



hbz

Wissen. Information. Innovation.

Dr. Silke Schomburg, Jörg Prante

Search Engine Federation in Libraries Suchmaschinenföderation in Bibliotheken

3 - 5 February 2009 in Bielefeld, Germany

9th International Bielefeld Conference

Übersicht

- Situation
- Reaktionen der Hersteller von Bibliothekssystemen
- Reaktionen der Bibliotheksverbände
- Reaktionen der Deutschen Forschungsgemeinschaft
- Herausforderungen
- Was ist eine Suchmaschinenföderation?
- Technik der hbz-Suchmaschinenföderation
- Perspektiven

Situation

- Bestände in wissenschaftlichen Bibliotheken Deutschlands sind zusammenhängend nicht in einer Suchmaschine verfügbar
- bisher keine deutschlandweite Suche, da Bibliotheken in regionalen Verbänden organisiert sind
- bisher Metasuche (Karlsruher Virtueller Katalog), regionale Ansätze (etwa: Dreiländerkatalog – alle deutschsprachigen Nachweise für Monografien)
- Metasuche reicht nicht aus, Google und Amazon sind selbst für Bibliothekare nützlicher

Reaktion der Hersteller von Bibliothekssystemen

- Suchmaschinenindexierung für lokale Bestände
- geschlossene Systeme, gekapselt durch Portale
- Konfiguration, Wartung und Betrieb, hohe Kosten stellen eine Herausforderung dar
- wachsende Bedeutung, aber fragmentierte Suchlandschaft begrenzt Akzeptanz durch Endnutzer, wenn nicht alle deutschsprachigen Nachweise abgedeckt werden
- kein gemeinsamer Sucheinstieg, keine Schnittstellen zu Bibliotheksdiensten wie Fernleihe

Aktivitäten der Bibliotheksverbände

- Hochschulbibliothekszentrum Nordrhein-Westfalen / Rheinland-Pfalz: FAST, Dreiländerkatalog, hbz stellte 2005 erfolgreich Antrag auf Erprobung von Suchmaschinentechnologie in vascoda (DFG): dezentrale Suchmaschinen (hbz Köln: FAST ESP, TIB Hannover und ZPID Trier: Lucene)
- Bibliotheksverbund Bayern: Infoguide FAST, Gateway Bayern
- Gemeinsamer Bibliotheksverbund: PSI - Pica Search Index, Kavia-Projekt
- Südwestverbund: Lucene, BAM-Portal (DFG)

Herausforderungen

- herstellerunabhängige Verfahren sind ein Muss
- zentrale Indexierung plus Suchmaschinenföderation
- Anpassung von Web-Suchmaschinentechnologie auf Bibliothekskataloge
- unterschiedliche Leistungsfähigkeit der Suchmaschinenprodukte am Markt
- keine technologischen Gemeinsamkeiten: Treffer-IDs, Anfragesprache, Trefferformat, Relevanzranking
- de-facto-Standard: Google, Amazon, daher Markt für Föderationsprodukte mit großer Schwankungsbreite und ohne Aussicht, Fragmentierungen zu überwinden

Was ist eine Suchmaschinenföderation?

- extrem kurze Antwortzeiten (unter einer Sekunde)
- Mechanismus für universell einheitliche Treffer-IDs (UUID)
- alle Teilnehmer haben eine gemeinsame Anfragesprache
- und ein gemeinsames Relevanz-Ranking
- und ein gemeinsames Trefferformat
- und ein gemeinsames Verfahren zum Treffer-Clustering (facettierte Suche, S.R. Ranganathan, 1933)
- und einen gemeinsamen Filtermechanismus ("selective dissemination of information", H. P. Luhn, 1958)
- wesentlich mehr als eine Suchföderation von Metasuchen

Technik der hbz- Suchmaschinenföderation

- herstellerneutrale Implementierung, offenes Suchprotokoll
- Web-Service (REST, HTTP, XML, Unicode)
- Anfragesprache: Contextual Query Language (CQL)
plus Erweiterungen für Suchmaschinenanfragen
- Filter-Kontext: Treffermengen einschränken, Auswahl von Kollektionen
- Facet-Kontext: Aktivieren der facettierten Suche („drill-down“)
- Option-Kontext: Suchmaschinensteuerung, z.B. linguistische Einstellungen

Technik der hbz- Suchmaschinenföderation

- Trefferformat jedes Teilnehmers und des Föderators ist gleich: Atom RFC 4287, erweitert um eigene Namensräume
- Atom UUID für Feed und Entry-Elemente
- Zusammenführung der Ranking-Relevanzwerte der teilnehmenden Suchmaschinen: lokale Werte müssen normalisiert werden
- Auswahl der relevanten Dokumente: hierarchisches Modell mit Hilfe von Klassifikationen und Geopositionen (ISIL, RVK, DDC, vascoda CLD)
- baut Worldcat API weiter aus: CQL, Atom, OpenSearch

Clients



Protokoll



HTTP Request



XML
Atom Syndication Format
RFC 4287

Federator



Protokoll



Retrieval-Systeme



3 - 5 February 2009 in Bielefeld, Germany

9th International Bielefeld Conference



Perspektiven

- Implementierung des hbz-Suchmaschinenförderators wurde 2005 begonnen und geht weiter
- 2007: Prototyp auf Basis Microsoft FAST ESP / FAST Unity / Lucene Solr wurde in vascoda eingebracht
- Einbindung weiterer kommerzieller Suchmaschinenprodukte nach Bedarf
- Suchmaschinenförderung in den Fällen mit Partnern weiter ausbauen, wenn Datenindexierung im Suchmaschinen-Index des hbz als nicht geeignet erscheint

Perspektiven

- Teilnehmer an der hbz-Suchmaschinenföderation müssen strenge Kriterien erfüllen: Hochverfügbarkeit, schnelle Antwortzeit, Suchsprachensyntax, Trefferformat
- Suchmaschinenföderator wird als "neuer Motor" der Digitalen Bibliothek des hbz von allen teilnehmenden Bibliotheken genutzt werden
- Föderator-Quellcodes werden offen gelegt
- Suchschnittstelle: Spezifikation wird offen gelegt