LNMIIT
The LNM Institute of Information Technology

The LNM Institute of Information Technology

Department of Computer Science and Engineering

Information Retrieval (IR)
Mid Semester Exam

Ujjawal Kumar
17UCS171.

Time: 1 hr 30 minute **Date: 30/09/2019** **Max. Marks: 20**

Instructions: 1) Look through the whole exam and answer the questions that you find easiest first.
2) If necessary, you may make assumptions that are reasonable, and if you do make an assumption, state it clearly.
3) You may use a calculator.

Q1. (a) Why are term frequency and inverse document frequency used so often in document scoring functions? Explain with suitable example.
(b) How do stopping and stemming reduce the size of an inverted index? Explain with suitable example.
(c) Discuss efficient way of referring inverted index for information retrieval **[2X3]**

Q2. (a) Why do commercial web search engines typically not provide relevance feedback functionality? Give at least 2 reasons.

(b) Comment on the statement **"Pseudo-feedback always increases precision and recall"** **[2X2]**

Q3. Consider a standard bag-of-words model for the document retrieval problem.
Suppose we query an IR system for the **query: "gold silver truck"**
The document corpus consists of three documents (D = 3) with the following content,

D1: "Shipment of gold damaged in a fire"
D2: "Delivery of silver arrived in a silver truck"
D3: "Shipment of gold arrived in a truck"

Sort and rank the documents in descending order according to the similarity values: **[5]**

Q4. The **20 ranked results** have been returned as a response to a query. We find results **2, 6, 8, 9, 10, 15, 16, 17, 18, and 20 relevant.** We know that there are in **total 20** relevant documents in the collection.
(a) Draw the precision-recall curve for the first 20 results shown above.
(b) What is F- measure at point 10 for this list of results? **[2.5X2]**

{Best of Luck}