

Information Retrieval Evaluation – Exercise 1

IST 441

This exercise is worth 6 points.

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

1. (4 points) 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.

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, Yahoo, MSN, ASK. Put in the queries:

jaguar
“your choice”

For “jaguar” do plots with the cat being the **relevant** returned content. Make a precision plot for each search engine as above for the first 10 returned links as a function of the number of documents returned.

Do the same for your query.

How did the search engines rank and why?