Global Hepatitis Outbreak and Surveillance Technology (GHOST)

Project History

Team 23:

Andrew Amontree
Jeongsoo Kim
Ernest Lai
Sarthak Mohapatra
Lovissa Winyoto
Georgia Institute of Technology

Final: April 25 2016

In partnership with:

David S. Campo, Ph.D.
Centers for Disease Control and Prevention (CDC)

Revisions History

Version	Primary Author(s)	Description of Version	Date Completed
0.1	Jeongsoo Kim	Introduction and content table are implemented	04/13/2016
0.2	Ernest Lai	Lessons Learned section is implemented	04/15/2016
0.3	Lovissa Winyoto	Project Data is implemented	04/16/2016
0.4	Andrew Amontree	Syntactical changes and style formating	4/18/2016
1.0	Sarthak Mohapatra & Lovissa Winyoto	Final edits and formatting	4/25/2016

Contents

Introduction

Hepatitis C is a contagious liver disease that is hard to detect and spreads rapidly. According to the Centers for Disease Control and Prevention (CDC), almost 3 million individuals in the US may be afflicted with chronic Hepatitis C. In response, the Division of Viral Hepatitis at the CDC has developed Global Hepatitis Outbreak and Surveillance Technology (GHOST), a system for the detection of Hepatitis outbreaks in the United States. The system is being used internally at the CDC as well as by affiliated State Departments of Health (SDH), who may submit “jobs” involving genomic analysis for the CDC’s high-performance computing (HPC) cluster to process. After the HPC cluster sequences the genomes, it returns a dataset to the user containing information on the relationship between potential patients, based on the similarity of those genomes.

GHOST accomplishes two goals: first, it is the front end for a database of previously sequenced patient genomes. Similarities between the genomes of two patients, A and B, imply that patient A may have infected patient B, vice versa, or that the two patients may be linked through a third patient, C. Second, once a researcher navigates to a specific genomic analysis, GHOST allows visualization of the data in the form of a node-link diagram of patient-to-patient relationships, where nodes are patients and links represent the strength of their relationship. This interactive visualization can be utilized by researchers and analysts to find important indicators in the data sets in order to better understand the outbreak.

1. Definitions and Acronyms

- **Hepatitis C:** An infection caused by a virus that attacks the liver and leads to inflammation
- **CDC:** The Centers of Disease Control
- **GHOST:** Global Hepatitis Outbreak Surveillance Technology, the system that is built to report Hepatitis C outbreak and analyze its spread across the United States
- **UI:** User Interface

2. References

- Statement of Work
- Feasibility Report
- Detailed Design Document
- Usability Study Report
- Source Code

Historical Overview

1. Phase 1

During the first planning phase, we formed our team for the project and were assigned to project GHOST. The main focus during this sprint was to establish contact with our client and set up meeting times throughout the semester. We were introduced to Dr. David Campo the main researcher from the CDC for project GHOST. In this phase we had several meetings with our client in order to pin out the main issues with the website that needed redesigning. During Phase 1, we worked on the Statement of Work and related documents for our project.

2. Phase 2

Throughout Phase 2, we narrowed down the scope of our project into three main areas: the landing page, the dashboard, and the node-link visualization. We also produced the Feasibility Report for Project GHOST where we decided to use d3.js for the node-link visualization due in part to its previous implementation by the CDC. In Phase 2, we started submitting our legal documents to the CDC to gain access to the their physical workspace and project source code.

3. Phase 3

In our last phase, we presented the designs for our solution to the client. After narrowing down the main issues, we explained to the client that the landing page needed to be altered from a technical database view to a more intuitive process flow that minimized the complexity for unexperienced users. We also suggested an upgrade to the dashboard that involved consolidating the various disparate parts into one seamless view. Finally, we recommended that the node-link visualization needed to be refactored to include valuable features that would help health practitioners gain more insight from the embedded visualizations. The client generally agreed with our approach and helped us by adding a few additional recommendations to be refactored into the design and we began prepping for the implementation phase.

4. Sprint 1

For our first sprint, we planned to implement the landing page for the GHOST website. The goal was to make the landing page intuitive enough for new users to familiarize themselves with the GHOST system and its functionality without needing to learn the complex data table style of the original. However, we had trouble getting access to the current GHOST code base that we were tasked with working on due part to immigration status issues. For that reason, this sprint was focused on back-and-forth discussion with the client to get access to the source code and its required dependencies.

5. Sprint 2

Throughout Sprint 2, we planned to implement the node-link visualization for the submitted raw file from GHOST user that has been processed into JSON files. The goal was to make the visualization more interactive and to give users more options on what they can do with the visualization to help them gain more information about the current Hepatitis C outbreak. However, after further communication with the client, we were told that we would not be able to gain access to the source code for several more months due to authorization issues with the CDC firewall. The solution proposed by the client was to have team members who were US citizens travel to the CDC Emory campus in order to develop the code base. As this would be difficult to do due to scheduling and travel constraints, we decided to compromise. At this point we decided to limit the scope of the project to focus solely on the node-link visualization in Sprint 3 as this was the only portion of the original project that the client team was able to provide to us. After communicating this revision with our professor and project coordinator, the client delivered us a large (> 14000 Lines of Code) HTML file that included a sandboxed version of their basic d3 implementation of the node-link visualization among a plethora of Javascript and CSS dependencies.

6. Sprint 3

During our final sprint, we planned to modify the dashboard of the GHOST website. The goal was to design the dashboard to look like a multi-step pipeline so that users could easily follow along the process of analyzing their submitted data. However, due to the aforementioned changes to the scope of our project, we worked on the node-link visualization instead. We developed the code delivered to us based upon on notes taken from project planning discussions with the client in the previous semester. Furthermore, we developed test cases and provided documentation related to our changes.

Project Data

1. Organizational Structure

The project client was Dr. David Campo from the Centers for Disease Control and Prevention, helped by Dr. Brian Gurbaxani, the Georgia Tech and the Centers for Disease Control and Prevention's liaison.
Project Manager - Robin Tracy, Sarthak Mohapatra
Documenter - Lovissa Winyoto
Point of Contact - Ernest Lai
UI Developer - Andrew Amontree, Jeongsoo Kim

2. Project Metrics

Actual Scheduling Documentation:

1. ***Status Reports***: Please refer to our status report lists from both the first and second semester.
2. ***Backlogs***: Please refer to our planning documentation.
3. ***Burndown Chart***:

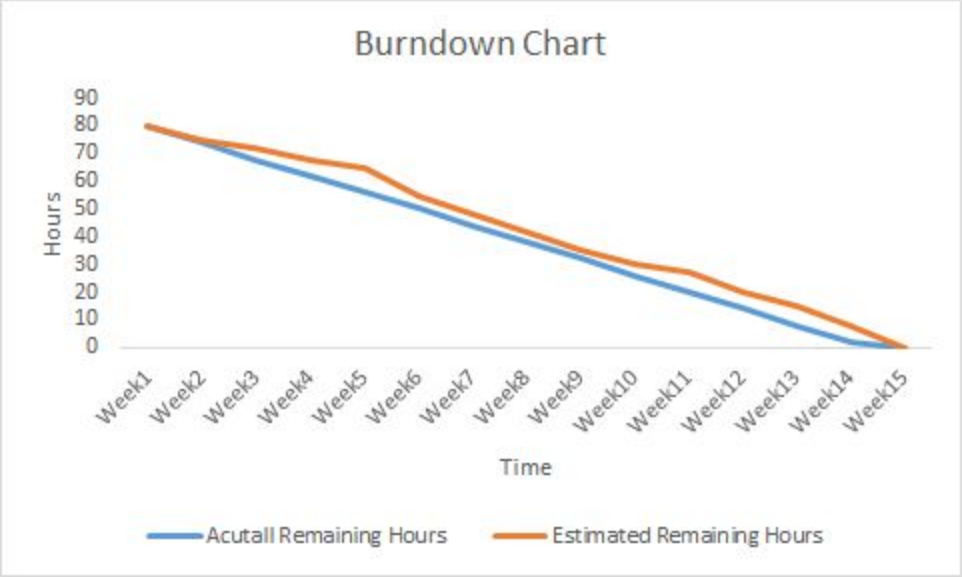

Figure 1. Burndown Chart.

Design Metrics:

1. User Study Tables

Participant #	Task #1	Task #2	Task #3	Task #4	Task #5	Task #6	Task #7	Task #8	Task #9
1	N/A	Yes	Yes	Yes	Yes	Yes	Yes	N/A	Yes
2	N/A	Yes	Yes	No	Yes	Yes	Yes	N/A	Yes
3	N/A	Yes	Yes	Yes	Yes	Yes	Yes	N/A	Yes

Table 1. Task Completion Breakdown.

Participant #	Task #1	Task #2	Task #3	Task #4	Task #5	Task #6	Task #7	Task #8	Task #9
1	N/A	0:03	0:10	0:07	0:08	0:08	0:04	N/A	0:06
2	N/A	0:15	0:05	0:08	0:10	0:10	0:08	N/A	0:17
3	N/A	0:14	0:09	0:06	0:06	0:02	0:05	N/A	0:12

Table 2. Task Time Breakdown.

Task #	Completion Rate	Average Time for Successes
1	N/A	N/A
2	100%	0:11
3	100%	0:08
4	66%	0:07
5	100%	0:08
6	100%	0:07
7	100%	0:06
8	N/A	N/A
9	100%	0:11

Table 3. Task Completion Rate and Time Summary.

Implementation Metrics:

1. Lines of Code: 14403 lines.

Testing Metrics:

For each feature that is implemented, the feature will be tested on its correctness and its usability. Testing its correctness to make sure that the feature does what it is supposed to do. Testing its usability by asking users to use the product and analyze whether the user interface is clear and intuitive enough for the users (Please refer to the Usability Study Report).

Lessons Learned

The past academic year has yielded a lot of experience and knowledge for our team. In the process of working as team and working with other teams, we learned how to create a concrete deliverable that exceeds the standards of all parties involved. When working with an organization as large as the CDC, there are a multitude of lessons to be learned; primarily revolving around clear communication and scheduling.

While the channel of communication was always kept open and emails were replied to as soon as was possible, a reply may have taken up to several days to be sent. Establishing a means of communication that was faster than email should have been one of the priorities at the beginning of the semester to allow for cross-team communication.

A bigger lesson learned from this semester was one of timing, or in a more general sense, time management. There were a fair number of people involved in this project, and scheduling meetings with all Team Casper members was a task in itself. Setting aside a block of time for meetings with the developers from the CDC should have been one of the points addressed when setting up initial contact. However, this was realized too late when a meeting had to be set up and the differing availabilities made scheduling significantly harder. Communication and time conflicts aside, learning to deal with setbacks in the project and having to downsize the scope of the project was an important lesson learned that will be extremely relevant to our future careers as scope creep is an extremely relevant issue in consulting practices.