**Query Expansion via thesaurus**

Adel Rahimi

95201936

mehr.sharif.edu/~adel.rahimi

adelr.ir

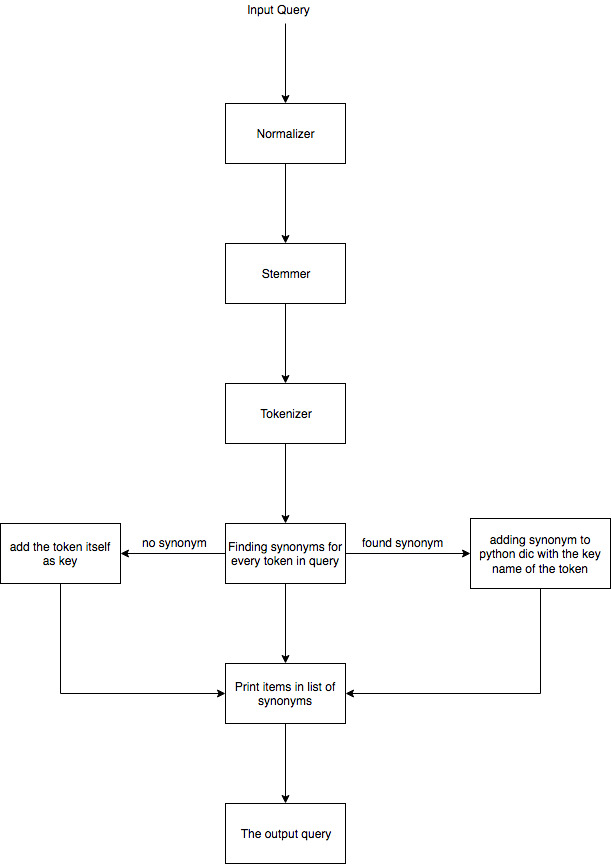
# Introduction:

Query expansion in a big part in information retrieval. It’s a big part of query understanding. Query understanding is done by inferring what does the user want from the query so that most related documents will be retrieved. In this project, I used several packages and several libraries for query expansion. NLP packages like Hazm.

# Methods:

Query expansion uses several processes from Natural Language Processing. In most cases when implementing query understanding systems, a simple or sometimes very complex spell checker is used. In this project; however, our first step in this project is normalization. In normalization process we use Hazm library normalizer. The output undergoes a stemming algorithm which is partly from Hazm algorithm used with some modifications. Some suffixes were removed from the algorithm. Then a simple tokenizer is applied to the output so that tokens will be divided. The most important part of the research is the lexicon. I extracted the vocabularies and their synonyms from the book: فرهنگ جامع واژگان مترادف و متضاد زبان فارسي written by: فرجاالله خداپرستي

After preprocessing of the texts from the book, and cleaning the data I dumped the words and their synonyms into a python pickle file. Python pickle is a module for serializing and de-serializing python objects. With pickle python objects will be converted into byte streams. Then I wrote a module to load this pickle file and get the input query. It will then process the input query and find the synonyms and show the nearest queries in the sense of meaning.



# Results

The results will be tested via Elasticsearch database and requests will be made through python requests.