

# Sentiment Analysis on Real Time Blog using R

Priya Jaiswal<sup>1</sup> Gayatri Hambarde<sup>2</sup> Shital Meshram<sup>3</sup> Prajakta Palpankar<sup>4</sup>

<sup>1,2,3,4</sup>Department of Computer Technology

<sup>1,2,3,4</sup>RTMNU, Rajiv Gandhi College of Engineering, Nagpur, Maharashtra, India

**Abstract**— Every time you release a product or service and you want to receive feedback from users, so you know what they like and what they don't. Sentiment Analysis can help you to create a simple application in R and Shiny to perform Twitter Sentiment Analysis in real-time. Real time analysis clears the concept of public by blogs. Here we are going to use R language in which R console window is needed for coding so R studio is used we are going to take any type of live data where data is based on live analysis and find the views or comments (which is called as blogs ) of public regarding it. Based on that we are going to make negative and positive analysis on that product and make a graph using it. This is nothing but the sentiment analysis on blogs.

**Key words:** Sentiment Analysis, Real time analysis, Blogs, R Studio, Keyword spotting

## I. INTRODUCTION

### A. Sentiment Analysis

Sentiment analysis (also known as opinion mining) refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. Sentiment analysis is widely applied to reviews and social media for a variety of applications, ranging from marketing to customer service.

Generally speaking, sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document. The attitude may be his or her judgment or evaluation (see appraisal theory), affective state (that is to say, the emotional state of the author when writing), or the intended emotional communication (that is to say, the emotional effect the author wishes to have on the reader).

Sentiment Analysis is the process of determining whether a piece of writing is positive, negative or neutral. It's also known as opinion mining, deriving the opinion or attitude of a speaker. A common use case for this technology is to discover how people feel about a particular topic.

#### 1) Types Of Sentiment Analysis

A basic task in sentiment analysis is classifying the polarity of a given text at the document, sentence, or feature/aspect level — whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative, or neutral. Advanced, "beyond polarity" sentiment classification looks, for instance, at emotional states such as "angry," "sad," and "happy."

- a) Subjectivity/objectivity identification
- b) Feature/aspect-based sentiment analysis

#### 2) Methods

Knowledge-based techniques classify text by affect categories based on the presence of unambiguous affect words such as happy, sad, afraid, and bored. Some knowledge bases not only list obvious affect words, but also assign arbitrary words a probable "affinity" to particular emotions.

Statistical methods leverage on elements from machine learning such as latent semantic analysis, support vector machines, "bag of words" and Semantic Orientation — Point wise Mutual Information (See Peter Turney's work in this area).

Hybrid approaches leverage on both machine learning and elements from knowledge representation such as ontologies and semantic networks in order to detect semantics that are expressed in a subtle manner, e.g., through the analysis of concepts that do not explicitly convey relevant information, but which are implicitly linked to other concepts that do so.

#### 3) Feature

Open source software tools deploy machine learning, statistics, and natural language processing techniques to automate sentiment analysis on large collections of texts, including web pages, online news, internet discussion groups, online reviews, web blogs, and social media. Knowledge-based systems, instead, make use of publicly available resources, e.g., Word Net-Affect, SentiWord Net, SenticNet, and Emoji Sentiment Ranking to extract the semantic and affective information associated with natural language concepts.

### B. Blogs

A blog (short for weblog) is a personal online journal that is frequently updated and intended for general public consumption. Blogs are defined by their format: a series of entries posted to a single page in reverse-chronological order. Blogs generally represent the personality of the author or reflect the purpose of the Web site that hosts the blog. Topics sometimes include brief philosophical musings, commentary on Internet and other social issues, and links to other sites the author favors, especially those that support a point being made on a post.

A blog is a Web page that serves as a publicly accessible personal journal for an individual. Typically updated daily, blogs often reflect the personality of the author. A blog (a truncation of the expression weblog)[1] is a discussion or informational site published on the World Wide Web consisting of discrete entries ("posts") typically displayed in reverse chronological order (the most recent post appears first). Until 2009, blogs were usually the work of a single individual [citation needed], occasionally of a small group, and often covered a single subject. More recently, "multi-author blogs" (MABs) have developed, with posts written by large numbers of authors and professionally edited. MABs from newspapers, other media outlets, universities, think tanks, advocacy groups, and similar institutions account for an increasing quantity of blog traffic. The rise of Twitter and other "micro blogging" systems helps integrate MABs and single-author blogs into societal new streams. Blog can also be used as a verb, meaning to maintain or add content to a blog. These blogs are referred to as edublogs.

### C. R language

R2 is a programming language and software environment for statistical computing and graphics. The R language has become a de facto standard among statisticians for the development of statistical software, and is widely used for statistical software development and data analysis. R is one of the major tools used in statistical research and in applications of statistics in science, social science, economics, and business. It is maintained and distributed by an international team of statisticians and computer scientists working in universities and industry.

R provides a wide variety of statistical (linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering ...) and graphical techniques, and is highly extensible. Not only is R free, but it's also open-source. That means anyone can examine the source code to see exactly what it's doing. This also means that you, or anyone, can fix bugs and/or add features, rather than waiting for the vendor to find/fix the bug and/or add the feature—at their discretion—in a future release. One of R's strengths is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed.

R is an integrated suite of software facilities for data manipulation, calculation and graphical display. It includes

- An effective data handling and storage facility,
- A suite of operators for calculations on arrays, in particular matrices,
- A large, coherent, integrated collection of intermediate tools for data analysis, graphical facilities for data analysis and display either on-screen or on hardcopy, and

#### 1) Packages

Package may refer to:

- Packaging and labeling, the science, art, and technology of enclosing or protecting products for distribution, storage, sale, and use
- Package testing, the measurement of a characteristic or property involved with packaging, including packaging materials, packaging components, primary packages, shipping containers, unit loads, and associated processes
- Mail item larger than a letter
- Chip package or chip carrier, in electronics, the material added around a component or integrated circuit to allow it to be handled without damage and incorporated into a circuit
- Electronic packaging, an area of study in electrical engineering regarding the enclosures and protective feature of an electronic device
- Automotive package, in automobile production and marketing, a group of related features sold as a bundle
- Term often used for burying a dead body (often illegally) Euphemism for a nuclear weapon

## II. LITERATURE REVIEW

Twitter Sentiment Analysis using Machine Learning and Knowledge-based Approach is the paper which we have read and in that paper we studied that Sentiment analysis is

mainly concerned with identifying and classifying opinions or emotions that are expressed within a text. These days, sharing opinions and expressing emotions through social networking websites has become very common. Therefore, a large amount of data is generated each day, on which mining can be effectively performed to retrieve quality information. Sentiment analysis on such data can prove to be instrumental in generating an aggregated opinion on certain products.

Twitter sentiment analysis often becomes a difficult task due to the presence of slangs and misspellings. Also, we constantly encounter new words, which make it more difficult to analyze and compute the sentiment as compared to the usual sentiment analysis. [10] Twitter restricts the length of a tweet to 140 characters. Thus, extracting valuable information from short texts is yet another challenge. Knowledge-based approach and machine learning can contribute considerably towards the analysis of sentiments from tweets. In this paper, we analyze people's sentiments in their tweets about certain companies. Computing a basic sentiment score and then classifying them as positive or negative would help to serve the company by providing them critical reviews about their products from people worldwide.

## III. WORK DONE

The data is been taken from twitter datasets which are live datasets. Selected datasets are ready for the further analysis. Sentiment analysis is then done on the selected datasets by the use of text analysis and keyword spotting which is nothing but the type of sentiment analysis. Keyword spotting technique distinguishes the positive, negative and neutral comments.

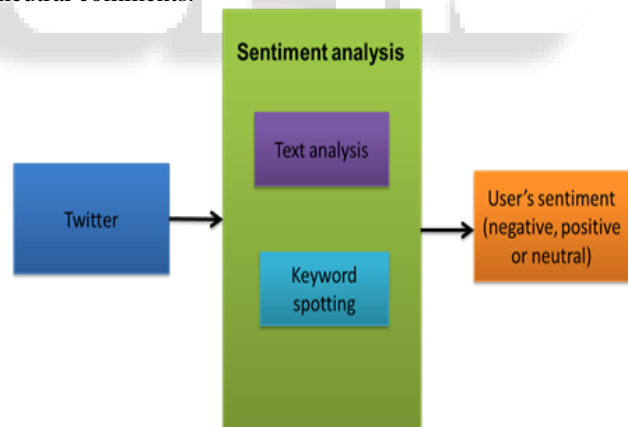


Fig. 1: Sentiment analysis on blog

This project works on sentiment analysis on blog using R. we are taking datasets from twitter and this datasets are prominently in the form of live datasets which works on real time. The instructions are coded in R language and perform using R studio. R studio is a platform which provide the interface between the R language and twitter live datasets. The purpose of sentiment analysis is to spot a keyword and perform opinion mining on that datasets. Keyword spotting helps to spot the negative, positive and neutral keywords from given tweets and provide histogram showing the statistics of the tweets.

#### IV. RESULT

##### A. Steps for Working With R

Step 1: Installation of packages

Command:

```
>install.packages('twitterR',dependencies = T)
```

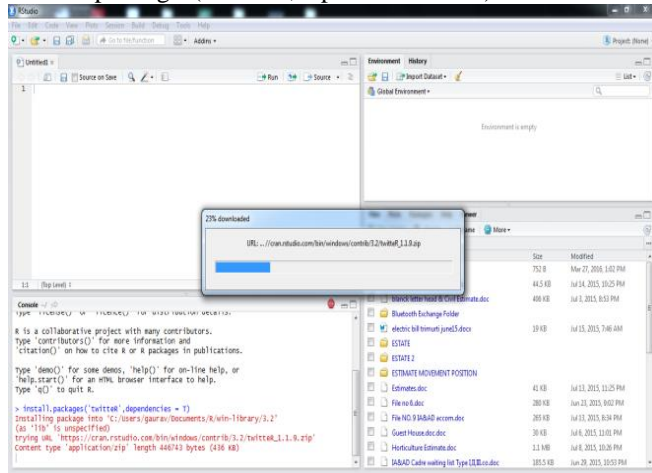


Fig. 2: Installation of packages in progress

Above instruction help to download the packages of twitter.

##### B. Providing Consumer\_Key And Consumer\_Secret

Command:

```
> library(twitteR)
> options(RCurlOptions=list(cainfo=system.file("CurlSSL",
"cacert.pem", package = "RCurl")))
> consumerKey <- "IfN2WO7TIJFGNpskIwQMlJQxo"
> consumerSecret<"c0zEL60E6MARI9udelwnyPDODOqFf
M6jQ9ANm90v8H913hZYCO"
> setup_twitter_oauth(consumerKey, consumerSecret,
"708830041585483776-
lp3p5qtiwDUg02B7FkfqiSNZrAgqOI",
"SVgiXIWRYGkHdLSy2D0kuNYqiNcQVmlPlg93emxhu
NCIip")
```

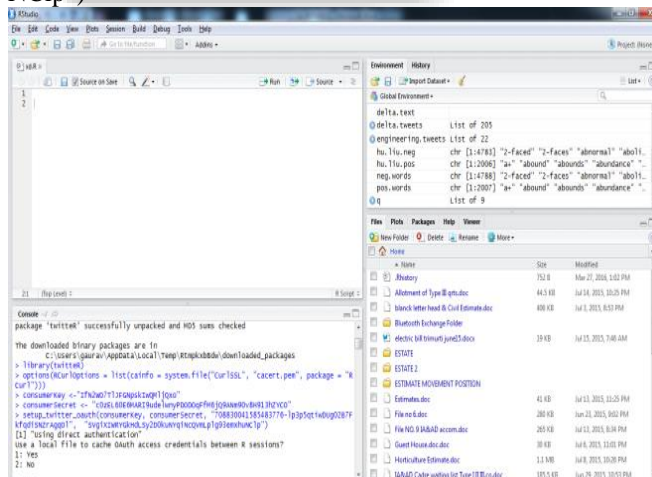


Fig. 3: providing consumer\_key and consumer\_secret

Here, Consumer\_key and consumer\_secret are created using twitter application management and then provided for authentication.

##### C. Getting Screen\_Name And Text

Command:

```
> tweet = cricket.tweets[[10]]
> tweet
[1] "Hy03953695: @Cricket IND V NWZ I'm sure"
```

```
> class(tweet)
[1] "status"
attr(,"package")
[1] "twitteR"
> tweet$getScreenName()
[1] "Hy03953695"
> tweet$text()
[1] "@Cricket IND V NWZ I'm sure"
```

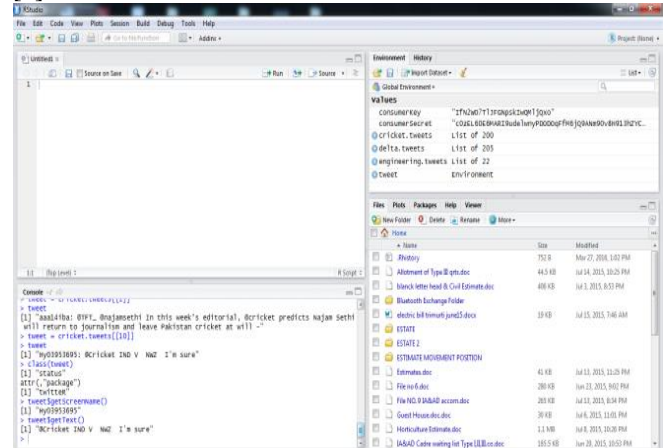


Fig. 4: Getting screen\_name and Text

The above code will help to get the latest tweet on the requested page, screen\_name of the user and tweet of the user.

##### D. Installation Of "plyr" Packages

Command:

```
> install.packages("plyr")
```

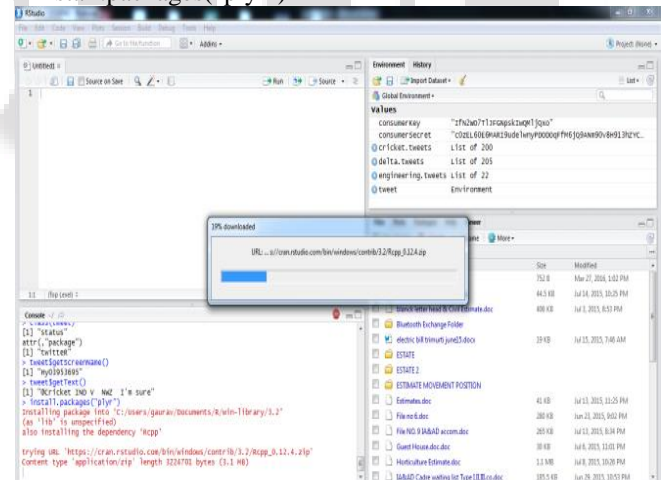


Fig. 5: installing "plyr" packages

Next step is installing "plyr" packages which is a set of tool need to break a big problem down into manageable pieces.

##### E. To See Top 10 Comments

Command:

```
> head(cricket.text, 10)
```

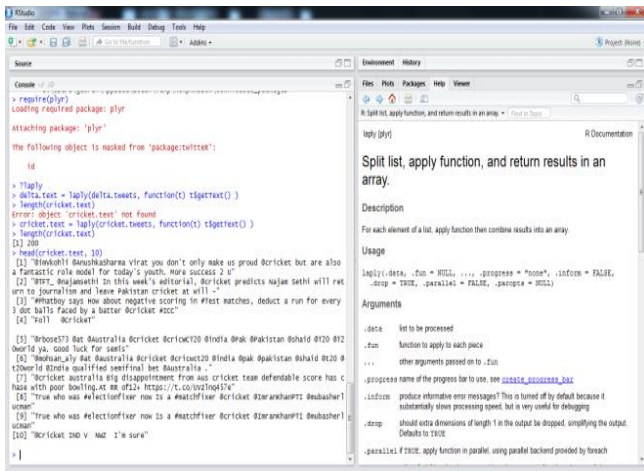


Fig. 6: Top 10 comments

The above comment tell us the top 10 live comments on cricket which tells the name and comment.

### F. To See Positive and Negative Comments

Command:

```
> hu.liu.pos = scan('D:/tweet/positive-words.txt',
+ what='character', comment.char=';')
Read 2006 items
> hu.liu.neg = scan('D:/tweet/negative-words.txt',
+ what='character', comment.char=';')
Read 4783 items
```

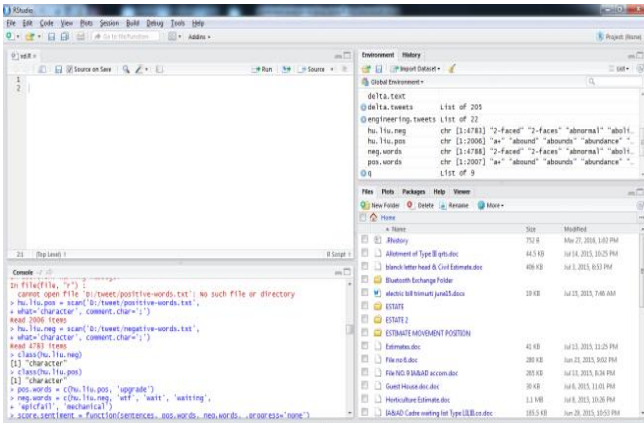


Fig. 7: Positive and negative comments

The above code gives us the positive and negative total countof by which we differtiate the different comment.

### G. Installation Of "Stringr" Packages

Command:

```
> install.packages('stringr')
```

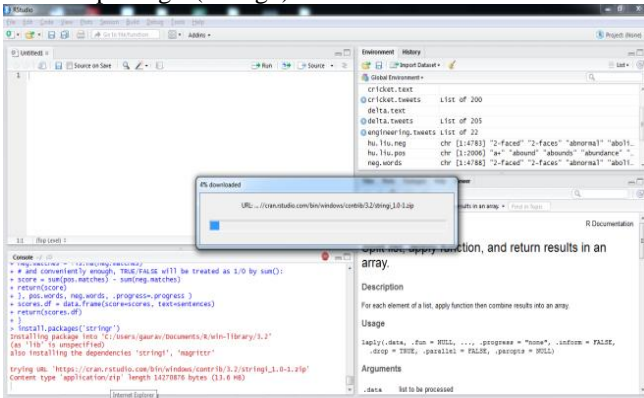


Fig. 8 Installing "stringr" packages

The next comments install the string packages and various dependencies to obtain 100% positive and negative comments.

### H. Loading Of "Stringr" Packages

Command:

```
Loading required package: stringr
```

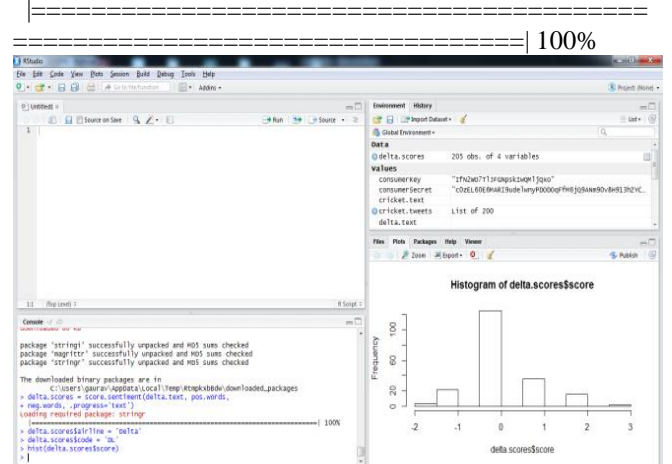


Fig. 9: Loading "stringr" packages

100% string packages are installed to run positive and negative comments

### I. To See Histogram

Command:

```
> hist(delta.scores$score)
```

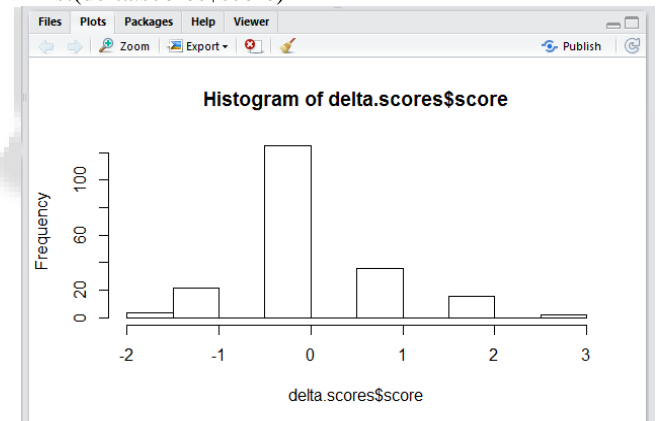


Fig. 10: Histogram

Histogram is obtained of cricket datasets based on 200 comments which are live and got from twitter.

### J. Importing Of The Dataset

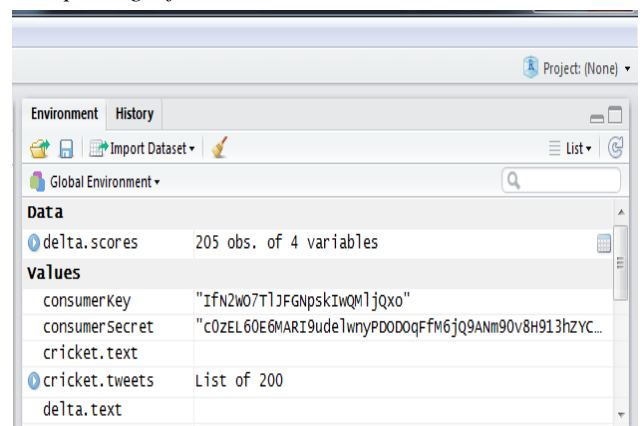


Fig. 11: Importing of the Dataset

This is import dataset which takes the codes and gives its value.

## V. CONCLUSION AND FUTURE WORK

### A. Conclusion

This project introduces the efficient and accurate method of sentiment analysis on blogs which we have done using R language. As we know that big data is current issue and there is huge amount of data to deal with, this project works on how this data can be analyzed that too in less time. In this project we deals & successfully executed the integration of R for sentiment-analysis on Tweets and its graphical visualization. Finally Sentiment analysis is done on live data (twitter blogs) by the use of text analysis and keyword spotting which is nothing but keyword spotting technique that distinguishes the positive, negative comments. Thus reviews on all comments are given on particular topic or product is represented by histogram. we figured out how the is.

### B. Future Scope

The future work that can be implemented here can be to gain the first older tweets on twitter.

## REFERENCES

- [1] Sruthi Kalarikkal, Remya P. C. "Sentiment Analysis and Dataset Collection: A Comparative Study" 978-1-4799-8047-5/15/\$31.00\_c 2015 IEEE
- [2] Krushikanth R. Apala, Merin Jose, Supreme Motnam, C.-C. Chan, Kathy J. Liszka, and Federico de Gregori "Prediction of Movies Box Office Performance Using Social Media" ASONAM'13, August 25-29, 2013, Niagara, Ontario, CAN Copyright 2013 ACM 978-1-4503-2240-9 /13/08 ...\$15.00
- [3] Huang Sui, You Jianping, Zhang Hongxian, Zhou Wei "Sentiment Analysis of Chinese Micro-blog Using Semantic Sentiment Space Model" -1-4673-2964-4/12/\$31.00 ©2012 IEEE
- [4] Aurangzeb khan, Baharum Baharudin "Sentiment Classification Using Sentence-level Semantic Orientation of Opinion Terms from Blogs"978-1-4577-1884-7/11/\$26.00 ©2011 IEEE
- [5] Despo Georgiou, Andrew MacFarlane, Tony Russell-Rose "Extracting Sentiment from Healthcare Survey Data: An Evaluation of Sentiment Analysis Tools" July 28-30, 2015
- [6] Prabhjot Kaur "Web Content Classification: A Survey" volume 10 number 2 – Apr 2014
- [7] M. Trivedi, S. Sharma, N. Soni, S. Nair "Comparison of Text Classification Algorithms" Vol. 4 Issue 02, February-2015 ISSN: 2278-0181
- [8] Umadevi V#1 "Sentiment Analysis Using Weka" ISSN: 2231-5381
- [9] Raj Kumar, Dr. Rajesh Verma "Classification Algorithms for Data Mining: A Survey" Vol. 1 Issue 2 August 2012 ISSN: 2319 – 1058
- [10] Sridhar Swamy, Riya Suchdev, Rahul Ravindran "Twitter Sentiment Analysis using Machine Learning and Knowledge-based Approach" Volume 103 – No.4, October 2014
- [11] <http://searchcloudcomputing.techtarget.com/definition/big-data-Big-Data>
- [12] <https://en.wikipedia.org/wiki/Blog>.
- [13] [https://en.wikipedia.org/wiki/History\\_of\\_blogging](https://en.wikipedia.org/wiki/History_of_blogging)
- [14] <http://searchwindevelopment.techtarget.com/definition/blog>.
- [15] [http://www.webopedia.com/TERM/B/Blog\\_and\\_Ping.html](http://www.webopedia.com/TERM/B/Blog_and_Ping.html)
- [16] [http://www.webopedia.com/quick\\_ref/index.asp](http://www.webopedia.com/quick_ref/index.asp)
- [17] [https://en.wikipedia.org/wiki/Apache\\_Spark](https://en.wikipedia.org/wiki/Apache_Spark)
- [18] <http://searchdatamanagement.techtarget.com/definition/NoSQL-Not-Only-SQL>
- [19] <http://www.cbronline.com/news/big-data/analytics/10-of-the-most-popular-big-data-tools-for-developers-4570483>.