University of Glasgow | School of Computing Science

Assessed Coursework

Course Name	Web Science (M & MSc level) COMPSCI5078 & COMPSCI5107			
Coursework Number	**Geo Localisation**			
Deadline	**Time:**	**4:30PM**	**Date:**	**8 March 2024**
% Contribution to final course mark	**20%**			
Solo or Group ✓	**Solo**	**x**	**Group**	
Anticipated Hours	**20**			
Submission Instructions	**Submission through Moodle** **Report and Python codes (zipped)**			
Please Note: This Coursework cannot be Re-Assessed				

Code of Assessment Rules for Coursework Submission

Deadlines for the submission of coursework which is to be formally assessed will be published in course documentation, and work which is submitted later than the deadline will be subject to penalty as set out below.

The primary grade and secondary band awarded for coursework which is submitted after the published deadline will be calculated as follows:

(i) in respect of work submitted not more than five working days after the deadline
   a. the work will be assessed in the usual way;
   b. the primary grade and secondary band so determined will then be reduced by two secondary bands for each working day (or part of a working day) the work was submitted late.

(ii) work submitted more than five working days after the deadline will be awarded Grade H.

Penalties for late submission of coursework will not be imposed if good cause is established for the late submission. You should submit documents supporting good cause via MyCampus.

Penalty for non-adherence to Submission Instructions is 2 bands

You must complete an "Own Work" form via https://studentltc.dcs.gla.ac.uk/ for all coursework - If you are not signing the declaration of originality, then your marks will not be released.

Individual Assessment: Geo-Localisation

Coursework is due on Friday, March 8, 2024, 430 PM (subject to LTC approval)

CW is marked out of 100 marks & Weighted 20% to the final marks.

All submissions are through Moodle

Around 10–20K tweets from London/UK are in the Data directory on Teams. It is in json format. **I recommend you use Jupyter notebook and submit code and outputs archived.**

(i) A dataset will be given to you (In the Data folder teams). Write python code to organise tweets into grids of 1km x 1km. Draw charts and/or figures to analyse the distribution of data.
The coordinate system we used to collect data is
London = [-0.563, 51.261318, 0.28036, 51.686031]

[25]

In the report:

Describe the algorithm with a pseudo-code

(5 marks)

Provide statistics of the data (total tweets, how many are on the cells, and how it is distributed etc.) and interpret the statistics – what does this mean?

(5 marks)

Provide charts and/or figures for the visualisation of the grid data

(10 marks – 5 for code and description; 5 for output shown)

Describe your views/interpretation on the data (and the resulting visualisation) - you may want to highlight issues with any potential geo-localisation techniques with this data.

(5 marks)

(ii) You will be given a set of high-quality, low-quality and background tweets. Develop newsworthy scoring method based on this dataset. Empirically adjust the thresholds to modify newsworthiness and discuss the results.

[30]

In the report:

Explain your newsworthiness computation method along with an algorithm/pseudo-code (15 marks).

Conduct data analysis & provide an analysis of various thresholds; data analysis may include for example, using or not using stop words, adapting thresholds etc.
(15 marks)

(iii) Use the above newsworthy scoring techniques to analyse the geo-tagged data set given (i) and discuss the results

[25]

In the report:

Apply your scoring method to data in task 1 -

Investigate tweets with low scores and high scores; find an appropriate threshold to separate them and remove tweets with low newsworthy scores

Justify the threshold used with any supportive information you can produce from your empirical analysis.

(10 marks)

Provide statistics of the data (total tweets, how many are with certain newsworthy scores, and how it is distributed etc. *how many removed*, see below)

(5 marks)

Apply the visualization you created on newsworthy data; Draw the figures/charts and compare them with results in (1). What can we say about the difference?

(10 marks)

(iv) [Open tasks]

Identify and discuss, with examples, issues for geo-localisation due to the nature of tweets or sources (10)

(v) Report – 10 marks

a. *Structuring and formatting*
b. *Articulation of ideas*
c. *Creativity in addressing the tasks*