SOLR & XAPIAN IN INVENIO

Patrick Glauner¹, Jan Iwaszkiewicz¹, Jean-Yves Le Meur¹, Tibor Simko¹, Nikos Kasioumis²

1/Author - 2/Speaker

Open Repositories 2013
Repository Island, Charlottetown, PEI, Canada
Bulletin du Cercle Vaudois de botanique

Founded in January 1949, the Vaudois Botanical Society or “CVB” (Cercle Vaudois de Botanique) is an association of botanists in the Canton of Vaud. Original articles, observations of the regional...

Institutions

- Neuchâtel (7'221)
- Valais (6'498)
- Fribourg (6'729)
- Jura (3'563)
- Vaud (3'175)
- Geneva (2'545)
- Tessin (577)

Contents

- Articles (14'947)
- Periodicals (5'465)
- Theses (2'978)
- Books (3'903)
- Dissertations (2'468)
- Research reports (339)
- Maps (178)
- Music Scores (74)
- Sound Recordings (46)

Latest news

- RERO DOC : new interface
  2013/11/02

- New collection: Journal « La Gruyère »
  2012/09/12

- RERO DOC & Multiwo : improved integration
  2011/11/21
Recent Uploads

25 June 2013  Dataset  Open access
VAARTIJDENBOEK M.S. CORNELIS PRINS, 1E HELFT: REGELGEVING.
Bijl, Leo
Het vaartijdenboek is bijgehouden tussen 1968 en 1981 [...]
Uploaded by LEO on 27 June 2013.

25 June 2013  Book  Open access
E-Books und E-Reader in Bibliotheken
Mumenthaler, Rudolf
Beiträge von Rudolf Mumenthaler zu den Themen E-Books, E-Reader und Tablets in Bibliotheken, die in Sammelbanden, Fachzeitschriften und in Blogs veröffentlicht worden sind (mehrheitlich auf Deutsch). Rudolf Mumenthalers articles and blogposts on the topics e-books, e-readers and tablets in libraries that have appeared in proceedings, journals and in blogs (partly in English).
Uploaded by m rudolf on 27 June 2013.
Atlantis Blogs
Preserving and disseminating blogs using the BlogForever Platform

Search 39028 records for

Narrow by collections:

Blog (1434)
Comments (0)

Posts (37594)
Pages (0)
ARCHITECTURE
INGESTION
Ingestion

Author

"My great research article"
Author

"My great research article"

Reference Extraction

Keyword Classification

...
Author -> "My great research article"

Reference Extraction

Keyword Classification

... ...

Full-text

Metadata

Processing
Author

"My great research article"

Reference Extraction

Keyword Classification

Full-text

Metadata

Indexing

Ranking

Author Disambiguation

...
Author -> "My great research article"

Reference Extraction

Keyword Classification

Full-text

Metadata

Indexing

Ranking

Author Disambiguation

...
SEARCHING
User

author:ellis AND fulltext:"higgs boson"
User

Query parser

author:ellis AND fulltext:"higgs boson"
User

author:ellis AND fulltext:"higgs boson"

ellis in author

AND operator

"higgs boson" in full-text
User

author:ellis AND fulltext:"higgs boson"

ellis in author

AND operator

"higgs boson" in full-text

Searcher

Database
User

author:ellis AND fulltext:"higgs boson"

ellis in author

AND operator

"higgs boson" in full-text

Searcher

Ranker

author index

citation index

...
FAST SEARCHING VS. SLOW INDEXING
CUSTOM RANKING METHODS
FULLTEXT INDEXING LIMITS
ADVANCED CAPABILITIES
Apache Solr
JAVA
HTTP / XML
VECTOR SPACE TF-IDF

Xapian
C++
LOCAL
PROBABILISTIC BM25
INGESTION
SEARCHING
User

author:ellis AND fulltext:"higgs boson"

ellis in author

AND operator

"higgs boson" in full-text

Searcher

Ranker

Result

Bridge

Apache Solr

Xapian
SCALABILITY
COMPARISON

FULLTEXT:"HIGGS BOSON"
<table>
<thead>
<tr>
<th></th>
<th>Apache Solr</th>
<th>Xapian</th>
</tr>
</thead>
<tbody>
<tr>
<td>RESULTS</td>
<td>16,903</td>
<td>15,945</td>
</tr>
<tr>
<td>SEARCH</td>
<td>0.06</td>
<td>0.56</td>
</tr>
<tr>
<td>RANK 10</td>
<td>0.06</td>
<td>0.40</td>
</tr>
<tr>
<td>RANK 100</td>
<td>0.12</td>
<td>0.42</td>
</tr>
<tr>
<td>RANK 1K</td>
<td>0.22</td>
<td>0.45</td>
</tr>
<tr>
<td>RANK 10K</td>
<td>1.46</td>
<td>0.70</td>
</tr>
</tbody>
</table>
CONCLUSION

OPTIMIZED INDEXING
LIMITATIONS
USE & REUSE
USE CASES
GENERIC BRIDGE
FUTURE
QUESTIONS

invenio-software.org
info@invenio-software.org