Discussion

Pre-workshop

• Revise the formulae and mechanisms for calculating the following effectiveness metrics (in an Information Retrieval context): Precision, Recall, F-score, Precision at $k$, Mean Average Precision.

Workshop

1. When evaluating the effectiveness of an IR system:
   
   (a) What simplifying assumptions do we make? How well do these map to user behaviour? Why do we not normally care about Recall, and what is different in fields (like Law) where we do?
   
   (b) What is “pooling”? Why is commonly used in a “shared task” environment like TREC?
   
   (c) What is a “P-R curve”? Why do we care about the area under it? One way of estimating this area is with MAP; another is with “Interpolated Precision”:

   $$p_{interp}(r) = \max_{r' \geq r} p(r')$$

   How can we use this to find the area under the P-R curve? In what ways is it superior or inferior to MAP?

2. What is “query expansion”? What are some mechanisms of leveraging it into an IR system? In what circumstances is it desirable or undesirable to do so?

   (a) What are the advantages and disadvantages of using a “manual thesaurus” to perform query expansion?
   
   (b) What are some mechanisms for constructing an “automatic thesaurus”, and what rationale(s) are they based on?

3. What is “relevance feedback”, and how does it help a system to return better results? (More on relevance feedback after the non-teaching period.)

Post-workshop

• Use an automatic thesaurus algorithm (for example, the similar(t) function in NLTK), and consider the set of results for one or more terms. What information do the results capture about the query term? The collection?
Programming

Pre-workshop

• With reference to a fixed set of relevant documents to a query, write a program that counts how many documents returned by an IR system are actually relevant. (You can use the project data.)

• Based on the previous program, write a program that calculates precision, Precision at 1, Precision at 10, and Average Precision.

Workshop

1. Using the “unit-normalised term similarity” approach, write a program which finds similar terms to a given query term, in a fixed document collection.

2. Compare the results from your program to those returned by the similar(t) function in NLTK.

Post-workshop

• Work on the project.