

# Information Retrieval Project: Search Engine

Yasin Boloorch

July 14, 2020

## 1 Introduction

The first exercise in information retrieval project is to make a search engine for the specific dataset that been given to us (further we will deal with it). for start we will install it's requirement that is persian stemmer library for stemm persian words. after that we will start indexing all of the words in the data and finally make an invert index out of it. next we make a query reader program that will read the invert index and user queries, and return the best result that it can.

## 2 Dataset

This exercise dataset is a set of about 7800 poem from different poet. my goal was to read all of it line by line and save each poem file in a list. after my program read all of the poem files I made a list of all poems so that i can access it more easily.

## 3 Implementation

### 3.1 Programming Language

The language that exercise recommended for doing this project was to use python programming or java. I myself used python so i can read files and work with them more easily and ofcourse I'm more comfortable with it.

### 3.2 Libraries

As I said in the introduction i used PersianStemmer wich is a python library for setemming persian words. I used it in two placees, one time in creating invert index program and another time for setemming queries words. the other library that I use is json library for writing python dictionary object in a file and also reading json from file as well.

### 3.3 Indexing

Indexing process was just like last exercise but with some modifications that required for this task.

### **3.4 Queries**

Users can enter their queries in the graphical interface and program will catch it. after it get the query, it will split it with spaces, then remove the stopwords and look for the words in the invert index. at the end with some policy that I used, it will finally return the result and show it to user in the graphical interface

### **3.5 Graphical User Interface**

What i used for GUI is pyqt, wich is a opensource framework for creating GUI with Qt and get the designed interface in python codes. at the end I merged the querie program functions with GUI codes and it's ready to use.