COMP 6751 # Assignment 2 **Fall 2018**

Goal Recognizing Textual Entailment using NLTK

Due date November 20, 2018

Environment NLTK

Description

1. Baseline System for Textual Entailment:

   Use the NLTK version of RTE datasets (from the RTE1, RTE2 and RTE3 challenges) and run the baseline module that comes with NLTK (see NLTK chap 6, section 2.2).

   Analyze the results. Are they all correct? Explain why some examples lead to a correct classification, while others don't.

2. Expand the baseline system of (1) to improve its performance. Experiment with different feature sets, classifiers and possible datasets. Do not train your system on the test sets of the RTE datasets. Use only the dev set for training and development.

   - 1 page: Quantitative Analysis: Report the overall performance of your various experiments in terms of precision, recall, F-measure and accuracy. Use the test sets provided in the package and the new test set that will be given to you a few days before the deadline. Tabular these results into a sigle table using the following format:

	RTE1-test				RTE2-test				RTE3-test				new-test			
	P	R	F	A	P	R	F	A	P	R	F	A	P	R	F	A
experiment 1																
experiment 2																
...																

   Compare the overall performances of the experiments. Make sure you report which modules you chose for each step and whether you tried and compared several different options.
   - $1\frac{1}{2}$ page: Qualitative Analysis: Report and explain 2 examples of errors that the baseline system made, but that your various experiments do not make. Report and explain 2 examples of errors that the baseline system made, and that your experiments still make. Discuss.

3. Test

   A few days before the deadline, a new test set will be posted on Moodle. You will process these instances with your systems and inspect the results the same way that you inspected the results in step (2) above.

   - 1 page: Report and analyze your results with this new test set in terns of qualitative assessments (error analysis). The quantitative assessment (i.e. precision, recall, ldots) should be included in the table mentioned above.

4. Interface

   Similarly to Project 1, make sure that you have a way to input a new test corpus that is then processed by the entire pipeline and displays all the annotations obtained and final classifications of the test instances on the screen without touching the code. This can be a command line call and pretty print output or a visualization. Provide an option to save/print this visualization.

Deliverables and Marking scheme

1. code (5pts, Grad Attr. 4,5,6)
2. 6 page report, one additional page for extra findings and experiments and references. All sources, papers, books that have influenced your work must be cited with proper references in addition to URLs, if possible. (5pts, Grad Attr. 1,6)
3. sample runs over training and challenge data

Thoroughness, creativity, and extra work that address pertinent issues in informed ways will be a bonus. That means number of words and colourful output do not make a difference, but considered discussion of the experiments might.