

# Question Paper

Exam Date & Time: 27-Nov-2018 (10:00 AM - 01:00 PM)



## MANIPAL ACADEMY OF HIGHER EDUCATION

SCHOOL OF INFORMATION SCIENCES

SECOND SEMESTER MASTER OF ENGINEERING - ME (BIG DATA AND DATA ANALYTICS)

### Text Retrieval and Search Engines [BDA 616.1]

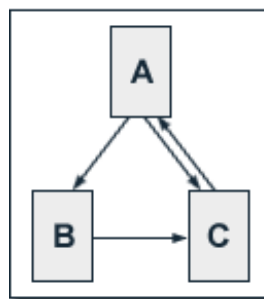
Marks: 100

Duration: 180 mins.

### END SEMESTER DEGREE EXAMINATION NOVEMBER 2018

**Answer all the questions.**

- 1) Describe the characteristics of the Web. 10 (10)
- 2) a) What is the feast or famine problem? Which Retrieval model faces this problem and why? 5 (10)  
b) Why don't we use grep for information retrieval? 5
- 3) Explain Single-pass in-memory Indexing. 10 (10)
- 4) Briefly explain the Porter Stemmer. 10 (10)
- 5) Briefly explain the steps involved in creating an inverted index with an example. 10 (10)
- 6) Explain the Map-Reduce Architecture with an example. 10 (10)
- 7) Calculate the Page Rank for the given graph with the teleportation factor  $\alpha=0.7$  (10)



- 8) Briefly explain the steps involved in creating an inverted index with an example. 10 (10)
- 9) You are given the following table that contains the term frequencies for certain novels. (10)

- i) Calculate the cosine score between each pair.
  - ii) What does the cosine score signify?
  - iii) List the document pairs in order of descending cosine scores.
- Take idf score = 1.

10

Term	SaS	PaP	WH
Affection	115	58	20
Jealous	10	7	11
Gossip	2	0	8
Wuthering	0	5	30

- 10) Briefly explain the different techniques used in duplicate detection. 10 (10)

-----End-----