Lab 13 – Mapreduce

Presented by Arnon Shimoni

To run, you have several options (this is for timing purposes mostly)

All runs will output two result files - results.txt and probabilities.txt.

When no file name specified - assumes that the split up files (a-z) already exist from previous runs. This means you must first run `mapreduce:main([],"354984si.ngl").`, and only after that can you run the other styles without a filename specified.

1. Single node, no spawning with NO file-splitting (for benchmarking)

   run `mapreduce:onlyOne("filename.txt").`

2. Single node/multinode with NO file splitting (for checking different timing issues):

   run: `mapreduce:main().` or `mapreduce:main(['a@Computer','b@Computer',…]).`

   This DOESN'T create onlya.txt through onlyz.txt but assumes they exist!

3. Single node/multinode with file name - meaning a file will now be split up. This is useful for the first time

   run: `mapreduce:main([],"354984si.ngl").` or `mapreduce:main(['a@Computer','b@Computer',...],"354984si.ngl").`

   This also creates the onlya.txt through onlyz.txt required.

Example test runs:

```
1> timer:tc(readFile,onlyOne,["354984si.ngl"]).
{7379000,true}


(a@Jengapad)1> timer:tc(readFile,main,[['a@Jengapad']]).
{6238000,ok}


(a@Jengapad)2> timer:tc(readFile,main,[['a@Jengapad','b@Jengapad']]).
{5476000,ok}


(a@Jengapad)3>
timer:tc(readFile,main,[['a@Jengapad','b@Jengapad','c@Jengapad']]
{4715000,ok}
```

```
 (a@Jengapad)4>
timer:tc(readFile,main,[['a@Jengapad','b@Jengapad','c@Jengapad','d@Jengapad'
]]).
 {5434000,ok}


 (a@Jengapad)2>
timer:tc(readFile,main,[['a@Jengapad','b@raspbmc','c@Jengapad','d@Jengapad']
]).
 {16023000,ok}
```

Running style	Single node, single process	Single node	Two nodes	Three nodes	Four nodes	Four nodes (3 on a PC, one on a Raspberry Pi)
Time until results	7.379 seconds	6.238 seconds	5.476 seconds	4.715 seconds	5.434 seconds	16.023 seconds

The larger number we got from running on three nodes and the Raspberry has to do with the network message passing as well as the limited power available on the raspberry.