Home Assignment 1: Introduction Into Natural Language Processing (5 points total)

Aim of this assignment is to give you a chance to practice some basics of Natural Language Processing with bare hands. Your assignment consists of two tasks: essay and practical part.

Essay (3 points)
Your task is to read the paper "Efficient Estimation of Word Representations in Vector Space" [link]. In the essay you will have to explain in your own words and your own examples provide a brief explanation of the key aspects of the paper. Main objective is to provide a clean explanation of how the model works (minimum 2 pages).

Practice (2 points)
As the practical part of the assignment you will have to practice similarity search task with your bare hands.

Part 1: Create the data. Your dataset should consist of four sentences:

"I really like {put your favorite dish}, it reminds me of {your favorite city}. ",
"{your least favorite dish} is ok, hard to forget {your hometown}. Wish to forget.",
"Sometimes I visit {some event}, reminds of how hard to get things done",
"This course is lame. Some {put your favorite dish} write this {your least favorite dish} lectures. Sometimes I'm thinking about {your hometown}.";

Part 2: Tokenize your sentences by words.

Part 3: Create The Bag of Words based on your sentences.

Part 4: Calculate cosine similarity between any of those two sentences.

Be aware:
- Your calculations have to be case sensitive,
- You have to do all calculations on the paper,
- All vector multiplications has to be done step-by-step,
- Do not forget to have the Bag of Words write down;

Submission:
- Deadline: June 10, 23:59,
- Submit single .pdf file;