

Manfred Weichert

„Gibt es auch Wahnsin“¹

Benutzungsuntersuchungen an einem Hamburger WebOPC



Der Beitrag fasst die Ergebnisse einer Untersuchung von über 50 000 Suchanfragen am Hamburger Campus-Katalog zusammen. Der Campus-Katalog ist ein WEB-Katalog auf der Basis des Pica-Lokalsystems LBS3, dessen Protokoll-dateien die Rohdaten lieferten. Über die statistische Analyse der Suchanfragen hinaus, wird ein Versuch unternommen, die einzelnen Schritte einer Katalogrecherche zusammenzufassen und die Reaktion der Benutzer auf erfolglose Suchanfragen zu untersuchen.

„Gibt es auch Wahnsin“. Using the OPAC of the University of Hamburg

The paper summarizes the results of an investigation of more than 50 000 retrieval queries to the Hamburg Campus-Catalogue. The Campus-Catalogue is a Web catalogue based on the Pica Local System LBS3, whose log file supplied the raw data. Beyond the statistic analysis of the retrieval queries, an attempt is made to summarize the individual steps of a catalogue search and to examine the reaction of the users to unsuccessful retrieval queries.

„Gibt es auch Wahnsin“. Usage du catalogue en ligne de l'Université de Hambourg

L'article résume les résultats d'une enquête de plus de 50 000 équations de recherche dans le catalogue de l'université de Hambourg. Ce catalogue en ligne s'appuie sur le système local de Pica LBS3, dont les journaux fournissent des données de base analysées ici. Au-delà de l'analyse statistique des équations de recherche, on essaie de résumer les différentes étapes d'une recherche dans le catalogue et d'examiner les réactions des usagers face aux recherches sans résultats.

1 Gesucht wurde der Titel: „Gibt es auch Wahnsinn, hat es doch Methoden ...“ von Eva Jaeggi. Die vollständige Recherche ist im Anhang abgedruckt.

An vielen Orten gehen Bibliothekare der Frage nach, wie in den von ihnen angebotenen elektronischen Katalogen gesucht wird und inwieweit der jeweils eingesetzte OPAC den Vorgehensweisen und Bedürfnissen der Suchenden entgegenkommt und deren Suchverhalten entspricht.

Neben direkten Benutzerbefragungen, die langwierig, aufwendig, teilweise subjektiv und in der Auswertung häufig wenig ergiebig sein können, bieten sich die Protokolldaten der Katalogrecherchen (Log-File) zur Auswertung an. Voraussetzung ist dabei, dass die gespeicherten Daten mittels eines geeigneten Programms in eine lesbare und damit verwertbare Form gebracht werden.

Auf sehr umfangreiches Datenmaterial kann Franz Obermeier² in Eichstätt zurückgreifen, wenn er die Benutzung des SIS-OPACs der Universitätsbibliothek mittels selbstentwickelter cgi-Skripte untersucht, wobei er vorangegangene Berichte aus Münster, Bonn, Düsseldorf und Österreich in seine Darstellung einbezieht. Der Hamburger Versuch, die Log-Files des WebOPCs auszuwerten, hatte zunächst – wie in Eichstätt – das Ziel, die Nutzung der verbalen Sacherschließung zu untersuchen. Ein Teil der Ergebnisse wurde bei einem Workshop der Facharbeitsgruppe Sacherschließung des Verbundes vorgetragen³. Eine wesentlich erweiterte Auswertung wird im vorliegenden Beitrag vorgestellt.

Der Campus-Katalog und seine Register

Der Campus-Katalog ist der WebOPC des Lokalen Bibliothekssystems (LBS3) Hamburg. Er enthält die ca. 3 Millionen EDV-erfassten Katalognachweise der Staats- und Universitätsbibliothek Hamburg (SUB), der Institutsbibliotheken und der Hochschule für Wirtschaft und Politik (HWP).

Die zum Zeitpunkt der Einführung des LBS vom GBV angebotene Standard-Indexierung für den OPAC entsprach in einigen Punkten nicht den lokalen Erfordernissen. Die in Hamburg für den Campus-Katalog programmierten Register weichen daher von dieser Standardlösung ab.

Kurzform	Bezeichnung	Inhalt
TNT	Titel Namen Themen	alle Stichwörter aus Sachtitel, verbaler Sacherschließung und die Nachnamen von Personen
PER	Personennamen	Ansetzungs- und Verweisungsformen von Personennamen, stringindexiert
KOR	Körperschaften	alle Wörter aus Körperschaften
TAF	Titel(anfang)	Titelstring aus allen Sachtiteln
SWO	Schlagwörter	alle ein- und mehrteiligen Schlagwörter aus zentraler und lokaler verbaler Sacherschließung, stringindexiert
SYS	Systematik	alle Systemstellen aus zentraler und lokaler klassifikatorischer Sacherschließung und ihre verbalen Benennungen, stringindexiert
NUM	Nummern	ISBN, ISSN, Reportnummer u.ä.
SIG	Signaturen	Aufstellungs- und Grundsignaturen, stringindexiert

Daneben bietet der WebOPC die Möglichkeit, bei der kombinierten Suche auch sekundäre Suchaspekte einzugeben: das Erscheinungsjahr, den Materialcode (Buch, Zeitschrift, Tonträger etc), den Erscheinungsort, den Verlag und das 3-stellige Bibliothekssigel der Instituts- und Fachbereichsbibliotheken zur Einschränkung der Suche auf den Bestand einer einzelnen Bibliothek.

Abweichungen gegenüber den bis dahin verbreiteten LBS-Indexierungen gibt es vor allem in zwei Bereichen:

- Das TNT-Register (ein Vorgriff auf den in den weiterentwickelten Systemen jetzt üblichen Basic- oder All-Index) sollte es ermöglichen, die im Vorgängersystem des WebOPC, einem Telnet-OPAC, noch recht komplizierte kombinierte Suche zu vermeiden. Die Standardsuche nach Verfasser(nachnamen) und Titelstichwort konnte so in einem Register stattfinden.
- Das TAF-Register ermöglicht die gezielte Suche nach den in stichwortbasierten Registern schwer zu findenden Titeln vom Typ „Spiegel“, „Journal of mathematics“ oder „Handbuch der deutschen Geschichte“.

Die Suche nach dem zuletzt genannten Titel ergibt im TAF-Register des Campus-Katalogs die drei vorhandenen Ausgaben mit genau diesem Haupt- bzw Ansetzungssachtitel. Im TNT-Register erzielt man sechs Treffer, unter denen die bei der ersten Recherche gefundenen Titel nicht vorhanden sind.

Hintergrund für diese zunächst nicht nachvollziehbare Unstimmigkeit sind die Probleme des WebOPCs beim Umgang mit hohen Treffermengen, die hier detailliert beschrieben sind, weil sie bei der Analyse der Suchergebnisse später eine wichtige Rolle spielen:

- Bei der Kombination mehrerer Suchbegriffe werden die Zwischenergebnisse zu den einzelnen Begriffen nicht vollständig abgearbeitet, wenn diese mehr als 15 000 Treffer umfassen. Es erscheint der Warnhinweis: „Während der Suche war ein internes Set auf 15 000 Treffer beschränkt“. Recherchen wie die oben beschriebene nach dem „Handbuch der deutschen Geschichte“ erzielen also unvollständige Ergebnisse.
- Ergibt eine Recherche mehr als 1 000 Treffer, so werden diese nicht mehr angezeigt, es erscheint der Fehlerhinweis: „Treffermenge zu groß. Bitte schränken Sie Ihre Suchanfrage mit einem weiteren Begriff ein. Verwenden Sie keine Wörter, die besonders häufig vorkommen wie z.B. Geschichte, Einführung etc.“

Datengrundlage

Die von den verschiedenen Clients kommenden Anforderungen werden in einer Log-Datei fortlaufend aufgezeichnet. Unsere Auswertung bezieht sich auf einen Ausschnitt von 86 Stunden (3½ Tage) aus dieser Datei. Die knapp 450 000 Zeilen dieses Protokolls wurden als

2 Obermeier, Franz: Schlagwortsuche in einem lokalen Opac am Beispiel der Universitätsbibliothek Eichstätt. Benutzerforschung anhand von OPAC-Protokolldaten. In: Bibliotheksforum Bayern 27 (1999) S. 296-319.

3 Petra Blödorn-Meyer/Manfred Weichert: 3 Tage im Mai. Schlagwortsuche im OPAC der SUB Hamburg; Vortrag auf dem Workshop „Ist mehr drin? Defizite und Chancen bei der Nutzung von Sacherschließungsdaten“ am 8.11.2000 in Hannover.

einzelne Datensätze in eine Allegro-Datenbank importiert, die Suchanfragen und andere Aktivitäten (z.B. Blättern in Kurzlisten, Vollanzeigen, Bestellungen aus dem Magazinbestand der SUB) isoliert und indexiert. Die Registergestaltung in der Allegro-Datenbank ermöglicht u.a. auch die Rekonstruktion der zeitlichen Abfolge der Benutzeraktivitäten und der Reaktionen des Webserver innerhalb einer Katalogsitzung (Session).

Eine besondere Schwierigkeit ergibt sich aus der Tatsache, dass die Log-Daten keinen direkten Hinweis auf die Rechercheergebnisse enthalten, es fehlen insbesondere Daten über den Umfang der Treffermengen. Um dennoch Aussagen über Null-Treffer-Recherchen machen zu können, werden zwei indirekte Kriterien genutzt.

1. Der Client fordert offensichtlich in Abhängigkeit von dem Suchergebnis die für den Aufbau des Folgebildschirms notwendigen unterschiedlichen graphischen Elemente (GIF-Dateien) vom Server an. Anhand dieser Zugriffe auf GIFs ist zu erkennen, ob auf eine Suche der Index-Bildschirm, eine Kurzliste oder eine Vollanzeige folgt. Der Index-Bildschirm ist in der Regel ein Hinweis auf eine vorangegangene fehlgeschlagene Recherche.
2. Mit jeder Suchanfrage wird intern eine Variable mitgeführt, die als Nummerierung der bisherigen Ergebnismengen (Sets) erkennbar ist. Ändert sich diese Zahl in aufeinander folgenden Recherchen nicht, so ist dieses ein weiterer Hinweis auf eine ergebnislose Anfrage.

Das Protokoll verzeichnet aber auch einzelne Fälle, in denen nach erfolgloser Recherche kein Index-Bildschirm angefordert wird und Fälle, in denen trotz erfolgreicher Recherche die Set-Nummerierung nicht ansteigt. Die Nutzung des „Zurück“-Button des Browsers mag einer der Gründe für diese Uneinheitlichkeit sein.

Alle quantitativen Ergebnisse müssen vor dem Hintergrund dieser Ungereimtheiten mit Vorsicht betrachtet werden. In Teilbereichen wurden zur Verifizierung der Daten und für weitere Analysen teils umfangreiche Stichproben programmunterstützt nachrecherchiert. Insbesondere alle Aussagen über den konkreten Umfang von Treffermengen beruhen auf diesen Nachrecherchen.

Statistik

Die vom selben Client abgeschickten Anfragen zwischen zwei Starts des WebOPCs definieren eine Session. Insgesamt fanden im untersuchten Zeitraum 12 595 Sessions statt, von denen einige nicht über den Katalogstart hinaus kamen oder nach wenigen Schritten abbrachen. Hauptgrund dafür war die große Belastung des Datenbankserver, die ein sehr schnelles automatisches Abschalten einer inaktiven Verbindung nötig machte (Timeout nach 150 Sekunden).

Die Auswertung der IP-Adressen in den Sessions zeigt, dass der überwiegende Anteil der Benutzung aus der Universität Hamburg (50,4 %) und von den Katalog-PCs der Bibliothek (28,1 %) stammt. Immerhin 2 700 Sessions (Ausland: 66) kommen von außerhalb dieses IP-Bereichs, obwohl die Datenbank zum Zeitpunkt der Untersuchung wegen der Serverbelastung für diese Adressen zwischen 10.00 und 16.00 Uhr nicht erreichbar war.

Zwei Drittel aller Sessions finden in dieser Kernzeit statt. Weniger als 50 Katalogstarts pro Stunde sind nur in der Zeit zwischen Mitternacht und 8.00 Uhr zu verzeichnen. Nur 73 Benutzer starten die englischsprachige Version des Campus-Katalogs, erstaunlicherweise auch nur 8 der 66 Nutzer aus dem Ausland.

Insgesamt 1 909-mal wird anstelle des Gesamtkatalogs einer der 18 Teilbestände (Fachbereiche) aufgerufen.

Die Bestellmaske des Campus-Katalogs wird 7 164-mal angefordert, es ist allerdings nicht erkennbar, ob der Bestellvorgang auch in jedem Fall bis zum Ende durchgeführt wird. Die Statistik des Ausleihsystems verzeichnet für den Untersuchungszeitraum ca. 5 000 Bestellungen. Die folgende Tabelle zeigt die Verteilung der 51 067 Suchanfragen auf Register und Suchart.

	Einfache Suche	Kombinierte Suche	Index-Suche	Link-Suche	Gesamt	%
TNT	20 032	3 679	1 120	670	25 501	49,9
PER	10 174	696	402	275	11 547	22,6
SWO	3 989	721	922	736	6 368	12,5
TAF	3 303	225	460	*	3 988	7,8
SIG	1 947	30	69	*	2 046	4,0
SYS	210	29	20	164	423	0,8
NUM	386	11	21	*	418	0,8
KOR	261	20	14	2	297	0,6
JAH		362			362	0,7
ORT		35			35	
VLG		35			35	
MAT		43			43	
BIB		4			4	
	40 302	5 890	3 028	1 847	51 067	

* Keine Links vorhanden

Einfache Suchen und kombinierte Suchen⁴ sind diejenigen, bei denen der Benutzer die Anfragen in die entsprechenden Formulare des WebOPC einträgt. Index-Suchen entstehen überwiegend im Zusammenhang mit fehlgegangenen Suchanfragen. Es wird in der Regel der Registerabschnitt aufgeblättert, der der Eingabe alphabetisch benachbart ist und aus diesem kann ausgewählt werden. Index-Suchen können auch direkt aus den beiden Suchformularen heraus aufgerufen werden, wenn statt der voreingestellten Funktion „Suchen“ aus einem Popup-Menü die Funktion „Indexliste“ ausgewählt wird. Link-Suchen sind die durch das Anklicken eines gelinkten Elements der Volltitelanzeige verursachten Suchen. Die statistischen Daten zeigen eine eindeutige Präferenz für die voreingestellten Werte:

- 80 % aller Suchen erfolgen über die einfachen Suchmaske, die auf fast jedem Bildschirm des WebOPC angeboten wird.
- Die Hälfte aller Suchen finden im voreingestellten TNT-Register statt.

4 Die kombinierte Suche des WebOPCs entspricht in anderen OPAC-Systemen der einfachen formulargesteuerten Suche. Die in diesen OPACs angebotene kommandoorientierte Expertensuche ist zwar im WebOPC möglich, wird aber nach unserer Kenntnis in keinem der im GBV installierten OPACs erläutert und genutzt.

- Nur 127 der 3 028 Indexsuchen erfolgen auf Grund von Änderung der Voreinstellung des Funktionsmenüs.
- Die ebenfalls in den Suchmasken auswählbaren Möglichkeiten, die erzielten Ergebnismengen nachzubearbeiten, werden nur 297-mal genutzt (242-mal durch „Einschränken“, 55-mal durch „Erweitern“).

Alle Möglichkeiten außerhalb der Voreinstellungen sind auf den Hilfeseiten erläutert, diese werden aber nur sehr sporadisch genutzt (287 Aufrufe).

Statt die im Popup-Menü wohl eher versteckte Möglichkeit zur Einschränkung großer Treffermengen zu nutzen, zeigen die Benutzer unseres Katalogs eine außerordentliche Geduld beim Abarbeiten der Kurzlisten. Viele Listen werden bis über den 100. Treffer hinaus durchgeblättert (Maximum: 977 Treffer). Von den im Protokoll verzeichneten 33 448 Vollanzeigen sind 1 549 aus 100. oder späterer Position in den Ergebnismengen ausgewählt.

In den folgenden Abschnitten wird in unterschiedlicher Ausführlichkeit auf einzelne Aspekte der Nutzung der verschiedenen Sucharten und Register eingegangen.

Kombinierte Suche

Das Formular für die kombinierte Suche wird 3 212-mal genutzt. In 962 Suchen ist nur ein Feld dieser Maske ausgefüllt, fast immer im Anschluss an eine misslungene Suche mit mehreren Feldern. Der WebOPC reagiert in diesen Fällen mit dem erneuten Angebot einer Kombi-Suche.

Im folgenden werden nur die 2 250 Recherchen betrachtet, in denen zwei und mehr Felder betroffen sind.

In nur 73 Fällen (3,2 %) wird bei der Verbindung der Suchbegriffe das logische ‚ODER‘ eingesetzt. Mit mäßigem Erfolg: 29 Recherchen erzielen kein Ergebnis, weil die Treffermenge über 1 000 liegt oder formale Fehler bei der Kombination vorhanden sind, 8-mal wird kein Treffer erzielt, weil beide Teile der ODER-Frage nicht vorhanden sind. Bei den erfolgreichen 36 Suchen liegt die Treffermenge bei durchschnittlich 168 Titeln. Der logische Operator „NICHT“ wird nur einmal ausgewählt.

Die sekundären Suchaspekte werden mit Ausnahme des Erscheinungsjahrs wenig genutzt: Ort und Verlag jeweils 35-mal, Materialcode 43-mal und Bibliothekssigel zur Einschränkung auf den Bestand einer Institutsbibliothek nur 4-mal.

Bei den 362 Suchanfragen mit Erscheinungsjahr ist in 79 Fällen diese zusätzliche Angabe überflüssig, weil die restlichen Teile der Anfrage (z.B. Person und Titel) ausreichen, um ein eindeutiges oder doch überschaubares Ergebnis zu erzielen. In 50 Fällen wird offensichtlich nach dem Erscheinungsjahr eines Zeitschriftenbandes gesucht.

Bei den Kombinationen der primären Suchaspekte machen die TNT/TNT-Recherchen mit 1 195 (53 %) und PER/TNT mit 245 (11 %) zusammen fast 3/4 aller Fälle aus. Nahezu sämtliche Anfragen (nämlich alle, in denen bei PER nur der Familienname der Person eingegeben wurde) hätten auch durch eine einfache Suche im TNT-Register ersetzt werden können.

In 185 Recherchen werden zwei oder mehr Schlagwörter kombiniert, alle anderen Verbindungen kommen weniger als 100 mal vor.

Bedenklich stimmt der Blick auf die erzielten Treffermengen bei den kombinierten Suchanfragen. In fast der Hälfte aller Fälle (47,4 %) wird kein Treffer gefunden oder die Treffermenge liegt über 1 000 (letzteres überwiegend bei den „ODER“-Recherchen). Bei 12 % aller Recherchen wird die Ergebnisanzeige von dem Warnhinweis „Während der Suche war ein internes Set auf 15 000 Treffer beschränkt“ begleitet, die Vollständigkeit des Ergebnisses bleibt also ungewiss.

Suchen nach nicht im Campus-Katalog nachgewiesener Literatur und Schreibfehler sind für mehr als die Hälfte der Null-Treffer-Ergebnisse verantwortlich. Formale Fehler – falsche Trenn- und Trunkierungszeichen, Nichtinvertierung der Namen u.ä. – spielen bei 20 % eine Rolle und in 15 % der Fälle werden wenig Erfolg versprechende Kombinationen verwendet:

- PER/SWO, wobei wohl eine Kombination aus Personen- und Sachschlagwort gemeint ist,
- PER/PER, in denen Familien- und Vornamen auf zwei Eingabefelder verteilt werden,
- Kombination des SIG-Index mit irgendeinem anderen Suchaspekt.

Häufig reagieren die Benutzer, indem sie bei der nachfolgenden Suche ein ganzes Feld weglassen oder die Menge der Begriffe in einzelnen Feldern reduzieren. Auffällig ist, dass Schreibfehler bei der kombinierten Suche deutlich seltener korrigiert werden als bei den einfachen Suchen. Ein Grund für diese unterschiedliche Reaktion liegt sicher in der Tatsache, dass auf die erfolglose Recherche ein leeres Suchformular folgt und die fehlerhafte Eingabe nicht mehr auf dem Bildschirm erscheint.

LINK-Suchen

Die vorherrschende Farbe des Vollanzeige-Bildschirms im WebOPC ist blau. Viele Elemente der Beschreibung sind blau unterstrichen, es gibt also eine Fülle von Links, die intern zu einer Suchanfrage verarbeitet werden. Die Gesamtstatistik zeigt Anzahl und Verteilung der über solche Links erzeugten Anfragen. Die Log-Datei verzeichnet 38 448 Vollanzeigen. Aus weniger als 5 % dieser Vollanzeigen heraus wird eine Link-Suche angestoßen.

Die relative Häufigkeit von Link-Suchen im Verhältnis zu Eingaben des Benutzers in die Suchformulare ist besonders hoch im SYS- und im SWO-Register.

Suchen über Schlagwort-Links (immerhin 11,6 % aller SWO-Suchen) haben in fast allen Fällen Erfolg, wenn es sich um einen Normsatz-Link auf einen SWD-Begriff handelt. Erfolglose Anfragen entstehen durch die ebenfalls gelinkten Forms Schlagwörter wegen der Treffermengenbegrenzung des WebOPC. Aber auch eine erfolgreiche Suche ist kein voller Erfolg: Die Tatsache, dass der Anteil der mit SWD-Vokabular erschlossenen Titel noch immer nicht sehr hoch ist, enttäuscht die Hoffnung, über einen Link alles Vorhandene zu einem engen Thema zu finden, allzu häufig.

Schwer zu interpretierende Suchergebnisse erzielen die Links auf die ebenfalls angezeigten verbalen Sacherschließungselemente aus den Fremddaten der Library of Congress. In der Anzeige erscheint die komplette Schlagwortkette gelinkt, tatsächlich berücksichtigt die sich anschließende Suchanfrage nur den ersten Teil dieser Kette.

Die genauere Analyse der Nutzung von Links im Sachtitelbereich lässt Zweifel aufkommen, ob solche großflächige Verlinkung sinnvoll ist. Fast 50 % der Aufrufe führen ins Leere: mehr als 1 000 Treffer, Suche mit Stopwörtern, die ebenfalls gelinkt sind, und 1-Treffer-Ergebnisse, die schwer interpretierbar sind, weil sie zur Vollanzeige eben des Titels führen, aus dem heraus der Link aufgerufen wurde. Die Hälfte der erfolgreichen Recherchen ergibt mehr als 90 Treffer, und die Zweifel an der Zweckmäßigkeit verstärken sich, wenn man einen Blick auf eine Auswahl der aufgerufenen Titellinks wirft (in Klammern die Anzahl der Treffer): Bismarcks (282), concise (333), Rueckblick (506), chronische (547), discussion (559), differences (571), Bewegungen (576), 69 (603), Sonderheft (619).

Suche nach Personen

Von den 10 174 Recherchen nach Personennamen mit der einfachen Suchmaske scheitern knapp 15 %. Dass dieser Wert so deutlich unter dem der Eichstätter Untersuchung liegt, mag einen Hauptgrund in der Tatsache haben, dass Eingaben in diesem Feld automatisch trunziert werden, sodass die vielen Suchen mit erstem abgekürzten Vornamen erfolgreich waren (auch wenn das Ergebnis in manchen Fällen recht umfangreich und damit nicht so leicht überschaubar gewesen sein mag).

Um der Frage nach den Gründen des Scheiterns etwas genauer nachzugehen, haben wir annähernd die Hälfte der fehlgeschlagenen Suchen in diesem Register detailliert untersucht. Die Eingaben wurden nach Fehlerarten klassifiziert und in einem nachlaufenden Programm wurde geprüft, wie der Benutzer auf die erfolglose Recherche reagiert.

Fehlerarten:

- nicht vorhanden (N),
- formale Fehler (FF): Hierunter fallen die Eingabe von Sachbegriffen oder anderen Zeichenketten (z.B. Leserausweisnummern), die erkennbar keine Namen sind, Eingabe von Namen in natürlicher Folge Vorname Familienname, Weglassen des Kommas zwischen Familien- und Vornamensgruppe oder Ersatz durch Punkte und andere Zeichen und fehlerhafter Umgang mit Präpositionen,
- erkennbare Schreibfehler (SF),
- Recherchen sind erfolglos, weil der Name nicht im ausgewählten Teilbestand, wohl aber im Gesamtbestand vorhanden ist (TB),
- Recherche scheitert an der Treffermengenbegrenzung (TMB),
- Eingabe von 2 und mehr Initialen in der Vornamengruppe (INI).

Reaktionen des Benutzers:

- neue Suche (NEU),
- Korrektur des eingegebenen Suchbegriffs (KOR),
- Wiederholung der erfolglosen Anfrage (WH),
- aus der nach einer fehlgeschlagenen Recherche gezeigten Indexliste wird eine Eintragung ausgewählt (IND).

Die Tabelle zeigt die relative Häufigkeit der Fehlerarten und der unmittelbar folgenden Benutzerreaktionen (alle Angaben in %).

Suche nach nicht vorhandenen Namen, formale Fehler und Schreibfehler sind in annähernd 80 % der Fälle der

Fehlerart	%	Reaktion			
		NEU	KOR	WH	IND
N	29,7	69,5	12,8	12,4	5,3
FF	25,7	31,9	51,1	14,3	2,7
SF	24,1	11,0	67,5	5,2	16,2
TB	7,1	34,0	10,0	14,0	42,0
TMB	6,8	20,8	33,3	45,8	*
INI	6,6	29,8	19,1	10,6	40,4
Mittelwerte		38,6	36,2	13,6	11,6

* Es erfolgt kein Aufblättern des Registers

Grund für das Scheitern. Bei den Reaktionen machen Neueingabe und Korrektur 75 % aus.

Die Korrekturquote ist besonders hoch bei den Form- bzw. Schreibfehler. Das ist sicher damit zu erklären, dass die angezeigte Registerseite und die auf derselben Seite noch vorhandene Eingabe des Benutzers genügend Hinweise für eine Korrektur geben. Auch der relativ hohe Wert der Indexsuche bei Initialenfehlern (INI) ist verständlich, wenn man bedenkt, dass das Register sehr nah bei dem gesuchten Namen aufgeschlagen wird.

Nachdenklich stimmen die zwei vom jeweiligen Mittelwert am stärksten abweichenden Werte bei „Wiederholung“ und „Indexsuche“.

– In fast der Hälfte der wegen der Treffermengenbegrenzung erfolglosen Recherchen wird die vorangegangene Suche wiederholt. Der Hinweis „Treffermenge zu groß. Bitte schränken Sie Ihre Suchanfrage mit einem weiteren Begriff ein. Verwenden Sie keine Wörter, die besonders häufig vorkommen wie z.B. Geschichte, Einführung etc.“, obwohl zentral in roter Schrift auf dem Bildschirm angezeigt, wird offensichtlich nicht verstanden.

– Nachvollziehbar ist hingegen die überdurchschnittlich häufige Indexsuche nach „Teilbestandsfehler“. Der Einstieg in den Teilbestand eines Fachbereichs wird technisch realisiert durch das Vorbelegen eines sekundären Suchbegriffs mit dem Fachbereichskürzel. Das geschieht bereits bei der Auswahl eines Teilkatalogs. Beim Aufblättern des Registers wird allerdings immer das Gesamtregister des Campuskatalogs angezeigt.

Die erläuternden Texte dienen eher der Verwirrung. So erscheint die Überschrift: „Indexliste FB 05 – Sozialwissenschaften“ in Großschrift und fett. Darunter etwas kleiner der Hinweis „Nichts gefunden“ und – durch andere Hinweise davon getrennt – der Text: „Achtung: Hier wird auch im Teilbestand immer die Indexliste des Gesamtbestandes angezeigt.“ Es folgt der Registerauschnitt, der den gesuchten Namen enthält. Nicht verwunderlich, dass in 42 % der Fälle dieser Begriff angeklickt wird, mit der Folge, dass der gleiche Bildschirm erneut präsentiert wird.

Recherchen im SWO-Register

Der Campus-Katalog erlaubt die einfache Recherche nach Sachaspekten mit dem TNT-Register, in dem alle Bestandteile der verbalen Sacherschließung stichwortindexiert sind. TNT-Suchen mit Sachbegriffen sind tatsächlich häufig nicht als gezielte Titelstichwortrecherchen sondern als Themenrecherchen gemeint, dieses ist aus dem jeweiligen Kontext deutlich zu erkennen.

Dass dennoch 12.5 % aller Recherchen direkt im stringindexierten SWO-Register stattfinden, ist erstaunlich, wenn man die Eichstätter Ergebnisse zum Vergleich heranzieht. Dort enthalten auch nur knapp 15 % aller Recherchen eine SW-Eingabe, obwohl das die einzige Möglichkeit ist, gezielt nach Sachverhalten zu suchen. Die ebenfalls umfangreiches Datenmaterial auswertende Studie, die der Verband Österreichischer Bibliothekare (VÖB) veröffentlicht hat, zeigt allerdings einen deutlich höheren Wert⁵: Fast 30 % aller Recherchen enthalten Suchen mit Schlagwörtern.

Eine Zuordnung des im SWO-Register des Campus-Katalogs gesuchten Wortmaterials zu den Schlagwortkategorien des SWD zeigt folgende Verteilung:

- 51,0 % sind einfache Sachschlagwörter oder Abkürzungen
- 19,7 % umfassen komplexere Themenformulierungen
- 13,4 % beziehen einen geographischen Aspekt mit ein
- 7,0 % Suche nach Personen
- 5,6 % Anfragen mit fremdsprachigen Begriffen
- 2,3 % enthalten einen Zeitaspekt.

Das SWO-Register enthält ein- oder mehrteilige Schlagwörter aus der SWD, die Library of Congress subject headings und die Begriffe aus der lokalen verbalen Sacherschließung. Eine Betrachtung von Erfolg oder Misserfolg der einzelnen Recherchen im Campus-Katalog ist also nicht vergleichbar mit den Ergebnissen der erwähnten Untersuchungen aus Eichstätt und Österreich.

Wir sind daher an einer umfangreichen Stichprobe (1 162 Recherchen der Alphabetabschnitte A-D, K-N, U-W) der Frage nachgegangen, ob und wie weit sich die eingegebenen Suchbegriffe mit dem Wortschatz der SWD decken. Dazu wurde ein Programm erstellt, mit dessen Hilfe diese Anfragen an der im zentralen Verarbeitungssystem des GBV (CBS) vollständig enthaltenen SWD nachgestellt wurden.

Das beunruhigende Ergebnis: Es werden insgesamt nur 37,1 % aller Eingaben über die Ansetzungs- oder Verweisungsformen der SWD gefunden (Österreich: 74,7 %). Bei Ein-Wort-Recherchen ist die Erfolgsquote besser, sie beträgt 63,5 %. Der Anteil der besonders fehlerträchtigen Mehrwort-Recherchen ist allerdings in unserer Stichprobe deutlich höher (42,7 %) als bei der VÖB-Studie (19,8 %). Vieles spricht dafür, dass die Benutzer des Campus-Katalogs Ein-Wort-Recherchen nach Sachbegriffen eher im TNT-Register versuchen.

Bei der Analyse der Ursachen für die SWD-Null-Treffer ergibt sich folgendes Bild:

- unzulässige Kompositabildungen (30 %),
- Pluralformen (12,7 %),
- Schreibfehler (12,6 %),
- nicht im Wortschatz der SWD enthaltene Terminologie (12,0 %),
- fremdsprachliche Begriffe (7,7 %).

Die restlichen 25 % verteilen sich auf kleinere Fehlergruppen, z.B. unvollständige Eingabe von mehrteiligen Schlagwörtern (insbesondere bei Personennamen), falsche Reihenfolge in Adjektiv-Substantivverbindungen, fehlende Homonymenzusätze, Eingabe von Abkürzungen und formale Fehler bei Trenn-, Verbindungs- und Trunkierungszeichen.

Zwei Schlussfolgerungen ergeben sich aus der Analyse des Rechercheverhaltens einerseits und der Fehlerüberprüfung andererseits:

1. Das Nebeneinander von stichwortindexierter und stringindexierter Sacherschließung in zwei verschiedenen Registern scheint für die Benutzer schwer nachvollziehbar zu sein. Bei der Indexgestaltung des im GBV zur Zeit noch überwiegend genutzten LBS3-Systems war eine Mischung beider Indexierformen in einem Register nicht möglich. Mit der Übertragung der neuen Search&Order-Funktionalität des Verbundkatalogs auf die lokalen Systeme ergeben sich hier erweiterte Möglichkeiten, die in die Überlegungen zur Neugestaltung des WebOPCs einbezogen werden sollten.
2. Das Parsing der Benutzereingaben, also die Nachbearbeitung etwa durch Vereinheitlichung von Groß- und Kleinschreibung, Auflösung der Umlaute etc. ist noch zu schwach entwickelt. Gerade im Bereich der Schlagwortsuche, in der es um deutschsprachiges normiertes Vokabular geht, könnten die heute schon verfügbaren semantisch-syntaktischen Nachbearbeitungsprogramme (z.B. Komponenten des in Osnabrück und Bremen eingesetzten Systems OSIRIS) einen erheblich verbesserten Erfolg erzielen. Die bei der Analyse ermittelten Fehlerbereiche Komposita, Plural und die leichteren Schreibfehler, die einem solchen Korrekturverfahren zugänglich sind, machen mehr als die Hälfte der erfolglosen Anfragen aus.

Die anderen Register

Die Suchanfragen im SIG-, SYS-, NUM- und KOR-Register des Campus-Katalogs spielen in der Gesamtmenge eine geringe Rolle (jeweils weniger als 5 %). TNT- und TAF-Suchen sind wegen der großen Datenmenge und der Unzulänglichkeiten des Log-Files nur mit erheblichem Aufwand genauer zu analysieren. Daher folgen hier nur wenige Bemerkungen zu diesen Recherchen.

Die Untersuchung der Gesamtmenge der TNT-Suchen hinsichtlich der Anzahl der in die einfache Suchmaske eingegebenen Begriffe zeigt folgendes Ergebnis:

- 36,4 % der Suchanfragen enthalten nur einen Begriff,
- in 30,7 % werden zwei Begriffe eingegeben,
- drei und mehr Begriffe kommen in 32,9 % der Anfragen vor.

Erfolg oder Misserfolg überprüft an einer kleinen Stichprobe (518 Recherchen) ergibt:

- Null-Treffer bei 25,8 % der Suchanfragen⁶,
- mehr als 1 000 Treffer in 6,7 % der Fälle,
- durchschnittliche Trefferanzahl bei den 67,5 % erfolgreichen Recherchen ist 89 Titel.

33 % der Null-Treffer werden durch die Eingabe von Vornamen oder Vornamensbestandteilen zusätzlich zu einem Familiennamen verursacht. Vornamen sind im TNT-Register nicht indexiert. Die zunächst gewählte missverständliche Bezeichnung in den Eingabeformularen („Titel Namen Themen“) wurde auf Grund dieser Beobachtung in „Titel Nachnamen Themen“ geändert.

5 Online verfügbar unter: <<http://info.uibk.ac.at/sci-org/voeb/kof-sesw.html>> (Abrufdatum 4.7.2001).

6 Das ist deutlich weniger als der in der Eichstätter Untersuchung (vgl. Anm. 2) angegebenen Wert für die Titelstichwortrecherche (36,33 %). Allerdings ist zu beachten, dass das TNT-Register fast ein Basic-Index ist, also erheblich mehr Eintragungen enthält.

Die zu erwartende Abhängigkeit der Null-Treffer-Quote von der Anzahl der eingegebenen Suchbegriffe wird durch die Auswertung der Stichprobe bestätigt: Sie steigt von 3,8 % bei einem Suchbegriff über 40 % bei zwei Wörtern auf über 63 % bei drei und mehr eingegebenen Wörtern.

Suchen im Titelstringregister TAF scheitern in 1/3 der Fälle (1 051 von 3 303). Zwei wesentliche Ursachen dafür sind:

- Artikel am Anfang eines Sachtitels werden 307-mal mit eingegeben. Der nach einer solchen ergebnislosen Anfrage aufgeblätterte Index-Bildschirm ist wenig hilfreich. Das Register enthält viele mit einem Artikel beginnenden Einträge, die überwiegend aus den Altdaten des SUB stammen. In dem Vorläufersystem wurde bei der Erfassung von Bandtiteln auf eine Kennzeichnung der einleitenden Artikel verzichtet.
- Häufig (110-mal) wird nach Titeln der Form „Journal of ...“, „Annals of ...“ gesucht. Die völlig unverständlichen Ergebnisse einer solchen Suche hängen mit der holländischen Herkunft der Pica-Systeme zusammen. Das holländische „of“, deutsch „oder“, wird als logischer Operator „ODER“ interpretiert, sofern es nicht in Hochkommata oder Anführungszeichen eingegeben wird. Da diese Zeichen in den meisten Fällen fehlen, ergibt z.B. die Anfrage *journal of algebra* eine Treffermenge, die alle Titel enthält, die entweder „Journal“ oder „Algebra“ heißen.

Es sind also unsaubere Register oder der kaum verständlich zu machende Umgang mit „reservierten“ Wörtern, die fast die Hälfte der ergebnislosen TAF-Recherchen verursachen.

Schlussfolgerungen und Ausblick

Die Untersuchung des Benutzerverhaltens an unserem Katalog gibt trotz der relativ schmalen Datenbasis Anlass zu einigen Bemerkungen und Anregungen.

- Die Daten zur Analyse des Nutzerverhaltens und der Reaktionen des WebOPCs müssen in einer Form zur Verfügung gestellt werden, die eine Auswertung erleichtert.
- Schwerpunkt der bisherigen OPAC-Studien waren statistische Auswertungen der einzelnen Suchanfragen und Betrachtungen über deren Erfolg oder Misserfolg. Zukünftige Untersuchungen sollten systematischer auf den gesamten Kontext einer Recherche eingehen. Die Komplexität und der Umfang unserer Kataloge ist so groß, dass ein Erfolg im ersten Rechercheversuch selten zu erwarten ist.
- Aus den Reaktionen des Benutzers auf Misserfolg (zu viele Treffer, zu wenige oder keine Treffer) lernen wir, ob das System in solchen Situationen vernünftige Anleitung zur Verbesserung der Anfrage gibt. Die beobachtete unterschiedliche Korrekturquote bei einfacher und kombinierter Suche hängt sicher mit den sehr verschiedenen Reaktionen des WebOPCs in diesen beiden Situationen zusammen.
- Auch die Präsentationsformen haben Einfluss auf das Rechercheverhalten. So liegt z.B. die Vermutung nahe, dass die bei Personenrecherchen besonders häufig vorgefundene „natürliche“ Folge von Vor- und Familiennamen damit zusammenhängt, dass der WebOPC in der Vollanzeige die Namen von Verfassern und sonstigen beteiligten Personen in eben die-

ser Form darstellt. Diese Anzeigeform ist in Bibliothekskatalogen unüblich und wird auch von den anderen deutschen Pica-Systemen nicht verwendet.

- Der alte Streit um index- oder maskengesteuerte Suche als primären Einstieg in elektronische Kataloge ist zugunsten der letzteren entschieden. Dennoch bieten fast alle modernen Systeme die Indexsuche zusätzlich an oder leiten den Benutzer bei erfolglosen Suchen in die Register. Die Eintragungen in diesen Registern, besonders in stringindexierten und den auf normiertem Vokabular beruhenden (Personen, Schlagwörter, Klassifikationen), sollen helfen, die Suchstrategie zu verbessern und müssen daher mit besonderer Sorgfalt von allen fehlerhaften oder irritierenden Eintragungen frei gehalten werden.
- Bei Suchen mit dem normierten Vokabular der Schlagwortnormdatei bedarf der Katalogbenutzer besonderer Unterstützung. Diese kann in der bereits erwähnten Nachbearbeitung der Suchanfrage mittels linguistischer Programme bestehen. Als Alternative bietet sich ein Verfahren an, das die meisten WebOPCs für die Basisklassifikation bereits realisiert haben: Die vollständige Normdatei als Einstiegsinstrument der Recherche, versehen mit den aus systematischer Gliederung und Thesaurusstruktur erzeugbaren Navigationshilfen.
- Die neue Suchoberfläche Search&Order (S&O) des GBV-Gesamtkatalogs wird in naher Zukunft auch in allen lokalen Systemen zur Verfügung stehen. Ein Teil der geschilderten Probleme, insbesondere im Hinblick auf die Treffermengenbegrenzung des WebOPCs, wird damit überwunden sein. Allerdings werden der Basic-Index und der im Katalog der Staatsbibliothek zu Berlin bereits realisierte Themenindex in großen Lokalsystemen zu sehr umfangreichen Ergebnismengen führen. Unsere Beobachtungen beim Umgang mit hohen Trefferzahlen lassen erkennen, dass viele Benutzer nicht in der Lage sind, die vorhandenen Einschränkungsmöglichkeiten gezielt zu verwenden. Die beiden Hinweise, die S&O hier anbietet – eine Statistik der Trefferzahl zu den einzelnen eingegebenen Suchbegriffen und, bei über 1 000 Treffern, die Empfehlung zur Umsortierung nach einer schwer zu durchschauenden Relevanz – können nur erste Schritte sein auf dem Weg zu einer gezielteren, dem jeweiligen Kontext angepassten Hilfestellung.
- Nicht nur im Hinblick auf große Treffermengen muss jede Redundanz in den Ergebnislisten vermieden werden. In den GBV-Datenbanken ist zu der allseits bekannten Redundanz durch Dubletten im Zuge der Strukturänderungen bei mehrbändigen Werken noch eine zusätzliche Redundanzquelle entstanden. Durch die maschinelle Ergänzung der Bandsätze um Personen- und teilweise auch Titeldaten aus der Gesamtaufnahme entstehen bei vielen Recherchen Trefferlisten, die eben diese Bände zusätzlich zu den Gesamtaufnahmen enthalten⁷.

7 Die Suche nach Luthers Werken im KVK erzielt im Durchschnitt der anderen Deutschen Bibliotheksverbände 46 Titel, der GBV liefert 229 Treffer. Darin enthalten sind 128 Bände aus mehrbändigen Werken, die ebenfalls in der Liste sind. Reduziert man die verbliebene Menge von 101 Treffern noch um die Dubletten und fehlerhaften Altdaten, so bleiben schließlich 42 unterschiedliche Ausgaben.

- Für die direkte Benutzerschulung können aus den in der Untersuchung gemachten Beobachtungen Konsequenzen gezogen werden, indem auf die immer wiederkehrenden Fehlerquellen besonders hingewiesen wird. Allerdings machen wir die Erfahrung, dass unsere wöchentlich angebotene Schulung in Kleingruppen (bis zu 6 Personen) immer weniger genutzt wird.

Viele der aufgezeigten Probleme sind keine Besonderheiten der Pica-Systeme. Wir haben einige der in unserem Katalog vorgefundenen Recherchen in anderen Verbund- und Einzelkatalogen nachgestellt, selten mit besseren Ergebnissen oder deutlicheren Hilfestellungen für den Benutzer. Suchmöglichkeiten und Präsentationsformen sind uneinheitlich, was unterstreicht, dass auch unter den heutigen OPAC-Systemen noch nicht die optimale Form gefunden ist.

Bibliothekare und Systementwickler gemeinsam sind weiterhin aufgefordert, die mit beträchtlichem wirtschaftlichen Aufwand erfassten Beschreibungs- und Erschließungselemente in einer Weise suchbar zu machen und zu präsentieren, die den Bedürfnissen der Benutzer nach schnellem Auffinden eines nach Titel und/oder Verfasser bekannten Werkes einerseits und der möglichst umfassenden Auskunft über vorhandenes Material zu einem Sachverhalt andererseits so weit wie möglich entgegenkommen.

Anhang

Aufbereitete Darstellung einer Katalogsitzung.

Nach mehreren Ansätzen hat diese mitternächtliche Recherche Erfolg:

```

USER=XXX.dip0.t-ipconnect.de
00:23:38 Gesamtbestand Session-Nr. 3093
00:25:43 PER: eva jaeggi                               Index (0 Treffer)
00:27:14 TNT: zu heilen die zerstoessenen herzen      [SET:0]
00:27:35 TNT: zu heilen                               [SET:1]
00:28:25 TNT: zu heilen die                           [SET:1]
00:29:22 TNT: gibt es auch wahnsin                   [SET:1]
00:29:32 TNT: gibt es auch wahnsin                   [SET:1]
00:29:42 TNT: gibt es auch                           [SET:2]
00:30:13 S2+3                                         Vollanzeige
00:30:26 LINK PER: Jaeggi,Eva                        [SET:3]
00:31:48 S3+2                                         Vollanzeige
00:32:16 BESTELLEN epn 295600225+mon

```

Anschrift des Autors:

Manfred Weichert
Staats- und Universitätsbibliothek
Carl von Ossietzki
Von-Melle-Park 3
D-20146 Hamburg