Author

Rui Diao
diaorui1987@gmail.com

Requirement

OS: Linux
Compiler: g++

Compilation

Run "make", or "g++ clonify.cc --std=c++0x -O2 -fopenmp -o clonify".
It generates a binary file "clonify".

Usage

./clonify <input_file> <output_file>

Required arguments

input_file: a JSON file
output_file: a text output of clustering result

Optional arguments

-n <N>: processed number of sequences from the input_file
-m: print out some memory usage information
-f <output_format>: specify the format of output

Examples

./clonify NBD4681.json test.out
./clonify NBD4681.json test.out -n 1000
./clonify NBD4681.json test.out -n 1000 -m
,/clonify NBD4681.json test.out -n 1000 -f seqs

Multithreading

The code is implemented with OpenMP.
Please make sure you put "-fopenmp" in compilation flags (already in Makefile).
The code uses all cpu cores for computation.

Distributed Computing

If multithreading is enough, please ignore this part.
Suppose we have one master computing node, and several slave computing nodes.
All computing nodes communicate with socket.
Run "make slave" or "g++ clonify.cc -DSLAVE --std=c++0x -O2 -fopenmp -o slave". It generates a binary file "slave".

Put the binary file "clonify" to the master computing node.
Put the binary file "slave" to all the slave computing nodes.
Run "./slave <#port>" on all the slave computing nodes. For example, "./slave 8888".
Create a text file "SLAVES" on master computing node, under the same directory with the binary file "clonify". "SLAVES" specifies the IP and #port of every slave computing nodes. One line for one slave computing node.
Do not list master computing node in "SLAVES".
Do not list one IP twice in "SLAVES", because we already have multithreading for every computing node.
See "SLAVES.sample" for reference.
Run "./clonify <input_file> <output_file>" on the master computing node.

Used Open Source Codes

JSON++
website: https://github.com/hjiang/jsonxx
license: https://github.com/hjiang/jsonxx/blob/master/LICENSE
fastcluster
website: http://danifold.net/fastcluster.html
license: http://cran.r-project.org/web/packages/fastcluster/LICENSE
SciPy
website: https://github.com/scipy/scipy
license: https://github.com/scipy/scipy/blob/master/LICENSE.txt