

Datenstruktur und Suchprinzip des Journal.Quick.Finders

GERHARD SCHMITZ-VELTIN

Der elektronische Zeitschriften-Katalog J.Q.F weist einige erstaunliche Eigenschaften auf. Da ist zunächst die völlige Loslösung der Suchbegriffe von ansonsten üblichen 'Feldern', desweiteren die Suchmöglichkeit nach beliebigen (!) Titelabkürzungen und dann ein umso schnelleres Finden, je kürzer die Suchanfrage ist. Seine Datenstruktur während des Aufbaus der Indizes und seine Methode, wie er Suchanfragen analysiert und die Wörter in den Indizes findet, sei kurz erläutert.

Aufbau des J.Q.F

Die ursprünglichen Daten stammen aus der Zeitschriften-Datenbank des Deutschen Bibliotheks-Instituts. Sie werden alle ca. 3 Monate komplett herunter geladen. Daher beginnen wir mit einem Beispiel einer (leicht veränderten) Titel-Aufnahme aus dieser Datenbank (das 1. Zeichen hat in der Datei ZDB-ORI.DAT die Position z.B. 12345678) (Abb. 1)

Diese Einträge werden nach Signaturen sortiert und in der Zwischen-Datei ZS-ZUS.TXT abgelegt (Abb. 2)

Aus dieser Datei wird die Datei JQF-

Z.DAT mit den kompakten Titelaufnahmen aufgebaut; außerdem alle Index-Dateien, diese in mehreren Schritten:

Abb.1:

1.00/010930

ND: 280533

CO: CJMAA

JT: The Canadian journal of
mathematics

PT: Journal canadien de
mathematiques

PP: Toronto, Ont.

PU: Univ. Pr.

PD: 1.1949 -

SS: 0008-414x

BAW: 352 1.1949 - (48.1998 - Ab-
best.2.3.97) : SIGN.: mat 2/c 16

Einträge o.g. Zeitschrift in der unsortierten Index-Datei JQF-IUS.DAT (Abb. 3)

Die ca. 240.000 Index-Einträge aller ca. 22.500 Zeitschriften werden alphabetisch sortiert in die Zwischen-Datei JQF-IS.DAT geschrieben. Das Sortieren solch großer Dateien erfolgt mit dem sort-Befehl von Windows 95. Die anschließenden Maßnahmen dienen dazu, mit

chen in der Datei JQF-3A.DAT im Arbeitsspeicher gesucht. Dabei macht das Programm von der sehr schnellen Array-Scan-Funktion von PowerBasic 3.2 Gebrauch. In der Datei JQF-3B.DAT (auf der Festplatte, um Arbeitsspeicher zu schonen) wird dann die Startstelle für die Suche im großen Index JQF-I.DAT ermittelt. Aus dem großen Index liest der J.Q.F dann die Positionen in

Zeitschrift sucht, wenn seine Eingabe ein Stoppwort enthält. Dann ist auch die Reihenfolge der Wörter wichtig. Am Vorliegen eines Stoppworts erkennt der J.Q.F, daß er zunächst nach den Initialen suchen muß. Auf diese Weise wird auch vermieden, einzelne Zeichen mit einer riesigen Trefferzahl zu suchen. Dieselbe Unterstellung, nämlich daß der Benutzer eine bestimmte Zeitschrift sucht, geschieht auch, wenn die Suchfrage einen oder mehrere Einzel-Buchstaben enthält:

Abb. 2:

123456780008-414x	<--Position und ISSN>
The Canadian journal of mathematics. Journal canadien de mathematiques.	
Toronto, Ont: Univ. Pr. 02413-6712*#ab	<--2.ISSN und evt. Abrufzeichen>
+mat 2/c16 1.1949 - (48.1998 - Abbestellt 2.3.97)	<--+Zeichen, falls ff-ZS>

Hilfe weiterer Index-Dateien die Voraussetzung zu schaffen, daß auch komplexe Suchen des J.Q.F schnell zum Ergebnis führen.

Gleichzeitig mit der Bildung zweier neuer Index-Dateien werden die ersten drei Zeichen aller Einträge der Datei JQF-IS.DAT abgeschnitten und die neue Datei JQF-I.DAT dadurch deutlich kompakter. Die abgeschnittenen ersten 3 Zeichen werden in einer neuen Datei mit Namen JQF-3A.DAT gespeichert. Sie enthält nur die reine Abfolge dieser 3 Startzeichen und die auch nur einmal. Diese Datei wird beim Start des J.Q.F eingelesen. Die weitere Datei JQF-3B.DAT enthält - in gleicher Reihenfolge wie JQF-3A.DAT - nur die Positionen des ersten Auftretens dieser 3 Zeichen in der Datei JQF-I.DAT.

Zum 'Betrieb' des J.Q.F sind insgesamt folgende Dateien erforderlich: JQF.EXE, JQF-3A.DAT, JQF-3B.DAT, JQF-I.DAT und JQF-Z.DAT. Soll auf die Original-Daten der ZDB zugegriffen werden können, ist auch die Datei ZDB-ORI.DAT nötig, und bei Nutzung der MEM-Funktion auch eine von Benutzern eingerichtete '.JMM'-Datei.

Suche im J.Q.F

Schritte zur Suche jedes Einzelwortes: Zunächst werden die ersten 3 (bei kürzeren 'Wörtern' auch 2) Zei-

chen der Datei JQF-Z.DAT. Aus dieser werden dann die Zeitschriften-Titelangaben entnommen. Ja, und auf der Suche nach den Original-Daten geht das Ganze noch einen Schritt weiter: In JQF-Z.DAT steht die Start-Position der Original-Daten in der Datei ZDB-ORI.DAT.

Die Suche nach mehreren 'Wörtern' setzt eine Vorab-Analyse der Suchbegriffe voraus:

Abb. 3:

cjmaa	12345678		
canadian			12345678
journal	12345678		
mathematics	12345678		
tcjom	12345678		= Initialen mit allen Stoppwörtern
cjom	12345678		= Initialen nur mit Binnen-Stoppwort
cjm	12345678		= Initialen ganz ohne Stoppwörter
journal	12345678		
canadien			12345678
mathematiques	12345678		
toronto	12345678	0008-414x	12345678
ont	12345678	mat2/c16	12345678
univpr	12345678	#ab	12345678

A. Suche: 'journal canadian mathe'. Gesucht wird nach allen 3 Einzelwörtern und diese werden mit 'and' verknüpft. Dabei ist die Reihenfolge der Wörter ohne Belang.

B. Suche: 'canadian journal of mathematics'.

Gesucht wird zunächst nach den Initialen 'cjom.' und dann erst nach den Einzelwörtern 'canadian.', 'journal.' und 'mathematics.'.

Dabei wird unterstellt, daß ein Benutzer immer dann eine konkrete

C. Suche: 'can. j mathem'. Gesucht wird zunächst nach 'cjm.' und dann nach 'can.' und 'mathem.'.

Der abschließende Punkt ist bei Zeitschriften-Titeln als Abkürzungszeichen gebräuchlich und kann - muß aber nicht - im J.Q.F ebenfalls so benutzt werden. Enthält die Suchanfrage mehr als ein Einzelwort, werden nämlich alle Wörter automatisch trunkiert, also unterstellt, der Benutzer habe sie abgekürzt. Diese Unterstellung erlaubt

auch eine Suche nach einer Kombination aus Initialen und sinntragenden Einzelwörtern (beide verkürzt):

D. Suche: 'cj.math.'. In diesem Fall wird nach 'cj.' und 'math.' gesucht und gefunden.

Die hohe Zahl findbarer Abkürzungen folgt aus der Tatsache, daß jedes der Titelwörter für sich und unabhängig von den anderen bis herunter zum 1. Buchstaben verkürzt werden kann. Stoppwörter können dabei

auch ganz weggelassen werden; wozu einschränkend zu sagen ist, daß dann alle Binnen-Stoppwörter wegzulassen sind. Da die Suche nach den kompakten Initialen relativ zu einer Suche nach sinntragenden Wörtern meist weniger Treffer bringt, ist die Suche nach verkürzten Sucheingaben sogar ganz außergewöhnlich schnell.

Da den Benutzern keine feldspezifische Suche abverlangt wird, kann die Eingabe der Suchwörter außerdem sehr simpel erfolgen und trotz-

dem ein gutes Ergebnis erzielen. Das ist natürlich bei einem sehr viel größeren Titelbestand wie bei Büchern so nicht möglich. Zeitschriften-Titel werden jedoch üblicherweise verkürzt zitiert, und der J.Q.F geht auf diese Gewohnheit gezielt ein.

Weitere Funktionen sind in den Info-Seiten des J.Q.F erläutert. Dort steht auch, wonach sonst noch alles gesucht werden kann und wie sich z.B. Listen laufender Zeitschriften ein-

zelner Fächer drucken lassen. Hier soll nur die Wirkungsweise der mehrstufigen Indizes zur Beschleunigung der Suche, das Prinzip der Initialen-Suche erläutert und die erstaunliche Tatsache erklärt werden, auf welche Weise es gelingt, weit über tausend mögliche Abkürzungen eines Zeitschriften-Titels aufzulösen. Wer mehr wissen und den J.Q.F auf seinem eigenen PC haben will, kann ihn über '<http://www.uni-konstanz.de/ZE/Bib/zs/jqf-ftp.htm>' zu sich holen.

CD-Rom-Zugriff online

Eine neue Dienstleistung der Bibliothek

REINER GRUTTKE

PETRA HÄTSCHER

Es ist auch in der Bibliothek der Universität Konstanz soweit: Mit Beginn der Vorlesungszeit im Sommersemester 1998 werden zahlreiche bibliographische Datenbanken sowie Fakten- und Volltextdatenbanken, die auf CD-ROM zur Verfügung stehen, den Universitätsangehörigen rund um die Uhr online angeboten.

Bislang konnten wir den Service des Online-Zugriffs auf Datenbanken nur für wenige Titel anbieten, nämlich für die SilverPlatter-Produkte EconLit, Medline, PsycLit, Psyn dex, Sociofile und WISO/WAO. Diese Datenbanken sind seit kurzem per Web-Browser direkt nutzbar, man muß also keine Anwendungssoftware (WinSpirs-Client) mehr auf seinem lokalen PC installieren. Der Ersteinstieg kann über die Homepage der Bibliothek bzw. direkt über

<http://isubkn.ub.uni-konstanz.de/cgi-bin/webspirs.cgi> erfolgen.

Mit Beginn der Vorlesungszeit im Sommersemester 1998 werden ca. 70 weitere bibliographische Datenbanken sowie Fakten- und Volltextdatenbanken, die auf CD-ROM zur Verfügung stehen, den Universitätsangehörigen online angeboten. Für diese Datenbanken wird vor dem ersten Zugriff einmalig ein ICA-Treiber aus dem Netz auf das eigene Endgerät heruntergeladen und gestartet werden müssen, wobei die Anweisung dafür auf der Internet-Einstiegsseite hinterlegt sein wird. Nach Installation der Datei auf dem eigenen Gerät kann ab dem Zeitpunkt immer direkt auf den CD-ROM-Server der Bibliothek zugegriffen werden. In der Regel wird sich der Endnutzer einen Icon hinterlegen, über den er die gewünschte Anwendung anklicken kann.

Die EDV-Abteilung der Bibliothek installiert derzeit die Serverkonfiguration, die es ermöglicht, von allen Endgeräten, die in der Universität an das Netz angeschlossen sind, auf CD-ROM zuzugreifen. Ebenso kann von externen Geräten, die über das Rechenzentrum an das Netz ange-

schlossen sind und die damit auch über eine IP-Adresse der Universität verfügen, auf die Installation zugegriffen werden. Die Wissenschaftler können also Literaturrecherchen zukünftig - bei vorhandener technischer Ausstattung - sowohl von ihrem Arbeitsplatz als auch von zuhause durchführen.

Es können folgende Hardware-Plattformen von uns bedient werden:

- Intel: Windows 3.x, Win95, NT
- MacOS
- Unix-Derivate

Für die Techniker hier eine kurze Beschreibung der Server-Konfigurationen:

- CD-Server unter NT 4
Pentium 200 MMX
128 MB RAM
45 Gbyte HD-Kapazität
7 CD-Drives
- WinCenter-Server unter NT 3.51
Pentium II 266 (2x)
256 MB RAM