Answers - Lab2

Lucas FIDON (lucas.fidon@student.ecp.fr)

Project TF-IDF: exercise 5.1

Top20 words and their TF-IDF values were:

1. buck@callwild 0.006827235
2. dogs@callwild 0.002442956
3. thornton@callwild 0.0017668009
4. myself@defoe-robinson-103.txt 0.0016665459
5. spitz@callwild 0.0013086796
6. sled@callwild 0.0013086796
7. francois@callwild 0.0011342764
8. bucks@callwild 0.0010251999
9. friday@defoe-robinson-103.txt 0.0010216236
10. trail@callwild 8.943081E-4
11. john@callwild 8.724928E-4
12. perrault@callwild 8.070469E-4
13. hal@callwild 6.543398E-4
14. team@callwild 6.543398E-4
15. thoughts@defoe-robinson-103.txt 6.3347816E-4
16. ice@callwild 6.107092E-4
17. traces@callwild 6.107092E-4
18. solleks@callwild 5.888939E-4
19. around@callwild 5.670786E-4
20. dave@callwild 5.23448E-4

Depending on the pre-processing the ranking could be slightly different.

Project PageRank: exercise 5.2

Top 10 ranking according to the pageRank* for the file /Data/soc-Epinions1.txt
(damping factor used: 0,85, number of iterations: 199)

```
Rank 1: nodeId=18, PageRank=347.6029
Rank 2: nodeId=737, PageRank=243.10344
Rank 3: nodeId=118, PageRank=163.19756
Rank 4: nodeId=1719, PageRank=161.73041
Rank 5: nodeId=136, PageRank=155.348
Rank 6: nodeId=143, PageRank=150.72754
Rank 7: nodeId=790, PageRank=150.51283
Rank 8: nodeId=40, PageRank=140.00058
```

```
Rank 9: nodeId=725, PageRank=124.71484
Rank 10: nodeId=1619, PageRank=121.212975
```

*Please note that page rank values of each node were initialized at 1, so the above page rank values are not normalized (it would not change the order of the ranking).

Project Trees of Paris: exercise 5.3

Part 1: Count the number of tree per type.
Note that I assumed that what was meant by "type" was the field "genre".
Basically, this is just a WordCount for the different type of trees.

Number of trees for each type:

```
araucana        1
atlantica       2
australis       1
baccata 2
bignonioides    1
biloba  5
bungeana        1
cappadocicum    1
carpinifolia    4
colurna 3
coulteri        1
decurrens       1
dioicus 1
distichum       3
excelsior       1
fraxinifolia    2
giganteum       5
giraldii        1
glutinosa       1
grandiflora     1
hippocastanum   3
ilex    1
involucrata     1
japonicum       1
kaki    2
libanii 2
monspessulanum  1
nigra   3
nigra laricio   1
opalus  1
orientalis      8
papyrifera      1
petraea 2
pomifera        1
pseudoacacia    1
sempervirens    1
serrata 1
stenoptera      1
suber   1
sylvatica       8
tomentosa       2
tulipifera      2
ulmoides        1
virginiana      2
x acerifolia    11
```

Part 2: Compute the height of the highest tree per type of trees.

araucana	9.0		
atlantica	25.0	japonicum	10.0
australis	16.0	kaki	14.0
baccata	13.0	libanii	30.0
bignonioides	15.0	monspessulanum	12.0
biloba	33.0	nigra	30.0
bungeana	10.0	nigra laricio	30.0
cappadocicum	16.0	opalus	15.0
carpinifolia	30.0	orientalis	34.0
colurna	20.0	papyrifera	12.0
coulteri	14.0	petraea	31.0
decurrens	20.0	pomifera	13.0
dioicus	10.0	pseudoacacia	11.0
distichum	35.0	sempervirens	30.0
excelsior	30.0	serrata	18.0
fraxinifolia	27.0	stenoptera	30.0
giganteum	35.0	suber	10.0
giraldii	35.0	sylvatica	30.0
glutinosa	16.0	tomentosa	20.0
grandiflora	12.0	tulipifera	35.0
hippocastanum	30.0	ulmoides	12.0
ilex	15.0	virginiana	14.0
involucrata	12.0	x acerifolia	45.0

Part 3: Compute the borough of the oldest tree.

The oldest tree I found was in the 5th borough of Paris.