

Group 6 - Phase 1

Kevin Dombrosky
Rochester Institute of Technology Student
610 Park Point Dr
Rochester, New York 14623
kfd6490@rit.edu

Brittany Purcell
Rochester Institute of Technology Student
107 Weldon Street
Rochester, New York 14611
blp6903@rit.edu

ABSTRACT

This paper discusses the progression of the group project for Group 6 in the Intro to Big Data class at Rochester Institute of Technology, CSCI-620. The paper will further explain the work that has been done so far, what will be done in the future, and what the expectations are for the finished project.

Keywords

Phase 1, Group 6, Plants

1. INTRODUCTION

The group project for CSCI-620, Intro to Big Data at Rochester Institute of Technology has been underway for 7 weeks now. It has been about a month since the initial submission of Phase-0 which laid out the data set and plan for the project.

Of the two requirements for the project, which include a database management system (DBMS), and having some form of database analytic, most of the work so far has been focused on the database analytic. During the last phase the dataset was chosen as well as a baseline for a visualization. As a reminder, Group 6 planned on creating an application that shows the geographical location of various types of plants in the United States and Canada using the Plants dataset from the UCI Machine Learning Repository. The database system is intended to be the relational database, MySQL. The analytics component is intended to be a visualization of what states and provinces have each plant within them.

There has been steady progression toward this goal, and so far there have been no changes to the visualization design, or the dataset.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...\$10.00.

2. BODY

2.1 Database

As a reminder, the dataset will be set up in a MySQL database, and will be inserted using a script of some sort. The dataset being used is the Plants dataset from the UCI Machine Learning Repository (<http://archive.ics.uci.edu/ml/datasets/Plants>). The information found in the database consists of a plant and its location in the United States/Canada. Below is the Entity-relationship(ER) diagram which was developed by Group 6. This diagram is the basis for how the information from the dataset will be stored in the database.

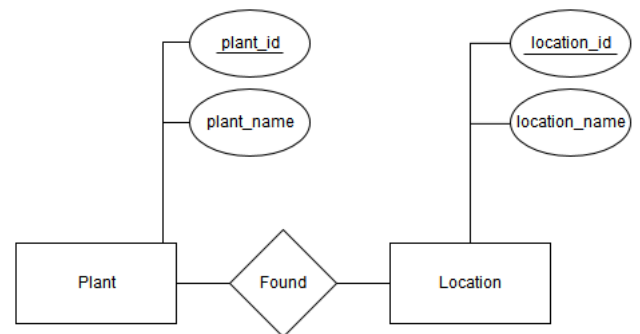


Figure 1: The ER diagram which represents the organization of the data in the database.

This is how the information will be stored within the database, and later when information needs to be accessed this will help making the proper calls to the database. There will be three tables within the database. The first, holding the information on the plants. This will consist of an ID for the plants, as well as the name of the plant associated with the ID. The locations will be very similar. The location chart will have a location ID, as well as the name of the location. Lastly, there will be a chart which compares the two items. This chart will map a plant-id to a location-id. That way, when a user is searching through the visualization the database will know what plants are in which location.

2.2 Application

Group 6 decided to create a visualization for this class, which was discussed in Phase-0. As a recap, the visualization will consist of a map(functional for only the United States and Canada), as well as a search box for the user to search for a certain plant. This has been created, and the screen shot of the application is in Figure 2 below.

This application is a web page that will be written in Jade Markup Language with CSS. This is the language that is most familiar to the group, after one of the members worked with the language all summer on an independent project. The basic blueprint for the web page has already been created (again, see Figure 2), but it has not yet been hooked up to the database.

Ultimately, there will be (on the left hand side of the website right below the search bar) a list of plants that are found in the database. The user can search for a specific plant (using the search bar) which will highlight the regions that plant is located. Otherwise, the user can choose a plant from the list, and it will do the same thing. Another option for users would be to select a region they are interested in, and the list on the left would change to be only plants that can be found in that area.

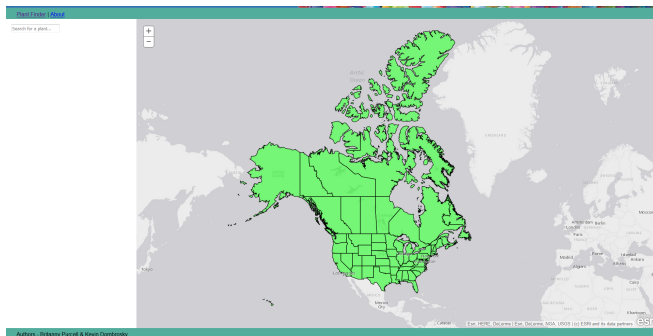


Figure 2: A screen shot of the website being used for the application.

2.3 Communication

In order for the website and database to communicate, the database will be setup on a Node.js server. This will allow the website to request information from the database in order for the user interface (UI) to change based on users requests. This is how the information will be transmitted from the database to the website.

2.4 Resources

There are many different resources being used in order to develop the project. These include certain files, softwares, and online databases all assisting in the creation of the website and database. As it was already covered, the database is being set up as a MySQL database, the application will be a website visualization that is coded using Jade and CSS, and the database and website will communicate with a Node.js server.

The code is being stored using GitHub, and is shared among all members of the group. This is a repository to store, download, and change the existing code. The code is being written in different platforms, some members are using Notepad++ to do the coding, while others are using Visual Studio Code. Once the code is written in each of these mediums, the code is added to the GitHub repository.

LaTeX is being used to write the ACM style templates, which is the document you are currently reading, and Visual Studio Code is used for the README (which will walk users through how to run the program and access the website).

The README will be provided to establish step-by-step

instructions on downloading and installing necessary software, and (once the software is installed), how to run the project and access the website.

3. CONCLUSIONS

The group has already made some progress towards finishing the project. This includes work on both the application as well as the database. There is still much work that needs to be done, but most of the thinking and setting up has been finished. Languages and applications have been chosen and are in the process of being set up. Below is a quick list of what was done, what needs to be done, and deliverables that have been set.

Accomplished:

- Developed ER Diagram
- Created Website Wire Frame
- Setup Website
- Conceptualized Application Utilization

To Be Done:

- Write Python Script for Database Input
- Set Up Database
- Establish Communication Between Website and Database
- Display Information on Website
- Change Display based on User Input

Resources

- GitHub
- LaTeX
- Node.js
- Notepad++
- Visual Studio Code

Files

- README
- Group6-Phase1
- CSCI620-master (Folder)