COMP3009J Information Retrieval

Worksheet 8: IR Evaluation - Programming

An Information Retrieval system has returned a set of results in response to a query. These results are (as Python list):

retrieved = ['d12', 'd1', 'd19', 'd15', 'd2', 'd4', 'd8', 'd9', 'd6', 'd14', 'd3', 'd5', 'd16', 'd18', 'd13', 'd20', 'd7', 'd11']

Relevance judgments have been performed and some documents have been judged to be relevant or non-relevant. These are presented as a Python dictionary below. For these judged relevant documents, the level of relevance is included after each (0 means that it has been judged non-relevant). Any document not included in this dictionary has not been judged.

relevant = { 'd1':3, 'd3':2, 'd7':1, 'd10':3, 'd11':2, 'd16':3, 'd17':2, 'd18':1, 'd2':0, 'd6':0, 'd8':0, 'd12':0, 'd13':0, 'd14':0, 'd15':0, 'd20':0 }

Write a program that calculates the following evaluation metrics.

1. Precision
2. Recall
3. P@10
4. R-Precison @ R=30%
5. MAP
6. bPref
7. NDCG@10

Advanced

Write a self-study tool: generate random list of document IDs and randomised relevance judgments. Present a problem to the user so that they can manually calculate the correct evaluation results. Then calculate correct answers so they can check their answers afterwards.