

## Assignment 3

Due by: midnight, 5th April, 2019

Q1. Download the word similarity-353 dataset from the web. I've put a link to where you can get it on the course home page. Now

(i) write code to calculate the Normalized Google Distance between any two words. Be careful you don't query Google too frequently while developing this, or your IP may be blocked. Use headers responsibly for your HTTP GET requests. (40 points)

(ii) reproduce the graph I showed in class plotting human similarity judgments for the 353 word pairs in the dataset against similarity judgments extracted using NGD. (20 points)

(iii) Use the word2vec-api service I have linked to on the course homepage to run a similar test for all 353 pairs using any of the pretrained models given on the webpage. (40 points)