

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -1 EXAMINATION- FEBRUARY 2018

M.Tech IV Semester

COURSE CODE: 11M1WCI431

MAX. MARKS: 15

COURSE NAME: Advanced Web Mining

COURSE CREDITS: 03

MAX. TIME: 1 Hrs

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

1.

- a) Explain Inclusion and overlapping between WWW and Internet.
- b) Write down the algorithmic steps of a large-scale web crawl.
- c) Explain Politeness and Refresh policy regarding web crawling.
- d) Demonstrate Mercator's crawler for large-scale web crawling.
- e) How Politeness and Refresh policy is handled in Mercator's crawler.

[1+2+1+4+2]

2. Write short notes on the following:

- a) Stemming.
- b) Robot exclusion.
- c) Small world network.
- d) Power law distribution.
- e) URL canonicalization.

[5x1]