

Available online at <http://www.mecs-press.net/ijeme>

A Web Based Application for Sentiment Analysis

Himanshu Jain ^a

^a *Master of Computer Applications(S.I.T.E), Vellore Institute of Technology, Vellore-632014, India*

Abstract

Sentiment analysis, also known as opinion mining, is referred to the studying and analyzing of people's opinions, emotions, sentiments, appraisals, evaluations and attitudes about an entity like an organization, product, service, issue, individual, event, topic, etc. and about their properties. It denotes a major problem part. While in Technical industry, the name, sentiment analysis is generally used, but in academic institutions, both terms, 'sentimental analysis' and 'opinion mining' are frequently used. They, unitedly represent the unique field of study. The term 'sentiment analysis' was first appeared in (Nasukawa and Yi, 2003) probably, and the term opinion mining was first appeared in (Dave, Lawrence and Pennock, 2003). In spite of the previous researches on sentiments and opinions published earlier, they become more prominent after this date.

Opinion mining and its related fields, for example, estimations, feelings, assessments, and states of mind are the subjects of investigation of slant examination and sentiment mining. The commencement and quick development of the field agree with those of the online networking on the Web, e.g., audits, gathering dialogs, web journals, microblogs, Twitter, and interpersonal organizations, in light of the fact that without precedent for mankind's history, we have an enormous volume of stubborn information recorded in computerized frames. We are working on accessing data from various data sources available online, including the various social networking sites, blogs or forums available. By using the provided APIs (of the respective site) and the web crawler mechanism, we will generate the outcomes and use them as a representation for users. Users can access the data represented in the forms of graphs and charts to analyze the behavior of the public on a particular brand, topic, issue or product.

Index Terms: Sentiment analysis, opinion mining, natural language processing (NLP), web-crawling.

© 2017 Published by MECS Publisher. Selection and/or peer review under responsibility of the Research Association of Modern Education and Computer Science.

1. Introduction

It is a web based application where user can search any term of his interest and the server will provide a result or sentimental analysis on that term based on the online data available on several platforms. The website

* Corresponding author. Tel.: +919413232423
E-mail address: himanshujain638@gmail.com

works in three phases: UI, Backend and Result Generation.

UI allows the user to interact with the system and login. It allows user to define the search term for analysis. While the backend carry on the analysis parts with the help of different modules and then generate results and Graphs.

Even with the presence of tools and technology it is really difficult to optimize the data available in raw form and make it useful for future references and forecast unless and until we don't have sufficient and efficient data. Innumerable data is scattered over the internet in various formats including text, images, audios and videos, but that is useless until reformed clustered purified and arranged according to user's needs. If, we focus only on one available source for data, it would be difficult to gain a requisite amount and type of data which will result in insufficient information or wrong information. However, if we combine various data sources and then input data combined together, it will give us best and optimum results with high accuracy and performance.

Basically user login to the website and enter the keyword to analyze. The Server uses the backend processes of crawling the web and database online using the REST APIs provided and the Web crawler code. After the server has finished searching for the keyword from various social platforms (social networking sites, blogs, forums, etc.), will copy and store the retrieved data in the database.

Then it will process the stored file using the NLP software and generate results for each statement. These results will be stored in separate files and then be used to plot a graph representing the negative, neutral and positive statistics of that keyword.

The results are very useful insights and recommendations which can help users and organizations in taking very reliable decisions and steps or to carry on a product or service, development functions or, a brand with the possibility of positive responses from its customers or according to public preferences (Fig.1).

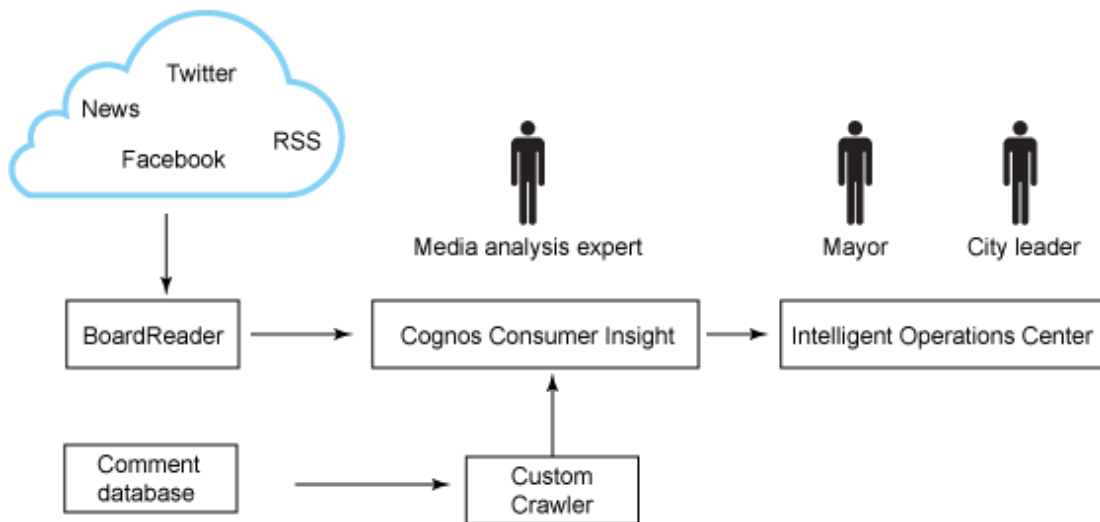


Fig.1. Uses of Sentiment Analysis through Social Networking Platforms in Real World

2. Related Works

There are a variety of tools and software which are used simultaneously for a desired graphical output, ranging from web development tools to APIs Web-crawlers, NLP etc. for achieving the results. But the present systems are limited to some extent. Such as the movie reviews and product ratings, which are given in terms of star ratings and user ratings as pre decided formats like excellent, good, average, bad or poor and a 1 to 5 star rating. Which only shows a rating not the fair review or the mood of the user. Also, the sources or the mode of the data collection is very limited or subjective. Generally, bound to a specific topic on a specific survey web

page. Thus the scope is very small and limited with almost single or minimum sources to conduct the proper feedback from users and that also, which are asked to leave a feedback, and normally people don't even bother about this. Therefore, there requires a system with more accuracy and wide area of study and fields, including more than one resource for proper user sentiment analysis.

In the present systems the major drawback is of their dependency on a single source of information (either on the social platform API or other raw text data available) which is not sufficient and completely reliable as the data is limited and may be biased in nature. So, here we developed a system which is using the Java APIs of different social networking sites, blogs portals and combine them with the data extracted using a web crawler mechanism and then produce the sentiment analysis result which provide high accuracy and better performance.

3. Working of the Proposed System and how it Works

It is a web based application which is designed using the Java and Stanford NLP software where the user search for any product, place, topic, stocks, brands, etc.. In return the application will provide in depth detail about the searched item and sentiment or opinion about that item. The result will be shown in the form of graphical representation like graphs or charts, etc. The chart will show that how many negative, positive and neutral related to that keyword are there. The application can be used for marketing it helps in judging the success of an ad campaign or new product launch, determine which versions of a product or service are popular and even identify which demographics like or dislike particular features.

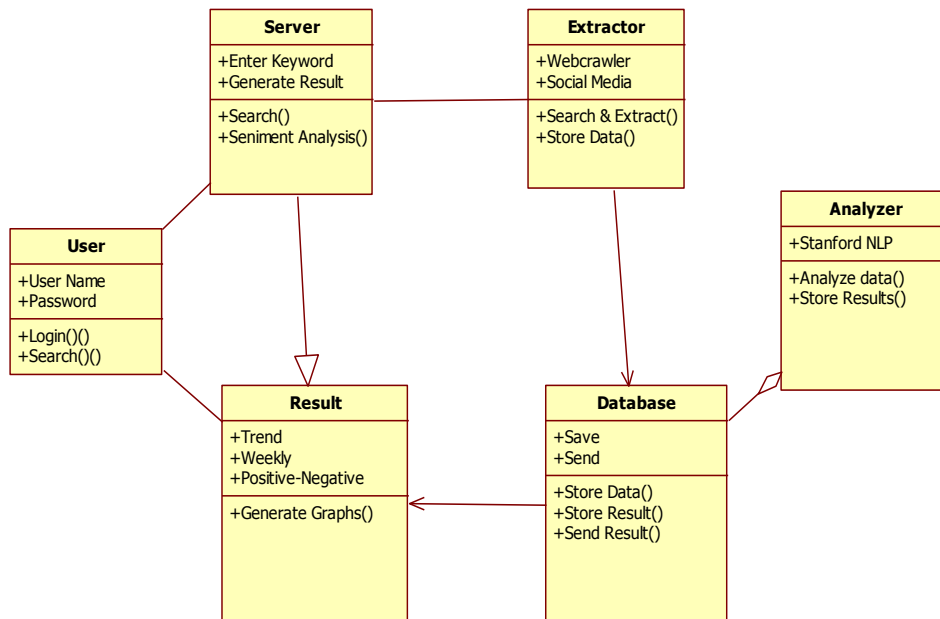


Fig.2. Class Diagram Showing the System Architecture Working of the Proposed System

The application will gather data for item from various review sites, eCommerce websites, blogs, discussion forums, social networking sites, etc. using the Java codes integrated with REST APIs and Web Crawler. That extracts the data from respected sites and store them for analytical process.

With this data we will do sentiment analysis which produces the result in form of positive opinion, negative opinion or neutral opinion. After this result from various sources we will present that result in the form of graphical representation. Sentiment Analysis is calculated by open source NLP software.

The proposed system can be divided into separate working modules integrated together with each module working simultaneously to analyze the text data and gain results. Each module as shown in Fig.2, works in an order to return output which is taken as input by the succeeding module. The following modules can be determined by their working:

3.1. Web API

Interface for users to input the keyword for Sentimental Analysis online the API validates the user and allows him to search for a topic using the web. The API, then passes the input keyword to the server for further processing and also publish the result after processing

3.2. Data Retrieval and Extraction

The server takes the input keyword and run it with the backend code for data retrieval. APIs and Web Crawlers crawl the web and the pre-determined web sites for extraction of data. Some web sites provide APIs to use data while others are accessible via a Web crawler, a special bot that go through a given link or website and access all the data (usually text) and copy it to the system. It is coded in Java.

3.3. Storing Results for Processing

The extracted data is rough and uneven. It needs to be formatted and cleaned before storing. So, all the data received is converted and formatted into plain text, removing unwanted images, links and other data. Then this clean data is stored in the JSON file for analysis.

3.4. Analyzing Data

The stored file is then passed to the NLP software in text sentences to analyze and produce results for each sentence. The Stanford NLP software analyzes the sentences for their emotions and sentiments of the user and return result in only Negative, Neutral or Positive terms.

3.5. Presenting Data

The acquired result is then stored in order to present it to the user. The available outcome of Negatives, Neutrals and Positives are statistically arranged and represented to user in forms of bar graphs, plots and charts to publish the result to user in an easily understandable format.

4. System architecture

The basic system works on a keyword input by the user through the UI and sending it for further processing to the server, which is working online using the internet and other resources and also the basic APIs and Web Crawlers for extracting data from the internet and then using the Graphs to plot the result:

- Basic Domains are UI for input of keyword, Java APIs for searching keyword related data online, web crawlers, NLP software for sentence analysis and Java graphs.
- Data is extracted using the APIs available for the different social media sites and blogs (like we have used the Twitter API) and also the Web crawler to extract data from relevant webpages (Fig.3).
- In the available data there is present a lot of unwanted data with errors and faults which needs to be rectified, aggregated, modified and arranged in sentences in order to be analyzed by the NLP software.

- These formed sentences are stored in the database for processing of NLP after which the NLP will generate its opinion for every sentence and that opinion is recorded for result generation.

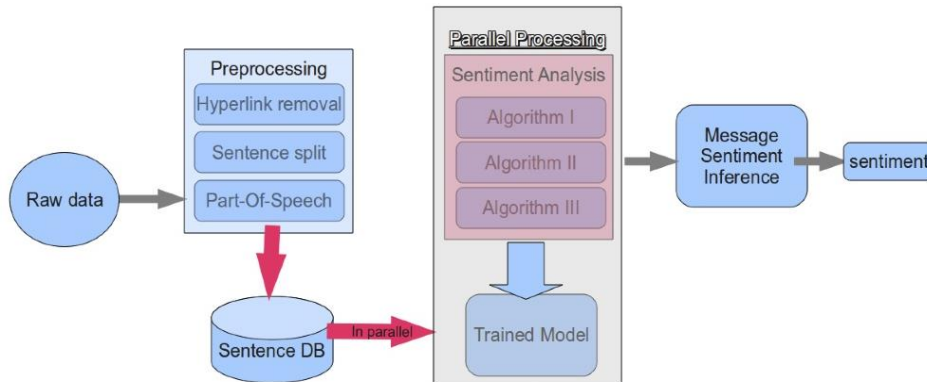


Fig.3. How the Proposed System Will Work using Web Crawler Mechanisms and Apis Together

5. Appearance and Components of a Working System

The sample screen shots of the websites and outputs working online and the API, streaming code and Web crawler code working on JAVA platform in NetBeans environment will give the graphical representation of the working of the original system being implemented and will discuss that how the system is used and working in the real time.

5.1. Web UI

The home page screen (Fig.4) will allow users to login and enter the keyword to search. It will also display the latest trending topics from different fields and user can select any topics from the list according to their choice for sentiment analysis.



Fig.4. Sample Home Page to Select or Search Most Trending Topics for Sentiment Analysis

5.2. Streaming APIs

Streaming APIs are provided by the particular sites to stream into their data with a limited access to only a part of user data with imposed restrictions and user tokens. Whenever, we want to access any data of users of a particular media platform, we have to access it by a token granted by that website and using the API provided by them to gain the information in order to access that site. For instance, we have shown here the use of Twitter Streaming API coded in Java and working in the NetBeans environment integrated with Stanford NLP to search any keyword online and analyzing it directly using NLP. It then automatically shows the results in the output window

5.3. Stanford NLP

Stanford NLP (Fig.5) is an NLP software designed and developed by members of the Stanford University team constantly working in the fields of linguistic processing and AI to make it better and better day by day. This software is basically working on the Java platform and is free to download from the university site. The user has to submit an input file, namely 'input.txt', in which all the input text to be analyzed is saved and then the NLP software is run with a command to analyze this input file and produce the opinion in return.

```

C:\WINDOWS\system32\cmd.exe
E:\STUDIES\Studies\INTERNSHIP\STANFORD NLP\stanford-corenlp-full-2015-12-09\stanford-corenlp-full-2015-12-09>java -cp "*" -mx5g edu.stanford.nlp.sentiment.SentimentPipeline -file input.txt
[main] INFO edu.stanford.nlp.pipeline.StanfordCoreNLP - Adding annotator tokenize
[main] INFO edu.stanford.nlp.pipeline.TokenizerAnnotator - TokenizerAnnotator: No tokenizer type provided. Defaulting to PTBTokenizer.
[main] INFO edu.stanford.nlp.pipeline.StanfordCoreNLP - Adding annotator ssplit
[main] INFO edu.stanford.nlp.pipeline.StanfordCoreNLP - Adding annotator parse
[main] INFO edu.stanford.nlp.parser.common.ParserGrammar - Loading parser from serialized file edu/stanford/nlp/models/lexparser/englishPCFG.ser.gz ... done [3.9 sec].
[main] INFO edu.stanford.nlp.pipeline.StanfordCoreNLP - Adding annotator sentiment
Apple is the world's largest information technology company by revenue, the world's largest technology company by total assets,[7] and the world's second-largest mobile phone manufacturer.
Negative
[8] In November 2014, in addition to being the largest publicly traded corporation in the world by market capitalization, Apple became the first U.S. company to be valued at over US$700 billion.
Negative
[9] The company employs 115,000 permanent full-time employees as of July 2015[4] and maintains 475 retail stores in seventeen countries as of March 2016.
Negative
[1] It operates the online Apple Store and iTunes Store, the latter of which is the world's largest music retailer.
Negative
There are over one billion actively used Apple products worldwide as of March 2016.
Negative
[10]
Apple's worldwide annual revenue totaled $233 billion for the fiscal year ending in September 2015.
Negative
[3] The company enjoys a high level of brand loyalty and, according to the 2014 edition of the Interbrand Best Global Brands report, is the world's most valuable brand with a valuation of $118.9 billion.
Negative
[11] By the end of 2014, the corporation continued to receive significant criticism regarding the labor practices of its contractors and its environmental and business practices, including the origins of source materials.
Negative
E:\STUDIES\Studies\INTERNSHIP\STANFORD NLP\stanford-corenlp-full-2015-12-09\stanford-corenlp-full-2015-12-09>p

```

Fig.5. Test Run of Stanford NLP in the Command Prompt using JAVA

5.4. Web Crawler

Apart from getting data sources from the social media, to be more efficient and precise, other data sources are also taken into consideration. It includes, various blogs and articles present on various web pages. But, accessing them is not that easy, as they do not provide any medium to access only data and easily extract the information for further analysis. Thus, we use web crawler, a special software to extract required information and data from a list of URLs and store them for further process. We can append any URL in the web crawler code which we want to access and then run the code. The web crawler will then automatically save the extracted information on the system in a predefined file format like HTML.

5.5. Statistical results

The final outcome of all the processes is a visual interpretation of Opinions extracted from various data sources in such a format that is easy to understand and quantifiable expressible for the user. So that, any normal person can determine the stats being represented in the form of graphs and charts and use the information for his own use.

- Weekly Analysis: A weekly analysis showing the proportions of positive, negative and neutral opinions about a product in the recent week in a form of a pie chart (Fig.6)

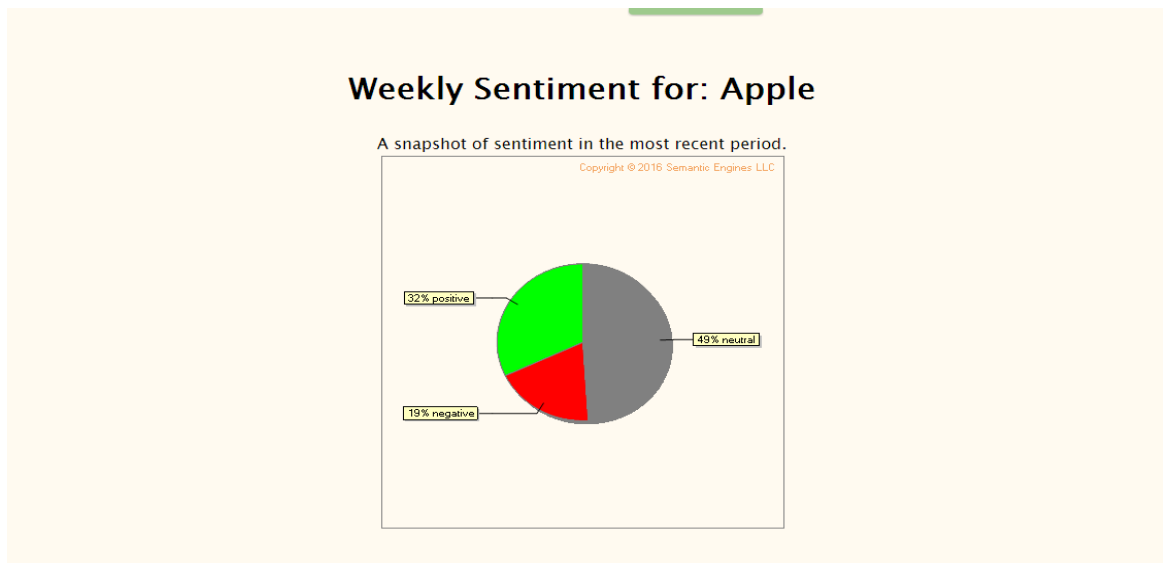


Fig.6. Weekly Sentiment Analysis in the Form of Pie Chart

- Trend Analysis: A product must be having a trend in the market over a period of time depending on its popularity, market value, user reviews and satisfaction. This is depicted using a trend analysis graph over such period of time (Fig.7).

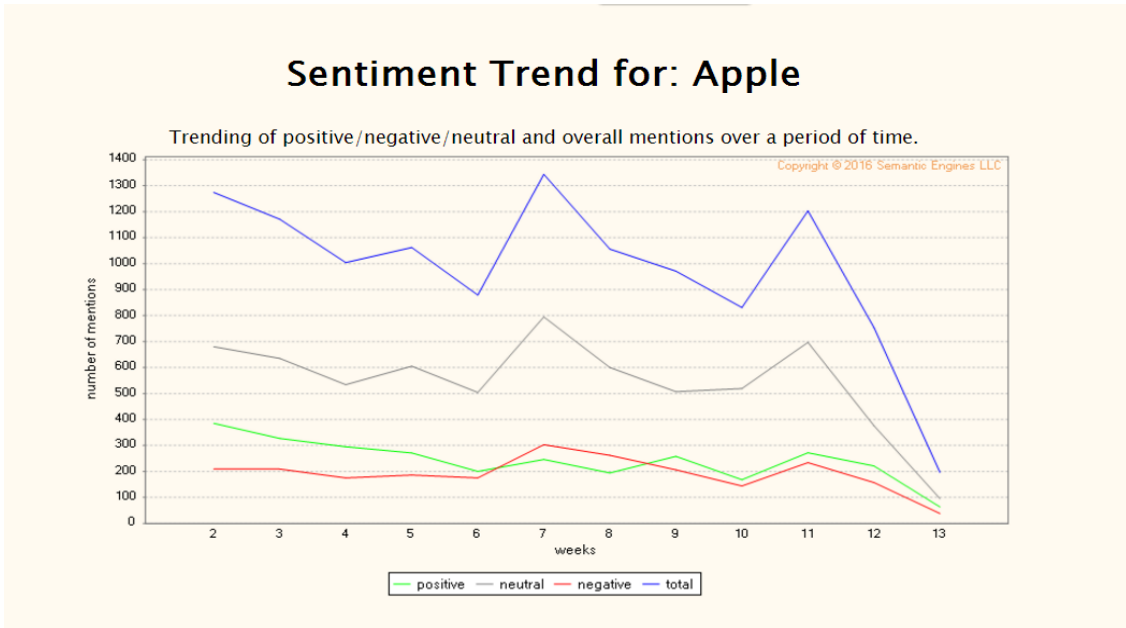


Fig.7. Graph Showing Positive, Negative and Neutral Trends Over a Period of Time

- Positive – Negative Ratio: A positive to negative ratio about a product indicates the shift of people or user about that product from negative to positive aspects and showing how the opinion is differencing between the positive and negative comments of users. It will show the ratio between the negative and positive opinions (Fig.8).

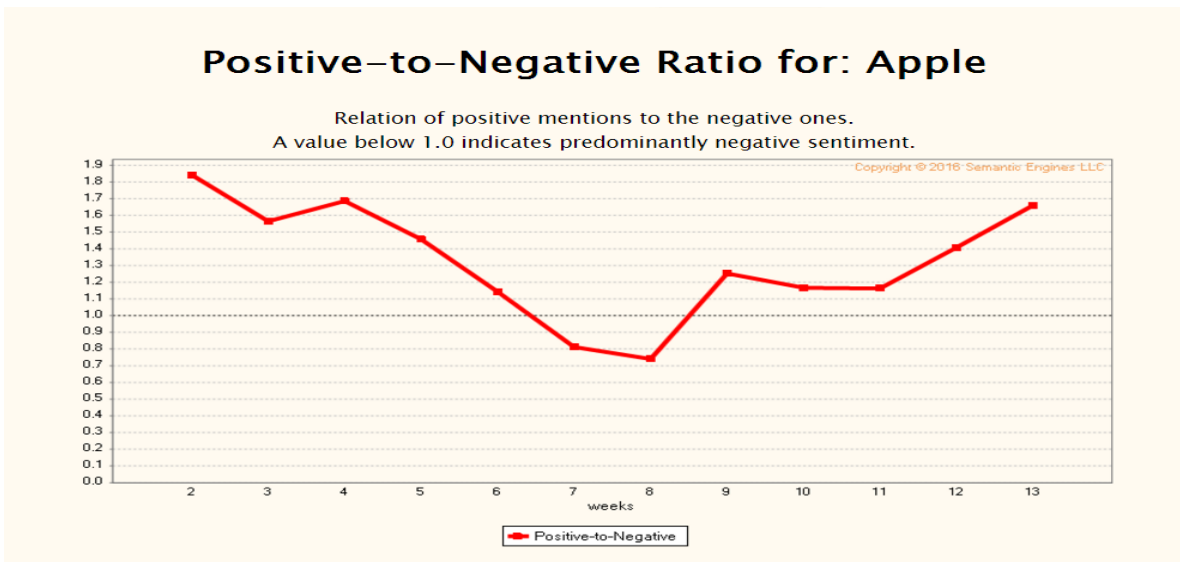


Fig.8. Plot Chart Showing the Difference Ratio between Positive and Negative

6. Conclusions

The ascent of online networking, for example, web journals and informal communities have energized enthusiasm for estimation investigation. With the multiplication of surveys, evaluations, suggestions and different types of online expression, online feeling has transformed into a sort of virtual money for organizations hoping to showcase their items, distinguish new open doors and deal with their notorieties. As organizations hope to computerize the procedure of sifting through the clamor, understanding the discussions, recognizing the pertinent substance and utilizing it fittingly, numerous are presently looking into the field of opinion mining. Users or companies can use this useful insight or information for their benefit and profits in following ways:

- **Advertising:** Sentiment examination can likewise help organizations create and refine their advertising methodology. For instance, organizations can utilize Sentiment Analysis to distinguish deals, leads and spot industry patterns.
- **Marketing:** Organizations are progressively utilizing the data found as a part of client created content on item audits and online networking destinations. For instance, suppose that Samsung needs to know how customers feel about its new Galaxy mobile phone. Rather than directing an overview, experts can go online and assess the remarks clients have left on major online commerce locales like the Amazon. Samsung can investigate the substance of these audits. For instance, Samsung could decide the tone of the remarks being left to pick up understanding into the feelings purchasers feel toward the item or dissect the remarks to make sense of the amount of information clients have about the item.
- **Data Mining:** Opinion extraction can likewise be utilized for data mining, or assembling aggressive knowledge about your rivals. For instance, a brand could without much of a stretch track online networking says or says of rivals in different spots over the web, and break down how shoppers feel about the contenders and their items. This is a superb approach to pick up an aggressive edge in today's exceedingly focused commercial center.
- **Political Analysis:** Investigations of conscious examination of tweets and microblogs have demonstrated that such examination can precisely show political supposition.

In future, the field of opinion mining will become a necessary part of the big companies and organizations. They will rely on the sentimental analysis result in any major institutional change, product updates, reviews, launches, strategic planning and investments etc. for gaining a competitive age in the modern era of internet and social platforms. The large amount of data present online and growing tendency of people showing and expressing their emotions to a product or topic online will force the institutions to gain accurate insight of vast data available. Which is not easy though, because of less number of reliable resources and tools in present day.

Further confusing the matter, is the ascent of unknown online networking stages, for example, 4chan and Reddit. On the off chance that the web was about democratizing distributed, then the following phase of the web may well be founded on democratizing information mining of all the substance that is getting distributed.

One stage towards this point is refined in exploration. A few exploration groups in colleges around the globe at present spotlight on comprehension the elements of slant in e-groups through Sentiment Analysis. The "CyberEmotions" venture, for example, as of late distinguished the part of negative feelings in driving informal communities' examinations.

The issue is that most Sentiment Analysis algorithms use basic terms to express conclusion around an item or administration. Be that as it may, social variables, semantic subtleties and varying connections make it to a great degree hard to transform a string of composed content into a basic expert or con assessment. The way that people frequently differ on the notion of content shows how huge an errand it is for PCs to get this privilege. The shorter the string of content, the harder it gets to be.

Despite the fact that short content strings may be an issue, Opinion Mining inside microblogging has demonstrated that Twitter can be seen as a legitimate online marker of political conclusion. Tweets' political

assessment exhibits close correspondence to gatherings' and lawmakers' political positions, showing that the substance of Twitter messages conceivably mirrors the logged off the political scene.

Acknowledgements

I would like to express my special gratitude and thanks to my internal guide Prof. Mary Mekala A, Assistant Prof. Sr., School of Information Technology and Engineering whose esteemed guidance and immense support encouraged to complete the project successfully. I also thank the Management of VIT University for permitting me to use the library resources. I also thank all the faculty members of VIT University for giving me the courage and strength I needed to complete my goals.

References

- [1] S. M. Al- Masum, H.Prendinger, and M.Ishizuka, "*SensNet: A linguistic tool visualize numerical valence based opinion of text data*" in Proceedings of the Int. Conf. on NLP (ICON), pp. 147-152, 2007.
- [2] Bo-Pang, LillianLee, "*Opinion Mining and Sentiment Analysis*" in Proceedings of the Foundations and Trends in Information Retrieval Vol. 2 Issue 1-2, Jan 2008 Pages 1-135.
- [3] Ronen Feldman, "*Techniques and applications for sentiment analysis*" in Proceedings of the Communications of the ACM CACM Vol. 56 Issue 4, April 2013 Pages 82-89.
- [4] Mingqing-Hu, Bing-Liu, "*Sentiments features in customer reviews*, Proceedings of the 19th Nat. conference on AI, p.755-760, July 25-29, 2004, San Jose, California.
- [5] Turney, Peter (2002). "*Thumbs Up or Thumbs Down? Semantic Orientation Apply to Unsupervised Classification of Reviews*". Proceedings of the Ass. for Computational Linguistics. pp. 417-424.
- [6] Snyder, Benjamin; Barzilay, Regina (2007). "*Multiple Aspect Ranking using the Good Grief Algo.*" Proceedings of the Joint Human Language Tech./North American Chapter of the ACL Conference (HLT-NAACL). pp. 300–307.
- [7] Thelwall, Mike; Buckley, Kevan; Paltoglou, Georgios; Cai, Di; Kappas, Arvid (2010). "*Opinion strength detection in short text*". Journal of the American Society for Information Sci. and Tech. 61 (12): 2544–2558.
- [8] Cataldi, Mario; Ballatore, Andrea; Tiddi, Ilaria; Aufaure, Marie-Aude (2013-06-22). "*Good location, terrible food: detecting mood in user-generated reviews*". Social Net. Analysis and Mining 3 (4): 1149–1163.
- [9] Dey, Lipika; Haque, S. K. Mirajul (2008). "*Sentiment Analysis from Noisy Text Data*". Proceedings of the second workshop on Analytics for noisy text data, p.83-90.
- [10] Zhai, Zhongwu; Liu, Bing; Xu, Hua; Jia, Peifa (2011-01-01). Huang, Joshua Zhexue; Cao, Longbing; Srivastava, Jaideep, eds. *Constrained LDA for Grouping Product Features in Opinion Mining*. Lecture Notes in Computer Science. Springer Berlin Heidelberg. pp. 448–459.
- [11] Titov, Ivan; McDonald, Ryan (2008-01-01). "*Modeling Online Reviews with Multi-grain Topic Models*". Proceedings of the 17th International Conference on World Wide Web. WWW '08. New York, NY, USA: ACM: 111–120.
- [12] Tumasjan, Andranik; O.Sprenger, Timm; G.Sandner, Philipp; M.Welpe, Isabell (2010). "*Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment*". "Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media".
- [13] Condliffe, Jamie. "*Flaming drives online social networks* ", NewScientist, 2010-12-07. Retrieved on 2010-12-13.
- [14] CORDIS. "*Collective emotions in cyberspace (CYBEREMOTIONS)*", European Commission, 2009-02-03. Retrieved on 2010-12-13.

- [15] Roebuck, K. (2012-10-24). *Sentiment Analysis: High-impact Strategies - What You Need to Know: Definitions, Adoptions, Impact, Benefits, Maturity, Vendors*. ISBN 9781743049457.
- [16] Ogneva, M. "How Companies Can Use Sentiment Analysis to Improve Their Business". Mashable. Retrieved 2012-12-13.
- [17] Galitsky, Boris; McKenna, Eugene William. "Sentiment Extraction from Consumer Reviews for Providing Product Recommendations". Retrieved 18 November 2013.
- [18] Twitter Developers. Doc. Available: <https://dev.twitter.com/docs>.
- [19] Twitter4j Community members. API Support matrix. Available: <http://twitter4j.org/en/api-support.html>.
- [20] Christopher D. Manning, Hinrich Schütze: *Foundations of Statistical Natural Language Processing*, MIT Press (1999), ISBN 978-0-262-13360-9, p. xxxi
- [21] Cambria, Erik; White, Bebo (2014). "Jumping NLP Curves: A Review of Natural Language Processing Research" (PDF). *IEEE Computational Intelligence Magazine*. 9 (2): 48–57.

Authors' Profiles



Himanshu Jain (born 2nd October, 1991) is a student of Master of Computer Applications in V.I.T. University, Vellore, India. He completed his BCA from University of Rajasthan, Jaipur in 2012 and then did a diploma In Hardware and Networking. He worked as an intern in a start-up (Jan – Apr, 2016) named Sentilyze, focusing on the field of Sentimental analysis.

How to cite this paper: Himanshu Jain, "A Web Based Application for Sentiment Analysis", *International Journal of Education and Management Engineering(IJEME)*, Vol.7, No.1, pp.25-35, 2017.DOI: 10.5815/ijeme.2017.01.03