

Information Retrieval

Beurteilung von IR-Systemen

Stefan Birkner

18. August 2010

Zweck der Beurteilung

- ▶ Vergleich verschiedener Systeme
- ▶ Auswirkung von Konfigurations- oder Konstruktionsänderungen

Auswirkung von Konfigurations- oder Konstruktionsänderungen

Beispiel: Einführung neuer Synonyme

Evaluierung bei Datenbanken

- ▶ Funktionsumfang
- ▶ Performance

Evaluierung bei IR-Systemen

- ▶ Funktionsumfang
- ▶ Performance
- ▶ Qualität der Ergebnisse

Effizient und Effektivität

- ▶ Effizienz
- ▶ Effektivität

Relevanz

- ▶ relevante und nicht relevante Dokumente
- ▶ subjektive Beurteilung
- ▶ Abhängigkeiten zwischen Dokumenten
- ▶ stark vereinfachend

Recall und Precision

- ▶ Recall - Trefferquote
- ▶ Precision - Genauigkeit

Recall und Precision

Das IR-System hat einen Gesamtbestand von 100.000 Dokumenten. Zur Anfrage „Wann begann die Ölkatastrophe im Golf von Mexiko?“ liefert das System 100 Dokumente zurück, von denen 70 wirklich relevant sind. Im Gesamtbestand gibt es jedoch 140 Dokumente, die für unsere Frage relevant sind.

Recall und Precision

Das IR-System hat einen Gesamtbestand von 100.000 Dokumenten. Zur Anfrage „Wann starb Dschingis Khan?“ liefert das System 60 Dokumente zurück, von denen 40 wirklich relevant sind. Im Gesamtbestand gibt es jedoch 50 Dokumente, die für unsere Frage relevant sind.

Recall und Precision

- ▶ Hits
- ▶ Noise
- ▶ Misses
- ▶ Rejected

Recall und Precision

$$F - Measure = \frac{Recall \cdot Precision}{(1 - \alpha) \cdot Recall + \alpha \cdot Precision}$$

Recall und Precision

Bestimmung des Recall

Suchbegriff: Anfrageerweiterung zur Bestimmung von Recall und Precision

- ▶ google.de
- ▶ bing.de
- ▶ ask.com
- ▶ yahoo.de

Recall und Precision

Bestimmung des Recall

- ▶ Stichprobe
- ▶ Dokument-Source-Methode
- ▶ Anfrageerweiterung
- ▶ Abgleich mit externen Quellen

Recall und Precision

Mittelwertbildung über mehrere Anfragen

- ▶ Makrobewertung
- ▶ Mikrobewertung

Testkolektionen

TREC-Kollektion

- ▶ mehrere Giga-Byte an Dokumenten
- ▶ Topics (Title, Description, Narrative)
- ▶ Relevante Dokumente manuell ausgewählt

Kriterienkataloge

Cleverdon, Mills und Keen

- ▶ Recall
- ▶ Precision
- ▶ Time Lag
- ▶ Effort
- ▶ Form of Presentation
- ▶ Coverage of the Collection

Kriterienkataloge

Vickrey (Verfügbarkeit und Selektivität)

- ▶ Coverage
- ▶ Recall
- ▶ Response
- ▶ Precision
- ▶ Usability
- ▶ Presentation

Literatur

- ▶ Andreas Henrich: Information Retrieval 1
- ▶ http://www.uni-bamberg.de/minf/ir1_buch/
- ▶ Kapitel 2