Contextproject Programming Life
Product Vision DRAFT

TU Delft

Ruben Bes - rbes - 4227492
Mathijs Hoogland - mhhoogland - 4237676
Jasper Denkers - jdenkers - 4212584
Robbert van Staveren - rhvanstaveren - 1527118
Willem Jan Glerum - wglerum - 4141040

May 8, 2014

Abstract

This is the product vision for the Programming Live Contextproject, a second year course from Computer Science at TU Delft. We will develop a tool for genetic biologist to determine the diseases of a person by analyzing DNA. This tool will visualize the mutations, trio data and gene interaction. The product will be intuitive to use and be able to handle existing data to compare the results to. The goal is to produce easier to use software then at hand and do this in a short fixed timeframe.

Contents

1 Who is going to buy the product? Who is the target customer?

The product will be developed for free, for the use of genetic biologists. These scientists use data from DNA to determine which diseases a person has, and how the mutations in their DNA affect other genes as well.

2 Which customer needs will the product address?

Our application is going to help our customers in visualizing DNA sequence variants and in linking SNP's with diseases. This is currently difficult because the human genome is extremely large and a lot is stil not known, by accepting raw data and making it understandable we will help our customer in understanding what effect certain SNP's have. We are also going to help by visualizing trio data and gene interactions.

3 Which product attributes are crucial to satisfy the selected needs, and therefore to the success of the product?

The product attributes that are crucial to satisfy the customer needs are an intuitive GUI for the user to use the application, an ability for the user to input data and a visual output of the results. There aren't any constraints on the so called non functional requirements; the applications could be a web service or a standalone desktop program. A file format for the used data is already available: VCF (Variant Call Format). The application should handle these files.

Furthermore, there should be existing information available (a reference sequence) so the application can compare input data and find annotations and report variants. Readed input data should be mapped to the reference data to detect variants like SNPs, deletions, insertions, etc. To distinguish true variants from sequencing errors, some statistics are needed to end up with a solid result. These data on mutations/variants is useful for the users.

One of the main topic users would like to treat is which mutations cause which kind of diseases. The applications could help the users by using one of the five given methods: looking for known 'disease mutations', use family data of the persons whose data is used as input, filtering, looking at nearby SNPs or lastly, looking at nearby genes. The power of the application could lie in expressing the results of these methods in a nice visual way.

In some of these methods it's needed to perform some calculations. The application should be able to execute these calculations fast. It would be possible to use a server-client model for the application where at client side, there's an interface for the user and where computations take place at server side.

4 How does the product compare against existing products, both from competitors and the same company? What are the products unique selling points?

There are no products to compare with from the same company, though several other product for DNA visualization have been made. Existing products like cBioPortal [1] and other software can either show relations between genes, or visualize specific genomic variants. For example Circos [2] visualizes data in a circular layout, useful for exploring relationships between objects.

Our product will focus on genomic variants at different locations. The product will visualize and assist in the interpretation of genomic variants on different genomes. The focus will be on trio data: genes of parents and children, and gene interactions: genes on different locations of the genome affecting the same attribute.

5 What is the target timeframe and budget to develop and launch the product?

The timeframe for this project is seven weeks, with a fixed deadline on the 26th of June. As this is a schoolproject we should finish the product before the deadline as we cannot continue the project after this date. Furthermore there is not a real budget for this project, again because this is a schoolproject, however each team member will receive 10 ECTS for this course.

References

[1] Jianjiong Gao, B. Arman Aksoy, et al. cbioportal for cancer genomics. `http://www.cbioportal.org/public-portal/index.do`, 2014.

[2] Martin I Krzywinski, Jacqueline E Schein, Inanc Birol, Joseph Connors, Randy Gascoyne, Doug Horsman, Steven J Jones, and Marco A Marra. Circos: An information aesthetic for comparative genomics. `http://genome.cshlp.org/content/early/2009/06/15/gr.092759.109.abstract`, 2009.