<div align="center">

Development of Citizen Science Builder Software
Ankita Kalkar, ECE department

</div>

Project Information

<u>Name:</u> Ankita Kalkar

<u>Research Title:</u> Software Architect Intern

<u>Researcher:</u> Dr. Pamela Gay, CosmoQuest

<u>Start date:</u> 6/15/2020

<u>End date:</u> TBD (continuing through fall semester possibly)

Internship Role

The intended purpose of this internship research project was to analyze and store NASA images in a more efficient and organized manner. In the past, many researchers after taking images of various celestial bodies, would spend hours analyzing each of those images individually to discover any unseen features or make any interesting topological observations. For example, a picture of volcanoes on Venus could take researchers several days to mark up and annotate on before analyzing the specific features that are contained in the image. CSB (Citizen Science Builder) software intends to alleviate this issue by outsourcing this task to volunteers of all different scientific backgrounds so researchers can focus on utilizing their expertise in their field to analyze the images and not spend time configuring them. There are several advantages for the researchers that include the image annotation process being speedy and shifting their focus to analyzing the images. For the volunteers, they are given the opportunity to participate in an interesting project that directly aids scientists everywhere in researching terrestrial bodies and to learn more about how scientists categorize images taken from space. Overall, this software

intends to bridge the gap between the public and researchers to create a way to collaborate on this interesting project.

Because this project started relatively recently, starting earlier this year in January, there is significant process to be made in structuring the software itself and creating the website so that it is accessible to everyone. One of our tasks as described by Dr. Gay was to figure out how to configure a feature on the website so that files can be imported along with other data associated with that file such as when the file was first inputted, the file description (what it contains), and other such information about the uploaded file. The format Priya, another intern working on the project, and I chose for this description of files was CSV (comma separated values) which is when each line in a data record or subject field is separated by commas. After researching other file types and their respective applications, we decided that CSV file types are typically easier than other file types to export and manipulate which is one of the intended capabilities of the website. After the file type was decided, this feature needed to be connected to a database so that all the images and the corresponding image information would be uploaded into a database that scientists would then have access to. The incorporation of a database allows for image versioning as well, so multiple volunteers could annotate the same image and different version of that image could be stored in a similar branch in the database. We are currently working on converting the CSV file types into an object that could be stored in a database and have spent approximately 15 hours so far on this task.

One of our other tasks involved manipulating the images by chopping them up before storing them in the database to be distributed to researchers. There are several advantages to chopping up the uploaded images. One of them include that the images can be processed by volunteers one feature at a time so that it is not overwhelming for volunteers to analyze the

images and provide annotations. Furthermore, the zoomed in section of the image might make it easier for volunteers to see the topographical features and annotate on it. Breaking up the larger images also makes it easier for them to be stored inside the database as the file size is smaller and thus the pictures will render at a better quality. To still preserve the image's content the picture's content will be cut up using an X-Y coordinate grid so the smaller images can be pieced back into the larger image based on specific coordinates. Since the satellite images are primarily two dimensional, it will be easy to implement this idea and overlay each image with a coordinate grid. To keep track of the smaller images that represent the same larger image, the smaller chopped up images will be tagged with the larger image's ID as well as that image's CSV file containing all its information.

Conceptualizing how we were to approach this task and come up with a strategy to how we were to approach this task ultimately took a large portion of our time. We encountered several issues for this seemingly straightforward task. For example, we had to account for and maintain clean edges on the chopped-up image pieces when they were sewn back together to make sure they didn't overlap. Ultimately this was just a rounding error, that was easily fixed. Another issue we encountered was if the image was abnormally shaped and one of the smaller squares only contained a small section of the original image and the rest of the square was simply blank. After researching some options and consulting how other researchers have approached this issue, we planned to either provide an option to chop up larger images into customizable shapes or to let the users determine how they wanted to approach chopping up the images to possibly allow for non-square chopped-up images. This task brought the number of hours spent on the project so far to approximately 21 hours.

Experience Description

  This experience was overall truly interesting as it was my first time using the knowledge I studied in my classes and apply that to a real-life application. I not only learned about new technologies, but I learned how I should approach problems I do not have a lot of background in. The learning curve was relatively steep, but I had the guidance of Priya, who has more experience working on the development aspect of websites than me, as well as Dr. Gay, who has been driving this project in addition to her other commitments. Throughout the entirety of this project, I gained two new mentors that explained a lot of the development process and the typical approach most coders take to approaching a project such as this. I am very fortunate to be under their guidance especially with this being my first application of my experience to a real-life project, it was truly beneficial to have people to guide me along the way and answer any questions.

  While the summer term for CMU is coming to an end, I found myself wanting to continue this project as I truly find working on this application interesting and is an excellent way to apply the skills I've been practicing in my classes into action while contributing to an amazing project that attempts to involve the public in a unique way to aid researchers. Because of various other commitments Priya and I had this summer, we were not able to put as much time as we wanted to regarding this project. However, I found the tasks we could accomplish truly rewarding and an interesting way for me to contribute to the project by utilizing my talents and skillsets.

  In this project, the setting up of the web server and configuring the system in our respective computers initially proved to be a roadblock, but Dr. Gay met with us several times to

help up set up the environment in our computers and patiently guided us throughout the whole

process.

Knowledge Gained

As a rising sophomore at CMU, the extent of my programming experiences has been

limited to the learning the programming languages Python and C, with not as much focus on the

web development aspect or working with databases. I learned a lot about how websites are

structured behind the scenes and I learned a lot about configuring databases and web servers.

Specifically, I learned about LAMP stacks which is software that combines Linux OS, Apache,

MySQL Database, and PHP environment. This software is commonly used to create websites

and other such web applications such as the Citizen Science Builder Software. Through seeing

the open-source code for this software, I learned a lot about the structure of how websites are

structured. For example, Apache provides a server that translates web browsers to the

corresponding website. After configuring the Apache environment, I set up the MySQL database

that can be configured by scripting to set up the website. The last layer was the PHP scripting

layer which consisted of PHP scripting language that connects the other parts of the LAMP

system. This layer is also where the main website or web application is run. Through working on

this project, I became more aware of how all these components are connected and how important

the concept of loose coupling is. Creating an environment where components are loosely

connected allows for greater flexibility and greater fault-tolerance as well because if one of the

components fails, then the others will not. Overall, working on this project gave me insight into

how my skills could be incorporated into a large-scale project such as developing Citizen

Science Builder software.

Timeline

Date and Time	Description
6/15/2020 – 6/27/2020	Attempting to configure the software through LAMP stack and configure the docker container to upload the code we work on to.
6/15/2020, 5:00 – 7:00 PM	First meeting with Dr. Gay and introduction to the project.
6/20/2020, 6:00 – 7:30 PM	Met with Priya to discuss how to install the web server and configure the docker.
6/23/2020, 6:00 – 8:00 PM	Met with Priya to continue trying to debug the configuration of the software download.
6/27/2020, 6:00 – 9:30 PM	Second meeting with Dr. Gay where she guided Priya and I through the entire installation process and explained the various components of the software.
6/27/2020 – 7/09/2020	Figure out how the website was structured and set up the web interface with specific tabs to contain the features.
7/04/2020, 6:00 – 8:00 PM	Met with Priya to set up the web interface for the install page.
7/07/2020, 5:45 – 7:00 PM	Met with Priya to conceptualize how to approach cutting up the images.
7/09/2020, 5:30 – 8:00 PM	Third meeting with Dr. Gay to get specific tasks Priya and I would work on.
7/09/2020 – 7/14/2020	Decided to use CSV file format, set up feature where files could be inputted into a database.

7/14/2020, 5:30 – 8:30 PM	Fourth meeting with Dr. Gay where we were given the additional task to chop up images.
7/14/2020 - present	Working on the stated tasks and keeping in touch with Priya to talk about how we should approach them.