Course Code : CST 419 – 1 / CST 409 – 1 **KOLP/RW – 19 /9625**

Eighth Semester B. E. (Computer Science and Engineering) Examination

Elective – IV

WEB INTELLIGENCE AND BIG DATA

Time : 3 Hours] [Max. Marks : 60

Instructions to Candidates :—

(1) All questions carry marks as indicated against them.

(2) Assume suitable data and illustrate answers with neat sketches wherever necessary.

1. (a) State and explain various V's that specify different characteristics of Big Data. 3(CO1)

OR

(b) Indexes are created for faster access of webpages over internet. Justify. What is the complexity of web index creation ? Explain. 3(CO1)

(c) Consider two websites as described below and suggest various methods for improving the page ranks of both the websites. Illustrate the improvement in page rank supported by calculations of page rank values.

Structure of Website 1 :

- Home page
- 3 other pages that have links to and fro from the home page.
- Also these 3 pages have 1 outlink each (may lead to external website)

Structure of Website 2 :

- Home page
- 2 other pages that have links to and fro from the home page.
- Also these 2 pages have 1 outlink each (may lead to external website) 7(CO1)

2. (a) Justify "TF – IDF can generate better search results over index created with only count of words in the document". 3(CO1)

(b) Where can the concept of bipartite graph be used. Illustrate with example. 2(CO1)

(c) There is a company that wants to judge the opinion of the customers about a newly launched product. The website contains the reviews about the product. The project manager intends to use Bayes classifier to find positive and negative sentiment of the customers. Justify how this can be done. State example for proper explanation. 5(CO1)

3. (a) Consider the following input documents and show the map – reduce process to find the word count of words in the set of documents. Use 3 mappers and 2 reducers.

D1 – w1 , w1 , w2 , w3 , w5 , w6
D2 – w1 , w3 , w4 , w5
D3 – w2 , w3 , w4 , w5 , w5
D4 – w2 , w2 , w3 , w3
D5 – w1 , w3 , w4 , w6
D6 – w1 , w2 , w3 , w4 , w5
D7 – w1 , w2 , w2 , w3 , w6
D8 – w1 , w3 , w2 , w5 , w6
D9 – w2 , w2 , w2 , w2
D10 – w3 , w4 , w4 , w5 , w6

5(CO2)

OR

(b) Describe the responsibilities of master and worker nodes in Map – Reduce process as a new user program arrives.
Also find the parallel efficency of Map – Reduce of Word counting task and

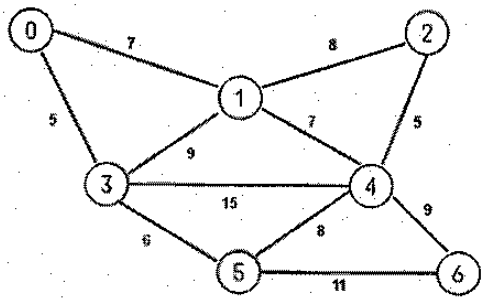
(i) State it's relation with number of processors.

(ii) What will happen if number of documents being processed is same and number of processors is increased ? 5(CO2)

(c) "Hadoop is fault tolerant, consistent, can cope with node failures and store data in distributed form" Justify. 5(CO2)

4. (a) Physical shops can store and display limited items. Whereas, any online e – commerce website can display lakhs of items. What is the problem that online websites have ? Explain and state its solutions. 3(CO3)

(b) Perform graph based clustering to cluster similar interest users for the given graph. Nodes represent the users and edges show their common interests.



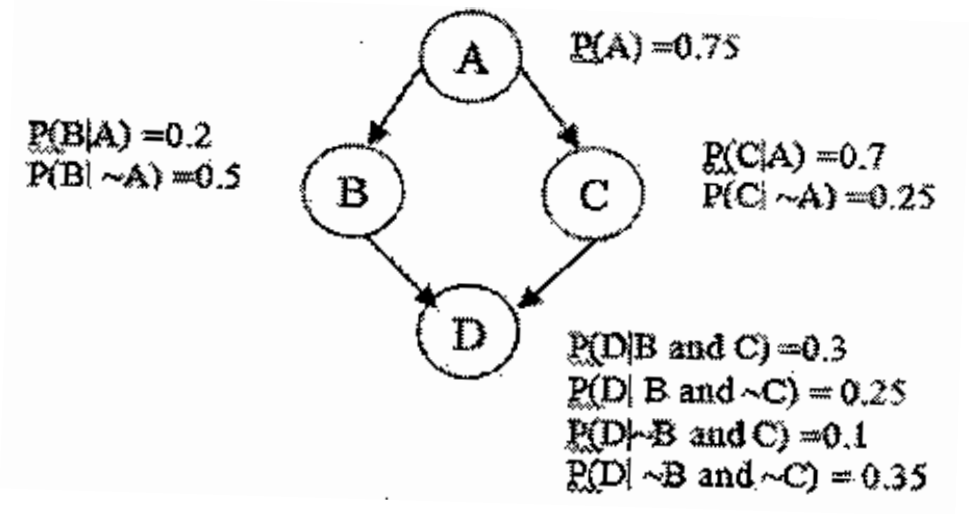
5(CO3)

(c) Discuss the use of clustering on web data. 2(CO3)

OR

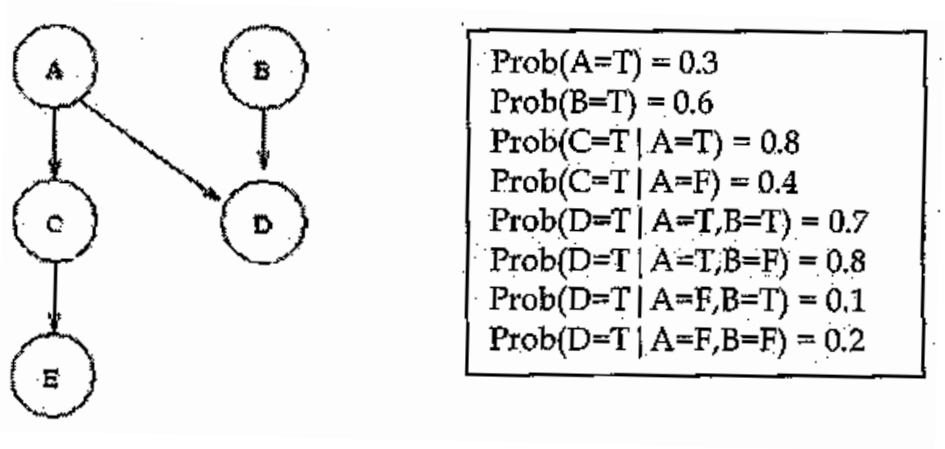
(d) Discuss the use of association rule mining. 2(CO3)

5. (a) Explain Bayesian nets and illustrate any application where Bayesian nets can be used.
Also consider the given Bayesian Network. A , B , C and D are Boolean random variables. If we know that A is true, what is the probability of D being true ?



6(CO4)

(b) Consider the following Bayesian network with probabilities :



The independence expressed in this Bayesian net are :
A and B are (absolutely) independent.
C is independent of B given A.
D is independent of C given A and B.
E is independent of A, B and D given C.
Compute the following :—

(a) P (D = T) (b) P (D = F , C = T) (c) P (A = T | C = T)
4(CO4)

OR

(c) Discuss the independence conditions in Bayesian Belief Networks ?
4(CO4)

6. (a) Design a Spare Distributed Memory (SDM).
Consider a Reference Address and a memory. Assume that there are 10 locations available to store the data.
Also assume any input data and show the memory with data and how to obtain the output same as what was given as input to be stored in the SDM.
Show the complete sparse memory and consider Radius = 3 and input, memory address of size 12 bits. 5(CO1)

(b) List the different components of Blackboard Architecture ? Explain them. Also design blackboard architecture for a mobile robot that is working as a chef in a kitchen environment of a hotel. 5(CO1)