CS483 Project Proposal

Team Googolplex

Bergstrom, Kevin kevin.bergstrom@wsu.edu
Singh, Shanpreet shanpreet.singh@wsu.edu
Nollette, Joe joseph.nollette@wsu.edu

1 Introduction

Our project is a database for finding landmarks. We identify a landmark as an object or feature that is easily recognizable and lets people determine their location. We are going to use buildings, bridges, and mountains as landmarks. The user will be able to search for landmarks in a certain country or continent. The database will return a ranked list of results based on the user's query. The returned results will contain a picture, description, and location of the landmark. It will show on a map where the landmark is using its longitude and latitude coordinates. The user will be able to select which country or continent they wish to search in and define a search range in miles. The user will be able to narrow their search to just bridges, mountains, or buildings with an option bar. The search will use BM25 [1] to rank documents.

2 Database Summary

In our database each tuple will hold 6 attributes:

- ID: Integer
- Name: String
- Location: String (Latitude, Longitude)
- Description: String
- Image: String (Link)
- Type: String

Schema:

Landmark

ID	Name	Location	Description	Image	Type

3 Indexing

The database will search in the names, descriptions, and locations of the tuples. Images will be a byproduct of the search. We are not including images in the index because they are not comprised of easily searchable text. Only the descriptions will be stopped and stemmed for the search. The stopped and stemmed data will be stored separately than the untouched descriptions of the tuples. The user will see the original descriptions and will not see the stopped and stemmed ones. This makes it easier to read as stopping and stemming mangles the text into something devoid of grammar.

4 Features

4.1 Images

Each tuple will have an image associated with it. The image will show up in the returned result when the user searches. If there is no image associated with a given tuple, then a default image will appear instead.

4.2 Locations

Each tuple has a location attribute. This attribute will include the longitude and latitude of the landmark. Each returned result will have an associated marker on the world map which will represent its location.

5 Member Responsibilities

We have outlined 6 tasks:

- Web Scraping: Shanpreet
- Indexing: Shanpreet
- Ranking: Kevin
- Website Frontend: Kevin
- Website Backend: Joe
- User Interface: Joe

References

[1] Robertson, Stephen, and Hugo Zaragoza. "The probabilistic relevance framework: BM25 and beyond." *Foundations and Trends® in Information Retrieva*l 3, no. 4 (2009): 333-389.