INLS 509 Information Retrieval
Homework 3

Due: Monday, November 2nd (by 6:40pm via Sakai)

1: Rank-based Evaluation Metrics [75 points]

Assume the following ranking for a given query (only results 1-10 are shown). The column 'rank' gives the rank of the document. The column 'docid' gives the document ID associated with the document at that rank. The column 'graded relevance' gives the relevance grade associated with the document (4 = *perfect*, 3 = *excellent*, 2 = *good*, 1 = *fair*, and 0 = *bad*). The column 'binary relevance' provides two values of relevance (1 = *relevant* and 0 = *non-relevant*). The assumption is that anything with a relevance grade of 'fair' or better is *relevant* and that anything with a relevance grade of 'poor' is *non-relevant*.

Also, assume that this query has only 7 documents with a relevance grade of *fair* or better. All happen to be ranked within the top 10 in this given ranking.

Answer the questions below. $\mathcal{P}$@K, $\mathcal{R}$@K, and average precision (AP) assume binary relevance. For those metrics, use the 'binary relevance' column. DCG and $NDCG$ assume graded relevance. For those metrics, use the 'graded relevance' column.

If you want partial credit, you must show your work. Otherwise, your grading will also be "binary" ;-).

rank	docid	graded relevance	binary relevance
1	43	4	1
2	531	0	0
3	183	3	1
4	102	2	1
5	10	1	1
6	1051	0	0
7	1031	1	1
8	332	2	1
9	573	0	0
10	128	2	1

(a) Compute $\mathcal{P}$@5 and $\mathcal{P}$@10. [10 points]

(b) Compute $\mathcal{R}$@5 and $\mathcal{R}$@10. [10 points]

(c) Provide an example ranking for this query that maximizes $\mathcal{P}$@5 [5 points]

(d) Provide an example ranking for this query that maximizes $\mathcal{P}$@10 [5 points]

(e) Provide an example ranking for this query that maximizes $\mathcal{R}$@5 [5 points]

(f) Provide an example ranking for this query that maximizes $\mathcal{R}$@10 [5 points]

(g) You have reason to believe that the users of this system will want to examine *every* relevant document for a given query. In other words, you have recent to believe that users want *perfect* recall. You want to evaluate based on $\mathcal{P}$@K. Is there a *query-specific* method for setting the value of K that would be particularly appropriate in this scenario? What is it? (**Hint:** there is an evaluation metric called *R-Precision*, which we did not talk about in the lectures. Your answer should be related to *R-Precision*. Wikipedia might help.) [10 points]

(h) Compute average precision (AP). [10 points]

(i) Provide an example ranking for this query that maximizes average precision (AP). [5 points]

(j) Compute DCG_5 (i.e., the discounted cumulative gain at rank 5). [10 points]

(k) $NDCG_5$ is given by

$$NDCG_5 = \frac{DCG_5}{IDCG_5},$$

where $IDCG_5$ is the DCG_5 associated with the *ideal* top-5 ranking associated with this query. Computing $NDCG_5$ requires three steps. [10 points]

(i) What is the *ideal* top-5 ranking associated with this query (notice that the query has 1 *perfect* document, 1 *excellent* document, 3 *good* documents, 2 *fair* documents, and the rest of the documents are *bad*)?

(ii) $IDCG_5$ is the DCG_5 associated with the *ideal* ranking. Compute $IDCG_5$. (**Hint:** compute DCG_5 for your ranking proposed in part (i).)

(iii) Compute $NDCG_5$ using the formula above.

2: Precision-Recall Curves [15 points]

A Precision-Recall (PR) curve expresses precision as a function of recall. Usually, a PR-curve is computed for each query in the evaluation set and then averaged. For simplicity, the goal in this question is to draw a PR-curve for a *single* query. Draw the PR-curve associated with the ranking above (same query, same results). **Hint:** Your PR curve should always go down with increasing levels of recall.