Product Vision Context Project
Desoxyribonucleïnezuur

Toine Hartman - tjbhartman - 4305655
Martijn van Meerten - mvanmeerten - 4387902
Iwan Hoogenboom - ihoogenboom - 4396634
Yannick Haveman - yhaveman - 4299078
Ivo Wilms - iwilms - 4488466

11 May 2017

Abstract

This document describes the vision of our product. It does so by analyzing the customer's needs and directly infers the product attributes from these needs. Furthermore, a comparison to existing products in this area is made. The key differences and similarities with existing products are highlighted. Finally, the timeframe and budget of the project are explained.

Table of Contents

1. Introduction 2
2. The Customer 2
3. Customer Needs 2
4. Product Attributes 3
5. Comparison to Previous Work 3
6. Project Organization 4
7. References 5

1. Introduction

This document is the Product Vision of the scrum team Desoxyribonucleïnezuur, further referred to as “we”, “us”, “our”. for the Programming Life context. The product vision acts as a means to align stakeholders and customers, and serves as a guideline to the scrum team. (Layton, n.d.).

First, we will list the customers of the project and identify the characteristics these groups have. Secondly, we will identify the needs of the customers and how our product solves these needs. Next, the crucial product features are listed and explained. Then we will compare our product to other existing products, to clarify where our product is unique. Finally, the timeframe in which the product has to be delivered is stated.

2. The Customer

The customer for this product is GenomeViz Inc. GenomeViz Inc. is doing research on differences in DNA sequences. To do this, they aim to visualize the differences between genomes with our product. Through these interactive visualizations they hope to gain a better understanding of the causes of differences between the phenotypes of organisms (Linthorst, 2017).

3. Customer Needs

Due to recent advances in DNA sequencing, there has been a massive drop in the cost of sequencing (The National Human Genome Research Institute, 2016). According to Linthorst (2017a, April 25), this has enabled scientists to carry out whole-genome sequencing for multiple members of the same species, which enables them to see the variance within that species. However, because these developments are so recent, there is a lack of programs for visualizing the differences in these dataset, and the few that do exist lack in certain areas (we will elaborate upon this in chapter 5). Our product aims to fill this gap by visualizing differences between genomes in an intuitive way, allowing users to select nodes and edges and view information about them. The genomes will be represented as a graph. The nodes will represent segments of similar dna and the edges will represent the differences between genomes. Furthermore, quality of life features, such as bookmarks and history functionality will be present.

4. Product Attributes

To satisfy the aforementioned needs, certain key attributes have to be present in the product. These attributes define when the product is viable, commonly known as a “minimum viable product”.

Visualization

According to the presentation about this project by Abeel (2017), our product needs to support visualization of genomes. The overall genome visualization will be in the form of a graph. This is done by representing similar segments of DNA in different genomes as a rectangle, which grows with the size of the segment. Differences in genomes should be visualized as an edge going from one rectangle node to multiple rectangle nodes depending on how many genomes differ from one another. The color of nodes should also carry some information. We have initially decided to visually encode the ratio of AT and GC pairs in a node with colors. If a segment contains too many unknown base pairs to be considered useful, the segment will be colored black. However, we may change this and use another source for deciding the color if we find something that would be more useful.

Interactivity

According to the presentation about this project by Abeel (2017), while visualizing the genomes, the product should allow the user to pan over the graph. When selecting a segment, information on that segment should appear in a console. The product should allow zooming and adapt properly to the amount of information shown. “Adapting properly” to zooming is a requirement, but as of now it is not defined how that should be done. Abeel (2017) suggested consolidating nodes into glyphs (symbols/nodes that don’t represent an actual node, but a set of multiple nodes), but this is not a hard requirement.

Data Scalability

The product should be able to load graphs containing at least 2 million nodes (Abeel, 2017b), and 6 million edges. (T. Abeel, personal communication, 2017-05-02). The product should be able to visualize subgraphs with 20,000 link traversals depth at a time.

5. Comparison With Previous Work

Since the ability to sequence entire genomes is quite new, there are few products available that are comparable with our product. The product that compares best with our product is Bandage (Wick, Schultz, Zobel, & Holt, 2015). Bandage has many of the same visualization aspects and many of the features. It does however differ in a few areas. The following observations were made by us by trying the software on several datasets.

First of all, it does not seem to be possible to see the sequence of a node, except by copying it to the clipboard and pasting it to some text editor. It does show the length of the sequence, just as our product will.

Furthermore, it does not seem to handle big files very well (human can be loaded, but drawing with a radius of 500 nodes completely freezes it). Loading big files is a must have for our product.

Last, our most distinctive feature will be the bookmarks and genome highlighting. Nodes in bandage can be labeled, but they can only be labeled individually. This means that you cannot include nodes in multiple bookmarks in Bandage. Our product allows the user to visualize genome paths and compare the mutations in the sequences.

6. Project Organization

Timeframe

The product will be developed over the course of 10 weeks. Weekly working versions will be delivered according to the Scrum methodology (Rubin, 2013). The final product will be delivered on friday, the 23rd of June 2017.

Budget

The product will be developed free of charge by a group of students studying computer science at Delft University of Technology.

7. References

Abeel, T. (2017a). *Briefing 1* [pdf]. retrieved from https://slack-files.com/T3GEK9ABD-F56RWSPLH-e846211fbb

Abeel, T. [Thomas Abeel]. (2017b, May 10th). Max base pairs per node? Sounds fair, why does this matter? Fail gracefully with informative information when you're hitting hard limits. [Slack message, reply to question in previous message]. https://tud-contextproject-17.slack.com/archives/C53R91Q95/p1494403996555214

Layton, M. C. (n.d.). *FOUR STEPS TO DEFINING YOUR PRODUCT VISION WITH AGILE MANAGEMENT*. Retrieved from http://www.dummies.com/careers/project-management/four-steps-to-defining-your-product-vision-with-agile-management/

Linthorst, J. (2017a, April 25). *Seminar Context project: Programming Life*. presentation.

Linthorst, J. (2017b, April 25). *Seminar Context project: Programming Life* [pdf]. Retrieved from https://blackboard.tudelft.nl/bbcswebdav/pid-2939373-dt-content-rid-10263164_2/courses/40221-161704/Seminar%20contextproject%202017.pdf

The National Human Genome Research Institute (2016). *The Cost of Sequencing a Human Genome*. Retrieved from https://www.genome.gov/sequencingcosts/

Rubin, K. S. (2013). *Essential Scrum: a practical guide to the most popular agile process*. Ann Arbor, Michigan: Pearson.

Wick R.R., Schultz M.B., Zobel J. & Holt K.E. (2015). *Bandage: interactive visualisation of de novo genome assemblies*. Bioinformatics, 31(20), 3350-3352.