5A. Consider the following English sentences. Represent these as predicate logic formulas. **5**

i. Anyone whom Mary loves is a football star.
ii. Any student who does not pass does not play.
iii. John is a student.
iv. Any student who does not study does not pass.
v. Anyone who does not play is not a football star.

Convert the predicate logic formulas into CNF and write the clauses.

5B. Using the clauses got in Q.5A, prove the conclusion "If John does not study, then Mary does not love John" using resolution. **3**

5C. Justify the following statement." Predicate logic is more expressible than propositional logic". **2**

Reg. No.

MANIPAL INSTITUTE OF TECHNOLOGY

MANIPAL

A Constituent Institution of Manipal University

VII SEMESTER B.TECH. (INFORMATION TECHNOLOGY/ COMPUTER AND COMMUNICATION ENGINEERING) END SEMESTER EXAMINATIONS, NOVEMBER 2017

SUBJECT: PROGRAM ELECTIVE IV – WEB INTELLIGENCE [ICT 4024]

REVISED CREDIT SYSTEM
(23/11/2017)

Time: 3 Hours MAX. MARKS: 50

Instructions to Candidates:

- Answer **ALL** the questions.
- Missing data may be suitably assumed.

1A. Consider the document and words along with their respective number of occurrences in the web as in Fig. Q.1A, Find the top keywords using TF-IDF. Assume that the total number of documents indexed in the web is 100 B. **5**

The past decade has witnessed the successful of application of many AI techniques used at `web-scale', on what are popularly referred to as big data platforms based on the map-reduce parallel computing paradigm and associated technologies such as distributed file systems, no-SQL databases and stream computing engines. Online advertising, machine translation, natural language understanding, sentiment mining, medicine, and national security are some examples of such AI based web-intelligence applications that are already in the public eye.

Words	Number of Hits
web-intelligence	0.6 B
Public	14 B
The	50 B
AI	4 B
map-reduce	0.4 B
Parallel	10 B

Fig. Q.1A

1B. Given the dataset in Table Q.1B, classify whether the comment "Students **enjoy** web-intelligence a **lot** since it is very **simple**" is positive or negative using Naïve Bayes classifier. **3**

Table Q.1B

Count	Comments	Sentiment
1000	Zeus **likes** dishu a **lot**	Positive
1800	Jerry **hates** tom, it feel it's a **waste** of time trying to be friends with tom	Negative
200	Athur has a **bad** fate, always ends up with a **wrong** name written on the documents	Negative
2000	Web intelligence subject is **simple** and **easy** to follow	Positive
2000	Students **like** a subject if its **simple** and they **enjoy** it	Positive
400	**Simple** subjects are not **interesting**	Negative
600	Ajay did not **enjoy** web intelligence	Negative

1C. Assuming n documents, m words, and w words per document and a balanced binary tree to store words, compute the complexity of index creation. **2**

2A. For the web graph given in Fig.Q.2A, compute the page rank using power method. Show the calculations for at least 5 iterations. **5**

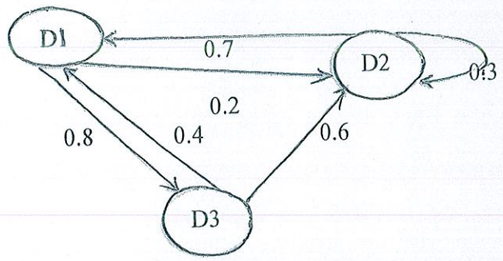


Fig.Q.2A

2B. Given the graph model as in Fig.Q.2B, write all RDF statements in Natural language sentence and Turtle notation. **3**

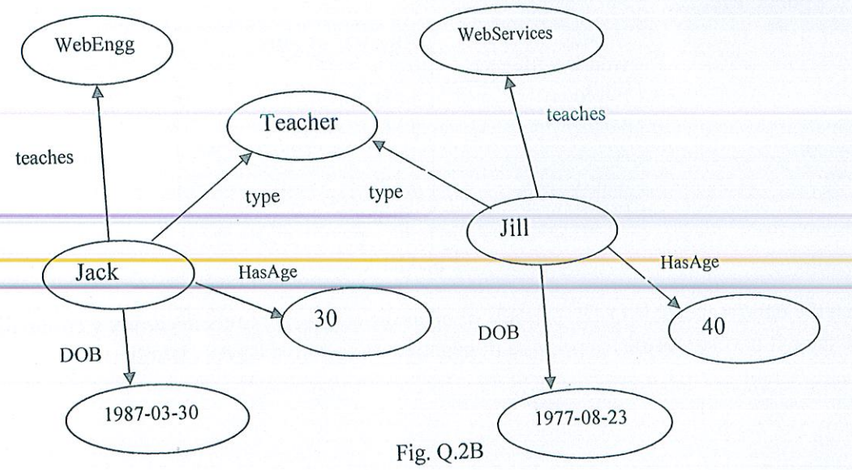


Fig. Q.2B

2C. While using a machine learning algorithm, how is mutual information helpful in choosing the features? Are more features always good? Justify. **2**

3A. Write a map reduce program to create an inverted index. **5**

3B. With a neat diagram explain the architecture of Hbase. **3**

3C. With an example, explain eventual consistency. **2**

4A. Consider the matrix given in Fig. Q.4A **5**

i. Compute the minhash signature for each column if we use the following three hash functions: $h1(x) = 2x + 1 \bmod 6$; $h2(x) = 3x + 2 \bmod 6$; $h3(x) = 5x + 2 \bmod 6$.
ii. Which of these hash functions are true permutations?
iii. How close are the estimated Jaccard similarities for the six pairs of columns to the true Jaccard similarities?

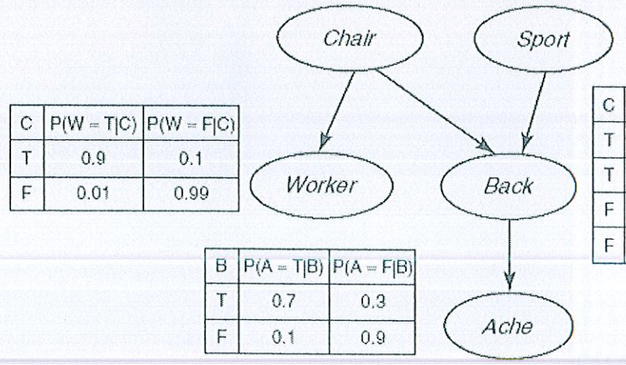
Element	S1	S2	S3	S4
0	0	1	0	1
1	0	1	0	0
2	1	0	0	1
3	0	0	1	0
4	0	0	1	1
5	1	0	0	0

Fig.Q.4A

4B. **3**

P(C = T)	P(C = F)
0.8	0.2

P(S = T)	P(S = F)
0.02	0.98



C	P(W = T|C)	P(W = F|C)
T	0.9	0.1
F	0.01	0.99

B	P(A = T|B)	P(A = F|B)
T	0.7	0.3
F	0.1	0.9

C	S	P(B = T|C,S)	P(B = F|C,S)
T	T	0.9	0.1
T	F	0.2	0.8
F	T	0.9	0.1
F	F	0.01	0.99

Fig.Q.4B

Consider the Bayesian belief network with the Probability distribution tables given in Fig. Q.4B. Compute the joint probability Distribution.

4C. How is classification different from regression? Explain how the best fit in linear regression is evaluated? **2**