Propose a vacation student project for Summer 2018

Deadline for making your proposal(s): 17:00 Friday 23 March. This is a hard deadline.

Please fill in and submit this form multiple times for multiple proposals. Any queries, do ask - Toby.

Project supervisor email *

s.sampaio@manchester.ac.uk

Title of the project *

Experimenting with and Comparing Existing Data Wrangling Tools and Packages using Big Data Sets

Source of funding *

☐ School funding requested

☐ You have your own funding (e.g. research grant)

Objective of the project *

(1) To develop new material for an existing MSc-level course unit (COMP60711);
(2) To complement the literature review work of an on-going research effort.

Number of students requested (justify if > 1) *

1

Start date, end date, total duration (weeks) *

Start date: Mon 11/06; End date: Mon 30/07; Duration: 6 weeks
The benefit to the School *

Improvement of the contents of an existing MSc-level course unit (COMP60711) with content relating to main stream-research and technologies.

The benefit to the student *

Opportunity to learn how to handle voluminous, raw, heterogeneous and time series data, performing complex data management tasks such as data sampling, profiling, integration and cleaning.

Skills needed by the student. *

Programming (in Python and R), data analysis, mastering of new tools and languages (such as Grel and the Trifacta language), experimental design, implementation, evaluation and results interpretation, writing communication skills.

Details of the work that the student would do *

Process large and raw data using tools such as Trifacta, Open Refine, Talend and MySQL, as well as specific packages for the Python and R languages, given a number of data wrangling jobs derived from a list of data analysis requirements. Writing of a document detailing the entire process in each of the tools, packages, emphasizing advantages and disadvantages of each tool/package.

Infrastructure requirements and any required staff support other than the project supervisor *

Open source tools and packages (currently installed in the MSc labs); Big data sets, currently available from open sources.
Propose a vacation student project for Summer 2018

Supervision arrangements throughout the duration of the project (named staff and dates covering the entire duration) *

One-hour meetings every week throughout the duration of the project. Day and time of the meetings are to be agreed between the supervisor and the student.

Location of the project work (building/room) *

MSc lab.

This content is neither created nor endorsed by Google.

Google Forms