



M.E DEGREE EXAMINATIONS:DEC 2015

(Regulation 2014)

Third Semester

COMPUTER SCIENCE AND ENGINEERING

P14CSTE61 : Web Mining

Time: Three Hours

Maximum Marks: 100

Answer all the Questions:-

PART A (10 x 1 = 10 Marks)

1. Match list I with list II and select the correct answer for the given below table

CO2 [K₁]

List I	List II
A. K-means clustering	i. Social Network analysis
B. Prestige	ii. String Matching
C. Spider traps	iii. Unsupervised Learning
D. Edit distance	iv. Crawler

	A	B	C	D
a)	i	iii	iv	i
b)	iii	i	iv	ii
c)	ii	iii	i	iv
d)	iv	ii	iii	i

2. The basic machine learning processes are _____

CO2 [K₁]

- a) Preprocessing and testing
- b) Training and testing
- c) Preprocessing and training
- d) Training and transformation

3. The main term in spam techniques

CO3 [K₂]

- i. Repeating some important terms
 - ii. Out-link spamming
 - iii. Dumping of many unrelated terms
- a) i only
 - b) i and ii only
 - c) ii and iii only
 - d) i and iii only

4. Hypertext Induced Topic Search(HITS) is

CO3 [K₂]

- a) Ranking algorithm
- b) Classification algorithm
- c) Data extraction algorithm
- d) Page extraction algorithm

5. The tasks to be performed in document object model tree using HTML code is

CO4 [K₂]

- a. HTML code cleaning

- | | | |
|--|-----|-------------------|
| 14. Outline the part of speech tagging. | CO3 | [K ₂] |
| 15. How is the intelligent crawler is ranked using the web pages? | CO3 | [K ₂] |
| 16. Summarize strengths and weakness of Hypertext Induced Topic Search (HITS). | CO4 | [K ₂] |
| 17. Compare list pages and detail pages. | CO4 | [K ₂] |
| 18. What are types of matching based on the input information? | CO5 | [K ₂] |
| 19. List the steps for web usage mining process. | CO1 | [K ₁] |
| 20. What is meant by sessionization? | CO5 | [K ₁] |

PART C (10 x 5 = 50 Marks)

- | | | |
|--|-----|-------------------|
| 21. Discuss how the sequential patterns are mined using GSP algorithm. | CO1 | [K ₂] |
| 22. Compare K-means clustering and hierarchical clustering. | CO2 | [K ₂] |
| 23. Explain about types of spam and spammers. | CO3 | [K ₂] |
| 24. Explain document sentiment classification based on supervised learning. | CO2 | [K ₂] |
| 25. Outline the citation analysis of scholarly publications. | CO4 | [K ₂] |
| 26. Explain high level architecture of a scalable universal crawler. | CO3 | [K ₂] |
| 27. Categorize schema level matching. | CO4 | [K ₂] |
| 28. Summarize the key elements of web usage data pre-processing. | CO1 | [K ₂] |
| 29. Explain about log feature extraction. | CO5 | [K ₂] |
| 30. Compare association rules based prediction filtering algorithm and matrix factorization filtering algorithm. | CO5 | [K ₂] |

PART D (2 x 10 = 20 Marks)

- | | | |
|--|-----|-------------------|
| 31. Explain data extraction based on a single list page. | CO3 | [K ₂] |
| 32. Compare the different page ranking algorithms. | CO4 | [K ₂] |
