

## ITEC 4305 Term Project

The term project can be done in a group of up to 3 members, and will be due and presented in classes in two stages on June 2 and July 7, respectively.

Stage 1[30 marks]: Preprocess the web page documents provided to select relevant terms to represent these documents. Your report and presentation (20 minutes) should include the methods or techniques and programs you developed in preprocessing, and results. All html tags should be cleaned, stop words be removed, and stemming be performed.

Stage 2[30 marks]: Implement a learning algorithm to classify test documents with the vocabulary selected in stage 1. Your report and presentation (20 minutes) should include learning programs developed and classification performance. You should use all possible techniques learnt in this course to achieve the best performance you can.