

kobv

■ ■ ■  
■ ■ ■ KOOPERATIVER BIBLIOTHEKSVERBUND  
■ ■ ■ BERLIN-BRANDENBURG  
■  
■  
■

# OPAC-Retrieval mit Nutch und Lucene

**Lavinia Hodoroaba**

**KOBV-Zentrale im Zuse-Institut Berlin (ZIB)**

**Lucene-Workshop, Stuttgart, 24. Januar 2006**



■ ■ Vision

■ ■ Von der Vision zur Implementierung

■ ■ Ausblick

■ ■ „Information Retrieval“-Lösung für die OPACs

■ ■ lokale und/oder zentrale Einsatzmöglichkeit

■ ■ Integration der lokalen Dienste

■ ■ Union Catalog (FU Berlin)

■ ■ MAB2, MABXML

■ ■ RVK, SWD, GKD

■ ■ Lucene 1.4.3

■ ■ Lucene Spell Checker 1.1

■ ■ Nutch 0.7.1

■ ■ OpenURL 1.0

# Von der Vision zur Implementierung - Prozesse

0) Aufbereitung der RVK-/GKD-/SWD-Normdaten

1) Export

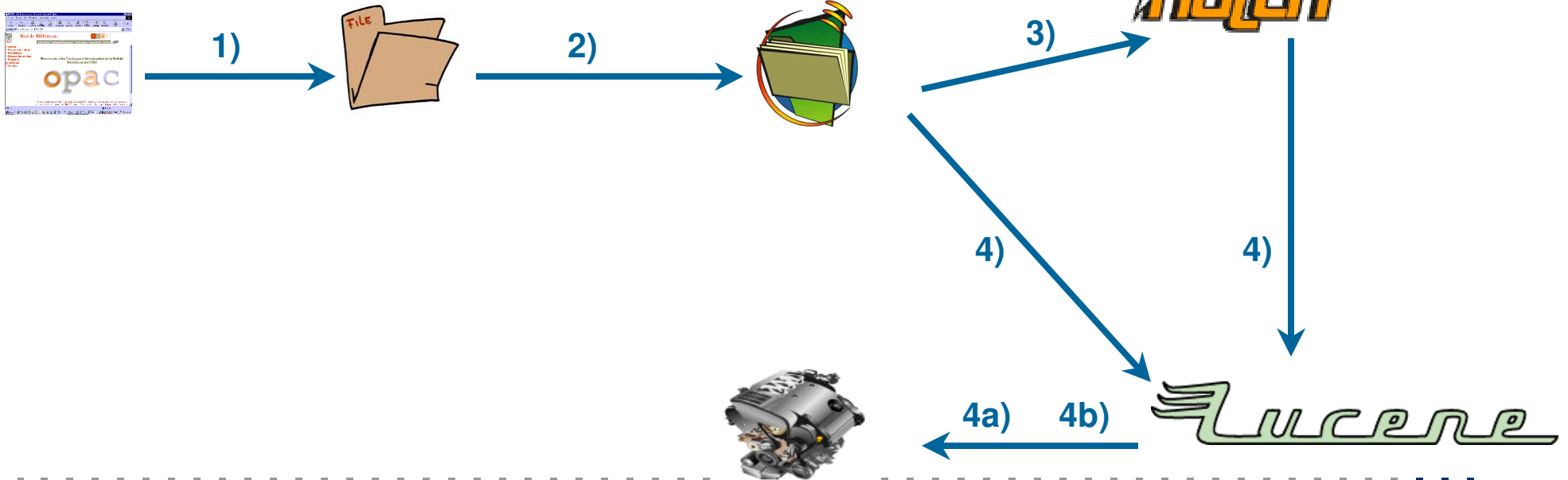
2) Umwandlung nach MABXML (MARCXML etc.)

3) Web-Crawling

4) Metadaten und Volltext-Indizierung

➤ 4a) Auflösung von RVK-Verweisungen

➤ 4b) Auflösung von GKD-/SWD-Verweisungen





### ■ ■ Lucene-Prototyp – mit den 1,8 Millionen Datensätzen der FU Berlin

- Indizierung des MABXML-Formates
- Einfache und erweiterte Suche
- Nahtloser Übergang zwischen einfacher und erweiterter Suche
- „Highlighting“ der Suchbegriffe - für das Content-Feld
- „Meinten Sie:“-Funktionalität – für die einfache Suche
- Sortierung der Trefferliste nach Relevanz
- Indizierung von vorhandenen Inhalten
- Statische Auflösung von RVK-/GKD-/SWD-Verweisungen für die Anzeige
- Wildcard-Suche / Boolesche Suche / Phrase-Suche (mit „“) / Fuzzy-Suche (mit ~) / Distanzsuche (mit „“ und ~) / Suche mit Verstärkungsfaktoren (mit ^)
- Sitzungsabhängige Einstellungen für die Anzahl der Treffer pro Seite und die Sprachauswahl (DE / EN)
- Sitzungsabhängige Merkliste
  
- Qualitätskontrolle

- ■ Update
- ■ Integration weiterer Online-Kataloge
- ■ Indizierung weiterer Formate (z.B. MARCXML)
- ■ Verbesserung der „Analyse“-Phase des Indizierungsprozesses
- ■ Verbesserung der Navigation
- ■ Weitere Sprache für die Oberfläche
- ■ Dynamische Auflösung von GKD-/SWD-Verweisungen für die Anzeige
- ■ Sortierung der Trefferliste nach weiteren Kriterien (Autor, Titel, Jahr etc.)
- ■ Relevanz-Verbesserung durch „Boosting“
- ■ „Drill-Down“-Suche
- ■ Bearbeitung der Suchergebnisse
- ■ Integration der lokalen Dienste
- ■ Barrierefreiheit
- ■ Browsing

## ■ ■ *Datenanreicherung*

**Vielen Dank**

**für ihre Aufmerksamkeit!**