UNIVERSITY OF MALAYA

EXAMINATION FOR THE DEGREE OF MASTER OF DATA SCIENCE

ACADEMIC SESSION 2019/2020 : SEMESTER II

WQD7005 : Data Mining

June 2020

INSTRUCTIONS TO CANDIDATES :

Answer **ALL** questions (50 marks).

(This question paper consists of 5 questions on 3 printed pages)

Mini-assignment (50 marks)

Instructions: Work individually, submission via Spectrum.

. 1. You are required to make a user-agent that will crawl the WWW (your familiar domain) to produce dataset of a particular website.

* the web site can be as simple as a list of webpages and what other pages they

link to

* the output does not need to be in XHTML (or HTML) form

a multi-stage approach (e.g. produce the xhtml or html in csv format )

(10 marks)

1. Draw snowflake schema diagram for the above dataset. Justify your attributes to be selected in the respective dimensions.

(10 marks)

1. It will be appeared in week 13.
2. marks)
3. It will be appeared in week 13.
4. marks)
5. It will be appeared in week 14.

( 10 marks)

Submissions:

The student is expected to submit answers to each question individually, and submit the document in PDF format. The student can include online materials, screenshots, videos and/or codes (ipynb format) to support your answer