

Total No. of Questions : 12]

SEAT No. :

P1441

[4759] - 195

[Total No. of Pages :3

B.E. (Information Technology)

INFORMATION RETRIEVAL

(2008 Pattern) (414449) (Semester - II)

Time :3 Hours]

[Max. Marks : 100

Instructions to the candidates:

- 1) *Answers to the two Sections should be written in separate answer books.*
- 2) *Answer Q1 or Q2, Q3 or Q4, Q5 or Q6 from Section -I & Q7 or Q8, Q9 or Q10, Q11 or Q12 from Section - II.*
- 3) *Neat diagrams must be drawn wherever necessary.*
- 4) *Figures to the right side indicate full marks.*
- 5) *Use of Calculator is allowed.*
- 6) *Assume suitable data, if necessary.*

SECTION - I

Q1) a) Describe 'index term weighing'. **[8]**

b) Explain the properties of dissimilarity coefficients used in information Retrieval. **[8]**

OR

Q2) a) Describe different matching coefficients. **[8]**

b) Let Document 1 =

{CPU, keyboard, RAM, VGA, SMPS, USB, CD-ROM, Printer}

Document 2 =

{CPU, VGA, Simulator, OS, Video, USB, Printer, Scanner, Compiler}

Find the similarity between two documents using different matching coefficients. **[8]**

Q3) a) Explain with example signature file structure. **[8]**

b) Compare with example suffix trees & suffix arrays. **[8]**

OR

Q4) a) Draw the generalized structure of an inverted file & Explain the algorithm for building an Inverted file of a given document. **[10]**

P.T.O.

- b) Generate an Inverted file for a given text. [6]

‘This is a text. A text has many words. Words are made form letters’.

- Q5)** a) Explain the architectural issues in digital libraries. [8]

- b) Explain the terms Precision and Recall and calculate the same for the following example. [10]

The set of relevant documents for query

$q = \{d3, d7, d8, d11, d14, d19, d23, d25\}$

A new retrieval algorithm returns following answer set

$= \{d1, d2, d3, d7, d9, d10, d14, d20, d23, d24, d25\}$.

OR

- Q6)** a) Explain TREC document collection with tasks & Evaluation measures at TREC conferences. [10]

- b) Explain different document models, its representations and access. [8]

SECTION- II

- Q7)** a) Explain with diagram and example how the MIMD architecture is used in parallel IR. [8]

- b) Explain with example how the ‘Query Processing’ is done in distributed IR. [10]

OR

- Q8)** a) Explain with diagram how inverted file is used in MIMD architecture. [8]

- b) Explain Collection Partitioning & Source Selection in Distributed IR. [10]

- Q9)** a) Explain multimedia data support in commercial DBMS. [8]
b) Write short note on MULTOS query language. [8]

OR

- Q10)** a) Explain how GEMINI is applied to color images. [8]
b) What is feature extraction in Multimedia IR? How is it helpful for data retrieval? [8]

- Q11)** a) What is role of web crawlers in search engine? How it works? [8]
b) Explain distributed architecture of search engine. [8]

OR

- Q12)** a) Write short note on: Search Engines. [8]
b) Write short note on: Characterizing the Web. [8]

