

Scheme Critic-An Automated Opinion Mining System for Policy Making

Haritha Akkineni^{#1}, P. V. S. Lakshmi^{*2} and B. Vijaya Babu^{#3}

ABSTRACT

Social has become mainstream and ubiquitous. Our day-to-day life has always been influenced by what people think and most of the conversations are being carried out online. At the same time, citizens are becoming more enthusiastically engaged in policy issues, more empowered, and more demanding in their relations with traditional ones. Researchers identify a hunger for and reliance upon peer advice and recommendations found online and this information hunger is strongly evident in the political sphere. In any policy making considering the opinion of the public is considered to be most crucial. Around the globe governments and policy makers are working on to improve their performance by creating systems which will facilitate them in understanding their performance.

However, people have difficulty, owing to their mental and physical limitations, producing consistent results when the amount of such information to be processed is huge. Automated opinion mining systems are thus needed, as subjective biases and mental limitations can be overcome with an objective opinion analysis system. The ever-rising expectations from ordinary citizens provide a continuing force for policy makers to provide more services and with higher standards of quality. These pressures are also reason enough to find more cost effective ways of operating so that policy makers can do more with less.

The system developed helps to tap into citizen generated content helping the policy makers more effectively engage with the citizens. It helps end users actively force social media for outreach, communication and engagement with the citizens and to identify the key initiatives required to enhance citizen delight by using the concepts of sentiment analysis.

In this paper, we showcase-Scheme Critic -An aspect based opinion mining system which has been developed to help the policy makers in framing their policies by collecting opinions from twitter. It is used to measure the quality of the schemes, their outputs and to measure the outcomes and impacts resulting from these outputs. We demonstrate our system based on different schemes launched in India.

1. INTRODUCTION

In the current scenario, the forefront confront for the governments is to focus on people centric applications. The government's triumph depends on how effectively it is able to take the schemes launched for the benefit of people and build strong alliances with them by empowering their participation in the decision-making process[1]. Governments around the globe are moving towards building electronic government applications and services[2]. So the application "Scheme Critic" which we are developing would be embedded into such systems to aid their decision making process.

Scheme Critic analyses on understanding the citizen's voice mostly related to schemes and includes information on how the schemes are perceived by the general public based on citizen sentiments expressed in social media mostly twitter. How do the citizens feel about the government policies and schemes. What do they say about them. The analysis can tell you what the most talked about aspect of the schemes in the

^{#1,3} Computer Science and Engineering , KL University, Guntur, Andhra Pradesh, India.

¹ E-mail: akkinenih@gmail.com; ³ vijay_gemini@kluniversity.in

^{*1,2} PVP Siddhartha Institute of Technology

E-mails: Vijayawada, Andhra Pradesh, India; ² papinenivsl@gmail.com

form of word cloud and the positives negatives and neutral opinions about them. By leveraging this information, you can build citizen goodwill by designing and executing on action items that are important to them. The policy makers can make corrective actions on the schemes of their choice.

However, when dealing with a large number of entities, even summaries would get confusing as users would still need to keep track of how well each entity fulfills their opinion requirements. Thus, to provide a more direct support for a policy makers decision making task, we have developed Scheme Critic, a novel system capable of Extracting opinions on the scheme specified by the policy maker. The scheme related tweets will be shown and the term frequency of the each word is depicted in the form of word cloud. The pulse of the citizens on the particular scheme is provided. The idea behind Scheme Critic is to allow policy makers select the scheme. What makes this system unique is that it gives the for and against the scheme on top of that, Scheme Critic allows policy makers to analyze these opinions and provides an abstractive summarization of these opinions. We demonstrate our current system in the context of certain main schemes launched by government of India.

Our proposed Framework is of three phases.

The first phase concentrating on user online comments, concerning a technique that detects and extracts phrases along with aspect orientation containing user opinion.

The second phase focuses on annotating the sentiment orientation of the identified opinion phrases depending on the opinion holders underlying stance against the issues they discuss.

The third phase focuses on formation of abstractive summaries. An informative and concise abstractive summary would be a better solution for the executive management to have compatibility with small screens like PDA's as well. A well formed summary is important for high level decisions about policy and strategies.

Our approach gives a better experience to the policy makers. As they will be knowing the pulse of the people in a more abstractive manner which not only save time but also gives better user experience. This idea leads to Development of Scheme Critic - An Opinion Mining System for Policy Making.

The rest of the paper is organized as follows. Section 2 gives a brief overview of the architecture used. The different modules used in the architecture is described in Section 3. Section 4 deals with DEMO of Scheme Critic. Section 5 sums up the conclusions of the paper.

2. ARCHITECTURE

In this section we will provide a brief overview of "Scheme-Critic" system architecture.

Scheme-Critic is a web application that enables policy makers to view abstractive summaries on the schemes. The goal of Scheme Critic is to first extract the relevant data for a particular scheme. Just like any other opinion mining systems, the Scheme Critic system consists of several key components ranging from collecting the user generated content component to an Abstractive summaries component as shown in Figure 1.

In brief, the system takes in tweets posted by the users related to the policy and gives results to the policy maker in the form of abstractive summaries to assist them with their decision making process. In the next few sections, we provide more information about some of the key components of the system.

The main process adapted in scheme critic is divided into 3 phases

Phase 1: Corpora Acquisition Phase.

The aim of this phase is to automatically extract tweets regarding the particular policy.

Phase 2: Sentiment Analysis Phase.

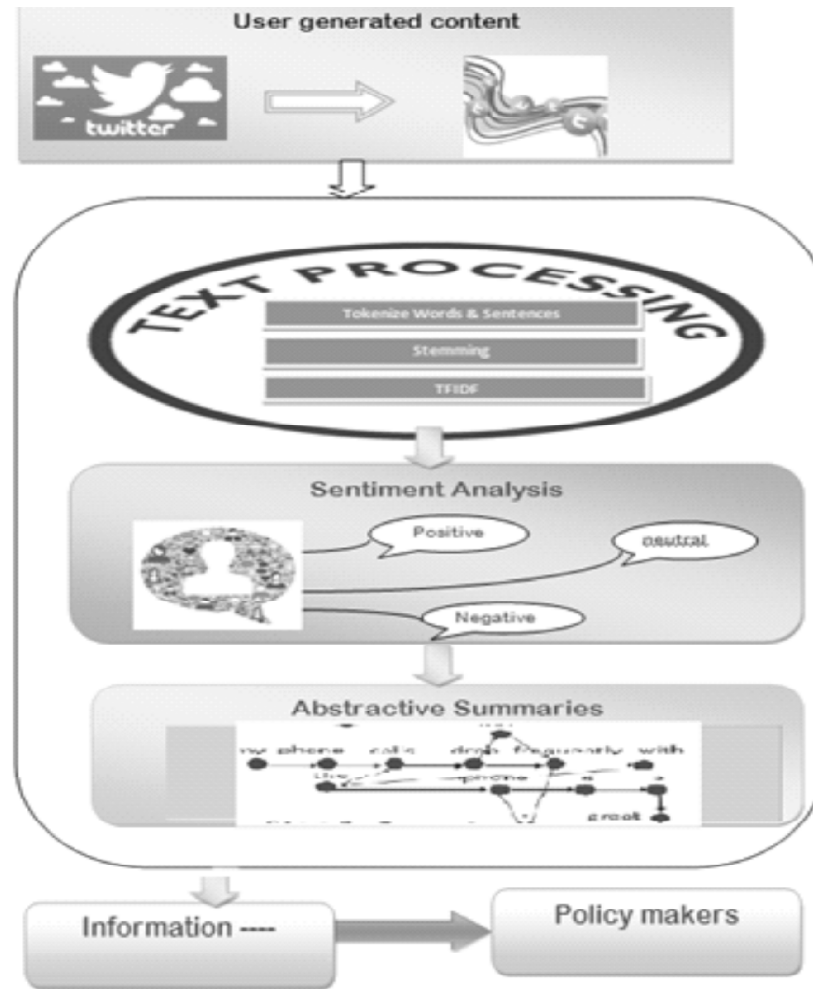


Figure 1

In this phase, we automatically extract sets of relevant positive, negative and neutral tweets.

Phase 3: Abstractive summaries Phase.

The aim of this phase is to focus on the visualization of summaries. These summaries need to be viewed on smaller screens like on a PDA. An informative and concise abstractive summary would be a better solution for the end users to take required measures.[3]

3. DIFFERENT MODULES

3.1. Corpora Acquisition phase: The twitter corpus

For the development of our application, we used a corpus of on the fly tweets. The Twitter Streaming API provides a continuous stream of public information from Twitter. These APIs use the push strategy for data retrieval. Once a request for information is made, the Streaming APIs provide a continuous stream of updates with no further input from the user. In order to get connected to streaming API we need to keep the persistent connection open. Using the method Create Streaming Connection, we can create a POST request to the API and fetch the search results as a stream. Twitter APIs can be accessed only via authenticated requests. Twitter uses Open Authentication and each request must be signed with valid Twitter user credentials. [4]

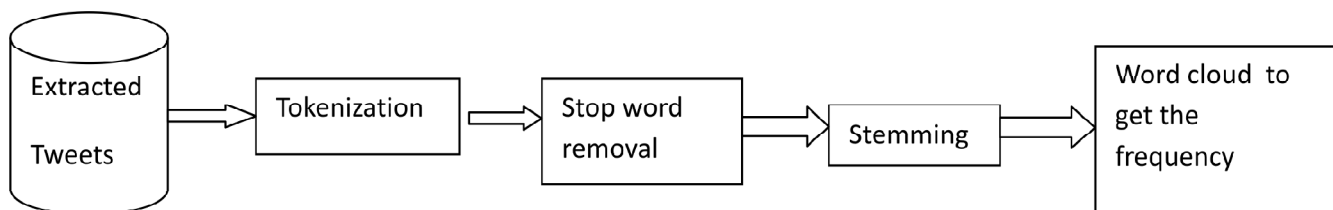
This process was used to extract tweets from twitter related to the schemes launched by the government in India.

We have different categories under which these schemes fall like pro poor, pro youth, pro development, pro farmer, pro women, pro senior citizen etc. Tweets were collected regarding the schemes. The current on the fly tweets are being collected and updated as soon as there are any active conversations regarding the particular schemes. According to a variety of relevant criteria (Make in India, Digital India and so on).

The tweets were collected in text format and then stored. The corpus contains the related tweets pertaining to the policy selected by the user.

The corpus contains not only the tweets themselves, but also a large amount of metadata associated with each tweet, such as its date and time, the number of followers of the person tweeting, the location and other information about the person tweeting and so on[5]. This information is useful for other predicting activities like which location of people are for the scheme and which location of people are against the scheme etc. The tweets are loaded in Scheme Critic, with the text and some of the metadata is highlighted.

Preprocessing of text is being done in the following manner



The extracted tweets from the storage have to pass these stages

Tokenization: The process of breaking a stream of text into words, phrases, symbols, or other meaningful elements called tokens. It is the process intended for exploration of the words in a sentence.

Stop Word Removal

The words which persist very frequently and that are essentially meaningless as they act as connectors for the sentences are called stop words. It is commonly understood that stop words do not contribute to the context or content of textual documents. Their frequency of occurrence is usually very high and they usually stand as the main obstacles in the information retrieval process. They prevent the exact content from being understood by the user. So stop words should be identified and must be removed[6].

Stemming

The process of reducing words to their root. The rule-based replacement of word suffixes is applied iteratively intending to reduce the length of the words until a minimum length is reached. This is a widely used procedure in text processing for information retrieval (IR) based on the assumption that posing a query with the stem term results in retrieval of all the relevant non stem terms.

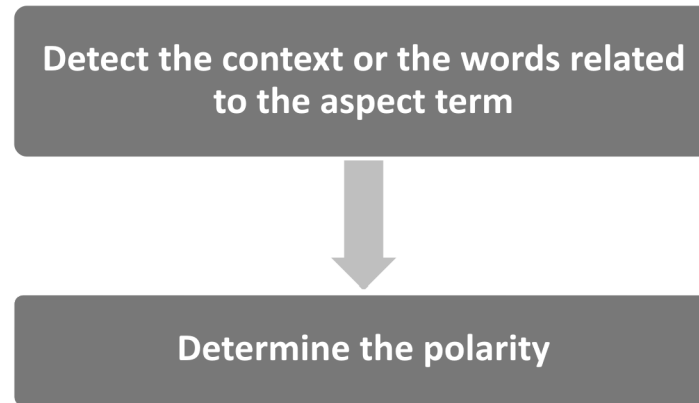
After this step we calculated the term frequency and inverse document frequency of the particular word in the document.[7]

Word cloud

Word cloud can be created to visualize word frequency of the text. It lets us find out the buzzing topic of the day on the particular policy.

3.2. Sentiment Analysis Phase

In this phase, we automatically extract sets of relevant positive and negative adjectives of the framework we annotate the sentiment orientation of the identified opinion phrases in order to assign them with a



positive or negative polarity label depending on their opinion holders underlying stance against the issues they discuss.

To identify how citizens evaluate governmental decisions, we rely on the notion of words semantic orientation and we try to discriminate between words of positive and negative sentiments as follows.

1. **Detect the context:** The context or the words related to the aspect terms are determined and then its polarity is computed according to these words. Dependency parsing or simple distance functions could be used to determine these words[8].
2. **Determine the polarity:** A Multinomial Naive-Bayes algorithm will be used to determine which among them could be positive, negative or neutral. We propose to use for learning a classifier based on different features:
 - The terms frequency
 - The Parts Of Speech features frequency
 - The pre-polarity features
 - Z-score features (the number of words which have Z-score more than three in each sentiment class)

Finally, the system aggregates the ratings of all sentiments expressed about the candidates of an aspect to estimate its rating. The aspects with positive orientation could be allowed to pass through the next level.

Based on the output of the above analysis, we may not only capture the citizen's viewpoints on concerned policy issues but with the proper tools and techniques we may also be able to build predictive models about how citizens assess public sector regulations.

3.3. Visualization of Summaries

Summaries are very informative but still harder for the user to understand why the topic received a particular rating.

The direct unstructured extraction would get all the tweets forcing users to read many often highly redundant sentences about schemes.

Therefore it is desirable to generate concise textual summary of redundant opinions.

Example:

1. Make in India drive catches momentum
2. Make In India Super Hit
3. Make in India is intended to make India a manufacturing hub of the world.

Sample abstractive summary from the above three tweets:

“Make in India drives catches momentum and it is regarded as a big hit intended for making India a manufacturing hub of the world.”

When the end user wants to view the summaries on smaller screens like on a PDA, an informative and concise abstractive summary would be a better solution. This is the main motive of going for abstractive summaries.

A graph data structure is used to produce abstractive summaries of highly redundant opinions. The key idea is to construct a textual graph that represents the text to be summarized. Then, three unique properties of this graph are used to explore and score various sub paths that help in generating candidate abstractive summaries [9].

Our graph data structure is different in that each word unit is regarded as a node and the structure of sentences are regarded as directed edges.

The algorithm used here is capable of finding out

- The redundant discussions
- Similar sequences with minor variations
- Nodes that look like hubs can be compressible

So finally we can view the summaries which will be of major help to the end users.


4. Demo

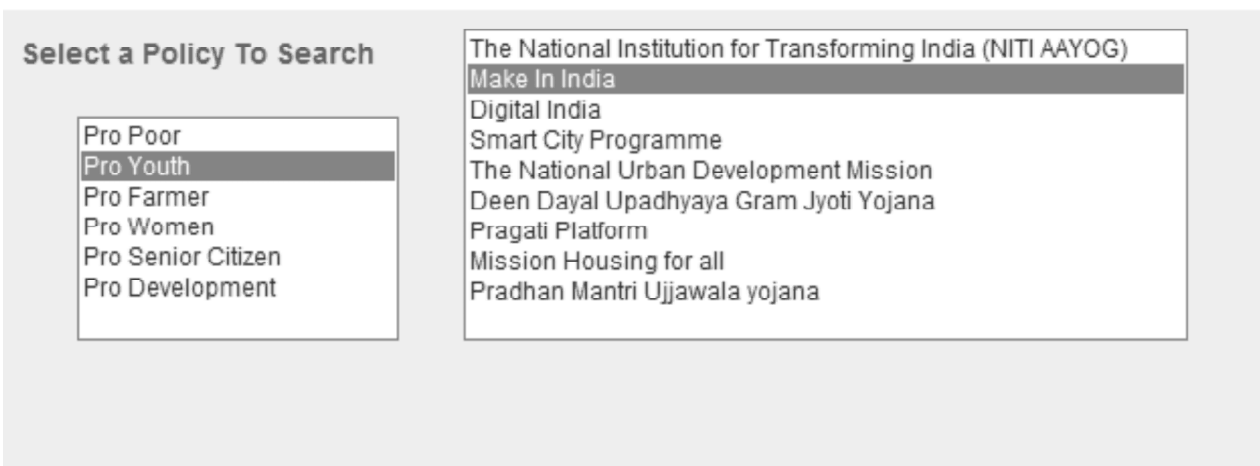
We will demonstrate our system in the context of schemes launched by the government. We will demonstrate the following features of Scheme critic:

The system is being divided into 4 modules

1. Policy search
2. Crowds Talk
3. Policy Summary
4. Sentiment

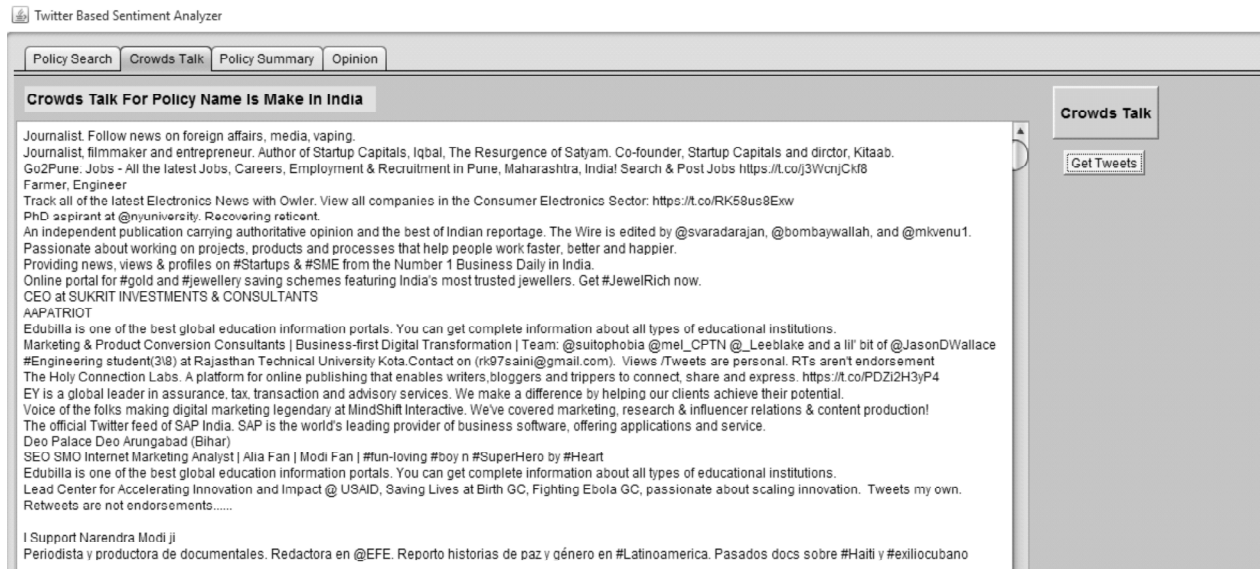
Scheme Critic enables users to select a category from the given number of categories.

 Twitter Analyzer

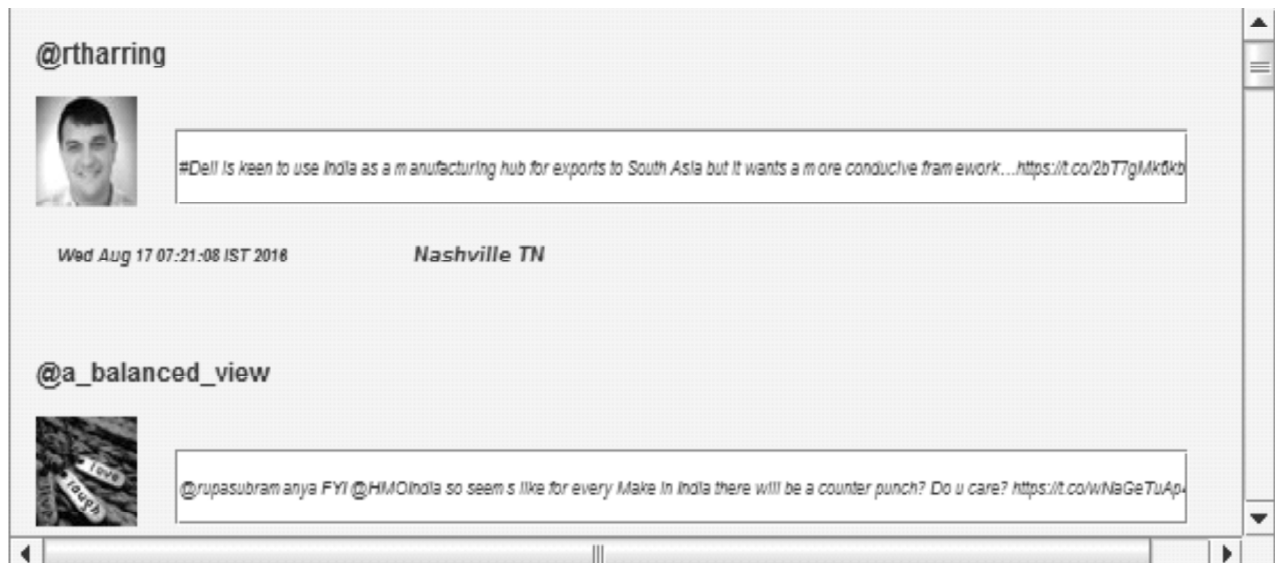


The screenshot displays the 'Twitter Analyzer' interface. On the left, under the heading 'Select a Policy To Search', there is a list of policy categories: Pro Poor, Pro Youth (highlighted), Pro Farmer, Pro Women, Pro Senior Citizen, and Pro Development. On the right, a list of specific policies is shown: The National Institution for Transforming India (NITI AAYOG), Make In India (highlighted), Digital India, Smart City Programme, The National Urban Development Mission, Deen Dayal Upadhyaya Gram Jyoti Yojana, Pragati Platform, Mission Housing for all, and Pradhan Mantri Ujjawala yojana.

Based on the users selection crowds talk about that policy is extracted. The tweets on the scheme “Make in India” are shown here



The tweets posted by which user and the date and time settings as well as the demographic location of the user will be extracted as shown. This gives us a clarity of



In order to find out the most trending topic or the topic which was discussed number of times by many people if we click on the word cloud button it shows the words and how they are associated along with their frequencies.

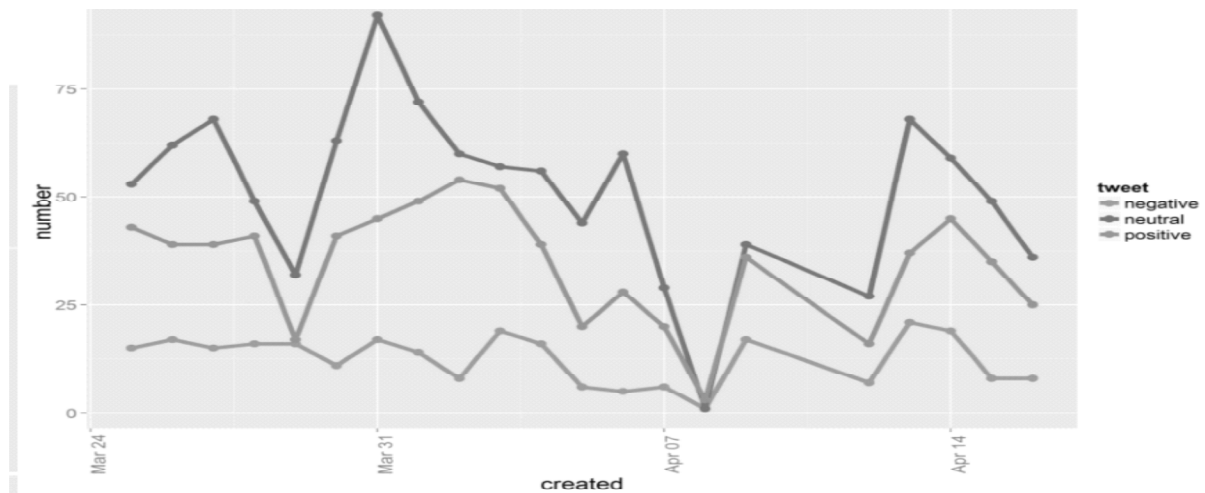
The sentiment of the collective opinion is being done using the algorithm said above

The opinions which are negative ,neutral and positive will be represented as a graph

Sentiment analysis is an excellent source of information about audience behavior and can provide us many insights to form a proper and effective marketing strategy. It allows you to measure what matters the most for your business. We can get a sample graph in this manner to allow the end user a comparative analysis.



Figure 3: Word cloud on Make in India



The final stage is preparing the summarized version of that particular scheme.

5. CONCLUSION

Scheme Critic is a system for performing real-time opinion mining about schemes launched by government over the enormous wealth of user-generated content being currently produced. This system differs from most previous opinion mining projects in multiple aspects. It aims at detecting opinions about specific schemes regarding an open set of topics or issues. It mainly concentrates on investigating highly reactive and short-lived user generated content such as tweets, which represent a great confront due to their unstructured nature. The output from the analysis helps us to develop your transformational social strategy roadmap giving the opinion on the schemes. The demo system gives a better experience to the policy makers. As they will be knowing the pulse of the people in a more abstractive manner which not only save time but also gives better user experience.

REFERENCES

- [1] Motaz Khorshid, Model-centered government decision support systems for socioeconomic development in the Arab world, The International Conference on Input-Output and General Equilibrium: Data, Modeling and Policy analysis September 2-4, 2004 Brussels, Belgium.
- [2] Stylios, G, Christodoulakis, D, Besharat, J, Vonitsanou, M, Kotrotsos, I, Koumpouri, A and Stamou, S. "Public Opinion Mining for Governmental Decisions" Electronic Journal of e-Government Volume 8 Issue 2 2010, (pp203-214).
- [3] Ali Harb, Gerard Dray, Mathieu Roche, François Troussset, Pascal Poncelet, Web Opinion Mining: How to extract opinions from blogs? "CSTST'08: International Conference on Soft Computing as Transdisciplinary Science and Technology, ", lirmm-00329525, version 1 - 11 Oct 2008.
- [4] Haritha Akkineni, P.V.S. Lakshmi, B. Vijay Babu, G. Lakshmi, Modeling and Visualizing the Extraction of Opinions from Twitter, International Journal of Innovations & Advancement in Computer Science, IJIACS, ISSN 2347 – 8616 Volume 5, Issue 2, February 2016.
- [5] Diana Maynard and Adam Funk, Automatic detection of political opinions in Tweets, EU FP7 project ARCOMEM, Chapter The Semantic Web: ESWC 2011 Workshops Volume 7117 of the series Lecture Notes in Computer Science pp 88-99.
- [6] Dr. S. Kannan, Vairaprakash Gurusamy, Preprocessing Techniques for Text Mining, <https://www.researchgate.net/publication/273127322>, Retrieved on: 02 August 2016
- [7] Jorge Carrillo de Albornoz, Laura Plaza, Pablo Gervás, and Alberto D'áz, A Joint Model of Feature Mining and Sentiment Analysis for Product Review Rating, P. Clough et al. (Eds.): ECIR 2011, LNCS 6611, pp. 55–66, 2011. Springer.
- [8] S. Moghaddam and M. Ester, "Opinion digger: an unsupervised opinion miner from unstructured product reviews". Proceedings of the 19th CIKM, pp. 1825–1828, Toronto, ON, 2010.
- [9] Kavita Ganesan and Cheng Xiang Zhai and Jiawei Han Department of Computer Science University of Illinois at Urbana-Champaign Opinions: A Graph-Based Approach to Abstractive Summarization of Highly Redundant Opinions, Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010), pages 340–348, Beijing, August 2010.