

Andrew Poll
anp28

*Audio and Visual representation of DNA
sequence differences*

Amanda Claire
afc

G400
Computer Science
CS39440

05/02/2015
Version 1.0

Release

Project Description

The aim of the project I have selected is to compare two DNA sequences and display the results. The results must be in a format that the general public will be able to understand and help them easily identify how different or similar one strand is compared to the other. The types of media I will be using to convey this information will be in the form of images, audio and animations. The target audience for this project is the general public in a learning environment such as a school, college or university as such the device for this program will be a portable school laptop. The files sizes of information that I will be working with can range from a few hundred characters to thousands and meaning that the file could be potentially megabytes in size. This means that the program will have to be robust enough that it can handle a lot of processing.

I will be writing this program in Java or HTML5 and JavaScript. I have decided that I would write in one of these languages as they both have easy to use Graphical User Interface design software, for Java I have the use of Swing and AWT and JavaScript and HTML I can use CSS. Another reason for choosing one of these as my language is because they both have a massive support for libraries. I found multiple audio and visual effect libraries that could be very helpful in my project as well as a number of audio APIs. It is my assumption however that because of the large file sizes that I will be working with and the amount of processing of that information, that HTML and JavaScript won't be powerful enough.

The end goal for my project is to have created a software program that accepts two DNA sequences, aligns them together and give the user a choice to see the resulting difference as a static image, an animation or listen to a musical comparison.

After talking to another lecturer I found that another direction in which this project could go would be to use robot controlled instruments to play the music. Some of these instruments were an organ piano that we could add paper strips or polystyrene foam balls into the pipes to visually represent the changes in the music. Another instrument I could use would be a glockenspiel with an actuator that strikes the wooden notes. If I use the piano I can use multiple octaves giving a wider range of notes and more complex music whereas the glockenspiel is restricted to one octave.

Proposed Tasks

My first major task of this project would be to read in the DNA sequences from a text file and align the information that I receive from them. This means that when reading through the data in the files I am reading from the same point in the sequence. I have looked briefly at this and have decided that I will need to look into it in further detail, specifically algorithms that I can manipulate for this purpose or I have found some Java libraries that can also aid in the task, I will also look at these.

I will also be using Image manipulation to generate the static images and the animated images. I will use a library of Image processing methods and classes that I created for a previous assignment. This will allow me to draw shapes, add existing images and alter images more freely. There are also a few API's I can use for this task such as d3.js which I can use to manipulate document information, in this case the sequences to create charts, graphs and images as well as animating them.

Another major task that I will have to complete is creating music from the sequence. I have found numerous packages and libraries for both Java and Javascript that allow me to create and generate music and tones. I have found numerous application on the web that have given me ideas about

how I might approach this task from drum machines to synthesisers. I will have to read around and write an algorithm that takes the information from the sequences and creates a structured piece of noise that can be interpreted as music, this will be applied to each sequence and play then both played to the user.

I will also be reading about MIDI messages. These will be used - if I choose this route - to send music information to a Raspberry Pi that will be controlling robotic musical instruments. This information will be instructions such as play a note and then stop playing a note along with status information about the instrument.

Project Deliverables

At the end of this project I hope to be able to deliver a piece of software that solves the initial problem. This means that program should be able to compare DNA sequences and display the results with a good aesthetic, and that catches the eye of the general public. It should also be able to create two separate pieces of music and be able to play them whether through speakers or by sending the music data to a robotic instrument. This program should also be able to run efficiently on the required machines.

With the Program I hope to deliver a guide on how to use the program as well as a set of UML diagrams including Class and Use case as well as a number of tests including unit tests.

The progress reports will document the progress made to the date and will include changes current status expected completion date of some tasks as well as any issues that may have occurred. If I follow a FDD approach these progress reports will be the mile stones that are used to plan and track progress in this method.

The final document I will produce will be the final report detailing the problem and my initial solution with details about my research into the subject. I will include changes that I made to the initial idea and the development process going into detail about problems I may have faced, solutions to them and unexpected developments that I found easier or more difficult than initially imagined. I will then write about technical aspects of the project and write a conclusion.

Initial Annotated Bibliography

Adenot, P., Wilson, C., & Rogers, C. (2015, January 06). *Web Audio API*. Retrieved from webaudio: webaudio.github.io/web-audio-api/

This website hosts the information relating to HTML web audio api. I plan to use this if I write the program in HTML. It holds the information about the methods and how to use them.

Bostock, M. (2013). *Data-Driven Documents*. Retrieved from d3js: <http://d3js.org/>

This site is a provider of d3.js software. I also documents detailed instruction on how to use the software. The site also contains examples that other people have used the software for and I found as a great source for ideas.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., . . . Marra, M. A. (2009, June 18). *Circos: An information aesthetic for comparative genomics*. Retrieved from Genome Research: <http://genome.cshlp.org/content/19/9/1639.full>

This paper is based on a visualization tool that they created to display the variations in genomes. They talk about how there program Circos can create different types of graphs. I hope to find ideas and methods of how I could achieve similar results but oriented more towards the general public.

Kumar, B. R. (2012, April 20). *DNA Representation*. Retrieved from Intech: <http://www.intechopen.com/books/dna-sequencing-methods-and-applications/dna-representation>

This paper is about how we read DNA sequences and the methods and techniques used to extract the information from them. In the paper he also demonstrates the types of graphs each of the methods procures. I can use this as a base for my work on representing the DNA sequence by observing what is out on the market already.