

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 6, Issue 10, October 2017

Survey on Intelligent Data Mining of Social Media for Better Decision Making

Shaikh Ashfak¹, Sonali Ransure¹, Anup Bhosale¹, Subhash Kumar¹, Dr.Arati Dandavate²

Students, Department of Computer Engineering, JSPM's Jayawantrao Sawant College of Engineering, Savitribai Phule
Pune University. Pune, India¹

Professor, Department of Computer Engineering, JSPM's Jayawantrao Sawant College of Engineering, Savitribai Phule
Pune University, Pune, India²

ABSTRACT: The advent of social media and the rapid development of mobile communication technologies have dramatically changed the way to express the feeling, attitude, mood, passion etc. People often express their reaction, fancies and predilections through social media by means of short texts of epigrammatic nature rather than writing long text. Many social websites like Twitter, Google Review, Just Dial, Book my show, etc enables people to share and discuss their thoughts, opinion and view in the form of short text, which can be useful for other unknown peoples, customers and service users to decide whether that service or product is good or not. In this paper, whole process is divided into two step. In first step through intelligent data mining data will be abstracted and in second step analysis frame work will be there, which will focuses on positive and negative opinion and through R-Programming the visualization will be created will be helpful for peoples to make decision on their subject (i.e. on institute, services, products, movies, tourist spot etc). In visualization section it will contain graph, pictures, etc. Comparison parameter can also be implemented which will be again very much helpful for user and peoples.

KEYWORDS- co-extracting algorithm, co-extracting model, opinion targets, Opinion Relation Graph, opinion words, Topical Word Trigger Model

I. INTRODUCTION

If any customer or service user wants to use or try some new or unknown product or services then before buying that product they are checking in various no of websites for review or opinion on the same. And in today's era of digitalization, internet and social websites it become easy to get opinion and review, but in unorganized form.

In the first stage of our current work, we employ exploratory analysis using the Self Organizing Maps to assess correlations between user posts and positive or negative opinion on the query (i.e. service, product, etc.). In a second stage, we model the users and their posts using a network-based approach. Conduct a manual qualitative analysis of a large sample of software-relevant data to determine the information value of software users' posts. Use text classification techniques to effectively capture and categorize the various types of actionable software maintenance requests present in such posts. Investigate the performance of various text summarization techniques in generating compact summaries of the common technical concerns raised in software systems' Twitter feeds. Our main objective is to lay down an infrastructure for a more responsive and a more adaptive software engineering process that can achieve user satisfaction in an effective and a timely manner.

Problem Statement:

If any customer or service user wants to use or try some new or unknown product or services then before buying that product they are checking in various no of websites for review or opinion on the same. And in today's era of digitalization, internet and social websites it become easy to get opinion and review, but in unorganized form. So therefore, this project will do the proper work of organization, easy access of review and opinion from different websites.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 6, Issue 10, October 2017

Objectives:

1. To perform intelligent data mining on social media.
2. To reduce the efforts of customer and common peoples of searching in different websites.
3. To provide all the opinion and review in organized form.
4. To conserve the time.
5. To present all the data at one webpage.

II. LITERATURE SURVEY

1. "A new approach towards co-extracting opinion- targets and opinion words from online reviews". 2017.Saru, Mamta Bhusry, Ketki. Efficient extraction of data can be done.
2. Mining Twitter Data for a More Responsive Software Engineering Process.2017. Grant Williams* and Anas Mahmoud. Better analyzing can be done.
- 3."Network-Based Modeling and Intelligent Data Mining of Social Media for Improving Care".2013. A. Akay, Member, IEEE, A. Dragomir, Member, IEEE, Björn-Erik Erlandsson, Senior Member, IEEE Grant Williams* and Anas Mahmoud. Effective treatments can derive from the opinion of large patient and doctors. Lack of advancement in technique of mining, sorting data and algorithms.
4. Artificial Societies and Social Simulation Using Ant Colony, Particle Swarm Optimization and Cultural Algorithms. A. Ochoa, A. Hernandez, L. Cruz, J. Ponce, F. Montes, L. Li, and L. Janacek.2010. It Proposes Ant Colony System Algorithm. It Improves good quality in a short time. It has better performance. Community of agents is not in application.
5. Collective Extraction for Opinion Targets and Opinion Words from Online Reviews. Xiangxiang Jiang*, Yuming Lin*, You Li†, Jingwei Zhang*. 2016. Proposes a method to extract opinion targets and opinion words collectively based on the word alignment model. Higher accuracy. Effectively ignore the problem of error propagation. Greatly reduce the work of manually labeling samples. It does not apply parallel extraction method.
6. Tracing Information Flow and Analyzing the Effects of Incomplete Data in Social Media. Gundeep Singh Bindra, Krishen Kant Kandwal, Prashant Kumar Singh, Seema Khanna. 2012. Proposes a k-tree model of cascades is generated from a balanced tree of height and branching factor. The information flow is unambiguous and precise. We can have the time, so it's easy to trace the information. In a very large network, it becomes easy to collect data. However, if data is incomplete cascades break into pieces. Many different diffusion mechanisms.. Not all the links transmits the information. Sometimes the links get missing due to Blogger forget to attach a link or mainstream media does not provide the source links. Not clear whether hashtags really diffuse. Due to "personalization" easier to argue URLs diffuse. Problem with all is that we do not know the "influencer".
7. Text and Structural Data Mining of Influenza Mentions in Web and Social Media. Courtney D. Corley 1,*, Diane J. Cook 2, Armin R. Mikler 3 and Karan P. Singh 4. Proposes a graph-based data mining technique. To identify trends in flu posts that correlate to real-world ILI patient reporting data. Content analysis does not provide.
8. TEXT MINING: PROMISES AND CHALLENGES. Ah-Hwee Tan. 1999. Proposes a text mining framework consists of two frameworks: Text refining and Knowledge distillation. Customer profile analysis, Patent analysis, Information dissemination, Company resource planning. There are issues in this paper, semantic analysis, Multilingual text refining, domain knowledge integration and personalized autonomous mining.
9. Social media competitive analysis and text mining: A case study in the pizza industry. Wu Hea,*, Shenghua Zhab,1, Ling Li a,c,2. 2013. Proposes competitive analysis for the user-generated data on Twitter and Facebook in three major pizza chains.Establishing Effective and realistic benchmarks. Mining the content of social media conversations.Does not track real-time data.
10. Mood Based Classification of Music by Analyzing Lyrical Data Using Text Mining. Nirbhay Kashyap1, Tanupriya Choudhury2, Dev kumar Chaudhary3,Roshan Lal 4. 2016. Proposes classification using Support Vector Machine algorithm:. The framework may be utilized to hunt down female craftsmen, content melodies, or hallucinogenic music. Content features may prompt higher correctness's for most mood classifications.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 6, Issue 10, October 2017

11. Text Mining for the Hotel Industry. Kin-Nam Lau, Kam-Hon Lee and Ying Ho. 2005. Proposes online text mining with analysis. Reduce the use of manual labor in identification, storage, and analysis of business intelligence. Fully automated system. Does not integrate text-mining tools with related technologies such as image recognition and Web-table mining.
12. Feature extraction and classification of proteomics data using stationary wavelet transform and naïve Bayes classifier. Liu Dan, Huang Yuan-yuan, Ma Chen-xiang School of Life Science and Technology, Xi'an Jiaotong University. 2006. Proposes Naïve Bayes Algorithm and stationary wavelet transformation. It requires a small amount of training data to estimate the parameters necessary for classification. High sensitivity, specificity and accuracy.
13. Semantic Data Analysis Algorithms Supporting Decision-making Processes. Lidia Ogiela, Marek R. Ogiela. 2015. Proposes semantic data analysis. Cognitive systems are very efficient. The structure of information record is too complex to perform full interpretation. The system has not enough knowledge to fully describe the semantic meaning of analyzed information record or complex structure.

III. PROPOSED SYSTEM

The objective of feeling extraction is to recognize where in reports opinion, review or tweets are installed. Opinions are covered up in words, sentences and records. A sentiment sentence is the littlest complete semantic unit from which opinions can be removed. The opinion words, the opinion holders, and the logical data ought to be considered as hints while separating sentiment sentences and deciding their inclinations. Subsequently, the extraction algorithm is developed base by recognizing opinion words at in the first place, then distinguishing the feeling polarities of sentences lastly reports a short time later. Feeling scores of words, which speak to their opinion degrees and polarities. We used the Twitter *Search API*, to collect our dataset of software relevant tweets. In our analysis, we limit our data collection process to tweets addressed directly to the Twitter account of a given software system. To automatically classify our data, we investigate the performance of two text classification algorithms, including Naïve Bayes (NB) and Support Vector Machines (SVM).

IV. SYSTEM ARCHITECTURE

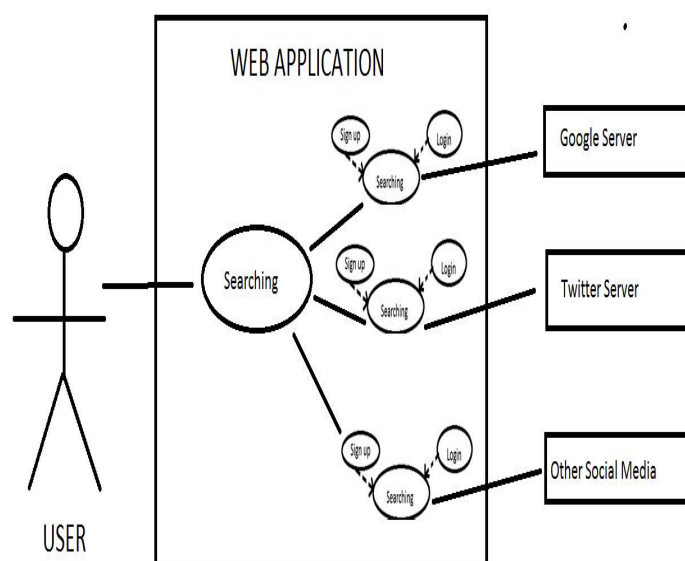


Fig. 01 Proposed system architecture

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 6, Issue 10, October 2017

Applications:

1. Data Scrapper
2. Data Miner

Advantages:

- Less time consumption.
- Easy access.
- Pervasiveness.
- Data from different websites can be found at one place.
- Data scraping only single click button.
- Automatic classification.

Disadvantages:

- Improper implementation of algorithms may effects the overall performance of system.
- Mined review may not be exactly true.
- License or authorization may require extracting data from some of the useful sites.

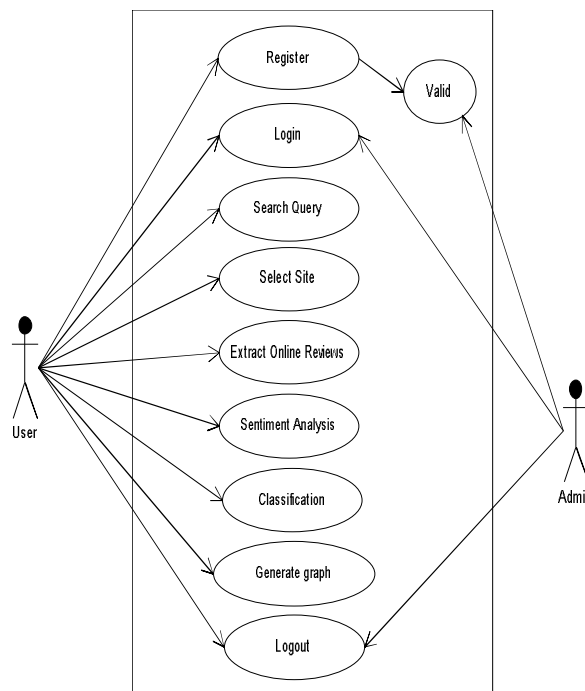
Future Scope:

- It can be made globalize, available all over the World.
- More advancement can be made in project, extracting data from various social media.
- Various classification algorithms can be implemented for better classifying the target words.

V. SYSTEM DESIGN

UML DIAGRAMS

1. USE-CASE DIAGRAM



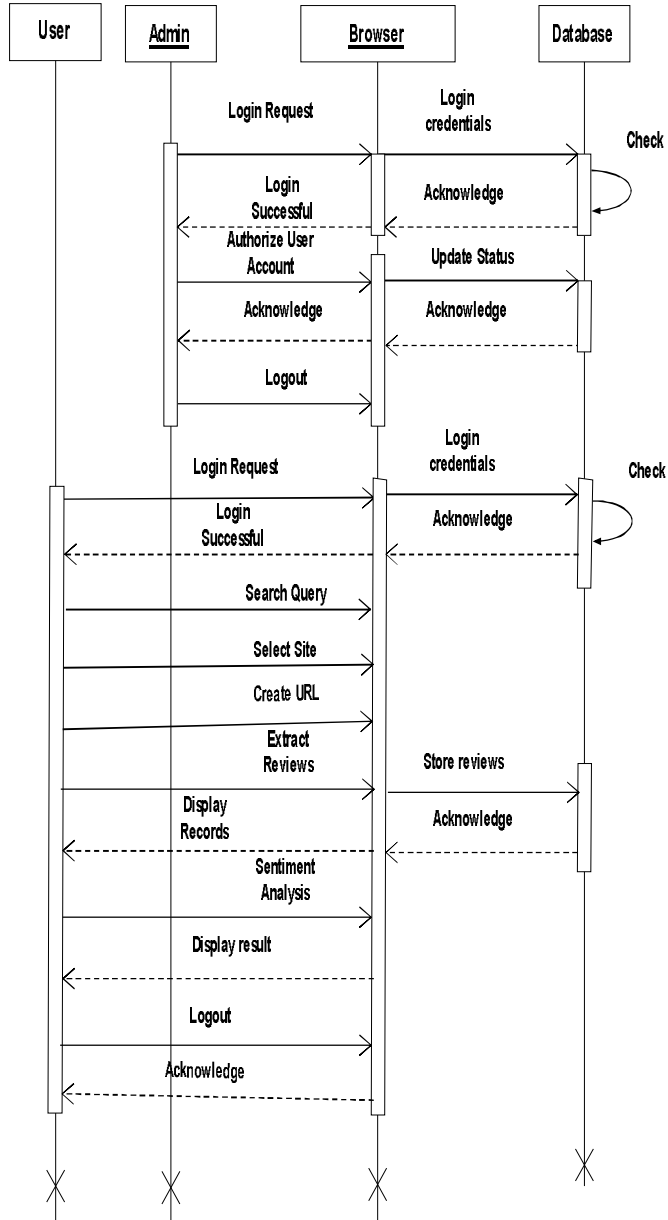
International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 6, Issue 10, October 2017

2. SEQUENCE DIAGRAM



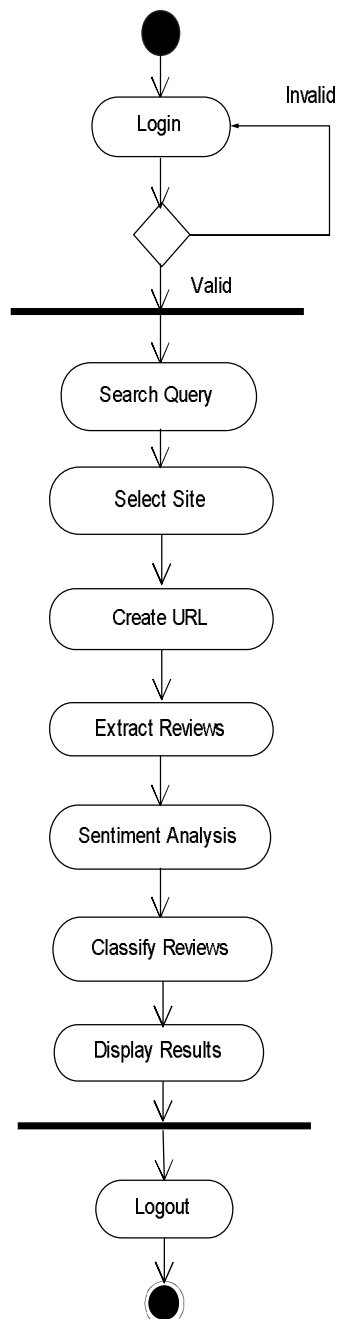
International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 6, Issue 10, October 2017

3. ACTIVITY DIAGRAM



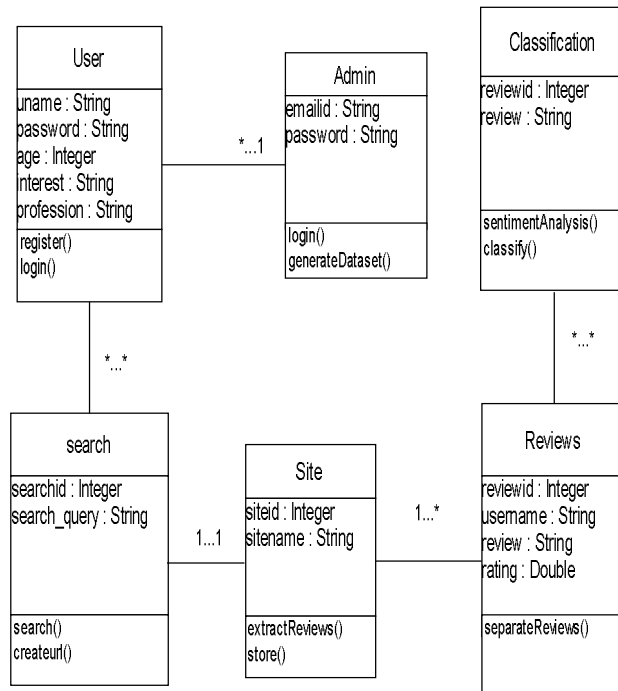
International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

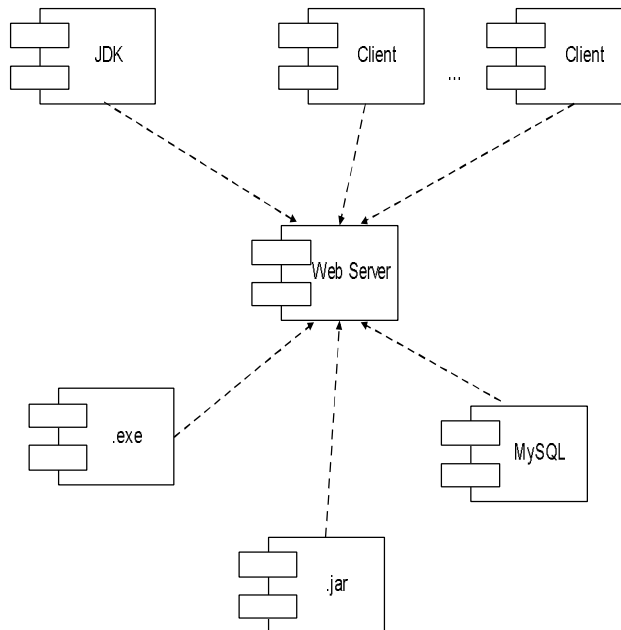
Visit: www.ijirset.com

Vol. 6, Issue 10, October 2017

4. CLASS DIAGRAM



5. COMPONENT DIAGRAM



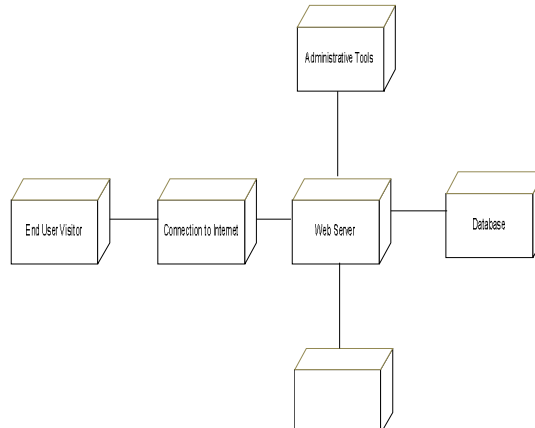
International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

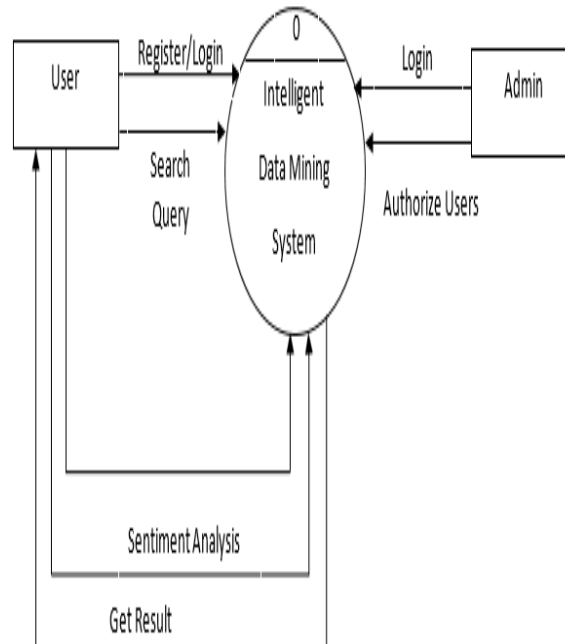
Visit: www.ijirset.com

Vol. 6, Issue 10, October 2017

6. DEPLOYMENT DIAGRAM



DFD Diagrams: DFD Level 0



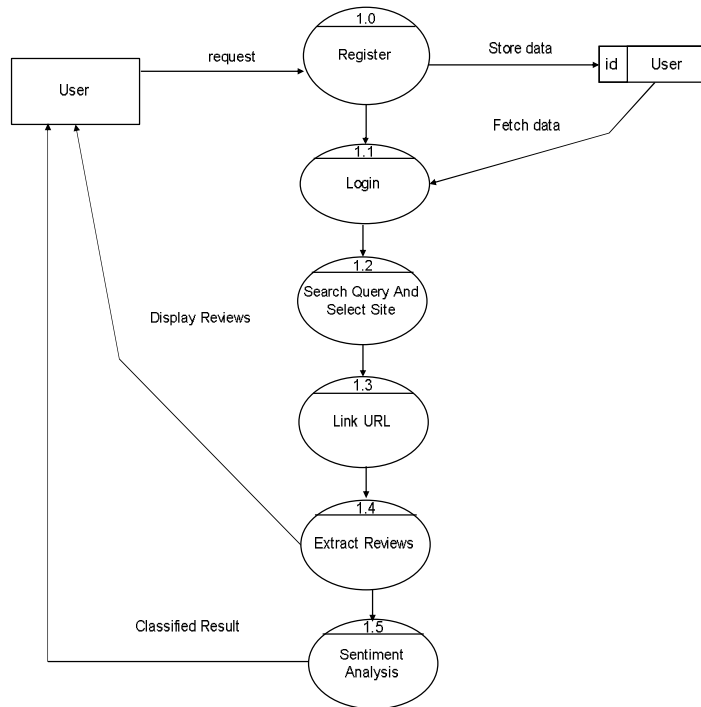
International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

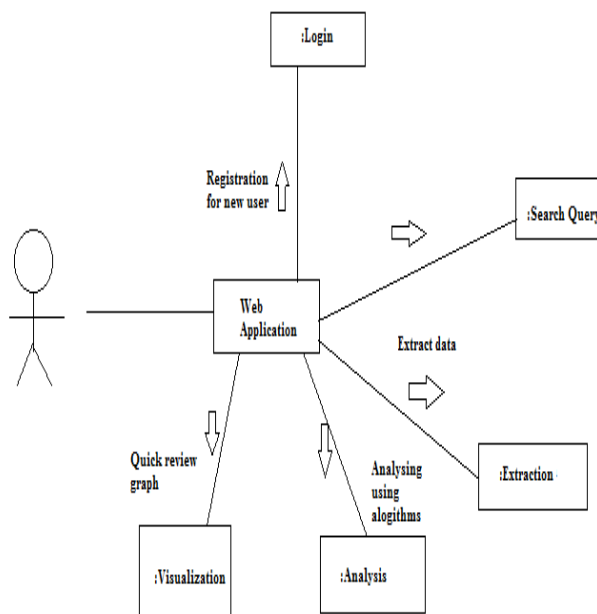
Visit: www.ijirset.com

Vol. 6, Issue 10, October 2017

DFD Level 1 Diagram:



7. Collaboration Diagram:



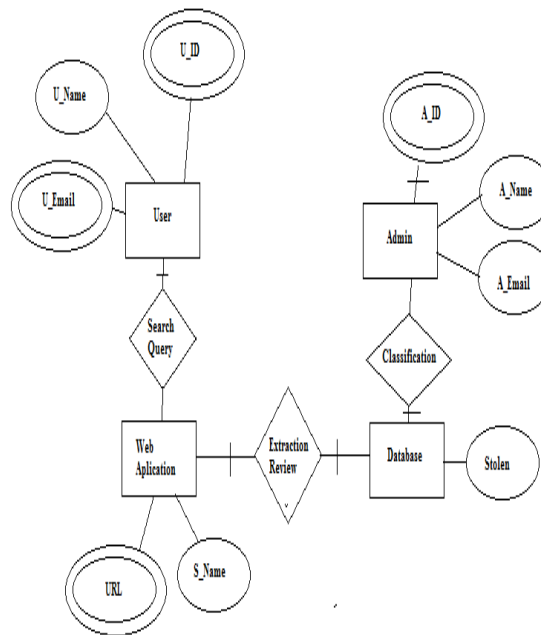
International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 6, Issue 10, October 2017

8. ER Diagram:



VI. CONCLUSION

This project is implement using data scraping technique. After data scraping, the number of posts are getting to user. Apply the sentiment analysis on every post which leads to user decide which one is good. This project reduces the time consumption and the work of searching in many website. This provides an easy access platform for peoples, which help them to take better decisions. Visualization and comparison add more efficiency for decision making.

REFERENCES

1. Artificial Societies and Social Simulation Using Ant Colony, Particle Swarm Optimization and Cultural Algorithms
2. Collective Extraction for Opinion Targets and Opinion Words from Online Reviews
3. Tracing Information Flow and Analyzing the Effects of Incomplete Data in Social Media
4. Text and Structural Data Mining of Influenza Mentions in Web and Social Media
5. TEXT MINING: PROMISES AND CHALLENGES
6. Social media competitive analysis and text mining: A case study in the pizza industry
7. Mood Based Classification of Music by Analyzing Lyrical Data Using Text Mining
8. Text Mining for the Hotel Industry
9. Feature extraction and classification of proteomics data using stationary wavelet transform and naïve Bayes classifier
10. Semantic Data Analysis Algorithms Supporting Decision-making Processes