Exercise 6: Word Vectors

You can earn up to 10 points on this exercise.
You may work as a group of up to 3 people, but please submit your own version.
You may use any programming language you wish, but any submission that we cannot run on our computers without installing things must be presented to the class.

Please email your solution to `claytong@coli.uni-saarland.de` *or submit before the tutorial by* **10:14:59 AM GMT+1, December 9, 2015**.

Task 1

Download and install the `word2vec` toolkit here:
`https://code.google.com/p/word2vec/`.

(a) Execute the `demo-word.sh` script. What files does it generate? What kind of corpus is this? What effect do you expect this to have on your vectors? (2 points)

(b) What are the ten most similar words to "monkey"? Write a few sentences analyzing the output. (2 points)

(c) What are the ten most similar words to your first name? Write a few sentences analyzing the output. (2 points)

Task 2

Explain the main differences between CBOW and Skip-gram. You can refer to sections 3.0 – 3.2 of the following paper: `http://arxiv.org/pdf/1301.3781.pdf`. (4 points)