PES University, Bengaluru

(Established under Karnataka Act 16 of 2013)

END SEMESTER ASSESSMENT (ESA) - May 2023

UE20CS332 - Algorithms for Information Retrieval and Intelligence Web

Total Marks : 100.0

1.a. Explain in detail, the steps involved in the inverted index construction. (6.0 Marks)

1.b. What is the minimum edit distance? Explain with an example. Compute the edit distance between 'cats' and 'fast' using the Levenshtein approach. (6.0 Marks)

1.c. Consider 2 documents as follows:
Doc1: I did enact Julius Caesar. I was killed I' the Capitol; Brutus killed me.
Doc2: So let it be with Caesar. The nobel Brutus hath told you Caesar was ambitious

For the above 2 documents construct the inverted index. Show sequence of (term, docid) pair and also show the output of the core indexing step. (6.0 Marks)

1.d. List and explain proximity operators in Westlaw. (2.0 Marks)

2.a. What is the basic idea of the BSBI algorithm?
Comment on how expensive BSBI is.
State limitations of BSBI. (6.0 Marks)

2.b. Convert 824 in variable byte encoding and 511 in gamma encoding. (4.0 Marks)

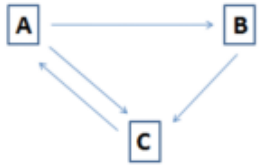
2.c. Consider the following documents and term frequencies. For simplicity do not consider Idf weighting. Compute log frequency weighting assuming idf=1. After length normalization, calculate cosine similarity between (Doc1, Doc2) and (Doc1, Doc3).

Term	Doc1	Doc2	Doc3
Term1	115	58	20
Term2	10	7	11
Term3	2	0	6
Term4	0	0	38

(10.0 Marks)

3.a. Consider the following graph. Let the initial page rank of all three nodes be 1/3.
According to the Basic page rank update rule, after 1st iteration, what is the page rank of Node C and Node B?

(4.0 Marks)



3.b. Compute precision, recall and F-score from the table:

	Relevant	Not relevant
Retrieved	20	40
Not retrieved	60	1,000,000

(3.0 Marks)

3.c. Explain in detail with the help of the diagram, the working of the URL frontier of a web crawler. (6.0 Marks)

3.d. What are IN and OUT components in a directed graph? In a Bow Tie structure of the web graph, what are the three components we see? (7.0 Marks)

4.a. Explain 4 common operational and technical goals of the recommender system. (8.0 Marks)

4.b. Briefly explain memory based collaborative filtering. (6.0 Marks)

4.c. Explain with examples:
1. Support
2. Confidence
3. Association rule (6.0 Marks)

5.a. What is RDF statement?
Explain RDF triple with example. (4.0 Marks)

5.b. Explain RDF triple graph notations with example. (4.0 Marks)

5.c. Explain with example, OWL datatype property and object property. (6.0 Marks)

5.d. Explain in details, classification of ontologies on the basis of semantic spectrum. (6.0 Marks)