Web Information Retrieval

Exam
January 27th, 2013
Time available: 90 minuti
5 points for each problem

Problem 1

1. Consider the Boolean query `x AND y AND NOT z`. Write out a postings merge algorithm that evaluates this query efficiently. *Motivate your answer.*

2. Assume the following case:
   x = {12, 27, 40}
   y = {27, 35, 46}
   z = {1, 3, 12, 18, 25}
   Work out the number of comparison needed to evaluate the above query in this case.

3. Next, consider the Boolean query `x AND NOT y`. Assume that x has length 1 and y has length some integer n. Give an upper bound on the overall number of comparisons needed to evaluate this query in the worst case. *Motivate your answer.*

Problem 2

The following list of R's and N's represents relevant (R) and nonrelevant (N) returned documents in a ranked list of a collection of 30 documents. The top of the ranked list (the document the system thinks is most likely to be relevant) is on the left of the list.

R R N R N N R R N N N R N N N N R N N R N N N N N N R N R N

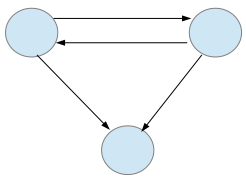
1. What are precision and recall of the system on the top 5?

2. Draw the precision-recall curve.

3. What is the interpolated precision at 50% recall?

Problem 3

1. Describe an external memory algorithm for the implementation of the power iteration method for pagerank computation.



2. Execute the algorithm on the graph of the figure with initial state $(1, 0, 0)$, $\alpha = \frac{1}{2}$ and 3 iterations.

3. Describe the set of liear equations for the computation of Pagerank of the example in figure and solve it to find its exact solution.

Problem 4

1. Describe the assumptions of a Naive Bayes classifier in the `bag of words` model.

2. Compute the coefficients of a boolean classifier without `smoothing` on the following 4 training documents:

(a) `browsing lion safari.  apple`

(b) `africa video lion.  not apple`

(c) `lion mountain osx.  apple`

(d) `mountain safari browsing tiger.  not apple`

3. Classify the query document: `lion mountain safari`

I consent to publication of the results of the exam on the Web

Firstname and Lastname in block letters.....

Signature