

Übung IR: German-Analyzer für Indizierung und Suche

May 18, 2009

- Gegeben sind folgende Dokumente
 - Doc1: Ich gebe Dir Rückmeldung
 - Doc2: Ich gebe Dir keine Rückmeldung
- Gegeben sind folgende *Queries*
 - Q1: "gebe Dir keine Rückmeldung"
 - Q2: Rückmeldungen
 - Q3: "gebe Dir eine Rückmeldung"

- Stoppworte (wie: *keine*, *dir*) sollen nicht indiziert werden
- Q1 soll nur Doc2 matchen (über QueryParser)
- Q2 soll beide Dokumente matchen

- Schreiben Sie einen eigenen Analyzer
- Benutzen Sie den GermanStemFilter (aus der Sandbox)
- Benutzen Sie die Stoppwortliste von GermanAnalyzer (aus der Sandbox)
- Wenden Sie den Analyzer bei der Indizierung und Suche an

- Laden Sie sich das Indexinspektionswerkzeug Luke aus der Lucene Sandbox
- Überprüfen Sie mittels Luke die Positionen der einzelnen Terme - Welche Terme sind indiziert und auf welchen Positionen in den beiden Dokumenten?

- Dokumentieren Sie kurz und präzise ihre Ergebnisse (pdf-Datei)
- Laden Sie ihre Ergebnisse ins Moodle-System: Dokumentation und Quellcode
- letzter Abgabetermin: 25.5.2009 - 12:00 Uhr

Beachten Sie

Die Qualität der Ergebnisdarstellung geht auch in die Bewertung ein.