Information Retrieval - Monsoon 2020

End Semester Examination - Online Mode (Set 2)

(For Students having Roll Nos. ending with EVEN numbers. This includes PhD Students as well)

Date: 16th December 2020 **Time: 09:00 AM - 11:00 AM**

Instructions: This End Semester Examination has two parts:

1) Written Exam - Descriptive Questions - 75 Minutes Duration followed by 15 mins break for uploading the scanned copy of the Answer Scripts
2) Quiz - Multiple Choice Questions - 15 Minutes Duration

Written Exam:

a) Read all questions carefully and answer them in A4 Sheets.
b) Answer all questions (no choice, unless otherwise mentioned) and avoid unnecessary/ trivial explanations.
c) Most importantly, NO answer should be written in Pencil. Final answers should be written using either a BALLPOINT pen or INK pen.
d) On the top right-hand corner of every A4 sheet (Answer Sheet), write your Roll No, Name, and keep the page number encircled. Answer Sheets having no student details would not be evaluated.
e) At the end of the examination, you will get a link to upload the scanned copy of the answer script in a single PDF format.
f) Late Submissions will not be accepted under any circumstances.
g) Most importantly, this is a proctored examination, and students are advised to keep their videos on during the entire duration of the examinations.
h) You must use the meet link sent to you only and you should not use the meet link sent to others (You can check Your Google Calendar for this link)

Descriptive Questions [5 Marks each]:

1) Consider the following Boolean Index

	d1	d2	d3	d4	d5	d6	d7
Brosnan	0	1	1	0	0	0	1
Sean	1	1	0	0	1	1	0
David	1	0	1	1	0	0	1
George	0	1	1	0	1	0	0
Roger	1	0	1	1	0	0	1
Timothy	0	0	1	0	0	0	1
Daniel	1	1	0	1	1	1	0

Now for each of the following boolean query, show their outcome:

a) Brosnan AND Sean NOT Roger
b) David OR George AND (NOT Daniel)
c) Timothy AND Sean OR Brosnan OR Daniel

2) Apply TFIDF to compute the scores for ranking of the documents (assume that there are N = 4 documents in the collection and apply log term frequency for computing TFIDF):
   d1: Fiji is an island country in the south
   d2: Seoul is closer to Japan
   d3: Japan is a country having more than one island
   d4: Fiji is reachable country from Japan
   query: "is Japan an island or country"

3) Answer the following questions:
   a) Draw the architecture of a Domain-Specific Crawler and briefly describe its parts
   b) What is query expansion? Illustrate it with an example

4) Define 11-pt interpolated average precision and apply it on the following ranked list (Assume: There are 16 relevant documents in the collection for the given query). Also, compute the cumulative gain of the given set of documents

Rank	Type
1	Non-relevant
2	Relevant
3	Relevant
4	Non-relevant
5	Non-relevant
6	Relevant
7	Relevant
8	Non-Relevant
9	Relevant
10	Relevant

5) Consider the query incidence vector: q = [1 1 0 0 0 1]
   Document Incidence Vectors:
   d1 = [1 1 0 1 0 1]
   d2 = [1 0 1 0 0 1]
   d3 = [0 0 1 0 1 1]
   d4 = [1 1 0 1 1 0]
   d5 = [1 0 0 1 0 1]
   Initialize p_i = 0.5 and consider the top 2 documents as the relevant documents.
   Now apply probabilistic ranking algorithm and derive the final ranking of the above documents.