

ASSIGNMENT 1

(Worth 15% of the total assessment for this course)

Aim

The aim of this assignment is to develop skills in implementing simple programming tasks using MapReduce.

This is an individual assignment and you are required to submit original work, any work copied or derived from another source must be identified and details provided in your written submission.

Component of assessment

This assignment will be marked out of 15. The assignment contributes 15% towards your final mark for the course. It comprises a single component..

This assignment should read in conjunction with the main assignment (assignment 3) and is meant a preliminary exercise to prepare for the main assignment.

Using MapReduce, process the 250 abstracts in the trial set to produce a list of the 10 most frequent words in each abstract. The output should be stemmed with the porter stemmer and be in a format similar to assignment 3.

Submission requirements

You are required to submit your source code, evidence of having running your program (e.g. diagnostics produced by MapReduce, together with output files for up to three runs.

Due Date

11.59pm Friday 20th August 2010

Late submissions

Late submission of assignments will be penalised as follows:

For assignments 1 to 5 days late, a penalty of 10% (of total available marks) per day.

For assignments more than 5 days late, a penalty of 100% will apply.