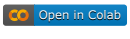
Homework 5: WordNet

INFO 159 / 259, Spring 2023

Due: Tuesday March 21, 2023 (11:59 PM)

In this homework, you will be working with WordNet synsets and exploring methods to align new words (not in WordNet) with an existing synset.

`https://github.com/dbamman/nlp23/blob/main/HW5/HW5.ipynb`

To use it, as usual, hit the "Open in Colab" button 

at the top, and be sure to save a copy in your Google Drive or Github *before* you start to work on the file.

1 Overview

WordNet is a great resource, but one of its downsides is *coverage*—many of the words in our vocabulary aren't in WordNet, but could conceivably be placed within existing synsets within it. Your task for this homework is to develop two methods to finding the closest synset for a given new word from Urban Dictionary.

2 Deliverables

There are three programming deliverables in this homework; no writeups. You'll find sample code and sanity checks in the notebook.

Please note: Do not change variable names or function signatures provided.
Put your answers between `# BEGIN SOLUTION` and `# END SOLUTION`.

Deliverabe 1. Find the WordNet synset with the highest cosine similarity between a) the average GloVe embeddings of its synset definition and b) the average GloVe embeddings of the new word definition.

Deliverable 2. Find the WordNet synset with the highest cosine similarity between a) the *sentence embedding* of its synset definition and b) the sentence embedding of the new word definition.

Deliverable 3. Implement a function `accuracy()` that assesses the quality of the dictionaries you return from those two methods.

3 How to submit

Submit your work to "HW5 (ipynb)" on Gradescope. Here are some recommended steps:

- Restart your Colab runtime and re-run all your cells; make sure everything runs as expected.
- Remember all your work **MUST BE PLACED** between `# BEGIN SOLUTION` and `# END SOLUTION`.
- Download your Colab notebook as an `.ipynb` file (File → Download `.ipynb`)
- Upload `HW5.ipynb` to Gradescope.