

Übung zur Vorlesung Information Retrieval

May 21, 2009

- Indizieren Sie die Dokumente der Übung Scoring und suchen Sie für die Fragen der letzten Übung Scoring mittels Vektorspace-Modell
- Benutzen Sie dabei einen Analyser, der nur Whitespaces trennt.
- Benutzen Sie `Explanation`
- Verwenden Sie den `QueryParser` über Kommandozeilen-Eingaben
- Vergleichen Sie die Ergebnisse mit den manuellen Ergebnissen und erklären Sie die Abweichungen

- Indizieren Sie Sie Dokumente und machen Sie dabei folgende Änderungen:
 - Boosten Sie Dokument 3 mit dem Faktor 2
 - Welche Änderungen ergeben sich jetzt bei der Suche nach "Krebs"
- Ändern Sie die Frage "Krebs OR Adenokarzinom" so, dass Sie als ersten Treffer ein Dokument mit dem Begriff "Krebs" erhalten

- Fügen Sie zu den Dokumenten ein Datumsfeld mit folgenden Einträgen hinzu:
 - Doc1: 10.1.2008
 - Doc2: 1.2.2008
 - Doc3: 1.2.2008
 - Doc4: 31.3.2008
 - Doc5: 12.4.2008
- Schränken Sie die Anfragen auf die Dokumente ein, die zwischen dem 15.1.2008 und dem 15.4.2008 liegen (ohne dass sie das Ranking ändert)
- Die Datums-Eingabe in der Query soll im (üblichem deutschen) Format erfolgen: dd.mm.yyyy
- Benutzen Sie zur Kontrolle `query.toString()`
- Der Datumanteil soll keinen Beitrag zum Score leisten (über Query)

- Dokumentieren Sie kurz und präzise ihre Ergebnisse (pdf-Datei)
- Laden Sie ihre Ergebnisse ins Moodle-System: Dokumentation und Quellcode
- letzter Abgabetermin: 18.5.2009 - 12:00 Uhr

Beachten Sie

Die Qualität der Ergebnisdarstellung geht auch in die Bewertung ein.