# Indexierung der HBZ-Verbunddaten mit FAST Data Search

8. InetBib-Tagung 2004 Bonn, 5.11.2004





Dr. Peter Kostädt, HBZ NRW

brought to you by

#### Katalog

- "Saubere" Daten
- Feldbezogene Suche
- Sacherschließung
- Klassifikation
- Normdaten

#### Suchmaschine

- Hohe Performanz
- Volltextsuche
- Ranking
- Kategorisierung
- Linguistik-Module



## Katalog + Suchmaschine

- Hohe Performanz
- "Saubere Daten
- Feldbezogene Suche
- Ranking
- Sacherschließung
- Klassifikation
- Kategorisierung
- Linguistik-Module
- Normdaten

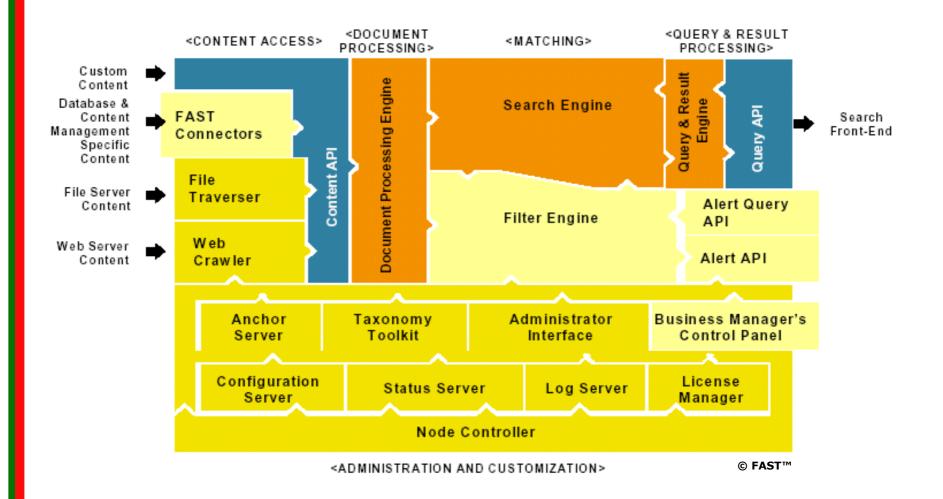
#### Wer oder was ist FAST?

- Die Software FAST Data Search™ ist ein Produkt der Firma Fast Search & Transfer™ (FAST)
- Die Firma wurde 1997 gegründet und ist ein Auswuchs einer Forschungs- und Entwicklungsabteilung der Norwegischen University of Science and Technology (NTNU)



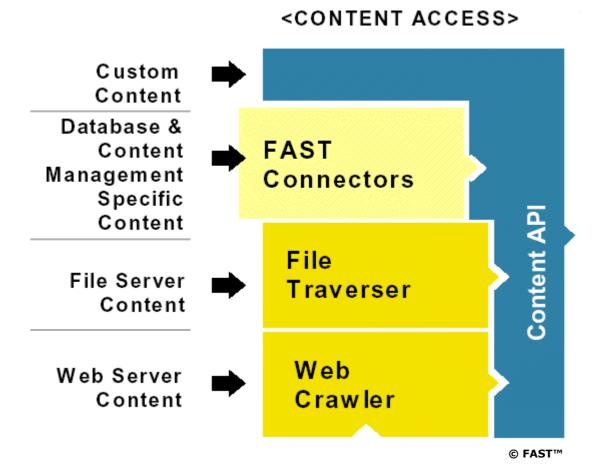


#### FAST Data Search - Systemarchitektur





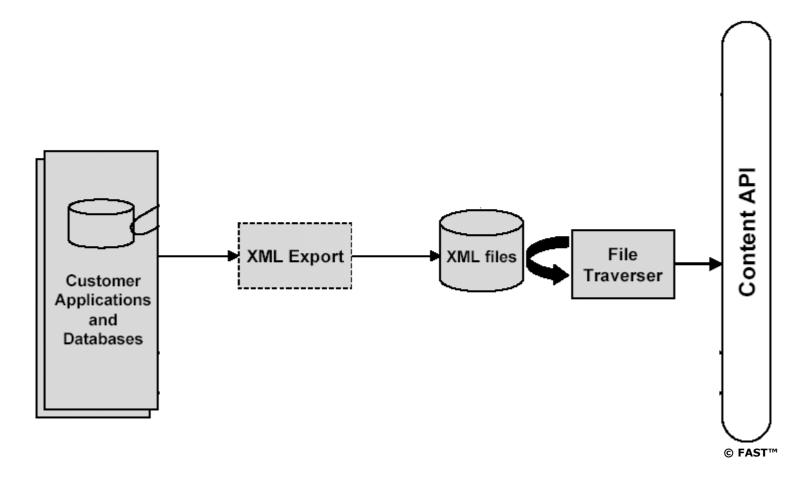
#### 1. Schritt: Daten einsammeln





#### Zugriff auf relationale Datenbanken

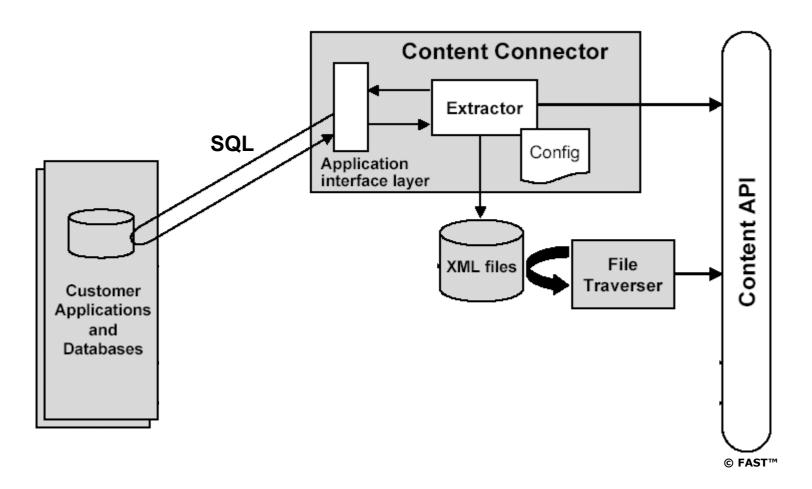
(i) Export → Umwandlung in XML → File Traversing





#### Zugriff auf relationale Datenbanken

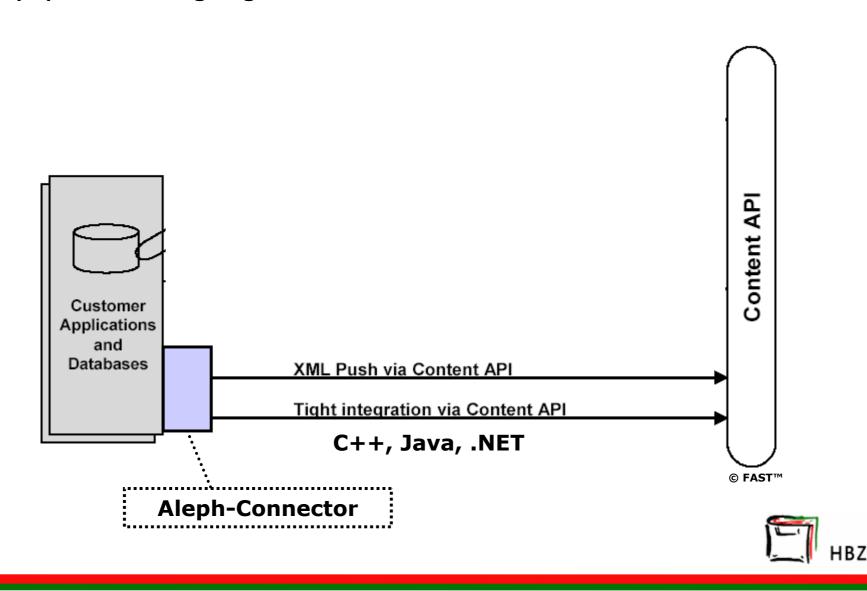
(ii) Nutzung der FAST Datenbank-Connectoren





#### Zugriff auf relationale Datenbanken

(iii) Anbindung eigener Connectoren über die Content API



#### Funktionsweise des Aleph-Connectors

- Auslesen der Titeldaten und Lokalsätze (über die HBZ-ID)
- Auflösen der BLOB bzw. LONG-Strukturen

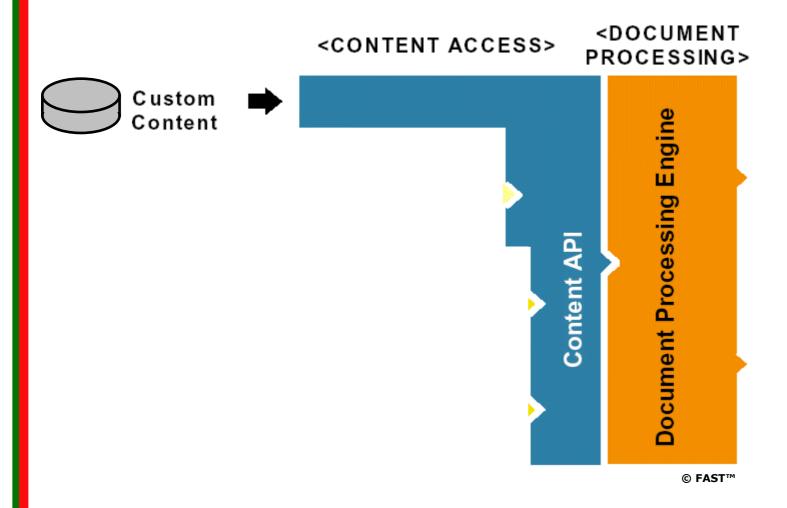
```
0030LDR L00351vM2.01200024-----h0008FMT LMH0020001
L$$aHT0003174160017002a L$$a199307220017003 L$$a200203130019030
Lh|5dc||||||70012037b L$$aeng0020050
La|||||||||||||||||0013051 Lm||||||0011070 L$$a610016070b L$$aHBZ/Red0011080
L$$a600044100 L$$aPope-Hennessy, John ...

$\int \frac{1}{4} \text{ } \text{ } \text{ } \frac{1}{4} \text{ } \te
```

• Konvertierung des Aleph-Internformats in XML-Strukturen

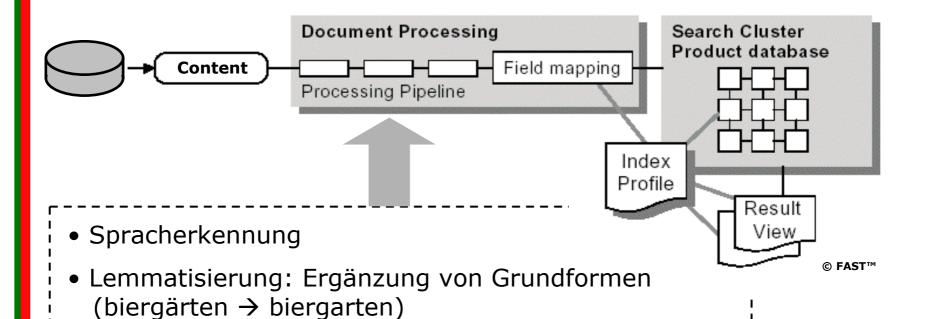


#### 2. Schritt: Daten analysieren und anreichern





## 2. Schritt: Daten analysieren und anreichern

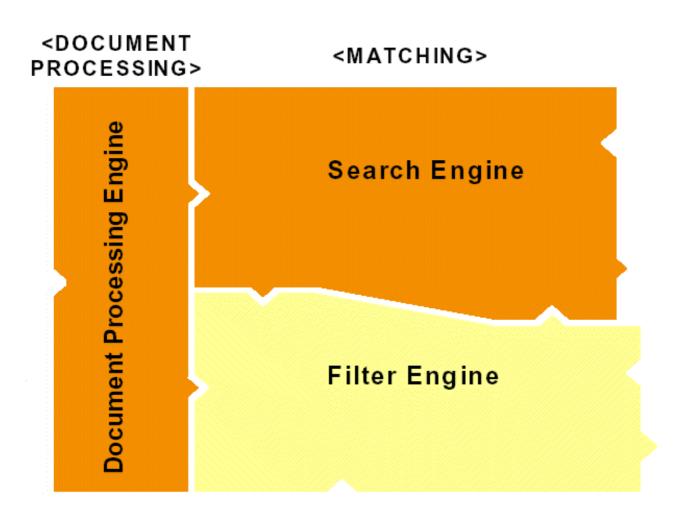


- Kategorisierung, z.B. für die Anzeige von Taxonomie-Bäumen in der Ergebnisanzeige
- Ermittlung von Ranking-Informationen
- Erzeugung von Vektorräumen für eine Ähnlichkeitssuche

• ...



## 3. Schritt: Indexierung





#### 3. Schritt: Indexierung

Es geht verdammt schnell ...

- Aleph-Connectoren:
   50-100 Datensätze pro Sekunde
- Document Processing Pipeline:
   ca. 20 Datensätze pro Sekunde
- Neuaufbau des Index: ca. 2-3 Tage
- ⇒ 11 Tage für die 14 Mio. Titel der HBZ-Verbunddatenbank



#### 3. Schritt: Indexierung

#### Indexgröße

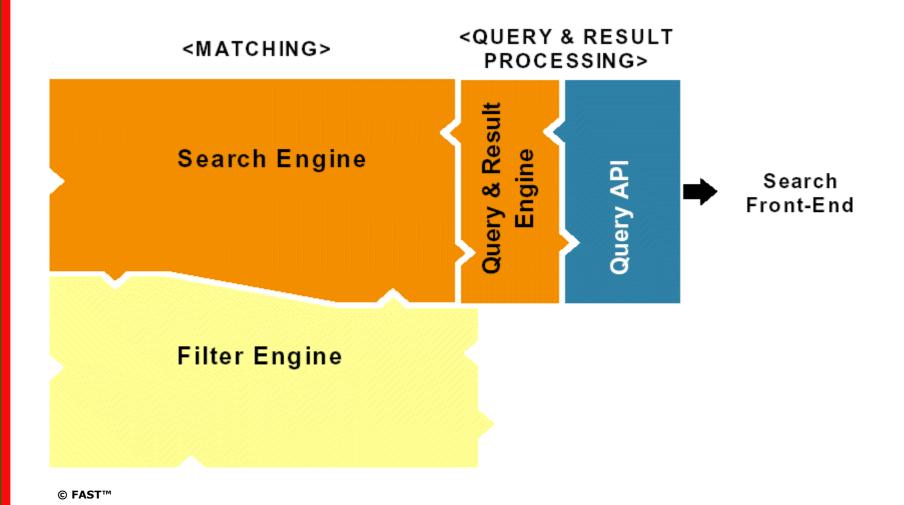
- FastXML: ca. 15 GB
- Binärindex: ca. 100 GB

#### Hardware:

Dell PowerEdge 6650
 4 x 1,9 GHz (Intel Xeon)
 4 GB RAM
 SCSI Festplatten (15.000 U/Min)

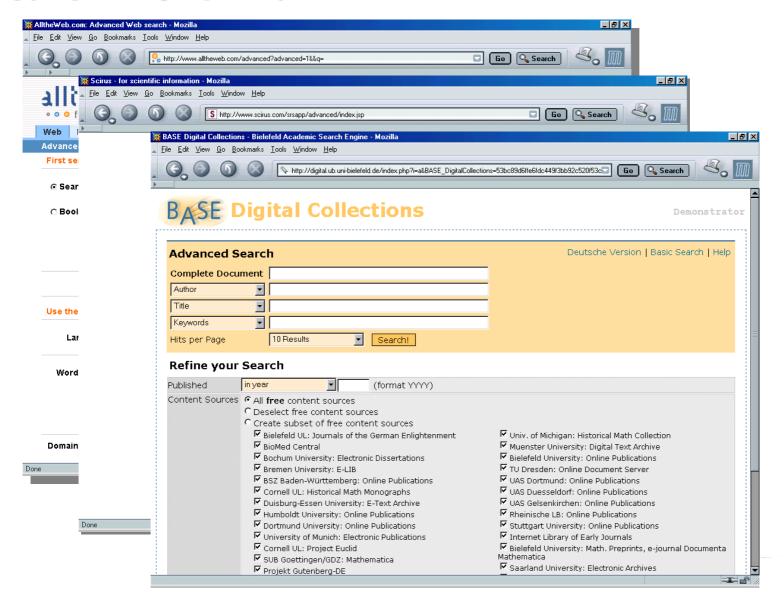


## 4. Schritt: Query & Result Processing





#### Search Front-End









\_ B ×



#### Demonstration der FAST-Suche im HBZ-Verbundkatalog

Suche nach		ausführen		
Titel:			HBZ-Nummer 🛭	
Autor:			ZDB-Nummer 🛚	
Körperschaft:			ISBN [	
Schlagwort:			ISSN [	
Treffermodus alle Wörter	Rechtschreib∨orschläge □			
Sortierung Relevanz	Vorschlag direkt ausführen □			
Sortierrichtung absteigend •	Sprache für Rechtschreib∨orschläge englisch			
Navigation ein/aus <b>☑</b>	Systematiken ein/aus □	Lemma-Index	∢□ Auf Sigel beschränk Alle	en ▼

© 2004 FAST Search and Transfer, Inc.









#### Bsp. 1: Rechtschreibvorschläge





#### Demonstration der FAST-Suche im HBZ-Verbundkatalog

Suche nach | acess | ausführen |
Wollen Sie vielleicht nach access suchen?

2 Treffer in 0,069 Sekunden

1 MacGuire, Christine | Clinical simulations / selected problems in patient management / ACESS-TM book | 1971

- H. (Hrsg.)

  2 Dreo. Gabriiela ¬A¬ frame work for supporting fault diagnosis in integrated network and systems management / methodologies for the 1995
- 2 Dreo, Gabrijela ¬A¬ frame work for supporting fault diagnosis in integrated network and systems management / methodologies for the correlation of trouble tickets and acess to problem solving expertise / Berichte aus der Informatik

© 2004 FAST Search and Transfer, Inc.

Wollten Sie vielleicht nach access suchen?





#### Bsp. 2: Suche ohne Lemmatisierung



1971



#### Demonstration der FAST-Suche im HBZ-Verbundkatalog

Suche nach | computer algebra systeme | ausführen |

3 Treffer in 0,094 Sekunden |

1 Bachmann, Franz; Schärer, Hans Rudolf; Willimann, Louis-Sepp |

2 Lemmen, Markus; Wey, Torsten; Jelali, Mohieddine |

NSAS, ein Computer-Algebra-Paket zur Analyse und Synthese nichtlinearer Systeme / Forschungsbericht Meß-, Steuer- und Regelungstechnik / Forschungsbericht / Forsc

3 Forsythe, George E.; Moler, Cleve B.

Computer solution of linear algebraic systems dt. / Computer-Verfahren für lineare algebraische Systeme / mit 5 Programmbeispielen / Verfahren der Datenverarbeitung

© 2004 FAST Search and Transfer, Inc.









## Bsp. 2: Suche mit Lemmatisierung



1996

1995

1990

1992

1971

1995

1995

1990

lben 1998



#### Demonstration der FAST-Suche im HBZ-Verbundkatalog

Suche nach | computer algebra systeme ausführer 9 Treffer in 0,177 Sekunden Mathematik mit MATLAB / Aufgaben und Lösungen / Computer-Algebra-Systeme

- Bachmann, Franz; Schärer, Hans Rudolf; Willimann, Louis-Sepp

NSAS, ein Computer-Algebra-Paket zur Analyse und Synthese nichtlinearer Systeme / Forschungsbericht Meß-,

Die Suche nach "computer algebra systeme" findet

2 Lemmen, Markus; Wey, Decelupactochnik: 05.00 (Forechungebericht Mod., Stouer, und Decelupactochnik (Mod., Stouer, Torsten; Jelali, Mobioddina

- 3 Kutzler, Bernhard; Lichtenberger, Fra Franz
- Computer Algebra Systeme
- 4 Kutzler, Bernhard: Bernhard; Winkler
- Computer Algebra Systemen
- 5 Scheuermann, Hel
- Computer Algebra Systems
- 6 Forsythe, George Cleve B.
- Computer Algebra System
- Steins, Andreas
- 8 Steins, Andreas Verifizierte Formelauswertung in Computer-Algebra-Systemen und objektorientierten Programmierumgebungen
- Johannes; Greither, Cornelius

9 Gollan, Holger W.; Grabmeier, Algorithms in representation theory and their realization in the computer algebra system scratchpad / Bayreuther mathematische Schriften; 33 / Generic cyclic extensions of prime power degree / Bayreuther mathematische Schriften / Generic cyclic extensions of prime power degree / C. Greither [u.a.]

© 2004 FAST Search and Transfer, Inc.







Done

## Bsp. 3: Suche verfeinern über Navigatoren





Demonstration der FAST-Suche im HE	3Z-\	/erbundkatalog		
Suche nach goethe			ausführen	
,-			adsidifferi	
15.810 Treffer in 0,287 Sekunden				
Navigation zurücksetzen  Navigation  Dokumenttypen	1	Maisak, Petra; Goethe, Johann Wolfgang ¬von¬; [III.]	Johann Wolfgang Goethe, Zeichnungen	1996
Alle Autoren	2	Goethe, Johann Wolfgang ¬von¬	¬Die¬ Wahlverwandtschaften / ein Roman / [15. Aufl.] / Insel-Taschenbuch ; 1 / Insel-Taschenbuch / Mit einem	1995
Alle			Essay "Goethes Wahlverwandtschaften" / von Walter Benjamin	
Körperschaften Alle	3	Goethe, Johann Wolfgang ¬von¬; Lavater, Johann Caspar; Funck,	Goethe und Lavater / Briefe u. Tagebücher / Schriften der Goethe-Gesellschaft ; 16 / Schriften der	1901
Schlagwörter		Heinrich (Hrsg.)	Goethe-Gesellschaft	
Alle	4	Runge, Philipp Otto; Goethe,	Briefwechsel mit Goethe / Philipp Otto Runges Briefwechsel	1940
Sprachen Alle		Johann Wolfgang ¬von¬; Maltzahn, Hellmuth ¬von¬ (Hrsg.)	mit Goethe / Schriften der Goethe-Gesellschaft ; 51 / Schriften der Goethe-Gesellschaft	
Erscheinungsjahre	5	Goethe, Johann Wolfgang ¬von¬;	Tagebücher und Briefe Goethes aus Italien an Frau von	1886
Alle Erscheinungsorte		Stein, Charlotte ¬von¬; [Adressat]; Herder, Johann Gottfried ¬von¬	Stein und Herder / mit Beilagen / Schriften der Goethe-Gesellschaft; 2 / Schriften der Goethe-Gesellschaft	1000
Alle		•	·	
Bände/Stücktitel/Zusätze  Alle   ▼	6	Schüddekopf, Carl (Hrsg.); Goethe, Johann Wolfgang ¬von¬	Goethe und die Romantik / Briefe mit Erl. / Schriften der Goethe-Gesellschaft ;	1898
	7	Goethe, Johann Wolfgang ¬von¬; Sauer, August (Hrsg.)	Goethe und Österreich / Briefe mit Erläuterungen / Schriften der Goethe-Gesellschaft ; $\dots$	1902
	8	Hansen, Volkmar (Hrsg.); Goethe, Johann Wolfgang ¬von¬	Ausstellung "Europa, wie Goethe es sah" / Düsseldorf, Saverne, Bologna, 1999 ; Katalog ; [zu Ehren Goethes 250. Geburtstag] / "Europa, wie Goethe es sah"	1999
	9	Goethe, Johann Wolfgang ¬von¬; Scheidig, Walther (Hrsg.)	Goethes Preisaufgaben für bildende Künstler / 1799 - 1805 / Schriften der Goethe-Gesellschaft : 57 / Schriften der	1958



## Bsp. 3: Suche verfeinern (Dokumenttyp)





Demonstration der FAST-Suche im HI	3Z-\	verbundkatalog			
Suche nach goethe			ausführen		
Į.			ausiunien		
15.810 Treffer in 0,287 Sekunden	1				
Navigation zurücksetzen  Navigation  Dokumenttypen	1	Maisak, Petra; Goethe, Johann Wolfgang ¬von¬; [III.]	Johann Wolfgang Goethe, Zeichnungen	1996	
Alle Alle Bücher (10287) Serien (3912)	2	Goethe, Johann Wolfgang ¬von¬	¬Die¬ Wahlverwandtschaften / ein Roman / [15. Aufl.] / Insel-Taschenbuch ; 1 / Insel-Taschenbuch / Mit einem Essay "Goethes Wahlverwandtschaften" / von Walter Benjamin	1995	
Stücke (3735) Dissertationen (995) Zeitschriften (261) Kataloge (123)	3	Goethe, Johann Wolfgang ¬von¬; Lavater, Johann Caspar; Funck, Heinrich (Hrsg.)	Goethe und Lavater / Briefe u. Tagebücher / Schriften der Goethe-Gesellschaft ; 16 / Schriften der Goethe-Gesellschaft	1901	
Festschriften (80) Musikalien (79) Konferenzschriften (74)	4	Runge, Philipp Otto; Goethe, Johann Wolfgang ¬von¬; Maltzahn, Hellmuth ¬von¬ (Hrsg.)	Briefwechsel mit Goethe / Philipp Otto Runges Briefwechsel mit Goethe / Schriften der Goethe-Gesellschaft ; 51 / Schriften der Goethe-Gesellschaft	1940	
Bibliograhien (62) Sonderdrucke (53) Biographien (46) Fortschrittsberichte (28)	5	Goethe, Johann Wolfgang ¬von¬; Stein, Charlotte ¬von¬; [Adressat]; Herder, Johann Gottfried ¬von¬	Tagebücher und Briefe Goethes aus Italien an Frau von Stein und Herder / mit Beilagen / Schriften der Goethe-Gesellschaft ; 2 / Schriften der Goethe-Gesellschaft	1886	
Wörterbücher (27) Enzyklopädien (17) Aufsätze (14)	6	Schüddekopf, Carl (Hrsg.); Goethe, Johann Wolfgang ¬von¬	Goethe und die Romantik / Briefe mit Erl. / Schriften der Goethe-Gesellschaft ;	1898	
Gesetze (1) ' Berichte (1)	7	Goethe, Johann Wolfgang ¬von¬; Sauer, August (Hrsg.)	Goethe und Österreich / Briefe mit Erläuterungen / Schriften der Goethe-Gesellschaft ; $\dots$	1902	
	8	Hansen, Volkmar (Hrsg.); Goethe, Johann Wolfgang ¬von¬	Ausstellung "Europa, wie Goethe es sah" / Düsseldorf, Saverne, Bologna, 1999 ; Katalog ; [zu Ehren Goethes 250. Geburtstag] / "Europa, wie Goethe es sah"	1999	
	9	Goethe, Johann Wolfgang ¬von¬; Scheidig, Walther (Hrsg.)	Goethes Preisaufgaben für bildende Künstler / 1799 - 1805 / Schriften der Goethe-Gesellschaft : 57 / Schriften der	1958	
				-III:- <u> </u>	





## Bsp. 3: Suche verfeinern (Dokumenttyp)





#### Demonstration der FAST-Suche im HBZ-Verbundkatalog

Suche nach goethe			ausführen	
62 Treffer in 0,307 Sekunden				
Navigation zurücksetzen  Navigation  Dokumenttypen	1	Dickson, Alexander J.	Goethe in England / 1909 - 1949 ; a bibliography / Publications of the English Goethe Society ; N.S.,19 / Publications of the English Goethe Society	1951
Bibliograhien (62)  Autoren  Alle  Körperschaften  Alle	2	Buchwald, Reinhard; Goethe, Johann Wolfgang ¬von¬; Sierks, Ruth; [Bearb.]	Goethe / der Mensch, der Dichter, der Denker ; Bücher von ihm und über ihn ; ein Verzeichnis / Goethe-Studienfuehrer / 2. Aufl. / neu bearb. von Ruth Sierks	1951
Schlagwörter Alle Sprachen Alle	3	Goethe, Johann Wolfgang ¬von¬; Dobel, Richard (Hrsg.)	Lexikon der Goethe-Zitate	1968
Erscheinungsjahre  Alle  Erscheinungsorte	4	Pick, Robert (Hrsg.)	Schiller in England / 1787 - 1960 ; a bibliography / Publications of the English Goethe Society ; N.S., 30 / Publications of the English Goethe Society	1961
Alle  Bände/Stücktitel/Zusätze  Alle	5	Pick, Robert (Hrsg.); Schiller, Friedrich (Hrsg.)	Schiller in England / 1787-1960 ; a bibliography / Publications of the English Goethe Society. New Series ; 30 / Publications of the English Goethe Society	1967
	6	Metzner, Günter	"Wir heißen euch hoffen" / Materialien zur musikalischen Goethe-Rezeption 1945 - 1968	1995
	7	Goethe, Elisabeth; Goethe, Friedrich	Bibliographie der vom Mellumrat betreuten Naturschutzgebiete	1982
	8	Kähmer, Inge	Goethe-Museum Düsseldorf, Anton-und-Katharina-Kippenberg-Stiftung:	1987

Katalog der Musikalien / Katalog der Musikalien / Schriften des

Done



Done

## Bsp. 3: Suche verfeinern (Erscheinungsjahr)



\_ B ×



#### Demonstration der FAST-Suche im HBZ-Verbundkatalog

		- · · · · · · · · · · · · · · · · · · ·		
Suche nach goethe			ausführen	
62 Treffer in 0,307 Sekunden				
Navigation zurücksetzen  Navigation  Dokumenttypen	1	Dickson, Alexander J.	Goethe in England / 1909 - 1949 ; a bibliography / Publications of the English Goethe Society ; N.S.,19 / Publications of the English Goethe Society	1951
Bibliograhien (62)  Autoren  Alle  Körperschaften  Alle	2	Buchwald, Reinhard; Goethe, Johann Wolfgang ¬von¬; Sierks, Ruth; [Bearb.]	Goethe / der Mensch, der Dichter, der Denker ; Bücher von ihm und über ihn ; ein Verzeichnis / Goethe-Studienfuehrer / 2. Aufl. / neu bearb. von Ruth Sierks	1951
Schlagwörter  Alle  Sprachen  Alle	3	Goethe, Johann Wolfgang ¬von¬; Dobel, Richard (Hrsg.)	Lexikon der Goethe-Zitate	1968
Erscheinungsjahre  Alle	4	Pick, Robert (Hrsg.)	Schiller in England / 1787 - 1960 ; a bibliography / Publications of the English Goethe Society ; N.S., 30 / Publications of the English Goethe Society	1961
Vor 1958 Zwischen 1958 und 1979 Zwischen 1979 und 1984 Nach 1984	5	Pick, Robert (Hrsg.); Schiller, Friedrich (Hrsg.)	Schiller in England / 1787-1960 ; a bibliography / Publications of the English Goethe Society. New Series ; 30 / Publications of the English Goethe Society	1967
	6	Metzner, Günter	"Wir heißen euch hoffen" / Materialien zur musikalischen Goethe-Rezeption 1945 - 1968	1995
	7	Goethe, Elisabeth; Goethe, Friedrich	Bibliographie der vom Mellumrat betreuten Naturschutzgebiete	1982
	8	Kähmer, Inge	Goethe-Museum Düsseldorf, Anton-und-Katharina-Kippenberg-Stiftung:	1987

Katalog der Musikalien / Katalog der Musikalien / Schriften des Arhaitel/raisas Salhetändigar Kultur Instituta : 6 / Schriftan das





## Bsp. 4: Suche verfeinern (Schlagwörter)





#### Demonstration der FAST-Suche im HBZ-Verbundkatalog

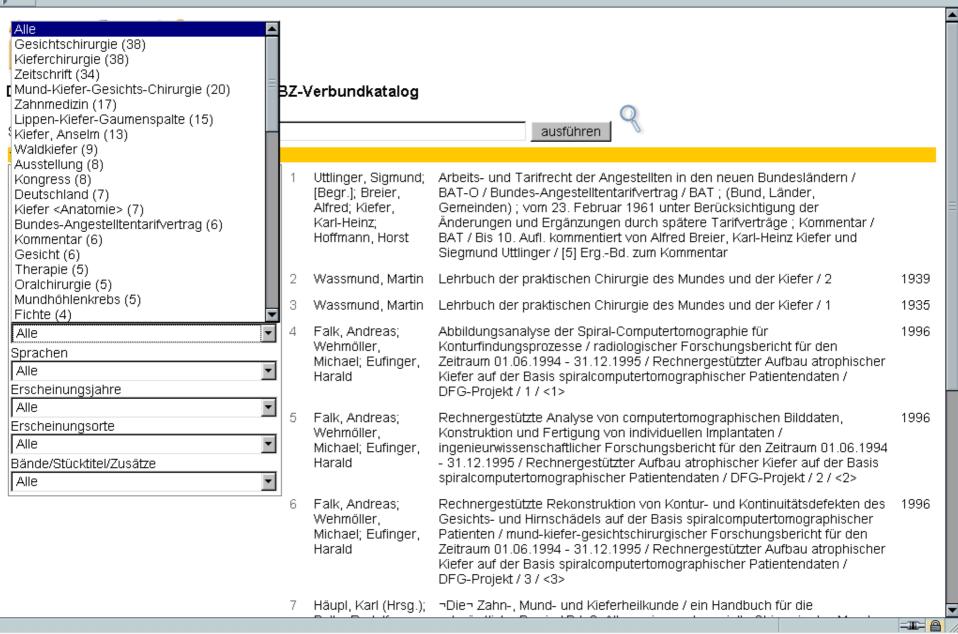
Suche nach | title: "kiefer" ausführen 1.543 Treffer in 0,286 Sekunden Navigation zurücksetzen Uttlinger, Sigmund; Arbeits- und Tarifrecht der Angestellten in den neuen Bundesländern / Navigation BAT-O / Bundes-Angestelltentarifvertrag / BAT; (Bund, Länder, [Begr.]; Breier, Dokumenttypen Alfred: Kiefer. Gemeinden): vom 23. Februar 1961 unter Berücksichtigung der Änderungen und Ergänzungen durch spätere Tarifverträge; Kommentar / Alle Karl-Heinz: BAT / Bis 10. Aufl. kommentiert von Alfred Breier, Karl-Heinz Kiefer und Hoffmann, Horst Autoren: Siegmund Uttlinger / [5] Erg.-Bd. zum Kommentar Alle Körperschaften Wassmund, Martin Lehrbuch der praktischen Chirurgie des Mundes und der Kiefer / 2 1939 Alle Wassmund, Martin Lehrbuch der praktischen Chirurgie des Mundes und der Kiefer / 1 1935 Schlagwörter Falk, Andreas; Abbildungsanalyse der Spiral-Computertomographie für Alle 1996 Konturfindungsprozesse / radiologischer Forschungsbericht für den Wehmöller. Sprachen Zeitraum 01.06.1994 - 31.12.1995 / Rechnergestützter Aufbau atrophischer Michael; Eufinger, Alle Kiefer auf der Basis spiralcomputertomographischer Patientendaten / Harald Erscheinungsjahre DFG-Projekt / 1 / <1> Alle Falk, Andreas: Rechnergestützte Analyse von computertomographischen Bilddaten, 1996 Erscheinungsorte Konstruktion und Fertigung von individuellen Implantaten / Wehmöller. l Alle ingenieurwissenschaftlicher Forschungsbericht für den Zeitraum 01.06.1994 Michael; Eufinger, - 31.12.1995 / Rechnergestützter Aufbau atrophischer Kiefer auf der Basis Bände/Stücktitel/Zusätze Harald spiralcomputertomographischer Patientendaten / DFG-Projekt / 2 / <2> l Alle Rechnergestützte Rekonstruktion von Kontur- und Kontinuitätsdefekten des Falk, Andreas; Wehmöller, Gesichts- und Hirnschädels auf der Basis spiralcomputertomographischer Michael; Eufinger, Patienten / mund-kiefer-gesichtschirurgischer Forschungsbericht für den Zeitraum 01.06.1994 - 31.12.1995 / Rechnergestützter Aufbau atrophischer Harald Kiefer auf der Basis spiralcomputertomographischer Patientendaten / DFG-Projekt / 3 / <3> Häupl, Karl (Hrsg.); ¬Die¬ Zahn-, Mund- und Kieferheilkunde / ein Handbuch für die





## Bsp. 4: Suche verfeinern (Schlagwörter)







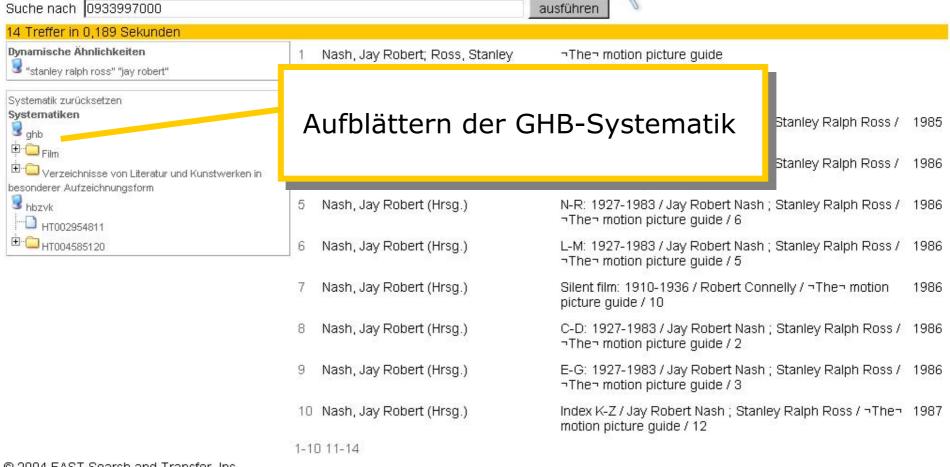
#### \_ B ×

## Bsp. 5: Suche verfeinern (Taxonomiebäume)

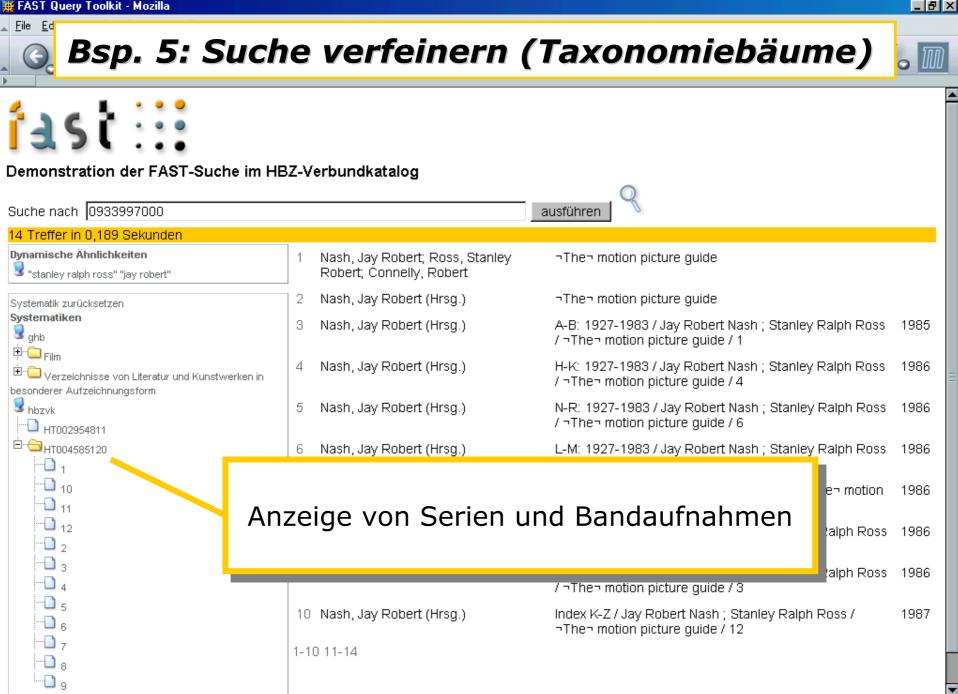




#### Demonstration der FAST-Suche im HBZ-Verbundkatalog



© 2004 FAST Search and Transfer, Inc.





#### Die nächsten Schritte ...

- Erstellung eines kompletten Abzugs des HBZ-VK
- Implementierung von Routinen für die Aktualisierung (ca. 20.000 Änderungen pro Tag im Titel- und Lokaldatenbereich)
- Testen, testen, testen ...
- Einbindung der PND
- Entwicklung von Algorithmen für ein Ranking nach Sigeln
- Erstellung weiterer Kollektionen (Jade, IBZ, PCI, ...)
- Anbindung des Bielefelder FAST-Knotens (BASE)
- Implementierung von Web-Services



#### Herzlichen Dank für Ihre Aufmerksamkeit!

Dr. Peter Kostädt kostaedt@hbz-nrw.de www.hbz-nrw.de



HBZ

Jörg Prante prante@hbz-nrw.de www.hbz-nrw.de



HBZ