Statement of Work

Introduction

Hepatitis C is a contagious liver disease that is hard to detect, but spreads rapidly. According to the Center for Disease Control and Prevention (CDC), almost 3 million individuals in the US may be afflicted with chronic Hepatitis C. 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 sequencing for the CDC's high-performance computing cluster to process.

GHOST accomplishes two things: first, it is the front end for a database of sequenced patient genomes. Similarities between the genomes of two patients, A and B, imply a possible connection in their disease. Second, once a researcher navigates to a specific genomic analysis, they can visualize a node-link diagram patient-to-patient relationships.

Problem Statement

GHOST enables public health researchers, who generally are not specialized bioinformaticians, to glean actionable insights from scientific data. However, feature and interface improvements are needed. For GHOST, a less intuitive, less easy-to-use system implies fewer tangible returns from health research. Users are unable to utilize the substantial computational power from the CDC's system either for lack of knowledge or because the options for computation job submission are not easily accessible.

The system also is inaccessible to the users without the technical background to understand the role of genomics in Hepatitis transmission, which creates a problem when users are overwhelmed with data they are unable to understand or make inferences from.

The specific problems with GHOST's user interface that our project will address are divided into the following two major categories:

<u>User interface issues</u>

1. The landing page does not allow the users to visualize the structure of the system/environment
2. The landing page does not display a broad array of options that allows job submissions by the user to CDC's high-performance computing environment, causing friction in the uploading process
3. The landing page does not allow users to personalize contents to suit their project's needs/workflow

<u>Data visualization issues</u>

1. The data visualization, in the form of a node-link diagram, does not encode enough data (geographic, demographic, or temporal) of the patients and their history causing poor functionality
2. The data visualization does not allow users to select/highlight patients and get details on demand on the selection, slowing users down by forcing them to read large volume of plain text

3. The data visualization does not have any option to reduce data through filtering or aggregation, causing time-consuming process of submitting new jobs
4. The nodes in the data visualization does not convey useful information based on their shape, color, size, or orientation

Objectives

User interface redesign is important for GHOST because a more user-friendly “sink” for the data pipeline - from molecular sequencing to processing to database querying to analytics, all the way to actionable insights, can move the needle for patient outcomes by enabling policymakers in high-risk areas where Hepatitis C is prevalent.

Functional Requirements

1. The landing page should provide a text introduction to first-time users, introducing them to the system
2. The landing page should be able to navigate to all the options for computational job processing for users
3. The landing page should allow users to manipulate user views, and remember the views when the user logs in again
4. The data visualization should encode all available geographic distribution of candidate genomes, demographic information on the candidate, and temporal data on the date of detection, where “encoding” refers to communicating the data through some channel (shape, color, size, or orientation of nodes) on the visualization
5. The nodes need to provide more detailed information on specific nodes
6. The web interface has to provide a feature to filter out nodes by characteristics and or by a user’s selection of a subset of nodes

Nonfunctional Requirements

1. The code for the web interface must follow the open-closed principle, where it needs to be open for extension but closed for modifications.
2. The code for the interface must be documented for all methods with top-level visibility
3. The web interface should take no more than 3 clicks from the main screen to navigate to any other scree
4. The web interface must be usable by multiple users by performing tasks concurrently with no decrease in job turnaround time from the original GHOST system.

Applicable Standards

Since CDC is a federal agency under the Department of Health and Human services, regulations for federal public websites are the only applicable standards, which affect our project. We will therefore be complying to Section 508 and Accessibility standards that ensure that all government owned websites or documents have to be accessible to people with disabilities.

References

http://www.cdc.gov/amd/pdf/factsheets/amd-projects-ghost.pdf
http://www.cdc.gov/hepatitis/hcv/cfaq.htm#cFAQ21