Mid Term

CSE 508, Information Retrieval

Question 1

Assume a collection of document where each document is also tagged with a location. Design an indexing mechanism which will allow Text only queries (this is traditional inverted index) and Text + Spatial queries? Each location is denoted by (x, y). (10 Points)

For extra credit, argue if the scheme will work if each document is tagged with multiple locations. If not then extend it to support multiple locations per document. (5 Points)

Question 2 – Programming

You are given a QA data set build out of Wikipedia articles. For each question there could be only one answer or multiple answers. Each row in a dataset is as follows:

```
WikipediaArticleTitle    Question        Answer
```

There are some questions that have multiple answers, and there are some questions that have only one answer. When there are multiple answers then there are multiple rows in the dataset where the value of question column remains same.

Part I:

Corpus: Question Answer pair

Input: Question

Output: ranked list of answers. The system does not know which is the most relevant answer. It must find the right answer

Evaluation: calculate Precision, Recall, NDCG

Note: Discard the questions whose answers are yes and no. (10 Points)

Part II:

Corpus: Question and documents

Input: question

Output: ranked list of documents. The system does not know which is the most relevant document. It must find the relevant document.

Evaluation: calculate Precision, Recall, NDCG (10 Points)

Part III:

Because you know the document (Wikipedia article), the question and the answer, you can therefore verify if for the same question the output of part I is same as output of part II. (5 points)