

Untersuchungen zur vergleichenden Evaluation einer natürlichsprachlichen Bibliothekssoftware

Claudia Moranz, Kai-Christoph Hamborg, Günther Gediga
Universität Osnabrück, FB Psychologie und Gesundheitswissenschaften,
Arbeits- und Organisationspsychologie, Institut für Evaluation und Marktanalysen

Zusammenfassung

In drei Evaluationsuntersuchungen (eine Feldstudie, zwei Laboruntersuchungen) wurden das natürlichsprachliche Bibliothekssystem OSIRIS (Osnabrueck Intelligent Research Information System) und das strukturiertsprachliche Bibliothekssystem OPAC (Online-Public-Access-Catalog) im Hinblick auf ihre Benutzbarkeit verglichen. Während sich bei erfahrenen Nutzern ein recht starres Suchverhalten bei der Nutzung des OPAC abzeichnete, konnte das OSIRIS-System von Computeranfängern bei der Themensuche effektiver und - bei bestimmten Sucharten - tendenziell effizienter als das OPAC-System genutzt werden. Allerdings ließ sich auch hier die verbreitete Vorstellung von der Einfachheit und unmittelbaren Verständlichkeit von NL-Systemen generell nicht bestätigen. Suchmöglichkeiten bei OSIRIS werden insbesondere von mit dem OPAC vorerfahrenen Nutzern nicht immer erkannt, was für eine Implementierung von Unterstützungskonzepten spricht, wie zum Beispiel adäquates Feedback, explizite Eingabehinweise oder Nutzertraining.

1 Einleitung

Literaturrecherchen mit strukturierten Abfragesprachen sind besonders für unerfahrene Nutzer häufig schwierig. Oftmals ist kein Wissen über die zugrunde liegende Datenstruktur vorhanden und noch seltener die Fähigkeit, Boolesche Operatoren zu nutzen. Gerade im Umgang mit OPACs (Online Public Access Catalog) wird von Nutzungsproblemen, wie zu viele Treffer, zu wenig Treffer oder Orientierungslosigkeit (Schulz, 1994) berichtet. Verschiedene Studien kommen zu dem Ergebnis, dass Nutzer bislang eher (Such-) Strategien mit Zusatzaufwand verfolgen (vgl. Dreis, 1994, Rossoll, 1996, Schulz, 1990, 1994). So wurde z.B. beobachtet, dass sehr große Treffermengen auf Relevanz durchgesehen werden, statt sie mit Booleschen Operatoren einzugrenzen. Weiterhin wenden OPAC-Nutzer u. U. nicht nur ineffiziente Suchstrategien an, sondern müssen darüber hinaus auch insgesamt *mehr* Strategien als erforderlich generieren, um zum gewünschten Ergebnis zu gelangen. Zum Beispiel formulieren Nutzer für *ein* Suchanliegen *mehrere* Abfragen, statt mit Hilfe der Trunkierungstechnik nur eine einzige einzugeben.

Als Alternative zum OPAC wurde an der Universität Osnabrück ein natürlichsprachliches System entwickelt (OSIRIS; Recker et al., 1996, Ronthaler, 1998) und parallel zu dem bestehenden OPAC-System für die Nutzer der Universitätsbibliothek, Studierende und Mitarbeiter der Universität, per WWW zugänglich gemacht. In drei Evaluationsstudien wurde untersucht, ob durch das natürlichsprachlich zu bedienende System die dargestellten Nutzungsprobleme des OPAC vermieden werden können.

Nach Auffassung zahlreicher Autoren ist natürliche Sprache gerade für unerfahrene Nutzer der beste Weg, mit Computern zu kommunizieren (s. Ogden & Bernick, 1997). Die vermutete Überlegenheit natürlichsprachlicher Systeme gegenüber durch strukturierte Abfragesprachen zu bedienende Systeme wie OPACs wird allerdings kontrovers diskutiert. Die Ergebnisse verschiedener Evaluationsstudien in diesem Bereich weisen keine einheitliche Tendenz auf (vgl. Bier-

mann et al., 1983; Fink et al., 1985; Jarke et al., 1985; Small & Weldon, 1983; Shneiderman, 1978; Turtle, 1994; Walker & Whittaker, 1989).

2 Gegenstand der Evaluation

Das Bibliothekssystem OPAC war standardmäßig über Telnet zugänglich und verfügte entsprechend über eine zeichenorientierte Benutzungsschnittstelle (CUI). Das System wird kommandosprachlich bedient. Suchabfragen werden mittels einer strukturierten Abfragesprache formuliert, die nicht auf einer natürlichen, sondern auf einer künstlichen Eingabesyntax basiert. Folgende Suchmodi stehen dem Benutzer zur Verfügung: *Titelstichwörter*, *Personennamen*, *Körperschaften*, *Kongressitel*, *Serientitel*, *Systematik*, *Nummern* und *Signaturen*. Durch die Verwendung Boolescher Operatoren lassen sich Suchabfragen modifizieren.

Das natürlichsprachlich zu bedienende Bibliothekssystem OSIRIS ist in einem gemeinsamen Projekt der Universitätsbibliothek Osnabrück (UB) und des Instituts für semantische Informationsverarbeitung der Universität Osnabrück ab 1996 entwickelt worden. Zum Zeitpunkt der hier dargestellten Untersuchungen lag OSIRIS in der Version 2.0 vor. Das System ist als WWW Anwendung realisiert und verfügt entsprechend über eine grafische Benutzeroberfläche. Als Suchmodi stehen dem Benutzer eine *Themensuche*, eine *Autorensuche*, eine *schnelle Suche* (Formalrecherche) und eine *Zeitschriftensuche* zur Verfügung. OSIRIS akzeptiert natürlichsprachliche Sucheingaben. Diese werden syntaktisch und semantisch analysiert, um die Abbildung der Anfrage auf die vorhandenen Daten zu verbessern. Komposita beispielsweise werden zerlegt, flektierte Formen normalisiert und Plural wird auf Singular zurückgeführt (Morphologiekomponente). Häufige und einfache Eingabefehler (zum Beispiel Verdoppelung, Buchstabendreher) werden erkannt und dem Benutzer konstruktiv rückgemeldet. Das heißt, es wird nicht nur auf einen eventuell vorliegenden Rechtschreibfehler hingewiesen, sondern man bekommt eine korrigierte Version der Eingabe vorgelegt. Bei der *Formalrecherche* nach einem bestimmten Buch eines Autors erledigt OSIRIS für den Benutzer die AND-Verknüpfung automatisch. Des Weiteren dürfen Titelstichwörter von der im Titel vorliegenden Form abweichen. Die Themensuche erfolgt durch die Vervollständigung des Halbsatzes „*Suche Literatur zum Thema...*“, wobei Entsprechendes für die anderen Suchmodi gilt. Zusätzlich zu den konkreten Titeln, aber auch in Null-Treffer-Situationen sieht OSIRIS ein Browsingangebot in den Klassenbezeichnungen der Systematik vor (vgl. Recker et al., 1996). Es sind drei charakteristische Eigenschaften, die OSIRIS im Gegensatz zum klassischen OPAC-System auszeichnen und die die Benutzbarkeit des Systems für die Literatursuche verbessern sollen:

- eine robuste, natürlichsprachliche Benutzerschnittstelle zur Datenbank,
- eine intelligente, automatische Aufbereitung des verfügbaren Datenbestandes in einer Wissensbasis,
- eine funktionale graphische Benutzeroberfläche.

Ausgangspunkt für die Untersuchungen war die Frage, ob das natürlichsprachliche Abfrageparadigma nach Benutzbarkeitsaspekten Vorteile gegenüber dem durch die strukturierte Abfragesprache zu bedienenden System bringt. Den Untersuchungen wurde das Benutzbarkeitskonzept nach ISO 9241/11 mit den Teilkomponenten: Effizienz, Effektivität und Akzeptanz zugrunde gelegt. Die erste Evaluationsuntersuchung wurde im Feld, die zwei folgenden im Labor durchgeführt.

3 Untersuchung 1

In dieser Evaluationsstudie wurden insgesamt 120 Nutzer der UB Osnabrück im Feld zu den Programmen befragt und Nutzungszeiten erfasst. Je 30 Personen arbeiteten mit:

- (1.) dem Standardsystem der Bibliothek (Telnet/OPAC),
- (2.) mit der zum Evaluationszeitpunkt aktuellen Version des OSIRIS-Systems (Version 2.0) und
- (3.) zusätzlich mit einer HTML Version des OPAC (OPAC/HTML) als auch
- (4.) einer älteren OSIRIS Version (Version 1.1), die hier nicht weiter berücksichtigt wird.

Bei der gewählten Stichprobengröße und einem α -Niveau von 0,05 kann (bei $1-\beta = 0.80$, $df = 2$) von mittleren bis großen Effekten ausgegangen werden (Bortz & Döring, 1995).

Die Untersuchung wurde von zwei wissenschaftlichen Hilfskräften in einem von der UB Osnabrück zur Verfügung gestellten Untersuchungsraum durchgeführt. Die Teilnehmer der Untersuchung wurden am Eingang der Universitätsbibliothek akquiriert. Die Teilnahme an der Untersuchung wurde mit 10 DM vergütet.

Zu Beginn der Untersuchung wurden Angaben zur Person (Alter, Geschlecht), Computervorerfahrung und der Rechercheabsicht des Bibliotheksbesuchs erhoben. Danach hatten die Untersuchungsteilnehmer bedarfsweise die Möglichkeit, das genutzte System fünf Minuten zu explorieren. Dies war insbesondere darum notwendig, weil der OPAC/HTML der Öffentlichkeit zum Zeitpunkt der Untersuchung nicht zur Verfügung stand. Dies gilt auch für OSIRIS 2.0, das zwar per Internet zugänglich war, aber zum Evaluationszeitpunkt noch nicht standardmäßig in der Bibliothek zur Verfügung stand. Nach der Explorationsphase führten die Untersuchungsteilnehmer die beabsichtigte Recherche durch. Hierfür gab es kein Zeitlimit. Die Bearbeitungszeit wurde von den Versuchsleitern gestoppt. Nach Beendigung der Recherche wurden die Untersuchungsteilnehmer zu der Qualität und Quantität der Rechercheergebnisse und der Bewertung des Systems nach ergonomischen Gesichtspunkten befragt.

3.1 Ergebnisse

Mit den untersuchten Systemen arbeiteten jeweils 14 weibliche und 16 männliche Benutzer. Die Gruppen wurden gematcht, um geschlechtsspezifische Effekte auszuschließen. Die Untersuchungsgruppen unterscheiden sich nicht signifikant in Bezug auf das Alter der Teilnehmer ($M_{\text{OPAC/Telnet}} = 27,57$, $M_{\text{OPAC/HTML}} = 27,73$, $M_{\text{OSIRIS}} = 28,10$; $F = 0,71$, $p = 0,93$). Von den Teilnehmern der Untersuchung hatten über 80% (OPAC/Telnet Gruppe) bzw. über 90% (OPAC/HTML, OSIRIS Gruppe) Vorerfahrung mit dem OPAC/Telnet, aber nur 3% aus der OSIRIS-Gruppe mit ihrem System 5 Personen aus der OPAC/Telnet Gruppe hatten keine Erfahrung mit einem der Systeme im Gegensatz zu einer Person aus der OPAC/HTML Gruppe und keiner Person aus der OSIRIS Gruppe. Eine Person aus der OPAC/HTML Gruppe hatte Vorerfahrung sowohl mit OPAC/Telnet als auch mit OSIRIS. Bezüglich der Vorerfahrung mit den untersuchten Systemen kann zusammengefasst werden, dass dem Großteil der Probanden der OPAC/Telnet bekannt war, nicht jedoch das OSIRIS-System (und auch nicht der OPAC/HTML).

Aus den von den Untersuchungsteilnehmern genannten Recherchezielen wurden vier Kategorien gebildet. Die Reliabilität der Kategorien wurde durch Zuordnung der Rechercheziele durch zwei unabhängige Rater überprüft. Sie kann als befriedigend bezeichnet werden ($Kappa = 0,63$). In den Fällen, in denen die zwei Rater nicht übereinstimmten, wurden ein Drittrating und eine Zuordnung durch „Mehrheitsentscheidung“ vorgenommen. In drei Fällen stimmten die drei Rater nicht überein. Diese Fälle wurden im Folgenden nicht weiter berücksichtigt. Die Kategorien lauten: 1.) Suche nach einem Thema, 2.) Suche nach einem speziellen Buch, Autor, Titel, 3. Suche sowohl nach 1. und nach 2. und 4.) Suche nach einer Zeitschrift.

Den Ergebnissen zufolge wollten die Nutzer zum größten Teil Themensuchen vornehmen (49,4%). An zweiter Stelle steht die Suche nach einem bestimmten Buch, Titel oder Autor (29,9%), gefolgt von komplexeren Suchabsichten, die sowohl Themensuche und die konkrete Suche nach einem bestimmten Buch, Titel oder Autor umfassten (19,5%). Nur in einem Fall bestand die Absicht, nach Zeitschriften zu suchen (1,1%). Die Häufigkeitsverteilung der Recher-

cheziele über die untersuchten Systeme zeigt keine bedeutsamen Abweichungen von den erwarteten Häufigkeiten

Die für die Recherche benötigte *Bearbeitungszeit* ist ein wichtiger Indikator für die Effizienz der Systeme. Die Mediane der Bearbeitungszeiten mit dem OPAC/Telnet (11,5 Min), OPAC/HTML (11,0 Min) und OSIRIS 2.0 (9,0 Min) unterscheiden sich nicht signifikant voneinander (Kruskal Wallis (KW-) Test, $\text{Chi}^2 = 1,11$, $\text{df} = 2$, $p = 0,57$).

Zur Bewertung der Effektivität der Systeme wurden die Untersuchungsteilnehmer mit einem Fragebogen (fünfstufigen Ratingskala mit den Antwortkategorien: „stimmt nicht“, „stimmt wenig“, „stimmt mittelmäßig“, „stimmt ziemlich“, „stimmt sehr“ und zusätzlich: „keine Angabe“) befragt, ob:

- alle Informationen, nach denen gesucht wurde, auch gefunden wurden (Frage R1),
- die gefundenen Informationen inhaltlich den Erwartungen entsprachen (Frage R3),
- die Informationen genau die Informationen enthielten, nach denen gesucht wurde (Frage R4).

Zusätzlich wurde die Angemessenheit der vom System gefundenen Informationsmenge beurteilt („zu wenig“, „genau richtig“, „zu viel“, Frage R2).

Die Beantwortung der Fragen R1 bis R4 ergibt keine statistisch unterschiedliche Bewertung der Systeme (KW- Tests mit $\text{df} = 2$, *Frage R1*: $\text{Chi}^2 = 1,38$, $p = 0,50$; *Frage R2*: $\text{Chi}^2 = 3,6$, $p = 0,17$, *Frage R3*: $\text{Chi}^2 = 3,16$, $p = 0,21$; *Frage R4*: $\text{Chi}^2 = 1,66$, $p = 0,44$). Die Effekte, die hier zwischen $\phi = 0,1$ und $\phi = 0,2$ variieren, zeigen, dass bei den Fragen R1 bis R4 und für die Bearbeitungszeit bestenfalls kleine – und damit sicherlich wenig relevante – Unterschiede zwischen den Systemen zu beobachten wären.

Die *ergonomische Bewertung* der untersuchten Programme wurde mit einer deutschen Übersetzung der Kurzversion des „Questionnaire for User Interface Satisfaction“ (QUIS) durchgeführt (Shneiderman, 1992). QUIS umfaßt fünf neunstufige Skalen mit geschlossenem Antwortformat. Die fünf Skalen lauten: 1.) Gesamtbewertung, 2.) Bildschirm, 3.) Fachwörter und Systeminformationen, 4.) Erlernen und 5.) Fähigkeiten des Systems. Die Reliabilität der Skalen wurde systemspezifisch bestimmt (Cronbachs α). Mit Ausnahme von Skala 5 erwiesen sie sich als befriedigend bis gut.

Abbildung 1 zeigt die Befragungsergebnisse. Hohe Skalenwerte der neunstufigen Skala drücken eine positive Systembewertung aus. Auf der ersten („Gesamteindruck des Systems“) und der letzten Skala („Fähigkeiten des Systems“) liegen die Systembewertungen weitgehend überein. Auf der zweiten („Bildschirm“), dritten („Fachwörter und Systeminformationen“) und vierten Skala („Erlernen“) zeigen sich Unterschiede bis zu ca. einem Skalenpunkt zugunsten des OSIRIS-Systems. Abgesehen von der Skala „Bildschirm“ (QUIS 2), auf der OPAC/HTML etwas besser bewertet wird, sind die Profile der beiden OPAC-Systeme weitgehend identisch.

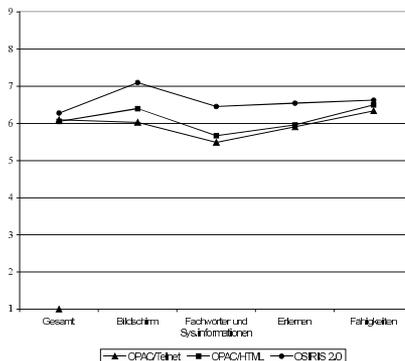


Abb. 1: Bewertung der Systeme mit dem QUIS-Fragebogen

Der statistische Vergleich von OSIRIS 2.0 und dem OPAC/HTML zeigt auf der Skala „Bildschirm“ (QUIS2) einen tendenziell signifikanten Effekt ($p = 0,075$). Auf den weiteren Skalen unterscheiden sich die beiden Systeme statistisch nicht bedeutsam. OSIRIS 2.0 und OPAC/Telnet unterscheiden sich auf der Skala „Bildschirm“ (QUIS2) signifikant ($p = 0,009$) und auf den Skalen „Fachwörter und Systeminformationen“ (QUIS3, $p = 0,059$) und „Erlernen“ (QUIS 3, $p = 0,058$) tendenziell.

3.2 Zusammenfassung

In Untersuchung 1 finden sich keine statistisch signifikanten Unterschiede zwischen den Systemen in Bezug auf deren Effektivität und Effizienz. Dies ist insofern bemerkenswert, als die Nutzer in der Studie über Vorerfahrung mit dem OPAC-, nicht jedoch mit dem OSIRIS-System verfügten. Nach ergonomischen Gesichtspunkten zeigen sich Stärken des natürlichsprachlichen OSIRIS-Systems im Verhältnis zum OPAC/Telnet auf den Skalen, die die Benutzungsschnittstelle, aber auch die Erlernbarkeit betreffen. Die Skalenprofile von OPAC/Telnet und OPAC/HTML verlaufen, mit Ausnahme der „Bildschirm“-Skala, nahezu parallel. OSIRIS wird jedoch auch bezüglich dieser Skala tendenziell besser als der ebenfalls mit graphischer Benutzungsschnittstelle versehene OPAC/HTML bewertet. Daher kann vermutet werden, dass die bessere ergonomische Bewertung von OSIRIS nicht alleine auf die graphische Benutzungsoberfläche zurückzuführen ist.

4 Untersuchung 2

In Anlehnung an die ISO 9241/11 wurde in dieser und der folgenden Untersuchung die Effizienz von OPAC/Telnet und von OSIRIS 2.0 in Laboruntersuchungen evaluiert (vgl. Moranz, 2000). Bei der *Effizienz* geht es zum einen darum, innerhalb welcher Zeit relevante Bücher gefunden werden (*zeitliche* Effizienz). Zum anderen aber - und hier liegt der Schwerpunkt der Untersuchungen - werden im Rahmen der Effizienzbetrachtung Aspekte der menschlichen Handlungsregulation beleuchtet. Dabei wird in Anlehnung an handlungstheoretische Modelle nach Norman (1986) und Hacker (1986) untersucht, inwieweit das eine oder andere Bibliothekssystem die Nutzung in der Weise unterstützt, dass die objektiv vorhandenen Funktionalitäten der Systeme erkannt und sinnvoll für die Generierung von Handlungs- respektive Suchstrategien genutzt werden können (*mentale* Effizienz). Die *Effektivität* wird daran festgemacht, ob relevante Bücher gefunden werden.

Die abhängigen Variablen der Untersuchungen 2 und 3 sind (a) die Suchstrategien, d.h. Strategiearten und Anzahl der Strategiewechsel, (b) die pro Aufgabe benötigte Zeit und (c) der Erfolg der Aufgabebearbeitung. Die unabhängige Variable ist die Art des Bibliothekssystems (OPAC versus OSIRIS).

4.1 Methode

Die Untersuchungsgruppen waren in bezug auf die genutzten Systeme unterschiedlich vorgefahren. Die Teilnehmer der OPAC-Gruppe ($N = 10$) verfügten im Umgang mit diesem Bibliothekskatalog über durchschnittliche Vorerfahrung. Die Teilnehmer der OSIRIS-Gruppe ($N = 10$) hatten dagegen mit diesem System noch nicht gearbeitet, da es zu dem Untersuchungszeitpunkt noch nicht öffentlich zugänglich war. Es wurden jeweils drei vorgegebene Suchaufgaben bearbeitet:

- eine Themensuche (Autor oder Titel sind hierbei nicht bekannt, lediglich das Thema ist vorgegeben),
- eine Autor-Titel-Suche (sowohl der Autor als auch der Titel des gesuchten Buches sind bekannt).

Darüber hinaus gab es eine Aufgabe zu Zeitschriftensuche, auf die an dieser Stelle nicht weiter eingegangen wird (vgl. dazu ausführlicher Moranz, 2000). Im Anschluss an die Aufgabenbearbeitung wurden die Versuchspersonen mit der Methode der Videokonfrontation (s. Hamborg & Greif, 1999) zu den verwendeten Suchstrategien befragt.

Tabelle 1: Kategoriensystem der Suchstrategien

Kategorie	Erläuterung
<i>Strategien bei der Ersteingabe</i>	<i>(Die in der Aufgabe gegebene Information wird in eine Suchabfrage transformiert.)</i>
Grober Filter	Die in der Aufgabe gegebene Information wird nicht vollständig genutzt. Z.B. wird nur der Autor eingegeben, obwohl Autor und Titel bekannt sind.
Direkte Übernahme	Alle in der Aufgabe gegebenen Informationen werden in der Suchabfrage berücksichtigt.
<i>Strategien im weiteren Suchverlauf</i>	<i>(Eine bereits getätigte Suchabfrage wird modifiziert.)</i>
Suche eingrenzen	Alle Aktionen, die dazu dienen, das Suchergebnis zu verkleinern oder zu spezifizieren, z.B. durch die Verwendung Boolescher Operatoren.
Suche erweitern	Alle Aktionen, die dazu dienen, das Suchergebnis zu vergrößern, z. B. durch Weglassen von zuvor in der Suchabfrage verwendeten Wörtern.
Umformulieren	Modifizieren einer Suchabfrage ohne Eingrenzen oder Erweitern.
Hinweise aufgreifen	Informationen aus vorherigen Suchabfragen für die Formulierung der neuen Suche verwenden, z. B. Beispiel Notationen oder Schlagwörter.
Fachgebiete durchsehen	Das Browsingangebot bei OSIRIS wird wahrgenommen.
Scrollen	Durch Scrollen oder Blätter in den angezeigten Dokumenten auf passende Titel stoßen.
Zusatzinformationen einholen	Alle Aktionen, die dazu dienen, mehr Informationen über bestimmte Dokumente einzuholen. Ein Beispiel ist das Anwählen des Vollanzeigemodus, um das Erscheinungsjahr herauszufinden. Aber auch das Lesen von Hilfetexten fällt hierunter.
Sonstige	Hierunter fallen alle die Strategien, die sich keiner der aufgeführten Kategorien zuordnen lassen.

Für die statistische Auswertung der Suchstrategien wurden die Strategienennungen der Versuchspersonen aus dem Interviewprotokoll extrahiert und mit Hilfe der *Qualitativen Inhaltsanalyse* (Mayring, 1997) zu Gruppen zusammengefasst. In Tabelle 1 ist das Kategoriensystem für die Strategien dargestellt. Die Reliabilität wurde durch ein Expertenrating geprüft. Sie kann nach Robson (1993) als gut bezeichnet werden ($Kappa = 0,71$). Es stellte sich bei der Kategorienbildung als sinnvoll heraus, zwischen Strategien der Ersteingabe, sprich solchen, die sich auf die zuerst getätigte Eingabe pro Aufgabe beziehen, und Strategien im weiteren Suchverlauf zu differenzieren. Zum einen wird die in den Suchaufgaben zugrunde gelegte Information in eine Suchabfrage transformiert und zum anderen wird die bereits getätigte Suchabfrage modifiziert.

4.2 Ergebnisse

Die Auswertung dieser Untersuchung hat hinsichtlich der *mentalen* Effizienz ergeben, dass OPAC-Nutzer bei der Autor-Titel-Suche signifikant häufiger von der ineffizienten Suchstrategie „Grober Filter“ Gebrauch machten und die Verknüpfungsfunktion über Boolesche Operatoren nicht ausnutzten ($\chi^2 = 5,625$, $df = 1$, $p = 0,018$, $\phi = 0,530$), das System also vergleichsweise wenig effizient nutzten. Darüber hinaus konnten bezüglich der verwendeten Strategiearten keine weiteren Unterschiede gefunden werden. Im Bereich der Themensuche wechselten OSIRIS-Nutzer signifikant häufiger ihre Strategien ($M_{OSIRIS} = 3,00$, $s_{OSIRIS} = 1,56$; $M_{OPAC} = 1,10$, $s_{OPAC} =$

1,29; $t = -2,967$, $p = 0,008$, $d = -0,7$). Abbildung 2 und Abbildung 3 dienen der Visualisierung der vorgenommenen Strategiewechsel.

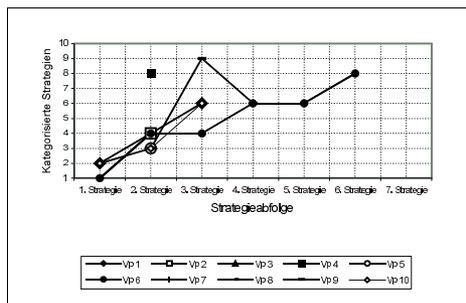


Abb. 2: Strategieabfolge:
Themensuche - OPAC (Studie I)

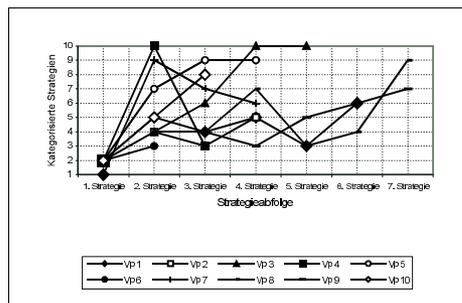


Abb. 3: Strategieabfolge:
Themensuche - OSIRIS (Studie I)

Auf der Y-Achse sind die verschiedenen kategorisierten Strategien abgetragen, auf der X-Achse die Abfolge der Strategien, wobei sich hinter „1. Strategie“ die Strategien der Ersteingaben verbergen, weshalb die Graphen in der Regel bei Kategorie 1 oder 2 respektive „Grober Filter“ oder „Direkte Übernahme“ beginnen. Ab der „2. Strategie“ handelt es sich dann um Strategien im weiteren Suchverlauf. Die einzelnen Graphen respektive Punkte sind den jeweiligen Versuchspersonen zugeordnet und überlagern sich teilweise. Einzelne Punkte bedeuten, dass die entsprechende Versuchsperson bis auf eine keine weiteren Strategien mitgeteilt hat.

Was die zeitliche Effizienz angeht, so konnte weder bei der Autor-Titel-Suche noch bei der Themen- oder Zeitschriftensuche ein statistischer Beleg für Unterschiede zwischen den Systemen bezüglich der Bearbeitungszeiten gefunden werden. Auch im Hinblick auf die Effektivität konnten keinerlei Unterschiede zwischen den Systemen ermittelt werden.

5 Untersuchung 3

In Untersuchung 3 verfügten die Probanden weder mit OPAC (N = 10) noch mit OSIRIS (N = 10) über Vorerfahrung. Das Ziel bestand darin, zu untersuchen, ob sich die in Untersuchung 2 ermittelten Ergebnisse auf Merkmale der Bibliothekssysteme zurückführen lassen oder von der unterschiedlichen Vorerfahrung der Untersuchungsgruppen abhängen. Es wurden dieselben Suchaufgaben wie in Untersuchung 2 bearbeitet.

Das Erhebungs- und Auswertungsverfahren entsprechen denen aus Untersuchung 2. Die Interraterreliabilität der Strategiekategorisierung kann wiederum als gut bezeichnet werden ($\text{Kappa} = 0,77$).

5.1 Ergebnisse

Die Auswertung zeigt in Bezug auf die mentale Effizienz auf dem 10%-Niveau für die Ersteingaben der Themensuche signifikante Unterschiede: OSIRIS-Nutzer haben hier häufiger die gesamte in der Suchaufgabe gegebene Information übernommen als OPAC-Nutzer ($\text{Chi}^2 = 2,813$, $\text{df} = 1$, $p = 0,094$, $\phi = 0,375$), das System also tendenziell effizienter genutzt. Auch haben OSIRIS-Nutzer im weiteren Verlauf der Themensuche erkennbar die Strategie „Fachgebiete durchsehen“ verwendet ($\text{Chi}^2 = 11,250$, $\text{df} = 1$, $p = 0,001$, $\phi = 0,75$). Was die Autor-Titel-Suche betrifft, so konnten keine Unterschiede im Hinblick auf die Strategiearten gefunden werden. Bei OPAC

und OSIRIS haben annähernd gleich viele Strategiewechsel stattgefunden (vgl. auch Abbildung 4 und Abbildung 5).

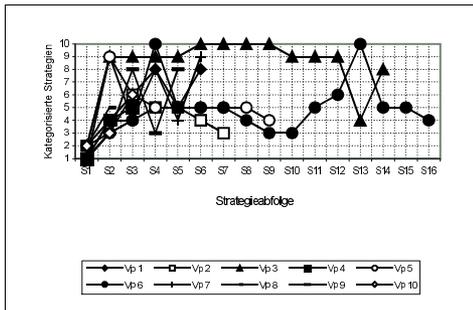


Abb. 4: Strategieabfolge:
Themensuche - OPAC (Studie II)

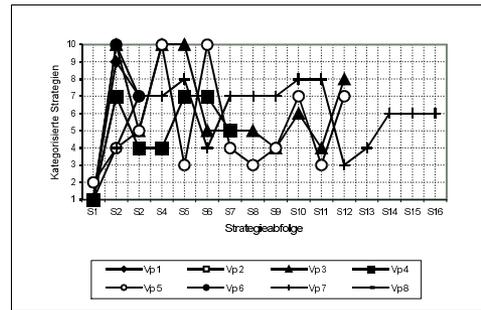


Abb. 5: Strategieabfolge:
Themensuche - OSIRIS (Studie II)

Die Systeme unterscheiden sich jedoch bei 10% Irrtumswahrscheinlichkeit hinsichtlich der benötigten Bearbeitungszeit (*zeitliche Effizienz*) (Themensuche: Median (Md)_{OPAC} = 238 Sekunden, Md_{OSIRIS} = 96,5 Sekunden, $p = 0,091$, zweiseitiger U-Test; Zeitschriftensuche: Md_{OPAC} = 393 Sekunden, Md_{OSIRIS} = 50 Sekunden, $p = 0,001$, zweiseitiger U-Test).

Weiterhin konnten mit OSIRIS insgesamt mehr Personen die in der Themensuche geforderte Titelzahl als mit OPAC finden ($\chi^2 = 3,60$, $df = 1$, $p = 0,058$, $\phi = 0,424$), hier zeigt sich also ein *Effektivitätsvorteil* zugunsten von OSIRIS.

6 Zusammenfassung und Diskussion

Die Untersuchungen 1 und 2 wurden mit Nutzern durchgeführt, denen das OPAC-System, nicht aber das OSIRIS-System vertraut war. In keiner der beiden Untersuchungen finden sich Unterschiede zwischen dem natürlichsprachlichen (OSIRIS 2.0) und dem durch die strukturierte Abfragesprache zu bedienenden System (OPAC/Telnet) in Bezug auf die zeitliche Effizienz und die Effektivität. Wohl aber wird die weniger effiziente Nutzung des OPAC-Systems bei der Realisierung von Suchstrategien (bei der Autor-Titel Suche) deutlich. Dies erscheint aufgrund der unterschiedlichen Vorerfahrung der Probanden in diesen Untersuchungen bemerkenswert. Für *gleichermaßen* unerfahrene Nutzer (Untersuchung 3) zeigen sich jedoch Vorteile des OSIRIS-Systems mit Bezug auf die Effektivität und bedingt für die Effizienz (zeitliche Effizienz, Effizienz der Formulierung von Suchabfragen) bei der am häufigsten verfolgten Suchart (Themensuche).

Die Befunde zu den Strategiewechseln aus Untersuchung 2 und 3 lassen vermuten, dass die, im Unterschied zu OSIRIS, geringe Varianz der Strategiewechsel bei der Bedienung des OPAC in Untersuchung 2 ein Routiniesierungseffekt der Vorerfahrung der OPAC Nutzer ist. Sie finden sich nicht bei den wenig vorerfahrenen Nutzern (Untersuchung 3).

Es finden sich weitere Belege dafür, dass weder unerfahrene noch gelegentliche OPAC-Nutzer in der Lage sind, Boolesche Operatoren zu benutzen (vgl. Dreis, 1994; Rossoll, 1996; Schulz, 1990,1994). Die Benutzer entscheiden sich bei der Autor-Titel-Suche mit OPAC entweder für die Komponente „Autor“ oder für die Komponente „Titel“, ohne beides zu verknüpfen. Bei OSIRIS hingegen wurde die gesamte in der Aufgabe gegebene Information, also Autor *und* Titel, von den Benutzern in der Suchformulierung übernommen. Eine mögliche Erklärung wäre, dass ein Beispiel für eine solche Suche unter dem Suchmodus *Quick-Info* explizit angegeben ist. OSIRIS übernimmt hier für den Benutzer automatisch die Verknüpfung von Autor und Titel.

Nach ergonomischen Gesichtspunkten wurde das natürlichsprachliche OSIRIS-System (Untersuchung 1) auf den Skalen zur Qualität der Bildschirmdarstellung und der Terminologie, aber auch, im Falle des OPAC/Telnet, in Bezug auf die Erlernbarkeit, (tendenziell) besser als die OPAC-Systeme bewertet. Da sich die Profile der ergonomischen Bewertung von OPAC/Telnet und dem OPAC/HTML mit graphischer Benutzungsoberfläche kaum unterscheiden, kann angenommen werden, dass die positive Bewertung von OSIRIS nicht alleine auf dessen graphische Benutzungsoberfläche zurückzuführen ist.

Als Fazit kann aus den Untersuchungen gezogen werden, dass sich für das untersuchte NL-System (OSIRIS) ergonomische Stärken (Bildschirm, Wording, Erlernbarkeit) im Vergleich zu dem mit der strukturierten Abfragesprache zu bedienenden System (OPAC/Telnet) bei Stichproben mit unterschiedlicher Vorerfahrung (Untersuchung 1) abzeichnen. Es finden sich jedoch keine Vorteile (aber auch keine Nachteile) in Bezug auf die Effektivität und Effizienz (Untersuchung 1 und 2). Nur für mit beiden Systemen gleichermaßen unerfahrene Nutzer schlägt sich die plausible Vorstellung bezüglich der Einfachheit und unmittelbaren Verständlichkeit von NL-Systemen in Form besserer Effektivität und Effizienz zugunsten des NL Systems nieder (Untersuchung 3). Für gelegentliche oder erfahrene Nutzer steht diese Analyse noch aus. Für die hier berücksichtigten Systeme könnte sie erst jetzt vorgenommen werden, nachdem OSIRIS hinreichend lange im Einsatz ist. Bei der Interpretation der Befunde sollte man jedoch in Betracht ziehen, dass nicht unbedingt die Option für eine natürlichsprachliche Eingabe bei OSIRIS für eine größere Effizienz respektive Effektivität des NL-Systems in Untersuchung 3 ausschlaggebend gewesen sein muss. Vielmehr dürfte auch die Aufbereitung und Präsentation der Ergebnisse zu der höheren Effektivität von OSIRIS beigetragen haben. Insbesondere muss hier die Browsing-Möglichkeit erwähnt werden, die durch die unerfahrenen Nutzer als sinnvolle Suchvariante erkannt und genutzt wurde (Untersuchung 3). Die tendenzielle Überlegenheit von OSIRIS in Bezug auf die zeitliche Effizienz fand sich in Untersuchung 3 u.a. im Bereich der Zeitschriftensuche, die in dieser Form durch OPAC nicht angeboten wurde.

Die Tatsache allein, dass es sich um ein natürlichsprachliches System handelt und dass infolgedessen auch „natürliche“ Eingaben möglich sind, reicht zur adäquaten Bedienung solcher Systeme nicht aus. Sowohl OPAC-vorerfahrenen als auch gänzlich unerfahrenen Nutzern ist keinesfalls klar, dass OSIRIS eine Verknüpfung von Autor und Titel bei der Formalrecherche übernimmt (Untersuchung 2 und 3). Nutzer, von denen man aufgrund ihrer Vorerfahrung annehmen kann, dass sich bei ihnen ein in bezug auf OPAC schon routinisiertes Suchverhalten herausgebildet hat, übertragen dieses auch bei der Themensuche auf OSIRIS und legen sich selbst im Sinne eines „negativen Transfers“ Restriktionen auf (Untersuchung 2). So halten sie sich bei der Formulierung ihrer Suchabfragen an die zwar bei OPAC, nicht aber bei OSIRIS vorgeschriebene Begrenzung der maximalen Anzahl einzugebender Wörter. Es wird ersichtlich, dass OSIRIS nicht selbsterklärend ist. Vielmehr muss das NL-System seine „Angebote“ besser vermitteln und deutlich machen, was möglich ist und was nicht. Explizite Hinweise auf die Eingabemöglichkeiten bei OSIRIS (wie sie ja bei der Autor-Titel-Suche im Ansatz schon vorhanden sind) wären als Unterstützung für den Nutzer hilfreich. Wünschenswert ist darüber hinaus ein für den Nutzer verständliches Feedback darüber, wie das System die Sucheingaben weiterverarbeitet. Auch ein entsprechendes Nutzertraining ist durchaus denkbar. Andernfalls werden die Bedienungsmöglichkeiten von OSIRIS insbesondere von den durch OPAC-Vorerfahrung geprägten Probanden nicht vollständig erkannt und genutzt.

7 Literatur

- Biermann, A.W., Ballard, B.W. & Sigmon, A.H. (1983): An experimental study of natural language programming. *International Journal of Man-Machine Studies*, 18, 71-87.
- Bortz, J. & Döring, N. (1995). *Forschungsmethoden und Evaluation für Sozialwissenschaftler*. 2. Auflage. Berlin: Springer.
- Dreis, G. (1994): Benutzerverhalten an einem Online-Publikumskatalog für wissenschaftliche Bibliotheken. Ergebnisse und Erfahrungen aus dem OPAC-Projekt der Universitätsbibliothek Düsseldorf. *Zeitschrift für Bibliothekswesen und Bibliographie*, [Sonderheft] (57). Frankfurt am Main: Vittorio Klostermann.
- Fink, P.K. & Sigmon, A.H. & Biermann, A.W. (1985): Computer control via limited natural language. *IEEE Transactions on Systems, Man, and Cybernetics*, 15, 54-68.
- Hacker, W. (1986). *Arbeitspsychologie: Psychische Regulation von Arbeitstätigkeiten*. Bern: Huber.
- Hamborg, K.-C. & Greif, S. (1999). Heterarchische Aufgabenanalyse. In H. Dunckel (Hrsg.). *Psychologische Arbeitsanalyse*. Zürich: VDF-Verlag (S. 147 – 177).
- ISO (1995). *ISO 9241. Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten. Teil 11: Anforderungen an die Gebrauchstauglichkeit - Leitsätze*. International Organization for Standardization.
- Jarke, M., Turner, J.A., Stohr, E.A., Vassiliou, Y., White, N.H. & Michielsen, K. (1985): A field evaluation of natural language for data retrieval. *IEEE Transactions on Software Engineering*, 11, 97-114.
- Mayring, Ph. (1997). *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. 6. Auflage. Weinheim: Deutscher Studien Verlag.
- Moranz, C. (2000). *OPAC oder OSIRIS? Evaluationsstudie zum Vergleich der Benutzbarkeit eines strukturiert- und eines natürlichsprachlichen Bibliothekssystems*. Unveröffentlichte Diplomarbeit am Fachbereich Psychologie und Gesundheitswissenschaften der Universität Osnabrück.
- Norman, D.A. (1986). Cognitive engineering. In D.A. Norman & S.W. Draper (Eds.). *User centered system design. New perspectives on human-computer interaction* (pp. 31-61). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Ogden, W.C. & Bernick, P. (1997). Using natural interfaces. In M. Helander, T.K. Landauer & P. Prabhu (Eds.). *Handbook of Human-Computer Interaction* (pp. 137-161). Elsevier Science.
- Recker, I., Ronthaler, M. & Zillmann, H. (1996): OSIRIS. *Bibliotheksdienst*, 30 (5), 833-848.
- Robson, C. (1993). *Real world research: a resource for social scientists and practitioner-researchers*. Oxford: Blackwell.
- Rossoll, E. (1996) Fortbildung: Round Table, Benutzeranleitung für den OPAC. *Bibliotheksdienst*, 30 (1), 107-111
- Ronthaler, M. (1998). Osiris: Qualitative Fortschritte bei der Literaturrecherche. In J. Dassow & R. Kruse (Hrsg.). *Informatik '98: Informatik zwischen Bild und Sprache; 28. Jahrestagung der Gesellschaft für Informatik*. Berlin: Springer.
- Schulz, U. (1990): Einige Aspekte zukünftiger Inhaltserschließung und Online-Benutzerkataloge. *Bibliothek*, 14 (3), 226-235.
- Schulz, U. (1994): Was wir über OPAC-Nutzer wissen: Fehlertolerante Suchprozesse in OPACs. *ABI-Technik*, 14 (4), 299-310.
- Shneiderman, B. (1978): Improving the human factors aspect of database interacting. *ACM Transactions on Database Systems*, 3, 417-439.
- Shneiderman, B. (1992): *Designing the User Interface. Strategies for Effective Human-Computer Interaction*. 2nd Edition. Reading, Massachusetts: Addison-Wesley Publishing Company.
- Small, D.W. & Weldon, L.J. (1983): An experimental comparison of natural and structured query language. *Human Factors*, 25 (3), 253-263.
- Turtle, H. (1994): Natural language vs. Boolean query evaluation: a comparison of retrieval performance. In *Proceedings of the 17th Annual International Conference on Research and Development in Information Retrieval* (pp. 212-221). London: Springer-Verlag.
- Walker, M. & Whittaker, S. (1989): *When natural language is better than menus: a field study*. Technical Report. Bristol: Hewlett Packard Laboratories.

Adressen der Autoren

Claudia Moranz
Zehntfeldstr. 199
81825 München

cmoranz@gmx.de

Dr. Kai-Christoph Hamborg
Universität Osnabrück
FB Psychologie und Gesundheits-
wissenschaften

Arbeits- und Organisationspsychologie
Seminarstr. 20
49069 Osnabrück
khamborg@uos.de

Günther Gediga
Universität Osnabrück
FB Psychologie und
Gesundheitswiss.
Institut f. Evaluation und
Marktanlysen
Brinkstr. 19
49143 Jegen