie eine personalisierte Einschränkung des Suchraumes vor Formulierung der Anfrage, sowie eine personalisierte Organisation der Suchergebnisliste (Navigation) und der Darstellung der Ergebnisse auf der Web-Site (Layout) [58].
3. Personalisierung des Sucherlebnisses: Diese sehr weite Auslegung des Personalisierungsbegriffs geht über die im Rahmen dieses Beitrags behandelten Techniken hinaus und wird nur in Tabelle 1 skizziert. Hierzu zählen z.B. Features, die dem Nutzer erlauben, den Aufruf der Suchfunktion aus seinem Arbeitskontext heraus persönlich anzupassen, z.B. über die Nutzung von Browser-Toolbars (Phase 1), sowie die Möglichkeit Suchergebnisse zu speichern, um die Art und Weise der Ergebnisnutzung an die eigenen Bedürfnisse anzupassen (Phase 6).

Die in den folgenden Unterkapiteln beschriebenen Personalisierungstechniken werden wiederum anhand der Phasen des Suchprozesses organisiert; eine Übersicht aller Techniken mit kurzen Beschreibungen und Literaturhinweisen findet sich in Tabelle 1. Die meisten Personalisierungstechniken basieren im Kern auf der Erstellung von Nutzerprofilen und der Erhebung persönlicher Daten der Nutzer. Deshalb wird im Weiteren zunächst kurz auf Nutzerprofile eingegangen.

### 3.1. Nutzerprofile als Basis der Personalisierung

Daten über den Nutzer und seine Präferenzen sind die Grundlage der automatisierten Personalisierung [44; 47]. Diese Daten werden in Profilen gespeichert, die vom Personalisierungssystem abgefragt und analysiert werden können [10; 18; 24]. Auf Basis des Nutzerprofils kann z.B. die Anzahl der für die Nutzer in Frage kommenden Themengebiete reduziert oder die Ergebnismenge anders organisiert und präsentiert werden [27].

#### 3.1.1. Explizite Gewinnung von Nutzerdaten

Bei der expliziten Datengewinnung wird dem Nutzer die Möglichkeit gegeben, seine Wünsche und Vorlieben zu äußern und so sein Profil direkt zu konfigurieren [54]. Diese Aufgabe kann durch das Ausfüllen von Fragebögen oder durch die Bewertung von Interessenkategorien gelöst werden [27]. Zudem könnte der Benutzer seinen Interessen entsprechende Webseiten (z.B. Bookmarks) bei der Suchmaschine registrieren, die dann Daten über diese Webseiten ins Nutzerprofil übernimmt [24]. Die explizite Profilerstellung hat den Vorteil, dass die Anwender ihre Präferenzen direkt angeben und sie die Kontrolle über die Personalisierung behalten können [3]. Andererseits ist dieser Ansatz problematisch, da Anwender oft nicht in der Lage sind, ihre Präferenzen richtig anzugeben [76], oder schlicht den zusätzlichen Aufwand ablehnen [24]. Auch veralten einmal hinterlegte Präferenzen im Zeitablauf [3].

#### 3.1.2. Implizite Gewinnung von Nutzerdaten

Die implizite Gewinnung von Nutzerdaten geschieht in der Regel durch die automatisierte Beobachtung des Nutzerverhaltens [48; 54]. Dabei kann sich die Beobachtung auf verschiedene Aspekte des Nutzerverhaltens beziehen [76]. Zum Einen kann die Suchhistorie des Nutzers ausgewertet werden, um Präferenzen zu ermitteln [24]. Diese Vorgehensweise liefert gute Daten für die Personalisierung, da Nutzer ihre Suchaktivitäten oftmals wiederholen [30]. Zum Anderen kann auch das Surfverhalten des Nutzers, wie die Zeit für das Lesen einer Seite, die Anzahl der Besuche, die Abstände zwischen Besuchen, sowie das Speichern oder Drucken einer Seite ausgewertet werden [23; 24]. Seine Interessen und Präferenzen können dann auf Basis der Worthäufigkeiten in den als wichtig erkannten Internetseiten abgeleitet werden [30]. Die Veränderung der Nutzerinteressen im Zeitablauf kann bei der impliziten Profilerstellung besser abgebildet werden, da die kontinuierliche Überprüfung und Speicherung des Nutzerverhaltens die Erstellung von dynamischen Profilen unterstützt [18; 62]. Da die Generierung der Daten automatisch geschieht, können die Nutzerprofile mit sehr geringen Kosten erstellt werden [16; 67; 82]. Es soll jedoch angemerkt werden, dass diese Modelle mit Unsicherheit behaftet sind. Die aus den Aktionen der Nutzer gezogenen Rückschlüsse können irreführend sein, da nicht alle vom Nutzer durchgeführten Aktionen notwendigerweise Ausdruck seiner persönlichen Präferenzen sind [17], z.B. wenn Internetrecherchen für andere Personen durchgeführt werden.

**Tabelle 1.** Zusammenfassende Übersicht der Personalisierungstechniken

Phase	Verfahren	Beschreibung	Quellen
1	Zugang über Toolbar	Initiierung der Suche aus dem Arbeitskontext heraus	
	Zugriff auf die Suchhistorie	Kombinationen von Suchbegriffen können wieder verwendet werden	
2	Automatisierte Vorschlagsfunktion	Gezielter Vorschlag von Treffern ohne eigentliche Suche	[53]
		Erstellen von Indexseiten	[60]
	Auswahl und Eingrenzung des Suchraums	Suchmaschinenauswahl, Selektion von Themenbereichen	[14; 44]
3	Iterative Anfrageverbesserung	Einordnung in thematischen Kontext und Ableitung verwandter Begriffe	[49; 66]
		Zweistufige Suche mit Feedback (Relevance Feedback)	[7; 21; 43; 68; 81]
	Profilbasierte Anfrageverbesserung (insbes. Lokalisierung)	Auf Basis des Nutzerprofils werden Suchwörter hinzugefügt	[22]
	Anfrageverbesserung mittels Kategorisierung	Suchanfragen werden zur Verfeinerung mit individualisierter Hierarchie von Kategorien abgeglichen	[47; 62], siehe auch [37; 57]
	Anfrageverbesserung mittels logischer Deduktion	Reformulierung der Suche durch Einordnung der Anfrage in ein System von logischen Aussagen	[46]
4	PageRank-Personalisierung	Themensensitiver PageRank auf Basis des ODP-Katalogs	[28]
		Personalisierter PageRank durch Höhergewichtung persönlicher Seiten	[12; 34]
	Collaborative Filtering	Relevanz von Seiten berechnet auf Basis von Ähnlichkeiten zwischen Nutzerprofilen	[73; 74], siehe auch [2; 55]
	Suchnetzwerke	PageRank gewichtet Seiten, Bookmarks von Netzwerkmitgliedern höher	[5; 73]

5	Ergebnis-Reranking	Höhere Gewichtung von Seiten, die in ein Interessengebiet im Profil fallen	[24]
		Basierend auf Aktualität/Popularität	[62]
		Nutzung von Themenbäumen	[57; 79]
		Kollaborativ basierend auf Gemeinschaftsprofilen	[71]
	Clusterung und interaktive Neuordnung	Clusterung und Sortierung der Ergebnismenge anhand von Themen	[42; 85]
6	Navigation & Layout	Individualisierung von Bedienung und Aussehen der Suchmaschine	
	Speicherung der Ergebnisse	Erlaubt eine spätere Nutzung und die Bildung persönlicher Linkseiten	

### 3.2. Personalisierung vor der eigentlichen Anfrage

#### 3.2.1. Automatisierte Vorschlagsfunktion

Diese Funktion nimmt eine Sonderstellung ein. Sie ist eigentlich nicht dem Suchprozess zuzuordnen, da hierbei einzelne Webseiten aktiv vom System vorgeschlagen werden, ohne dass eine Suche gestartet werden muss [2]; sie basiert auf der Erkenntnis, dass Nutzer oft keine genaue Vorstellung davon haben, welche Art von Information sie genau suchen [8]. Die Auswahl der vom System vorgeschlagenen Webseiten kann dabei aus den im Profil hinterlegten Nutzerpräferenzen oder Themengebieten mittels sachlogischer Regeln abgeleitet werden (regelbasiertes Matching) [53; 54; 65]. Es ist auch möglich, anstatt einzelne Seiten vorzuschlagen, eine Indexseite mit den für den Nutzer interessanten und relevanten Links bereitzustellen [60].

#### 3.2.2. Auswahl und Eingrenzung des Suchraumes

Die Personalisierung des Suchraumes bzw. der Suchdomäne stellt die erste Technik im eigentlichen Suchprozess dar. Suchmaschinen können dem Nutzer die Möglichkeit geben, zur Einschränkung des Suchraumes aus einer Liste von Kategorien seine Interessen oder die Art der gesuchten Information auszuwählen noch bevor die eigentliche Suchanfrage gestellt wird [75]. Die Technik ist dabei insbesondere auf Meta-Suchmaschinen anzuwenden, bei denen der Nutzer nicht nur Themenkategorien, sondern einzelne oft themenspezifische Suchmaschinen für die Suche auswählen und festlegen kann [26; 44]. Dabei kann der Nutzer Suchmaschinen so miteinander kombinieren, dass die Abdeckung des Internets verbessert und so der Suchraum gezielt ausgeweitet wird, oder aber im Gegensatz dazu der Suchraum bewusst auf themenspezifische Suchdienste eingegrenzt und damit der Suchraum verringert wird, um die Anzahl der Treffer zu reduzieren und die Relevanz der Ergebnisse zu erhöhen [14].

### 3.3. Personalisierung der Suchanfrage

Wie oben bereits erläutert, haben Nutzer typischerweise Probleme bei der Formulierung sinnvoller und geeigneter Suchanfragen. Hier setzen Techniken an, die versuchen, den Nutzer bei der Verbesserung seiner Suchanfragen zu unterstützen, indem sie dem Nutzer Begriffe vorschlagen und die Begriffe wenn möglich automatisiert herleiten [81]. Eine aktuelle Studie belegt, dass Nutzer tatsächlich Suchinterfaces bevorzugen, die ihnen eine personalisierte Anpassung der Suchanfrage erlauben [79]. In der Suchmaschinenforschung ist dieser Bereich seit langem vertreten, so dass sich hier zahlreiche Ansätze finden.

#### 3.3.1. Iterative Anfrageverbesserung

Suche ist meist ein iterativer Prozess, der in mehreren Schleifen abläuft [83]. Hier setzt die iterative Anfrageverbesserung an. Unter Bezeichnungen wie „Query Modification“ [43] oder „Relevance Feedback“ [68; 81] finden sich in der Literatur zahlreiche technische Verfahren und Algorithmen für eine schrittweise Verbesserung und Verfeinerung einer vom Nutzer gestellten Suchanfrage. Dabei ergeben sich prinzipiell zwei Möglichkeiten. Zum einen kann das System die vom Nutzer gewählten Suchwörter in den thematischen Kontext verwandter Themen und Schlagwörter einordnen, die die im ersten Schritt gefundenen Ergebnissen beschreiben. Lundquist et al. diskutieren verschiedene Ansätze, wie das System dem Nutzer aus einer Analyse der gefundenen Ergebnisse thematisch verwandte Wörter zur Verfeinerung der Suchanfrage vorschlagen kann [49]. Ein weiteres Beispiel findet sich bei [66]. Zum anderen kann dem Nutzer die Möglichkeit gegeben werden, die gelieferten Ergebnisse hinsichtlich ihrer Relevanz zu bewerten und hieraus neue Suchwörter zu ermitteln [7]. Dieser Ablauf kann auch automatisiert erfolgen; Fang und Salvendy demonstrieren, wie durch einfaches Markieren relevanter Ergebnisse in der Suchseite neue Schlüsselwörter automatisch abgeleitet werden können [21]. Die Ausführungen machen deutlich, dass diese Verfahren auf einem sehr flüchtigen Verständnis von Personalisierung basieren [75], Nutzerprofile müssen hierfür nicht angelegt werden.

#### 3.3.2. Profilbasierte Anfrageverbesserung – insbesondere zur Lokalisierung

Während die iterative Anfrageverbesserung ohne Profile auskommt, können als für den Nutzer relevant erkannte Schlagwörter vorangegangener Suchen auch permanent im Nutzerprofil gespeichert und zukünftigen Suchanfragen automatisiert beigelegt werden. Ein solches Verfahren wird von Fitzpatrick und Dent beschrieben; ihre Untersuchung belegt, dass die automatisiert hinzugefügten Begriffe in der Lage sind, das Suchergebnis signifikant zu verbessern [22]. Neben den aus der Suchhistorie des Nutzers entnommenen Schlüsselwörtern können auch andere im Profil gespeicherte Nutzercharakteristika, wie beispielsweise der Wohnort des Nutzers der Suchanfrage beigelegt werden. Bei einer solchen Lokalisierung können die Standortinformationen dazu dienen, dem Nutzer regionale oder lokal interessante Informationen bereitzustellen [77]. So könnten bspw. Objekte wie Restaurants, Ärzte, Geldautomaten, Apotheken oder Tankstellen gezielt in der Umgebung des Nutzers gesucht werden.

#### 3.3.3. Anfrageverbesserung mittels Kategorisierung

Bei dieser Technik werden die Suchanfragen des Nutzers intern mit einer Hierarchie von Kategorien (Ontologie) abgeglichen, um die Anfrage durch das gezielte

Hinzufügen von Schlagworten zu verfeinern [57]. Liu et al. beschreiben ein Verfahren, bei dem die Suchhistorie des Nutzers als Set von Themen-Kategorien durch das System implizit erfasst und mit einem allgemeingültigen Kategoriensystem abgeglichen wird [47]. Als allgemeines Kategoriensystem wird dabei das offene Projekt „Open Directory Project (ODP)“ genutzt.<sup>1</sup> Konkret wird dabei die Suchhistorie des Nutzers ausgewertet und mit dem ODP-Kategoriensystem verglichen. Anschließend werden die in der Nutzerhistorie repräsentierten Themen markiert und gewichtet. Bspw. könnte aus der Nutzerhistorie ermittelt werden, dass das Wort „Jaguar“ in die Kategorie „Natur/Raubkatze/Jaguar“ fällt. Diese Kategorie würde markiert, während „Automobil/Jaguar“ unmarkiert bliebe. Die identifizierten Kategorien werden dann zur Einschränkung des Themas bei der Suche verwendet [47]. Ein ähnliches Verfahren beschreiben Pitkow et al. [62]. Sie nutzen ebenfalls das ODP, um ein kategoriebasiertes Nutzerprofil zu erstellen, indem die Top-1000 Einträge des ODP-Kataloges für jeden Nutzer individuell gewichtet werden; dies kann entweder implizit anhand der Suchhistorie oder durch explizite Einspeisung einer Reihe favorisierter Links geschehen [37]. Gibt der Nutzer eine Anfrage ein, wird deren Übereinstimmung mit dem individuellen Kategoriemodell berechnet. Beinhaltet die Anfrage z.B. ein Schlagwort („Jaguar“) aus einem Themenbereich („Natur“), der im Profil hervorgehoben ist, so fügt das System gezielt die entsprechenden Kategorienamen der Anfrage hinzu, um die Anfrage in diese Richtung zu gewichten. Auf diese Weise kann durch Verfahren der Kategorisierung dem Problem der Doppeldeutigkeit bei Suchanfragen begegnet werden [37].

#### 3.3.4. Anfrageverbesserung mittels logischer Deduktion

Diese Form der Personalisierung basiert auf der Nutzung semantischer Zusammenhänge in Kombination mit der Suchanfrage [46]. Sie nutzt intern das Prinzip der logischen Deduktion, das von Expertensystemen bekannt ist. Hierfür wird ein System logischer Aussagen im System hinterlegt. Zur Verdeutlichung der Arbeitsweise dient folgendes Beispiel: Im System könnten die Aussagen „Ein Dackel ist ein Hund“, „Ein Hund ist ein Haustier“, „Eine Erkältung bedeutet, dass man krank ist“, „Tierärzte können kranke Haustiere behandeln“ und „Tierärzte haben eine Praxis in einem bestimmten Gebiet“ gespeichert sein. Bei der Suchanfrage „Erkältung Dackel“ können intern dann folgende Schritte durchgeführt werden: Die Suchanfrage wird zunächst in eine semantische Anfrage abgewandelt, diese wird dann einem der gespeicherten Themenbereiche zugeordnet. Daraufhin wird die Suchanfrage mit Hilfe der logischen Zusammenhänge neu formuliert. Die Suchmaschine ändert die Suchanfrage dann bspw. in „Tierärzte, Köln, NRW“, wobei die Ortsangabe wiederum dem Nutzerprofil entnommen werden kann. Der größte Nachteil dieses Verfahrens ist, dass zunächst alle logischen Zusammenhänge im System gespeichert sein müssen, was einen sehr hohen Aufwand erfordert.

#### 3.4. Personalisierung des Suchalgorithmus

Verfahren zur Personalisierung der Suchanfrage (s.o.) können auch auf der Ebene von Meta-Suchmaschinen angewendet werden, noch bevor die Suchwortkombination an die

<sup>1</sup> Das ODP hat sich zum Ziel gesetzt, aufbauend auf einer ständig gepflegten Hierarchie von Themen-Kategorien „der umfassende Katalog des Internets zu werden“ (siehe <http://www.dmoz.org/World/Deutsch/about.html>).

eigentliche Suchmaschine weitergeleitet wird. Die Personalisierung des Suchalgorithmus dagegen greift in den Kern einer Suchmaschine ein, indem z.B. der PageRank-Algorithmus zur Identifikation und Relevanzbestimmung von Suchergebnissen gezielt verändert wird. Hierzu gibt es verschiedene Verfahren, manche nutzen die in Profilen gespeicherten Interessen der Nutzer, andere basieren auf kollaborativen Techniken wie Nutzerähnlichkeitsvergleichen und sozialen Suchnetzwerken.

#### *3.4.1. PageRank-Personalisierung*

Der PageRank-Algorithmus wurde von Page et al. – den Gründern von Google – entwickelt und beschrieben [59] und dann in zahlreichen Varianten weiter ausgearbeitet; eine umfassende Übersicht findet sich bei [1]. Der PageRank einer Seite versucht, deren Relevanz auszudrücken; er wird im Prinzip wie folgt bestimmt: Ein einfacher PageRank kann als Anzahl aller Seiten gebildet werden, die auf diese Seite verlinken, d.h. eine wichtige Seite wird häufiger durch andere referenziert als weniger wichtige. Da dieser Wert durch Spammer gezielt beeinflusst werden kann, geht in den PageRank zudem die Relevanz der verlinkenden Seiten ein. Ein Link einer wichtigen Seite ist also mehr wert als der Link einer unwichtigen. Der PageRank ist also ein rekursiv gebildeter Wert, da die Relevanz einer Seite sowohl die Relevanz anderer Seiten beeinflusst, als auch von deren Relevanz abhängt [1].

Anstatt allgemeine PageRanks für Seiten zu bilden, können nun verschiedene Verfahren genutzt werden, um die Relevanz individuell für einzelne Nutzer zu bestimmen. Haveliwala beschreibt ein Verfahren, das themensensitive PageRanks („topic-sensitive PageRanks“) errechnet [28]. Hierbei werden für eine Seite zu jeder der 16 Top-Level-Kategorien des ODP-Katalogs separate PageRanks berechnet, indem gezielt nur die Verlinkungen von Seiten aus diesen Bereichen berücksichtigt werden. Der PageRank einer Seite ist dann ein Vektor mit Relevanzwerten und symbolisiert die Kompetenz einer Seite in den 16 Themenbereichen [37]. Mittels der aus den Suchanfragen von Nutzern extrahierten Themen können dann personalisierte PageRanks errechnet und so die Suchergebnisse gezielt verbessert werden [28]. Eine weitere Form der PageRank-Personalisierung beschreiben Jeh und Widom: Bei ihrem Verfahren werden die PageRanks konsequent anhand einer vom Nutzer explizit vorgegebenen Anzahl von Webseiten (z.B. persönliche Bookmarks) ausgerichtet [34]. Da den Bookmarks des Nutzers eine höhere Gewichtung zugesprochen wird, können auf diese Weise seine Interessen in den Algorithmus einfließen und damit personalisierte Ergebnisse erzielt werden [12; 34]. Diese Form der Personalisierung funktioniert natürlich auch, wenn nicht Seiten vorgegeben, sondern gezielt solche Seiten bei der Berechnung des PageRanks höher gewichtet werden, die Schlagworte enthalten, die als Themen im Profil des Nutzers hinterlegt sind.

#### *3.4.2. Collaborative Filtering*

Während die gerade diskutierten Verfahren auf dem Nutzerprofil eines einzelnen Nutzers basieren, kann für die personalisierte Berechnung der Relevanz auch der Vergleich von Profilen mehrerer Anwender herangezogen werden. Ein solches Verfahren wurde von Sugiyama et al. beschrieben [73]. Ähnlichkeiten zwischen Nutzerprofilen können dazu beitragen, die Suchergebnisse für einen speziellen Nutzer zu verbessern. Die Annahme hierbei ist, dass, wenn sich zwei Profile in einigen Bereichen stark ähneln, die Nutzer sich mit hoher Wahrscheinlichkeit auch in anderen

Bereichen ähneln und damit ähnliche Suchergebnisse als relevant erachten. Internetseiten, die ähnliche Nutzer als relevant eingestuft haben, werden dann entsprechend mit einer höheren Gewichtung beim Ranking versehen [73]. Diese Methode basiert auf dem sog. Collaborative Filtering, das auch z.B. bei Amazon.com für die Empfehlung von Produkten angewandt wird [54; 65]: „Kunden, die dieses Produkt gekauft haben, haben auch ... gekauft.“ Für konzeptionelle Hintergründe zu dem Konzept, sowie Algorithmen für die Berechnung von Ähnlichkeiten zwischen Nutzerprofilen siehe [2]; Mild und Natter vergleichen zudem die Vorteilhaftigkeit typischer Collaborative Filtering Verfahren mit Regressionsverfahren [55].

### 3.4.3. Suchnetzwerke

Eine weitere Möglichkeit kollaborativer Personalisierung ist der Aufbau von Suchnetzwerken [5]. Dabei wird die Ermittlung derjenigen Nutzer, die als Referenz zur Ermittlung von Ähnlichkeiten dienen, nicht vom System vorgenommen. Stattdessen wählt der Suchende diese Referenznutzer selbst aus, indem er ein Suchnetzwerk aufbaut oder einem Netzwerk beiträgt. Anwender können sich dann z.B. explizit solche Webseiten anzeigen lassen, die andere Mitglieder ihres Netzwerks als relevant empfinden. Der wahre Nutzen ergibt sich aber erst in Kombination mit der oben beschriebenen PageRank-Personalisierung. Dabei können Internetseiten anderer Nutzer des Netzwerks implizit mit einer höheren Gewichtung versehen werden [73]. Ein PageRank-Algorithmus wird also solche Seiten höher gewichtet, auf die die Nutzer des Netzwerks direkt oder indirekt verlinkt haben. Auf diese Weise wird die Relevanz der Suchergebnisse zugunsten der Interessen des Suchnetzwerks beeinflusst. Stimmen Nutzer- und Netzwerkinteressen überein, findet also Personalisierung statt.

## 3.5. Personalisierte Präsentation der Suchergebnisse

### 3.5.1. Ergebnis-Reranking

Um dem Nutzer nur relevante Ergebnisse anzuzeigen, ist es möglich, die von der Suchmaschine ermittelten Ergebnisse personalisiert zu filtern bzw. die Suchergebnisse entsprechend der persönlichen Relevanz des Nutzers zu ordnen [61; 62]. Ein solches Verfahren eignet sich dabei wiederum auch für Meta-Suchmaschinen. Dabei können z.B. solche Internetseiten, die auf Basis des Nutzerprofils einem spezifischen Interessengebiet zugeordnet werden können, bei der Festlegung der Anzeigereihenfolge höher gewichtet werden [24]. Eine weitere Möglichkeit des Reranking basiert auf Kriterien wie Aktualität und Popularität; hierbei kann dem Nutzer die Möglichkeit gegeben werden, gezielt neue und beliebte Informationen anzeigen zu lassen [62]. Ein weiteres Verfahren wird von Makris et al. beschrieben; die Autoren nutzen wiederum Kategoriebäume auf Basis des ODP [57]. Für jeden Nutzer wird dabei implizit auf Basis seiner Suchhistorie lokal ein eigener Teilbaum des ODP berechnet und mit PageRanks versehen. Dieser Themenbaum wird dann mit dem Suchmaschinenergebnis verknüpft, um eine Ergebnisliste zu generieren, die den Vorlieben des Nutzers entspricht [57]. Ein weiteres graphenbasiertes Verfahren beschreiben [79]. Smyth und Balfe schließlich beschreiben ein kollaboratives Verfahren für Meta-Suchmaschinen, bei dem zunächst das Nutzerverhalten bei der Auswahl von Suchergebnissen festgehalten und in sog. Community-Profilen gespeichert wird [71]. Dabei wird einerseits die Privatsphäre des einzelnen Nutzers gewahrt, da Informationen zum Verhalten nur aggregiert gespeichert werden; andererseits werden Ähnlichkeiten im

Nutzungsverhalten der Suchenden dazu genutzt, die Ergebnisse des Rerankings zu verbessern.

### 3.5.2. Clusterung und interaktive Neuordnung

Eine weitere Möglichkeit der Neuordnung von Suchmaschinenergebnissen basiert auf der Idee der thematischen Clusterung [42; 85]. Dabei werden die Ergebnisse mittels Cluster-Techniken in thematische Gruppen eingeteilt und dem Nutzer präsentiert. Der Nutzer kann dann auf Cluster-Namen klicken, um die Ergebnismenge thematisch auf Ergebnisse in seinem Interessenbereich einzuschränken [42]. Die thematische Clusterung erlaubt dem System zudem, doppelte und irrelevante Informationen zu entfernen und so zur Reduzierung des Information Overload beizutragen. Auf diese Weise können Nutzer die für sie relevanten Webseiten einfacher finden, da die Ergebnisse übersichtlicher und besser voneinander abgegrenzt sind [85]. Darüber hinaus kann dem Nutzer auch die Möglichkeit gegeben werden, erneut innerhalb von Clustern zu suchen. Dies kann dann in einer neuen Suchanfrage mit reduziertem Suchraum münden; dementsprechend ist diese Technik eng verwandt mit der oben beschriebenen iterativen Anfrageverbesserung.

## 4. Marktüberblick: Einsatz der Techniken in der Praxis

Im vorangegangenen Kapitel wurde der Stand der Forschung zu Personalisierungsverfahren für die Web-Suche dokumentiert. Alle gefundenen Techniken und Verfahren wurden anhand des Suchprozesses strukturiert präsentiert. Ziel dieses Kapitels ist es nun, zu untersuchen, ob die diskutierten Techniken bereits in der Praxis zu finden sind, d.h. welche der Techniken von Suchmaschinenbetreibern bereits eingesetzt werden. Dabei konzentriert sich die Untersuchung auf Suchdienste, die eine generelle Web-Suche unterstützen; spezielle themen- und domänenbezogene Suchdienste werden nicht betrachtet. Im weiteren wird zunächst die Auswahl und Klassifikation der Suchdienste beschrieben; die Übersicht über die untersuchten Dienste und die unterstützten Techniken findet sich in Tabelle 2. Dem schließt sich eine Diskussion der gefundenen Ergebnisse an.

### 4.1. Auswahl und Klassifikation der Suchdienste

Ziel der Untersuchung war es, eine möglichst breite Palette von Suchdiensten auf die Unterstützung von Personalisierungstechniken zu untersuchen. Für die Auswahl der Suchmaschinen wurde dabei zunächst die Referenz-Web-Site „Search-Engine-Watch“ herangezogen; diese wurde ergänzt durch Literatur- und Web-Recherche nach Suchdiensten mit Personalisierungsfunktion. Zunächst wurden alle bei Search-Engine-Watch unter „Major Search Engines“ gelisteten Suchdienste auf das Vorhandensein von Personalisierungsfunktionen überprüft; wurden entsprechende Techniken unterstützt, so wurde der Dienst in die Untersuchung einbezogen<sup>2</sup>.

---

<sup>2</sup> Anmerkung: Suchdienste, die lediglich die Suchraumeinschränkung bieten (wie z.B. AOL) wurden nicht aufgenommen, da diese Funktion mittlerweile von fast allen Diensten geboten wird und die Personalisierungswirkung eher begrenzt ist. Die Tabelle erhebt nicht den Anspruch auf Vollständigkeit, soll aber einen umfassenden Überblick über Suchdienste mit Personalisierungsfunktion geben.

Dabei erreichen allein die vier großen Suchdienste *Google* (49,2%), *Yahoo!* (23,8%), *MSN* (9,6%) und *AskJeeves* (2,6%) zusammen 85,2% der Nutzer<sup>3</sup>. Zusammen mit den im Weiteren erfassten Suchdiensten deckt die Stichprobe also einen Großteil des Marktes ab. Neben den vier großen Suchdiensten wurden weitere Suchdienste erfasst. Dabei wurden von den bei Search-Engine-Watch gelisteten Meta-Suchmaschinen die fünf „Award-Winners“ *Dogpile*, *Vivisimo*, *Kartoo*, *Mamma* und *Surfwax* aufgenommen sowie einige weitere Meta-Suchmaschinen, die spezielle Personalisierungsfunktionen bieten, die sich von den bei Meta-Suchdiensten üblichen absetzen (siehe Diskussion unten). Als Ergänzung zu den Meta-Suchmaschinen wurden mit *Copernic* und *Webferret* zudem zwei bekannte Desktop-Tools aufgenommen, die ähnliche Funktionen wie Meta-Suchmaschinen bieten. Schlussendlich wurde die Menge der Suchdienste um solche ergänzt, die kollaborative Personalisierungstechniken bieten. Diese wurden auf Basis von Web-Recherche (*Filangy*) und Literaturrecherche (*Amazon A9* und *Eurekster*) identifiziert [5].

Tabelle 2. Klassifikation der Suchdienste

Suchdienste	P1: Profil-Bildung		P2: Vor-Anfrage		P3: Anfrage-verbesserung				P4: Algorithmus			P5: Prä-sentation	
	Explizit	Implizit	Vorschlagsfunktion	Eingrenzung des Suchraums	iterative Anfrageverbesserung	Profilbasiert (Lokalisierung)	Kategorisierung	Logische Deduktion	PageRank-Personalisierung	Kollaborative Filtering	Suchnetzwerk	Ergebnis-Reranking	Clustering & Neuordnung
Mooter													v
Vivisimo/Clusty (Meta)				v									v
iZito (Meta)				v									v
Gigablast				v	v								
Surfwax (Meta)					v								
Kartoo (Meta)					v								v
Infospace Engines (Meta)				v	v								
Mamma (Meta)				v	v								
Copernic (Tool)	v			v									v
lboogie (Meta)	v			v									v
Exalead	v			v	v								
Webferret (Tool)	v			v	v								v
Fazzle (Meta)	v			v									v
AskJeeves	v			v	v	v							
MSN (Windows Live)	v	v	v	v		v							v
Google Beta		v		v					v				
Amazon A9	v	v	v	v		v				v			
Eurekster	v	v	v	v					v		v		
Filangy	v	v		v					v		v		
Yahoo! MyWeb beta	v	v		v		v					v		
n=	11	6	3	17	8	4	0	0	3	1	3	3	6

<sup>3</sup> Nielsen-Rating, Stand August 2006, siehe <http://searchenginewatch.com/showPage.html?page=2156451>.

Alle Suchmaschinen wurden benutzt und getestet; zudem wurden die vom Anbieter zur Verfügung gestellten Beschreibungen und Hilfe-Informationen herangezogen, um die Suchmaschinen zu klassifizieren. Dabei war es nicht das Ziel, Aussagen über die Qualität der Suchergebnisse zu machen, sondern das Vorhandensein von Personalisierungstechniken zu überprüfen. Unterstützt ein Suchdienst eine Technik oder legt er ein Profil über den Nutzer an, so wurde dies in Tabelle 2 entsprechend vermerkt. Die Struktur der Tabelle folgt dabei der aus dem vorangegangenen Kapitel bekannten Unterteilung.

#### 4.1.1. Diskussion der Klassifikationsergebnisse

Anhand der Klassifikation in Tabelle 1 lassen sich zwei Gruppen von Suchdiensten unterscheiden: Die erste Gruppe umfasst Suchdienste, die nur Techniken in den Phasen vor und nach der eigentlichen Durchführung der Suche bieten, die also auch von Meta-Suchmaschinen und Such-Tools unterstützt werden können. Bei diesen Techniken handelt es sich um die Eingrenzung des Suchraumes, die iterative Anfrageverbesserung, das Reranking, sowie die Clusterung und Neuordnung der Ergebnisse. Die zweite Gruppe umfasst Suchdienste, die Personalisierung im engeren Sinn anbieten. Diese basieren entsprechend auf implizit gebildeten Nutzerprofilen und bieten PageRank-Personalisierung oder kollaborative Techniken. Im Weiteren werden die beiden Gruppen sowie ihre typischen Funktionen und die beispielhafte Umsetzung durch einzelne Suchdienste beschrieben.

#### 4.1.2. Personalisierung auf der Meta-Ebene

Fast alle untersuchten Suchdienste bieten eine Form der Suchraum-Eingrenzung. Die Umsetzung und Detaillierung dieser Funktion unterscheidet sich jedoch teils deutlich. Das Spektrum reicht von einer simplen Eingrenzung auf Kategorien wie „Web“ und „Bilder“ (*Exalead*), über eine breite Auswahl an Eingrenzungen wie sie auch *Google* bietet, bis hin zur Möglichkeit, spezielle Suchdienste auszuwählen oder gar eigene Suchdienste hinzuzufügen. Letzteres bietet *Iboogie*; hier kann man Suchdienste anmelden und detailliert festlegen, wie die gelieferten Resultate dargestellt werden sollen. *Iboogie* legt dabei ein Profil für den Nutzer an, in dem er die getroffene Auswahl verwalten kann. Eine solche Speicherung der Suchraumeingrenzung ist dabei gleichzeitig die einzige Funktion für die die Nutzerprofile in dieser Gruppe der Suchdienste genutzt werden.

Die iterative Anfrageverbesserung, bei der dem Nutzer neben der Liste der Ergebnisse eine Liste mit Schlagworten für die Formulierung einer neuen Anfrage geliefert wird, wird von mehr als der Hälfte der Suchdienste in dieser Gruppe unterstützt. Eine gelungene Umsetzung bietet *Exalead*. Nach Durchführung der initialen Suche kann der Nutzer mittels „Verbundene Begriffe“ seine neue Anfrage hinsichtlich verwandter Schlagworte verfeinern und zudem durch die Auswahl einer „Kategorie“ auf thematisch disjunkte Bereiche eingrenzen. Letzterer Punkt dient der Behandlung von Mehrdeutigkeiten. Zudem kann er die neue Anfrage auf eine Sprache, einen Dokumenten- („pdf“, „doc“), wie auch Medientyp („Audio“, „Video“) eingrenzen.

Nur drei Suchdienste bieten eine Form des Ergebnis-Reranking an. *Fazzle* und *Webferret* bieten dabei nur sehr begrenzte Möglichkeiten, das Ranking zu beeinflussen. Der Nutzer kann lediglich auswählen, nach welchen Kriterien die Ergebnisse gelistet werden sollen (z.B. Popularität, URL, Titel). Einzig *MSN (Windows Live)* bietet ein

weitergehendes Ergebnis-Reranking. Hier stehen dem Nutzer Schieberegler für die Kriterien „Kürzlich aktualisiert“, „Häufig besucht“ und „Ungefähre Übereinstimmung“ zur Verfügung, mit denen er die jeweilige Gewichtung der Ergebnisse explizit vorgeben kann.

Die Clusterung und Neuordnung der Ergebnisse ist eine weitere typische Technik der Personalisierung auf der Meta-Ebene. Bekannt für die Clusterung ist die Meta-Suchmaschine Clusty von *Vivisimo* [42]. Über die Funktion „Show in Clusters“ können Suchergebnisse in den thematischen Clustern gezeigt werden; durch Aufklappen der Cluster werden die Ergebnisse neu geordnet. Die Themen-Cluster werden dem Nutzer ähnlich präsentiert wie die Schlagworte zur Anfrageverbesserung. Daher stellen sich die beiden Techniken aus Nutzersicht sehr ähnlich dar, auch wenn sie intern unterschiedlich funktionieren. Eine integrierte Umsetzung der beiden Techniken in Kombination mit einer grafischen Flash-Visualisierung bietet die Suchmaschine *Kartoo*. Der Dienst präsentiert Schlagworte für die Verfeinerung sowie grafisch visualisierte Cluster, die die Links zwischen den gefundenen Web-Sites darstellen. Diese Cluster kann der Nutzer gezielt navigieren und so die Ergebnismenge erkunden.

#### 4.1.3. Personalisierung im engeren Sinn – Profilbildung und Algorithmusebene

Während die von der ersten Gruppe unterstützten Techniken dem Nutzer lediglich helfen, eine bessere Anfrage zu formulieren oder einen persönlichen Zugang zu ansonsten unpersonalisierten Menge der Suchergebnisse zu finden, versucht die zweite Gruppe von Suchdiensten, von vornherein auf den Nutzer persönlich zugeschnittene Ergebnisse zu liefern. Hierfür werden die im Nutzerprofil gespeicherten Daten genutzt. Dabei werden – teils zusätzlich zu den oben beschriebenen – auch Techniken auf der Algorithmusebene eingesetzt.

Eine reine PageRank-Personalisierung wird bisher nur von Google angeboten. Die als Betaversion vorliegende Personalisierung wird zum jetzigen Zeitpunkt lediglich auf die Suche im Web angewendet. Sie basiert auf der Anpassung der internen Gewichte des Algorithmus durch einen Ähnlichkeitsvergleich mit den Daten im Nutzerprofil. Die genaue Funktionsweise des Algorithmus wurde nicht offen gelegt, es kann jedoch davon ausgegangen werden, dass sie den in [12] und [34] beschriebenen Verfahren ähnelt. Nutzerprofile werden auf Basis der Suchhistorie implizit gewonnen, indem Suchergebnisse und angeklickte Internetseiten ausgewertet werden, um relevante Themen zu extrahieren.

Vier Dienste bieten kollaborative Personalisierung, die auf den Techniken Collaborative Filtering oder Suchnetzwerke basiert (*Amazon A9*, *Eurekster*, *Filangy*, *Yahoo!* ). Die *Amazon A9*-Personalisierung basiert auf Collaborative Filtering. Die Technik entstammt dem Produktvorschlagsystem bei *Amazon.com* [74]. Dabei können existierende Profile von *Amazon.com* und die Bookmarks des Nutzers explizit ins Profil importiert werden. Darüber hinaus speichert der Dienst alle Aktionen der Nutzer in einer Suchhistorie und erfasst zusätzlich das Surfverhalten, sofern Nutzer die *A9*-Toolbar nutzen. Die anderen drei Dienste basieren auf Suchnetzwerken; dabei kombinieren *Filangy* und *Eurekster* die PageRank-Personalisierung mit Nutzer-Profilen auf Netzwerkbasis. Bei *Yahoo! myWeb* dagegen findet keine Algorithmus-Personalisierung, sondern lediglich eine Verschlagwortung und Empfehlung von Webseiten durch die Mitglieder des Netzwerks statt; in diesen Webseiten kann der Nutzer auch explizit suchen.

Eurekster stellt ein gutes Beispiel komplexer Personalisierung mit Suchnetzwerken dar. Nutzer können eine sog. SearchParty gründen oder einer existierenden beitreten. Dabei passt *Eurekster* bei der Suche die internen Relevanzgewichte seines Algorithmus anhand der Bewertung der Webseiten durch andere Netzwerkmitglieder an. Internetseiten, die von einem Nutzer (z.B. durch längeren Verbleib auf der Seite) als nützlich klassifiziert wurden, werden bei anderen Nutzern im gleichen Suchnetzwerk stärker gewichtet [5]. Der Algorithmus von *Eurekster* lernt dabei kontinuierlich vom Nutzerverhalten, welche Suchresultate relevant sind und welche Filtertechniken am besten funktionieren. Die Ermittlung der Nutzerdaten erfolgt also im Wesentlichen implizit auf Basis der Suchhistorie. Falls Nutzer ein eigenes Netzwerk gründen, müssen sie relevante Schlagwörter und Webseiten jedoch explizit eingeben.

Neben den genannten drei Kerntechniken unterstützen einige der Suchdienste in dieser Gruppe auch eine Lokalisierung und eine automatisierte Vorschlagsfunktion. Bei *MSN (Windows Live)* z.B. kann der Nutzer seine Suchanfrage gezielt auf eine vorher im Profil hinterlegte geographische Region eingrenzen. Zudem unterstützt *MSN* die Vorschlagsfunktion. Der Dienst erfasst und speichert die Suchhistorie des Nutzers, um ihm nach Auswertung personalisierte Nachrichtenbeiträge auf der Web-Site liefern zu können. Das implizit gewonnene Profil wird bei *MSN* im Übrigen nur für die personalisierte Anzeige der Nachrichten genutzt, nicht jedoch für die Suchpersonalisierung. Auch *A9* unterstützt die Vorschlagsfunktion, indem zu einer bekannten Webseite ähnliche Seiten gesucht oder auf Basis des Nutzerprofils einzelne als relevant eingestufte Internetseiten vorgeschlagen werden.

#### 4.2. Status-Quo und weitere Entwicklung

Während das Thema Personalisierung von Suchdiensten in der Forschung bereits einen hohen Grad an Differenzierung erreicht hat – es finden sich zahlreiche, teils neuartige Techniken und Verfahren in den verschiedenen Phasen des Suchprozesses (siehe Kap. 3) – steckt die Umsetzung in der Praxis noch in den Kinderschuhen. Nur drei der oben diskutierten Techniken haben sich insbesondere bei Meta-Suchmaschinen bereits zu Standard-Funktionen entwickelt: die Suchraum-Eingrenzung, die iterative Anfrageverbesserung und die Clusterung und Neuordnung. Diese Techniken sind gleichzeitig jedoch die am wenigsten komplexen Techniken; alle kommen ohne das Anlegen von Nutzerprofilen aus. Die iterative Anfrageverbesserung ist zudem die älteste und am besten in der Literatur beschriebene, so dass deren Verbreitung in der Praxis nicht überrascht.

Zwei der oben beschriebenen Techniken (Anfrageverbesserung mittels Kategorisierung und mittels logischer Deduktion) finden sich in der Praxis noch gar nicht wieder und auch das Ergebnis-Reranking findet sich nur in sehr rudimentärer Form, ohne dass für die Technik Nutzerprofile genutzt würden. Diese Beobachtungen machen deutlich, dass hier von den Suchmaschinen-Betreibern zukünftig noch weitere Potenziale genutzt werden können. So könnten sich Anbieter von Meta-Suchmaschinen durch die Umsetzung einer weitergehenden profilbasierten Personalisierung von ihren Wettbewerbern differenzieren, ohne dass dafür Techniken auf der Algorithmusebene nötig wären. Dies könnte z.B. durch den Einsatz von Anfrageverbesserung mittels Kategorisierung und einer ausgefeilten Reranking-Technik geschehen, wie bereits von Pitkow et al. anhand einer prototypischen Umsetzung demonstriert wurde [62].

Weitergehende Techniken wie PageRank-Personalisierung oder kollaborative Techniken finden sich bisher nur bei sehr wenigen Suchdiensten. Hinzu kommt die

Tatsache, dass die Personalisierung bei den dominierenden Suchmaschinen-Anbietern wie Google und Yahoo! bisher nur als Beta-Funktion existiert. In ihrer Grundversion bieten die beiden Dienste, genau wie der Suchdienst von AOL als drittgrößter am Markt, gar keine Personalisierungstechniken außer der Suchraumeinschränkung. Aus diesem Grund wird bisher vermutlich nur ein verschwindend geringer Anteil von Web-Nutzern überhaupt mit Personalisierungsfunktionen in Kontakt gekommen sein. Erst wenn sich die Marktführer zu einer Einführung und Vermarktung solcher Techniken auf breiter Basis (und nicht versteckt in den Tiefen der Menüs) entscheiden sollten, dürfte mit dem Durchbruch der Suchpersonalisierung in der Praxis zu rechnen sein. Die Tatsache, dass entsprechende Beta-Dienste bereits angeboten werden, lässt einen solchen Schritt für die nähere Zukunft erwarten.

Kollaborative Techniken haben mit einem kritische Masse-Problem zu kämpfen: Collaborative Filtering funktioniert nur zuverlässig, wenn genügend Vergleichsprofile anderer Nutzer vorliegen; Amazon versucht bei A9 das Problem durch den Import von Profilen aus dem Amazon-Shop zu mildern. Suchnetzwerke benötigen ebenfalls eine kritische Masse an Nutzern; zudem muss der Nutzer bei Systemen wie Yahoo! und Eureka das Suchnetzwerk selbst pflegen. Als Gegenleistung versprechen diese Systeme den Nutzern jedoch ein mächtiges Werkzeug, das besonders auch für die gemeinsame, kontextunterstützte Suche in (Arbeits-)Gruppen genutzt werden kann und sich so für die Nutzung im professionellen (z.B. universitären) Umfeld anbietet. Den besten Gegenwert aus Aufwand für den Nutzer und Personalisierungsleistung versprechen jedoch Systeme, die selbständig lernen, implizit Profile bilden und dies mit mächtigen Techniken wie Kategorisierung oder PageRank-Personalisierung verbinden. Hier ist wieder einmal Google mit seinem Ansatz sehr gut positioniert.

## 5. Implikationen für die weitere Forschung

Nachdem der Markt für Suchdienste auf die Verbreitung der vorgestellten Personalisierungstechniken untersucht, Beispiele vorgestellt und ein kurzer Marktausblick gegeben wurde, sollen nun einige Implikationen für die weitere Forschung zur Suchpersonalisierung diskutiert werden. Dabei kann in Anlehnung an Abb. 1 wieder eine Unterscheidung in gestaltungsorientierte und erkenntnistheoretische Studien vorgenommen werden.

### 5.1. Studien zur Gestaltung der Suchpersonalisierung

Drei Bereiche können identifiziert werden, die die weitere Forschung in diesem Bereich prägen könnten. Erstens wird die Forschung an neuartigen Verfahren und Algorithmen zur Personalisierung fortgesetzt werden, zweitens rückt die Art und Weise der Datengewinnung über den Nutzer und damit besonders das Erfassen des Suchkontextes in den Vordergrund und drittens müssen auch die Nutzerinterfaces der hinzugekommenen Komplexität durch die Personalisierung Rechnung tragen.

Wie die Ausführungen oben gezeigt haben, werden bereits einige der auf Basis der Literatur identifizierten Personalisierungstechniken im Markt eingesetzt. Jedoch sind weitere Personalisierungstechniken möglich und mit fortschreitender Verbesserung der Technik auch umsetzbar. Es ist abzusehen, dass sich dieser Bereich in Zukunft weiter stark entwickeln wird [45]. Aktuelle Arbeiten zu neuartigen Algorithmen, wie z.B. zu kategoriebasierten [57], graphenbasierten [79] oder kollaborativen Verfahren [71]

machen deutlich, dass auch weiterhin mit der Entwicklung neuartiger Verfahren in Ergänzung zu den in dieser Arbeit skizzierten zu rechnen ist.

Eine verbesserte Datengewinnung zur Profilbildung ist ein weiterer zentraler Forschungsbereich, um die Qualität von Suchpersonalisierung zu verbessern. Ein Thema ist hierbei die Erfassung und Nutzung der Suchhistorie des Nutzers. So erörtert eine aktuelle Studie von Komlodi et al. neue Wege, die Suchhistorie des Nutzers in die Suche einfließen zu lassen [41]. Darüber hinaus sind neuartige Verfahren der Profilbildung denkbar, die nicht den einzelnen Nutzer, sondern Gruppenprofile oder Profilcluster bei der Personalisierung zugrunde legen; ein erstes Beispiel hierfür liefert [71]. Des Weiteren wird die Personalisierung zukünftig nicht nur auf der Suchhistorie, sondern auf einem reichhaltigeren Verständnis des Kontextes des Nutzers basieren. Um den Nutzungskontext des Suchenden zu bestimmen, könnten z.B. Inhalte von Dokumenten des lokalen Rechners ausgewertet werden [13; 44]. Hier bietet die Kombination von Websuche und Desktopsuche einige Potenziale. Die bei der Desktopsuche anfallenden Daten können entsprechend zur Ermittlung des Kontexts oder der Nutzerpräferenzen ausgewertet werden [75]. Eine aktuelle Übersicht über Kontextdimensionen und die Rolle von Kontext in der Suchmaschinenforschung gibt Kelly [38; 39].

Die durch die Personalisierung der Suche für den Nutzer steigende Komplexität sollte sinnvoll durch interaktive Interfaces unterstützt werden. Neue Techniken zur Interface-Entwicklung wie das Ajax-Framework bieten Entwicklern neue Möglichkeiten, interaktive Personalisierungstechniken wie die Anfrageverbesserung besser abzubilden [84].

## 5.2. Studien zur Nutzung und Akzeptanz der Suchpersonalisierung

Um die Entwicklung von Techniken und Verfahren und insbesondere den erfolgreichen Einsatz von Suchpersonalisierung in der Praxis sinnvoll zu unterstützen, ist ein besseres Verständnis für die Sichtweise des Nutzers beim Einsatz solcher Techniken notwendig. Nur so kann ein Konzept wie die Personalisierung, dessen Ziel es gerade ist, sich konkret an den Bedürfnissen des Nutzers zu orientieren, erfolgreich sein. Studien zur Nutzung von Personalisierungsdiensten, zu deren Vorteilhaftigkeit und Akzeptanz durch die Nutzer sind jedoch bisher kaum vorhanden; für erste aktuelle Beispiele siehe [42; 81].

In der aktuellen Phase des Forschungszyklus bieten sich insbesondere Studien an, die aufbauend auf Prototypen Gruppen von Nutzern mit neuartigen Techniken konfrontieren, um aus der Interaktion mit dem System in experimentellen Studien erste Erkenntnisse zu Verhalten und Akzeptanz seitens der Nutzer zu gewinnen. Dabei fließen solche Erkenntnisse wieder als Anforderungen in den Gestaltungsprozess ein. Experimente mit entsprechenden Prototypen wurden beispielsweise von Pitkow et al. anhand des von ihnen entwickelten Outride-Systems durchgeführt [62]. Ein weiteres Beispiel liefern Smyth und Balfe, die ihr Verfahren des Ergebnis-Reranking ebenfalls als Prototyp (i-Spy) implementiert und mit Nutzern experimentell getestet haben [71].

Ein besonders wichtiger Akzeptanzaspekt bei der Personalisierung betrifft die Speicherung und Nutzung von Daten über den Anwender. Diese hat neben einer rechtlichen Dimension auch eine Vertrauensdimension. Zum einen müssen Datenschutzbestimmungen berücksichtigt werden, darüber hinaus ist ein offener und transparenter Umgang mit den Daten zwingende Voraussetzung, um das Vertrauen der Anwender nicht zu verlieren [69; 80]. Erste Studien in diesem Bereich belegen, dass es

für Nutzer wichtig ist, dass sie die Kontrolle über ihre Daten und die vorgenommene Personalisierung behalten [36; 53]. Insbesondere wenn Anwender nicht sicher sind, wie Systeme funktionieren (Blackboxes), werden sie diese seltener nutzen, besonders wenn ihre Resultate nicht intuitiv, konsistent und vorhersehbar sind [50].

## 6. Fazit

Entwickler von Web-Suchmaschinen stehen traditionell vor einem Trade-off: Einerseits sollen Suchmaschinen eine möglichst breite und vielfältige Abdeckung des Korpus WWW gewährleisten und dabei neutral sein und Bias vermeiden. Andererseits sollen Suchmaschinen möglichst relevante Ergebnisse liefern und diese gut nutzbar präsentieren. Aus Anwendersicht stellt sich das Problem in ähnlicher Weise: Nutzer möchten sicher gehen, dass keine potenziellen Ergebnisse bei der Suche vergessen werden und andererseits möglichst genau die richtigen und relevanten Ergebnisse geliefert bekommen. Personalisierung von Suchmaschinen kann helfen, den zweiten Aspekt zu verbessern, d.h. die subjektiv empfundene Relevanz zu steigern, indem persönliche Charakteristika des Nutzers als zusätzlicher Filter in die Suche und bei der Relevanzbestimmung eingehen. Der Beitrag liefert eine Übersicht über existierende Techniken und Verfahren der Personalisierung in der Literatur, sowie deren Verbreitung in der Praxis. Er bietet sich als Basis für weitere Forschung zu Gestaltung, Nutzung und Akzeptanz der Suchpersonalisierung an.

## Literaturangaben

- [1] *Arasu, A.; Cho, J.; Garcia-Molina, H.; Papcke, A. Raghavan, S.*: Searching the Web. In: ACM Transactions on Internet Technology 1 (2001) 1, S. 2-43.
- [2] *Ansari, A.; Essegai, S.; Kohli, R.*: Internet Recommendation Systems. In: Journal of Marketing Research 37 (2000), S. 363-375.
- [3] *Ansari, A.; Mela, C. F.*: E-Customization. In: Journal of Marketing Research 15 (2003), S. 131-145.
- [4] *Anthes, G. H.*: Search for tomorrow. In: Computerworld 38 (2004) 14, S. 26.
- [5] *Balas, J. L.*: When every search engine knows your name. In: Computers in Libraries 25 (2005) 2, S. 33-35.
- [6] *Bar-Ilan, J.; Mat-Hassan, M.; Levene, M.*: Methods for comparing rankings of search engine results. In: Computer Networks 50 (2006), S. 1448-1463.
- [7] *Belkin, N. J.*: Helping People Find What They Don't Know. In: Communications of the ACM 43 (2000) 8, S. 58-61.
- [8] *Berreby, D.*: Getting to know you. Special Report: The Rise of E-Business. In: IBM Systems Journal (1999) 1, S. 1-4.
- [9] *Bhargava, H. K.; Feng, J.*: Paid Placement Strategies for Internet Search Engines. In: Proceedings of the WWW2002. Honolulu, Hawaii, 2002, S. 117-123.
- [10] *Blom, J.*: Personalization - a taxonomy. In: Proceedings of the CHI 2000 Workshop on Designing Interactive Systems for 1-to-1 E-commerce. 2000, S. 313-314.
- [11] *Brewer, E. A.*: The Consumer Side of Search. In: Communications of the ACM 45 (2002) 9, S. 41.
- [12] *Brin, S.; Page, L.*: The Anatomy of a Large-Scale Hypertextual Web Search Engine. Brisbane, Australia, 1998.
- [13] *Budzik, J.; Hammond, K. J.*: User interactions with everyday applications as context for just-in-time information access. In: Proceedings of the 2000 International Conference on Intelligent User Interfaces. ACM Press: New Orleans, Louisiana, 2000.
- [14] *Chen, H.; Fan, H.; Chau, M.; Zeng, D.*: MetaSpider: Meta-Searching and Categorization on the Web. In: Journal of the American Society for Information Science and Technology 52 (2001) 13, S. 1134-1147.

- [15] Choo, C. W.; Detlor, B.; Turnbull, D.: Information Seeking on the Web: An Integrated Model of Browsing and Searching. In: *first monday* 5 (2000) 2, online verfügbar: [http://firstmonday.org/issues/issue5\\_2/choo/index.html](http://firstmonday.org/issues/issue5_2/choo/index.html) [Abruf: 2006-11-22].
- [16] Claypool, M.; Le, P.; Waseda, M.; Brown, D.: Implicit Interest Indicators. In: Proceedings of the 6th international conference on intelligent user interfaces (ACM) (2001), S. 33-40.
- [17] Deng, L.; Chai, X.; Tan, Q.; Ng, W.; Lee, D. L.: Spying out Real User Preferences for Metasearch Engine Personalization. In: Proceedings of the 6th WEBKDD workshop: Webmining and Web Usage Analysis (WEBKDD'04), in conjunction with the 10th ACM SIGKDD conference (KDD'04). Seattle, Washington, USA, 2004, S. 71-82.
- [18] Datta, A.; Dutta, K.; D., V.; Ramamritham, K.; Navathe, S. B.: An Architecture to Support Scalable Online Personalization on the Web. In: *VLDB Journal* 10 (2001) 1, S. 104-117.
- [19] Elgin, B.: Web Search for Tomorrow. In: *Business Week* (2004) 3883, S. 48.
- [20] Ellwein, C.: Suche im Internet für Industrie und Wissenschaft. Oldenbourg Industrieverlag GmbH, München 2002.
- [21] Fang, X.; Salvendy, G.: Keyword Comparison: A user-centered feature improving web search tools. In: *International Journal of Human-Computer Studies* 52 (2001), S. 915-931.
- [22] Fitzpatrick, L.; Dent, M.: Automatic Feedback Using Past Queries: Social Searching? In: Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM Press: New York, NY, 1997, S. 306-313.
- [23] Gauch, S.; Chaffee, J.; Pretschner, A.: Ontology-Based Personalized Search and Browsing. In: *Web Intelligence and Agent Systems* 1 (2004) 3-4, S. 219-234.
- [24] Gauch, S.; Speretta, M.: Personalizing Search Based on User Search Histories. In: Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI'05). Compiegne University of Technology, France, 2005, S. 622-628.
- [25] Gordon, M.; Pathak, P.: Finding information on the World Wide Web: the retrieval effectiveness of search engines. In: *Information Processing and Management* 35 (1999), S. 141-180.
- [26] Gravano, L.; Garcia-Molina, H.; Tomasic, A.: GLOSS: Text-source discovery over the Internet. In: *ACM Transactions on Database Systems* 24 (1999) 2, S. 229-264.
- [27] Greco, G.; Greco, S.; Zumpano, E.: Collaborative Filtering Supporting Web Site Navigation. In: *AI Communications* 17 (2004), S. 155-166.
- [27] Griesbaum, J.: Evaluations of three German search engines: Altavista.de, Google.de and Lycos.de. In: *Information Research* 9 (2004) 4, online verfügbar: <http://informationr.net/ir/9-4/paper189.html> [Abruf: 2006-11-28].
- [28] Haveliwala, T. H.: Topic-sensitive PageRank. In: *IEEE Trans. Knowl. Data Eng.* 15 (2003) 4, S. 784-796.
- [29] Hearst, M.; Elliott, A.; English, J.; Sinha, R.; Swearingen, K.; Yee, K.-P.: Finding the Flow in Web Site Search. In: *Communications of the ACM* 45 (2002) 9, S. 42-49.
- [30] Hirsh, H.; Basu, C.; Davison, B. D.: Learning to Personalize. In: *Communications of the ACM* 43 (2000) 8, S. 102-106.
- [31] Hevner, A. R.; March, S. T.; Park, J.; Ram, S.: Design Science in Information Systems Research. In: *MIS Quarterly* 28 (2004) 1, S. 75-105.
- [32] Jansen, B. J.; Molina, P. R.: The effectiveness of Web search engines for retrieving relevant ecommerce links. In: *Information Processing and Management* 42 (2006), S. 1075-1098.
- [33] Jansen, B. J.; Spink, A.; Saracevic, F.: Real life, real user, and real needs: A study and analysis of user queries on the web. In: *Information Processing and Management* 36 (2000), S. 207-227.
- [34] Jeh, G.; Widom, J.: Scaling Personalized Web Search. In: Proceedings of the 12th International World Wide Web Conference. Computer Science Department, Stanford University: 2002, S. 1-24.
- [35] Kalbach, J.: "I'm feeling lucky": The Role of Emotions in Seeking Information on the Web. In: *Journal of the American Society for Information Science and Technology* 57 (2006) 6, S. 813-818.
- [36] Karat, C. M.; Brodie, C.; Karat, J.; Vergo, J.; Alpert, S. R.: Personalizing the user experience on ibm.com. In: *IBM Systems Journal* 42 (2003) 4, S. 686-701.
- [37] Ke, Y.; Deng, L.; Ng, W.; Lee, D.-L.: Web dynamics and their ramifications for the development of Web search engines. In: *Computer Networks* 50 (2005), S. 1430-1447.
- [38] Kelly, D.: Measuring Online Information Seeking Context, Part 1: Background and Method. In: *Journal of the American Society for Information Science and Technology* 57 (2006) 13, S. 1729-1739.
- [39] Kelly, D.: Measuring Online Information Seeking Context, Part 2: Findings and Discussion. In: *Journal of the American Society for Information Science and Technology* 57 (2006) 14, S. 1862-1874.
- [40] Koutrika, G.; Ioannidis, Y.: A Unified User Profile Framework for Query Disambiguation and Personalization. In: Proceedings of the Workshop on New Technologies for Personalized Information Access (PIA). Edinburgh, UK, 2005, S. 44-53.

- [41] *Komlodi, A.; Marchionini, G.; Soergel, D.*: Search history support for finding and using information: User interface design recommendations from a user study. In: *Information Processing and Management* 43 (2006), S. 10-29.
- [42] *Koshman, S.; Spink, A.; Jansen, B. J.*: Web Searching on the Vivisimo Search Engine. In: *Journal of the American Society for Information Science and Technology* 57 (2006) 14, S. 1875-1887.
- [43] *Kruschwitz, U.; Al-Bakour, H.*: Users want more sophisticated Search Assistants: Results of a Task-based Evaluation. In: *Journal of the American Society for Information Science and Technology* 56 (2005) 13, S. 1377-1393.
- [44] *Kramer, J.; Noronha, S.; Vergo, J.*: A User-Centered Design Approach to Personalization. In: *Communications of the ACM* 43 (2000) 8, S. 45-48.
- [44] *Lawrence, S.*: Context in Web Search. In: *IEEE Data Engineering Bulletin* 23 (2000) 3, S. 25-32.
- [45] *Levy*: All Eyes on Google. In: *Newsweek* 143 (2004) 13, S. 48-58.
- [46] *Liu, H.; Lieberman, H.; Selker, T.*: GOOSE: A Goal-Oriented Search Engine With Commonsense. In: *Proceedings of the 2002 International Conference on Adaptive Hypermedia and Adaptive Web Based Systems* (2002), S. 253-263.
- [47] *Liu, F.; Yu, C.; Meng, W.*: Personalized Web Search by Mapping User Queries to Categories. In: *Proceedings of the 11th International Conference on Information and Knowledge management*. McLean, Virginia, USA, 2002, S. 558-565.
- [48] *Liu, F.; Yu, C.; Meng, W.*: Personalized Web Search for Improving Retrieval Effectiveness. In: *IEEE Transactions on Knowledge and Data Engineering* 16 (2004), S. 28-40.
- [49] *Lundquist, C.; Grossman, D. A.; Friedler, O.*: Improving relevance feedback in the vector space model. In: *Proceedings of the 6th International Conference on Information and Knowledge Management*. New York, 1997, S. 16-23.
- [50] *Manber, U.; Patel, A.; Robison, J.*: Experience with Personalization on Yahoo! In: *Communications of the ACM* 43 (2000) 8, S. 35-39.
- [51] *Marchionini, G. N.*: *Information Seeking in Electronic Environments*. Cambridge University Press, Cambridge 1995.
- [52] *Martzoukou, K.*: A review of Web information seeking research: considerations of method and foci of interest. In: *Information Research* 10 (2005) 2, online verfügbar: <http://informationr.net/ir/10-2/paper215.html> [Abruf: 2006-11-28].
- [53] *Mesquita, C.; Barbosa, S.; de Lucena, C.*: Towards the Identification of Concerns in Personalization Mechanisms Via Scenarios. In: *Proceedings of the 1st Conference on Aspect-Oriented Requirements Engineering and Architecture Design*. 2002, S. 25-31.
- [54] *Mertens, P.; Stöflein, M.; Zeller, T.*: *Personalisierung und Benutzermodellierung in der betrieblichen Informationsverarbeitung - Stand und Entwicklungsmöglichkeiten*. Arbeitsbericht Nr. 2, Bereich Wirtschaftsinformatik I, 2004.
- [55] *Mild, A.; Natter, M.*: Collaborative filtering or regression models for Internet recommendation systems? In: *Journal of Targeting, Measurement and Analysis for Marketing* 10 (2002) 4, S. 304-313.
- [56] *Mowshowitz, A.; Kawaguchi, A.*: Bias on the Web. In: *Communications of the ACM* 45 (2002) 9, S. 56-60.
- [57] *Makris, C.; Panagis, Y.; Sakkopoulos, E.; Tsakalidis, A.*: Category ranking for personalized search. In: *Data & Knowledge Engineering* 60 (2007), S. 109-125.
- [58] *Notess, G. R.*: Customization Options for Web Searching. In: *Online* 25 (2001) 1, S. 55-57.
- [59] *Page, L.; Brin, S.; Motwani, R.; Winograd, T.*: *The Pagerank Citation Ranking: Bringing Order to the web*. Technical Report, 1998.
- [60] *Perkowitz, M.; Etzioni, O.*: Towards adaptive Web sites: Conceptual framework and case study. In: *Artificial Intelligence* 118 (2000), S. 245-275.
- [61] *Prakash, K. S. S.; Raghavan, S. V.*: *Intelligent Search Engine: Simulation to Implementation*. In: *Proceedings of the iiWAS 2004 - The sixth International Conference on Information Integration and Web-based Applications Services*, 27-29 September. Austrian Computer Society: Jakarta, Indonesien, 2004.
- [62] *Pitkow, J.; Schütze, H.; Cass, T.; Cooley, R.; Turnbull, D.; Edmonds, A.; Adar, E.; Breuel, T.*: Personalized Search. In: *Communications of the ACM* 45 (2002) 9, S. 50-55.
- [63] *Ravindran, D.; Gauch, S.*: Exploiting Hierarchical Relationships in Conceptual Search. In: *Proceedings of CIKM 2004* (2004), S. 238-239.
- [64] *Resnick, M. L.; Vaughan, M. W.*: Best Practices and Future Visions for Search User Interfaces. In: *Journal of the American Society for Information Science and Technology* 57 (2006) 6, S. 781-787.
- [65] *Riemer, K.*: Grundzüge der internetbasierten Personalisierung. In: *Piller, F. T.; Stotko, C. M. (Hrsg.) Mass Customization und Kundenintegration*. Symposium, Düsseldorf 2003, S. 1-17.
- [66] *Rose, D. E.*: Reconciling Information-Seeking Behavior with Search User Interfaces for the Web. In: *Journal of the American Society for Information Science and Technology* 57 (2006) 6, S. 797-799.

- [67] *Runte, M.*: Personalisierung im Internet. DUV, Wiesbaden 2000.
- [68] *Salton, G.; Buckley, C.*: Improving retrieval performance by relevance feedback. In: *Journal of the American Society for Information Science* 41 (1990) 4, S. 288-297.
- [69] *Schackmann, J.; Schü, J.*: Personalisierte Portale. In: *Wirtschaftsinformatik* 43 (2001) 6, S. 623-625.
- [70] *Semeraro, G.; Degemmis, M.; Lops, P.; Palmisano, I.*: WordNet-based User Profiles for Semantic Personalization. In: *Proceedings of the Workshop on New Technologies for Personalized Information Access (PIA)*. Edinburgh, UK, 2005, S. 74-83.
- [71] *Smyth, B.; Balfe, E.*: Anonymous personalization in collaborative web search. In: *Information Retrieval* 9 (2006) 2, S. 165-190.
- [72] *Spink, A.; Xu, J. L.*: Selected Results from a large study of Web searching: the Excite study. In: *Information Research* 6 (2000) 1, online verfügbar: <http://informationr.net/ir/6-1/paper90.html> [Abruf 2006-11-22].
- [73] *Sugiyama, K.; Hatano, K.; Yoshikawa, M.*: Adaptive Web Search Based on User Profile Constructed without Any Effort from Users. In: *Proceedings of the International WWW Conference*. New York, USA, 2004, S. 675-684.
- [74] *Swartz, N.*: Amazon's New Offering Manages Searches. In: *Information Management Journal* 38 (2004) 6, S. 16.
- [75] *Teevan, J.; Dumais, S. T.; Horvitz, E.*: Beyond the commons: Investigating the value of personalizing Web search. In: *Proceedings of the Workshop on New Technologies for Personalized Information Access (PIA)*. Edinburgh, UK, 2005, S. 84-92.
- [76] *Teevan, J.; Dumais, S. T.; Horvitz, E.*: Personalizing search via automated analysis of interests and activities. In: *Proceedings of SIGIR '05* (2005), S. 449-456.
- [77] *Teuteberg, F.*: Agentenbasierte Informationserschließung im World Wide Web unter Einsatz von künstlichen neuronalen Netzen und Fuzzy-Logik. Eul, Lohmar 2001.
- [78] *Tonta, Y.*: Internet and electronic information management. In: *Information Services & Use* 25 (2005), S. 3-12.
- [79] *Vassilvitskii, S.; Brill, E.*: Using Web-Graph Distance for Relevance Feedback in Web Search. In: *Proceedings of the SIGIR '06*. Seattle, 2006, S. 147-153.
- [80] *Volokh, E.*: Personalization and Privacy. In: *Communications of the ACM* 43 (2000) 8, S. 84-88.
- [81] *White, R. W.; Ruthven, I.*: A Study of Interface Support Mechanisms for Interactive Information Retrieval. In: *Journal of the American Society for Information Science and Technology* 57 (2006) 7, S. 933-948.
- [82] *Wiedmann, K.-P.; Buxel, H.; Walsh, G.*: Customer profiling in e-commerce: Methodological aspects and challenges. In: *Journal of Database Marketing* 9 (2002) 2, S. 170-184.
- [83] *Wildemuth, B. M.*: Evidence-based Practice in Search Interface Design. In: *Journal of the American Society for Information Science and Technology* 57 (2006) 6, S. 825-828.
- [84] *Wusteman, J.; O'hleada, P.*: Using Ajax to Empower Dynamic Searching. In: *Information Technology and Libraries* 25 (2006) 2, S. 57-63.
- [85] *Zeng, H.-J.; He, Q.-C.; Chen, Z.; Ma, W.-Y.; Ma, J.*: Learning to cluster web search results. In: *Proceedings of SIGIR '04* (2004), S. 210-217.