Projects Requirements Specification

Mert DOĞAN
Tuğba Nur KORKMAZ
Sevcan DOĞRAMACI
Süleyman Emre ÇELİK
Mehmet Gürsel ARSLAN

Table of Contents

1. Introduction **3**

1.1 Description **3**

1.1.1 Project's Purpose 3

1.1.2 Project's Scope 3

1.1.3 Access to Software 3

1.1.4 Target Audience 3

1.1.5 Project's Mission 3

1.2 Constraints **4**

1.2.1 Hardware Constraints 4

1.2.2 Communications Constraints 4

1.2.3 Software Constraints 4

1.2.4 Design Constraints 4

1.2.5 Time Constraints 4

1.2.6 Privacy Constraints 4

1.2.7 Operational Constraints 4

1.3 Professional and Ethical Issues **5**

2. Requirements **6**

2.1 Functional Requirements 6

2.1.1 Variant Calling 6

2.1.2 Variant Annotation 7

2.2 Non-Functional Requirements 8

2.2.1 Variant Calling 8

2.2.2 Variant Annotation 8

3. References **9**

1. Introduction

1.1 Description

1.1.1 Project's Purpose

Variant calling, annotation and visualization are the top three tasks in the bioinformatics area. However, the researchers utilize different tools to perform each. Therefore, they need to deal with each tool separately and reach the final conclusion by giving the result of one to the other. By this system, users can do variant calling, annotation and visualization in one program enabling them to spend less time and effort.

1.1.2 Project's Scope

System consists of three separate modules : variant calling, variant annotation and variant visualization. It is responsible for performing variant calling, variant annotation by getting required files from the users and visualizing the results. Each module can run individually or all can be executed together. System can not be used to explore genetic variation of genomes other than human genomes.

The system will be available as a command line tool . The users can also use it by the system's web interface to execute their actions in a clear user interface.

1.1.3 Access to Software

The system will operate on its users' local computers. It is available on GitHub. Necessary installation and usage guidelines are provided on the repository page.

1.1.4 Target Audience

The system is intended to be used by researchers who are interested in exploring genetic variation for human diseases and population genetics of humans.

1.1.5 Project's Mission

The objective of this project is to create a tool which combines variant calling, variant annotation and variant visualization to make the researcher’s jobs easier.

1.2 Constraints

1.2.1 Hardware Constraints

a. There should be enough space on the user's computer memory to download the system.
b. To save alignment, calling and annotation files into the user's computer, the user should have enough memory space.

1.2.2 Communications Constraints

c. Data fetching time depends on the user's internet connection quality.
d. Data fetching time depends on the servers' response time.
e. Data shall be fetched via REST APIs or FTP.

1.2.3 Software Constraints

a. System must run on Linux.
b. Python, Node.js and npm must be installed on the operating system.
c. Internet connection must exist to fetch data from data sources.

1.2.4 Design Constraints

a. Front end software shall be created by using JavaScript and ReactJS.
b. Back end software shall be created by using Python and Flask.
c. System shall be designed to work from the command line and web interface.
d. System shall use SQlite to store data.
e. System shall allow users to switch between modules (variant calling, annotation, visualization).
f. System shall be open source.
g. System shall be stored in GitHub.

1.2.5 Time Constraints

a. System shall be scheduled to be completed at the end of May 2021.

1.2.6 Privacy Constraints

a. No data files can be shared with a person or a third-party company.

1.2.7 Operational Constraints

a. System's performance depends on the server capacity.

b. The inputs of the algorithm received over GitHub must be compatible with the user's input.

1.3 Professional and Ethical Issues

The project should have some restrictions and regulations to protect the social environment and privacy of users, in order for them to benefit from it with mind at peace .There are professional and ethical issues to be considered for this.

Especially, since the project is a study in which human DNA information is processed and includes the field of health, it must work very accurately and consistently. The probability of error must be exceptionally low.

In addition, since DNA information is private information of humans, the issue of privacy is of great importance. Biological weapons specific to individuals or races can be produced with DNA information today or users may have concerns regarding privacy. These situations increase the importance of privacy even more.

For all these reasons, the project has certain restrictions and requirements to be met. These are listed below:

a. Users' name, age, gender, nationality, location and other personal information cannot be shared and used for other purposes.
b. The files uploaded by the users to the system and the files resulting from the processing of these files cannot be shared and used for other purposes.
c. The program must give accurate results so that users are not misinformed.
d. Protecting files is the user's responsibility.
e. The application will be developed to support researchers in the field of genetics and to facilitate their work.

2. Requirements

2.1 Functional Requirements

1. The system shall allow the users to go to the desired step and perform the process, if the user has the input in the appropriate format.
2. The system shall display information messages if there are any conditions or explanations.
3. The system shall allow the user to download the system program from GitHub repository.

2.1.1 Variant Calling

A. Alignment/Assembly

a. The system shall allow users to upload FASTQ or FASTA files from their computer.
b. The system shall allow users to upload at least two files one for reference and one for sample.
c. The system shall allow users to decide read length option for alignment.
d. The system shall allow users to select an algorithm for assembly.
e. The system shall allow users to select output format(s) for assembly.
f. The system shall allow users to save, view the output or send it into the next step.
g. The system shall display a success message if the file is loaded successfully.
h. The system shall display an error message if the loading of the file fails.

B. Calling

a. The system shall allow users to upload a BAM file from their computer.
b. The system shall allow users to fetch a BAM file from Alignment/Assembly step.
c. The system shall allow users to decide sequencing method.
d. The system shall allow users to decide calling method.
e. The system shall allow users to select multiple algorithms of calling method.
f. The system shall allow users to add a new algorithm from GitHub for calling.
g. The system shall show an output list of selected algorithms.
h. The system shall allow users to view each output.
i. The system shall allow users to select outputs from the list to save or send into the merge step.
j. The system shall display a success message if the file is loaded successfully.

k. The system shall display an error message if the loading of the file fails.

C. Merge SV

a. The system shall allow users to upload VCF files from their computer.
b. The system shall allow users to upload at least two files.
c. The system shall allow users to fetch VCF files from Calling stage.
d. The system shall allow users to decide the algorithm to merge.
e. The system shall allow users to save or view the output.
f. The system shall display a success message if the file is loaded successfully.
g. The system shall display an error message if the loading of the file fails.

2.1.2 Variant Annotation

A. Load input

a. System shall allow users to upload a VCF file from their computer.
b. System shall allow users to go to Variant Calling step to acquire a VCF file.
c. System shall display a success message if the file is loaded successfully.
d. System shall display an error message if the loading of the file fails.

B. Fetch Annotations from Annotation Sources

a. System shall use annotation sources named 1000 Genomes, dbSNP, dbVar, ClinVar, OMIM, Gencode, Ensembl, COSMIC, ENCODE, HPO to fetch annotations.
b. System shall connect to the internet to make annotation requests from the annotation sources.
c. System shall send GET requests to RESTful services of the sources.
d. System shall show loading while it is fetching the annotation data.
e. System shall show an error message if it cannot connect to the internet.
f. System shall show an error message if the sources' services return error responses.
g. System shall show a success message if annotation fetching is performed successfully.

C. Show Variant Annotations

a. System shall allow users to save the results of annotation as a text file.
b. System shall show variant information and annotations for each variant in the VCF file in a table.
c. System shall allow users to go to the Visualization step to visualize the results of annotation.

2.2 Non-Functional Requirements

1. The system shall be available for 7/24 hours.
2. Downtime of the system should not exceed 10 seconds. ?
3. The system shall be controlled via command line or the web interface.
4. The web interface has to be user-friendly and easy to use.
5. The web interface pages' load time should not be more than one second. ?
6. The system shall be open source.
7. The system should allow downloading the files such as reference genomes or bam files from the database.
8. The system should have a database to keep variants.
9. The duration of the process in variant calling should be close to the sum of the specified average running times of the algorithms used in the process.
10. The duration of annotation fetching should not exceed 60 seconds.

3. References

1. https://web.cs.dal.ca/~hawkey/3130/srs_template-ieee.doc
2. http://web.cse.ohio-state.edu/~bair.41/616/ReqDoc_Outline.htm
3. https://winatalent.com/blog/2020/05/what-are-non-functional-requirements-types-and-examples/
4. https://www.srs.fs.usda.gov/pubs/VT_Publications/01t7.pdf
5. https://www.acm.org/code-of-ethics