Patrick Kelly

12/3/2021

COMPSCI 446

Project 5 Report

Project 5 was written in Python using the python standard library. My implementation reads in the qrels file for judgements and stores them in a dictionary containing a key to a list of tuples. Then my program reads the trecrun files into their own dictionary and each entry in the dictionary holds a tuple in the form of rank, document, score. My program then handles all the six different measures that are required for this project. Each measure has its own function that follows the description/formulas in the book and class. My program then runs all the necessary calls that are outlined in the project documentation on moodle.

	NDCG@15	MRR	P@5	P@10	Recall@10	F1@10	MAP
BM25	1.9056	0.2238	0.1548	0.1393	0.0417	0.0641	0.0766
QL	5.7008	0.7004	0.4755	0.4274	0.1387	0.2094	0.2517
SDM	5.8454	0.6730	0.4900	0.4339	0.1423	0.2143	0.2586
Stress	2.6734	0.5673	0.3073	0.2297	0.0548	0.0884	0.0495

The data in the above table is what I would expect. QL and SDM are similar, while BM25 is based on a different language model and stress is unrelated to all 3. The MAP (mean average precision) of a query with no retrieved documents would be 0. This is because the precision score would be 0 therefore it would be 0/n where n is the number of docs.

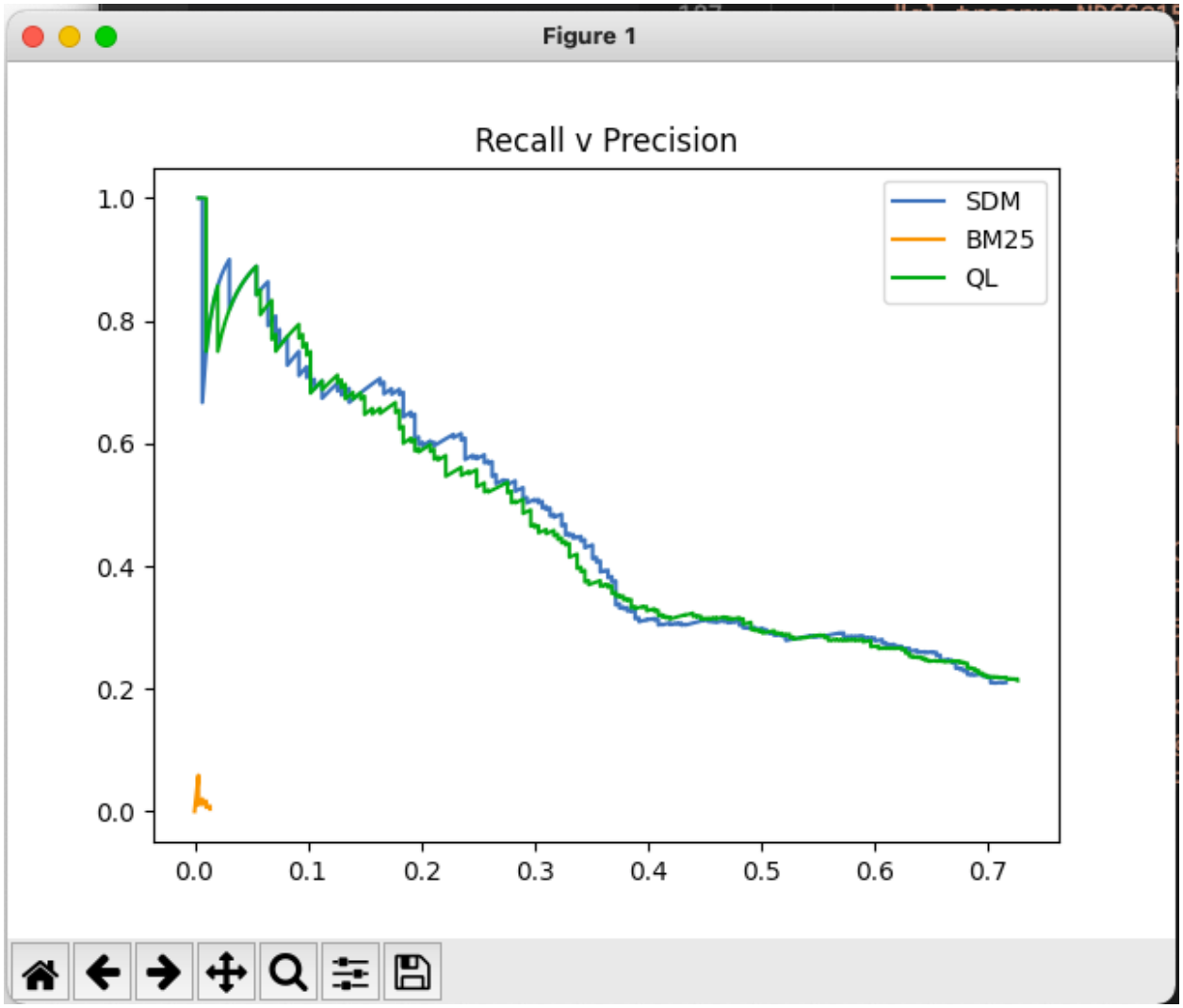
Figure 1
Recall v Precision
SDM
BM25
QL
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7