

Genome Similarity using Frequency Vectors

Requirements:

In order to compile and run this assignment, you need either g++ version 4.8 or clang++ version 6.10 or superior. Furthermore, the *PThread* library must be correctly installed on your environment. The whole project was developed on an OSX 10.10 machine, but the compilation and execution should also work file on Linux machines as well.

Instructions to compile

To compile, use the command according to your compiler:

- (g++)
\$ g++ mods.cpp Bacteria.cpp -O3 -lpthread -o cvtree
- (clang++)
\$ clang++ mods.cpp Bacteria.cpp -O3 -lpthread -o cvtree

Instructions to run

To run the program, ensure that the supplied file `list.txt` and the folder containing the specific bacteria files `data/` are on the same folder as `cvtree` binary file (obtained on the previous step). Then just run the program passing as argument the list file. For example:

```
$ ./cvtree list.txt
```

The program will start running and as soon as it finishes the processing, the output is displayed on the screen. Failure on providing the list file as argument will cause a Segmentation Fault.

Input dataset

The dataset consists on the file `list.txt` and folder `data/`, which are the same files supplied to us with the original code.

Code versions

Three different code versions are provided. The following table contains details about them

| Version | Comments | Related Code |
|---------|--|---|
| Before | Initial supplied naïve version | <code>tidy.cpp</code> |
| Before | Initial supplied improved version | <code>improved.cpp</code> |
| After | The version produced during the assignment development | <code>mods.cpp</code> <code>Bacteria.cpp</code> <code>Bacteria.h</code> |

If you want to compile the “Before” versions, to compile you don’t need to append the `-lpthread` flag. The running instructions are the same.

The produced code was tagged along the development. To see the produced result after each step, checkout a tag (The list of tags can be obtained by using `git tag` then `git show <tag_name>`).

Known issue

If the computer used to run this code hasn't enough RAM, the OS may kill its process due to high usage of resources (Happened on a Ubuntu 15.04 with 1GB of RAM)