

Computational Semantics 2019-2020

Week 4 assignment

This week you will have an assignment that will help you with the second graded assignment.

Exercise 1. Check the script `inclass_word2vec_tsne.py`, and understand what it does.

Exercise 2.

A) **Linguistics students:** Do as instructed in `phrase_simA.py`.

B) **DTIC students:** Create a Python script, named `phrase_simB.py`, that does the following, with data from `data/word_pairs_pruned.txt`:

- create synthetic phrase vectors by averaging and normalizing the respective word vectors
- plot the word vectors and the phrase vectors in a single t-sne graph. Try using different colors for word and phrase vectors.

Analyze the results:

- what kinds of expressions are rightly close together or far apart?
- what kinds of similar expressions appear far apart?
- what kinds of dissimilar expressions appear close?