

Multimedia Information Extraction and Retrieval

Summer Term 2012

Exercise Sheet 5

Ralf Möller, Karsten Martiny

Exercise Session:
Thursday, May 24, 2012, 8.00-8.45, SBS95-D1025

1. Suppose that an IR system contains 1000 documents. It is known that for one particular query 27 documents are relevant, as listed below:

$$\{d1, d5, d6, d10, d88, d150, d200, d210, d250, d300, \\ d399, d400, d405, d450, d472, d500, d501, d530, \\ d545, d590, d600, d635, d700, d720, d800, d888, d900\}$$

Two different approaches are used to retrieve ranked documents for this query. Each system only returns the top 10 ranked documents in order of ranking. Approaches 1 and 2 each retrieves documents one at a time in the following order with all 10 documents eventually returned:

- Approach 1: d2, d5, d150, d250, d11, d33, d50, d600, d500, d720.
 - Approach 2: d250, d400, d150, d210, d999, d1, d501, d800, d200, d300.
- (a) Plot the Precision and the Recall graphs for each approach as a function of the number of documents returned (for 1 document returned, 2 documents returned, etc).
 - (b) Calculate the Precision versus Recall for approaches 1 and 2 using these query results as a function of the number of documents returned.
 - (c) Compute the F_1 -measure.
 - (d) Which approach is best? Justify your answer.
2. Use the following search engines: Google, Bing, Yahoo, DuckDuckGo. Put in the queries:
 - (a) jaguar
 - (b) jaguar cat (without quotation marks!)

For the queries and for each search engine:

- (a) Generate plots with jaguar the cat being the relevant returned content. Make a precision plot for each search engine as above for the first 20 returned links as a function of the number of documents returned.

- (b) Calculate the average precision over both queries. Compare the results for each query. How did the search engines rank?
3. Consider the a set of four retrieved documents $\{d1, d2, d3, d4\}$. The relevance of the retrieval documents has been evaluated by 2 judges the following way (+ means relevant, - means non-relevant):
- Judge a: d1+, d2+, d3-, d4-
- Judge b: d1+, d2-, d3-, d4-
- Give a measure of agreement between the judges.