

## Sentiment Analysis of YouTube Comments on Koha Open Source Software Videos

**Lambodara Parabhoi**

Professional Assistant  
Indian Institute of Advanced Study  
Shimla  
E-mail: suresh19871987@gmail.com

**Payel Saha**

Library Assistant  
KIIT University  
Bhubaneswar, Odisha  
E-mail: Payel.saha@kiit.ac.in

***Abstract-** YouTube is a popular video sharing website and used by many users for different purposes. Peoples watch, share, comments and likes on videos uploaded by the creator. Sentiment analysis is very useful topics in the field of information science. However, many commercial website, online shopping sites are frequently reviewing the comments posted by the user. In this paper we examine sentiment analysis of comments pertaining information to Koha ILS videos. In total 404 commented were found in this study. The comments were analyzed based on predetermine objective such as, Intention, Subjectivity, Emotion, Sentiment and word frequency.*

Keywords: Sentiment Analysis, Koha, YouTube Comments, YouTube Videos.

### Introduction

Social media playing an important role for dissemination of thought and ideas in worldwide. Peoples are frequently using these social media platform for different purposes i .e. stay connected with peoples. Now a day's commercial firms and online shopping sites are connected with social media platform. They used for promoting products and services. Peoples are giving their comments or opinion or feedback on online shopping websites. Amazon and Flipchart are frequency connect with social media platform. YouTube is an online popular video site that, have uploaded a huge amount of videos on different areas that may be description about a product, their services or Tutorial about a software, like wise YouTube received a huge amount of comments from the viewers. The comments and opinion is very much important part when someone buying product. The reviews and feedback found in i.e. Amazon reviews areplaying an important role to making decision. The sentiment analysis is conducted by the commercial firms, industries, online shopping sites to know where the services provided by the company is good or bad. Sentiment analysis is ongoing and rapidly popularity on different areas of study. The main motto of the sentiment analysis is to find out where the text or sentence is positive or negative emotion. (Medhat, Hassan and Korash, 2014) sentiment analysis or opinion mining researches constant in text mining field. According to (Liu, 2012) "Sentiment analysis, also called opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes". Sentiment analysis or opinion mining is a process where the text or groups of text are analyzed using tools. The software detect where the text is positive andnegative or neutral. (Liu, 2012a) stated that "sentiment analysis is a highly restricted NLP

problem because the system does not need to fully understand the semantics of each sentence or document but only needs to understand some aspects of it, i.e., positive or negative sentiments and their target entities or topics".

### **Objectives**

The main objective is to analysis the sentiment of the comments of YouTube videos on Koha ILS. The other objectives are given below:

1. To identify the intention behind a comment which are expressed by commenter?
2. To find out the subjectivity of the comments.
3. To identify the sentiment of comments whether, positive, negative or neutral?
4. To determine emotion which has expressed by users?
5. To find out mostly words used in the comments.

### **Related Work**

Several studies, on sentiment analysis have been conducted by various author in worldwide. However, some of the notable paper has been published on sentiment analysis on Twitter comments and YouTube comments or social web.(Thelwall et. al, 2012) Sentiment analysis is related with the automatic extraction of sentiment-related information from text. The sentiment analysis deal with commercial task however the role of sentiment analysis has been increasingly in social web especially in Twitter.(Prabowo and Thelwall,2009) Sentiment analysis play an important role in current research. Sentiments are found in feedback or critique; comment and it can be useful for many purposes. These sentiments are categorized into positive and negative; or into an n-point scale. More recently (Jianqiang, & Xiaolin, 2017) Twitters sentiment analysis provides of an organization with the ability to monitor public feelings that is related products and events in real time of an individual organization. In this paper the authors study about the effect of text preprocessing method of sentiment analysis on twitter comments. The study also unfold that the accuracy of twitter sentiment analysis was improved by using preprocessing method of expansion. But if various preprocessing method applied then the naive Bayes and random forest classifiers are more sensitive. (Bouazizi, Ohtsuki and Tomoaki 2017) published a paper in which authors explain that opinion mining and sentiment analysis is the most popular topic in nowadays. Most of the recent works happens on sentiment analysis and mining of texts. In this study the authors introduced SENTA open-source tool for the first time for sentiment analysis. This approach is very much accurate in binary and ternary classification. Another study conducted by (Pandey, Rajpoot, Saraswat 2017) on Sentiment analysis. The study based on hybrid cuckoo search method. The authors proposed a method called metaheuristic based clustering method which is based on K-means and Cuckoo search. They used this method for analysis of twitter comment. Schumaker, Robert P (2016) conducted a study of sentiment analysis of twitter data in Twenty Club of English Premier League. Saif, Hassan (2016) demonstrated that sentiment analysis is a recent topic and it is appreciated because of its wider application. It is appreciated both in public and commercial sector. This study present SentiCircles, it is a lexicon based approaches for sentiment analysis of twitter comments. Kharde, Vishal A. & Sonawane, S.S (2016) stated that Twitter data is very much popular in these days for sentiment analysis. This study is about sentiment analysis of twitter comments. The authors used various algorithms like, Naive bayes, Max Entropy, support vector machine. Saif, Hassan, He, Yulan & Alani, Harith (2012) Twitter comment, sentiment analysis is a very much recent topic for research. This study is about the sentiment analysis in two different ways. One is semantic feature stand another is sentiment topic feature set. Pak, Alexander

&Paroubek, Patrick (2011) sentiment analysis and affective lexicons are a useful tool for emotion studies. This study was sentimental analysis of twitter comments. (Mike, et. al, 2010) extract sentiment strength form the informal English using SentiStrength software.(Vilares, Thelwall, & Alonso, 2015) analyzed the sentiment of 2,704,523 tweets related to Spanish politicians and parties from a month in 2014-15 using SentiStrength. state that YouTube is a popular website that hold both professional and unprofessional videos. Previous studies (Bhuiyan, Hanif, Ara ,Jinat , Bardhan , Rajon& Islam , Md. Rashedul (2017) have reported that YouTube is the most popular site for watching videos, and videos are continuously uploaded. Users share and rate videos on the quality of the videos are depends on the rating. But all time not the popular videos are getting more likes or shares furthermore study present natural language processing for sentiment analysis or user comments retrieves form YouTube. This process helps to find out the relevant, popular and high-quality videos.(Thelwall et al., 2013 ) at present sentiment analysis programs are used to find out patterns of sentiment use over time in online correspondence and to help machine-controlled frameworks interact better with user's.

### Data Set

For this study, the data set was retrieved from the YouTube using Webometrics Analyst software and manually checked the comments. The authors only considered English comments for this study. Finally, we found 404 comments on Koha ILS videos on English language. The sentiment analysis was done by using Parallel Dots API and Google Spreadsheet using AYLIEN Text Analysis API. The sentiment analysis has done on categories like intention, subjectivity and sentiments, emotion and world frequency.

### Result

#### Intention

(Kröll and Strohmaier, 2009) Intent analysis is similar to sentiment analysis. It is a type of document classification that different from traditional topic categorization by focusing on classification by intent. Intent analysis provides deeper information whether a string of text is a complaint, a suggestion or a query. The table 1 depicts different intention on comments given on Koha ILS videos. Our study analyzed user intention behind a comment. Intention can be related to feedback/opinion, marketing, news, query, and spam/Junk. The data inform that majority of 164 comments on Koha ILS videos are feedback/opinion intention. However, 23 comments have their intentions for marketing, 31 comments for news intention, 145 comments for some queries and 41 comments for spam or junk intentions.

| Intention          | Comment    |
|--------------------|------------|
| Feedback/Opinion   | 164        |
| Marketing          | 23         |
| News               | 31         |
| Query              | 145        |
| Spam/Junk          | 41         |
| <b>Grand Total</b> | <b>404</b> |

**Sentence level classification**  
**Subjectivity**

In subjectivity classification of a sentence divide into two classes one is objective and other subjective. In objectivity sentences its offers factual evidence whereas in subjective sentence. An objective sentence presents some factual evidence, while a subjective sentence expresses personal feelings, views, emotions, or beliefs. Subjective sentence analysis can be done through different methods like Naïve Bayesian classification(Katrekar, 2014). After analysis of subjectivity classification, it found 338 commentsrelated to subjective whereas 66 comments are objective. It can be said that most of the comments related to objective sentence (personal feeling, opinion and emotions).

Example: 1. Useful demo as early as possible (objective) 2. Ok sure. let u know soon (subjective)

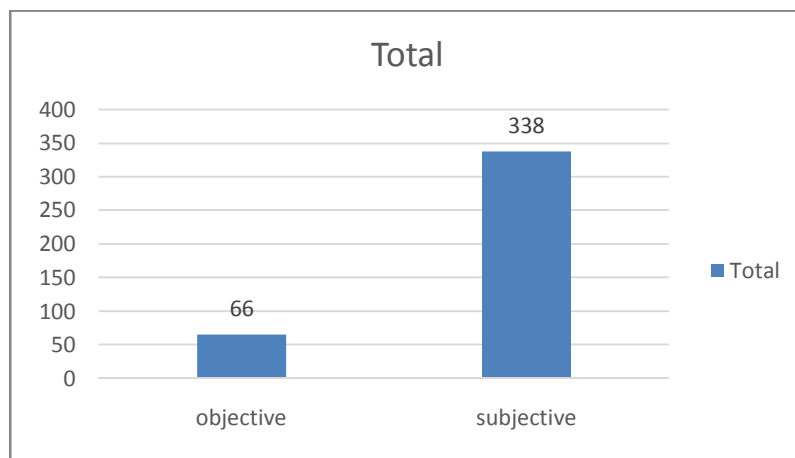


Figure.1

**Sentence level classification**  
**Sentiment-classification**

In this level of classification sentences or comments are analysed and determine whether a sentence is positive, negative, or neutral opinion. It is closely associated with subjectivity classification (Liu, 2012c). For instance, 1. Sir can you please send the Koha full pack(Neutral), 2. Good job (Positive) and 3. Link is broken (Negative). After analysis of data it reveals that Majority of 203 comments neutral sentiment followed by 81 comments for negative sentimentand 120 comments for positive sentiment out of total404comments.

| Row Labels  | Count of Comments |
|-------------|-------------------|
| Negative    | 81                |
| Neutral     | 203               |
| Positive    | 120               |
| Grand Total | 404               |

Table.2

**Emotion**

This sub section deals with emotion analysis of sentence whether an expression is “happy, sad, angry, fearful, excited, bored or sarcastic”. From thefigure.2it can befound that majority of comments deal with the emotion “Happy” with 173, followed by 75 excited, 83 sad, 34

comments is about sarcasm, 23 comments have anger emotion and 16 comments related with fear emotion.

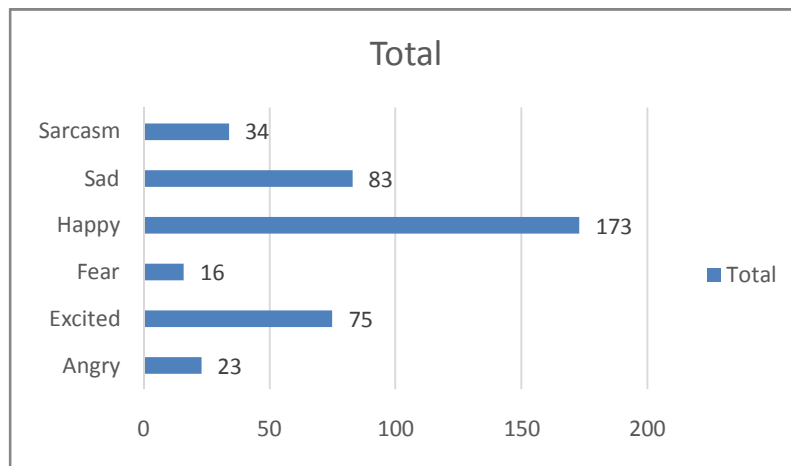


Figure.2

Example: 1. Thank you sir, it is very nice (Excited) 2. Your mail ID is not valid (Angry), 3.Can I install this on server? I mean online (Sarcasm)

### Word Frequency

This table shows the top five-word frequency of comments received on Koha ILSvideos. After analysis of data found that out that "KOHA " word occur 149 times, "Depends" word come 104 times,

"Perl" word come in this comment 104 times, Perl 104 times and installable 97 times and thanks words use 63 times in comment lines.

### Conclusion

The above study on sentiment analysis of Koha ILS videos found on YouTube. About 404 total comments were found, which were discussed in this study. YouTube is online platform that have billions of videos beenuploaded and many new users are connected with YouTube and watch videos related to his/ her interest. The current studies on Koha open source software ILS, which is very popular in the world and used by many libraries? This study gives information about how and what kinds of comments posted by the user of Koha ILSYouTube videos.

### References

1. Asghar, M. Z., Ahmad, S., Marwat, A., &Kundi, F. M. (2015). Sentiment Analysis onYouTube: A Brief SurveyMAGNT Research Report,3(1), 1250–1257.
2. Bhuiyan,H., Ara, J., Bardhan, R., & Islam, M. R. (2017). Retrieving YouTube Video by Sentiment Analysis on User Comment. In prof. of the 2017 IEEE international Conference on Signal & Image Processing Applications (IEEE ICSIPA 2017), Malaysia, September 12-14,2017 (pp. 474–478). IEEE.

3. Bouazizi, M., Ohtsuki, T., & Member, S. (2017). A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter. *Rapid Review*, 5, 20617–20639.
4. Jianqiang, Z., & Xiaolin, G. U. I. (2017). Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis. *Rapid Review*, 5, 2870–2879.
5. Kharde, V. A., & Sonawane, S. S. (2016). Sentiment Analysis of Twitter Data: A Survey of Techniques. *International Journal of Computer Applications*, 139(11), 5–15.
6. Liu, B. (2012). *Sentiment Analysis & opinion Mining*. Morgan & Claypool Publishers.
7. Pak, A., & Paroubek, P. (2011). Twitter for Sentiment Analysis: When Language Resources Are Not Available. In 2011 22nd International Workshop on Database and Expert systems Applications (pp. 111–115). IEEE. <https://doi.org/10.1109/DEXA.2011.86>.
8. Pandey, A. C., Rajpoot, D. S., & Saraswat, M. (2017). Twitter sentiment analysis using hybrid cuckoo search method. *Information Processing and Management*, 53(4), 764–779. <https://doi.org/10.1016/j.ipm.2017.02.004>.
9. Saif, H., He, Y., & Alani, H. (2012). Alleviating Data Sparsity for Twitter Sentiment Analysis. In MSM2012, 2nd Workshop on Making sense of Microposts, April 16 2012, (pp. 2–9). Lyon, France.
10. Saif, H., He, Y., Fernandez, M., & Alani, H. (2016). Contextual semantics for sentiment analysis of Twitter. *Information Processing and Management*, 52(1), 5–19. <https://doi.org/10.1016/j.ipm.2015.01.005>.
11. Schumaker, R. P., Jarmoszko, A. T., & Jr, C. S. L. (2016). Predicting wins and spread in the Premier League using a sentiment analysis of twitter. *Decision Support Systems*, 88, 76–84. <https://doi.org/10.1016/j.dss.2016.05.010>.
12. Statistical Cybermetrics Research Group. (2018). *Webometrics Analyst*. Wolverhampton, UK: University of Wolverhampton.
13. Thelwall, M., Sud, P., & Vis, F. (2012). Commenting on YouTube Videos: From Guatemalan Rock to EI big bang. *Journal of the American Society for Information Science & Technology*, 63(3), 616–629.
14. Uryupina, O., Plank, B., Severyn, A., Rotondi, A., & Moschitti, A. (2014). SenTube: A corpus for sentiment analysis on YouTube social media. In Ninth International Conference on Language Resources & Evaluation (LREC-2014) (pp. 4244–4249). Retrieved from [http://www.lrec-conf.org/proceedings/lrec2014/pdf/180\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2014/pdf/180_Paper.pdf).
15. Globallogic Inc. (2014). An Introduction to Sentiment Analysis. Retrieved from <https://globallogic.com/wp-content/uploads/2014/10/Introduction-to-Sentiment-Analysis.pdf>.

16. Kroll, M., &Strohmaier, M. (2009). Analyzing human intentions in natural language text.Proceedings of the fifth International Conference on Knowledge Capture -K-CAP'09. doi:10.1145/1597735.1597780.
17. Prabowo, R., &Thelwall, M. (2009). Sentiment analysis: A combined approach.Journal of Informetrics, 3(2), 143-157. doi:10.1016/j.joi.2009.01.003.
18. Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., &Kappas, A. (2010). Sentiment strength detection in short informal text.Journal of the American Society for Information Science & Technology, 61(12), 2544-2558. doi:10.1002/asi.2141.
19. Thelwall, M., Buckley, K., &Paltoglou, G. (2011).Sentiment strength detection for the social web.Journal of the American Society for Information Science & Technology, 63(1), 163-173. doi:10.1002/asi.21662.
20. Thelwall, M., Buckley, K., Paltoglou, G., Skowron, M., Garcia, D., Gobron, S., ... Holyst, J. A. (2013). Damping Sentiment Analysis in Online Communication: Discussions, Monologs and Dialogs.ComputationalLinguistics& Intelligent Text Processing, 1-12.doi:10.1007/978-3-642-37256-8\_1.
21. Vilares, D., Thelwall, M., & Alonso, M. A. (2015). The megaphone of the people? Spanish SentiStrength for real-time analysis of political tweets. Journal of Information Science, 41(6), 799-813. doi:10.1177/0165551515598926.

