

Gebrauchstauglichkeit der Ergebnisseiten von Suchmaschinen

Lars Kaczmirek

Zentrum für Umfragen, Methoden und Analysen (ZUMA)

Zusammenfassung

Bei Ergebnisseiten von Suchmaschinen nutzen Anwender die Informationen in den Linkbeschreibungen, um zu entscheiden, ob ein Link für die Suche relevant ist und weiterverfolgt werden soll. Die Linkbeschreibung ist damit ein Kernelement im Selektionsprozess der Suche. Ein neues Modell für eine systematische und differenzierte Forschung über Ergebnisseiten von Suchmaschinen wird vorgestellt. Dazu werden die drei Faktoren Mensch, Maschine, Aufgabe und ihre Implikationen erläutert. Für eine Einbettung der bisherigen Forschung und ihrer Methoden werden verwandte Untersuchungen skizziert. Die aus dem Modell heraus entwickelte Untersuchungsmethode erlaubt die Erfassung verschiedener Kriterien und die Kontrolle störender Einflussfaktoren bei gleichzeitig geringer Untersuchungsdauer. In einem Experiment wurde die Art der Zusammenfassungen der verlinkten Internetseiten und die Spezifität der Suchaufgabe variiert. Die Daten von 191 Personen erlaubten einen Vergleich der Gebrauchstauglichkeit verschiedener Linkbeschreibungen mit der Möglichkeit einer statistischen Kontrolle der Unterschiede zwischen den Teilnehmern. Die Ergebnisse zeigen, dass eine Linkbeschreibung durch 'Anfangstext mit 400 Zeichen redaktionellem Inhalt' den anderen Bedingungen überlegen ist und die Entscheidung am besten unterstützt.

1 Einleitung

Ein aktueller Bericht des Pew Internet Project zeigt eindrucksvoll die große Bedeutung von Suchmaschinen für den Beruf und Alltag von Internetnutzern. Demnach haben acht von zehn Amerikanern bereits Suchmaschinen zum Finden von Informationen verwendet, und mehr als jeder Vierte, d.h. 33 Millionen Menschen, verwenden täglich Suchmaschinen (Fox 2002). Daraus folgt eine sehr hohe Zeitspanne, die aus der Beurteilung von Treffern auf Ergebnisseiten entsteht. Die verlorene Zeit, die bei Fehlentscheidungen mit dem Betrachten irrelevanter Seiten vergeudet wird oder die Verluste durch abgewanderte Kunden, die die gesuchten Informationen auf den Firmenseiten nicht finden konnten, erhöhen diesen Betrag deutlich. Gute Ergebnisseiten helfen Zeit und Kosten sparen, erhöhen die Kundenbindung, reduzieren Fehlentscheidungen und verringern Frustrationen. Der vorliegende Beitrag¹ verfolgt daher zwei Ziele. Im theoretischen Teil wird ein Rahmenmodell für die Erforschung zur Gebrauchstauglichkeit (Usability) der Ergebnisseiten von Suchmaschinen herausgearbeitet, um die Bestimmungsstücke zu identifizieren, die eine erfolgreiche Forschung konstituieren. Der empirische zweite Teil beschreibt wie die daraus entwickelte Methodik in einer experimentellen Untersuchung angewandt wurde und stellt ausgewählte Ergebnisse zur optimalen Ergebnisseitengestaltung dar. Die Forschung auf diesem Gebiet entwickelte sich erst in den letzten Jahren, so dass noch bedeutende Veränderungen erwartet werden können, wobei die abschließende Diskussion zeigen wird, wohin die Entwicklung gehen könnte.

¹ Der Beitrag basiert auf den Kernthesen und -ergebnissen der Diplomarbeit von Lars Kaczmirek (2003).

2 Das Rahmenmodell Mensch, Maschine, Aufgabe

Eine Analyse der Literatur zur Gebrauchstauglichkeit (Usability) von Suchmaschinen und den Standardwerken zur Mensch-Maschine-Interaktion lässt drei zentrale Ebenen erkennen, die als die größten Einflussfaktoren in der Forschung zu Suchmaschinen gelten können. Diese werden im Folgenden mit Mensch (*user*), Maschine (*system/machine*) und Aufgabe (*task*) bezeichnet (Abbildung 1).

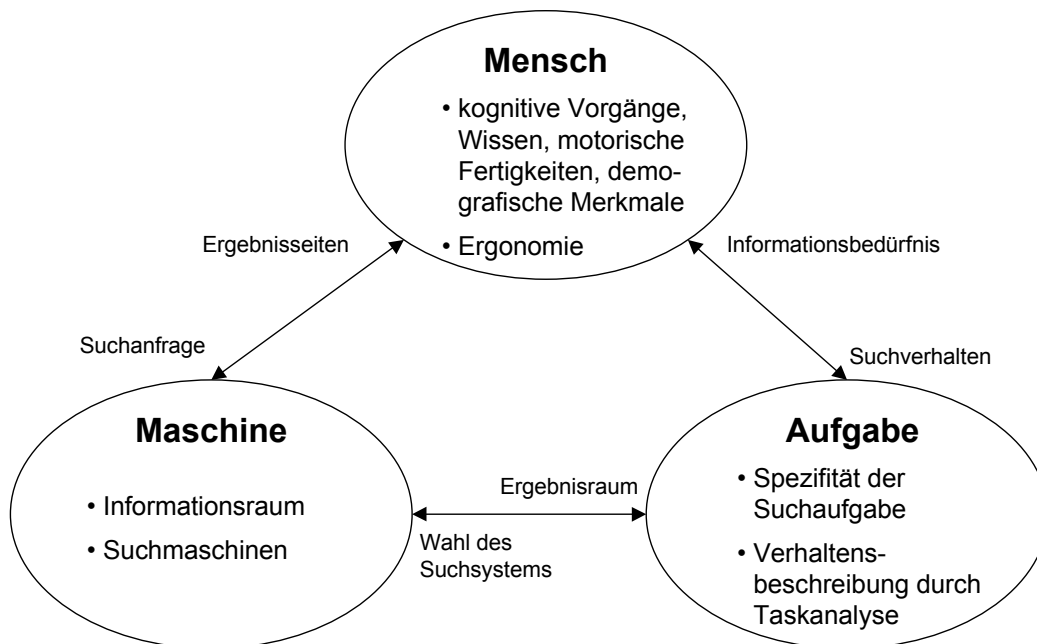


Abbildung 1: Rahmenmodell zur Erforschung der Gebrauchstauglichkeit der Ergebnisseiten von Suchmaschinen.

2.1 Ebene Mensch

Die Ebene Mensch beinhaltet eine Struktur von Objekten, die Beschreibungen der Anwender selbst ermöglicht. Eine Dimension enthält dabei diejenigen Variablen, durch die sich Unterschiede zwischen Anwendern beschreiben lassen. Dies sind insbesondere kognitive Strategien und Prozesse beim Umgang mit Ergebnisseiten. Anwender unterscheiden sich jedoch auch hinsichtlich ihres Wissens über Suchmaschinen, dem Internet und dem Suchgegenstand. Des Weiteren finden sich Unterschiede in den motorischen Fertigkeiten, zum Beispiel die Sicherheit, mit der die Maus bedient wird, sowie in allgemeinen demografischen Merkmalen wie Alter, Geschlecht usw. Die zweite Dimension enthält allgemeinspsychologische Gesetzmäßigkeiten, wie sie für alle Menschen gelten. Diese münden in Ergonomierichtlinien, aus denen wiederum allgemeine Strukturen für Ergebnisseiten abgeleitet werden können. Entscheidungspsychologische Aspekte finden sich bei Wirth und Schweiger (1999), sowie Jungermann, Pfister und Fischer (1998).

Der wichtigste Mediator beim Such- und Entscheidungsverhalten wird durch die großen Unterschiede in der Erfahrung der Anwender gebildet (Shneiderman 1992). Weber und Groner (1999)

unterscheiden drei Formen von Wissen, die für einen effizienten Suchprozess im Internet von Bedeutung sind: Wissen über das Web umfasst die Kenntnis von Strukturen und Inhalten und ermöglicht dem Anwender damit eine Modellbildung seiner Funktionsweise. Zum Wissen über Suchinstrumente gehört die Kenntnis, zu welchem Zweck welche Suchmaschine eingesetzt werden sollte und wie diese zu bedienen ist. Die dritte Form umfasst Wissen über den Suchgegenstand und wird auch als Domänen- oder Bereichswissen bezeichnet (Marchionini 1995).

2.2 Ebene Maschine

Die zweite Ebene, Maschine, umfasst allgemein betrachtet das gesamte System, auf dem und mit dem die Suche durchgeführt wird. Erstens ist dies der Informationsraum, in dem die Suche erfolgt. Dies entspricht der Struktur des Internets mit all seinen Webdokumenten und Angeboten. Zweitens enthält sie eine Beschreibung des Werkzeugs, mit dem gesucht wird, und befasst sich daher mit der Funktion und dem Aufbau von Suchmaschinen. Dies schließt auch die formalen Eigenschaften zur Definition von Ergebnisseiten sowie die Dialoge, mit deren Hilfe Anwender mit Suchmaschinen interagieren, mit ein.

Nach Kleinfeldt und Baphna (2000) lassen sich Treffer auf Ergebnisseiten als Ordnungspunkte mit mehreren Eigenschaften beschreiben:

- Die Position in der Liste. Je weiter oben ein Dokument platziert ist, desto höher war die vom Suchmaschinenalgorithmus geschätzte Relevanz des Links.²
- Der Titel des Dokuments, der auch einen Verweis (Link) auf das Zieldokument enthält.
- Eine Beschreibung des Dokuments. Üblicherweise eine Zusammenfassung, in der der Anfang des Dokuments oder die gesuchten Wörter im Kontext der umliegenden Wörter dargestellt werden.
- Zusätzliche Informationen über das Dokument, wie URL, Dokumenttyp, Dokumentgröße, Autor, Datum der letzten Änderung usw.

Der vorliegende Beitrag legt seinen Schwerpunkt darauf, inwieweit die Zusammenfassung der Dokumente den Anwender bei dieser Entscheidung unterstützen können. Der häufigste Ansatz zur Beschreibung von Dokumenten besteht in der Darstellung von Text aus dem Anfang der repräsentierten Dokumente (Methode: Textanfang). Eine andere Methode stellt die gesuchten Wörter im Textkontext dar (Methode: Wörter-im-Kontext). Weitere Ansätze versuchen mittels Techniken der Textsummarisation die wichtigsten Sätze oder zentrale Wörter zu erfassen und diese darzustellen (vgl. Drori 2001), um dem Anwender einen Vergleich seines Informationsbedürfnisses mit den Kernaussagen des Dokuments zu ermöglichen (Methode: Kernaussagen oder Schlüsselwörter).

2.3 Ebene Aufgabe

Die Ebene Aufgabe schließlich beinhaltet alle mit der Suchaufgabe im direkten Zusammenhang stehenden Aspekte von Ergebnisseiten. Dies ist zunächst die Klassifizierung von Suchaufgaben selbst, die allgemeiner auch als Informationsbedürfnisse aufgefasst werden können. Hierbei spielt insbesondere eine Rolle, wie breit bzw. wie spezifisch eine Suchaufgabe ausfällt. Shneiderman (1997) klassifiziert die Suchaufgaben auf dem Kontinuum von „spezifischem Faktenfinden“ bis

² Für einen Ansatz, der Ergebnisse in Kategorien darstellt und damit gute Erfolge erzielt, vgl. Dumais, Cutrell und Chen (2001).

hin zu „unstrukturiertem freiem Browsen bekannter Datenbanken“ und „Überprüfung der Verfügbarkeit von Informationen“. Einfache Fakten sind klar strukturierte und abgrenzbare Informationseinheiten wie zum Beispiel ein Geburtsdatum. Dem Anwender ist klar, wonach er sucht, und sobald er die Information gefunden hat, ist die Suche beendet, wobei er großes Vertrauen in die Richtigkeit und Vollständigkeit der gefundenen Antwort hat. Am anderen Ende des Kontinuums, bei Meinungen und Interpretationen, ist die Bestimmung einer ausreichenden Antwort komplizierter und damit auch ein Ende der Suche nicht klar bestimmbar. Im Folgenden fasst der Begriff der Spezifität diese Dimension zusammen. Der zweite Aspekt betrifft die Aufgabenanalyse (task analysis), womit das gezeigte Suchverhalten in elementare Bestandteile zerlegt und so eine genauere Betrachtung der Anforderungen an Ergebnisseiten ermöglicht wird.

3 Forschungsparadigma

Anhand dreier Untersuchungen wird das aktuelle Forschungsparadigma zur Ergebnisseitengestaltung mit dem Schwerpunkt der Methoden zur Zusammenfassung der gelinkten Dokumente dargestellt.

Unz, Capstick und Erbach (1999) stellten ihren Teilnehmern in Saarbrücken verschiedene Suchaufgaben und ließen mit einer Suchmaschine passende Ergebnisse finden und notieren. Insgesamt benötigte ein Teilnehmer drei Stunden. Sie variierten die Art der Organisation der Suchergebnisse, die Art der Zusammenfassung und erhoben als abhängige Variablen Performanz, Akzeptanz und Variablen des Nutzungsverhaltens. Bei den Zusammenfassungen waren dabei drei Variationen vorhanden: Die „ersten 200 Zeichen“, „die Kernaussagen“ und „Wörter im Kontext“. 84 Teilnehmer wurden auf sechs Gruppen verteilt, wodurch sich in einer Post-hoc-Analyse³ eine Teststärke von 0.36 errechnen lässt (ANOVA, Effektstärke Medium, Alphafehler 5%). Dies und die fehlende Kontrolle der Spezifität der Suchaufgaben könnte insgesamt zu dem uneinheitlichen Ergebnisbild der Untersuchung beigetragen haben. Performanz- oder Akzeptanzunterschiede konnten zwischen den Arten der Zusammenfassungen nicht festgestellt werden.

Drori (2001) ließ in Jerusalem mit einer Suchmaschine vier Suchaufgaben unterschiedlicher Spezifität bearbeiten. Variiert wurde die Art der Zusammenfassung in drei Bedingungen: „Nur Dokumenttitel ohne weitere Beschreibung“, „einige erste Zeilen“ und die „Zeilen mit den für die Suche relevanten Inhalten“. Erhoben wurden Maße der Performanz und Zufriedenheit. 128 Teilnehmer bearbeiteten die vier Aufgaben und wurden dabei allen drei Bedingungen in zufälliger Reihenfolge zugewiesen. Mittels Post-hoc-Analyse ergibt sich eine Teststärke von 0.94 (abhängiger T-Test, Effektstärke Medium, Alphafehler 5%, unspezifische Hypothese). Für das Performanzmaß „aufgewendete Zeit“ ergab sich ein Interaktionseffekt mit der Aufgabenspezifität. Bei unspezifischer Aufgabe war die Wörter-in-Kontext-Bedingung am praktikabelsten, wohingegen bei sehr spezifischer Aufgabe der Titel allein für das schnellste Ergebnis sorgte. Insgesamt war die Wörter-in-Kontext-Bedingung den anderen Bedingungen überlegen. Problematisch bei dieser Untersuchung ist jedoch das Messwiederholungsdesign einzuschätzen, dass jedem Teilnehmer alle Bedingungen gegenüberstellte und damit möglicherweise Fehlervarianz erzeugt, die durch die Methode der erzwungenen Differenzierung hervorgerufen wird.

Die beiden vorherigen Studien verlangten von den Teilnehmern das Durchlaufen von mehreren kompletten Suchprozessen. Solche Untersuchungen sind sehr aufwendig durchzuführen, da nicht nur die Teilnehmer über einen langen Zeitraum gebunden sind, sondern häufig auch einen be-

³ Berechnung der Teststärken im vorliegenden Beitrag wurden mit GPOWER von Faul und Erdfelder (1992) durchgeführt.

stimmtm Untersuchungsraum aufsuchen müssen. Darüber hinaus sind umfangreiche technische Vorbereitungen zu treffen, die mit dem Untersuchungsgegenstand selber wenig gemeinsam haben: Eine Datenbank muss aufgebaut, ihre Dokumente für die Untersuchung im günstigsten Falle in nur zwei Kategorien „relevant“ und „nicht relevant“ klassifiziert und eine Suchmaschine bereit gestellt werden. Mag dies für eine Modellierung des gesamten Suchprozesses noch notwendig erscheinen, so lässt sich das relevante Verhalten auf Ergebnisseiten deutlich einfacher untersuchen, wie Lergier und Resnick (2001) zeigten. Teilnehmer gaben ihre Suchworte ein, erhielten dann jedoch vordefinierte Ergebnisseiten unabhängig von ihrer Eingabe. Sobald sie einen Link ausgewählt hatten, war die Suchaufgabe beendet und die Abschlussfragen folgten. Die Dokumente hinter den Links wurden nicht dargeboten. Ihre Ergebnisse zeigen, dass die Schwellenwertmethode im Gegensatz zur erschöpfenden Methode der Entscheidung häufiger bei spezifischen als unspezifischen Aufgaben angewendet wird.

Die dargestellten Untersuchungen zeigen, welche theoretische und praktische Unterstützung das neue Rahmenmodell Mensch, Maschine, Aufgabe für eine erfolgreiche, systematische Forschung über Ergebnisseiten haben kann. Im Folgenden wird eine Untersuchung dargestellt, die sich an der Systematik des Rahmenmodells orientiert und neue Untersuchungsmethoden implementiert.

4 Fragestellung und Methode

Im vorliegenden Beitrag werden lediglich die Ergebnisse zur Kernfrage der Untersuchung präsentiert: Wie ist das optimale Beschreibungsattribut gestaltet, d.h. welche Art der Zusammenfassung der gelinkten Dokumente unterstützt den Anwender am optimalsten bei seiner Selektionsentscheidung?⁴ Die Faktoren des Rahmenmodells wurden im Design einer einfaktoriellen Varianzanalyse mit einem Messwiederholungsfaktor umgesetzt. Jeder Teilnehmer wurde randomisiert einem von vier Stufen des Faktors „Art der Zusammenfassung“ zugewiesen und durchlief beide Stufen des Faktors „Aufgabenspezifität“, wobei die Sequenz permutiert wurde, um Reihenfolgeeffekte zu kontrollieren. Im Faktor „Art der Zusammenfassung“ wurden vier Methoden operationalisiert.

1. Wörter im Kontext: Die Suchwörter mit den im Text vorkommenden umliegenden Wörtern. Die Darstellung erfolgte in der Art von Google.
2. Textanfang: Die ersten 400 Zeichen redaktioneller Inhalt.
3. Schlüsselwörter: Die sieben häufigsten semantisch sinnvollen Substantive im Dokument (HTML-Code wurde herausgefiltert).
4. Maximale Informationsmenge aller drei Methoden zusammen (MaxInfo). Die Angaben jeder Methode erfolgt untereinander in der Reihenfolge Schlüsselwörter, Wörter im Kontext, Textausschnitt des Anfangs.

Die Aufgabenspezifität wurde durch die beiden Suchaufgaben „Welche Größe haben Euroscheine?“ (spezifisch) und „Welche Folgen hat die Währungsumstellung?“ (unspezifisch) operationalisiert. In einer Untersuchungsbedingung (bspw. Spezifisch–Textanfang) wurden die Teilnehmer aufgefordert sich vorzustellen, auf die entsprechende Frage eine Antwort zu suchen. Ihnen wurde die Art der Zusammenfassung und die verwendeten, vorher festgelegten Suchwörter mitgeteilt. Auf der präsentierten Ergebnisseite entschieden sie für jede der sieben Optionen, ob sie den Link auswählen würden oder nicht. Für dieselbe Ergebnisliste wurde anschließend abgefragt, für wie wahrscheinlich sie es hielten, dass die einzelnen Links zu den gesuchten Informationen führen

⁴ Für eine ausführliche Darstellung des Untersuchungsdesigns und weiterer Ergebnisse sei auf Kaczmarek (2003) verwiesen.

würden.⁵ Auf zwei weiteren Seiten wurden Zufriedenheits- und andere Einstellungsmaße zur Ergebnisseite erhoben. Am Ende beider Bedingungen zur Spezifität erfolgten Gegenüberstellungen aller vier Arten der Zusammenfassung, wobei die Teilnehmer ihre Präferenzen angaben. Mit offenen Fragen und einem Debriefing inklusive optional auswählbarer Antwortdokumente zu den Suchanfragen endete die Untersuchung.

Um in der Auswertung Effekte aufgrund von Unterschieden zwischen den Teilnehmern (Faktor Mensch) statistisch konstant halten zu können, wurden verschiedene Erfahrungs-, Wissensmerkmale und demografische Daten erhoben. Darüber hinaus gab es eine Übungsphase, in der die Teilnehmer die in dieser Untersuchung spezielle Auswahltechnik von Links erprobten. Zusätzlich zu den vom Teilnehmer angegebenen Entscheidungen und Skalenwerten wurde die Zeit, die zwischen jedem Klick auf einen radio-button verstrich, erfasst. Dies diente als Maß für die Entscheidungsdauer auf den Ergebnisseiten und soll die Effektivität einer Ergebnisseite quantitativ und objektiv erfassbar machen. Einen Überblick wie die verschiedenen Konstrukte operationalisiert wurden, zeigt Tabelle 1.

Tabelle 1: Variablen im Untersuchungsdesign.

Theoretische Konstrukte und ihre Teilaspekte	Operationalisierungen, gemessene Variablen
<i>Mensch</i>	
Demografische Angaben	Alter, Geschlecht, Rekrutierungsweg
Erfahrung und Wissen	Acht Teilfragen zu: Welche Suchmaschinen wurden in der letzten Woche wie lange, bei wie vielen Suchanfragen verwendet? Wie viele Tage wurde in der letzten Woche am PC und im Internet gearbeitet? Drei Fragen zur Selbsteinschätzung: Wissen über Suchmaschinen, Währungsumstellung und Wirtschaft.
Manuelle Geschwindigkeit mit der Maus	Zeitmessung in den Übungsdurchgängen zur Linkauswahl.
<i>Maschine</i>	
Attribut Beschreibung	Randomisierte Zuweisung zu den Faktorstufen des Faktors „Art der Zusammenfassung“.
Andere Attribute	Lediglich der Titel wurde angezeigt
<i>Aufgabe</i>	
Aufgabenklassifizierung	Messwiederholungsfaktor „Spezifität der Aufgabe“.
Verhalten	Es sind Schlussfolgerungen anhand der Selektionsreihenfolge und dem Effektivitätsmaß Vollständigkeit möglich.
<i>Abhängige Variablen</i>	
Effektivität	Zeitdauer der Entscheidungen
	Genauigkeit (Anzahl relevanter und selektierter Links dividiert durch Anzahl selektierter Links) und Vollständigkeit (Anzahl relevanter und selektierter Links dividiert durch Anzahl rele-

⁵ Dieser Ansatz weist Ähnlichkeiten mit dem Konzept des Pre-Click-Confidence von Lergier und Resnick (2001) auf. Unter Pre-Click-Confidence verstehen die Autoren die Sicherheit, mit der Anwender annehmen, dass ihre Auswahlentscheidung zu einer richtigen Antwort führen würde.

	vanter Links) ausgewählter Links
	Subjektive Sicherheit in der Linkauswahl
Effizienz	Erfolgsrate = Genauigkeit x Vollständigkeit / benötigte Zeit
Zufriedenheit	Direkter Vergleich der vier Arten der Zusammenfassung durch die Frage „Welche Darstellung würden Sie bei einer Suchmaschine bevorzugen. Beantworten Sie bitte: Wie gut oder schlecht ist die jeweilige Darstellung des Links?“ Vierzehn Items zur Einstellung nach jeder Ergebnisseite mittels einer fünfstufigen Likertskala von „lehne stark ab“ bis „stimme stark zu“.

5 Beschreibung der Stichprobe

Von April bis Juni 2002 konnte online an der etwa zwanzigminütigen Untersuchung teilgenommen werden. In die Auswertung gingen 191 Fälle ein. Die Teilnehmer waren 127 Männer und 61 Frauen (drei ohne Angabe). Das Durchschnittsalter betrug 30,9 Jahre (SD 10,2; Min 17; Max 78; Mn 27).⁶ Die meisten Teilnehmer wurden durch einen Link von anderen Seiten (72 Personen), durch E-Mail (55) oder eine Suchmaschine (22) auf die Untersuchung aufmerksam und unterlagen einer Selbstselektion (zu den Verzerrungen bei elektronischen Befragungen vgl. Vogt 1999). Es handelt sich um Anwender, die fast täglich mit Computern arbeiten. Durchschnittlich sind sie 6,2 Tage (SD 1,1; Mn 7) in der Woche am Computer und davon 5,7 Tage (SD 1,5; Mn 6) im Internet. Ihr Wissen über Suchmaschinen wird von ihnen auf einer fünf-stufigen Likertskala leicht überdurchschnittlich (Mn 4; M 3,6; SD 1,0), ihr Wissen über die Währungsumstellung zum Euro und ihr Wirtschaftswissen durchschnittlich eingeschätzt (Mn 3; 3; M 3,3; 3,0 SD 1,0; 1,0). In den vier Gruppen zur Art der Zusammenfassung fanden sich 58, 51, 38 und 44 Teilnehmer.

6 Ergebnisse

Im Sinne der ISO 9241-11 wird in diesem Beitrag als Effektivitätsmaß die Erfolgsrate in Anlehnung an Wolfram und Dimitroff (1997) berichtet, sowie ein Zufriedenheitsmaß. Zum Einsatz kamen einfaktorielle Varianzanalysen mit Messwiederholung, sowie im Falle der Zufriedenheitsskala der multivariate Ansatz. Als Post-hoc-Test für Mehrfachvergleiche wurde die *ehrllich signifikante Differenz nach Tukey* verwendet.

⁶ SD = Standardabweichung, M = Mittelwert, Mn = Median, Min = kleinster Wert, Max = größter Wert.

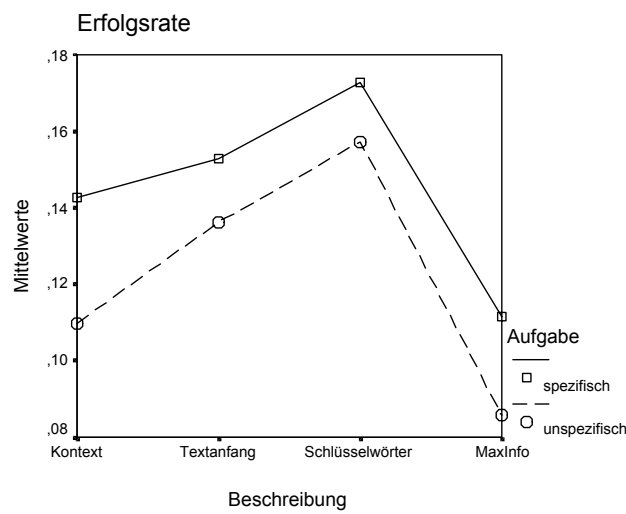


Abbildung 2: Erfolgsrate hinsichtlich Beschreibung und Spezifität.

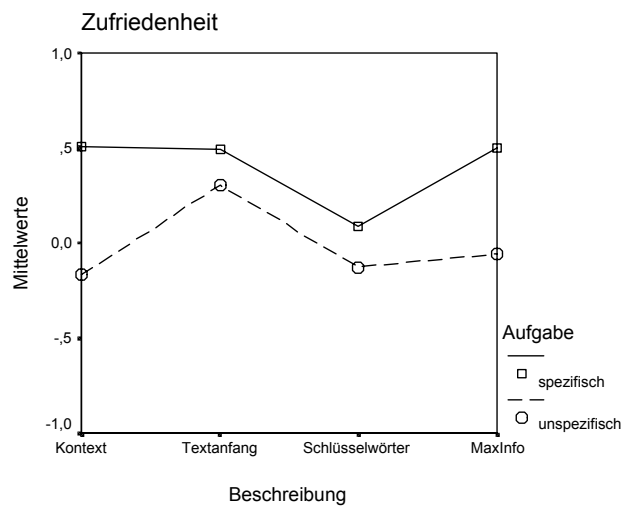


Abbildung 3: Zufriedenheit hinsichtlich Beschreibung und Spezifität.

Bei der Erfolgsrate sind die Hauptfaktoren Spezifität und Beschreibung höchst signifikant ($F=12,5$; $p=0,001$; $F=7,1$; $p=0,000$). Post-hoc-Mehrfachvergleiche zeigen, dass sich die Bedingung Schlüsselwörter signifikant von den beiden Bedingungen Kontext und MaxInfo unterscheidet ($p=0,31$; $p=0,000$). Weiterhin ist der Unterschied zwischen den Bedingungen Textanfang und MaxInfo signifikant ($p=0,005$). Aus Abbildung 2 lässt sich erkennen, dass eine Beschreibung der Treffer mit Schlüsselwörtern den Beschreibungsvarianten Kontext und MaxInfo überlegen ist. Außerdem ist eine Beschreibung durch Textanfang der Konglomeratvariante MaxInfo überlegen. Spezifische Suchaufgaben ergeben eine höhere Erfolgsrate als unspezifische Suchaufgaben.

Für die Zufriedenheit (Abbildung 3) erweist sich der Faktor Spezifität als höchst signifikant ($F=61$; $p=0,000$), der Faktor Beschreibung als sehr signifikant ($F=4,0$; $p=0,009$). Auch der Interak-

tionseffekt zwischen Beschreibung und Spezifität ist höchst signifikant ($F=5,9$; $p=0,001$). Post-hoc-Mehrfachvergleiche zeigen, dass sich die Bedingung Schlüsselwörter von den Bedingungen Kontext ($p=0,003$) und Textanfang ($p=0,003$) signifikant unterscheidet.

7 Interpretation

Vergleiche der Erfolgsrate zeigen, dass Trefferbeschreibungen mit Schlüsselwörtern den beiden Beschreibungsvarianten Wörter-im-Kontext und der Konglomeratbedingung MaxInfo überlegen sind. Weiterhin ist Textanfang ebenfalls der Variante MaxInfo überlegen. Zwischen den Varianten mit Schlüsselwörtern und Textanfang zeigen sich keine signifikanten Unterschiede. Für das Effizienzmaß Erfolgsrate können demnach die Beschreibungsvarianten Textanfang und Schlüsselwörter als Favoriten angesehen werden. Bei einer Bewertung der Ergebnisseite im Hinblick auf die Anforderung Zufriedenheit kann ein anderes Bild beobachtet werden. Beschreibungen durch Schlüsselwörter erhalten bei den Items zur Zufriedenheit die schlechteste Bewertung durch Anwender und unterliegen signifikant der Variante mit Textanfang. Im Sinne eines Interaktionseffektes ist eine Beschreibung mit Wörtern im Kontext lediglich bei spezifischer Suchaufgabe der Schlüsselwortmethode überlegen. Der Favorit beim Kriterium Zufriedenheit ist demnach die Variante der Textanfang.

Bei einer genauen Betrachtung der Anforderungen zur Effizienz und Zufriedenheit fällt auf, dass Textanfang als Beschreibungsvariante in beiden Anforderungen gute Ergebnisse erzielt, wohingegen die Wörter im Kontext Methode nur sehr eingeschränkt empfohlen werden kann. Die Variante Schlüsselwörter wiederum ergibt zwar ein gutes Effizienzmaß, wird aber durch die Anwender selbst als eher schlecht bewertet. Die Zufriedenheit ist allerdings eine wesentliche Anforderung der ISO 9241-11. Bei gleichwertiger Effizienz mehrerer Beschreibungsvarianten ist es angebracht, der Meinung der Anwender im Sinne freier Entscheidungswahl den Ausschlag zu geben. Da sich die Effizienz bei Schlüsselwörtern von Textanfang nicht signifikant unterscheidet, kann die aufgeworfene Fragestellung zu Beschreibungsvarianten daher wie folgt beantwortet werden: Erstens existieren beobachtbare Unterschiede zwischen den verschiedenen Beschreibungsvarianten. Zweitens wird der Entscheidungsprozess bei der Bewertung von Treffern auf Ergebnisseiten am optimalsten durch redaktionellen Anfangstext aus den repräsentierten Dokumenten unterstützt.

Sehr viel eindeutiger ist die Befundlage zur Spezifität der Suchaufgabe. Bei allen Kriterien schneiden spezifische Suchaufgaben besser ab als unspezifische.

8 Diskussion

Es zeigten sich keine unerwünschten Effekte durch Dropouts, Missings oder Reihenfolge der Suchaufgaben. Es konnte gezeigt werden, dass das Rahmenmodell Mensch, Maschine, Aufgabe eine systematische und differenzierte Forschung über Ergebnisseiten von Suchmaschinen unterstützt. Die entwickelte Erhebungsmethode erlaubt die Kontrolle störender Varianzquellen bei geringer Untersuchungsdauer. Die erhobenen Variablen und das spezielle Design sichern die Vergleichbarkeit mit anderen Untersuchungen.

Redaktioneller Anfangstext von verlinkten Dokumenten ist anderen komplizierteren Methoden der Darstellung überlegen. Obwohl ausführlichere Informationen über das verlinkte Dokument zu einer Verbesserung der Entscheidungen führen, ergibt sich ein Trade-Off aufgrund der längeren Bearbeitungszeit. Designer von Ergebnisseiten sehen sich daher in einem Konflikt. Sollen sie die

Attribute der Links kurz und gering halten, um eine schnelle und oberflächliche Evaluation zu ermöglichen, oder sind ausführlichere Informationen sinnvoller, die die Ergebnisseite deutlich länger werden lassen, aber auch eine genauere Darstellung der erwarteten Inhalte erlauben? Für eine Generalisierung der Ergebnisse in dem Sinne, dass die verwendete Methode mit 400 Zeichen Anfangstext für alle Suchaufgaben optimal ist, bedarf es weiterer Untersuchungen mit vielen verschiedenen Suchaufgaben. Es wurde die hohe Bedeutung des Faktors Aufgabenspezifität deutlich. Das in dieser Untersuchung gefundene Ergebnis widerspricht anderen Untersuchungen (vgl. Drori 2001). Jedoch wurden mit 400 Zeichen vom Textanfang die sonst üblichen 50-200 Zeichen deutlich überschritten. Dadurch scheint der Anwender in der Lage zu sein, sich ein genaueres Bild vom verlinkten Dokument zu machen, als dies bei den Methoden „Wörter im Kontext“, „Kernaussagen oder Schlüsselwörter“ oder „Anfangstext mit 200 Zeichen“ möglich ist. Unklar bleibt, ab welcher Zeichenmenge sich die Vorteile in einen Nachteil umkehren. Im Sinne des Grundsatzes der Steuerbarkeit von Dialogen erscheint es wünschenswert, dem Anwender die Entscheidung zu überlassen, wie viele Zeichen aus dem Anfangstext dargestellt werden sollen, wobei eine sinnvolle Menge deutlich über den bisher üblichen 50-200 Zeichen liegt.

Weitere Forschung erscheint bei den Algorithmen zur Zeichenextraktion aus den verlinkten Dokumenten notwendig. Häufig kann der Anwender bei der Wörter-im-Kontext-Methode bei unspezifischen Suchaufgaben den Bezug zur Fragestellung nur schwer feststellen. Bei der Zeichenextraktion von redaktionellem Inhalt aus dem Textanfang sind spezielle Algorithmen notwendig, um keinen Programmcode oder Tabellenüberschriften, die ohne ihren grafischen Zusammenhang nicht sinnvoll interpretierbar sind, zu erhalten.

Der vorliegende Beitrag stellte Komponenten verschiedener Disziplinen in einem neu entwickelten Rahmenmodell vor. Das aktuelle Forschungsparadigma zu Ergebnisseiten von Suchmaschinen wurde skizziert und neue methodische Ansätze in der hier vorgestellten Untersuchung angewandt. Die Ergebnisse zeichnen ein neues Bild optimaler Ergebnisseiten, entgegen den bisherigen Bemühungen der Forschungsgemeinde die optimale Zusammenfassung eines Dokuments mit komplizierten Algorithmen zu suchen, die eine möglichst kurze und präzise Repräsentation des Dokuments erzeugen sollen. Bisher kann der Anwender selbst am sichersten entscheiden, ob ein Dokument für seine Suche relevant ist. Dafür verwendet er bevorzugt zusammenhängende Auszüge wie den Textanfang aus dem (für ihn geschriebenen) redaktionellen Inhalt der Dokumente.

Literatur

- Drori, O. (2001). Improving Display of Search Results in Information Retrieval Systems – User's Study. In: P. T. Isaías (Ed.), *New Developments in Digital Libraries, Proceedings of the 1st International Workshop on New Developments in Digital Libraries. NDDL 2001. In conjunction with ICEIS* (pp. 20-33). Setúbal, Portugal: ICEIS Press.
- Faul, F. & Erdfelder, E. (1992). *GPOWER: A priori, post-hoc, and compromise power analyses for MS-DOS* [computer programm]. Bonn University, Dep. of Psychology.
- Fox, S. (2002). *Pew Internet Project Data Memo*. [Webdokument, 22.09.2002]: Pew Internet & American Life. URL: <http://www.pewinternet.org/reports/index.asp>
- Jungermann, H., Pfister, H.-R., & Fischer, K. (1998). *Die Psychologie der Entscheidung*. Heidelberg: Spektrum Akademischer Verlag.
- Kaczmirek, L. (2003). *Information und Selektion. Gebrauchstauglichkeit der Ergebnisseiten von Suchmaschinen*. Unveröffentlichte Diplomarbeit, Universität Mannheim.

- Kleinfeldt, S., & Baphna, J. (2000). *A Commercial Perspective on Hypertext Search Results*. Paper presented at the Information Doors. Where Information Search and Hypertext Link, San Antonio, Texas.
- Lergier, R., & Resnick, M. (2001). *A framework for evaluating user strategies in Internet search and evaluation*. Paper presented at the 7th Conference on Human Factors and the Web.
- Marchionini, G. (1995). *Information Seeking in Electronic Environments*. Cambridge: Cambridge University Press.
- Shneiderman, B. (1992). *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Santiago: Addison Wesley Longman.
- Shneiderman, B. (1997). Designing information-abundant web sites: issues and recommendations. In: *International Journal of Human-Computer Studies*, 47, 5-29.
- Unz, D. C., Capstick, J., & Erbach, G. (1999). Damit die Suche (schneller) zum Ziel führt: Interface-Design einer Suchmaschine zur Unterstützung von Selektionsentscheidungen. In: W. Wirth & W. Schweiger (Eds.), *Selektion im Internet. Empirische Analysen zu einem Schlüsselkonzept* (pp. 249-270). Opladen: Westdeutscher Verlag.
- Vogt, K. (1999). Verzerrungen in elektronischen Befragungen? In: B. Batinic, A. Werner, L. Gräf & W. Bandilla (Eds.), *Online Research. Methoden, Anwendungen und Ergebnisse*. Göttingen: Hogrefe.
- Weber, C., & Groner, R. (1999). Suchstrategien im WWW bei Laien und Experten. In: W. Wirth & W. Schweiger (Eds.), *Selektion im Internet. Empirische Analysen zu einem Schlüsselkonzept* (pp. 181-196). Opladen: Westdeutscher Verlag.
- Wirth, W., & Schweiger, W. (1999). Selektion neu betrachtet: Auswahlentscheidungen im Internet. In: W. Wirth & W. Schweiger (Eds.), *Selektion im Internet. Empirische Analysen zu einem Schlüsselkonzept* (pp. 43-74). Opladen: Westdeutscher Verlag.
- Wolfram, D. & Dimitroff, A. (1997). Preliminary Findings On Searcher Performance and Perceptions of Performance in a Hypertext Bibliographic Retrieval System. In: *Journal of the American Society for Information Science*, 48 (2), 1142-1145.