Bielefeld University Library

DLF Spring Forum 2004

Session: Search Engine Technology and Digital Libraries

Beyond Digital Libraries - The Use of Search Engine Technology to Create Next Generation Scholarly Portals

Norbert Lossau, Bielefeld UL
Bjørn Olstad, FAST
Friedrich Summann, Bielefeld UL

DLF Spring Forum, 21 April 2004

Our Session

- *Part 1: Digital library / scholarly portals and search engine technology (Norbert Lossau, Bielefeld UL)*
- Part 2: State-of-the-art search engine technology and their future challenges (Bjørn Olstad, Fast)
- Part 3: From theory to praxis: early implementations of SE-technology (Friedrich Summann, Bielefeld UL)

Bielefeld University Library

DLF Spring Forum 2004

Bielefeld University Library, Germany – background information

- Located in North Rhine-Westphalia State (Mid-Germany), between Hanover and Cologne
- University: 14 faculties with a mission to interdisciplinarity (such as the “Centre for Interdisciplinary Research”, ZIF)
- New Research Centre for Biotechnology, Genome Research, Bielefeld History School
- Library founded in 1967
- Active registered users in 2003: 27,000
- 180 FTEs and PTEs

One central on-site library divided into groups of subject libraries

2,675 reader workplaces

2 Mio books and other media items, the majority on open shelves

Committed to Innovation – The Development of Electronic Services

- 1977 Implementation of *IBAS* as the first **online cataloguing system** in Germany
- 1988 **CD-ROM OPAC**, the first German university library catalogue on CD-ROM (OPAC: Online Public Access Catalogue), use of PC networks.
- 1993 Introduction of the document delivery system **JASON** (**J**ournal **A**rticles **S**ent **O**n **D**emand), collaboration of North Rhine-Westphalian university libraries.
- 1998 – 2001 Main coordinator of the **Digital Library North Rhine-Westphalia** (a major grant of the NRW State Ministry)

Electronic Services in Operation

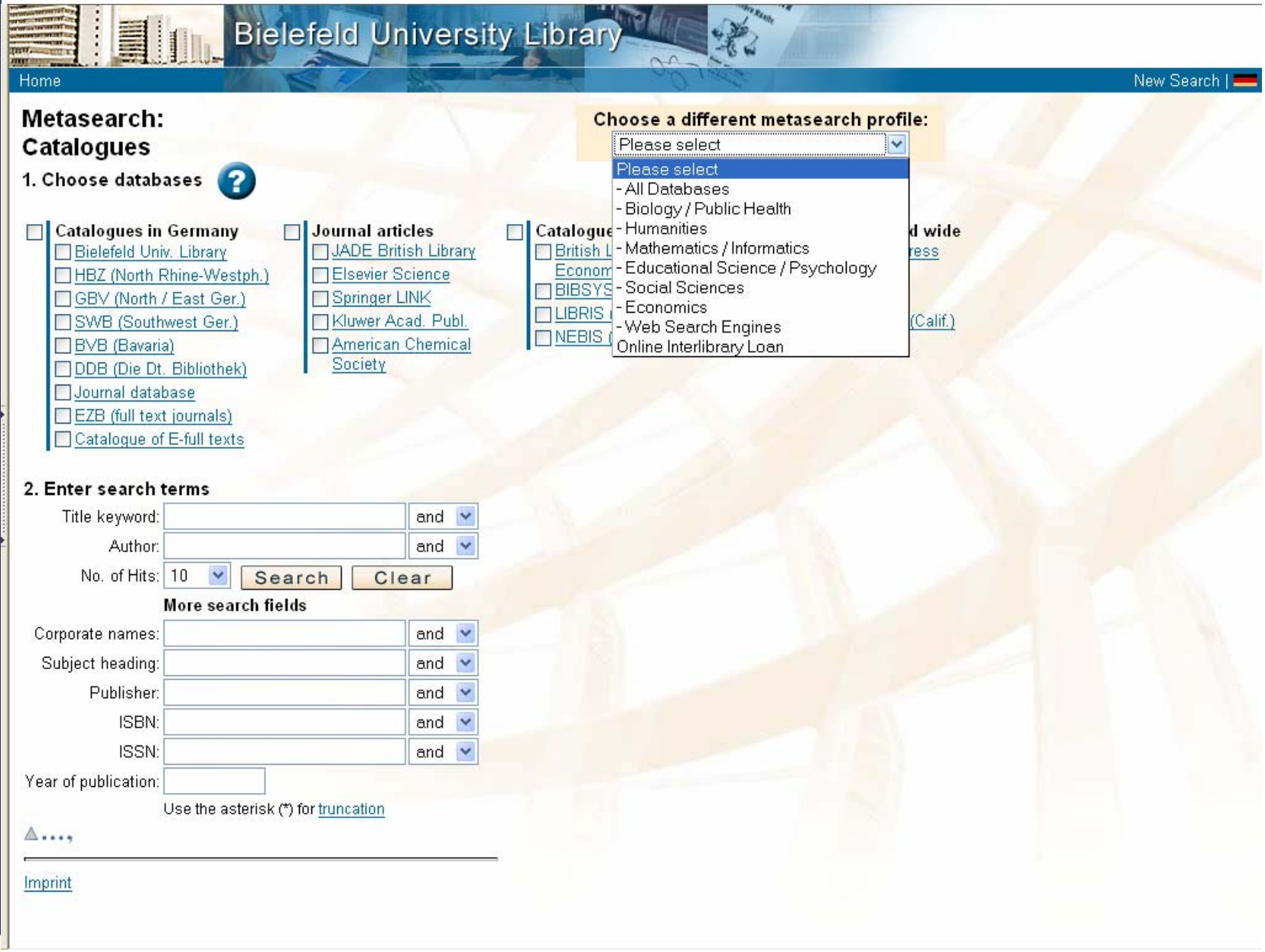
- *MetaSearch* (Digital Library Portal)
- Journal articles database (*JADE*, more than 25Mio articles)
- Full text mirror site of Elsevier (until 03/2004), Springer and Kluwer with more than 1,5 TB of data
- Digitised collection of 18th/19th century German key scientific review organs and literary magazines (2004: 1/2 Mio, 2006: 1 Mio pages)
- eScholarship repository (*BieSOn*)
- Access to state-wide (JASON) and national (SUBITO) document delivery services

Digital Library North Rhine-Westphalia – Bielefeld UL's *MetaSearch*

- State-wide system developed from 1998-2001 (IPS-product from IHS Inc.)
- Local implementation of the Digital Library North Rhine-Westphalia (integrates BRS full text database)
- Simultaneous search across more than 50 core library resources
- Clustering of resources by content type and subject

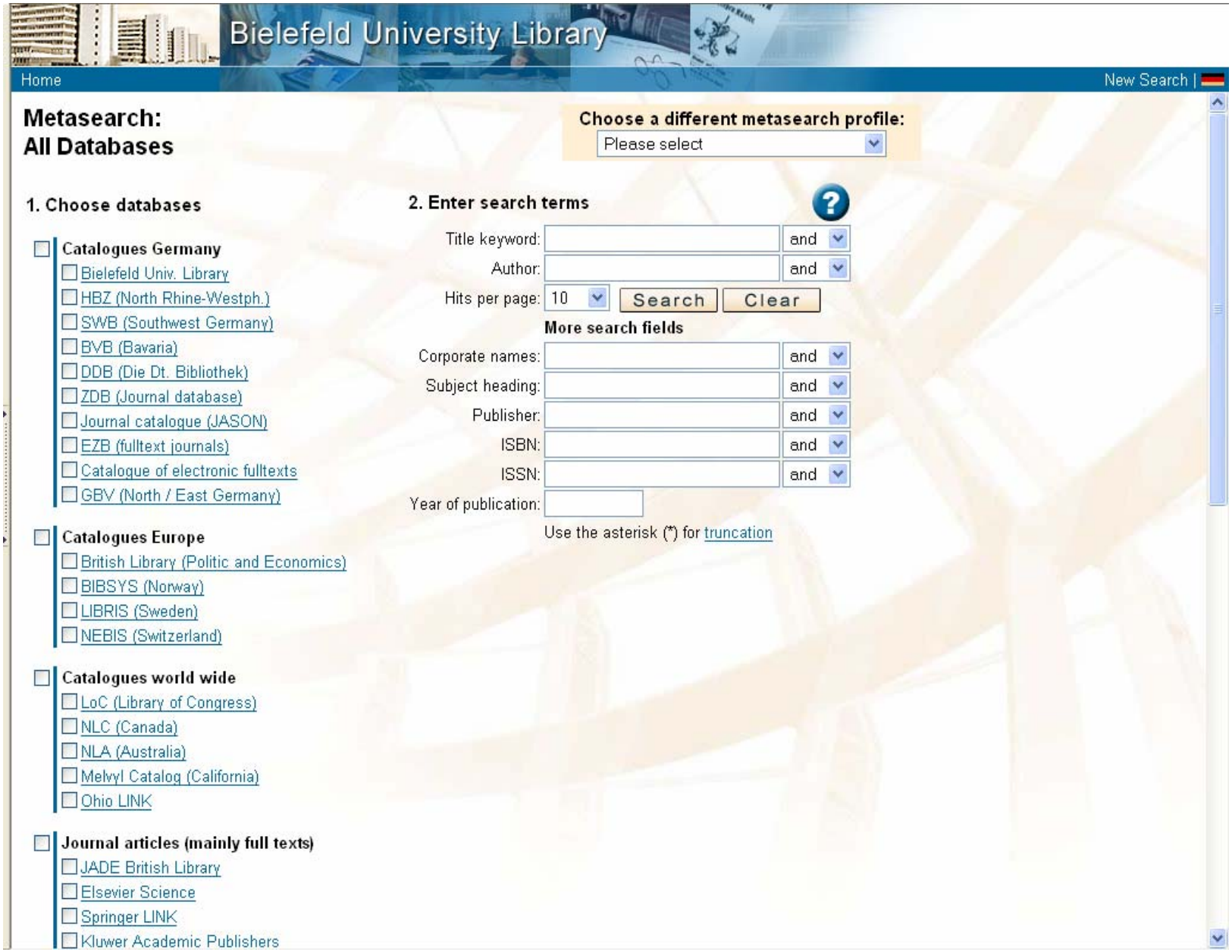
Bielefeld University Library

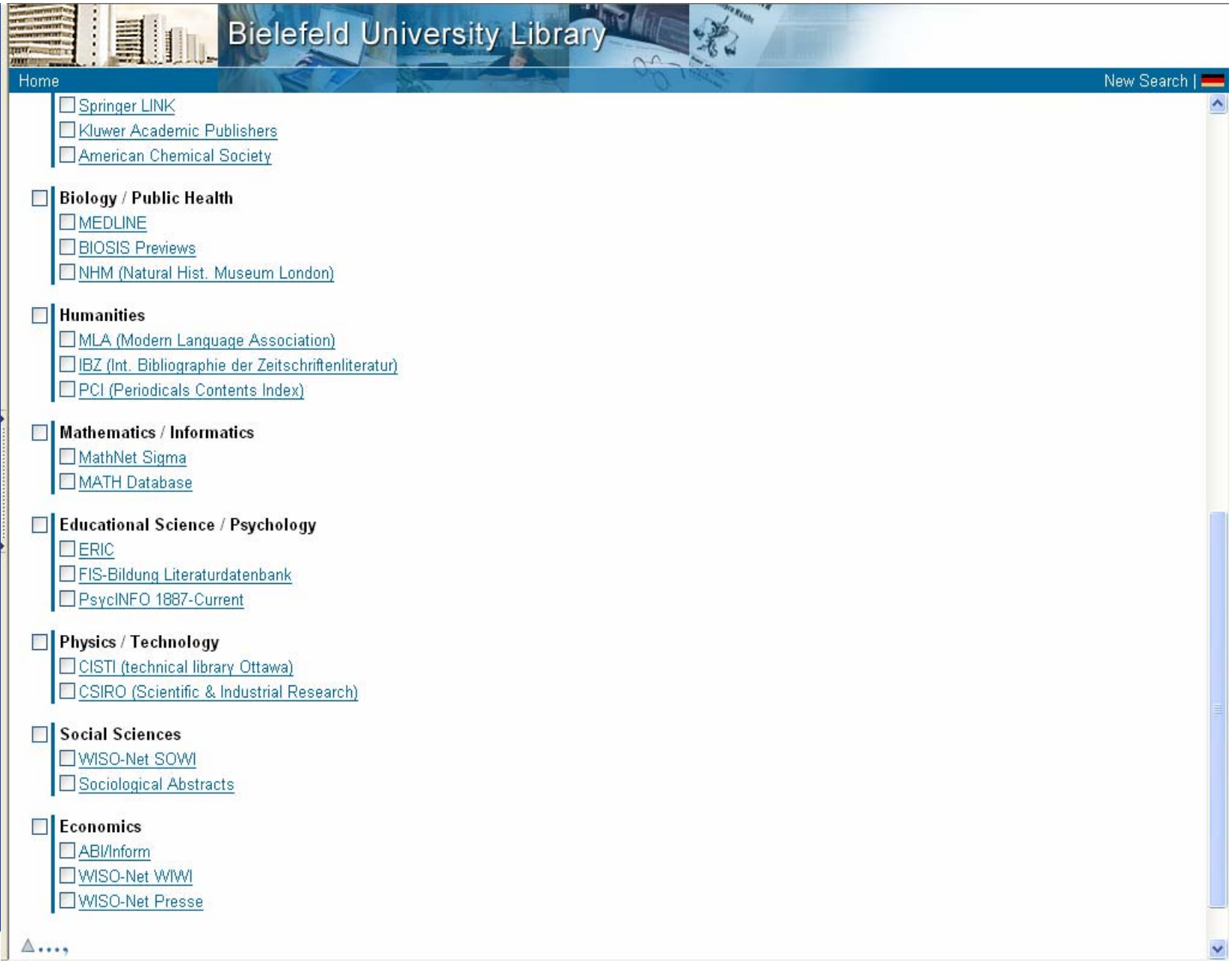
DLF Spring Forum 2004

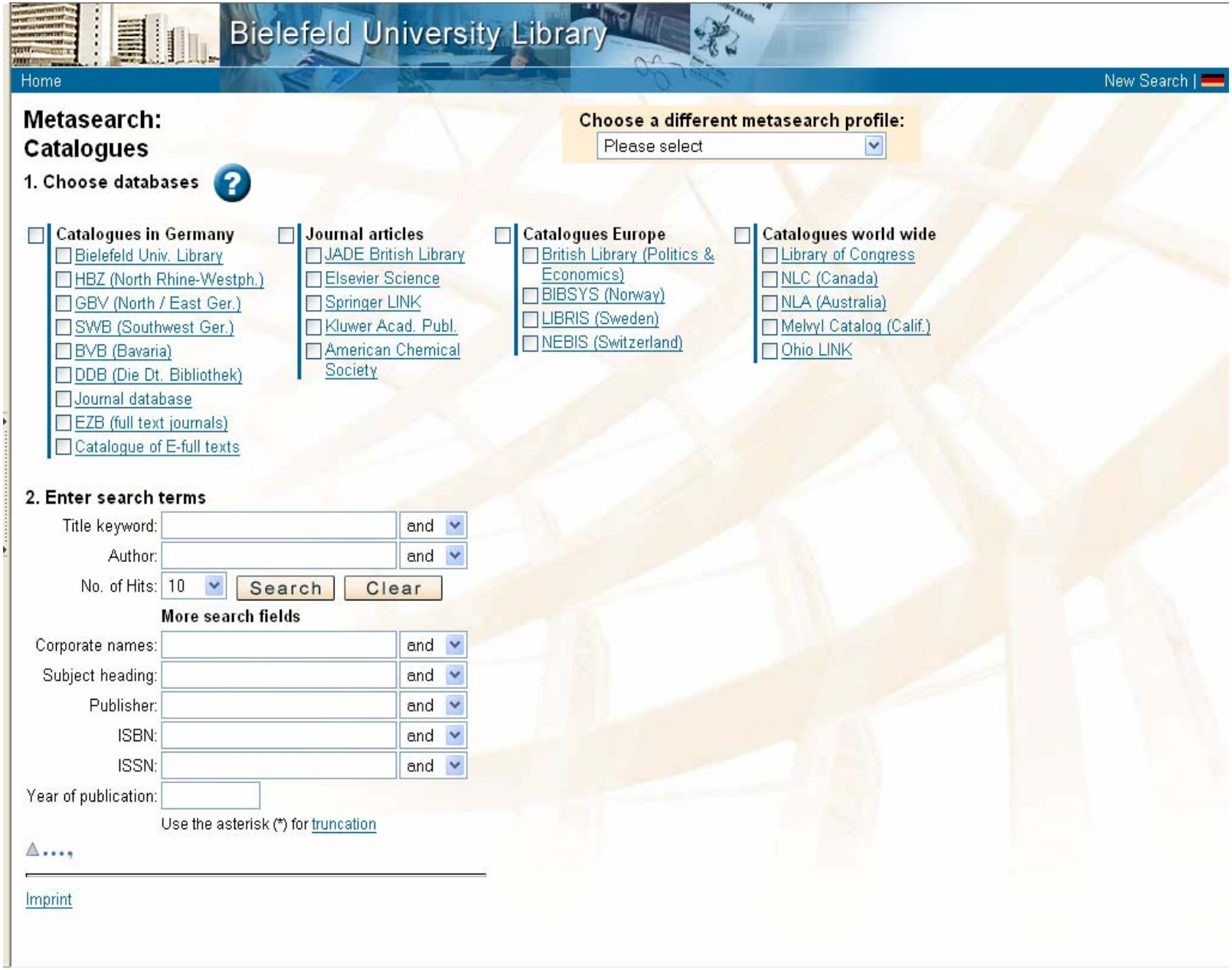

Bielefeld University Library
Home
New Search |
Metasearch:
Catalogues
1. Choose databases
Catalogues in Germany
Bielefeld Univ. Library
HBZ (North Rhine-Westph.)
GBV (North / East Ger.)
SWB (Southwest Ger.)
BVB (Bavaria)
DDB (Die Dt. Bibliothek)
Journal database
EZB (full text journals)
Catalogue of E-full texts
Journal articles
JADE British Library
Elsevier Science
Springer LINK
Kluwer Acad. Publ.
American Chemical Society
Choose a different metasearch profile:
Please select
Please select
- All Databases
- Biology / Public Health
- Humanities
- Mathematics / Informatics
- Educational Science / Psychology
- Social Sciences
- Economics
- Web Search Engines
Online Interlibrary Loan
2. Enter search terms
Title keyword:
Author:
No. of Hits:
10
Search
Clear
More search fields
Corporate names:
Subject heading:
Publisher:
ISBN:
ISSN:
Year of publication:
and
Use the asterisk (*) for truncation
Imprint

Bielefeld University Library

DLF Spring Forum 2004




Bielefeld University Library
Home
New Search |
Springer LINK
Kluwer Academic Publishers
American Chemical Society
Biology / Public Health
MEDLINE
BIOSIS Previews
NHM (Natural Hist. Museum London)
Humanities
MLA (Modern Language Association)
IBZ (Int. Bibliographie der Zeitschriftenliteratur)
PCI (Periodicals Contents Index)
Mathematics / Informatics
MathNet Sigma
MATH Database
Educational Science / Psychology
ERIC
FIS-Bildung Literaturdatenbank
PsycINFO 1887-Current
Physics / Technology
CISTI (technical library Ottawa)
CSIRO (Scientific & Industrial Research)
Social Sciences
WISO-Net SOWI
Sociological Abstracts
Economics
ABI/Inform
WISO-Net WIWI
WISO-Net Presse


Bielefeld University Library
Home
New Search |
Metasearch:
Catalogues
Choose a different metasearch profile:
Please select
1. Choose databases
Catalogues in Germany
Bielefeld Univ. Library
HBZ (North Rhine-Westph.)
GBV (North / East Ger.)
SWB (Southwest Ger.)
BVB (Bavaria)
DDB (Die Dt. Bibliothek)
Journal database
EZB (full text journals)
Catalogue of E-full texts
Journal articles
JADE British Library
Elsevier Science
Springer LINK
Kluwer Acad. Publ.
American Chemical Society
Catalogues Europe
British Library (Politics & Economics)
BIBSYS (Norway)
LIBRIS (Sweden)
NEBIS (Switzerland)
Catalogues world wide
Library of Congress
NLC (Canada)
NLA (Australia)
Melvyl Catalog (Calif.)
Ohio LINK
2. Enter search terms
Title keyword:
Author:
No. of Hits: 10
Search
Clear
More search fields
Corporate names:
Subject heading:
Publisher:
ISBN:
ISSN:
Year of publication:
and
Use the asterisk (*) for truncation
Imprint

Metasearch:
Biology / Public Health

Choose a different metasearch profile:
Please select

1. Choose databases

- ☐ **Subject databases**
  - ☐ MEDLINE
  - ☐ BIOSIS Previews
  - ☐ NHM (Natural Hist. Museum London)
- ☐ **Books and journals**
  - ☐ Bielefeld Univ. Library
  - ☐ Catalogue of electronic fulltexts
  - ☐ Journal catalogue (JASON)
  - ☐ HBZ (North Rhine-Westph.)
- ☐ **Journal articles**
  - ☐ JADE British Library
  - ☐ Elsevier Science
  - ☐ Springer LINK
  - ☐ Kluwer Academic Publishers
  - ☐ American Chemical Society

2. Enter search terms

Title keyword: [ ] and
Author: [ ] and
Hits per page: 10 Search Clear

More search fields

Corporate names: [ ] and
Subject heading: [ ] and
Publisher: [ ] and
ISBN: [ ] and
ISSN: [ ] and
Year of publication: [ ]

Use the asterisk (*) for truncation

Imprint

Online Interlibrary Loan

Books

HBZ (North Rhine-Westph.)

Please order articles via JASON.

- Search the book title via the search form.
- After the search has been completed, you receive a list of titles from which you can choose and directly order the desired title.
- **Costs**: Upon placing an order, your loan account is automatically charged 1.50 Euro.
  add. 0.55 EUR postal charge (e-mail notification is **free**!)
  Please pay this fee at the Central Loan Department.

> **Frequently Asked Questions (FAQ)**

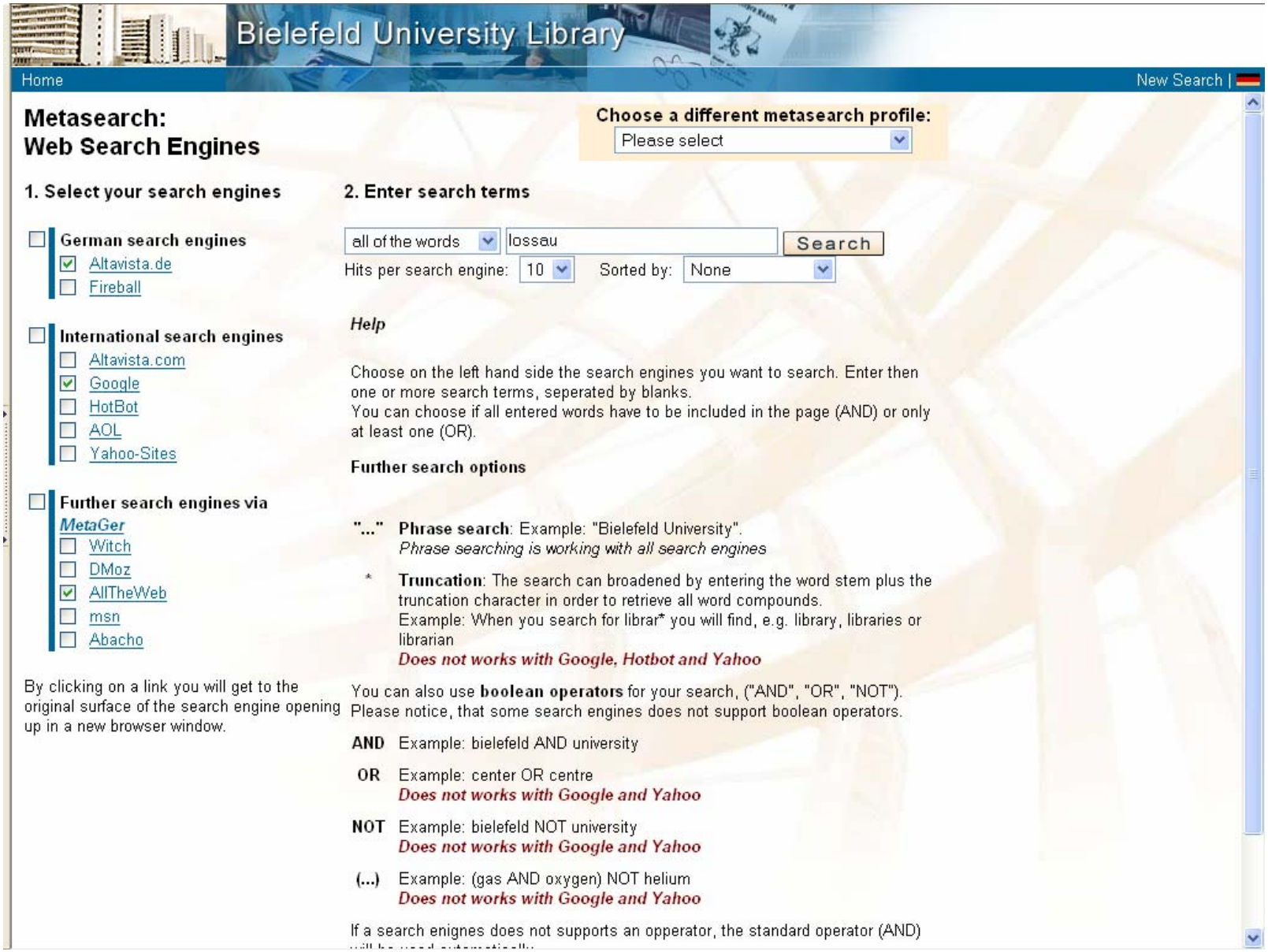
Enter search terms

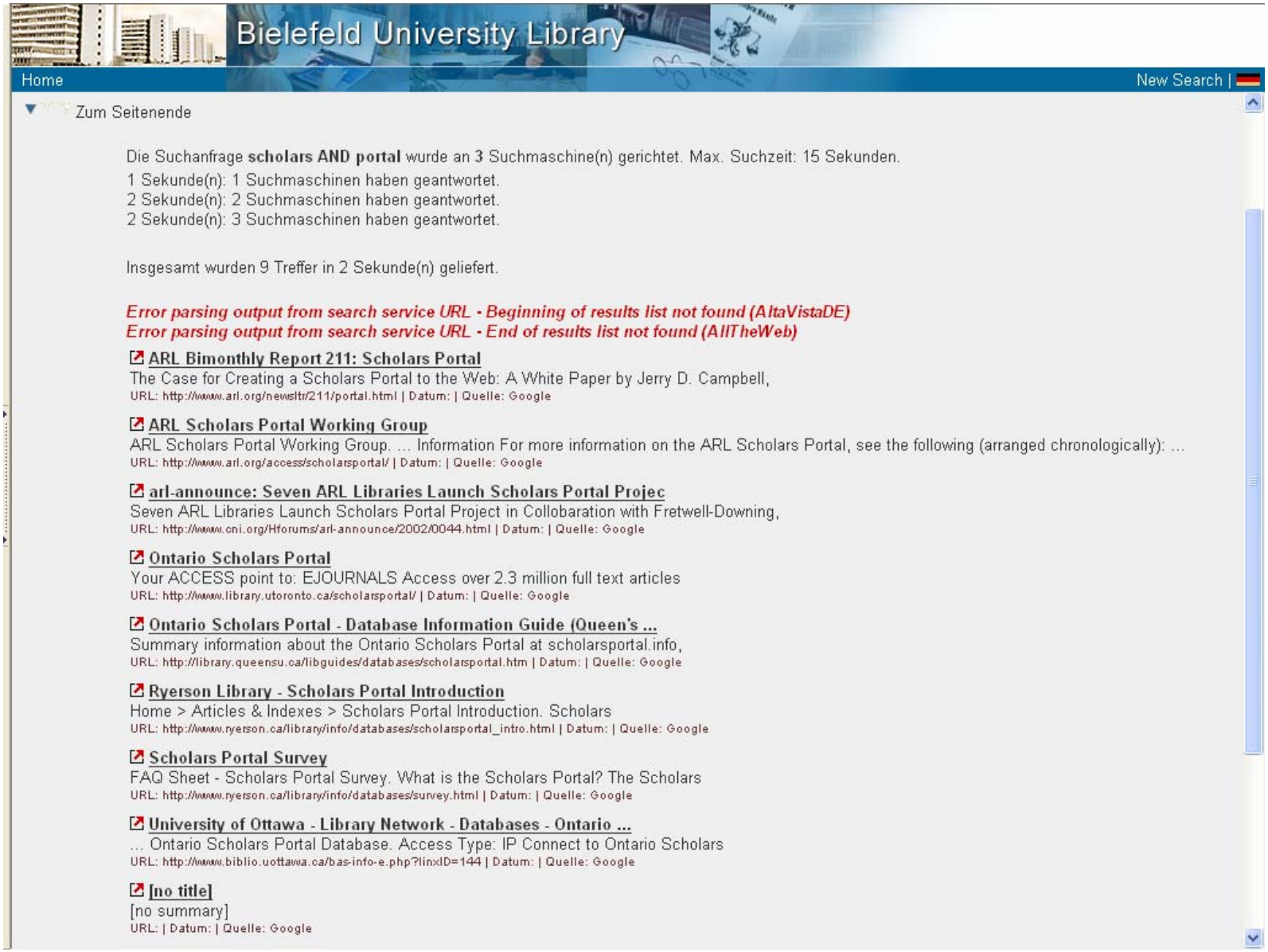
Title keyword:		and
Author:		and
Hits per page:	10	Search Clear

More search fields

Corporate names:		and
Subject heading:		and
Publisher:		and
ISBN:		and
ISSN:		and
Year of publication:		

Use the asterisk (*) for truncation


Bielefeld University Library
Home
New Search
Metasearch:
Web Search Engines
Choose a different metasearch profile:
Please select
1. Select your search engines
German search engines
Altavista.de
Fireball
International search engines
Altavista.com
Google
HotBot
AOL
Yahoo-Sites
Further search engines via
MetaGer
Witch
DMoz
AllTheWeb
msn
Abacho
By clicking on a link you will get to the original surface of the search engine opening up in a new browser window.
2. Enter search terms
all of the words
lossau
Search
Hits per search engine: 10
Sorted by: None
Help
Choose on the left hand side the search engines you want to search. Enter then one or more search terms, seperated by blanks.
You can choose if all entered words have to be included in the page (AND) or only at least one (OR).
Further search options
"..." Phrase search: Example: "Bielefeld University".
Phrase searching is working with all search engines
* Truncation: The search can broadened by entering the word stem plus the truncation character in order to retrieve all word compounds.
Example: When you search for librar* you will find, e.g. library, libraries or librarian
Does not works with Google, Hotbot and Yahoo
You can also use boolean operators for your search, ("AND", "OR", "NOT").
Please notice, that some search engines does not support boolean operators.
AND Example: bielefeld AND university
OR Example: center OR centre
Does not works with Google and Yahoo
NOT Example: bielefeld NOT university
Does not works with Google and Yahoo
(...) Example: (gas AND oxygen) NOT helium
Does not works with Google and Yahoo
If a search enignes does not supports an opperator, the standard operator (AND)


Bielefeld University Library
Home
New Search |
Zum Seitenende
Die Suchanfrage scholars AND portal wurde an 3 Suchmaschine(n) gerichtet. Max. Suchzeit: 15 Sekunden.
1 Sekunde(n): 1 Suchmaschinen haben geantwortet.
2 Sekunde(n): 2 Suchmaschinen haben geantwortet.
2 Sekunde(n): 3 Suchmaschinen haben geantwortet.
Insgesamt wurden 9 Treffer in 2 Sekunde(n) geliefert.
Error parsing output from search service URL - Beginning of results list not found (AltaVistaDE)
Error parsing output from search service URL - End of results list not found (AllTheWeb)
ARL Bimonthly Report 211: Scholars Portal
The Case for Creating a Scholars Portal to the Web: A White Paper by Jerry D. Campbell,
URL: http://www.arl.org/newsltr/211/portal.html | Datum: | Quelle: Google
ARL Scholars Portal Working Group
ARL Scholars Portal Working Group. ... Information For more information on the ARL Scholars Portal, see the following (arranged chronologically): ...
URL: http://www.arl.org/access/scholarsportal/ | Datum: | Quelle: Google
arl-announce: Seven ARL Libraries Launch Scholars Portal Projec
Seven ARL Libraries Launch Scholars Portal Project in Collobaration with Fretwell-Downing,
URL: http://www.cni.org/Hforums/arl-announce/2002/0044.html | Datum: | Quelle: Google
Ontario Scholars Portal
Your ACCESS point to: EJOURNALS Access over 2.3 million full text articles
URL: http://www.library.utoronto.ca/scholarsportal/ | Datum: | Quelle: Google
Ontario Scholars Portal - Database Information Guide (Queen's ...
Summary information about the Ontario Scholars Portal at scholarsportal.info,
URL: http://library.queensu.ca/libguides/databases/scholarsportal.htm | Datum: | Quelle: Google
Ryerson Library - Scholars Portal Introduction
Home > Articles & Indexes > Scholars Portal Introduction. Scholars
URL: http://www.ryerson.ca/library/info/databases/scholarsportal_intro.html | Datum: | Quelle: Google
Scholars Portal Survey
FAQ Sheet - Scholars Portal Survey. What is the Scholars Portal? The Scholars
URL: http://www.ryerson.ca/library/info/databases/survey.html | Datum: | Quelle: Google
University of Ottawa - Library Network - Databases - Ontario ...
... Ontario Scholars Portal Database. Access Type: IP Connect to Ontario Scholars
URL: http://www.biblio.uottawa.ca/bas-info-e.php?linxID=144 | Datum: | Quelle: Google
[no title]
[no summary]
URL: | Datum: | Quelle: Google

Strategic questions I

- How do we define academic relevant (online) content?
- Are we aware of the incredible volume of this content that is available on the web?
- Who among our users (and fellow colleagues) knows about the wealth of digital resources we as libraries/universities make available? (“visibility”)
- Do we consider “Information Search & Navigation” as a mission-critical service for academic libraries?

Strategic questions II

- Will libraries be able to maintain their role as key information service provider at their universities?
- Can we afford to neglect general / mainstream search services?
- Where do we want / need specific local developments and where should we share efforts with colleagues from other institutions?

The vision…

- One virtual resource of unprecedented comprehensiveness to any type and format of academically relevant (online) information
- A resource that is built and maintained by the library community, ensuring long-term access
- A variety of customisable search & navigation interfaces that meet local (institutional), subject specific and individual demands
- A resource with the popular search & system comfort of Google-like services but the relevance and proven quality of content which is accessible through libraries

Digital Libraries and Scholarly Portals

- 1999 ARL : “The Scholars Portal”
- Germany over the last years – “digital library portal” discussion, recent launch of “Vascoda”, the German Digital Library

Collaborative Efforts to Create Single Points of Access to Distributed Repositories

Samples:

- Virtual subject libraries, Subject guides, *Vascoda*, Germany
- RDN, UK
- Renardus, EU
- Scout, U.S.
- The Scholarly Portal, U.S.
- CGM
- *OAI-Registries

The Impact of Internet Search Engines on Libraries

- Information “search” has grown to a significant business sector of a global, competitive and commercial market
- Libraries are only one player among powerful stakeholders
- Usage behaviour is changing
  - The „all-inclusive“ search
  - The “empowered” user
  - Search comfort (Query, performance, result presentation)

Challenges for Search Systems of Current Portal Products

- Coverage of data formats, full text search
- Coverage of content types
- Scalability/ information retrieval performance
- Search comfort
- Display of result lists, relevance ranking and ordering of hits

Search Engine Technology to Build NG Search Services in Scholarly Portals – the Strengths

- “Unlimited” scalability (indexes billions of documents)
- Superior performance
- Handles a wide range of document (text) formats
- Provides comfortable search interfaces and flexible result presentation
- Handles highly structured AND unstructured data

The Shortcomes of “Information.com” Internet Search Indexes

- Purely commercially driven
- No guarantee for long term accessibility of content resources
- Focus on the “visible” web, that can be automatically indexed
- No authoritative assessment of content resources
- “Mix” academic and general content
- Monolithic index architecture – at significant costs

Search Engine Technology – the Myths

- Can’t handle structured, “high quality” data
- Provide only simple search boxes

Commercial Internet Search Indexes vs. Search Engine Technology

- Major objectives against search engines are bound to the business model / concept of commercial indexes
- Build new developments on the “technology” itself and rely on proved principles in the academic information discovery environment (like “quality” vs. take everything, “long-term reliability” vs. ad-Hoc availability, “professional assessment criteria” vs. commercially driven ranking)

Additional Requirements for Search Services in Scholarly Portals

- Indexing only of qualified (certified) content resources
- Advanced navigation / browsing functionality /scientific taxonomies, thesauri, cross-concordances
- Handling of data heterogeneity: mark-up of various content types (e.g. metadata, full text, image, binary data)
- Flexible ranking and ordering schemes for result displays

Specific Requirements for Search Services in Scholarly Portals

- Aggregation and indexing capabilities for any type of academic online information data
- Automatic extraction of metadata

Scholarly Portals – Potential Scenarios

1. Expose data to commercial internet search engines (e.g. OCLC, HBZ/Germany, MIT + other Libraries)
2. Buy-in into commercial internet search engines
3. Take the technology and build a qualified, community driven academic search index (network)

Future Architecture of Scholarly Portals

- Modular system that consists of various service components, a central one being “search & navigation”
- The service concept of current portals can still apply, but single services may be replaced
- Looking beyond the library systems – interdependencies and interaction with other university systems (such as courseware applications, central research database, central user administration)
- Use of open source AND commercial systems
- “Web Services” as (one possible) technology to connect service components

Moving from Theory to Practice: Early Implementations of Search Engine Technology

- Objectives: proof-of-concept, first real-life service

Part 3: Friedrich Summann

Next Steps at Bielefeld University Library

- Upgrade to new FAST ESP 4.0 version (May 2004)
- Implement linguistic tools – module (e.g. approximate search) (May 2004)
- Launch public prototype of subject based search service („Math-Demonstrator“, June 2004)
- Launch first public cross-digital collections search service („nucleus“, July 2004)
- Continue and extend partnership within Germany (*Distributed Document Server*/Vascoda), National Library of Norway, Oxford, open for further partners

Synergies from Related Projects for Additional Services / Features

- CiteSeer
- Vivisimo: ClusterMed (taxonomy, clustering of results with pre-defined or on-the-fly vocabulary)
- Jakarta Lucene
- Grid technology (open network architecture)

Outlook – Collaborative Approach?

- Participate in an open federated search index network
- Potential roles for partners
  - Exposing data (e.g. your digitised collections) to the index engines of partners
  - Take yourself responsibility for a segment of content repositories (e.g. index digitised collections from X universities) and connect your index to the network
  - Make technical tools available to partners (such as data processing scripts/„connectors“ or navigation tools like online taxonomies, classifications)
- Create a critical mass!

Our Session Continues…

- Part 1: Digital library / scholarly portals and search engine technology (Norbert Lossau, Bielefeld UL)
- *Part 2: State-of-the-art search engine technology and their future challenges (Bjørn Olstad, Fast)*
- Part 3: From theory to praxis: early implementations of SE-technology (Friedrich Summann, Bielefeld UL)