Image Search with Elasticsearch

Burak Enes BEYGO
Master's Degree Student
Student Number: N19134235
Computer Engineering Department
Hacettepe University
Ankara, Turkey
n19134235@cs.hacettepe.edu.tr

***Abstract*—This document represents the project proposal for CMP681 Information Retrieval class.**

I. Introduction

My track choise for the class project is software track. I am planning to create a software by using the basic information retrieval techniques. I believe that creating a software project like this will help my aim to understand the information retrieval techniques. I will create this project as an individual.

Information retrieval techniques are mostly focused on text data. But there are works and needs for other kinds of data such as image, video and audio. I am planning to create a software which will be simply an image search engine through folders.

II. Project Details

A. Function

I am planning to use Java as the programming language. I will use landscape pictures as my data set. First of all, I will give a folder of images to the software. The software will save the images in the format of Elasticsearch. Later, I will give one random image and the software will tell me if that image exists in the set of previously saved images.

B. Why is This a Standalone Tool

The software will be a standalone tool. I decided to create a standalone tool because I believe the performance will be higher than a customized tool. A contribution might need extra tools or extra communication steps. Also, the software can be a remote server which responds to the requests as a Web service if I create an extra layer between systems.

C. Users of the Software

System admins might use this software to authenticate the system users. If a person's picture is in the system, this software may check and authorize the user. Also, collectors can use this software. People can save the pictures of collectibles in the Elasticsearch system and check if the next random collectible is in their collection.

D. Related Work

There are works about categorizing images by their contexts. There are also works about performing queries on image sets to gain information. I will do a more detailed research about the topic. The challenge I will take is to make the queries give results in a very short time by using helpful tools and algorithms.

E. Technologies

The most important tool I will be using is Elasticsearch. Elasticsearch is a search engine developed by REST principles. I will use Elasticsearch library for Java. I will also use image libraries for Java in order to convert images to vectors.

F. Usefulness of the Tool

For example, this software might be used for authentication in a server. A user uploads his image to the server. In this case, this image will be his or her key. Anytime the user wants to log in to the system, this software checks the image if it exists in the system. By that way, system users will know that each user is authorized to be a member in the domain.

G. Timeline

First of all, I will install Elasticsearch to my computer. I will also install Java development kits to compile. I am planning to learn how to use Elasticsearch with Java until the April 12, 2020. To convert images to vectors and save it in Elasticsearch, I will need to design my data structures and create an algorithm. I will do this until 17 May, 2020. I will prepare a progress presentation until 25 May, 2020. Finally I will write a document explaining the project with all the details until 10 June, 2020.