COURSE CODE: 15B1WCI833   MAX. MARKS: 35
COURSE NAME: Big Data Analytics
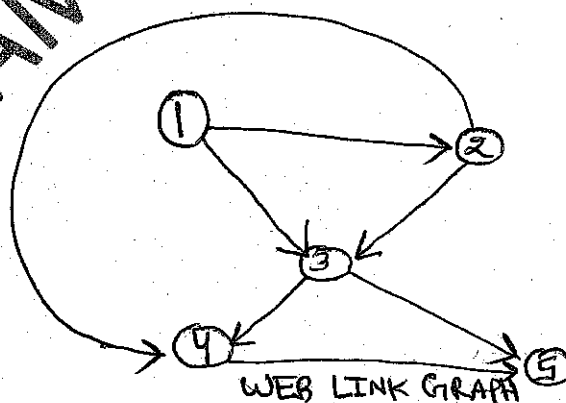COURSE CREDITS: 03   MAX. TIME: 2 HRS

Note: All questions are compulsory. Carrying of mobile phone during examinations will be treated as case of unfair means.

1. What is content based search in search engines? How it is better than other traditional search approaches? Explain search procedure of google with suitable example.   [1+1+5]

2. What do you mean by **a priori** property? Explain with suitable example. Consider a shopping mall basket as shown below:

Customer	Items
C1	Milk, egg, bread, chip
C2	Egg, popcorn, chip, beer
C3	Egg, bread, chip
C4	Milk, egg, bread, popcorn, chip, beer
C5	Milk, bread, beer
C6	Egg, bread, beer
C7	Milk, bread, chip
C8	Milk, egg, bread, butter, chip
C9	Milk, egg, butter, chip

Let the minimum support is 30% and minimum confidence is 80%. Generate the information using Apriori algorithm to describe the behavior of customers. Why association rules mining is a better target of big data analytics?

[10]

3. Consider the given below link graph and compute the page rank of each web page using PageRank (matrix algebra) and HITS (Hyperlink Induced Topic Search) method.   [10]

WEB LINK GRAPH

3. Consider 8 persons showing interest in personal loan from a reputed public sector bank. Let point(x,y,z) represents their age(in years), salary(in Indian rupees) and income from other sources(in Indian rupees). So the data is represented as Ram(20,35K,10K), Sita(30,25K,5K), Laxman(50,20K,12K), Anil(40,20K,2K), Prem(45,50K,15K), Sunil(21,46K,15K), Sourabh(23,32K,11K) and Meenu(32,27K,13K). The bank manager's task is to check whether the person is rich, medium earning or poor based on the given data. Use k-means clustering algorithm technique based on Euclidean distance to ease manager's task. What changes would you suggest in k-means algorithm in case points are assumed as evolving data?   [8]