**Community Questions Answering - Question Comment Similarity with the use of Unsupervised ranking model**

4 study points project work on Text-Based Information Retrieval

by,

Marimuthu Ananthavelu (r0652832) Monika Filipčiková – (r0683254)

**Summary**

This project work demonstrates the use of the Unsupervised model for ranking the comments with respect to a Question in the Community Questions Answering forums. The model which is used in this task is a Vector Space Model. The text pre-processing of the questions and comments were done with the help of available libraries for stop words, punctuations marks. The relevancy between each question and all the comments were referred as independent Vectors and their similarities were calculated with the help of Cosine scoring. The evaluation of model is done on the development set and compared with the baseline results and they are discussed. The output prediction file is generated for the ‘test\_input.xml’ is submitted for evaluation.

Introduction to the Project task:

Community Questions Answering forums are helpful to know the information a user wants to know. The comments in this task are the different answers for a specific question which is presented the in the file with the .xml format. The objective of this task is to rank the very relevant answers to a question as higher and rank less to the least relevant answers.