



Staatsbibliothek  
zu Berlin  
Preußischer Kulturbesitz

## **„Kombinierte Power von Open-Source: Pazpar2 und Solr auf dem Slavistik-Portal“**

43. Internationale Arbeits- und Fortbildungstagung  
der der ABDOS e.V.

Hannover, den 6. Mai 2014





## Überblick

- ➔ Slavistik-Portal (Genese, Ziele)
  
- ➔ Was bietet das Slavistik-Portal?
  - Fachspezifische Informationen
  - Schnelle und benutzerfreundliche Suche
  
- ➔ Ausblick





## Slavistik-Portal

- ➔ Entstanden als Virtuelle Fachbibliothek Slavistik, ein von der DFG gefördertes Projekt (2005-2011)
- ➔ seit Mai 2007 online
- ➔ baut(e) auf Sondersammelgebiet (SSG) Slawistik auf
- ➔ richtet sich an
  - Wissenschaftler und Studierende der Slawistik, Lehrer, Übersetzer etc.
- ➔ bietet Zugriff auf fachrelevante Literatur ohne Medienbrüche
- ➔ im Internet unter

[www.slavistik-portal.de](http://www.slavistik-portal.de)



## Fachspezifische Informationen

- ➔ Internetquellen
- ➔ Fach- und Spezialbibliographien
- ➔ Regionalbibliographien
- ➔ Kataloge
- ➔ Volltextdatenbanken
- ➔ Datenbanken mit Nationallizenz



## Parallele Suche

- ➔ Neues Suchsystem, seit Januar 2013 online
- ➔ Basiert auf der Suchmaschinentechologie SOLR in Kombination mit der Metasuchsystem PAZPAR2





► **Suche**

- Aktuelles
- Neuerwerbungen
- Zeitschriften
- Datenbanken
- Online-Tutorium
- Wir über uns

## Suche

in slawistischen Datenbanken, Bibliographien und Katalogen

SUCHE

?

Datenbanken-Info

Sortierung nach  Anzeige  Treffer pro Seite.

Ein Projekt der



Gefördert von der



Partner von



Folgende Quellen können durchsucht werden:

**Slavistik-Guide - Fachinformationsführer Slavistik**

**BibSlavArb - Bibliographie slavistischer Arbeiten aus deutschspr. Fachzeitschriften 1876-1983**

**MatSlavBib - Materialien zu einer slavistischen Bibliographie 1945-1983**

**BibSlavPub - Bibliographie slawistischer Publikationen aus der DDR 1946-1986**

**BibDatSlav - Bibliographie der deutschsprachigen Slavistik 1983-1992**

**OlBiSlav - Online-Bibliothek der deutschsprachigen Slavistik 1993-2006**

**BibSlavKon - Bibliographie der internationalen Slavistenkongresse 1929-2003**

**BibCzechLing - Bibliographie der tschechischen Linguistik**

**KempgenDB - Slavistische Arbeitsbibliographie von Prof. Sebastian Kempgen**

**EBSEES - Europäische Bibliographie zur Osteuropaforschung**

**RussGus - Regionalwissenschaftliche Datenbank**

**SorBib - Sorbische Bibliographie (Sorbisches Institut Bautzen)**

**GVK - Gemeinsamer Verbundkatalog**

**OLC Slavistik - Online Contents Slavistik**

**UDB-EDU - Universal Database of Social Sciences & Humanities**

**PSSS - Primary Sources for Slavic Studies**

und viele andere ...



- ▶ **Suche**
- ▶ Aktuelles
- ▶ Neuerwerbungen
- ▶ Zeitschriften
- ▶ Datenbanken
- ▶ Online-Tutorium
- ▶ Wir über uns

## Suche

in slawistischen Datenbanken, Bibliographien und Katalogen

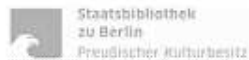
?
Datenbanken-Info

Sortierung nach  Anzeige  Treffer pro Seite.

<< 1 2 3 4 5 6 7 8 9 10 11 12 13 ... >>

1 bis 100 von **2282+**

Ein Projekt der



Gefördert von der



Partner von



**Die Sud-Slawen und deren Lander in Beziehung auf Geschichte, Cultur und Verfassung. Von J. F. Neigebaur**

<b>Titel</b>	Die Sud-Slawen und deren Lander in Beziehung auf Geschichte, Cultur und Verfassung. Von J. F. Neigebaur
<b>Verfasser</b>	Neigebaur, J. F.
<b>Jahr</b>	1851
<b>Sprache</b>	ger
<b>Verlag</b>	Costenoble, Rimmelman
<b>Umfang</b>	391 c
<b>Mediumart</b>	electronic
<b>URL</b>	PDF
<b>Quelle</b>	RSL (Digital)
<b>Export</b>	Text-Format, RIS, BibTex, XML, DublinCore

### Kataloge

- BibDatSlav (200+)
- BibSlavArb (200+)
- BibSlawPub (200+)
- EBSEES (200+)
- KemppenDB (200+)
- Litdok (200+)
- MatSlavBib (200+)
- OLC Slawistik (200+)
- OLBiSlav (200+)
- GVK (100+)
- RussGus (81)
- SLK (76)
- BibSlavKon (73)
- SorBib (66)
- OAI-Data (24)
- PAO (18)
- GBooks (Slavica) (16)
- Archive.org (Slavica) (13)
- Slavistik-Guide (11)

**Hans-Ulrich Voss (Berlin): "Die vielen Väter Österreichs. Römer - Germanen - Slawen. Eine Spurensuche" (Book Review) Herwig Friesinger**

**Historya Słowian : nasze podanie dziejowe**

**Die Sage von Heinrich dem Löwen bei den Slawen**

**Die ersten Begegnungen der Slawen mit dem Christentum und den christlichen Völkern vom VI. bis VII. Jahrhundert. Die Slawen zwischen Byzanz und Abendland**



## Solr-Technologie

- ➔ Ist ein Projekt der Apache Software Foundation, das auf der Open Source Suchmaschinentechologie LUCENE basiert
- ➔ Geschrieben in plattformunabhängigen Sprache Java
- ➔ Entwickler: Doug Cutting, Otis Gospodnević & Co.
- ➔ Ein Werkzeug für Indizierung und schnelle Durchsuchung strukturierter Daten sowie Datentypen (arbeitet in erster Linie mit XML, aber auch PDF, TXT, DOC u.a.)
- ➔ Erlaubt die Sortierung der Ergebnisse nach ihrer Relevanz
- ➔ Bietet Einsatz von Facetten (Mediumart, Themen/Schlagwörter, Klassifikationen, Jahre, Autoren usw.)
- ➔ Bietet eine Reihe flexibler Clients JSON, XML, PHP u.a.





## Solr-Technologie auf dem Portal

- ➔ Einsatz auf dem Portal seit Mitte 2009
- ➔ Solr-Server im Einsatz
- ➔ Umfangreiche Erfahrungen in der Content-Aufbereitung
- ➔ Indizierung von bibliographischen Daten (XML-Format) aus Retrokonversionsprojekten
  - Stemming-, Stopword- und SynonymFilter
  - MappingCharFilterFactory (Suche kyrillisch und lateinisch)
  - ASCIIFoldingFilterFactory (Suche mit und ohne Diakritika)
  - AnalyzingQueryParser (Kombination von Trunkierung und Suchbegriffsanalyse) – (Source Code angepasst)
- ➔ Solr und Lucene-Indizes sind kompatibel
  - JSP-Scripte, die Lucene-Bibliothek bedienen, kommen zum Einsatz für spezielle Belange (Fuzzy-Search via URL-Parameter)



## Pazpar2

- ➔ Ist eine Metasuchmaschine, die von dem dänischen Unternehmen „Index Data“ entwickelt und unter GPL vertrieben wird (auch Entwicklung von Yaz-Software)
- ➔ Entwickler: Mark Taylor & Co.
- ➔ Kennzeichen: geschrieben in C+, Server XML-basiert, Konfiguration über XSLT, Webclient arbeitet mit JavaScript; sehr anpassungsfähig
- ➔ Ist in der Lage die wichtigsten bibliothekarischen Schnittstellen zu bedienen: Z39.50, SRU/SRW und Solr-XML
- ➔ Bietet Einsatz von Facetten (Solr-Facetten können durchgereicht werden)
- ➔ Ausgefeilte Suchsyntax (Korrespondierend mit Solr)
- ➔ Eine aktive Webclient-Entwicklung findet an der SUB Göttingen durch Sven Prost statt



## Pazpar2 auf dem Portal

- ➔ Pazpar2-Server im Einsatz
- ➔ Webclient ist sehr leichtfüßig, ca. 1000 Zeilen lang
- ➔ Typo3 und Pazpar2 sind getrennt, was eine besondere gesteigerte Performance erlaubt
- ➔ Einsatz von Javascript(-Frameworks) bedeutet enorme Flexibilität bei der Ausgestaltung sowohl des Webclients als auch eines benutzerfreundlichen Designs
- ➔ Ajax-Elemente: Google-Cover, EZB-Verfügbarkeit, Event-Statistiken, Authentifizierungsmechanismen (im Aufbau)
- ➔ Vor allem: Freiheit im Bereich Content-Aufbereitung (das was man gewöhnlich nicht sieht ...)



## Welche Daten kommen über welche Schnittstelle rein?

### ➔ Solr-Indizes

- Slavistik-Guide: XML-Abzug->XML-Konversion
- Slavistische Bibliographien: SQLite-Datenbanken->XML-Konversion
- Slavistische Bibliographien II: TXT/XML-Abzug->XML-Konversion
- Slavistische Bibliographien III: MySQL-Abzug->XML-Konversion
- OAI-Metadata: OAI-Harvesting->XML-Konversion
- Europeana-OpenData: RDF/XML-StreamParsing/Konversion (demnächst mit SPARQL-Schnittstelle)
- RSL-Digital: Z39.50-Harvesting->XML-Konversion
- Google-Slavica: Google-API->SQLite-DB->XML-Konversion
- Slavica-Archive.org: Solr-XML->XML-Bereinigung+Konversion
- Otto-Sagner-Digital: Solr-XML->XML-Konversion
- BASE (Slavica): Solr-XML->Filter->XML-Konversion
- UDB-EDU (NLZ): XHTML->XML-StreamParsing/!Bereinigung!/Konversion



## Welche Daten kommen über welche Schnittstelle rein?

### ➔ Solr-Indizes

- OLC-Slavistik: mntl. MARC-Datenabzug->Yaz-XMLDump->XML-Bereinigung->XML-Konversion (Einsatz über LUCENE-URL-Fuzzy)
- UBO: Csv-Abzug->XML-Anreicherung+Konversion

### ➔ Pazpar2 direkt

- PSSS (NLZ): VZG-Z39.50 (Datenabzug bereits erhalten)
- PAO (NLZ): VZG-Z39.50
- Slovanská knihovna (SLK): Z39.50
- GVK: VZG-Z39.50



## Und wie kommen sie wieder raus?

### ➔ Export-Funktion

- Benutzer wollen die aufgefundenen Daten komfortabel mitnehmen
- Web-Client ist XML-basiert
- Daher: Export-Formate jeder Art können leicht implementiert und angepasst (!) werden
- Momentan implementiert: Text, RIS, BibTex, XML-RDF, DublinCore
- RIS ist für den Einsatz im Zotero oder Citavi oder EndNote ausgelegt

### ➔ Feedback sehr positiv



## Vorteile zusammengefasst

- ➔ Software ist Open Source
- ➔ Rege Entwickler-Community
- ➔ Kosten minimal
- ➔ Sehr flexibel, erlaubt schnelles Eingreifen (statt eines monatelangen Wartens auf das nächste Update eines Discovery-System-Anbieters)
- ➔ Design nach Erfordernissen der Benutzer leicht anpassbar
- ➔ Content in unterschiedlichen Formaten handelbar
- ➔ Retrieval individuell gestaltbar
- ➔ Fazit: Ideal für kleine, hochspezialisierte Portale wie das der Slawisten



## Ausblick

- Im Aufbau: Mobile-Version mit responsiven Webdesign
- Aufbau einer slawistischen Synonyme-Datenbank (für den Einsatz in Solr.SynonymFilterFactory)
- Neues Verfahren der Content-Gewinnung
  - Indizierung von einzelnen PDF-Seiten (Solr-Tika) von frei verfügbaren bibliographischen Werken mit der Erstellung von Fragmenten (Snippets) zur Anzeige in den Metadaten
    - (z.B. Bibliografija staroruskoj literatury, Puškinskij Dom)
  - Indizierung einzelner hochwertiger HTML-Seiten (mit überwiegend bibliographischen Angaben)
    - (z.B. Sajt Ekateriny Kislovoj)
- Ausbau der Kompetenzen im Bereich Informationsvermittlung, Benutzerschulungen; Kooperation mit Berliner Slawistik





Staatsbibliothek  
zu Berlin  
Preußischer Kulturbesitz

Vielen Dank für Ihre Aufmerksamkeit!

Vladimir Neumann  
Osteuropa-Abteilung, SBB

[Vladimir.Neumann@sbb.spk-berlin.de](mailto:Vladimir.Neumann@sbb.spk-berlin.de)

Tel. 030 – 266 435640

Partner von

