

CS221: Information Retrieval

Assignment 7

1. NDCG@5 for 1 and 2:

			NDCG 1	NDCG 2	
1	mondego		0.81699	0.918769	
2	machine learning		0.03304	0	
3	software engineering		0	0.00699	
4	security		0	0.044481	
5	student affairs		0	0.807879	
6	graduate courses		0	0.862662	
7	Crista Lopes		0.5681	0.966954	
8	REST		0.6654	0.665448	
9	computer games		0	0.343503	
10	information retrieval		0.39	0.190129	
		AVG	0.247	0.4806814	

2. What we did to improve NDCG@5:

We included various ranking factors on top of the basic search engine developed for Milestone 2. The different ranking factors are as follows:

- Including a separate Fancy Index for Special HTML tags such as <title>, <h1>, <a>, etc. to make use of the important information embedded in the html.
- Used Pagerank as one of the factors for ranking.
- Made use of Abbreviation expansion to get better results.
- Included Proximity information to find the documents that match the query as closely as possible.
- Made use of information in the URL of the search engine to determine the relevance factor.

A linear model of the above mentioned factors was tuned appropriately to give best results.