R.V.R.&J.C. COLLEGE OF ENGINEERING, GUNTUR-522019
(AUTONOMOUS)

III/IVB.Tech., (CSE) **CS314 (CSEL02) - Information Retrieval** Time: 45 Mts.

2022-2023(5th Semester) **ASSIGNMENT TEST # I** Max.Marks:12

1. (a). Explain how to build an inverted index — CO1 3M
   (b). write algorithms for processing Boolean queries. — CO1 3M

2. Explain the process of determining vocabulary of terms — CO1 6M

3. (a). Explain how to process wild card queries. — CO1 3M
   (b. Explain forms of spelling correction. — CO1 3M

4. Explain the following algorithms.
   (a). BSBI — CO2 3M
   (b). SPIMI — CO2 3M

5. (a). Explain dictionary compression techniques. — CO2 3M
   (b). Explain posting compression techniques. — CO2 3M

6. (a). Explain vector space model for scoring. — CO3 3M
   (b). Explain weighted zone scoring. — CO3 3M

SESSIONAL TEST#1

1.(a). What are the challenges of Information Retrieval? — CO1 1M
(b).Difference between Parametric index & Zone index — CO3 1M
(c). Differentiate lossy compression & lossless compression. — CO2 1M
(d). What is the gamma code for 12? — CO2 1M
(e). Consider the documents — CO1 1M

D1: "a dog will bark at a cat on a tree"
D2: "ants eat the bark of a tree"
Find term-document incidence matrix

(f). Write down the entries in the permuterm index dictionary that are generated by the term "mama". — CO1 1M

2. (a). List out the indexes used in information retrieval. Explain the purpose and functionality of each index. — CO1, CO2, CO3 6M

(Or)

(b). Define edit distance. Compute the edit distance between "paris" & "alice". CO1 2M

(c). Consider the following documents.

doc1 phone ring person happy person
doc2 dog pet happy run jump

420C529

doc3 cat run pet person happy
doc4 lite smile run happy
doc5 like laugh happy run run

(i). Construct the inverted index. CO1 2M

(ii). Explain how the processing of the following Boolean query
Happy AND run AND pet CO1 1M

(iii). How to optimize the query? CO1 1M

3. Differentiate

(a). Dynamic indexing & static indexing. CO2 2M

(b). Dictionary compression & Postings compression CO2 2M

(c). Boolean queries & Phrase queries. CO1 2M

(Or)

(d). Differentiate the two classic models in information retrieval CO1, CO3 2M

(e). Given the query "Elvis music" and the following term frequencies for the three documents doc1, doc2 and doc3: CO3 2M

Calculate the cosine similarity between the query and each document in order to rank

0.2170 0.417 0.7833

docs	Elvis	Presley	Mississippi	pop	music	life
doc1	3	4	0	6	0	0
doc2	4	0	4	0	0	3
doc3	5	3	0	4	4	0

(f). Explain various measures for finding relevance in vector space model. CO3 2M