

# Suchmaschinentechnologie – ein Weg zu nationalen Nachweissystemen

Reiner Diedrichs  
Verbundzentrale des GBV (VZG)

95. Deutscher Bibliothekartag in Dresden, 22. 3.2006

# Suchmaschinentechnologie

## Was ist das?

- Traditionell = Metasuche
  - Durchsuchen externer Datenbanken (Z39.50 etc.)
  - KvK, P7+ (GBV), DigiBib (NRW)  
Software: iPort, Elektra, Metalib, IPS, usw.
  - Nachteile:
    - Abhängig von der Verfügbarkeit externer Systeme  
„der langsamste bestimmt das Tempo“
    - Abhängig von externer Indexierung
    - Dublettenzusammenführung bei Verbunddatenbanken
    - Indexbrowsing
  - Vorteile:
    - Für Standardprotokolle sofort und mit wenig Aufwand verfügbar

# Suchmaschinentechnologie

- Einsammeln der zu indexierenden Inhalte
- Aufbereitung nach eigenen Regeln
- Typisch für Internetsuchmaschinen: Google, Yahoo, usw.
- Software: FAST, Lucene, PSI (Pica Search&Index)
- Vorteile:
  - Homogener Index
  - Gute Integrierbarkeit in eigene Anwendungen und Oberflächen
- Problemstellung:
  - Große Datenmengen
  - Aufbereitung der Daten

# Suchmaschinentechnologie – ein Weg zu nationalen Nachweissystemen?

- JA!
- Aber ...
- Nicht DAS NATIONALE NACHWEISSYSTEM
- Mehrere Implementierungen:
  - Innerhalb der gewohnten Umgebungen
  - Mit den gewohnten Dienstleistungen:  
Fernleihe, Lieferdienste, TOC, OPAC-Verlinkung, usw.
- Zwei Ansätze:
  - Bibliothekarische Suchmaschine PSI (OCLCPICA, GBV)
  - Internetsuchmaschine FAST (hbz)

Suchen | Suchergebnis | Erweiterte Suche | Zwischenablage | Benutzer-Info | Hilfe © 1998-2006 OCLC PICA

suchen [und] | Alle Wörter [ALL] | ? | sortiert nach Erscheinungsjahr | unscharfe Suche

suchen | Alle Verbünde

Benutzergruppe: 8000 | IP : 195.37.139.61

Über KaViA | Willkommen

Abmelden

- **KaViA - Katalog der Verbünde GBV, HBZ und SWB**

Der Katalog der Verbünde enthält zurzeit die kompletten Titeldaten aus den Verbundkatalogen des GBV, HBZ und SWB. Vor jedem Treffer wird in der Kurztitelliste angezeigt, aus welchem Verbund der Titel stammt. Gibt es Treffer aus mehreren Verbundkatalogen, die gleich bzw. ähnlich sind (gleiche ISBN, ISSN oder ZDB-Nr), finden Sie einen Link "Ähnliche Titel" in der Kurzanzeige. Ein bekanntes Problem ist, dass in bestimmten Fällen - trotz des angezeigten Links "Ähnliche Titel" - nur ein Treffer angezeigt wird. Dieser Katalog wird sukzessive ausgebaut und mit Titeldaten weiterer Verbünde ergänzt. Regelmäßige Aktualisierungen erfolgen in Kürze. Die nachgewiesenen Bestände sind weitgehend (bis auf SWB Nachweise) über die [Online-Fernleihe](#) oder den Dokumentlieferdienst [GBVdirekt/subito](#) bestellbar.

**Verbundkatalog des Gemeinsamen Bibliotheksverbundes (GBV)**  
(ca. 27,1 Mio Titel und 52,7 Mio Besitznachweise - Stand: Januar 2006)

**Verbundkatalog des Hochschulbibliothekszentrums Nordrhein-Westfalen (HBZ)**  
(ca. 13,7 Mio Titel und 29,8 Mio Besitznachweise - Stand: November 2005)

**Verbundkatalog des Südwestdeutschen Bibliotheksverbundes (SWB)**  
(ca. 10 Mio Titel und 29,5 Mio Besitznachweise - Stand: August 2004)

**Wichtiger Hinweis:**  
Bei diesem Katalog handelt es sich um eine Testversion, die laufend weiterentwickelt wird. Es kann also sein, dass Sie auf Inkonsistenzen oder Fehler stoßen oder der Katalog gelegentlich nicht erreichbar ist bzw. Ihre Sitzung unvermittelt abbricht. Wir bitten dies zu entschuldigen, doch es ist uns wichtig, Ihnen schon jetzt diesen Katalog vorzustellen.

Haben Sie Anregungen, Kritik oder Fragen, können Sie uns [über diesen Link](#) eine E-Mail zukommen lassen.

# KaViA – Technische Plattform

- Logische Teildatenbanken: GBV, BSZ, HBZ, ...
- Gemeinsame bibliographischer Index
- Multithread Search:
  - Partitionierung des Index
  - Beliebig viele Search Nodes
  - Optimal für Rechnercluster aus preiswerten (L)unix-Rechnern
  - Zurzeit: Demo-Version mit 4 Prozessoren
  - In Kürze: Produktionsversion mit 16 Prozessoren

# KaViA – Indexierung

Standard-  
indexierung  
für Verbund-  
datenbanken

## IKT ILN Index Key Phrase

IKT	ILN	Index	Key	Phrase
8617 (TIW)	0	1	C	berlin
8617 (TIW)	0	1	C	bibliothekswissenschaft
8617 (TIW)	0	1	C	fuer
8617 (TIW)	0	1	C	humboldt
8617 (TIW)	0	1	C	humboldtuniversitaet
4 (TIT), 8617 (TIW)	0	1	A	im
8617 (TIW)	0	1	C	institut
4 (TIT), 8617 (TIW)	0	1	A	internet
8617 (TIW)	0	1	C	konrad
4 (TIT), 8617 (TIW)	0	1	A	suchmaschinen
8617 (TIW)	0	1	C	umlauf
8617 (TIW)	0	1	C	universitaet
8617 (TIW)	0	1	C	von
1008 (PUS)	0	5	G	berlin
1008 (PUS)	0	5	G	berlin inst. fuer bibliothekswiss.
1008 (PUS)	0	5	G	inst. fuer bibliothekswiss.
8060 (TPK), 8062 (TAF)	0	10	A	suchmaschinen im internet
8061 (TPL), 8062 (TAF)	0	10	B	suchmaschinen im internet von konrad umlauf humboldtuniversitaet
5 (SER), 8541 (VSE)	0	13	A	58
5 (SER), 8541 (VSE)	0	13	A	berliner
5 (SER), 8541 (VSE)	0	13	A	bibliothekarausbildung
5 (SER), 8541 (VSE)	0	13	A	bibliothekswissenschaft
5 (SER), 8541 (VSE)	0	13	A	h
5 (SER), 8541 (VSE)	0	13	A	handreichungen
5 (SER), 8541 (VSE)	0	13	A	zur
1007 (NUM), 8530 (SNR)	0	14	Z	00b270209
1007 (NUM), 8530 (SNR)	0	14	Z	959099980
1 (PRS), 1004 (PER)	0	16	A	umlauf,konrad
8063 (GTI), 8540 (RTI)	0	19	A	berliner handreichungen zur bibliothekswissenschaft und bibliotheka
1016 (ALL)	0	20	A	00b270209
1016 (ALL)	0	20	A	58
1016 (ALL)	0	20	A	959099980
1016 (ALL)	0	20	A	berlin
1016 (ALL)	0	20	A	berliner
1016 (ALL)	0	20	A	bibliothekarausbildung
1016 (ALL)	0	20	A	bibliothekswiss
1016 (ALL)	0	20	A	bibliothekswissenschaft
1016 (ALL)	0	20	A	h
1016 (ALL)	0	20	A	handreichungen
1016 (ALL)	0	20	A	inst
1016 (ALL)	0	20	A	internet
1016 (ALL)	0	20	A	konrad
1016 (ALL)	0	20	A	suchmaschinen
1016 (ALL)	0	20	A	umlauf
59 (PLC), 8580 (PUB)	0	22	A	berlin
8538 (VLG), 8580 (PUB)	0	22	B	bibliothekswiss
8538 (VLG), 8580 (PUB)	0	22	B	fuer
8538 (VLG), 8580 (PUB)	0	22	B	inst
5041 (THM)	0	23	J	internet
5041 (THM)	0	23	J	suchmaschinen

# KaViA – Suchoptionen

Suchen | Suchergebnis | **Erweiterte Suche** | Zwischenablage | Benutzer-Info | Hilfe © 1998-2006 OCLC PICA

**KaViA**  
GBV

**Suchfilter** | Sachgebiete

Abmelden

**Suchfilter**

Füllen Sie das Formular aus, stellen Sie ggf. weitere Optionen ein und starten die Suche mit einem Klick auf die Schaltfläche **suchen**. [Formular zurücksetzen](#)

Alle Wörter [ALL]  und

Person, Autor [PER]  und

Schlagwörter [SLW]  und

Ort,Verlag (Stichwort) [PUB]

sortiert nach Relevanz

Erscheinungsjahr  z.B.: "1948-1987", "1996-", "1965, 1984, 2001" oder "2003"

Bestand -- Alle Verbände --

Sprache -- Alle Sprachen --

Land -- Alle Länder --

Materialart [alle](#) | [keine](#)

<input checked="" type="checkbox"/> Bücher	<input checked="" type="checkbox"/> Briefe	<input checked="" type="checkbox"/> Musikalien
<input checked="" type="checkbox"/> Zeitschriften/Serien (ohne Online-Zeitschr.)	<input checked="" type="checkbox"/> Filme, Videos, etc.	<input checked="" type="checkbox"/> Tonträger
<input checked="" type="checkbox"/> Online Zeitschriften	<input checked="" type="checkbox"/> Bilder	<input checked="" type="checkbox"/> Datenträger
<input checked="" type="checkbox"/> Online Ressourcen (ohne Zeitschr.)	<input checked="" type="checkbox"/> Kartenmaterial	<input checked="" type="checkbox"/> Mikroformen
<input checked="" type="checkbox"/> Aufsätze	<input checked="" type="checkbox"/> Manuskripte	<input checked="" type="checkbox"/> Spiele, Skulpturen, etc.



# KaViA – Spezifische Funktionalitäten

- Quellenanzeige und Titel-Clustering

The screenshot shows the KaViA search interface. At the top, there are navigation tabs: Suchen, Suchergebnis (selected), Erweiterte Suche, Zwischenablage, Benutzer-Info, and Hilfe. The search bar contains 'hbz' and is set to 'Alle Wörter [ALL]'. The results are sorted by 'Erscheinungsjahr'. A red circle highlights the 'Alle Verbünde' dropdown menu. Below the search bar, there are tabs for Suchgeschichte, Kurzliste, Titeldaten, and Nachweisinformationen. The search results are displayed in a list format, with each item showing a source code (e.g., GBV, SWB, HBZ) and a title. A red circle highlights the 'Ähnliche Titel' link next to item 14. The interface also includes a sidebar with 'Download', 'Trefferanalyse', and 'Abmelden' options, and a footer with the URL 'http://www.gbv.de'.

Suchen | Suchergebnis | Erweiterte Suche | Zwischenablage | Benutzer-Info | Hilfe

suchen [und] | Alle Wörter [ALL] | sortiert nach Erscheinungsjahr | unscharfe Suche

hbz | suchen | Alle Verbünde

Benutzergruppe: 8000 | IP : 195.37.139.61

Suchgeschichte | Kurzliste | Titeldaten | Nachweisinformationen

Ihre Aktion suchen [und] (Alle Wörter [ALL]) hbz 11 - 20 von 98

siehe auch: [hbz](#) | [hochschulbibliothekszentrum](#) | [zknw](#) | [hbzknw](#) | [hbzsk](#) | [zentrakatalog](#) | [verbundkatalog](#) | [bibliotheksverbund](#) | [abloesesystem](#) | [verbunddatenbank](#)

GBV 11. [Die Anwendung der überregionalen Personennamendatei in Verbundsystemen : Das Beispiel HBZ-Verbund](#)  
/ Schmidt, Ronald Michael

GBV 12. [50 Jahre Zentralkatalog Nordrhein-Westfalen : 20. März 1947-29. März 1997](#)  
/ Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen <Köln>. - Köln : HBZ Hochschulbibliothekszentrum d. Landes Nordrhein-Westfalen, 1997

\*\*\* 13. [Fortbildungsprogramm : Fortbildung der Beschäftigten der Hochschulbibliotheken und bibliothekarischen Zentraleinrichtungen des Landes Nordrhein-Westfalen](#)  
/ Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen <Köln>. - Köln : HBZ, 1997-

SWB 14. [50 Jahre Zentralkatalog Nordrhein-Westfalen : 20. März 1947 - 20. März 1997](#)  
Köln : HBZ - Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen, 1997

HBZ 15. [Handwerkskammer-Bildungszentrum Münster : HBZ ; Berufsausbildung, Meisterschule, Weiterbildung, Institut für Umweltschutz, Akademie für Unternehmensführung, Akademie für Gestaltung](#)  
Münster, [ca. 1997]

\*\*\* 16. [Weiter auf dem Weg zur virtuellen Bibliothek! : Bibliotheken nutzen das Internet](#)  
/ Jedwabski, Barbara. - 2., erw. Aufl.. - Dortmund, 1996

GBV 17. [Weiter auf dem Weg zur virtuellen Bibliothek! : Bibliotheken nutzen das Internet](#)  
/ INETBIB-Tagung <1, 1996, Dortmund>. - Dortmund, 1996

HBZ 18. [Weiter auf dem Weg zur virtuellen Bibliothek! : Bibliotheken nutzen das Internet ; Fortbildungsveranstaltung im Rahmen der Fortbildung des HBZ](#)  
Dortmund : Univ.-Bibliothek, 1996

GBV 19. [Leihverkehrsrecherchen in der PICA- und HBZ-Datenbank](#)  
/ Hilger, Horst

HBZ 20. [Die bibliographischen Hilfsmittel des HBZ: Prinzipien der bibliographischen Verzeichnung und ihr Nutzen für den Leihverkehr](#)  
/ Lingemann, Sabine. - Köln, 1995

# KaViA – Titel-Clustering

Suchen | Suchergebnis | Erweiterte Suche | Zwischenablage | Benutzer-Info | Hilfe © 1998-2006 OCLC PICA

suchen [und] | Alle Wörter [ALL] | ? | sortiert nach Erscheinungsjahr | unscharfe Suche

suchmaschine | suchen | Alle Verbünde

Benutzergruppe: 8000 | IP : 195.37.139.61

Suchgeschichte | Kurzliste | Titeldaten | Nachweisinformationen

Download  
Trefferanalyse  
Abmelden

Ihre Aktion suchen Gefunden werden im Internet 1 - 4 von 4

siehe auch: [hessenmedia](#) | [alpar](#) | [pfuhl](#) | [gefunden](#) | [landesentwicklung](#)

- GBV 1. [Gefunden werden im Internet](#)  
/ Alpar, Paul. - 2. Aufl. - Wiesbaden : Hessisches Ministerium für Wirtschaft, Verkehr und Landesentwicklung, 2005
- GBV 2. [Gefunden werden im Internet](#)  
/ Alpar, Paul. - 1. Aufl. - Wiesbaden : Hessisches Ministerium für Wirtschaft, Verkehr und Landesentwicklung, 2002
- HBZ 3. [Gefunden werden im Internet](#)  
/ Alpar, Paul. - 1. Aufl.. - Wiesbaden : Hessisches Ministerium für Wirtschaft, Verkehr und Landesentwicklung, 2002
- SWB 4. [Gefunden werden im Internet](#)  
/ Alpar, Paul. - 1. Aufl.. - Wiesbaden : Hess. Ministerium für Wirtschaft, Verkehr u. Landesentwicklung, 2002

Suchen | Suchergebnis | Erweiterte Suche | Zwischenablage | Benutzer-Info | Hilfe © 1998-2006 OCLC PICA

suchen [und] | Alle Wörter [ALL] | ? | sortiert nach Erscheinungsjahr | unscharfe Suche

suchen | Alle Verbünde

Benutzergruppe: 8000 | IP : 195.37.139.61

Suchen | Suchergebnis | Erweiterte Suche | Zwischenablage | Benutzer-Info | Hilfe © 1998-2006 OCLC PICA

suchen [und] | Alle Wörter [ALL] | ? | sortiert nach Erscheinungsjahr | unscharfe Suche

suchen | Alle Verbünde

Benutzergruppe: 8000 | IP : 195.37.139.61

Suchen | Suchergebnis | Erweiterte Suche | Zwischenablage | Benutzer-Info | Hilfe © 1998-2006 OCLC PICA

suchen [und] | Alle Wörter [ALL] | ? | sortiert nach Erscheinungsjahr | unscharfe Suche

suchen | Alle Verbünde

Benutzergruppe: 8000 | IP : 195.37.139.61

Suchgeschichte | Kurzliste | Titeldaten | Nachweisinformationen

Ihre Aktion suchen Gefunden werden im Internet

Quelle: HBZ  
Titel: [Gefunden werden im Internet](#) / Paul Alpar; Markus Pfuhl, [Philipps-Universität Marburg, Ins Wirtschaftsinformatik]. - 1. Aufl. - Wiesbaden : Hessisches Ministerium für Wirtschaft, Ve 2002  
Schriftenreihe: Hessen-Media ; Bd. 32  
ISBN: 3-933732-32-8 (geh., EUR 10.00)

Besitzende Bibliothek(en): Klicken Sie auf einen Bibliotheksnamen oder auf **Detailansicht**

[Koeln, Universitaet zu Koeln Bibliothek](#) <38>

# KaViA – Verbundübergreifende FL

Suchen | Suchergebnis | Erweiterte Suche | Zwischenablage | Benutzer-Info | Hilfe © 1998-2006 OCLC PICA

**KaViA**  
GBV

Abmelden

Suchgeschichte | Kurzliste | Titeldaten | Nachweisinformationen

Leihbestellung: 00074711

Die Bestelldaten sind korrekt

Titel:	Gefunden werden im Internet
Autor:	Alpar,Paul
Körperschaft:	
Jahr:	2002
ISSN:	
ISBN:	3-933732-32-8
Externe DB (ID):	port:HBZ01
Externe DB (NR):	HT013951637

Bestellparameter

Bemerkung:

Lokale Bestellnr.:

Benutzerdaten

Name:	Markus Mustermann
Adresse:	
Postleitzahl:	Keine Straße 1000
Ort:	12345
Email:	
Fax:	
FTP-aktiv:	Ort

Adressformat: [ftp://]user:password@host[:port][path]

Lieferantenliste

HBZ

Letzte Bestätigung

bestellen  Geben Sie Ihr Passwort ein und schicken Sie dann dieses Formular ab.


# KaViA – Suchvorschläge

- Vorschlagsfunktion: Analyse des Treffersets liefert weitere relevante Suchbegriffe:  
Suchmaschine → onlinerecherche, google, spider, rankingverfahren, metasuchmaschine, ...

The screenshot shows the KaViA search interface. At the top, there are navigation tabs: Suchen, Suchergebnis, Erweiterte Suche, Zwischenablage, Benutzer-Info, Hilfe. The search bar contains 'suchen [und]' and 'Alle Wörter [ALL]'. The search results are sorted by 'Relevanz'. The search term is 'suchmaschine'. Below the search bar, there are tabs for 'Suchgeschichte', 'Kurzliste', 'Titeldaten', and 'Nachweisinformationen'. The search results are displayed in a list format. A red circle highlights the 'siehe auch' section, which contains the following links: [suchmaschine](#), [onlinerecherche](#), [spider](#), [suchmaschinen](#), [google](#), [literaturrecherche](#), [informationsrueckgewinnung](#), [informationsretrieval](#), [informationswiedergewinnung](#), [informationsrecherche](#). The search results list includes 10 items, each with a GBV logo and a title. The first item is 'Recherche d'information distribuée: approches pour la selection de collections et la fusion de listes de résultats' by Rasolofo, M. Omega Yves. The second item is 'UNICO-WebRoboter : Konzept einer spezialisierten Suchmaschine' by Nitsche, Dirk. The third item is 'The Google story : [inside the hottest business, media and technology success of our time]' by Vise, David A.. The fourth item is 'Forschungsportal 2nd Edition : Sachbericht ; Projektzeitraum: 15. Feb. 2001 - 28. Feb. 2005' by Hennig, Dirk. The fifth item is 'Forschungsportal 2nd Edition : Sachbericht ; Projektzeitraum: 15. Feb. 2001 - 28. Feb. 2005' by Hennig, Dirk. The sixth item is 'Fuzzy-Suchmethoden im Information-Retrieval' by Homann, Ingo R.. The seventh item is 'Benutzerprofile für die Anfrageverarbeitung in verteilten Digitalen Bibliotheken' by Schmitt, Bethina. The eighth item is 'Fuzzy-Suchmethoden im Information-Retrieval' by Homann, Ingo R.. The ninth item is 'Challenging the invisible web : improving web meta-search by combining constraint-based query translation and adaptive user interface construction' by Huang, Lieming. The tenth item is 'Google hacks : [100 industrial-strength tips & tools]' by Calishain, Tara.

# KaViA – Setanalyse

Suchen | **Suchergebnis** | Erweiterte Suche | Zwischenablage | Benutzer-Info | Hilfe © 1998-2006 OCLC PICA

 suchen [und]  ? sortiert nach  unscharfe Suche

suchen

Benutzergruppe: 8000 | IP : 195.37.139.61









**Suchgeschichte** | **Kurzliste** | **Titeldata** | **Nachweisinformationen**

Abmelden

## Trefferanalyse

Sie können das Suchergebnis nach folgenden Kriterien analysieren:

- ▶ [Materialart](#)
- ▶ [Sprachcode](#)
- ▶ [Ländercode](#)

Materialart	Wert	Treffer
 <a href="#">Aufsätze</a>		38
 <a href="#">Bücher</a>		278
 <a href="#">Tonträger</a>		1
 <a href="#">Online Ressourcen (ohne Zeitschr.)</a>		33
 <a href="#">Online Zeitschriften</a>		2
 <a href="#">Datenträger</a>		14
 <a href="#">Zeitschriften/Serien (ohne Online-Zeitschr.)</a>		2
 <a href="#">Filme, Videos, etc.</a>		14

max: 278 insgesamt: 382

# KaViA – Ausblick

- Regelmäßige Aktualisierung der Verbunddaten
- Einbeziehung weiterer Verbünde in KaViA
  - Integration in den Index
  - Über Z39.50 (PSIport → Hybride Suchmaschine)
- Integration weiterer Datenbanken
  - Nationallizenzen und ZVDD für Digitalisierungen
  - Online Contents: 25.000.000 Artikel
  - Anreicherung der Daten mit TOC für Monos etc.
- Verbesserung der Cluster-Verfahren
  - Ergebnisse der AG Matchkey
  - Abgleich der Verbunddatenbanken: Titelerkennung
- Verbesserung der thematischen Erschließung
  - Navigation über DDC
  - Abgleich der Verbunddatenbanken: Sacherschließung

Herzlichen Dank für ihre  
Aufmerksamkeit