Dylan Diaz Moreno

Final Project Proposal: Blast Search Tool

I'm a Computer Engineer with a passion for artificial intelligence, who for the longest time, didn't know in what to apply to field of machine learning to. I always knew that learning artificial intelligence had to be combined to something to make the most use of this field. So, after watching a TedxTalk on youtube about the future of Bioinformatics, I got instantly hooked. First semester wasn't so good for me, because I was taking Molecular Biology plus other two prerequisites Biochemistry and Bio Organic Chemistry, without every taking a biology course in my life. Sure I struggled a lot but it helped me with the foundations of what this master's is really about.

I took Introduction to Bioinformatics and the first Practical Computer Concepts course and it was then when I started grasping the pattern of how this field rolled out. While taking Introduction to Bioinformatics, I always wondered how all these NCBI, Pubmed and Galaxy actually worked. We performed many NCBI Blast searches, but I wanted to see how everything was running in the back end. Then I stumbled upon something interesting, running a Blast search using only the ubuntu terminal. In that course, I switched over, and with the ubuntu server knowledge I gained from Practical Computer Concepts, I started performing BLAST searches on the terminal.

With this, BLAST is faster, it runs on my own database, there is automation with running it on the server, downstream analysis, and can be part of other programs. With the biggest pro being the speed. This is what I started being more passionate about, developing tools using the ubuntu server. The command line looks simple but intimidating at the same time, and I'm up for taking that challenge with my knowledge in the Computer Engineering field and my expanding horizons with Biology.

That is the foundation for this project, I attempted to recreate the NCBI Blast search from the web server, on my own, using the ubuntu server commands.

For that I installed NCBI-Blast+ on my own local machine and performed a blast search on a test sequence from a mouse. The command for installing it is:

```
sudo apt-get install ncbi-blast+
```

The test DNA sequence I used was:

>test sequence
AGGATTAAAAGTAAATCACCATGACCTACTACTGTTTGTCTTTTTCCTGTTGCTTTGT
TTTGTTTTGTGTTCTTCAGACGAGGTCTTAATCTACAGTTCTGGCTGGCCTAGAGAT
CAGTGTGTAGAACAGGCAGGTCTCATCGCTTAGATATCCACTCCTTTTGCCTCAGG
AATGATAGGATTTCAGGTATGTGTAACCACTCCCGACTTTTCATTTTGAAGTTTTGT
GGTTACTCAAGATCGCTTATCTGCCAGGTGGGTCTGATTAAGTTGACCAAGCTCTC
TAGCTATGAAAAATTTAGTTGTTGTTTTATGTGTCTCAGCCCTAGAGCTGGCTTACG
TGGAACTCTCAGTTTAGACCTGGCTCACCATGACCTCATACCCTAAGTGTATCCCC
CTGTTGCTTCAGCCTTCTGAATAGCTTGGTTCAGGGGAATCAGGAATGGGCTCTTG
TCTTTCTCCCTAGTCTGGGTTCACTCTGTCCTGAGTTCAATGAGGGGTGGAGCTCT
ACCAATTCAGTGAGGTGGCAGCCTGTGGTTGGCCAATTTGGCTGCTAAGGGAAGG
TGTGAAAGGTAGGGTGCATAAAAGGCTCAGAGAGTCAGAGAAGACCATAGGAATC
TTGGAGTTACCTCCTCGCCCTACTGCTCTAGGAGACCTGTCACCAGCGTCTGTGTC
TGGTAAGCCCCTCATCTGCATCAGGCATGACACCTGTTCTCCAGCAGGCCAGCCC
CTCCCCATTGGAGACATTTTAGCACAGATGACACAAGACTCTAATAAGTGCCGATT
AAAGGCGTGGTGGCGCACGCCTTTAATCCCAGCACTCGGGAGGCAGAGGCAGGT
GGATTTCTGAGTTCGAGGCCAGCCTGGTCTACAACGTGAAGGTCCAGGACAGCCA
GGGCTATACAGAGAAACCCTGTCTCGAAAAACCAAAAAAAAAAAAAAAAAAAACA
ACAAAAAAAAAAAAAACAAAAAAAAAACTCTAATAAGTGATTTTGG

After that, I incorporated it into the server with all the files needed, and created a web form using HTML, CSS and JavaScript to try and replicate the NCBI Blast search experience.

Difficulties came in actually performing the shell commands using a cgi script, which is something that I don't have much experience with, but It got resolved. In the end, I was able to parse the results from the BLAST search and incorporate them into the HTML page using JQuery successfully, and later styled the website in an unique manner using CSS (didn't want to make it look exactly like NCBI).

The tool can be run at the following link:

http://bfx3.aap.jhu.edu/ddiaz8/final/finalproject.html

It was an overall fun experience, which reiterated my liking even more into the field of Bioinformatics. With this I hope to accomplish more interesting tasks like creating gene prediction algorithms after taking the Applied Machine Learning course, and be on the path of graduating as a Bioinformatics Scientist.