



## Google Chrome Extension for Topic Summarization

Shardul Katare<sup>1</sup>, Saurabh Bawdhankar<sup>2</sup>, Sourab Patil<sup>3</sup>, Aniket Deshpande<sup>4</sup>, Prof. Shrikat Nagure<sup>5</sup>  
Savitribai Phule Pune University, B.E. Department of Computer Engineering, RMD Sinhgad School of Engineering,  
Warje, Pune, Maharashtra, India

**Abstract-** Recent studies indicate that comments are an important part of any information site. They give an overall view of the topic from the people's perspective. People or customers response on a particular product or information would be very useful for facilitating various surveys on a particular product or information. The information gathered can be used for various marketing strategies, statistical analysis and research surveys. There are not many software that are available for summarizing topics online. The proposed Chrome Extension would give the overall summary of the topic under discussion by scanning the comments and performing sentence extraction and sentimental analysis. The domain chosen is Google Chrome Browser as it is one of the most popular Web Browsers. Also, it is very easy to develop Google extension as most of the instructions, software and packages for developing an extension are readily available on internet

**Keywords-** Web Browser Extension, Data Mining, Software Libraries, Languages, Pattern, Sentiment analysis..

### I. INTRODUCTION

The Chrome extension will summarize the information of mobiles and gadgets and any other information taking Tech Crunch blog as the data set. Extensions are software programs that modify and improve the functionality of Chrome browser. Extensions are known as Add-ons in Internet Explorer and Mozilla Firefox. The proposed Chrome extension will be written using Web Technologies such as Java Scripts, HTML, and CSS. Extensions will provide functionality without diving deeply into native code.

The extension will provide a summary about the topic which is taken as input from the site. It will also include the summarization of comments as comments are an integral part of the topic. The benefit of this Chrome Extension is that it will be useful for reader's feedback as well as the sentiment analysis.

Given a topic on any information site which is under discussion and has received any comments, our model consists of three modules – sentence detection which will split the topic post into sentences, word weighing measures weighs the words appearing in comments and sentence selector computes the representative score for each sentence based on the frequency of occurrence of its contained words, the model will take websites which have Facebook comments associated with the topic, the benefit of using Facebook comments is that, the comments under the topic will not be spammed by the malicious users and any abusive content can be reported. The Facebook APIs are readily available and can be used for accessing the comments.

### II. LITERATURE SURVEY

Data from the Comments oriented Blog Summarization Blog summarization CIKM 07:

Based on previous studies, it has been observed that comments do affect one's understanding about any topic. The solution measured word representativeness using information hidden in comments, and then selects sentences based on the representativeness contained in the sentences. ReQuT model gives the flexibility to measure word representativeness to three aspects which are Reader, Quotation and Topic.

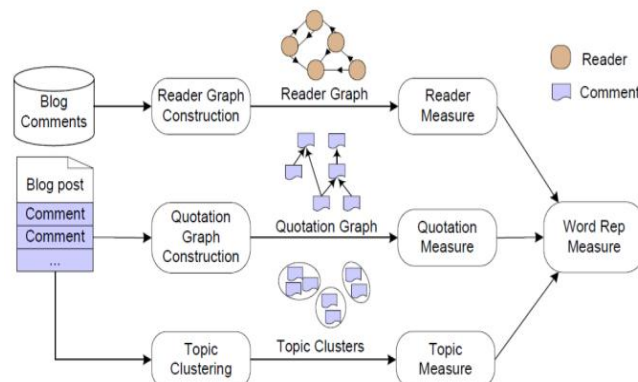


Figure 2: ReQuT Model

Sentiment Analysis is the use of natural processing, text analysis along with computational linguistic to identify and extract subjective information in source material in general sentiment analysis gives the attitude of a speaker or a writer with respect to some topic or the overall summary of the document.

The existing software's only analyze the data provided but not the comments below it as this information can be exaggerated by the author only to convince the readers. From the recent studies it is found that users read other comments for response of other people but it is a very lengthy and time consuming task because many sites have more than hundred of comments.

### III. PROPOSED SYSTEM

#### 3.1 Disadvantages of Existing software:

- There are not many software which summarize the topic along with its comments.
- Blog summarization only focuses on its data provided by the author but not the comments which are given by the users.
- It is not readily available on Chrome extension even though it is one of the most used Browsers.

#### 3.2 Proposed System Introduction:

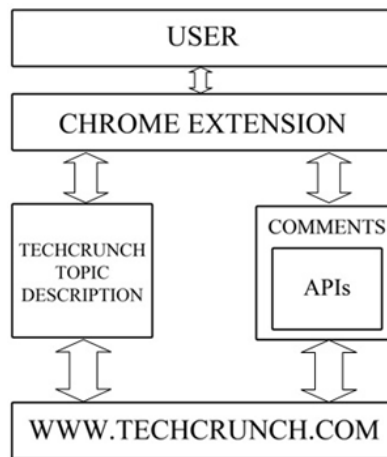
##### 3.2.1 Modules:

**Time Limit Module** - This module enables the user to get the summary in stipulated amount of time.

**Topic Module** - This module scans all the information given by the author or the publisher in the Topic.

**Sentence Detection Module** - This module divides all the sentences in the topic.

##### 3.2.2 System Architecture



##### 3.2.3 Algorithm

1. Start
2. Open Google Chrome Browser
3. Open TechCrunch.com
4. Search for the Topic on the website
5. Open the topic
6. Open the Extension in browser
7. If (comments not found) "no comments available" If (comments found) "Open Chrome Extension"
8. Click on "Get Summary" button
9. Summary will be displayed.
10. "Download as PDF" button will download the summary in PDF format

##### 3.2.4 Feasibility study:

P type problem -

If the running time is some polynomial function of the size of the input, for instance if the algorithm runs in the linear time or cubic time, then we say the algorithm runs in polynomial time and the problem it solves is in class P[3] The P(polynomial time) class problems can be solved using inputs that are traceable and are easy to solve.

NP type problem -

A problem is assigned to the NP (nondeterministic polynomial time) class if it is solvable in polynomial time by a nondeterministic Turing machine. A is always also NP.[4]

Taking into consideration the above mentioned points and after analysis of our project algorithm it can be inferred that the problem is a P-type problem.

##### 3.2.5 Mathematical Model:

1. Let  $S$  be a system that describes the execution of the Extension.  
 $S = \{ \dots \}$
2. Identify the modules as  $M$

$S = \{M \dots\}$

Where,

$M = \{E, R\}$

E = selection of topic given by author.

R = selection of comments under the topic.

3. Identify the modules of R as Mr

$Mr = \{Tl, To, FC, SD\}$

Where,

Tl= Time limit module.

To= Topic module.

SD=Sentence module.

FC=Facebook comments module..

[A] Input to Tl is T\_limit

Where,

T\_limit = Time limit specified for the module.

[B] Input to is topic given by the author

4. Identify the Processes as P

$S = \{M, P \dots\}$

$P = \{Pt, Pr, Pl, Pcl, P\_rep, Pe\}$

Where,

Pt = Process of evaluating time.

Pr = the reader's authority is calculated as the number of distinct user that post a reply to the users comment.

Pl = number of likes on the comment

Pcl = Algorithm that clusters the comments and a weight is assigned to a cluster

P\_rep = the number of replies a comment has got = Rep (Ci)

Pe = All the named entities of topics are identified.

5. Identify the output as O.

$S = \{M, P, O\dots\}$

$O = \{Os, Op\}$

Where,

Os = Output is summary of the topic and comments

Ow = Download as PDF option is available.

6. Identify the success as Su.

$S = \{M, P, O, Su\dots\}$

Where,

Su =Success is when the summary is successfully displayed in the extension

7. Identify the failure as F.

$S = \{M, P, O, Su, F\dots\}$

Where,

F = When nothing is displayed in the extension.

#### IV. FEATURES

- The Extension has simple and easy GUI for non-technical users.
- Extract valuable information from the thousands of comments.
- Get the overall idea of the product and opinions from the customers.
- To extract the sentiments from comments.
- “Download as PDF” will download the extracted information in the PDF format for records

#### 3.2.5 Advantages of the proposed System:

- The extension can be installed on a PC or Laptop with Google Chrome Browser and used to summarize the Topics on techcrunch.com website.
- The Plugin will reduce the time and efforts needed to read comments of trending topics.
- The Plugin will provide the overall description of the topic along the with People’s opinion which most blogs fail to consider.
- The Plugin has a very a user-friendly interface that makes it possible to be used by inexperienced users as well.
- The source code can be further modified to extend to other browsers like Mozilla Firefox, Internet explorer and Opera.
- The source code can be modified further so that it can gather user comment’s summary from multiple websites about the same product or information topic.

#### V. CONCLUSION

Our proposed solution gives a good score for long blog posts and sufficient number of comments associated with it. It helps to easily read the summary from add-on than to read all the comments..

## VI. FUTURE SCOPE

The Extension can also be extended to other web browsers such as Mozilla Firefox and Internet Explorer. The number of Blog Sites or Web Sites under the domain can be increased to benefit greater number of users by modifying the source code.

## REFERENCES

- [1] Google Chrome official website <https://www.google.com/chrome/>
- [2] Google Chrome Extensions <https://chrome.google.com/webstore/category/extensions>
- [3] Feasibility Study - P type problems <https://www.quora.com/What-are-P-NP-NP-complete-and-NP-hard>
- [4] Feasibility Study - NP type problems <http://mathworld.wolfram.com/NP-Problem.html>
- [5] TechCrunch website as dataset <http://techcrunch.com/>
- [6] Sentiment analysis [https://en.wikipedia.org/wiki/Sentiment\\_analysis](https://en.wikipedia.org/wiki/Sentiment_analysis)
- [7] Comment oriented blog summarization *CIKM'07*, November 6–8, 2007, Lisboa, Portugal. Copyright 2007 ACM 978-1-59593-803-9/07/0011
- [8] “Mining the User Clusters on Facebook Fan Pages Based on Topic and Sentiment,” *IEEE IRI 2014*, August 13-15, 2014, San Francisco, California, USA.
- [9] Google Chrome Developer <https://developer.chrome.com/extensions/getstarted>
- [10] Make Google Chrome extension <https://chrome.google.com/webstore/category/extensions>
- [11] Natural language processing [https://en.wikipedia.org/wiki/Natural\\_language\\_processing](https://en.wikipedia.org/wiki/Natural_language_processing)
- [12] Web mining <http://www.scaleunlimited.com/about/web-mining/>

## ABOUT AUTHORS



**Shardul Katare** is currently pursuing the degree of B.E in Computer Engineering from the RMD Sinhgad School of Engineering which is affiliated to the SavitribaiPhule Pune University (formerly The University of Pune)



**Saurabh Bawdhankar** is currently pursuing the degree of B.E in Computer Engineering from the RMD Sinhgad School of Engineering which is affiliated to the SavitribaiPhule Pune University (formerly The University of Pune)



**Sourab Patil** is currently pursuing the degree of B.E in Computer Engineering from the RMD Sinhgad School of Engineering which is affiliated to the SavitribaiPhule Pune University (formerly The University of Pune)



**Aniket Deshpande** is currently pursuing the degree of B.E in Computer Engineering from the RMD Sinhgad School of Engineering which is affiliated to the SavitribaiPhule Pune University (formerly The University of Pune)



**Prof. Shrikant Nagure** received the B.E. and M.E. Degree in Information Technology. He now is working with RMDSSOE, Warje, Pune, as Asst. Professor. He has experiences of 06 years and his Area of specialization – Distributed Systems, Open source, Networks