SRN | | | | | | | | | | | | |

PES University, Bengaluru

(Established under Karnataka Act No. 16 of 2013)

UE18CS332

AUGUST 2021: END SEMESTER ASSESSMENT (ESA) B TECH VI SEMESTER

UE18CS332 – Algorithms for Intelligence Web and Information Retrieval

Time: 3 Hrs	Answer All Questions	Max Marks: 100

1	a)	Why is "grep" not the solution to construct Term Document Incidence Matrix?	6
	b)	How can we enumerate all terms meeting the wild-card query pro*cent? How can we enumerate all documents containing such terms?	6
	c)	Doc 1: I did enact Julius Caesar I was killed i' the Capitol; Brutus killed me. Doc 2: So let it be with Caesar. The noble Brutus hath told you Caesar was ambitious For the above documents construct an Inverted Index	8
2	a)	List the Basic Idea of the BSBI Algorithm.	6
	b)	How Expensive is BSBI ? Compare and contrast the BSBI VS Merge sort.	6
	c)	Why Compression in General and in Information Retrieval? Compare and contrast the Lossless VS Lossy Compression.	8
3	a)	Consider an information need for which there are 4 relevant documents in the collection. Contrast two systems run on this collection. Their top 10 results are judged for relevance as follows (the leftmost item is the top ranked search result): System 1 R N R N N N N N R R System 2 N R N N R R R N N N a. What is the MAP of each system? Which has a higher MAP? b. Does this result intuitively make sense? What does it say about what is important in getting a good MAP score? c. What is the R-precision of each system? (Does it rank the systems the same as MAP?)	6
	b)	Give three reasons why relevance feedback has been little used in web search.	6

SRN | | | | | | | | | | | | |

	c)	Below is a table showing how two human judges rated the relevance of a set of 12 documents to a particular information need (0 = nonrelevant, 1 = relevant). Let us assume that you've written an IR system that for this query returns the set of documents {4, 5, 6, 7, 8}.	8

docID	Judge 1	Judge 2
1	0	0
2	0	0
3	1	1
4	1	1
5	1	0
6	1	0
7	1	0
8	1	0
9	0	1
10	0	1
11	0	1
12	0	1

a. Calculate the kappa measure between the two judges.
b. Calculate precision, recall, and F 1 of your system if a document is considered relevant only if the two judges agree.
c. Calculate precision, recall, and F 1 of your system if a document is considered relevant if either judge thinks it is relevant.

4	a)	Which Model is Most Risky or Suitable for Advertiser?	6
	b)	Compare Locality Sensitive Hashing with General Hashing.	6
	c)	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?	8
5	a)	Briefly explain Memory based collaborative filtering.	6
	b)	With a neat diagram briefly explain six possible general classes for contextual factors	6
	c)	Explain four common operational and technical goals of recommender systems.	8