

Interaktives Retrieval und situationsabhängige Vorschläge

Sascha Kriewel

Zusammenfassung In diesem Artikel wird beschrieben, wie ein adaptives Vorschlagssystem für Suchstrategien Interaktives Retrieval unterstützen kann. Ein Benutzerexperiment mit 24 Teilnehmern zeigte, dass ein solches System Suchenden hilft, erfolgreichere Strategien einzusetzen, als Suchende ohne Unterstützung. Die Ergebnisse lassen eine Korrelation zwischen dem Sucherfolg der Teilnehmer (gemessen an der Zahl relevanter gespeicherter Dokumente) und dem Einsatz von Vorschlägen erkennen. Durch Vorschläge unterstützte Suchende setzten zudem signifikant öfter fortgeschrittene Werkzeuge und Optionen des Suchsystems ein – auch nach Abschaltung der Vorschläge während der letzten Aufgabe des Experiments.

1 Einleitung

Noch bis in die 80er Jahre waren die Hauptnutzer von Information-Retrieval-Systemen meist Suchexperten oder Informationsmittler [27], die für ihre Arbeit effektive Suchstrategien entwickelten. Heute lösen Endbenutzer ihre Informationsprobleme in direkter Interaktion mit dem System [34]. Doch selbst mit der Domäne und dem Suchwerkzeug vertrauten Endnutzern fehlt oft das Verständnis effektiver Suchstrategien, um die angebotenen Werkzeuge effizient und effektiv einzusetzen [26,8]. Moderne IR-Systeme bieten Benutzern zahl-

reiche zusätzliche Funktionalitäten zum erfolgreicherem Suchen von und Arbeiten mit Information. Selten aber unterstützen sie Benutzer bei der Wahl der richtigen (nützlichsten) Taktik in einer bestimmten Situation einer langen, interaktiven Suche [9,10,17].

Suchanfänger versuchen oft, mehrere Konzepte zugleich auszudrücken, und über- oder unterspezifizieren ihre Anfragen [29]. Wenn sie umformulieren, dann meist unter Einsatz paralleler Taktiken oder durch Hinzunahme von Begriffen zur ursprünglichen Anfrage [30]. Es fällt ihnen schwer, Suchbegriffe zu finden, die ihr Informationsbedürfnis hinreichend beschreiben und dabei hilfreiche Resultate erbringen [13]. Nur selten setzen sie ausgefeiltere Suchstrategien ein [11], stattdessen werden aus Unkenntnis teils kontraproduktive Aktionen unternommen. Auch können Benutzer die fortgeschrittenen Suchoptionen und Werkzeuge eines Systems selten erfolgreich einsetzen [27]. Suchexperten hingegen verfügen über effektive Informationssuchstrategien, bei denen sie in Interaktion mit dem System und abhängig von ihrer bisherigen Interaktion und den aktuellen Ergebnissen zu einer erfolgreichen Suche gelangen [13].

Für komplexe Aufgaben, etwa die Recherche unterschiedlicher Standpunkte zu einer kontroversen Fragestellung, interagieren Suchende in langen Suchsitzungen mit dem System und erwarten nicht, alle ihre Antworten als Ergebnis einer einzelnen perfekten Anfrage zu finden. Zwar gibt es zahlreiche Arbeiten, die Benutzer bei der Umsetzung einzelner Aktionen, Taktiken oder Strategeme unterstützen, doch umfangreiche, interaktive Sitzungen erfahren kaum direkte Unterstützung.

Dieser Artikel soll eine solche situative Suchunterstützung für interaktives Information Retrieval vorstellen. Es werden dabei auch die Ergebnisse zweier früher veröffentlichter Studien präsentiert [23,24]. In Benutzerexperimenten wurden drei Fragen untersucht: sind Benutzer, die Unterstützung erhalten, erfolgreicher

Dr.-Ing. Sascha Kriewel
Universität Duisburg-Essen
Abteilung für Informatik und Angew. Kognitionswissenschaft
Tel.: +49-203-3793944
Fax: +49-203-3793557
E-Mail: sascha.kriewel@uni-due.de

über die Dauer einer kompletten Sitzung (im Gegensatz zu einer einzelnen Anfrage), verwenden sie dabei fortgeschrittene Funktionalitäten und können sie aus Vorschlägen des Systems lernen?

2 Strategische Unterstützung

Mit dem Begriff Taktik beschreibt man Benutzeraktionen mit dem Ziel, eine Suche voranzubringen (zum Beispiel den Austausch eines Suchterms durch einen gebräuchlicheren). Strategeme sind komplexere Aktivitäten und bestehen meist aus einer Reihe von Taktiken und Aktivitäten, um einen klar definierten und strukturierten Suchraum (beispielsweise ein Zitationsnetz) effizient zu durchsuchen [3]. Suchstrategien sind schließlich Pläne zur Bearbeitung von Suchaufgaben, die zahlreiche für Informationssuchen nützliche Taktiken und Strategeme umfassen können [5]. Um eine solche Strategie umzusetzen, müssen Suchende die für ihr Problem passenden Aktionen aus den vom Suchsystem bereitgestellten auswählen. Dabei ist die Suche selbst ein interaktiver und opportunistischer Prozess, bei dem die Suchziele sich während der Aufgabe ändern können [33]. Das fruchtbarste Vorgehen ist oft, vielversprechende Gelegenheiten zu verfolgen, die sich während der Suche aus den bisherigen Interaktionen ergeben. Bates beschreibt dies in ihrem *berry picking model* der Informationssuche [4]. Järvelin [20] hat gezeigt, dass interaktive Suchsitzungen aus für sich ineffektiven Anfragen insgesamt effektiver sein können als einzelne, umfassende Anfragen (abhängig von der Berücksichtigung der Kosten und Nutzen interaktiver Suchsitzungen bei der Bewertung des Sucherfolgs).

Es ist daher wichtig, dass Suchende diese strategischen Gelegenheiten während einer Suchsitzung erkennen und die strategischen Optionen nutzen, um „Sackgassen“ zu vermeiden. Leider setzen Benutzer aber nur selten die fortgeschrittenen Möglichkeiten ihres Systems ein, selbst wenn diese ihre Suche voranbringen würden, entweder weil sie diese nicht kennen, sie nicht verstehen oder nicht abschätzen können, in welcher Situation sie sich effektiv nutzen lassen. Suchsysteme mit strategischer Unterstützung können die Sucheffektivität steigern, indem sie solche fortgeschrittenen Möglichkeiten in geeigneter Situation automatisch vorschlagen [18].

Drabenstott [12] untersuchte, ob von Domänenexperten eingesetzte Suchstrategien durch Nichtexperten genutzt werden und kam zu dem Schluss, dass Systemunterstützung nötig ist, um Nichtexperten von ihrer üblichen Suchweise zu effektiveren Strategien zu führen. Auch Wildemuth [32] wies darauf hin, dass gegenwärtige Suchsysteme wenig Unterstützung bieten, um Suchenden beim Gestalten effektiver Strategien zu helfen.

Brajnik et al. beschreiben in [10] ein strategisches Hilfssystem, das auf kollaborativem *Coaching* aufbaut und versucht, Benutzer mit Vorschlägen und Hinweisen in sogenannten kritischen oder verbesserbaren Situationen zu helfen. Das System benutzt eine Wissensbasis aus 94 Produktionsregeln, die ursprünglich von Marcia Bates beschriebene Taktiken und Strategeme [3,5] vorschlagen. Zu den Taktiken gehören etwa das Trunkieren von Suchtermen oder das Finden verwandter Begriffe, als Strategem wird beispielsweise die Suche in zur Anfrage passenden Journaltiteln angeboten. Das Strategie-modul wurde in FIRE integriert, einem Benutzerinterface für ein Boolesches Retrievalsystem. Nur sechs Personen nahmen an einer Benutzerevaluation teil, aber sie lieferte Hinweise auf die Nützlichkeit strategischer Vorschläge.

Jansen und McNeese [19] evaluierten die Effektivität automatischer Suchunterstützung in einer Studie mit 40 Teilnehmern (von denen jedoch nur 30 auch die angebotene Unterstützung nutzten). Sie fanden dabei heraus, dass automatische Unterstützung die Leistung der Suchenden verbessern kann, abhängig davon, an welchen Kriterien man diese misst. Die Unterstützung brachte signifikante Verbesserungen bei der Auswahl relevanter Dokumente durch die Teilnehmer, aber nicht bei der Konstruktion präziser Anfragen. Es zeigte sich aber auch, dass die Hälfte der die Unterstützung nutzenden Suchenden nicht von dieser profitierte und dass automatische Assistenz zielgerichtet und auf den einzelnen Suchenden und seine Situation angepasst sein sollte, um die besten Ergebnisse zu erzielen.

3 Das Vorschlagswerkzeug

Zur strategischen Hilfe bei komplexen, interaktiven Suchen wurde daher ein adaptives Vorschlagswerkzeug für das Suchsystem DAFFODIL entwickelt. DAFFODIL ist ein Suchsystem für Digitale Bibliotheken, das an der Universität Duisburg-Essen entwickelt und genutzt wurde [14,25].¹ Es bietet ein reiches Angebot an Werkzeugen, die erfahrene Benutzer in ihrer Suche einsetzen können. Frühere Benutzerexperimente hatten jedoch gezeigt, dass unerfahrene Benutzer Probleme bei der effektiven Nutzung dieser Werkzeuge hatten [21].

Für die Experimente wurde zur Herstellung einer kontrollierten Umgebung eine reduzierte Version DAFFODILS verwendet. Diese umfasst ein Suchwerkzeug, das Digitale Bibliotheken über ein Anfrageformular mit vier Suchfeldern abfragt („Titel“, „Autor“, „Jahr“ und „Freitext“). Die Suche unterstützt Boolesche Ausdrücke und

¹ das System wird inzwischen unter dem neuen Namen EZDL weiterentwickelt, <http://www.ezdl.de>

Phrasensuche. Resultate aus verschiedenen Bibliotheken werden zusammengeführt und in eine gemeinsame Rangfolge gebracht. Eine Suchhistorie zeigt alle bisherigen Suchen des Benutzers und erlaubt die Wiederverwendung und die Bearbeitung früherer Anfragen. Relevante Dokumente, Anfragen, Suchbegriffe oder Autoren können in einer Ablage gespeichert werden. Eine Detailansicht zeigt eine kurze Zusammenfassung einzelner Dokumente aus dem Resultat oder der Zwischenablage mit Links zu den Volltexten. Die Dokumentdetails sind interaktiv und Autoren oder Schlüsselbegriffe können direkt benutzt werden, um neue Anfragen zu generieren, die aktuelle Anfrage zu modifizieren oder andere Werkzeuge aufzurufen. Zusätzlich werden eine Reihe unterstützender Werkzeuge angeboten:

- Extraktion häufiger Begriffe und Autoren aus Resultat oder relevanten Dokumenten in der Ablage
- Verwandte Begriffe zur aktuellen Suchanfrage
- ein Thesaurus mit Synonymen, Ober- und Unterbegriffen, sowie Definitionen
- ein Autorennetzwerk zur Darstellung von Koautorenbeziehungen eines ausgewählten Autoren
- ein Klassifikationsbrowser für die Suchdomäne

3.1 Vorschlagssystem

Das eigentliche Vorschlagssystem wird in [22] und [23] beschrieben und besteht aus drei Hauptkomponenten. Die Beobachtungskomponente sammelt Informationen über die aktuelle Benutzersituation, indem sie Nachrichten des Systems auswertet. Die Inferenzkomponente hält eine Fallbasis mit Suchsituationen in einer relationalen Datenbank vor und vergleicht eine vorliegende Situation mit den gespeicherten Fällen (siehe auch Abschnitt 3.2).

Die gewichteten Vorschläge werden dann an die Benutzerschnittstelle geschickt, wo das Vorschlagswerkzeug diese an die aktuelle Situation bezüglich verwendeter Suchbegriffe oder vorhandener Ergebnisse anpasst und den Benutzer über das Angebot neuer Vorschläge informiert. Suchende können die angebotenen Vorschläge entweder ignorieren oder sie sich in einer Rangliste präsentieren lassen. In der Rangliste zeigt ein graphischer Balken für jeden Vorschlag eine Wahrscheinlichkeitsschätzung an, ob dieser Vorschlag in der aktuellen Situation hilfreich ist (siehe Abb. 1). Die Vorschläge sind mit einer kurzen Erklärung ihres Zwecks und der Verwendung versehen. Die meisten der Vorschläge erlauben die direkte Ausführung durch einen Doppelklick.

Die bereitgestellten Vorschläge stammen größtenteils aus der Literatur [3,5,15,16]. Dazu wurden auch Vorschläge aufgenommen, die spezifische Probleme

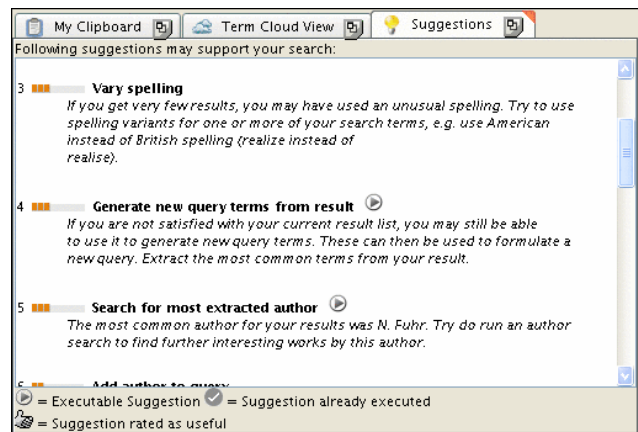


Abb. 1 Das Vorschlagswerkzeug [23]

adressieren, die bei früheren Evaluationen das DAFFODIL-Systems beobachtet wurden [21,31]:

- terminologische Vorschläge (Schreibvarianten, verwandte Begriffe, Synonyme, Unter-/Oberbegriffe),
- Vorschläge bezüglich der Suchoperatoren und Felder (Disjunktion oder Konjunktion, Phrasenoperatoren, Titel- oder Freitextfelder, Einschränkungen nach Jahren oder Autoren),
- strategische Vorschläge (Zerlegung der Anfrage entlang in ihr vertretener Konzepte in einzelne Facetten, Pearl growing von einem relevanten Dokument², Vermeidung von Überspezifikation, Autorensuche),
- Vorschläge zu fortgeschrittenen Werkzeugen des DAFFODIL-Systems (Koautorennetzwerk, Klassifikationsbrowser, Filterung auf der Resultatliste, Extraktion von Begriffen aus Ergebnismenge oder relevanten Dokumenten).

3.2 Auswahl der Vorschläge

Fallbasiertes Schließen (*case-based reasoning*, CBR) wird benutzt, um die angemessensten Vorschläge für die Suchsituation des Benutzers zu finden und in eine Rangordnung zu bringen. Ein solcher CBR-Ansatz für ein interaktives IR-System wurde bereits früher von Belkin et al. vorgeschlagen [7]. Sie nutzten den Ansatz im MERIT-System, um skriptbasierte Interaktion mit dem System zu ermöglichen, basierend auf vorgegebenen Fällen für bestimmte Suchstrategien [6].

In dem hier beschriebenen Vorschlagssystem sind Fälle gespeicherte Suchsituationen. Ihre Lösungen sind

² Bei der Suchstrategie des Pearl growing wird zunächst ein relevantes Dokument identifiziert. Dieses nutzt der Suchende dann, um über Deskriptoren, Klassifikationsbegriffe, Titeltermine, Zitationen oder Referenzen weitere relevante Dokumente zu finden. Mit den neuen Dokumenten als Quelle kann die Strategie iterativ werden.

Vorschläge, die andere Benutzer in solchen Situationen nützlich fanden. Benutzer können die Vorschläge bezogen auf ihre aktuelle Situation als nützlich oder nicht nützlich bewerten, wodurch sie in die Fallbasis eingehen. Jeder Fall besteht aus Aspekten, die benutzt werden, um Fälle zu vergleichen und Ähnlichkeiten zu berechnen. Grob lassen sie sich in drei Kategorien unterteilen: numerische Aspekte (z. B. Anzahl der Resultate), Termmengen (z. B. die Anfrageterme) und gewichtete Termvektoren (z. B. häufige Begriffe aus dem aktuellen Ergebnis mit Anzahl des Vorkommens).

Nach jeder Suchaktion wird eine neue Situation zusammengestellt und die Inferenzkomponente ermittelt die ähnlichsten Fälle für alle verfügbaren Lösungen. Die Ähnlichkeit zwischen zwei Situationen wird dabei als das gewichtete Mittel der einzelnen Ähnlichkeitswerte der Aspekte einer Situation berechnet. Die technische Umsetzung des Werkzeugs und die Berechnung der Ähnlichkeiten werden ausführlicher in [22] behandelt.

Um eine Vorschlagsrangliste zu präsentieren, werden sowohl positive als auch negative Fälle genutzt: ähnliche Situationen, in denen der Benutzer einen bestimmten Vorschlag als hilfreich bzw. nicht hilfreich bewertet hat. Eine Variante der von Ontañón und Plaza [28] vorgeschlagenen Methode zur Berechnung eines Vertrauenswertes aus positiven und negativen Fällen wird benutzt.

4 Benutzerexperimente

Es wurden zwei Benutzerexperimente zur Evaluation des Unterstützungsansatzes durchgeführt. Das erste Experiment [23] war eine leichtgewichtige Pilotstudie mit zwei wichtigen Zielen:

1. einen ersten Eindruck über die Benutzerakzeptanz für strategische Vorschläge zu gewinnen, und
2. die Angemessenheit des Rankings der Vorschläge bezüglich der Ähnlichkeit der Benutzersituation zu den Situationen früherer Benutzer zu evaluieren.

Zur Vorbereitung des Experiments wurde das System mit einer Suchaufgabe trainiert, die thematisch keinen Bezug zur Suchaufgabe hatte, die für das Experiment selbst genutzt wurde. So wurde eine Fallbasis bereitgestellt, in der für jeden Vorschlag zumindest eine Situation (ein Fall) existiert, für den der Vorschlag ein passendes Vorgehen darstellt. Es wurden für die erste Evaluation keine negativen Fälle verwendet, sowie stets alle verfügbaren Vorschläge ungeachtet des ermittelten Gewichts angezeigt. Es wurden sechzehn Vorschläge umgesetzt.

Insgesamt zwölf Teilnehmer bearbeiteten eine komplexe, umfangreiche Suchaufgabe mit Hilfe der DAFFO-

DIL-Software und des Vorschlagswerkzeugs. Dazu wurde ihnen eine DAFFODIL-Version mit einer Teilmenge der existierenden Werkzeuge zur Verfügung gestellt, und die Suchen selbst wurden über mehreren Artikelkollektionen aus der Domäne der Informatik durchgeführt.

Das Experiment zeigte einige Verbesserungsmöglichkeiten auf, aber insgesamt wurde das Vorschlagswerkzeug von den Benutzern positiv aufgenommen. Die Teilnehmer des Experimentes empfanden die automatischen, aber unaufdringlichen Suchtipps als hilfreich und konnten die vorgeschlagenen Taktiken erfolgreich umsetzen, um ihre Suche voranzubringen.

Die gefundenen Probleme in der Bedienbarkeit und der Darstellung wurden behoben. Die Zahl der verfügbaren Vorschläge wurde seitdem erweitert, um eine breitere Spanne von Taktiken und Strategien abzudecken (auf insgesamt 22). In einem zweiten Benutzerexperiment [24] wurde dann untersucht, ob ein adaptives Vorschlagswerkzeug Benutzern helfen kann, besser und erfolgreicher zu suchen und dabei zu lernen, die fortgeschrittenen Möglichkeiten eines komplexen Suchsystems einzusetzen.

4.1 Versuchsteilnehmer

Für das Experiment wurden 20 männliche und 4 weibliche Studierende der Angewandten Informatik, Angewandten Kognitions- und Medienwissenschaften und verwandter Studiengänge an der Universität Duisburg-Essen rekrutiert. Die freiwilligen Teilnehmer stammten aus unterschiedlichen Semestern und erhielten zusätzliche Übungspunkte für ihre Teilnahme. Das Alter der Teilnehmer lag zwischen 22 und 48 Jahren (durchschnittlich 27,25 Jahre).

In einem einleitenden Fragebogen wurde die bisherige Sucherfahrung der Teilnehmer erhoben. Auf einer fünf-stufigen Likertskala bewerteten sich zwei Teilnehmer als unerfahren, sechs als etwas erfahren, 15 als erfahren und ein Teilnehmer als sehr erfahren. Niemand schätzte sich selbst als sehr unerfahren ein.

Da Selbstevaluation von Suchfertigkeit für einige Aufgabentypen nur ein schwacher Indikator für tatsächliche Suchleistung ist [1], wurden die Teilnehmer auch gefragt, seit wann und wie häufig sie Suchsysteme einsetzen. Die durchschnittliche Erfahrung betrug 4,75 Jahre. Nur sechs Teilnehmer benutzten Suchmaschinen seltener als täglich. Die Suchen sind jedoch vielfach recht einfach und die derart gesammelte Erfahrung nicht auf komplexe Suchen übertragbar.

Nach vertrauten Suchsystemen gefragt, nannten 23 der Teilnehmer Google, vier Yahoo und zwei MSN Search bzw. Bing. Drei der Studenten benutzten

Desktop-Suchsysteme wie etwa Google Desktop oder Beagle für Linux. Drei Studierende benutzten Digitale Bibliotheken. Alle Benutzer waren mit DAFFODIL unerfahren oder kannten das System nicht. Keiner benutzte es regelmäßig.

4.2 Versuchsaufbau

Für die Studie kamen zwei Varianten der DAFFODIL-Software zum Einsatz. Die Systeme waren bis auf den Einsatz des Vorschlags-Werkzeuges in einem der beiden identisch. Es wurde in sechs Digitalen Bibliotheken der Informatik-Domäne gesucht.

Alle Teilnehmer erhielten drei komplexe Suchaufgaben, für deren Bearbeitung ihnen jeweils 20 Minuten zur Verfügung standen. Während dieser Zeit waren so viele relevante Dokumente wie möglich zu finden und zu speichern. Die Hälfte der Teilnehmer nutzte das System mit Vorschlägen für die ersten beiden Aufgaben und ohne für die letzte (Gruppe A). Die andere Hälfte benutzte das System ohne Unterstützung für alle Aufgaben (Gruppe B). Die Teilnehmer wurden vorab nicht über den Zweck der Studie informiert und es wurde ihnen nicht mitgeteilt, dass einige von ihnen Suchunterstützung durch Vorschläge erhielten und einige nicht. Die Zuordnung der Teilnehmer zu den beiden Gruppen erfolgte zufällig. Es gab keine signifikanten Unterschiede zwischen den Gruppen bezüglich durchschnittlicher Sucherfahrung (4,83 zu 4,67 Jahre, Welch-Test: $p = 0,872$) oder ihrer Selbsteinschätzung.

Die Aufgaben wurden passend zur Kollektion ausgewählt, so dass sie realen Arbeitsaufgaben ähnelten. Die Reihenfolge der Aufgabenbearbeitung wurde rotiert, so dass alle Aufgaben gleich oft als erste, zweite oder dritte Aufgabe von den Teilnehmern bearbeiteten wurden:

- a. **Gesundheit:** „Suchen Sie nach Artikeln, um zu beantworten, ob und inwieweit die ständige Arbeit an einem Computer- bzw. Bildschirmarbeitsplatz zu gesundheitlichen Problemen führen kann. Relevante Dokumente sind Studien oder Berichte über gesundheitliche Probleme bei der Bildschirmarbeit. Dazu zählen etwa Augenüberbeanspruchung, Haltungsschäden, Muskel- oder Sehnenscheidenprobleme oder psychische Probleme (auch Stress).“
- b. **Plagiate:** „Suchen Sie Artikel, die Methoden beschreiben oder bewerten, um automatisch Software-Plagiate zu erkennen. Relevante Dokumente stellen Methoden zur Plagiaterkennung von Software-Quellcode vor, schildern wie existierende Verfahren für diese Aufgabe genutzt werden können oder beschreiben Evaluationen bzw. Vergleiche verschiedener Ansätze. Relevant sind auch Methoden zur

Feststellung von Autorenschaft von Programmcode. Nicht relevant sind Methoden, die sich hauptsächlich oder ausschließlich mit der Plagiaterkennung bei digitalen Textdokumenten wie z. B. journalistischen oder wissenschaftlichen Artikeln befassen.“

- c. **Java:** „Suchen Sie zum Aufbau und zur Untermauerung einer Argumentation für oder gegen Java als Unterrichtssprache im ersten Semester einer Informatikerausbildung geeignete Artikel. Relevant sind Artikel, die sich mit Java als Lehr- oder Unterrichtssprache auseinandersetzen. Ebenfalls relevant sind Vergleiche mit anderen Programmiersprachen für diese Zwecke, sowie Artikel zu allgemeinen Kriterien und Anforderungen an eine solche Lehrsprache. Nicht relevant sind Texte, die sich mit dem Erlernen und Unterrichten von Java beschäftigen.“

Alle Teilnehmer erhielten eine identische Einführung in das Suchsystem DAFFODIL. Zusätzlich wurde den Teilnehmern in der unterstützten Gruppe der Aufruf und das Ausführen der Vorschläge erklärt. Das Bewerten der Vorschläge wurde für das Experiment deaktiviert. Vor jeder Aufgabe erhielten die Teilnehmer eine schriftliche Aufgabenbeschreibung. Falls notwendig, wurden Unklarheiten über die Aufgabe beantwortet, bevor der Teilnehmer mit der Suche begann. Während der Suche wurden vom Protokollanten nur technische Fragen beantwortet und keine Unterstützung in Form von Ratschlägen oder Suchbegriffen gegeben.

4.3 Auswertung

Alle Benutzeraktivitäten und Anfragen während der Suche wurden in einem Aktivitätslog festgehalten. Zusätzlich wurden die Suchenden angehalten, während der Bearbeitung der Aufgaben laut zu denken. Ein Protokollant war während des Experiments anwesend, um diese Bemerkungen festzuhalten. Die Protokolle wurden später benutzt, um uneindeutige Logeinträge zu klären.

Aus den Suchlogs wurden zwei Teilhistorien extrahiert: (a) eine Historie aller Anfragen, die mit dem Suchwerkzeug von DAFFODIL gestellt wurden, (b) eine Historie aller Werkzeugverwendungen. Die Anfragehistorie wurde dann kodiert (siehe Tabelle 1). Die Werkzeughistorie wurde benutzt, um die durchschnittliche Werkzeugnutzung pro Aufgabe zu berechnen. Die Resultate sind in Tabelle 2 zu sehen.

Alle gespeicherten Dokumente wurden pro Aufgabe zusammengeführt. Zwei unabhängige Personen beurteilten diese Dokumente blind auf Relevanz. Die Vereinigungsmenge der als relevant beurteilten Dokumente wurden dann genutzt, um die Zahl der von jedem Su-

Code	Aktion	\bar{x} (Stdabw.)		Welch-Test		
		Gruppe A	Gruppe B	t	df	p
T	Term hinzu/weg	7,50 (5,95)	12,00 (7,62)	-1,61	20,78	0,122
S	Term ersetzen	11,08 (4,89)	16,92 (8,54)	-2,05	17,51	0,055
R	ganze Anfrage ersetzen	3,33 (2,57)	6,00 (4,05)	-1,93	18,64	0,069
O	Anfrage wiederholen	0,83 (1,11)	1,08 (1,31)	-0,50	21,44	0,620
F	zus./anderes Feld nutzen	7,08 (5,04)	0,92 (1,78)	3,99	13,71	0,001
B	Boolsch. Op. hinzu/weg	6,25 (6,16)	6,66 (6,92)	-0,16	21,71	0,877
C	Schreibweise korrigiert	2,25 (1,82)	2,42 (2,81)	-0,17	18,82	0,865
F_3	Felder nutzen (Aufg. 3)	1,75 (1,42)	0,25 (0,62)	3,35	15,05	0,004

Tabelle 1 Aktionen zur Änderung der Anfrage, Gruppe A erhielt Unterstützung für die ersten beiden Aufgaben [22]

Aktivität	\bar{x} (Stdabw.)		Welch-Test		
	Gruppe A	Gruppe B	t	df	p
Grundlegende Aktivitäten					
Anfrage ausführen	46,50 (9,76)	51,33 (19,89)	-0,82	17,01	0,423
Details betrachten	79,17 (45,47)	51,67 (28,50)	-1,78	18,49	0,092
Dokument speichern	38,17 (14,86)	23,25 (14,18)	2,52	21,95	0,019
Vorschläge					
Vorschläge anfordern	8,33 (3,11)	— (—)	—	—	—
Vorschlag ausführen	9,50 (4,01)	— (—)	—	—	—
Fortgeschrittene Aktivitäten					
Begriffe extrahieren	10,33 (6,53)	3,75 (3,89)	3,02	18,00	0,007
Thesaurus benutzen	4,91 (4,60)	0,50 (1,73)	2,99	12,58	0,011
Anfrage aus anderem Werkzeug ändern	1,17 (1,75)	0,92 (1,56)	0,37	21,73	0,716
Fortg. Aktivitäten (Aufg. 1)	4,67 (2,87)	2,33 (2,31)	2,19	21,03	0,039
Fortg. Aktivitäten (Aufg. 2)	3,42 (2,39)	1,16 (1,34)	2,84	17,26	0,011
Fortg. Aktivitäten (Aufg. 3)	4,25 (1,24)	1,33 (1,44)	3,24	16,52	0,005
Fortg. Aktivitäten (Aufg. 3)	2,67 (1,78)	1,25 (1,29)	2,48	21,52	0,021

Tabelle 2 Benutzeraktivitäten, Gruppe A erhielt Unterstützung für die ersten beiden Aufgaben [22]

Aufgabe	\bar{x} (Stdabw.)		Welch-Test		
	Gruppe A	Gruppe B	t	df	p
Gesundheit	8,25 (5,64)	3,00 (2,95)	2,86	16,61	0,011
Plagiate	14,92 (5,30)	8,25 (7,62)	2,49	19,63	0,022
Java	2,42 (1,78)	1,42 (1,62)	1,44	21,81	0,165
Gesamt	25,58 (9,61)	12,67 (8,95)	3,41	21,89	0,002
pro Anfrage	0,58 (0,24)	0,29 (0,26)	2,80	21,92	0,010
pro Aktion	0,14 (0,04)	0,09 (0,06)	2,30	19,55	0,032

Tabelle 3 Von Suchenden gespeicherte relevante Dokumente, Gruppe A erhielt Unterstützung [22]

chende für jede Aufgabe gespeicherten relevanten Dokumente zu bestimmen.

4.4 Ergebnisse

Ein Fragebogen im Anschluss an das Experiment sollte der qualitativen Erhebung der Nutzerzufriedenheit dienen. Die Teilnehmer gaben auf einer fünfstufigen Likertskala an, wie zufrieden sie mit dem Verlauf ihrer Suche und ihren Ergebnissen waren. Bei gleichbleibendem Sucherfolg wäre ein Suchsystem vorzuziehen, bei dem die Suchenden insgesamt mit der Benutzung zufriedener waren.

Von zwölf unterstützten Teilnehmern antworteten sieben, sie seien mit dem Verlauf der Suche zufrieden oder sehr zufrieden gewesen, und fünf waren weder zufrieden noch unzufrieden. Drei der zwölf nicht unter-

stützten Teilnehmer gaben an, sie seien zufrieden oder sehr zufrieden gewesen, vier weder noch und fünf antworteten, sie seien unzufrieden oder sehr unzufrieden mit dem Verlauf der Suche. Der Exakte Test nach Fisher ergab mit $p = 0,067$ eine positive Korrelation zwischen den beiden Variablen Unterstützung und Zufriedenheit. Da $p > 0,05$ ist, konnte die Null-Hypothese jedoch nicht verworfen werden, und es muss davon ausgegangen werden, dass keine Abhängigkeit vorliegt. Pearsons Chi-Quadrat-Test ergab eine Korrelation mit $p = 0,040 < 0,05$, jedoch ist dieser aufgrund der geringen Anzahl an Beobachtungen als unzuverlässig anzusehen.

Bei der Einschätzung der Ergebnisse zeigt sich ein deutlicheres Bild. Neun der unterstützten Teilnehmer waren zufrieden oder sehr zufrieden mit ihren Ergebnissen, drei waren weder zufrieden noch unzufrieden. Von

den Teilnehmern ohne Unterstützung erklärten sich nur zwei als zufrieden oder sehr zufrieden, fünf waren weder zufrieden noch unzufrieden und ebenso viele waren entweder unzufrieden oder sehr unzufrieden mit ihren Ergebnissen. Der Exakte Test nach Fisher ergab hier eine positive Korrelation zwischen den Variablen ($p = 0,004$), so dass davon ausgegangen werden kann, dass eine Abhängigkeit zwischen Unterstützung und Zufriedenheit mit den Ergebnissen existiert.

Dieser Zusammenhang wurde durch die quantitative Auswertung der Suchergebnisse bestätigt (aufgrund der Varianzheterogenität wurde zur Auswertung der Welch-Test herangezogen).

Zunächst wurde untersucht, ob die unterstützten Suchenden bei ihren Suchen erfolgreicher waren. Unterstützte Suchende stellten weniger Anfragen und betrachteten mehr Dokumentdetails, die Unterschiede waren jedoch nicht signifikant (siehe Tabelle 2). Sie speicherten signifikant mehr Dokumente ($p = 0,019$), es ist aber nicht klar, ob dies einen guten Indikator für tatsächlichen Sucherfolg darstellt.

Daher wurden Relevanzurteile von unabhängigen Gutachtern herangezogen. Unterstützte Teilnehmer speicherten signifikant mehr *relevante* Dokumente für alle Aufgaben zusammen, sowie für die Aufgaben „Gesundheit“ und „Plagiate“ (siehe Tabelle 3). Für die Suchaufgabe „Java“ konnte kein signifikanter Unterschied festgestellt werden. Dies lässt sich möglicherweise auf die geringe Zahl für die Aufgabe relevanter Dokumente in der benutzten Kollektion zurückführen.

Betrachtet man die gespeicherten Ergebnisse der Suchenden nicht isoliert, so muss man, wie von Järvelin [20] vorgeschlagen, auch die Kosten der Benutzer, die zur Erzielung des Sucherfolgs notwendig waren, in die Bewertung mit einbeziehen. Hierzu kann man das Verhältnis relevanter Ergebnisse (Nutzen) zur Anzahl gestellter Anfragen oder inspizierter Dokumente betrachten (Kosten). Dieses drückt aus, wieviel Aufwand der Suchende betreiben musste, um ein einzelnes relevantes Ergebnis zu erhalten.

Es fehlen Kostenfaktoren der einzelnen Aktivitäten. Man kann annehmen, dass der Aufwand für eine Suchanfrage deutlich höher ist als für andere Aktionen, insbesondere aufgrund des Zeitaufwands für das Warten auf die Ergebnisse und für das Scannen der Resultatliste. Da die tatsächlichen Kosten unbekannt sind, lassen sich hier nur schwer Schlüsse ziehen. Betrachtet man allein das Verhältnis relevanter Ergebnisse zu gestellten Anfragen, so zeigt sich, dass unterstützte Benutzer doppelt so viele relevante Dokumente pro Anfrage fanden ($p = 0,01$). Bezieht man alle Aktionen mit ein und geht von einheitlichen Kosten für alle Aktionen aus, so führten Benutzer mit Unterstützung im Schnitt zwar mehr

Aktionen aus, waren aber im Verhältnis zum Nutzen noch immer effizienter als die Benutzer ohne Unterstützung ($p = 0,032$, siehe Tabelle 3).

Suchvorschläge in der präsentierten Form scheinen also grundsätzlich geeignet, Benutzer erfolgreicher und mit weniger Kosten suchen zu lassen. Allerdings hilft die Unterstützung offenbar nicht bei allen Aufgaben.

Es wurde weiter untersucht, ob unterstützte Benutzer die fortgeschrittenen Werkzeuge und Suchoptionen von DAFFODIL signifikant häufiger benutzen als nicht unterstützte Benutzer. Die Evaluation der Suchhistorien zeigte, dass es keine statistisch signifikanten Unterschiede zwischen beiden Gruppen gab bezüglich Anfrageveränderungen durch Hinzufügen oder Wegnehmen von Suchbegriffen oder booleschen Operatoren, durch Korrektur der Schreibweise, Wiederverwendung einer früheren Anfrage, Ersetzung eines oder mehrerer Suchbegriffe oder Ersetzung der gesamten Anfrage. Allerdings waren die Unterschiede in den beiden letzten Fällen nur marginal nicht signifikant: Benutzer, die *keine* Vorschläge erhielten, ersetzten Anfragebegriffe durchschnittlich 16,92 mal pro Suche im Gegensatz zu 11,08 mal ($p = 0,055$) und ersetzten die gesamte Anfrage im Schnitt 1,31 mal pro Suche zu 0,83 mal ($p = 0,062$).

Andererseits verwendeten Benutzer, die Vorschläge erhielten, signifikant häufiger das Jahresfeld zur Einschränkung der Anfrage, wechselten zwischen Freitext- und Titelfeld oder benutzten das Autorenfeld ($p = 0,001$). Obwohl beide Gruppen die gleiche schriftliche Einführung in das Suchformular des Programms erhielten und alle Suchfelder während des gesamten Experiments durchgehend sichtbar waren, war zu beobachten, dass die Benutzer ohne Vorschläge sich grundsätzlich auf ein Suchfeld beschränkten, welches sie zu Beginn für ihre erste Anfrage gewählt hatten (entweder „Titel“ oder „Freitext“). Bei diesem Suchfeld blieben sie dann für alle Suchen während aller drei Aufgaben, auch wenn die Ergebnisse nicht zufriedenstellend waren.

Die Benutzung der fortgeschrittenen Werkzeuge während des Experiments zeigt ein ähnliches Bild. Keine signifikanten Unterschiede zeigten sich bei der Benutzung des Thesaurus zum Nachschlagen allgemeinerer oder speziellerer Begriffe, von Synonymen oder Wortdefinitionen. Unterstützte Teilnehmer benutzten aber signifikant häufiger die Extraktionsmöglichkeiten von DAFFODIL, um neue Begriffe oder häufige Autoren aus Suchresultaten oder gespeicherten, relevanten Dokumenten zu extrahieren ($p = 0,01$). Sie nutzten auch mehr Begriffe aus anderen Werkzeugen (wie der Darstellung verwandter Begriffe, den extrahierten Begriffen, dem Thesaurus oder der Klassifikation) direkt in ihrer Suche ($p = 0,039$).

Für die jeweils zuletzt bearbeitete Aufgaben wurde in der unterstützten Gruppe das Vorschlagswerkzeug abgeschaltet, so dass keine Vorschläge angezeigt wurden. Man kann bei dieser Gruppe einen klaren Rückgang der Benutzung fortgeschrittener Aktionen erkennen (siehe Tabelle 2). Trotzdem verwendeten zuvor unterstützte Benutzer in der dritten Aufgabe signifikant häufiger solche fortgeschrittenen Aktionen als nie unterstützte Benutzer ($p = 0,02$). Außerdem benutzten Teilnehmer der Gruppe, die zuvor Vorschläge erhalten hatten, häufiger zusätzliche Suchfelder ($p = 0,004$).

Das lässt sich als ein schwacher Lerneffekt interpretieren, der aus dem Erhalt situativ passender Suchvorschläge resultiert. Benutzer konnten zuvor vorgeschlagene Aktionen später unabhängig für eine andere Aufgabe einsetzen. Natürlich muss dabei berücksichtigt werden, dass die drei Aufgaben in direkter Folge ausgeführt wurden. Es bleibt ungewiss, ob die Benutzer das erworbene Wissen für spätere Aufgaben wieder abrufen könnten. Bei einer Langzeitstudie, bei der Suchende das Vorschlags-Werkzeug über einen langen Zeitraum hinweg einsetzen, zeigt sich aber möglicherweise auch ein ausgeprägter Effekt.

5 Zusammenfassung und Ausblick

Es wurde ein Vorschlagssystem vorgestellt, das mit Hilfe von CBR adaptive Suchunterstützung leistet und dabei durch Vorschlagsbewertung der Benutzer lernen kann. Das System wurde im Rahmen des DAFFODIL-Suchsystems für Digitale Bibliotheken umgesetzt, das Grundprinzip ist jedoch unabhängig vom konkreten Suchsystem und wurde bereits erfolgreich für die Google-Webuche angepasst [2]. Benutzerexperimente belegen den Nutzen des Vorschlagssystems, um nicht professionell Suchende beim erfolgreichen Einsatz fortgeschrittener Werkzeuge und Suchtaktiken zu unterstützen. Unterstützte Benutzer fanden mehr relevante Dokumente, setzten mehr Möglichkeiten des Suchsystems ein und griffen auch auf unterschiedliche Suchfelder zurück. Die neuen Taktiken wurden bei einer späteren, nicht unterstützten Suche auch ohne Vorschläge eingesetzt.

Derzeit wird an einer Weiterentwicklung des adaptiven Vorschlagssystems im Rahmen des EU-Projekts Khresmoi³ gearbeitet. Das in Khresmoi entwickelte Interface für die medizinische Informationssuche soll sowohl Domänenexperten als auch Laien bei ihrer Suche unterstützen und über die Webseite von *Health*

*on the Net*⁴ der Öffentlichkeit zur Verfügung gestellt werden. Dabei wird das Vorschlagssystem als eine von zahlreichen Unterstützungsmöglichkeiten zum Einsatz kommen. Im Langzeiteinsatz kann sich dann zeigen, ob die Unterstützung zu einer dauerhaften Verbesserung im interaktiven Suchverhalten von Endnutzern führen kann. Eine entsprechende Evaluation ist beabsichtigt.

Literatur

1. Aula A, Nordhausen K (2006) Modeling successful performance in web searching. *J Am Soc Inf Sci Technol* 57(12):1678–1693
2. Awasm M (2008) Vorschläge für Google-Suchen als Firefox-Erweiterung. Bachelorarbeit, Universität Duisburg-Essen
3. Bates MJ (1979) Information search tactics. *J Am Soc Inf Sci* 30(4):205–214
4. Bates MJ (1989) The design of browsing and berrypicking techniques for the online search interface. *Online Review* 13(5):407–424
5. Bates MJ (1990) Where should the person stop and the information search interface start? *Inf Process Manage* 26(5):575–591
6. Belkin NJ, Marchetti P, Cool C (1993) BRAQUE: Design of an interface to support user interaction in information retrieval. *Inf Process Manage* 29(3):325–344
7. Belkin NJ, Cool C, Stein A, Thiel U (1995) Cases, scripts, and information-seeking strategies: On the design of interactive information retrieval systems. *Expert Systems with Applications* 9(3):379–395
8. Bhavnani SK, Drabenstott K, Radev D (2001) Towards a unified framework of IR tasks and strategies. In: *Proceedings of the ASIST*, pp 340–354
9. Brajnik G, Mizzaro S, Tasso C (1996) Evaluating user interfaces to information retrieval systems: A case study on user support. In: *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, New York, pp 128–136
10. Brajnik G, Mizzaro S, Tasso C, Venuti F (2002) Strategic help in user interfaces for information retrieval. *J Am Soc Inf Sci Technol* 53(5):343–358
11. Carstens C, Rittberger M, Wissel V (2009) How users search in the german education index - tactics and strategies. In: *Proceedings of the workshop IR at the LWA 2009*
12. Drabenstott KM (2003) Do nondomain experts enlist the strategies of domain experts. *J Am Soc Inf Sci Technol* 54(9):8836–854
13. Fields B, Keith S, Blandford A (2004) Designing for expert information finding strategies. In: Fincher S, Markopoulos P, Moore D, Ruddle RA (eds) *BCS HCI*, Springer, pp 89–102
14. Fuhr N, Klas CP, Schaefer A, Mutschke P (2002) Dafodil: An integrated desktop for supporting high-level search activities in federated digital libraries. In: *Research and Advanced Technology for Digital Libraries. 6th European Conference, ECDL 2002*, Springer, Heidelberg et al., pp 597–612

³ gefördert durch das European Union Seventh Framework Programme (FP7/2007-2013), grant agreement 257528 (KHRESMOI), <http://khresmoi.eu/>

⁴ http://www.hon.ch/pat_de.html

15. Harter SP (1986) Online information retrieval: concepts, principles, and techniques. Academic Press Professional, Inc., San Diego, CA, USA
16. Harter SP, Peters AR (1985) Heuristics for online information retrieval: a typology and preliminary listing. *Online Review* 9(5):407–424
17. Hsieh-Yee I (1993) Effects of search experience and subject knowledge on online search behavior: Measuring the search tactics of novice and experienced searchers. *J Am Soc Inf Sci* 44(3):161–174
18. Jansen BJ (2005) Seeking and implementing automated assistance during the search process. *Inf Process Manage* 41(4):909–928
19. Jansen BJ, McNeese MD (2005) Evaluating the effectiveness of and patterns of interactions with automated searching assistance: Research articles. *J Am Soc Inf Sci Technol* 56(14):1480–1503
20. Järvelin K (2009) Explaining user performance in information retrieval: Challenges to IR evaluation. In: Azzopardi L, Kazai G, Robertson SE, Rüger SM, Shokouhi M, Song D, Yilmaz E (eds) *ICTIR 2009*, Springer, Lecture Notes in Computer Science, vol 5766, pp 289–296
21. Klas CP, Fuhr N, Schaefer A (2004) Evaluating strategic support for information access in the DAFFODIL system. In: *Research and Advanced Technology for Digital Libraries. Proceedings of the ECDL*, Springer, Heidelberg et al., pp 476–487
22. Kriewel S (2010) Unterstützung beim Finden und Durchführen von Suchstrategien in Digitalen Bibliotheken. PhD thesis, University of Duisburg-Essen
23. Kriewel S, Fuhr N (2007) Adaptive search suggestions for digital libraries. In: *Asian Digital Libraries: Looking Back 10 Years and Forging New Frontiers (ICADL 2007)*, pp 220–229
24. Kriewel S, Fuhr N (2010) An evaluation of an adaptive search suggestion system. In: *32nd European Conference on Information Retrieval Research (ECIR 2010)*, Springer, Heidelberg et al.
25. Kriewel S, Klas CP, Schaefer A, Fuhr N (2004) Daffodil - strategic support for user-oriented access to heterogeneous digital libraries. *D-Lib Magazine* 10(6)
26. Marchionini G (1989) Information-seeking strategies of novices using a full-text electronic encyclopedia. *J Am Soc Inf Sci* 40(1):54–66
27. Markey K (2007) Twenty-five years of end-user searching, part 1: Research findings. *J Am Soc Inf Sci Technol* 58(8):1071–1081
28. Ontañón S, Plaza E (2003) Justification-based multiagent learning. In: Tom Fawcett NM (ed) *The Twentieth International Conference on Machine Learning (ICML 2003)*, AAAI Press, AAAI Press, pp 576–583
29. Pollock A, Hockley A (1997) What's wrong with internet searching. *D-Lib Magazine*
30. Rieh SY, Xie HI (2001) Patterns and sequences of multiple query reformulations in web searching: a preliminary study. In: *Proceedings of the 64th Annual Meeting of the ASIST*, vol 38, pp 246–255
31. Schaefer A, Jordan M, Klas CP, Fuhr N (2005) Active support for query formulation in virtual digital libraries: A case study with DAFFODIL. In: *Research and Advanced Technology for Digital Libraries. Proceedings of the ECDL*, Springer, Heidelberg et al.
32. Wildemuth BM (2004) The effects of domain knowledge on search tactic formulation. *J Am Soc Inf Sci Technol* 55(3):246–258
33. Xie HI (2000) Shifts of interactive intentions and information-seeking strategies in interactive information retrieval. *J Am Soc Inf Sci* 51(9):841–857
34. Xie I, Cool C (2009) Understanding help seeking within the context of searching digital libraries. *J Am Soc Inf Sci Technol* 60(3):477–494