

Multimedia Information Extraction and Retrieval SoSe 2010 Exercise Sheet 4

Prof. Dr. Ralf Möller, Sebastian Wandelt

19.05.2010

1. Suppose that an IR system contains only 1000 documents. A query is known to generate 27 relevant documents as listed below:

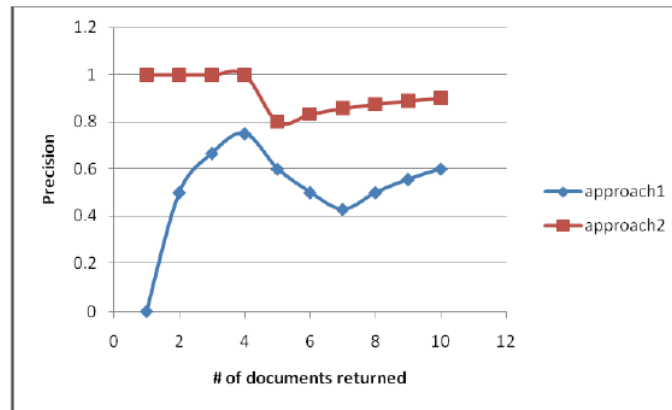
$\{d1, d5, d6, d10, d88, d150, d200, d210, d250, d300, d399, d400, d405, d450, d472, d500, d501, d530, d545, d590, d600, d635, d700, d720, d800, d888, d900\}$

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:

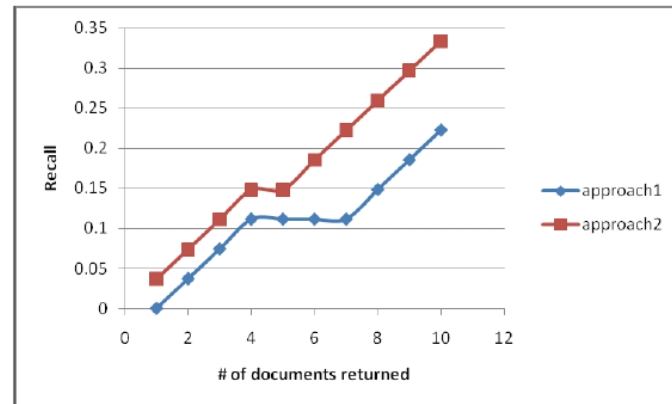
- Approach 1: d2, d5, d150, d250, d11, d33, d50, d600, d500, d720.
 - Approach 2: d250, d400, d150, d210, d999, d1, d501, d800, d200, d300.
- (a) Plot the Precision and the Recall graphs for each approach as a function of the number of documents returned (for 1 document returned, 2 documents returned, etc).

Solution:

Precision:

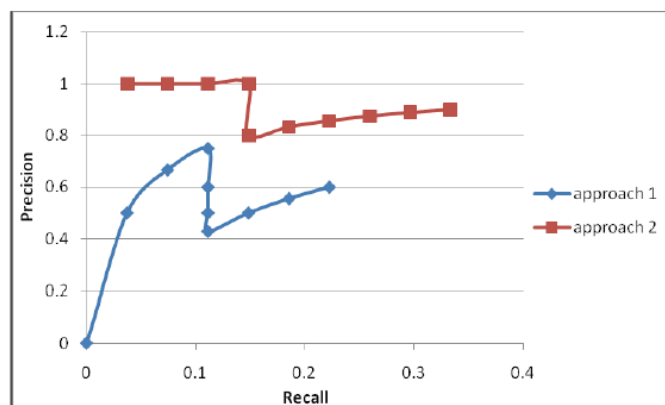


Recall:



- (b) Calculate the Precision versus Recall for approaches 1 and 2 using these query results as a function of the number of documents returned.

Solution:



- (c) Compute the F_1 -measure.

Solution:

We compute the F_1 -measure for the maximum (10 documents):

$$F_1 = \frac{1}{\frac{0.5}{P} + \frac{0.5}{R}}$$

Approach 1: $F_1 = 0.322$

Approach 2: $F_1 = 0.483$

- (d) Which approach is best? Justify your answer.

Solution:

Depends on the point of view. Comparing graph-intuitions with the F -measure yields different results.

2. Use the following search engines: Google, Bing, True Knowledge, Freebase. Put in the queries:

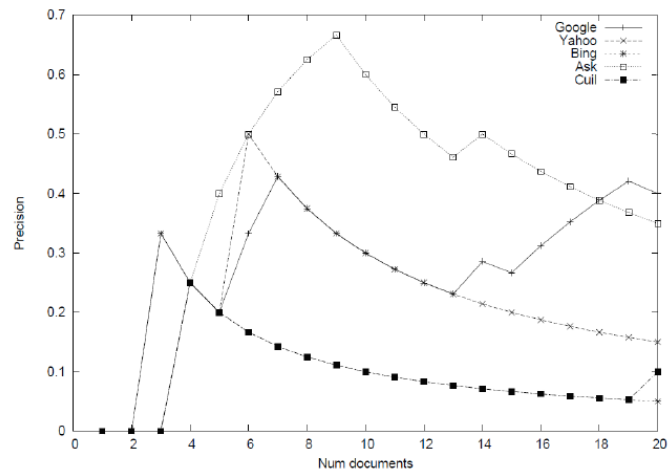
- (a) jaguar
- (b) jaguar cat (without quotation marks!)

For the queries and for each search engine:

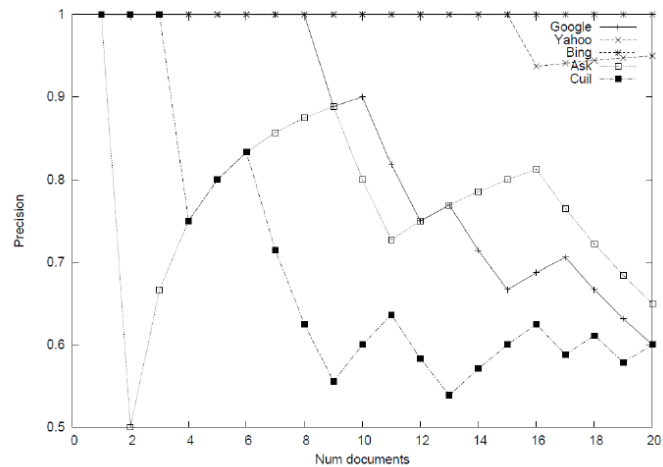
- (a) Generate plots with jaguar the cat being the relevant returned content. Make a precision plot for each search engine as above for the first 20 returned links as a function of the number of documents returned.

Solution:

jaguar:



jaguar cat:



- (b) Calculate the average precision over both queries. Compare to that for each query. How did the search engines rank?

Solution:

Ranking as follows: $A > G > Y > B > C$ for jaguar and $B > Y > G > A > C$ for jaguar cat

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.

Solution:

Judge a	Judge b			
		Yes	No	Total
	Yes	1	1	2
	No	0	2	2
	Total	1	3	4

$$kappa = \frac{P(A) - P(E)}{1 - P(E)}$$

$$P(A) = 3/4$$

$$P(\text{rel}) = (2+1)/8 = 3/8$$

$$P(\text{non-rel}) = (3+2)/8 = 5/8$$

$$P(E) = P(\text{rel})^2 + (P(\text{non-rel}))^2 = \frac{9}{64} + \frac{25}{64} = \frac{34}{64} = \frac{17}{32} = 0.5313$$

$$p(A) = 3/4 = 24/32$$

$$kappa = \frac{\frac{24}{32} - \frac{17}{32}}{1 - \frac{17}{32}} = \frac{\frac{7}{32}}{\frac{15}{32}} = \frac{7}{15} = 0.4667 \quad \text{not a good agreement}$$