ADAMAS UNIVERSITY PURSUE EXCELLENCE	**ADAMAS UNIVERSITY** **END-SEMESTER EXAMINATION : MAY 2021** (Academic Session: 2020 – 21)		
Name of the Program:	B.Tech	**Semester:**	VIII
Paper Title :	Information Retrieval	**Paper Code:**	ECS44104
Maximum Marks :	40	**Time duration:**	3 Hrs
Total No of questions:	8	**Total No of Pages:**	2
(Any other information for the student may be mentioned here)			

Answer all the Groups

Group A

Answer all the questions of the following $5 \times 1 = 5$

1. a) Define Posting.
   b) What do you understand by biword indexes.?
   c)What is the requirement of skip pointers?
   d)Define precision
   e)What is grepping?

GROUP –B

Answer *any three* of the following $3 \times 5 = 15$

2. State Rocchio's Classification Algorithm.

3. State Soundex algorithm for phonetic correction with example**.**

4. State the differences between dynamic and distributed indexing.

5. Why skip pointers are not useful for the queries in the form x or y? Discuss porter stemmer.

GROUP –C

Answer *any two* of the following $2 \times 10 = 20$

6. Given the query “elvis music” and the following term frequencies for the three documents doc1, doc2 and doc3 :

	elvis	presley	mississippi	pop	music	life
DOC 1	3	4	0	6	0	0
DOC 2	4	0	4	0	0	3
DOC 3	5	3	0	4	4	0

The Rocchio algorithm is a classic algorithm for implementing relevance feedback. Use Rocchio to compute the new query vector for “elvis music” using doc3 for relevance feedback (i.e., doc3 has been marked as relevant). Give suitable values for Rocchio’s weight parameters. As above, calculate cosine similarity (you can ignore the idf term) in order to rank the documents in order of relevance. Show your workings.

7. State Levenshtein distance algorithm and its necessity, and also calculate the distance between ‘**final**’ and ‘**first**’

8. Consider the following documents:

DOC1	phone ring person happy person
DOC2	dog pet happy run jump
DOC3	cat purr pet person happy
DOC4	life smile run happy
DOC5	life laugh walk run run

a) Construct the inverted index required for ranked retrieval for these five documents. Assume that no stemming or stop-word removal is required.
b) What is the complexity of processing a two-term conjunctive query using standard postings lists?
c) Briefly describe one technique that can improve this efficiency.
d) Relating to the sample documents above, outline how the processing of the following Boolean query can be optimised:

happy AND run AND pet

———————