

Was nachher können?

- Data Mining in Strings
- Index-Kompression
- verteilte Indizes
- dynamische Textkollektionen
- Suche mit regulären Ausdrücken
- Boole'sche Suche
- Analyse der fundamentalen DSen
- Lucene/Xapian anwenden können
- mit großen Alphabeten umgehen können
- linguistische/semantische Vorverarbeitung von Dokumenten
- Autocomplete
- komplexere reguläre Ausdrücke
-

Anwendungen

- Desktopsuche (Dokumente oder besuchte Webseiten indexieren) 5
- Bücher indexieren: RFID?? 2
- (Dateisystemsuche)
- automatische Kreuzworträtselerstellung 3
- automatische Übersetzung 2
- Plagiaterkennung 4
- (Metadaten­suche)
- semantische Bedeutung von Wörtern *automatisch* finden (auf Basis von Texten) 5 **Ko-
Okkurrenzen**
- automatische Thesauruserstellung

Bis zum 10.6.

- Gruppen bilden (2-4 Personen)
- für eine Anwendung entscheiden
- Überlegen, wie umsetzen
- ca. 10-15 Minuten Präsentation
 - ▶ Problemstellung
 - ▶ Lösungsansätze (inkl. zu verwendende Tools)
 - ▶ evtl. Bedarf nach weiterer Theorie