

Der Einsatz von Suchmaschinentechnologie in Bibliothekskatalogen zeigt, dass durch optimierte Indexierungsverfahren Verbesserungen der Informationsversorgung für Bibliotheksbenutzer erreicht werden. Im Dezember 2006 wurde von OCLC PICA die Suchmaschine FAST-Instream in das lokale Bibliotheksportal „SISIS-InfoGuide“ integriert. In diesem Artikel werden die Vorteile dieser Erweiterung anhand der Erfahrungen aus den Piloteinsätzen an der Bayerischen Staatsbibliothek und der Universitätsbibliothek Bayreuth vorgestellt. Zusammenfassend kann festgestellt werden, dass sich die Investition des Freistaates Bayern in eine Landeslizenz für den Einsatz von FAST in den bayerischen Universitätsbibliotheken und der Bayerischen Staatsbibliothek in deutlichen Vorteilen für die Benutzer auszahlt.

Am offensichtlichsten ist die Auswirkung der Einbindung der Suchmaschine bei den Antwortzeiten auf Suchanfragen. Insbesondere bei der Suche nach häufig in den Bibliotheksdaten vorkommenden Begriffen, z. B. Deutschland oder Geschichte, werden selbst sehr große Treffermengen in Bruchteilen von Sekunden geliefert. Das Scheitern von Anfragen wegen erreichter Zeitlimits gehört damit der Vergangenheit an. Aus dieser Tatsache ergibt sich auch, dass Benutzer nicht mehr spezifisch suchen müssen, da nun auch große Treffermengen in kurzer Zeit präsentiert werden.

Das alleine wäre kein Vorteil, wenn nicht auch Methoden zur Verfügung gestellt würden, mit diesen großen Treffermengen umzugehen. Auch hier liefert die Suchmaschinentechnologie ad-

äquate Hilfsmittel, insbesondere die so genannten „Drill-Downs“, ein Verfahren, um die großen Treffermengen schrittweise durch gezielte Sucheinschränkungen zu begrenzen. Bestimmte, vordefinierte Felder werden beim Aufbau der Trefferlisten analysiert und die Feldinhalte nach Häufigkeit sortiert als Sucheinschränkung angeboten. Die Eingrenzung auf eine Sprache, einen Erscheinungszeitraum oder ein bestimmtes Schlagwort ist so mit einem einzigen Mausklick möglich. Die daraus neu entstehende Treffermenge wird wiederum dieser Analyse unterworfen. Mit wenigen Mausklicks gelangt der Benutzer so von einem oftmals unüberschaubaren Ergebnis zu einer begrenzten Menge an Treffern, die sinnvoll bearbeitet werden kann.

Eine gezielt eingegrenzte Suche war vor der Einführung der Drill-Downs durch die differenzierte Verwendung der unterschiedlichen Suchkategorien zwar auch schon möglich, jedoch erlaubt es die neue Technik, in den Sucheinschränkungen zu „browsen“. Der Benutzer muss also nicht schon vorweg wissen, welche Einschränkungen, z. B. welche Schlagworte, im Bereich seines Themas tatsächlich einschlägig sind, sondern kann sie in Bezug auf seine Treffermenge aus einer Liste herausuchen. Auch das Blättern oder die Suche im Schlagwortindex ist mit den Drill-Downs nicht vergleichbar. Denn es wird bei der Verwendung des Schlagwortindex immer der vollständige und damit sehr mächtige Index angezeigt, wogegen bei den Drill-Downs nur die Begriffe angezeigt werden, die in den Treffern tatsächlich vorkommen.

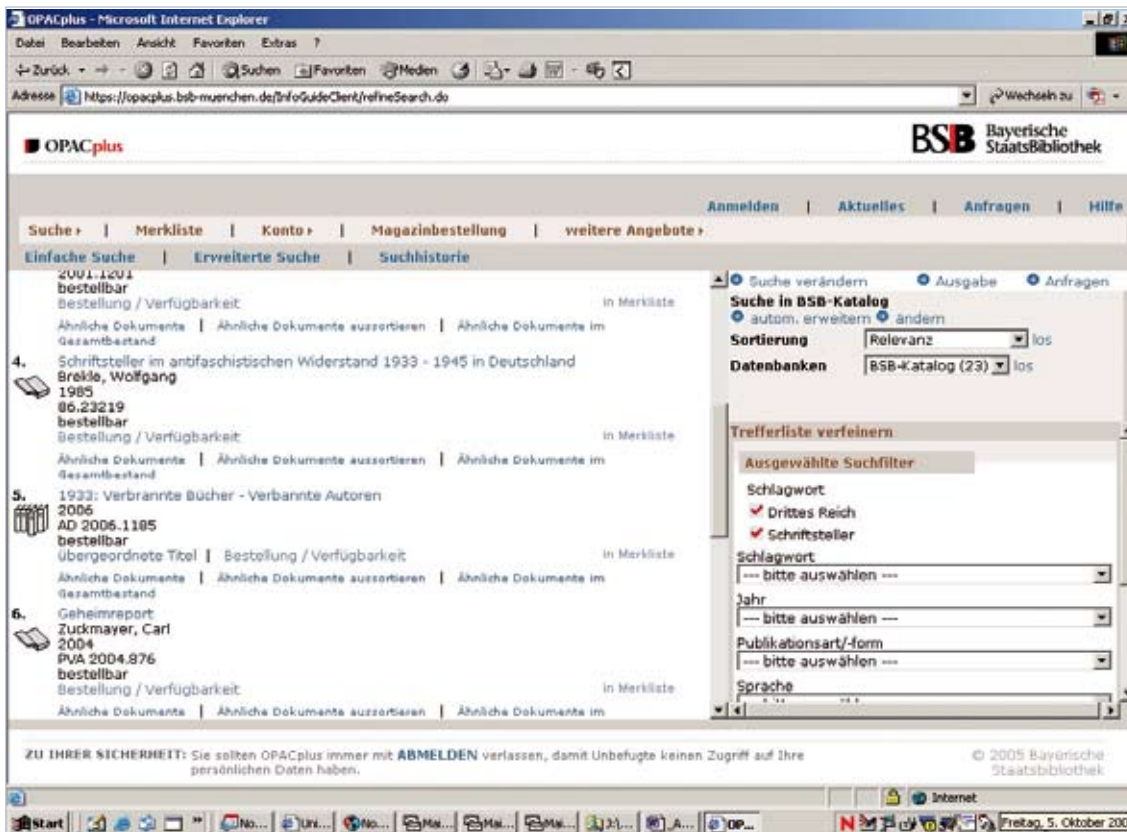
Folgendes Beispiel soll die neuen Möglichkeiten verdeutlichen: Die einfache Suche („Google-like“-Suche über alle Kategorien) nach den Begriffen „Deutschland“ und „Geschichte“ im Katalog der Bayerischen Staatsbibliothek liefert ca. 75.000

Treffer. Dieser Sucheinstieg alleine wäre vor der Einführung von FAST nicht möglich gewesen, aber auch die 75.000 Treffer hätte ein Benutzer nicht vernünftig nutzen können. Durch die schrittweise Eingrenzung auf die Schlagworte „Drittes Reich“ und „Schriftsteller“ kommt der Benutzer sehr schnell zu 23 relevanten Treffern, die mit der Rolle der Schrift-

Des OPACs neue Kleider – vom Benutzerkatalog zur Suchmaschine

Neue Möglichkeiten für Bibliothekskataloge durch den Einsatz von Suchmaschinentechnologie

Von Berthold Gillitzer und Andreas Weber



Der OPAC der BSB: Mit der neuen Funktion der Suchfilter können auch große Treffermengen effektiv auf die relevanten Titel begrenzt werden.

steller im Dritten Reich zu tun haben, ohne dass diese Begriffe so im Titel eines Werkes zu finden sein müssen, wie z. B. „1933: Verbrannte Bücher – Verbannte Autoren“ usw. Ohne eine genaue Kenntnis der Sacherschließung an den Bibliotheken wäre eine solche thematische Suche bisher kaum möglich gewesen. Damit wird die

Neue Methoden für den Umgang mit großen Treffermengen: Suchfilter und Relevance-Ranking

aufwändige Beschlagwortung der Titel durch die Bibliotheken, die den meisten OPAC-Benutzern bisher nur schwer zugänglich war, intuitiv nutzbar.

Eine zweite Strategie, um auch bei großen Treffermengen noch zu brauchbaren Ergebnissen zu kommen, besteht im so genannten Relevance-Ranking. Standardmäßig werden damit die Treffer nicht mehr nach dem Alphabet oder den Erscheinungsjahren sortiert, sondern es wird eine von den Internetsuchmaschinen bekannte Sortierung nach der Relevanz der Treffer angeboten. Die Treffer, die wahrscheinlich dem beabsichtigten Suchergebnis am besten entsprechen, werden

dabei zuerst angezeigt. Um dies zu erreichen, werden zwei Faktoren herangezogen, nämlich das Erscheinungsjahr bzw. die Aktualität der Katalogaufnahme, und textstatistische Faktoren, wie das Vorkommen der Suchbegriffe in bestimmten Feldern und die Häufigkeit der Suchbegriffe in den Titelaufnahmen. Diese unterschiedlichen Faktoren können wiederum untereinander gewichtet werden.

Obwohl mit dieser Methode schon in vielen Fällen eine verbesserte Sortierung der Suchergebnisse zustande kommt, sind die zur Anwendung gebrachten textstatistischen Relevanzverfahren, die aus der Umgebung der Volltextindizes stammen, nicht problemlos auf den Index von Katalogdaten zu übertragen. Kriterien wie die Häufung von Suchbegriffen in einer Titelaufnahme oder auch die Frage, wie nah die verschiedenen Vorkommen eines Suchbegriffs innerhalb einer Katalogkategorie zusammenstehen, führen oftmals zu keinen wirklich sinnvollen Sortierungen. Die Aktualität eines Titels

alleine könnte dagegen sehr einfach durch eine nachträgliche Sortierung der Treffermenge nach dem Erscheinungsjahr zur Geltung kommen.

Die Situation wird sich aber grundlegend ändern, sobald mehr echte elektronische Volltexte oder Daten aus den Kataloganreicherungen, wie Klappentexte, Inhaltsverzeichnisse usw. in den FAST-Index einbezogen werden. Hier können die textstatistischen Verfahren ohne Einschrän-

das zu indexierende Dokument errechnet werden. Aufgrund des Vergleichs dieser Vektoren können zu einem bestimmten Treffer diejenigen Titel ermittelt werden, deren Dokumentvektoren diesem Treffer am ähnlichsten sind. Diese Auswahl kann auf die aktuelle Treffermenge oder auch den kompletten Bestand angewendet werden. Mit dieser Methode können aus einer Treffermenge auch Treffer entfernt werden, die einem ausgewählten Treffer nahekommen. Obwohl es bei diesem Verfahren auch zu Ergebnissen kommt, die nicht sofort nachvollziehbar sind, ist sie als erste Eingrenzung einer großen Treffermenge, um z. B. einen verbesserten Überblick über Literatur zu einem Thema zu gewinnen, durchaus praktikabel.

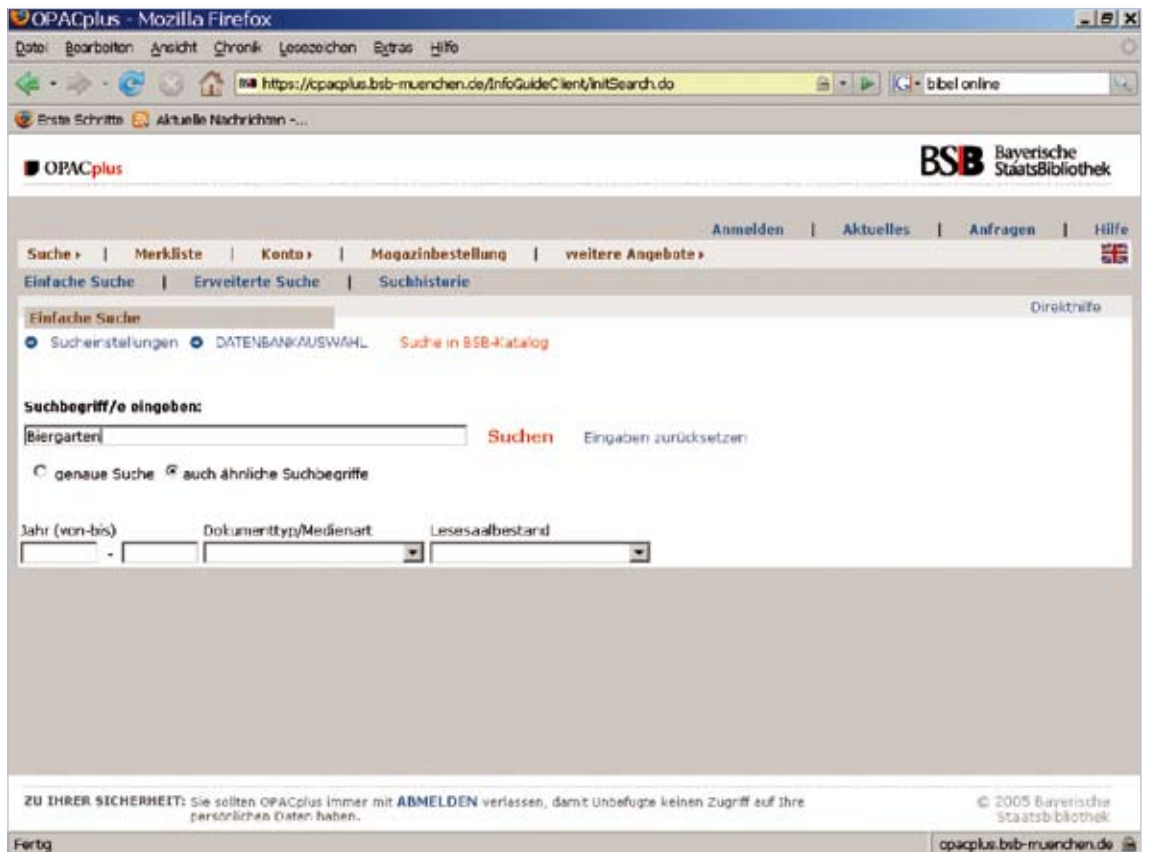
Eine zusätzliche Unterstützung der Recherche bietet die auf Wörterbüchern basierende linguistische Komponente der Suchmaschine. Die Begriffe können beim Indexieren der Daten lemmatisiert, also auf eine Grundform zurückgeführt werden. Eingegebene Suchbegriffe werden in der gleichen Art behandelt, so dass auch Treffer ermittelt werden, wenn der Begriff nicht exakt eingegeben wurde. Mit dem Suchbegriff „Biergarten“ werden beispielsweise auch die „Biergärten“ gefunden und somit die „Erfolgsquote“ gesteigert. Vor der

Linguistische Methoden und der Aufbau eigener Suchstrategien ermöglichen eine intuitivere Recherche

kungen zur Anwendung gebracht werden. Auch ist es denkbar, dass zukünftig nutzungsstatistische Daten wie die Ausleihhäufigkeit, die Anzahl der Volltitelanzeigen für ein bestimmtes Dokument oder sogar Verweisungsstrukturen zwischen Dokumenten über Zitate in Volltexten in die Relevanzberechnung Eingang finden. Noch steht das Relevance-Ranking im Bibliothekskatalog am Anfang, doch die jetzt schon vorhandenen Erfolge lassen auf große Verbesserungen für den Benutzer in diesem Bereich hoffen.

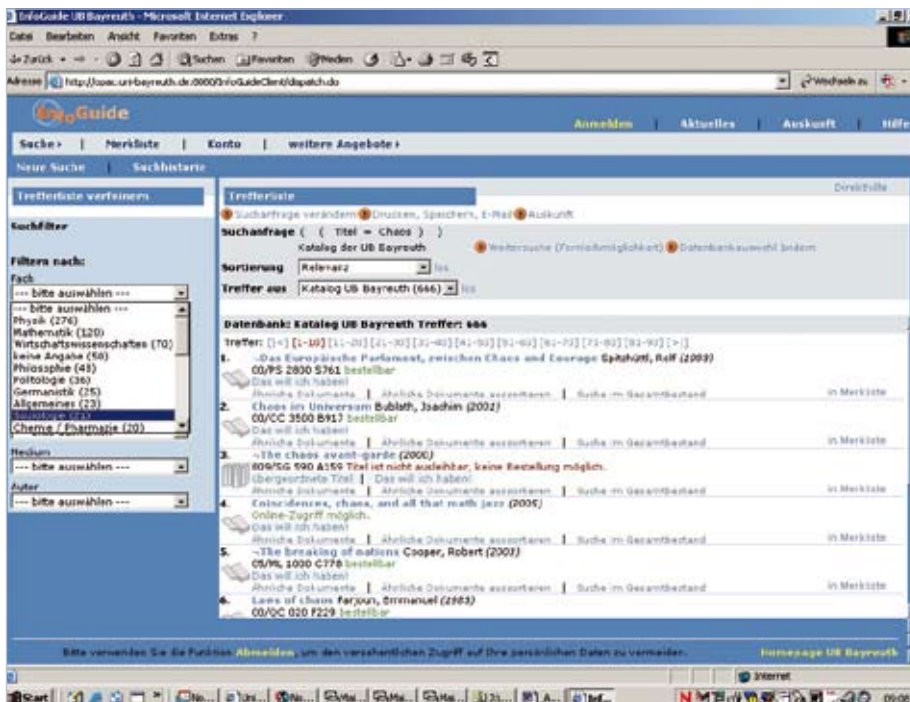
Weiterhin ermöglicht die FAST-Technologie eine Ähnlichkeitssuche, die auf Dokumentvektoren aufbaut, die an Hand statistischer Verfahren für

Die „Lemmatisierung“ ermöglicht eine Ähnlichkeitssuche: Mit dem Suchbegriff „Biergarten“ werden auch „Biergärten“ gefunden.



Einführung von FAST wären dafür sehr komplexe Rechercheeingaben oder vermutlich in den meisten Fällen mehrere Suchen notwendig gewesen. Mit Hilfe der Linguistikfunktion werden zusätzlich eventuell vorhandene Rechtschreibfehler korrigiert oder ggf. alternative Begriffe vorgeschlagen („Meinten Sie?“).

Ein wesentlicher Vorteil des Einsatzes der FAST-Suchmaschine besteht darin, dass ein zusätzlicher Index zur Verfügung steht, der von den internen Bibliotheksanwendungen unabhängig ist. Die für die bibliotheksinternen Abläufe optimiert gespeicherten Daten können so vor der Indexierung durch die Suchmaschine verändert werden, ohne dass interne Arbeitsvorgänge davon betroffen werden. Auf diese Weise lassen sich viele, für den Benutzer nicht transparente, fachspezifische Daten aussondern oder transparenter gestalten. Das einfachste Beispiel dafür ist das Filtern von Sonderzeichen, z. B. im Erscheinungsjahr. Weiterhin können aus verschiedenen kodiert gespeicherten Kategorien neue, mit sprechenden Bezeichnungen versehene Kategorien gebildet werden, z. B. für die Erscheinungsform eines Werkes. Die Bildung eines zusätzlichen Feldes „Fach“ aus der in der Bibliothek verwendeten Klassifikation hat sich an der Universitätsbibliothek Bayreuth als effektiver Filter für die Selektion relevanter Treffer bei der Suche nach unspezifischen Suchbegriffen, wie z. B. Chaos, herausgestellt. Die Möglichkeit, nach Schlagwortketten, also der Kombination von Schlagworten, einzugrenzen, führt den Benutzer



sehr einfach zu spezifischen Treffern und ist auch bei kleineren Treffermengen wirkungsvoll.

Die Einbindung der FAST-Suchmaschine in den lokalen Bibliothekskatalog bietet nicht nur eine wesentliche Verbesserung der Suchgeschwindigkeit, sondern sie eröffnet dem Benutzer durch die vielfältigen Methoden, die Treffer während der Anzeige zu analysieren, auch ganz neue Wege im Umgang mit Treffermengen. Eine Kehrseite dieser Errungenschaften liegt jedoch darin, dass dem Benutzer nun in der Ergebnisanzeige zusätzliche Möglichkeiten angeboten werden müssen, ohne dass dadurch die bisher bestehenden Optionen verloren gehen dürfen. Sowohl die Bayerische Staatsbibliothek als auch die Universitätsbibliothek Bayreuth mussten deshalb Änderungen an der Präsentation der Ergebnisse vornehmen, um den Benutzern übersichtliche Ergebnisse zu bieten.

Die Firma OCLC arbeitet daher derzeit an einer vollständigen Neugestaltung der Oberfläche der Anwendung, so dass die neuen Möglichkeiten zusammen mit den bislang gewohnten Funktionen ergonomisch und vor allem barrierefrei präsentiert werden. Die Wandlung des OPAC zur modernen Suchmaschine hat erst begonnen, aber die Anfänge sind verheißungsvoll und lassen für die nächsten Weiterentwicklungen noch viel erwarten.

Der verständliche Suchfilter „Fach“ im InfoGuide der UB Bayreuth wurde erst durch den Aufbau einer neuen Indekskategorie in der Suchmaschine möglich.

DIE AUTOREN



Dr. Berthold Gillitzer ist Leiter des Referats Dokumentlieferung und IT-Anwendungen in der Benutzungsabteilung der Bayerischen Staatsbibliothek. **Dr. Andreas Weber** ist Leiter der Abteilung Digitale Informationsversorgung und elektronisches Publizieren am Rechenzentrum der Universität Bayreuth.