S.T.A.R.K. – Simultaneous Tracking And Ranking Using Word Cloud

Dr. A. Umamageswari

Head and Professor, Computer Science and Engineering
DMI College of Engineering, Palanchur Chennai India

Madhan. B

Student, Computer Science and Engineering
DMI College of Engineering, Palanchur Chennai India

Philip Joel Dcruz. R

Student, Computer Science and Engineering
DMI College of Engineering, Palanchur Chennai India

SanthanPrakash. T

Student, Computer Science and Engineering
DMI College of Engineering, Palanchur Chennai India

Abstract - **STARK (Simultaneous Tracking and Ranking) is a web application which comes handy with Telecommunication companies for knowing their product's major faults and most frequently appearing comments using a Big Data methodology called the Word Cloud. A Word cloud (tag cloud or weighted list in visual design) is a visual representation of text data, typically used to depict keyword metadata (tags) on websites, or to visualize free form text. The Word Cloud is an algorithm commonly used in big data to bring an image of words with varying font size based on their number of occurrence in a text. Tags are usually single words, and the importance of each tag is shown with font size or color. This format is useful for quickly perceiving the most prominent terms and for locating a term alphabetically to determine its relative prominence. When used as website navigation aids, the terms are hyperlinked to items associated with the tag. Thus the web application wisely employees the Word Cloud algorithm to form the word cloud image based on the comments and faults provided by the users of the web application.**

KEY WORDS - Word Cloud; Multi word cloud; Text Analysis; Clustering; visualization; Data Analysis.

I. INTRODUCTION

Text analysis is a procedure that is implemented whenever there is a need to find out certain specific aspects, characteristics of some large collection of textual data. The results of these analysis can be represented in a variety of ways depending on the need to showcase any particular aspect of the data obtained. The usual way of showing the results of the analysis is by again using plain text which can be a little monotonous. Word clouds on the other hand provide a visually pleasing way to display the main points and a kind of brief summary of the analyzed text. This can be very useful in analyzing and categorizing a large amount of text such as the comments made by users on a specific product or service. STARK therefore represents the results of the analysis as a word cloud.

Telecommunication companies are constantly innovating and are providing a great service to people everywhere. But they aren't as successful in fault analysis and rectification which can lead to poor customer satisfaction. By employing STARK, companies can find the most frequent or prevailing opinion that their customers have about their products or services. Since most users use similar words in describing their thoughts or complaints - such as good, bad, easy, cheap, etc. - STARK can be used to easily analyze all their data as a whole. The more a certain word is used by the customers the bigger they'll appear on the word cloud that STARK provides as the result of the analysis. The company will obtain a clear understanding about the views of their customers and can act or respond appropriately. If a company employs this particular application for analyzing they can get a clearer view of their product's impact and appeal. Consequently, they can improve their services with this information which can significantly increase their customer satisfaction level.

II. RELATED WORKS

There are certain works on data analysis that utilized word cloud algorithm in order obtain the required result of it. ShailajaJayashankar utilized word cloud algorithm in her work to easily employee eLearning process by displaying the word cloud chart of the content that's needed to be learnt [2]. Peng Wei, on his work of analyzing 110 police intelligence incident data to identify the most occurred crime activities he used the word cloud algorithm in order to faster find the mostly performed criminal cases on the data base [3]. Masahiko Itoh on his work Word-Clouds in the Sky: Multi-layer Spatio-Temporal Event Visualization from a Geo-Parsed Microblog Stream proposes a method of visualizing spatio-temporal events with a multi-layered geo-locational word-cloud representation from a geo-parsed microblog stream [4]. Steffen Lohmann on his work ConcentriCloud proposed that text data from different documents can be easily analyzed by forming concentric circles of word clouds within each other based on the document analyzing order [5].

III. SYSTEM STUDY

As a web application to satisfy the users by rectifying the complaints provided by the users, STARK possess five modules to handle all the process that are made for different category of users available with the STARK application. Basically, the STARK application have three different users of it, they are the network user who puts complaint into the application, the Admin who sits on the side of ISP who are available with the web application in order to accept and process the complaints that are being posted by the customers of the ISP and the third category is the Technician who works under a particular ISP and would do the field works in order to process and complete the complaints that has been posted by their customer.

The user of the application cannot only post complaints but also can receive temporary solutions from the application based on the complaint provided by them, thus this is possible by two modules namely the pattern detection and message retrieval modules. The pattern identification is where many words which have their own tag numbers are being individually searched in the complaint provided by the user such that their identification order is noted by means of their tag numbers, these tag numbers are being arrange in such a way that there able to be processed by other modules that are going to use the tag arrangement, let the tag number arrangement be called as a pattern that is being found out from the complaint that has been processed. The next module is the message display module where the temporary solution for the complaint posted by the user will be displayed as a message to them. This module involves utilization of the pattern that has been identified by the pattern identification module from the complaint provided by the user. The maximum possible combination of the tag numbers to form a meaningful pattern are being calculated and are clustered into similar problem thus there forms many clusters which may have the same patterns or different patterns from each other, thus there would be a message assigned for each cluster that is being formed from the patterns that are available. Thus, when a pattern is detected from the user complaint by the pattern identification module then the message display module would match the pattern with the patterns that are available with the clusters, when a particular match has been found then the particular cluster for the match would be choose and the message that has been assigned to that particular cluster would be displayed to the user who posted the complaint as a temporary and immediate solution, if the solution dose not rectify the problem of the user then the user would be able to submit the complaint to the particular ISP who would be able to process the complaint.

The next module that's to be discussed is Complaint management module, this module will be available only with the Admin of any particular ISP. Using this module, the Admin will be able to manage all the complaints that are being posted by the users of the ISP in which he is working for. The process that are being able to be done in this module is editing a comment to the company specified requirement, deleting a complaint if it is not required and changing the status of a complaint in order to process it once again for the user's betterment or admin's satisfaction. Another module which is being used in this particular application is the Technician management module, this module is also available only with the admin of any particular ISP. Using this module, the Admin will be able to add any number of Technician profile to the ISP by just providing their credentials details, such that the details of the particular technician must be known to the Admin before the creation process, an important credential to be provided by the admin to the technician is the area in which the respective technician would work for, this may enable to display the particular complaint that are available with that particular area only. Thus, other functions that can be added to this module is that deletion of a technician profile, modification of the technician profile, assigning a job from the complaint list to any of the technician available and checking the progress of the technicians on processing the complaints taken by them. The last module that is going to be discussed is the Word Cloud generator module, this module uses the word cloud algorithm in order to generate a word cloud chart based on the complaints that are being provided by the users. This module is available only with the admin of any ISP in the application available. This module would enable the admin to provide criteria such as data or location or both based on which the complaints from the database would be fetched. Using the complaints fetched the module would generate a word cloud chart which would be displayed to the admin as a result from which the admin could come up with a problem that needs to be solved. The entire architecture discussed in above context has been clearly represented in the figure 1.0 given below.
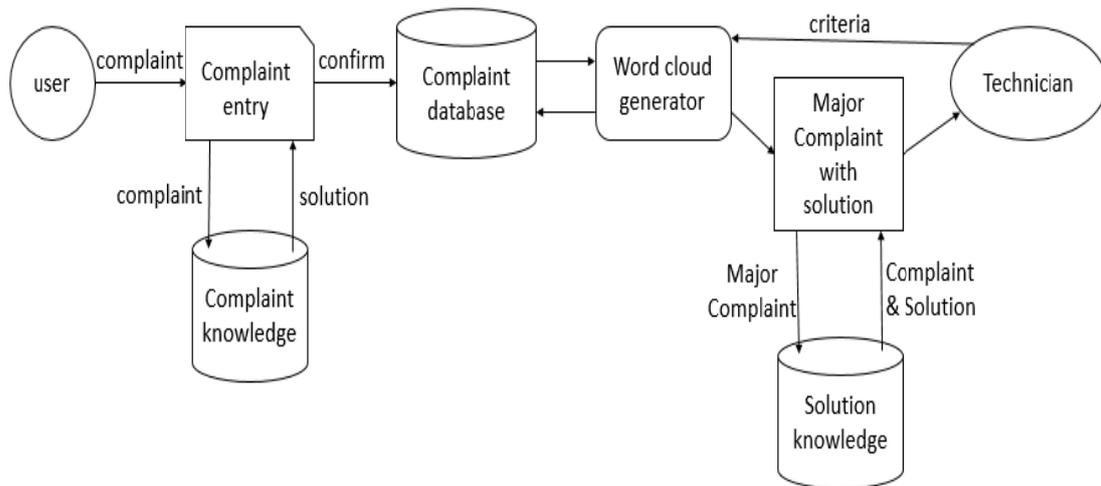
figure 1.0

IV. METHODOLOGY

The algorithm that is being used in this web application is the Word Cloud algorithm. The basic construct of Word Cloud algorithm is as follows.

1) Text Processing

The data that are needed for the word cloud generation are being fetched first. Then the occurrence of the stop words (it, is, are, etc.) are identified and removed from the context that has undergone to the text processing function.

2) Word Occurrence Frequency

The next step is where two arrays are being created, one for the storage of the words that are being used in the text provided and another is for storing the frequency of occurrence of the words. Thus, the word and its frequency are related using the index number of the array.

3) Word Cloud Generation

Here, after finishing the frequency process the maximum and minimum of the frequencies are being considered in order to find a value called spread. Then a maximum and minimum font size are being chosen which would be already chosen and available, then the following formula will be used to calculate the font size a particular word is to be displayed.

$$size = minFontSiz + (count - minimumCount) * (maxFontSize - minFontSize) / spread$$

Where,

minFontSiz = Minimum Font Size,

maxFontSize = Maximum Font Size,

minimumCount = Minimum Count,

count = frequency.

When displaying the word according to the font size that is being calculated based on its frequency a word cloud chart will be generated where most frequent words will be in larger font and low frequent words will be in the most smaller font.

V. RESULT AND DISCUSSIOIN

The process where an immediate solution is provided to the user who made a complaint involves machine learning process, in this process the pattern obtained will be matched with the patterns that are available with different clusters and the matched cluster's message will be provided as an immediate solution to the user. The figure 2.0 given below show the graph of efficiency calculated for machine learning.

The word cloud chart that show in the below figure 3.0 is a result from analyzing 300 complaints from the application. Thus, we can say from the chart that the most reported complaint is about the ISP's network connectivity and network speed, on solving these problem, approximately 70% of the complaint can be rectified using this method
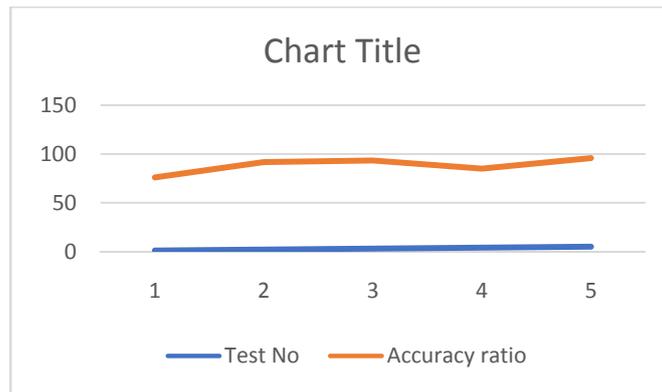
figure 2.0
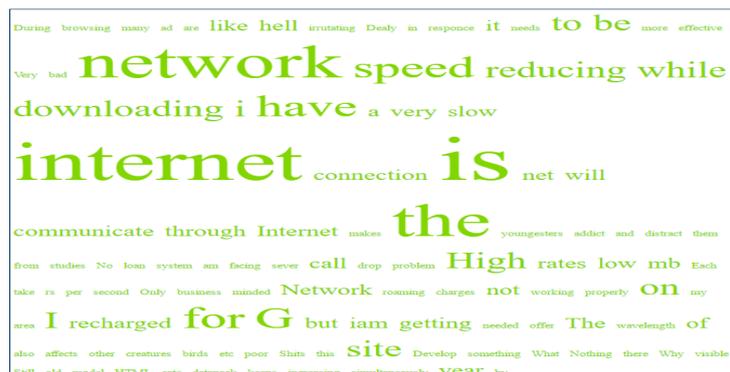
figure 3.0

VI. CONCLUSION AND FUTHER ENHANCMENT

Customer satisfaction will be greatly increased throughout the industry. ISP's won't have to spend as much money as training their customer care employees as they do currently. The work of the employees will be reduced significantly because of using STARK. About 65-95% of the complaints analyzed using STARK Word Cloud algorithm are solved. The most significant problems faced by the customers are easily identified and are solved.

STARK can be adapted to suit any industry. Offers and discounts provided by ISP can be directly advertised to customers through STARK. Google Maps functionality can be added so that the technician can easily identify the customer's location. Customers can rate the service provided by the technicians based on their personal satisfaction. Based on the rating provided by the customers the technician will be provided points which will be taken into consideration by the administrator.

REFERENCE

[1] R. Kaptein and J. Kamps. "Word clouds of multiple search results". In IRFC 2011, volume 6653 of LNCS, pages 78–93, 2011.
[2] ShailajaJayashankar, R. Sridaran, "Moving word cloud from visual to text analysis to endow eLearning", Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference, March 2016.
[3] Peng Wei, TonggeXu, Xi Qin, Chao Wang "Visualization of Police Intelligence Data Based on Word Clouds", Computational Intelligence and Security (CIS), 2014 Tenth International Conference, Nov. 2014.
[4] Masahiko Itoh, Naoki Yoshinaga, Masashi Toyoda "Word-Clouds in the Sky: Multi-layer Spatio-Temporal Event Visualization from a Geo-Parsed Microblog Stream", Information Visualisation (IV), 2016 20th International Conference, July 2016.
[5] Steffen Lohmann, Florian Heimerl, Fabian Bopp, Michael Burch, Thomas Ertl "Concentri Cloud: Word Cloud Visualization for Multiple Text Documents", Information Visualisation (iV), 2015 19th International Conference on, July 2015.
[6] Carla Binucci, Walter Didimo, Enrico Spataro, "Fully dynamic semantic word clouds", Information, Intelligence, Systems & Applications (IISA), 2016 7th International IEEE Conference, 19 December 2016.
[7] Ming-Te Chi, Shih-Syun Lin, Shiang-Yi Chen, Chao-Hung Lin, Tong-Yee Lee, "Morphable Word Clouds for Time-varying Text Data Visualization", IEEE Transactions on Visualization and Computer Graphics Volume: 21, Issue: 12, Dec. 1 2015.
[8] Florian Heimerl, Steffen Lohmann, Simon Lange, Thomas Ertl, "Word Cloud Explorer: Text Analytics based on Word Clouds", System Sciences (HICSS), 2014 47th Hawaii International Conference, 10 March 2014.
[9] R. Kaptein, D. Hiemstra, J. Kamps, "How different are language models and tag clouds", ECIR-2010, Volume 5993 of LNCS, Pages 556-568, 2010.