ΑΣΚΗΣΗ – Υλοποίηση με MapReduce

2.1

+---------+--------------+

|dayofyear|SearchesPerDay|

+---------+--------------+

|60 |46199 |

|61 |47676 |

|62 |41796 |

|63 |44450 |

|64 |51358 |

|65 |46295 |

|66 |46293 |

|67 |44487 |

|68 |45885 |

|69 |40205 |

|70 |41702 |

|71 |49048 |

|72 |46631 |

|73 |45875 |

|74 |45996 |

|75 |44054 |

|76 |40176 |

|77 |43593 |

|78 |51868 |

|79 |46023 |

|80 |45960 |

|81 |46805 |

|82 |44447 |

|83 |42100 |

|84 |43730 |

|85 |47799 |

|86 |44812 |

|87 |45595 |

|88 |43357 |

|89 |41886 |

|90 |37625 |

|91 |40064 |

|92 |42404 |

|93 |43696 |

|94 |33676 |

|95 |31826 |

|96 |32078 |

|97 |29517 |

|98 |31884 |

|99 |33608 |

|100 |32155 |

|101 |32351 |

|102 |26196 |

|103 |21435 |

|104 |30228 |

|105 |29385 |

|106 |30879 |

|107 |34250 |

|108 |31342 |

|109 |31202 |

|110 |30423 |

|111 |30496 |

|112 |32126 |

|113 |35361 |

|114 |33491 |

|115 |32383 |

|116 |31619 |

|117 |30590 |

|118 |27158 |

|119 |29749 |

|120 |32293 |

|121 |31523 |

|122 |32498 |

|123 |37641 |

|124 |39857 |

|125 |35768 |

|126 |38387 |

|127 |43781 |

|128 |43692 |

|129 |42996 |

|130 |42116 |

|131 |41788 |

|132 |39048 |

|133 |39718 |

|134 |40884 |

|135 |44437 |

|136 |35393 |

|137 |317 |

|138 |41455 |

|139 |38850 |

|140 |39402 |

|141 |44484 |

|142 |46346 |

|143 |43360 |

|144 |39281 |

|145 |38482 |

|146 |37256 |

|147 |35715 |

|148 |34305 |

|149 |42689 |

|150 |43563 |

|151 |45788 |

+---------+--------------+

+----------+---------------+

|weekofyear|SearchesPerWeek|

+----------+---------------+

|9 |231479 |

|10 |313915 |

|11 |318193 |

|12 |316864 |

|13 |295743 |

|14 |236285 |

|15 |202629 |

|16 |225200 |

|17 |217283 |

|18 |259455 |

|19 |290242 |

|20 |244338 |

|21 |274745 |

|22 |132040 |

+----------+---------------+

+-----+----------------+

|month|SearchesPerMonth|

+-----+----------------+

|3 |1393726 |

|4 |963865 |

|5 |1200820 |

+-----+----------------+

Total searches: 3558411

Success searches percentage: 53.1295569848452%

Unsuccess searches percentage: 46.8704430151548%

+--------------------------------+----------------+

|url |distinctVisitors|

+--------------------------------+----------------+

|http://www.google.com |8438 |

|http://en.wikipedia.org |5166 |

|http://www.amazon.com |5121 |

|http://www.mapquest.com |4737 |

|http://www.imdb.com |3730 |

|http://www.myspace.com |3557 |

|http://www.yahoo.com |3321 |

|http://www.geocities.com |2361 |

|http://www.bizrate.com |1888 |

|http://profile.myspace.com |1818 |

|http://www.answers.com |1808 |

|http://cgi.ebay.com |1686 |

|http://www.tripadvisor.com |1676 |

|http://www.nextag.com |1670 |

|http://www.superpages.com |1648 |

|http://www.ask.com |1538 |

|http://dir.yahoo.com |1452 |

|http://www.angelfire.com |1397 |

|http://shopping.msn.com |1338 |

|http://www.tv.com |1324 |

|http://www.ebay.com |1322 |

|http://www.walmart.com |1318 |

|http://travel.yahoo.com |1304 |

|http://www.southwest.com |1235 |

|http://www.findarticles.com |1197 |

|http://www.city-data.com |1187 |

|http://mail.yahoo.com |1176 |

|http://www.irs.gov |1167 |

|http://www.epinions.com |1163 |

|http://www.switchboard.com |1116 |

|http://www.nlm.nih.gov |1111 |

|http://www.azlyrics.com |1071 |

|http://www.target.com |989 |

|http://disney.go.com |980 |

|http://shopping.yahoo.com |968 |

|http://music.myspace.com |953 |

|http://www.microsoft.com |948 |

|http://www.cooks.com |944 |

|http://www.msnbc.msn.com |942 |

|http://www.bankofamerica.com |924 |

|http://www.cnn.com |923 |

|http://www.ehow.com |906 |

|http://www.msn.com |888 |

|http://www.webmd.com |888 |

|http://www.weather.com |868 |

|http://www.ncbi.nlm.nih.gov |861 |

|http://www.usps.com |821 |

|http://www.topix.net |802 |

|http://members.aol.com |799 |

|http://www.sing365.com |784 |

|http://www.greatschools.net |777 |

|http://www.infoplease.com |769 |

|http://www.kbb.com |760 |

|http://www.drugs.com |749 |

|http://www.shopping.com |740 |

|http://www.idolonfox.com |737 |

|http://www.nbc.com |712 |

|http://maps.yahoo.com |701 |

|http://dictionary.reference.com |699 |

|http://www.medicinenet.com |698 |

|http://www.ticketmaster.com |695 |

|http://www.flickr.com |680 |

|http://www.gamespot.com |679 |

|http://www.lowes.com |665 |

|http://www.hotmail.com |665 |

|http://dmoz.org |642 |

|http://www.pbs.org |635 |

|http://experts.about.com |633 |

|http://www.whitepages.com |629 |

|http://www.homedepot.com |628 |

|http://www.mayoclinic.com |628 |

|http://www.wunderground.com |614 |

|http://hotel-guides.us |610 |

|http://groups.msn.com |606 |

|http://www.mtv.com |605 |

|http://www.bbc.co.uk |603 |

|http://www.cduniverse.com |589 |

|http://www.usatoday.com |587 |

|http://www.craigslist.org |586 |

|http://www.automotive.com |575 |

|http://southernfood.about.com |564 |

|http://www.epodunk.com |560 |

|http://www.starpulse.com |556 |

|http://www.xanga.com |556 |

|http://www.enchantedlearning.com|555 |

|http://abc.go.com |554 |

|http://www.wrongdiagnosis.com |552 |

|http://www.fotosearch.com |551 |

|http://www.cdc.gov |549 |

|http://www.partypop.com |548 |

|http://www.washingtonpost.com |547 |

|http://www.foodnetwork.com |547 |

|http://www.rottentomatoes.com |543 |

|http://www.dealtime.com |540 |

|http://www.apple.com |537 |

|http://kidshealth.org |534 |

|http://www.lyricsdownload.com |530 |

|http://www.bedbathandbeyond.com |525 |

|http://www.hgtv.com |522 |

|http://www.homes.com |519 |

+--------------------------------+----------------+

only showing top 100 rows

+--------+-----------+

|keywords|Appearances|

+--------+-----------+

|of |110575 |

|- |104052 |

|in |91521 |

|the |82961 |

|for |70107 |

|and |66675 |

|to |45168 |

|free |45149 |

|a |36220 |

|google |34970 |

|new |26856 |

|http |24394 |

|on |24113 |

|yahoo |23491 |

|county |22435 |

|pictures|21416 |

|how |19377 |

|my |19082 |

|lyrics |18476 |

|school |18091 |

|myspace |16549 |

|sex |16511 |

|florida |15779 |

|ebay |15603 |

|com |15051 |

|sale |14612 |

|american|14479 |

|city |14331 |

|with |14228 |

|home |13794 |

|state |13622 |

|what |12466 |

|is |12252 |

|www |12171 |

|music |12161 |

|pics |11704 |

|.com |11691 |

|games |11639 |

|york |11556 |

|i |11234 |

|girls |10980 |

|beach |10887 |

|bank |10707 |

|texas |10666 |

|online |10664 |

|black |10567 |

|high |10506 |

|nude |10457 |

|by |10118 |

|aol |10083 |

+--------+-----------+

only showing top 50 rows

Distinct keywords: 580462

+-----------+-----------+-------------------+

|firstLetter|appearances|frequency |

+-----------+-----------+-------------------+

|!Symbol |1917269 |4.968702112169262 |

|0Number |1277659 |3.3111197766849316 |

|A |2597540 |6.731659852483162 |

|B |1941726 |5.032084009199284 |

|C |2849172 |7.38377683814439 |

|D |1651073 |4.278841349258351 |

|E |1037739 |2.689354632815055 |

|F |1430292 |3.7066760218236663 |

|G |1226079 |3.1774475270983684 |

|H |1372201 |3.5561300738278727 |

|I |1171199 |3.0352231560764493 |

|J |820905 |2.127417849941212 |

|K |816228 |2.1152972651326354 |

|L |1630389 |4.22523768872232 |

|M |2263605 |5.866249848361816 |

|N |964946 |2.5007076880487755 |

|O |1767027 |4.579342371405313 |

|P |1869143 |4.843981290342527 |

|Q |118624 |0.30742025852959437|

|R |1606566 |4.163499457601135 |

|S |3416790 |8.854788994933179 |

|T |2241762 |5.809642475453533 |

|U |493808 |1.2797290853788603 |

|V |584449 |1.5146299456885866 |

|W |1035340 |2.683137396024668 |

|X |59226 |0.1534872557970879 |

|Y |259430 |0.6723263224164812 |

|Z |166730 |0.4320894566414829 |

+-----------+-----------+-------------------+

Persentage of searches with results from Wikipedia: 71.67793826351146