Big Data Analytics
Lab 02 - Wikipedia

Damien Rochat, Dorian Magnin, and Nelson Rocha
Master of Science in Engineering
University of Applied Sciences and Arts Western Switzerland

April 2, 2019

1 Results

Here is the number of articles where each programming language has been found. Note, the comparison is made by lowercase.

Rank	Language	Number of articles
1	JavaScript	1'721
2	C#	707
3	Java	618
4	CSS	400
5	C++	335
6	Python	315
7	MATLAB	307
8	PHP	302
9	Perl	167
10	Ruby	125
11	Haskell	56
12	Objective-C	47
13	Scala	44
14	Clojure	26
15	Groovy	26

Table 1: Wikipedia lab ranking

Rank	Language
1	JavaScript
2	Java
3	Python
4	PHP
5	C#
6	C++
7	CSS
8	Ruby
9	C
10	Objective-C
11	Swift
12	Scala
13	Shell
14	Go
15	R
16	TypeScript
17	PowerShell
18	Perl
19	Haskell
20	Lua

Table 2: RedMonk top-20 ranking (June 2018)

Except for JavaScript, large winner in both cases, the Wikipedia ranking doesn't match RedMonk one. However, the top-5 of the two rankings is almost the same, in different order.

2 Performances

Iteration	Time
Naive ranking	32'798 ms
Ranking using inverted index	13'250 ms
Ranking using reduceByKey	10'607 ms

Table 3: Processing time comparison

TODO: explain...