

## Web Blog Application for Visualization of Movie Reviews Polarity

Mayuri R.Lahamage<sup>1</sup>, Prof.Santosh R.Durugkar<sup>2</sup>

<sup>1</sup>Department of Computer Engineering, SNDCOE &RC, Yeola, [mayu\\_lahamage23@yahoo.in](mailto:mayu_lahamage23@yahoo.in)

<sup>2</sup>Department of Computer Engineering, SNDCOE &RC, Yeola, [santoshdurugkar@gmail.com](mailto:santoshdurugkar@gmail.com)

---

**Abstract-** Web platform is the biggest library for sharing regular diaries of people emotions, experience, aim and opinion throughout the world. The same criteria use the number of movie industry for their blockbuster and to show the popularity of current releasing movie on user interest. In the proposed system we introduced Architecture, Implementation and Report Visualization for movie review system and we have implement web blog application with three components. First component is Web Blog Crawler: Is a software program use to extract URL of related movie. Second is Sentiment Analyzer: is used to classify movie reviews in bipolar orientation. And third component of this system is Report Visualize: this component to visualize final result of the system to user. This application is use for user entertainment.

**Keywords-** Blog Mining, Jsoup Library, sentiment Analysis, BIRT Tool, User Interface.

---

### I. INTRODUCTION

Internet is the world biggest library for sharing any kind of data throughout the world. Consumers use to share their views, aim, emotions and their regular diaries with other people via internet is in the form of text, audio, video. So it is very hard grab this kind of data for opinion mining. Basically we design system for movie reviews. In the movie industries there are different review sites for different movies. In this system we use web blog crawler with JSOUP jar for searching for particular movie review blogs. In the past there are number of author used number of methods to classify reviews those reviews are classifying on their sentiment orientation. Turney's [8] work on classification of reviews He applied a specific unsupervised learning technique based on the mutual information between document phrases and the words "excellent" and "poor", where the common information is computed using statistics gathered by a search engine. Zhongchao Fei [2] described the sentiment classification application at the document classification phase, the authors add some labels to certain words in the text, and then match the label within a sentence with predefined phrase patterns to get the sentiment orientation of the sentence under consideration. In this study we have to use Natural Language processing methodology for classification of movie reviews.

### II. PROPOSED SYSTEM

#### 2.1 Problem Defination

As per our survey I had been seen it is hard task to grab people opinion from Internet. People use internet to share their emotions or regular diaries with other people throughout the world. There are number of blogs to show opinion of people related to some product. But it's a difficult task to classify their sentiments in a proper way for better understanding. In the movie industries they use same criteria to show the popularity of current releasing movie, but they got number of difficulties to collect total review in a single site and calculate the popularity of best movie from movie database.

## 2.2 Solution

To explore this problem in our proposed system we have implement Web Blog Application to show result. In our system we introduced component related to this application and calculating average score of total sentiments from movie blogs and display the top most movie from database on their rating out of 10.

### III. WEB BLOG APPLICATION

Web blogs are nothing but web logs. These blogs are used for opinion mining; here we introduced their architecture, implementation and report visualization for user. There are three major components related to this application. First is Web Blog crawler: It is computer program to extract URL, indexed them and stored into the database. Second is Sentiment Analyzer: Used to classify sentiment orientation in bipolar orientation. And last but not the list component is Report visualize: in this component we have implement user interface for report visualization as shown in bellow figure1.

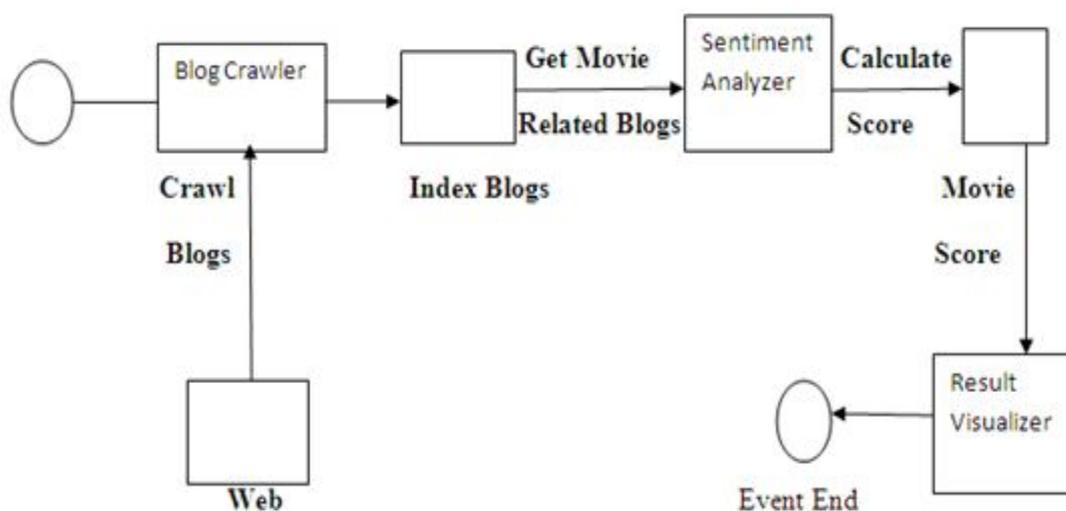


Figure 1. Component diagram of web blog application.

#### 3.1 Web Blog Crawler

Web blog crawler is a tool for automatically gathering web pages. Web crawlers are developed to perform verity of jobs like: download web pages, parse through the downloaded pages and retrieve all the hyperlinks and for each link retrieved, repeat the process. Here we use JSOUP as web crawler, which is use to provides a very convenient API for extracting and manipulating data, using the best of DOM, CSS and jquery-like methods. Crawlers can also be used for automating maintenance tasks on a Web site, such as checking links or validating HTML code.

#### 3.2 Sentiment Analysis

The Sentiment analyzer is a crucial component of the proposed system. It calculates the sentiment scores for a product for different keywords by mining the comments from blogs. In this study, we use an unsupervised approach in sentiment analysis. We use OPEN-NLP to find types of words. In our approach, we use a keyword database, which contains specific words about a movie domain. We use a keyword algorithm, which searches the keywords in the text. If a keyword under consideration is found in the database, then the algorithm identifies whether the keyword is an adjective or an adverb and calculates the score. NLP is restricted to converting graphical language into natural language or visual representation of text processing languages. It also uses a

grouping of language dictionaries, linguistic constructs like parts of speech, and noun phrases along with a range of operators.

### **3.3 Report Visualization**

In this component we implement user interface to show the result. Here we generate report on the movie reviews in bipolar orientation and we display top most movie from our movie database on their calculated rating. In this proposed system we calculate movie reviews and provide rating along with their positive or negative score.

## **IV. IMPLEMENTATION**

In our proposed system we use personal computer with windows, 4GHz processor, and 80 GB hard disk and 250MB RAM, here we used JAVA programming to implement the system.

### **4.1. Dataset**

In our system we use review dataset from [www.greatbong.com](http://www.greatbong.com) web site and for detailed information related to movie, we parse the data from Wikipedia. Here we parse all the data related to the particular movie and stored them into our database, for calculating score we use paraWord list and stored keyword dataset.

### **4.2. Mathematical Model**

The main focus of research work is on the preprocessing stage which involves calculating attribute weights and training the model with given attributes by weight. Using following process we implement our system:

**Crawling-→Data Extraction-→ Data feeding into database-→ Sentiment Analysis + Report Generation**

According this process model we have to calculate movie rating system and analyze the popularity of the given movie on their rating from stored database. This system used in movie industries for their blockbuster and to show the popularity of current releasing movie.

## **V. RESULT**

As per section 4 we calculate sentiment score using given methods, after calculating positive, negative score we calculate total score and then calculate rating for each movie using following pseudo code. In this code I am going to use rating out of 10 , here if rating is less than 0 then there is no rating for movie and if it is less than or equal to 10 then we will see large number of user interest.

```
Function: var rating = row["rating"]
Input: rating in 0 to 10
Output: apply sequence of top most movies from database.
for each(movie){ if(rating <= 0){
    "No Rating";
} else if (rating <= 2){
    "Disappointed";
} else if (rating <= 4){
    "Not a fun";
} else if (rating <= 6){
    "It's okay";
} else if (rating <= 8){
    "Like it";
} else if (rating <= 10){
    "Love it";
}
```

Figure 2. Pseudo code for movie rating

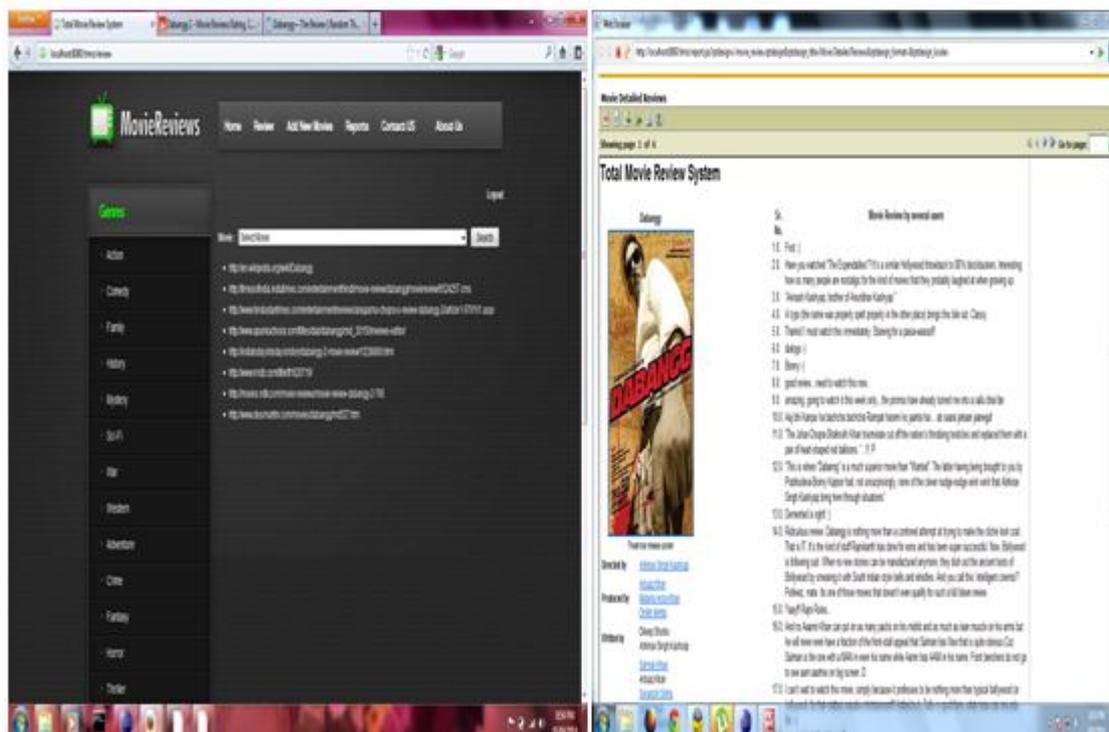


Figure 3. Screen shot for web crawler

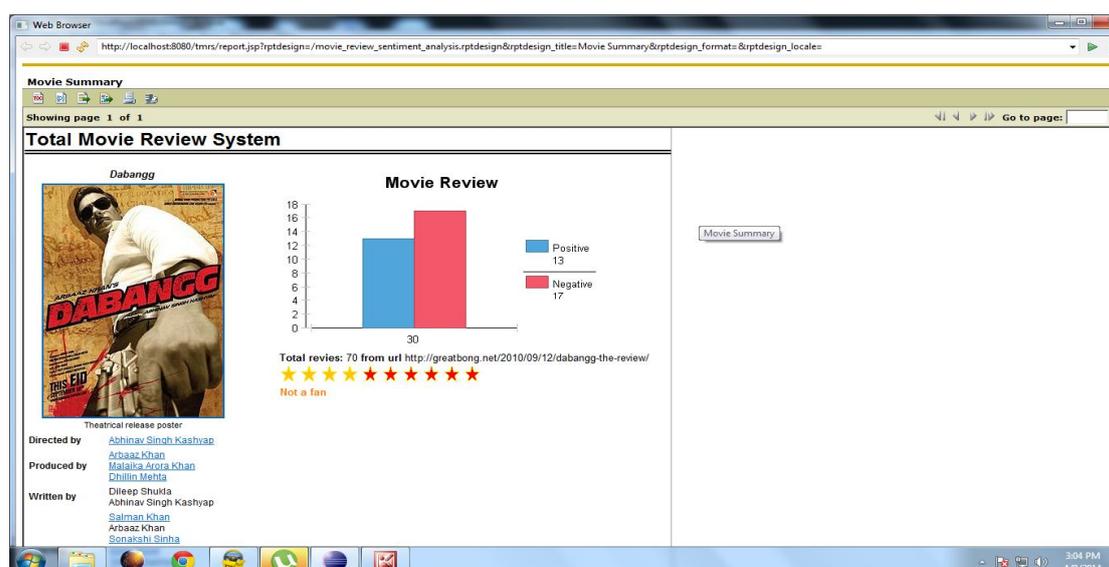


Figure 4. Result of movie review in the form of bipolar orientation

## CONCLUSION

In this project we proposed a web mining application which is used for calculating movie scores from web blogs and display top most move from stored database on their rating. We used web crawling, Sentimental analysis and visualization approach to get the expected result. For the future study, we can improve this application by adding extra features like Spell Check and movie rating system to verify manipulated reviews for improving the performance and accuracy.

## REFERENCES

- [1] K. Dave, S. Lawrence & D. Pennock, "Mining the Peanut Gallery-Opinion Extraction and Semantic Classification of Product Reviews", Proceedings of the 12th International World Wide Web Conference, pp. 519-528, 2003.
- [2] A. Esuli & F. Sebastiani, "Determining the Semantic Orientation of terms through gloss analysis", Proceedings of CIKM-05, 14<sup>th</sup> ACM International Conference on Information and Knowledge Management, pp. 617-624, Bremen, DE, 2005.
- [3] B. Pang, L. Lee & S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques", Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 79-86, Philadelphia, US, 2002.
- [4] S.M. Kim & E. Hovy, "Determining sentiment of opinions", Proceedings of the COLING Conference, Geneva, 2004.
- [5] K.T. Durant & M.D. Smith, "Mining Sentiment Classification from Political Web Logs", Proceedings of WEBKDD'06, ACM, 2006.
- [6] F. Sebastiani, "Machine Learning in Automated Text Categorization", ACM Computing Surveys, 34(1): 1-47, 2002.
- [7] P. Turney & M.L. Littman, "Unsupervised Learning of Semantic Orientation from a Hundred-Billion-Word corpus", NRC Publications Archive, 2002.
- [8] P. Waila, Marisha, V.K. Singh & M.K. Singh, "Evaluating Machine Learning and Unsupervised Semantic Orientation Approaches for Sentiment Analysis of Textual Reviews", NProceedings of International Conference on Computational Intelligence and Computing Research, Coimbatore, India, Dec. 2012.
- [9] V. K. Singh, R. Piryani, A. Uddin & P. Waila, "Sentiment Analysis of Movie Reviews and Blog Posts: Evaluating SentiWordNet with different Linguistic Features and Scoring
- [10] Zhongchao Fei, et al., Sentiment Classification Using Phrase Patterns Proceedings of the Fourth International Conference on Computer and Information Technology (CIT'04), 2004.

