

WIKIPEDIA SEARCH RATES EVALUATION USING R AND SHINY

1st Author

G.Sai Divya

M.Tech,CSE ,KMIT.

sdivyagurrapu.11@gmail.com

2nd Author

C.Jagadeeswari

M.Tech,CSE,KMIT

jmerugumala@gmail.com

ABSTRACT

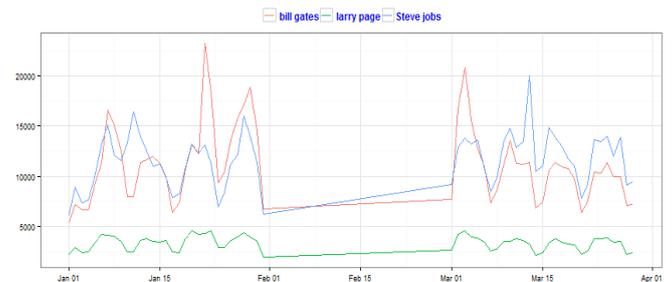
The project allows representing the daily number of Wikipedia searches for any subject(s) - animal, vegetable or mineral - over recent months.

The data is available from December 2007 to the present day. The time covered can be amendable with the help of range parameters. Increasing the number of subjects and extending the time period will impact processing time.

The data can also be downloaded in a csv format which can be useful for further analysis.

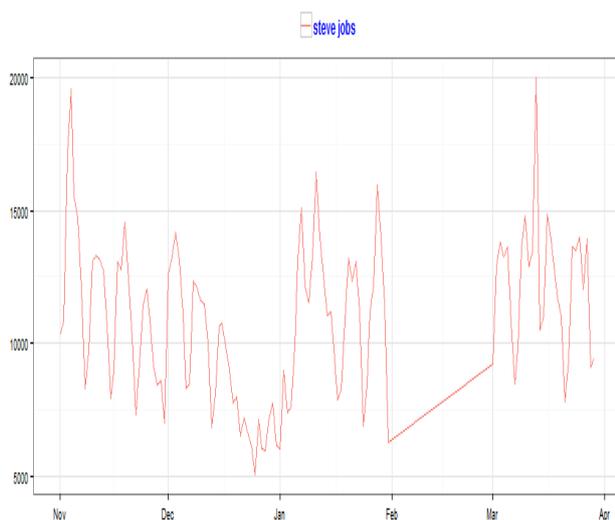
The Search rates are customizable to an extent of multiple subjects such as animal, vegetable or mineral

Daily rates from Jan to Mar 2015



	Mean	Median	Min	Max	Max Date
1 Steve jobs	11827.76	12119	6030	19983	13 Mar, 2015
2 bill gates	10948.18	10737	5306	23220	22 Jan, 2015
3 larry page	8308.81	8449	1916	4555	3 Mar, 2015

Daily rates from Nov 2014 to Mar 2015



	Mean	Median	Min	Max	Max Date
1 steve jobs	11190.03	11195.50	5022	19983	13 Mar, 2015

Advantages of the System:

The Wikipedia search rates can be useful for analysis of specific subject rates and multiple subjects rates over time. The data can be used for further analysis.

Keywords

Data Mining, Wikipedia, R, Shiny, Search Rates

1. INTRODUCTION

Recent technological advancements have led to a deluge of data from distinctive domains such as Health care and scientific sensors, user-generated data, Internet and financial companies, and social media websites over the past two decades. Lately one of the most popular online encyclopedias – Wikipedia which

is used worldwide by users for acquiring and sharing knowledge, has offered a web page with daily searches for each month in JSON format, which simplifies matters for data analysis of single and multiple search rates that takes places over the time. By employing the idea of data analysis, we visualize the wiki search rates over the time using R and its web interface Shiny. The interactive and visual parameters can be useful for comparison of different search rates on Wikipedia and also helpful for further data analysis.

Wiki Search rates helps to explore the cooffending comparisons between single and multiple subjects- over the time. In order to test the functionality of this project, the data is downloaded from wiki web page in JSON format. This project allows representing the daily number of Wikipedia searches for any subject(s) - animal, vegetable or mineral - over recent months using R and its Shiny package.

1.1 EXISTING SYSTEM

With the advancement of technology, many social media websites such as LinkedIn, twitter and Facebook with the help of their API'S have been used for data analysis and visualization. Wikipedia being the largest encyclopedia has never been used for data analysis which could give rise to many challenges for further analysis in data science.

1.2 PROPOSED SYSTEM

In the proposed system the system allows representing the daily number of Wikipedia searches for any subject(s) - animal, vegetable or mineral - over recent months. The data is available from December 2007 to the present day. The time covered can be amendable with the help of range parameters. Increasing the number of subjects and extending the time period will

impact processing time using R and the Shiny package in R.

Given Subject name(s) and range of time interval, the graphs come up within few seconds by displaying the search rates over time. The data can also be downloaded in a csv format which can be useful for further analysis. The Search rates are still more customizable to an extent of multiple subjects such as animal, vegetable or mineral.

2. METHODOLOGY

Shiny Package

Shiny makes it super simple for R users to turn analyses into interactive web applications that anyone can use. Let your users choose input parameters using friendly controls like sliders, drop-downs, and text fields. Easily incorporate any number of outputs like plots, tables, and summaries.

Loading data from Wiki in using RJSONSIO Package

This package provides facilities for reading and writing data in JSON - [Javascript Object Notation](#) - format. This allows us to read data in this format, e.g. from Web service applications, and also to export objects from R for use in Javascript/ECMAScript/ActionScript code. This is used for generating dynamic plots with SVG (Scalable Vector Graphics), or Flash and ActionScript.

Converting JSON format into R objects

The fromJSON() function will allow to read in and parse the data. It loads the data from wiki in JSON format and converts it into R objects.

Download output as .csv

The resulting data visualizations represented as wiki search rates are downloaded in a .csv format which can be useful for further analysis.

Fig2:The search rates are compared using multiple subjects

Daily rates from Feb to Apr 2015

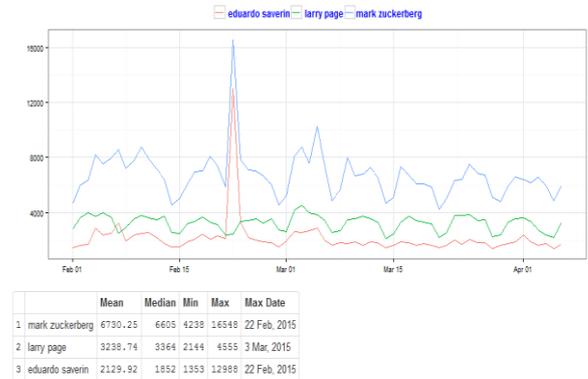
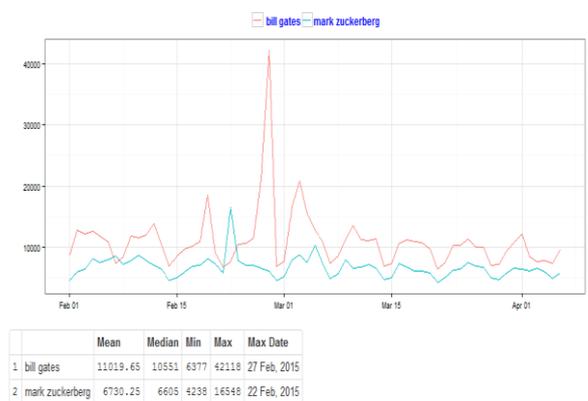


Fig3: The search rates are compared using multiple subjects and statistics values.

Daily rates from Feb to Apr 2015



3. RESULTS

Wikipedia Search Rates

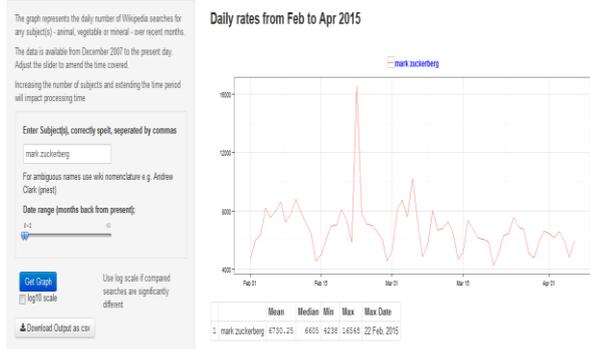


Fig1:The graph results shows the daily search rate for a particular subject.

Wikipedia Search Rates

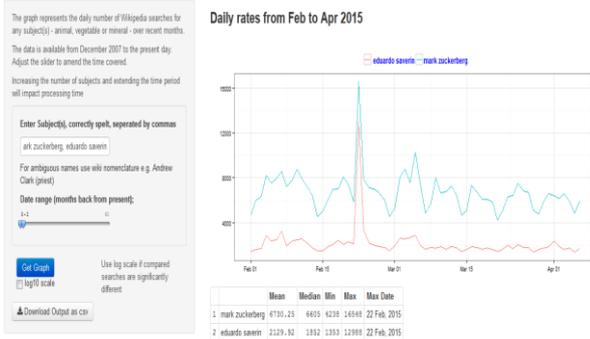


Fig4: Comparison of Bill Gates and Mark Zuckerberg for daily search rates.

4. CONCLUSION

In this project, we created a system to analyze and visualize Wikipedia search rate using the R environment and using the R shiny package. This project makes it super simple for R users and others to turn analyses into interactive web applications that anyone can use. The interactive and visual features of this project can be useful in discovering and analyzing both relational and temporal patterns of search rates of different subjects over recent time.

Given Subject name(s) and range of time interval, the graphs come up within few seconds by displaying the search rates over time. The data can also be downloaded in a csv format which can be useful for further analysis. The Search rates are still more customizable to an extent of multiple subjects such as animal, vegetable or mineral.

5. REFERENCES

1. Ihaka, Ross (1998). R: Past and Future History. Interface98 (Technical report). Statistics Department, The University of Auckland, Auckland, New Zealand.
2. Fox, John and Andersen, Robert (January 2005). "Using the R Statistical Computing Environment to Teach Social Statistics Courses" (PDF). Department of Sociology, McMaster University. Retrieved 2006-08-03.
3. Vance, Ashlee (2009-01-06). "Data Analysts Captivated by R's Power". New York Times. Retrieved 2009-04-28. R is also the name of a popular programming language used by a growing number of data analysts inside corporations and academia. It is becoming their lingua franca...
4. David Smith (2012); R Tops Data Mining Software Poll, Java Developers Journal, May 31, 2012.
5. Karl Rexer, Heather Allen, & Paul Gearan (2011); 2011 Data Miner Survey Summary, presented at Predictive Analytics World, Oct. 2011.
6. Robert A. Muenchen (2012). "The Popularity of Data Analysis Software".
7. Tippmann, Sylvia (1 Jan 2015). "Programming tools: Adventures with R". *Nature* (517): 109–110. doi:10.1038/517109a.
8. Morandat, Frances; Hill, Brandon (2012). "Evaluating the design of the R language: objects and functions for data analysis". ECOOP'12 Proceedings of the 26th European conference on Object-Oriented Programming.