

Homework #1

Consider the following pairs of expressions.

(1) Extol, praise highly

(2) Destitute, very poor

In (1), the verb *extol* is contrasted to the verb *praise* in its intensity. In (2), the adjective *destitute* is contrasted to the adjective *poor* in its intensity. Notice that both relations are indicated by the presence of the intensity-modifying adverbs *highly* and *very*.

We have two underlying questions: First, can you think of a systematic way to identify such pairs above from the available text corpora? Second, can you think of intensity-modifying adverbs other than, or including, *highly* and *very*?

A Specify your own problem definition that abstracts over the two questions above, so that your problem asks you to B write a Python code to access some text corpora and to output results that contrast two expressions, such as *extol* and *praise*, or *destitute* and *poor*, together with the kind of intensity modification, such as *highly*, *very*, and so on.

The output must include at least 50 pairs of expressions of reasonable quality. We will discuss later how to assess the quality of the output with measures such as *precision* or *recall*, but for the purpose of this homework, you should C use your own argument in English to show why your output is of reasonable quality, and how you can improve its quality further by an extra stage of curation (automatic or manual correction/filtering). Please note that your output does not have to be perfect, but your argument matters.

All the requirements as underlined above must be composed by yourself and without help from anyone else, including the related resources on the Internet, if any. Any similarity of the results will be flagged for plagiarism and, if found sufficiently similar, penalized, up to, but not limited to, a failure to this homework.

Deadline for uploading your homework at KLMS: 79 April (11:59pm)

Homework Submission Guidelines

1. Submission files

- CS372_HW1_report_[your ID].doc for **A** and **C**
- CS372_HW1_code_[your ID].py for **B**
- CS372_HW1_output_[your ID].csv for the initial 50 triples

2. Remarks

- Use only 1 page for **A** your own problem description and **C** output justification.
- The code should include comments about your implemented idea.
- For the implementation, external models are not allowed.
- Your code should be runnable in our environment.
- For the output, use slicing [:50] to produce the initial 50 triples.
- You may use any NLTK corpora for the input.
- Use 11pt font size and default margin/line spacing for your report.
- Do not use a cover page.