**The LNM Institute of Information Technology**

mid Semester Exam-2015

Information Retrieval and Web Search(IRWS)

Max Marks: 20

Q1. What is the role and type of relevance feedback? Why do commercial web search engines typically not provide relevance feedback functionality? Give at least 2 reasons. **[4]**

Q2. Describe TF-IDF weighting scheme for text representation. Discuss its advantages and disadvantages. Can you suggest some variations to TF-IDF weighting? **[4]**

Q3. Which are the parameters used for retrieval performance evaluation? Explain with suitable example. [**4]**

Q4. Consider a collection with the documents: **[2+2+4]**

**document1**: the way to the school is long and hard when walking in the rain

**document2**: the rain has not stopped in days and the school has closed

**query**: school closed rain

1. What would be the similarity score of the query with each document given above using Jaccard coefficient,
2. if you assume there are no stop words for this question.
3. The list of stop words is (the, to, is, and, in, has, not )
4. Construct an efficient inverted index for the above IR system. How can you reduce the size of the inverted index?

For your convenience, you can use the table above to note the frequencies.

|  |  |  |  |
| --- | --- | --- | --- |
| **term** | **document 1** | **document 2** | **query** |
| the |  |  |  |
| way |  |  |  |
| to |  |  |  |
| school |  |  |  |
| is |  |  |  |
| long |  |  |  |
| and |  |  |  |
| hard |  |  |  |
| when |  |  |  |
| walking |  |  |  |
| in |  |  |  |
| rain |  |  |  |
| has |  |  |  |
| not |  |  |  |
| stopped |  |  |  |
| days |  |  |  |
| closed |  |  |  |