This article provides the reader with some outputs and facts observed through the LSA/SVD assignment.

Through this assignment, a combination of functions from Linux and Matlab have been deployed to reach the final goal. The text, which is in polish, was refined using tr command in Linux. The following commands were used in order to have a text without any punctuation where space is the only delimiter. Using the command: tr -d [:punct:] All the English punctuation marks have been removed. Besides, as the text is in Polish, there are some punctuations which are not in English, and can not be removed using the above command. In order to dispose of those punctuation the tr command must be used to delete those specific characters.

The input to the Matlab script (which is attached to this report) is "refinedWord.txt" file, where there is no punctuation in the text. Using the commands in Matlab are words are read and stored in a vector called data (for further details it is recommended to refer to the appropriate Matlab script).

Finally the procedure is broken down to two main functions: mCalculator and plotting. The first function calculates the M matrix as proposed in the assignment. In other words, using two for loops, all the entries for vector M is calculated and stored. Thus each entry in vector M, represents the frequency of occurrence of bi-grams in all words of the text. It is noteworthy that, in this context the term bi-gram refers to a combination of two symbols in the alphabet. The final step will be conducted using the plotting function, where the projection and svd calculations are performed and the final results are depicted.

As the final observation about this project, by looking carefully at the figure available in the attachment, there are lots of point clustered together. Generally, as mentioned in the class, the two points which are close, illustrate the resemblance in semantic. In this context, we are dealing with letters, thus it seems reasonable enough, to consider this distance measure as the resemblance between letters. I believe that the clustered letters in the figure, are those who may appear in the text together.