Enhanced BLAT

Rushi Patel – ruship   Arash Khoshparvar - arashkh

Abstract — *Christine from BE755 needed an updated version of the BLAT software which would help reduce the amount of time she spends searching for 20 base pair sequences in a specific genome for use in her FISH probes. Our API driven version of BLAT was created as a starting method to improve on the current limitations of BLAT and also allow the application to be customized further for other unique uses.*

1. Introduction

Enhanced BLAT is a simple to use interface connecting the current UCSC BLAT search engine to an interface that can run multiple BLAT sequence jobs and store the results locally or display them in an easy to read and understand table. This project enhances BLAT towards performing batch queries without the need for constant user control. These queries can be independent of one another, or they can be related. An example of a related batch of queries is sequence expansion, which is query of a sequence encapsulated in a larger sequence, particularly its overlay in the genome. Since the encapsulation for the expansion is part of the selected genome, a first query is needed to attain the encapsulation itself. Using the Enhanced BLAT project user can throw in jobs and leave the desk while the time-consuming process of alignment is being performed. By running code on a powerful servers instead of the local machine, the user does not have to be worried about not being able to use the local machine.

2. Project Details

a. Java application

1. **Servlet-based Web Application**
   The outcome of this application is web interface that takes a number of BLAT jobs form an HTML form, parses and runs them, and finally shows the results in the browser. The input form has been designed to be as similar as possible to the original BLAT form, with only one extra input which is the expansion options. The added functionality includes the ability to run expansions (base pairs surrounding a query sequence), the ability to run unlimited number of query sequences, and a mix of these two. Results are shown in a table that can easily be copied to an excel sheet for further review.

2. **CommandLine JAR Application**
   The outcome of this application is a commandline interface that runs batches of BLAT jobs, saves results in file with the JSON format, and shows the results in the commandline. All the options that are possible in the web application are supported here as well. Everything is included in a single Java Archive file which makes it independent of any library installation. The final use case for this app is its integration in Node.js which is going to be integrated into OWL.

b. Node JS framework (See /node/run.js)

The Node JS file uses a simple exec call to run any java jar object that is passed through command line arguments. This makes the Node JS generic to run for any external java jar application. Once the application is finished a signal is sent to a node function which then looks for the JSON file that was created by the java application and stores the information locally to be used further. At the moment as proof that the JSON is found, I print out the file after being stored.

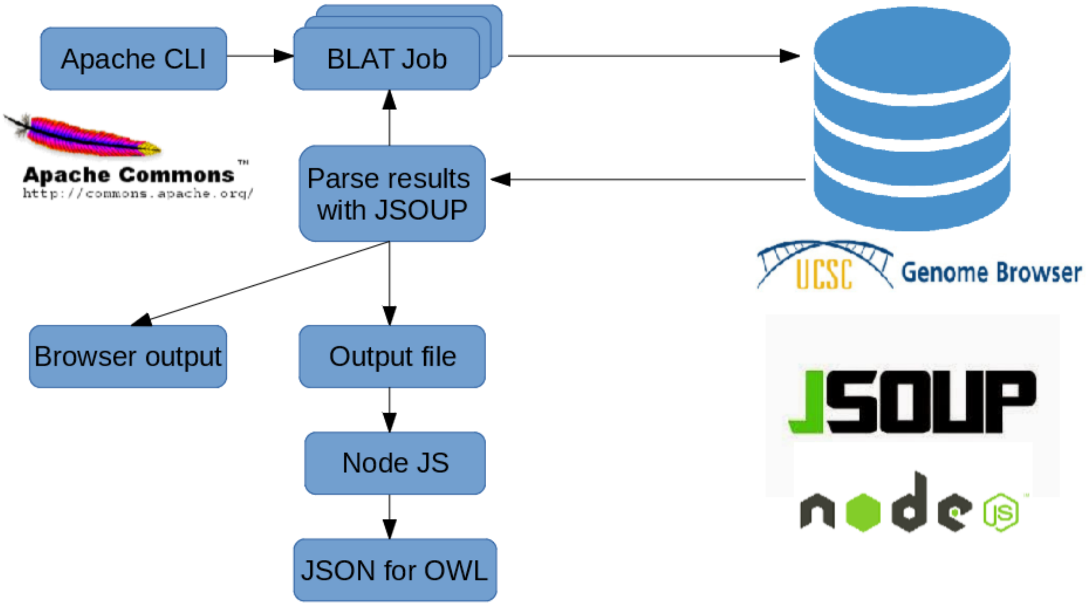

Figure 1: Shows the basic diagram of how the Enhanced BLAT is created.

The figure above shows how the BLAT system can take parameters using the Apache Commons CLI interface creates a number of BLAT jobs and runs these jobs through the UCSC Genome Browser. Once the data is returned a new set of BLAT jobs may be created. After all BLAT jobs are finally completed, the results are either displayed through the web app browser interface or can be stored as a JSON object through the help of a Node JS application layer.

4. Summary

The final product Enhanced BLAT was well received by Christine and she said it will help her with future BLAT work. The goals of the application to give Christine a better interface toward BLAT was successful and also helped us to learn more about synthetic biology and what types of applications are being created and used by those who are working in the genetic engineering fields today.