IST 441 - Information Retrieval Evaluation – Exercise 1

This exercise is worth 6 points. (Due Feb 9)

In this exercise, you will familiarize yourself with standard methods for evaluating information retrieval systems with an emphasis on precision and recall.

1. (3 points) Suppose that an IR system contains only 1000 documents. A query is known to generate 29 <u>relevant</u> documents as listed below:

$\{d_1, d_5, d_6, d_{10}, d_{88}, d_{150}, d_{200}, d_{210}, d_{250}, d_{300}, d_{399}, d_{400}, d_{405}, d_{450}, d_{472}, d_{500}, d_{501}, d_{530}, d_{545}, d_{590}, d_{600}, d_{635}, d_{674}, d_{700}, d_{720}, d_{800}, d_{888}, d_{900}, d_{917}\}$.

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: $d_3, d_5, d_{150}, d_{250}, d_{11}, d_{33}, d_{50}, d_{600}, d_{500}, d_{720}$.

Approach 2: $d_{250}, d_{400}, d_{150}, d_{210}, d_{999}, d_1, d_{530}, d_{800}, d_{200}, d_{300}$.

Answer the following:

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. Which approach is best? Justify your answer.

2. (2 points)

Use the following search engines: Google, Bing, blekko, DuckDuckGo and one of your choosing. Put in the queries (make sure you are logged out of google and/or bing):

1. jaguar
2. jaguar cat (without quotation marks!)

For the queries and for each search engine:

1. 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.
2. Calculate the average precision over both queries. Compare to that for each query.
3. How did the search engines rank? Which search engine did you like best and why?

3. (1 point) Exercise 8.2 in Manning, Raghavan, Schutze