

Reg. No. :

--	--	--	--	--	--	--	--	--	--

Question Paper Code: 42933

M.E.DEGREE EXAMINATION, MAY 2016

Elective

Computer Science and Engineering

14PCS508 - WEB DATA MINING

(Regulation 2014)

Duration: Threehours

Maximum: 100 Marks

Answer ALL Questions

PART A - (5 x 1 = 5 Marks)

1. Metadata describes _____
 - (a) Contents of database
 - (b) Structure of contents of database
 - (c) Structure of database
 - (d) Database itself
2. K-nearest neighbor is one of the _____
 - (a) Learning technique
 - (b) OLAP tool
 - (c) Purest search technique
 - (d) Data warehousing tool
3. Odd one out?
 - (a) Popularity
 - (b) Authority
 - (c) Prestige
 - (d) Impact factor
4. The extract process is which of the following?
 - (a) Capturing all of the data contained in various operational systems basic path
 - (b) Capturing a subset of the data contained in various operational systems graph testing
 - (c) Capturing all of the data contained in various decision support systems dataflow
 - (d) Capturing a subset of the data contained in various decision support systems

5. Which of the following is not a part of the web log preparation process?
(a) De-spidering (b) Sessionization (c) Path completion (d) Reporting

PART - B (5 x 3 = 15 Marks)

6. What is the use of meta search?
7. What are the two methods of ensemble classifier and explain it?
8. Compare co-citation and bibliographic coupling.
9. Write the crawler algorithm.
10. What you meant by pre-processing of web usage data?

PART - C (5 x 16 = 80 Marks)

11. (a) Write any two algorithm for association rule mining with examples. (16)

Or

(b) Discuss the following

(i) Latent semantic indexing (6)

(ii) GSP mining algorithm (10)

12. (a) Compare Naive Bayesian classification with support vector machines. (16)

Or

(b) Discuss in detail about unsupervised learning models. (16)

13. (a) Describe the working of web community discovery. (16)

Or

(b) How to analysis social networks and evolution of social networks? (16)

14. (a) How to extracting evolution of web community from a series of web archive? (16)

Or

(b) Summaries the automatic wrapper generation with examples. (16)

15. (a) Discuss the latent dirichlet allocation model and its applications. (16)

Or

(b) Explain in details about web usage mining using probabilistic latent semantic analysis. (16)