Product Vision
Programming Life: Group 3
Thunder Wolf Squadron

Contents

1. Introduction

Our product is a genomics visualizer. It will be able to parse in data about a collection of genes given by the user, and display that information in a comprehensive and intuitive manner. To achieve this, our visualizer will use a few key features; A ribbon visualization of the genes at the upper level view to improve how the user is able to understand the bigger picture of the data[1]; A semantic zooming functionality so the user is able to dive deep into the data and look at specifics without getting lost; A graph visualization of the data at the lower levels of the semantic zoom so the user is able to look at specific mutations and sequences of the genes[1]; An internal database functionality that allows for quick and efficient data storage and retrieval; And lastly, a visualization of the corresponding phylogenetic tree[2].

2. The Client

The main customers for this product are going to be genetic researchers. Specifically those from the Broad Institute and The KwaZulu Natal Research Institute as they are the ones we will be corresponding with throughout production for feedback and general impressions of the product. Ideally, any genetic researcher should be able to find use in our product. Meaning that genetic researchers as a whole can be considered our target audience as well as our potential customers who will be buying our product.

2.1 Client Needs

The need that this product will address is the need that genetic researchers have for an intuitive and easy to use visualizer for genome data. There were several implementations that were specifically requested by the customers from Broad Institute and The KwaZulu Natal Research. These include:

- Interactivity with the visualization
- A semantic zooming functionality

- Phylogenetic tree
- Ability to see specific mutations and annotations within the data

These are the 'must haves' for the product, but we envision even more. Due to our use of an internal database, we are able to query the data very efficiently. This means that implementing components such as query tools, meta-data analysis, and data integration is also in our sights. But, as previously stated, the 'must haves' for our product are the features that were specifically requested by the customer and will therefore be placed at a higher priority than all other aspects of the product.

3. Existing Products

When we compare our product with those that are already on the market, such as the Integrative Genomics Viewer[3] that Broad institute already uses, ours is easier to use and understand at a glance. Though it is worth mentioning that even with our envisioned analytical functionality, IGV offers much more in terms of in depth analysis of genome data. Our program is easier to use and understand because from the outset we assumed that our customers may not be very experienced with computers and/or complicated user interfaces. Meaning that, without a simple and easy to understand user interface, it is possible for them to become lost in the application, making it immediately lose its usefulness as a product.

This differs from what is currently on the market because, from what we have seen from visualization programs such as GenomeSpace[4] and GENE-E[5], standard user interfaces are very complex and have a pretty steep learning curve associated with them. Apart from an intuitive user interface, our product is unique when compared to both those in and out of our company due to the fact that we use a ribbon visualization of the data at the higher levels that transitions into a graph representation when zoomed into. This means that, through smart obstruction of data, we are able to dynamically show information to the user so that they are not overwhelmed but are still able to accurately derive and understand the meaning of the data.

4. Timeframe, Budget, and Launch

4.1 Timeframe

The product will take approximately 10 weeks to create. Using SCRUM we will divide up the work into sprints lasting one week each. At the end of each sprint we will present our working version to the customers (For much of the time this role will be played by Thomas Abeel, though researchers from Broad and KwaZulu will be able to give their feedback as well). In addition to these meetings, we will be having weekly sprint meetings to decide on what our goals are for the coming week. During the same meeting we will also be discussing the previous week, any problems we encountered, and possible improvements that we are wanting to make for the following sprint.

4.2 Budget

The budget for the project will be measured mainly in hours worked. This is because we are still students and there is a large variation in the amount developers are paid. Meaning that an exact monetary amount has a lot of variance, though an estimate will be given.

   Each student is expected to work a minimum of 28 hours in a week. Across the 5 students in the group we will be working 140 hours in a week and around 1400 hours throughout the entirety of the 10 week process.

   With regards to the monetary cost of the project. If one were to use the median hourly rate of a software developer in the Netherlands, the total cost of the project would be €28,300 with an hourly rate of €17.00.

4.3 Launch

The launch of the product is rather normal. Because the product is designed to be used anywhere with minimal setup or installation, the client only needs to download the .zip file and run the executable after extracting the files. It is even possible for the user to keep the program on a flash drive and use it on any computer they wish, online or offline.

5. References

[1] Schroeder, M. P., Gonzales-Perez, A., & Lopez-Bigas, N. (2013). Visualizing Multidimensional Cancer Genomics Data. *Genome Medicine*. Retrieved from http://www.biomedcentral.com/content/pdf/gm413.pdf

[2] Hudson, D. H., Richter, D. C., Rausch, C., Dezulian, T., Franz, M., & Rupp, R. (2007). Dendroscope: An interactive viewer for large phylogenetic trees. *Genome Medicine*. Retrieved from http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-8-460

[3] Robinson, J. T., Thorvaldsdóttir, H., Winckler, V., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative Genomics Viewer. *Nature*. Retrieved from http://www.nature.com/nbt/journal/v29/n1/abs/nbt.1754.html

[4] Qu, K., Garamszegi, S., Wu, F., Thorvaldsottir, H., Liefeld, T., Ocana, M., . . . Mesirov, J. P. (2016). Integrative genomic analysis by interoperation of bioinformatics tools in GenomeSpace. *Nature Methods*. Retrieved from http://www.nature.com/nmeth/journal/v13/n3/full/nmeth.3732.html

[5] Gould, J. (n.d.). Retrieved from http://www.broadinstitute.org/cancer/software/GENE-E/index.html