Project Proposal CS582

Piervincenzo Ventrella 664665713

Option: Search Engine.

Smart Component: Query Expansion.

Abstract: I would like to investigate the applicability of Word Embeddings in the context of Information Retrieval. The purpose of the project is to study the performance of the System when the query is expanded with the most similar terms according to Word2Vec algorithm.
Word Embedding makes use of a Neural Network that needs to be trained so it could be interesting to compare also the performances that will be obtained when the training is performed globally (using given datasets that Google itself provides i.e. from Wikipedia) and locally (using as dataset the sentences extracted from the UIC domain pages that will be crawled).