COURSE CODE(CREDITS): L-11M1WCI431   MAX. MARKS: 15

COURSE NAME: Advanced web mining

COURSE INSTRUCTORS: Dr. Nishant Sharma   MAX. TIME: 1 Hour

Note: (a)All questions are compulsory.

(b)Marks are indicated against each question in square brackets.

(c) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems

Q1. (a)What is the significance of web crawling? What are the responsibilities of a load manager in a web crawler?

(b) Describe the working of a large-scale web crawler by using Mercator's crawler as example.

<div align="right">

[CO-1]   [2 + 4 marks]

</div>

Q2. Write a short note on following:

1. Spider traps
2. Pre-fetching
3. Multi-threading   [CO-2]   [1 + 1 + 2 marks]

Q3. Explain the role of index and inverted index in context of Information Retrieval? Create an index and inverted index for the documents provided below.

Document 1: Covers breaking news from around the world

Document 2: International news, analysis and commentary

Document 3: Internationally published free city news paper

<div align="right">

[CO-2]   [3 marks]

</div>

Q4. Why is index compression used in information retrieval? Describe a few index compression techniques in brief?   [CO-3]   [2 marks]