

[Total No. of Questions: 12]

[Total No. of Printed Pages: 2]

UNIVERSITY OF PUNE

[4364]-802

B. E. (IT) Examination - 2013

Information Retrieval (2008 Course)

[Time: 3 Hours]

[Max. Marks: 100]

SECTION -I

- Q.1 A Explain Conflation Algorithm in detail and explain in short “steps to conflate” the following words: Here, Hereby, Hereafter, Herein, Hereupon. 10
- B Explain exhaustively and specificity with respect to Index term weighting. 8

OR

- Q.2 A Describe various classification methods based on the relations among properties, classes and objects. 10
- B Explain single Link Algorithm in detail. 8

- Q. 3 A Explain various IR models in detail with their advantages and disadvantages. 8
- B Explain the differences between suffix array and suffix tree. 8

OR

- Q. 4 A Explain the concept of Inverted index file. How it can be used Information Retrieval. 8
- B Explain the different kinds of search strategies. 8

- Q. 5 Consider a reference collection and its set of example information requests. If q is the information request and a set $R_q = \{d_3, d_5, d_9, d_{25}, d_{39}, d_{44}, d_{50}, d_{70}, d_{80}, d_{120}\}$. Now, consider new retrieval algorithm has been designed and has been evaluated for information request q returns, ranking of the documents in the answer set as. 16

- | | | |
|--------------------------------|--------------------------------|-----------------------------|
| 1. <u>d_{120}</u> | 6. <u>d_9</u> | 11. d_{38} |
| 2. d_{84} | 7. d_{58} | 12. d_{48} |
| 3. <u>d_{50}</u> | 8. d_{129} | 13. d_{230} |
| 4. d_6 | 9. d_{143} | 14. d_{113} |
| 5. d_8 | 10. <u>d_{25}</u> | 15. <u>d_3</u> |

The documents that are relevant to the query q are underlined. Calculate precision and recall for the documents that are relevant to the query q . Also, explain

average precision at seen relevant documents and R-Precision with the help of above example.

OR

- | | | | |
|------|---|--|---|
| Q. 6 | A | Discuss the architectural issues in Digital Library. | 8 |
| | B | Explain the terms
Harmonic mean, E-measure, R- precision and precision histogram. | 8 |

SECTION II

- | | | | |
|------|---|--|----|
| Q. 7 | A | Explain distributed IR with the help collection partitioning, source selection and query processing. | 10 |
| | B | Explain the differences between parallel IR and distributed IR. | 8 |

OR

- | | | | |
|------|---|--|----|
| Q. 8 | A | Describe MIMD architecture with respect to parallel IR.
How is inverted file used for MIMD? | 10 |
| | B | What is parallel computing? Discuss commonly used taxonomies of parallel architectures. | 8 |

- | | | | |
|------|---|---|---|
| Q. 9 | A | Explain GEMINI approach for multimedia IR. | 8 |
| | B | What is multimedia IR? Explain the steps multimedia IR. | 8 |

OR

- | | | | |
|-------|---|--|---|
| Q. 10 | A | What is multimedia information retrieval? Explain query specification query processing with respect to multimedia information retrieval. | 8 |
| | B | Explain the feature extraction and distance function for 2D color image. | 8 |

- | | | | |
|-------|---|--|---|
| Q. 11 | A | Explain the different components of web crawler. | 8 |
| | B | Explain centralized search engine architectures, comment on drawbacks of harvest distributed architecture. | 8 |

OR

- | | | | |
|-------|---|---|----|
| Q. 12 | A | Write short notes on
i) Challenges in web search
ii) Digital Libraries. | 16 |
|-------|---|---|----|