

# Discovery-Systeme - die OPACs der Zukunft?

Zukunft der lokalen Bibliothekssysteme  
Neue Strukturen und Anforderungen im wissenschaftlichen Bibliothekswesen

Gerald Steilen

101. Deutscher Bibliothekartag  
Bibliotheken - Tore zur Welt des Wissens  
Hamburg 22. - 25. Mai 2012

# Discovery in der VZG

- VuFind seit 2007
- Produktiver Betrieb und Entwicklung seit 2008:  
„Suchkiste“ für DFG Nationallizenzen:
- <http://finden.nationallizenzen.de/>
- Ergebnis aus positiven Erfahrungen:
  - „GBV Discovery“ Index: Solr-Index mit allen Daten des GBVs, derzeit ca. 101 Mio. Datensätze
  - Angebot von VuFind-Dienstleistungen
- TouchPoint von 2009 - 2011

# Ziel

Finden von Bibliotheksbeständen für Nutzer

- Nutzer verwenden außerbibliothekarische Recherchesysteme zur Vorbereitung von Literatursuchen in Bibliotheksangeboten (Akselbo et al. 2006, S. 17)
- schlechte Nutzbarkeit, hohe Komplexität, fehlende Integration elektronischer Ressourcen (Wong et al. 2009, S. 5 f.)

# OPAC

- direkter Nachfolger des Zettelkatalogs ( $\approx$  70er Jahre)
  - ähnliches Konzept
    - wenige Einsteigpunkte (Autor, Titel, Thema)
    - Präkoordination
  - kein Stöbern mehr möglich
- Erweiterte Suche ( $\approx$  80er Jahre)
  - kombinierte Suche in verschiedenen Indices
  - formale Anfragesprache

## Suchparadigma: exakte Treffer

### Ziele

- möglichst wenige Treffer (Idealfall ein Treffer)
- lösen eines Mengenproblems

### Voraussetzung

- spezielle Anfragesprache
- Expertenwissen / Schulungen

## Bool'sche Algebra

### Operatoren für den Nutzer

- UND (AND)

*= alle Elemente müssen enthalten sein*

- ODER (OR)

*= min. 1 Element muss enthalten sein*

- NICHT (NOT)

*= kein Element darf enthalten sein*

- exklusive ODER (XOR)

*= entweder das eine Element oder das andere, aber nicht beide*

# Bool'sche Algebra

## Gesetze

- Kommutativgesetze

$$a \text{ AND } b = b \text{ AND } a$$

- Assoziativgesetze

$$(a \text{ AND } b) \text{ AND } c = a \text{ AND } (b \text{ AND } c)$$

- Distributivgesetze

$$a \text{ AND } (b \text{ OR } c) = (a \text{ AND } b) \text{ OR } (a \text{ AND } c)$$

- Absorptionsgesetz

$$a \text{ OR } (a \text{ AND } b) = a$$

- ...

## Bool'sche Algebra

### Ergebnis

- wahr
- falsch
- Menge prinzipiell unsortiert
- sehr präzise Suchen möglich

### Probleme

- 0 Treffer
- zu viele Treffer



OPAC > Suchparadigma: exakte Treffer > Bool'sche Algebra >  
Zielgruppe

- Informationsspezialisten
- geschulte Nutzer

# Discovery Systeme

- direkte Konsequenz der Web-Suche
- große Datenbestände
  - Viele sortierte Treffer (Relevance Ranking)
    - Ergebnisse „gut genug“
    - Postkoordination
  - Stöbern möglich (z.B. Facetten)
- intuitiv nutzbar
  - keine formale Anfragesprache
  - Assistenz
    - Rechtschreibhilfe
    - alternative Suchanfragen

## Suchparadigma: beste Treffer

### Ziele

- beste Treffer im Blickfeld des Nutzers
- lösen eines Sortierungsproblems

### Voraussetzung

- Ermittlung der Relevanz
- Grundannahmen über den Nutzer

## Vektorraum Modell

### Relevanz

- Ähnlichkeit zwischen Suchanfrage und Dokument
  - = *berechneter Messwert*
- Relevanzsortierung
  - = *Sortierung nach Messwerten*

### Boosting

- Gewichtungsfaktor zur Relevanzberechnung
- individuell beeinflussbar

## Vektorraum Modell

### Suchvorschläge

- während der Eingabe (autocomplete)
- ähnliche Dokumente
- ...

### Facetten

- Ergebnisgruppen (Materialtyp, Standort, ...)
- Kombinationen von Gruppen
- Stöbern möglich
- eingrenzen der Ergebnisliste

## Vektorraum Modell

### Ergebnis

- errechneter Messwert der Ähnlichkeit zwischen Suchanfrage und Dokument
- prinzipiell sortierte Liste
- sehr gut geeignet für große Dokumentmengen

### Probleme

- sinnvolle Elemente für Boosting
- Datengrundlage für Facetten

## Zielgruppe

- Web-affine Nutzer
- nicht speziell geschulte Nutzer

# Vergleich

OPAC (exakte Treffer)	Discovery (beste Treffer)
präzise Suchanfrage	Suchanfrage beschreibt gewünschtes Dokument
Treffer erfüllen Suchanfrage exakt	Treffer erfüllen Suchanfrage so gut wie möglich
unsortierte Ergebnismenge	nach Rangfolge sortierte Ergebnisse
bekannte Dokumente finden	unbekannte Dokumente finden („Entdecken“)
spezielle Suchsprachen (Boolsche Operatoren, Klammerung)	einfache Suchanfragen
Erlernen notwendig	intuitiv benutzbar
gut für abgegrenzte, überschaubare Dokumentbestände	gut für offene, große Dokumentbestände
geringe Fehlertoleranz („0 Treffer“)	Fehlertoleranz, kann aber unpassende Treffer liefern



# Vergleich

OPAC (exakte Treffer)	Discovery (beste Treffer)
Nutzer verwendet Mathematik (Boolsche Algebra)	Maschine verwendet Mathematik (Vektorraum)

# Voraussetzungen für den Einsatz von Discovery-Systemen

- Verständnis der prinzipiellen Unterschiede zwischen den unterschiedlichen Suchparadigmen exakte und beste Treffer
- Kenntnis der eigenen Metadaten
  - für Facetten geeignete, einheitlich belegte Felder
  - Erschließungsgrad
  - Bereitschaft Fehler in den Metadaten zu beheben
- Bereitschaft das System permanent zu optimieren.

# Voraussetzungen für Discovery-Systeme



# VuFind-Service der VZG

- [http://www.gbv.de/wikis/cls/VuFind-Service der VZG](http://www.gbv.de/wikis/cls/VuFind-Service_der_VZG)
- <http://www.gbv.de/wikis/cls/Vufind>

# Literatur

- Akselbo, J. L., Arnfred, L., Barfort, S., Bay, G., Christiansen, T. B., Hansen, J. H., Jensen, H. T., Markussen, G. B., Morthorst, A. M. e Nielsen, M. P. (2006), «The hybrid library: from the users' perspective. A report for the DEFF project "The loaners' expectations and demands for the hybrid library".», Rap. tecn., The National Library and Copenhagen University Library. The Aarhus School of Business Library. The State and University Library. The University Library of Southern Denmark., URL <http://www.statsbiblioteket.dk/summa/fieldstudies.pdf>
- Wong, W., Stelmaszewska, H., Barn, B., Bhimani, N. e Barn, S. (2009), «JISC User Behaviour Observational Study: User Behaviour in Resource Discovery. Final Report», Rap. tecn., JISC, URL <http://www.jisc.ac.uk/media/documents/publications/programme/2010/ubirdfinalreport.pdf>

Herzlichen Dank für Ihre Aufmerksamkeit

Kontakt:

Gerald Steilen  
steilen@gbv.de  
0551 39-91272