iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** IX **Month of publication:** September 2021

DOI: https://doi.org/10.22214/ijraset.2021.38325

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com

A Study on Sentimental Analysis of Novels and Shakespeare's Play

Rohan Baldha
Computer Science, The LNM Institute of Information Technology

Abstract: The current work is a sentimental analysis study of Shakespeare's play, depicting the character-to-character analysis. The emotional sentimental analysis of Harry potter's novels is also analyzed. The current work exemplifies a complete evaluation of the techniques required for sentimental analysis, as well as their practical application. Because of the planned discourse of this literary format, it is possible to make assumptions about who is taking part in a conversation. Once it is known to whom a character is speaking, the emotions in his or her speech can be assigned to that individual, enabling for the generation of lists of a character's foes and allies as well as the pinpointing of situations crucial to a character's emotional development. In addition, the Serendio group's framework for estimating utilizing dictionary-based methods is employed.

I. INTRODUCTION

The method of determining whether a block of text is good, negative, or neutral is known as sentiment analysis. Sentiment analysis is the contextual mining of words that reveals the social sentiment of a brand and assists businesses in determining whether the product that they are creating will be in demand in the market. Sentiment analysis attempts to achieve the goal of analyzing people's opinions in order to assist businesses in expanding. It is concerned not just with polarity (positive, negative, and neutral), but also with emotions (happy, sad, angry, etc.). It employs a variety of Natural Language Processing algorithms, including rule-based, automatic, and hybrid [1].

For example, if we want to determine whether a product meets client needs or whether there is a market need for this product. We can utilize sentiment analysis to keep track of the product's reviews. Sentiment analysis is also useful when there is a huge quantity of unstructured data that needs to be classified by automatically tagging. Net Promoter Score (NPS) surveys are widely used to learn how a client evaluates a product or service. Sentiment analysis has also grown in popularity due to its ability to analyse huge numbers of NPS data rapidly and consistently [2].

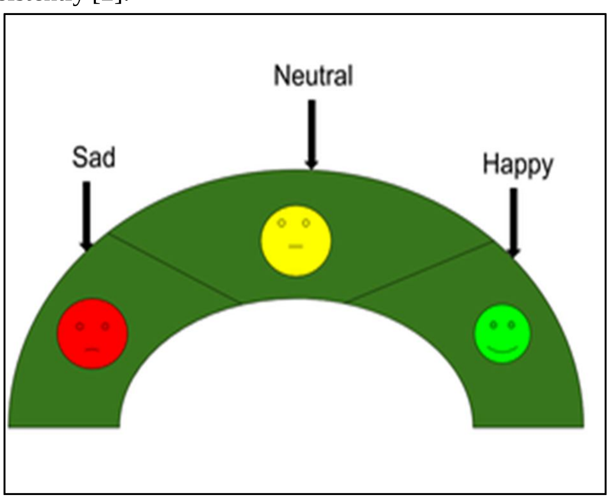


Figure 1: Sentimental Analysis

According to the survey, unstructured data accounts for 80% of the world's data. The data must be examined and structured, whether it is in the form of emails, messages, documents, articles, or anything else. The need for sentimental analysis arises from the fact that it stores data in an efficient and cost-effective manner. Moreover, sentiment analysis handles real-time concerns and can assist you in solving all real-time circumstances [2].

II. CLASSIFICATION OF SENTIMENTAL ANALYSIS

The sentimental analysis can be classified into four following types [3]:

A. *Fine-grained Sentiment Analysis*

This is determined by the polarity. This category might be highly positive, very positive, neutral, negative, and very negative. The rating is given on a scale of 1 to 5. A rating of 5 indicates that it is very positive, a rating of 2 indicates that it is negative, and a rating of 3 indicates that it is neutral.

B. *Emotion Detection*

Emotion detection includes the sentiments happy, sad, angry, upset, merry, pleasant, and so on. It is also known as a sentiment analysis lexicon approach.

C. *Aspect Based Sentimental Analysis*

It concentrates on a specific aspect, such as if a person wants to check the feature of a cell phone, it checks the aspect such as the battery, screen, camera quality, and so on.

D. *Multilingual Sentimental Analysis*

It refers to multiple languages that must be classified as good, negative, or neutral. This is a demanding and challenging task.

III. APPROACHES OF SENTIMENTAL ANALYSIS

Sentimental Analysis can be approached in the following three-way [4]:

A. *Automatic Approach/System*

Machine learning algorithms like clustering are used in automated sentiment analysis approaches. The classifier is fed large chunks of text and provides negative, neutral, or positive results. We'll look at two main procedures that make up automatic systems right now.

B. *Rule-Based Approach/System*

Rule-based techniques, unlike automated models, rely on custom rules to classify data. Tokenization, parsing, stemming, and a few more techniques are popular. The example we looked at previously can be considered a rule-based approach. The capacity to customize rule-based systems is one of their advantages. By establishing smarter rules, these algorithms may be tailored to the circumstances. Keep in mind that these rule-based models will need to be updated on a regular basis to provide consistent and improved results.

C. *Hybrid Approach/System*

For sentiment analysis, hybrid techniques are the most contemporary, efficient, and extensively utilized strategy. Hybrid systems that are well-designed can combine the advantages of both automatic and rule-based systems. Hybrid models combine the strength of machine learning with customization flexibility.

IV. LITERATURE REVIEW

Sentiment analysis (SA) is now commonly utilized in the business world to infer customer opinions from product reviews and social media posts [1]. When the data is labeled (e.g., IMDB's user reviews are labeled with one to ten stars, which are supposed to correspond with the text's polarity), traditional machine learning techniques on n-grams, parts of speech, and another bag of words features can be employed [2]. However, content that is annotated with its real sentiments is difficult to come by, thus labels are frequently collected through crowdsourcing.

Knowledge-based solutions (which frequently rely on crowdsourcing) offer an alternative to labeled data [3]. Sentiment lexicons, which are fixed lists that associate words with "valences" (signed numbers that reflect positive and negative feelings), are at the heart of these systems [4, 5]. Some lexicons allow for the examination of specific emotions by linking words with degrees of fear, joy, surprise, rage, anticipation, and so on [6]. Unsurprisingly, approaches that lack deep knowledge, such as these, tend to operate better as the length of the input text grows.

When it comes to automatic semantic analysis of fiction, it appears that narrative modification and summarization have received the most attention. Elson and McKeown [7] developed a platform that can symbolically represent and reason over narratives, while Chambers and Jurafsky [8] described a system that can learn (without supervision) the sequence of events described in a narrative.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com

Character interactions have also been represented as networks in order to study narrative structure. Mutton [9] modified Internet Relay Chat (IRC) methodologies for extracting social networks to mine Shakespeare's plays for their networks. Elson and McKeown [7] extended this line of research to novels, developing a reliable method for speech attribution in unstructured texts and successfully extracting social networks from Victorian novels [1].

While the structure is unquestionably vital, we feel that evaluating a narrative's emotions is critical to conveying the "reading experience," a viewpoint shared by others. Alm and Sproat (2005) looked at the 'emotional trajectories' of Brothers Grimm fairy tales and discovered that emotion rises as the story goes. Mohammad [6] took their work to the next level by employing a crowd-sourced emotion lexicon to track emotion dynamics throughout a variety of texts and plays, including Shakespeare's. Elsner [7] examined emotional trajectories at the character level in Pride and Prejudice, revealing how Miss Elizabeth Bennet's feelings vary throughout the course of the novel.

V. METHODOLOGY

Two different sentimental analysis is performed in the present study. The first part explains the sentimental analysis done over Shakespeare's play and the later part will explain the sentimental analysis of harry potter's novel.

A. Shakespeare Play

This section presents the character-to-character sentimental analysis of Shakespeare's play. The play "Othello" is sentimentally analyzed in this article to determine the sentiment of the protagonist "Othello," whereas the other characters in the play are positive, negative, or neutral. The Lexicon-based sentimental analysis was utilized in the current work (Figure 2). The approach used in the analysis is depicted in figure 3.

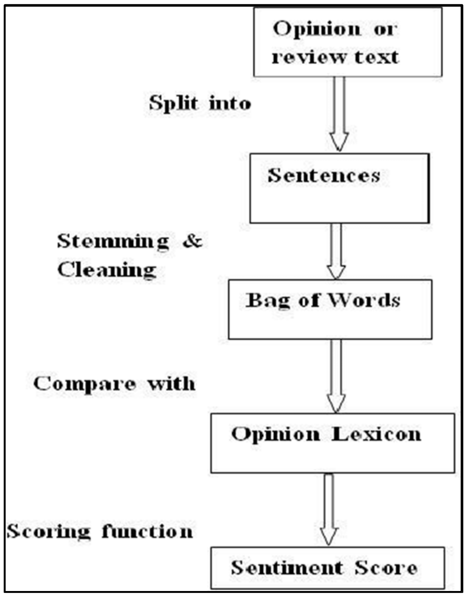


Figure 2 Lexicon Based Sentimental Analysis

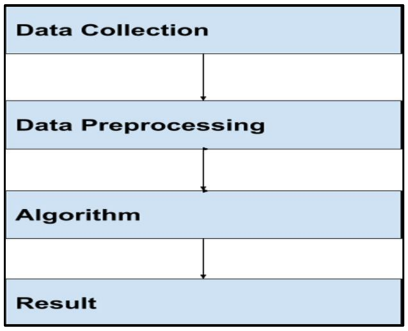


Figure 3: Approach of Analysis

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com

B. Data Acquisition

FASTA files taken are the files from the Gutenberg file that looks like the depicted picture (Figure 4). Where the character's name is represented by the word following the symbol ">". The numbers following the character's name represent the act and scene numbers, respectively. The nth number of dialogues in that scene is represented by the incremental counter.

```
>roderigo11_1
Tush, never tell me; I take it much unkindly
That thou, Iago, who hast had my purse
As if the strings were thine, shouldst know of this,--


>iago11_1
'Sblood, but you will not hear me:--
If ever I did dream of such a matter,
Abhor me.
```

Figure 4 Data Acquisition

C. Data Preprocessing

The file format was selected such that it is easy and efficient to read the dataset. Initially, the spaces are replaced by known characters along with the removal of tabs as demonstrated in figure 5. Two dictionaries were created, one for the character list and the other for the speech list (Figure 6(a)). All the reductant words were either replaced or removed. The character was the key, and speech was the value, on a map that was developed (Figure 6(b)). In the end, the algorithm was implemented as depicted in figure 7.

```
>roderigo11_4
I would not follow him, then.

  Is converted to --

'>roderigo11_4@',
'i#would#not#follow#him,#then.@@@'
```

Figure 5: Modified Dataset

```
a  o Character list -- othello13_3', 'duke13_17', 'othello13_4'
   o Speech list -- "'sblood, #but#you#will#not#hear#me:
b      'iago33_48': 'would! nay,  i will.    ',
```

Figure 6: (a) Dictionary (b) map

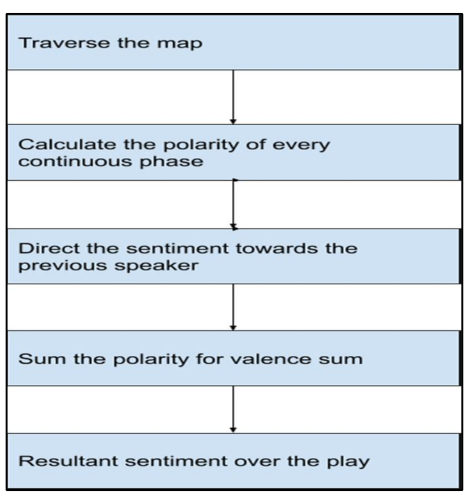


Figure 7: Algorithm of Shakespeare Play

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com

D. *Sentimental Analysis of Harry Potter's Novel*

This section describes how to discover the emotional sentimental arc throughout Harry Potter's novel using emotional sentimental analysis. The VADER sentimental analysis method was used to perform the sentimental analysis in the present work. The VADER word stands for "Valence Aware Dictionary and Sentiment Reasoned". It's the type of vocabulary and principle-based emotion assessment tool that's specifically sensitive to assumptions transmitted in web-based life. VADER employs a variety of techniques. A presumption dictionary is a list of lexical highlights (e.g., words) that are classified as positive or negative based on their semantic orientation. VADER notifies the positive and negative sentiment, as well as the positivity and negativity score.

E. *Data Preprocessing*

The data set contains all of the novels in the Harry Potter series in pdf format. Preprocessing entails replacing and deleting all unneeded words. By using regex and modifying the format of the data such that it is easily readable by the script. Figure 8 demonstrates the format of the Harry Potter hp dictionary.

```
{book 1 title: {
        'Chapter 1': (chapter title, chapter text),
        'Chapter 2': (chapter title, chapter text),
        'Chapter 3': (chapter title, chapter text),
        ...
        }
'book 2 title': {
        'Chapter 1': (chapter title, chapter text),
        ...
       }
 ...
}
```

Figure 8: Format of Harry Potter hp Dictionary

F. *Algorithm*

VADER works best on short messages (a few sentences at most) and applying it to an entire part without a second's hesitation resulted in ludicrous and generally meaningless results. Instead, I circled each sentence individually, calculated the VADER ratings, and then took a normal of all sentences in a segment.

G. *Emotional Analysis*

The NRC emotion lexicon is employed. The algorithm goes over each word in a chapter, looks it up in the lexicon, and outputs the emotions associated with that word. Each chapter was then given a score for each emotion based on the number of words related to that emotion in the chapter relative to the total number of words in the chapter (for normalization). The constructed dictionary is now turned into a data frame with the addition of eight columns, each corresponding to a different emotion. The data frame is depicted in figure 9.

Out[50]:

book	chapter_title	text	anger	anticipation	disgust	fear	joy	negative	positive	sadness	surprise	trust	word_count
Harry Potter and the Sorcerer's Stone	THE BOY WHO LIVED	Mr. and Mrs. Dursley, of number four, Privet D...	0.006786	0.015138	0.005046	0.010266	0.009744	0.022795	0.027145	0.009048	0.009744	0.022098	5747
	THE VANISHING GLASS	Nearly ten years had passed since the Dursleys...	0.013738	0.020843	0.015632	0.021317	0.013027	0.036002	0.029607	0.016106	0.015869	0.024159	4222
	THE LETTERS FROM NO ONE	The escape of the Brazilian boa constrictor ea...	0.014153	0.024715	0.008872	0.011196	0.008450	0.025982	0.024715	0.011196	0.008872	0.018589	4734
	THE KEEPER OF THE KEYS	BOOM. They knocked again. Dudley jerked awake....	0.012528	0.017252	0.008831	0.022797	0.007599	0.032039	0.024440	0.014377	0.014377	0.016020	4869
	DIAGON ALLEY	Harry woke early the next morning. Although he...	0.008068	0.017657	0.009471	0.010290	0.009705	0.020931	0.028648	0.007951	0.009822	0.018007	8552

Figure 9: Data Frame

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com

VI. RESULTS

A. *For Shakespeare Play*

The graph in figure 10 depicts how Othello and Iago's sentiments for each other change during the play. The red dots represent Othello's feelings for Iago, while the blue dots represent Iago's feelings for Othello on the opposite side of the green area/line. The graph in figure 11 depicts how Desdemona and Othello’s sentiments for each other change during the play. The red dots represent Desdemona’s feelings for Othello, while the blue dots represent Othello's feelings for Desdemona on the opposite side of the green area/line.

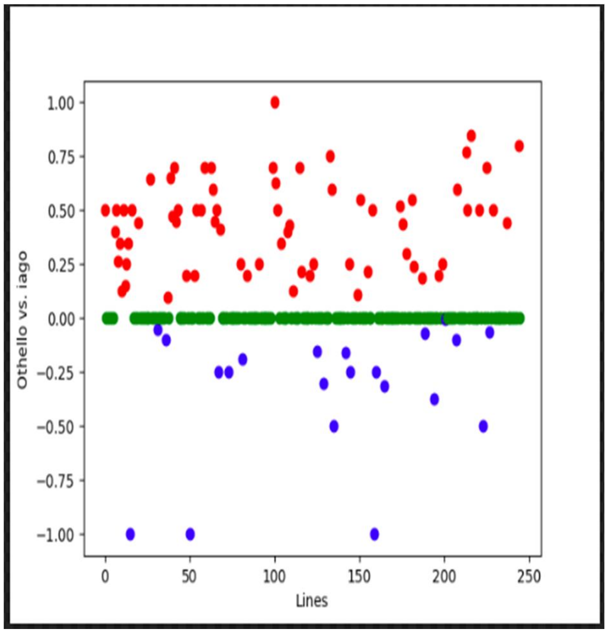


Figure 10: Sentiments of Othello towards Iago

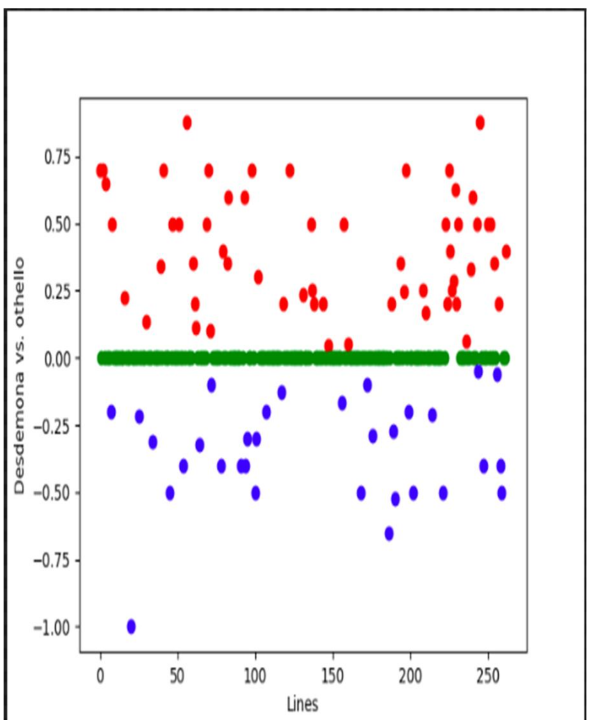


Figure 11: Sentiments of Desdemona and Othello

B. For Harry Potter's Novel

It can be stamped the occasions in the books by calculating the VADER compound score for each part of each book. Harry Potter being chosen by the Goblet of Fire around section 70 of the arrangement, Cedric Diggory's death around section 88, and Dumbledore's death around section 160 are the three biggest spikes in that diagram. Figure 12 shows the emotional sentiments of the harry potter series. Each emotion of Harry Potter's sentimental analysis is shown in figure 13.

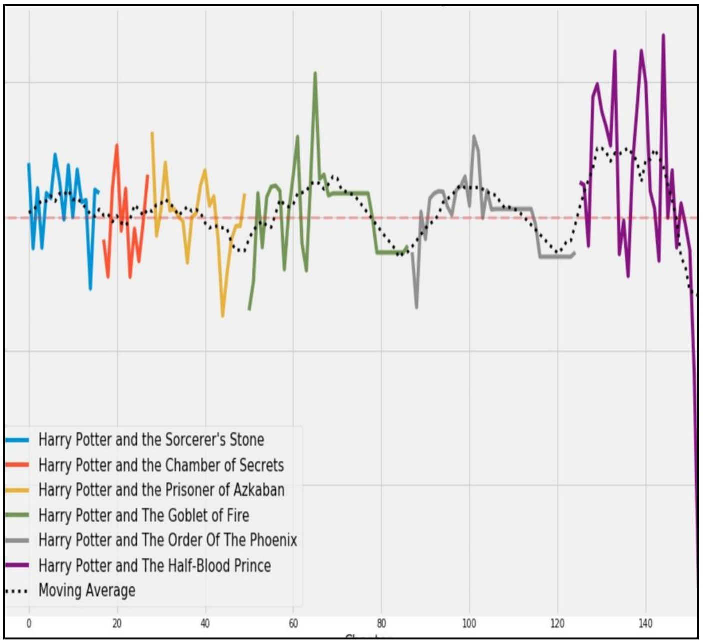


Figure 12 Emotional Sentiment of the Harry Potter Series

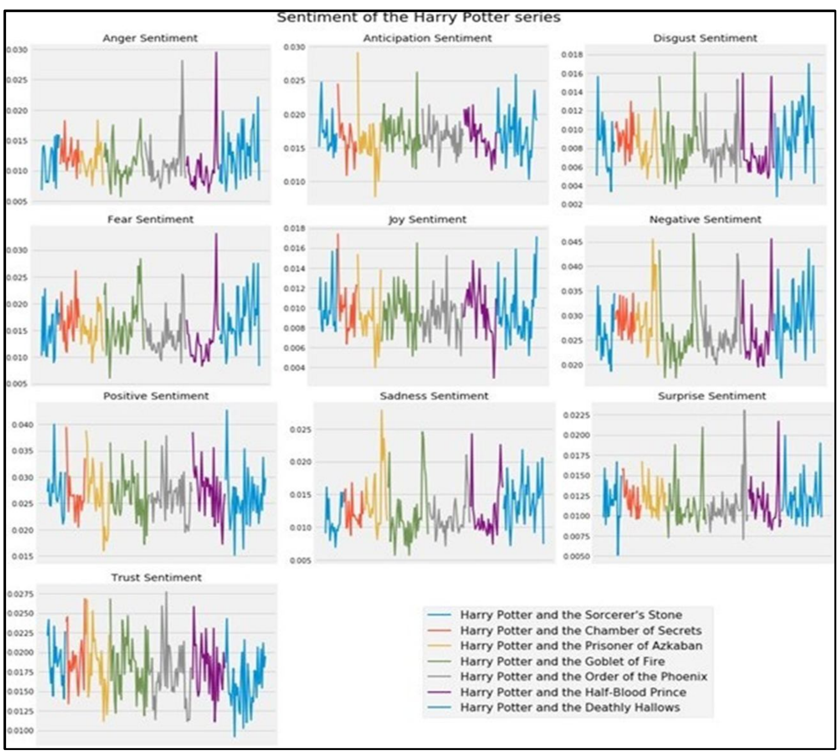


Figure 13 Sentimental Analysis of the Harry Potter Series

VII. CONCLUSION

As shown, the superficial, non-customized sentimental analysis combined with text structure can be utilized to evaluate interpersonal relationships depicted in a play and produce an interpretation that matches reader expectations. We noticed the character-to-character sentiment analysis of each character in Shakespeare's play "Othello" concerning the protagonist. As can be seen from the data, the outcomes are the pure justification of the protagonist's relationships with contemporary characters. We've also looked at the emotional sentiment analysis of the "Harry Potter" novels, and based on the results and graph, we may conclude that the sentiment analysis results are a reproduction of the novel's genuine timeline. This analysis aids the reader in comprehending the material in a new and more visual manner. Future studies would entail reading more texts by different authors and employing various machine learning and artificial intelligence approaches to ensure that the results are more authentic. Future studies could also include training the classifiers on old English to improve the accuracy of the results.

REFERENCES

[1] Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. Foundations and trends in infor- mation retrieval, 2(1-2):1–135.

[2] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on Empirical methods in natu- ral language processing - Volume 10, EMNLP ’02, pages 79–86, Stroudsburg, PA, USA. Association for Computational Linguistics.

[3] Alina Andreevskaia and Sabine Bergler. 2007. Clac and clac-nb: knowledge-based and corpus-based ap- proaches to sentiment tagging. In Proceedings of the 4th International Workshop on Semantic Evalu- ations, SemEval ’07, pages 117–120, Stroudsburg, PA, USA. Association for Computational Linguis- tics.

[4] Soo-Min Kim and Eduard Hovy. 2004. Determin- ing the sentiment of opinions. In Proceedings of the 20th international conference on Computational Linguistics, COLING ’04, Stroudsburg, PA, USA. Association for Computational Linguistics.

[5] Strapparava and A. Valitutti. 2004. Wordnet-affect: an affective extension of wordnet. In Proceedings of LREC, volume 4, pages 1083–1086.

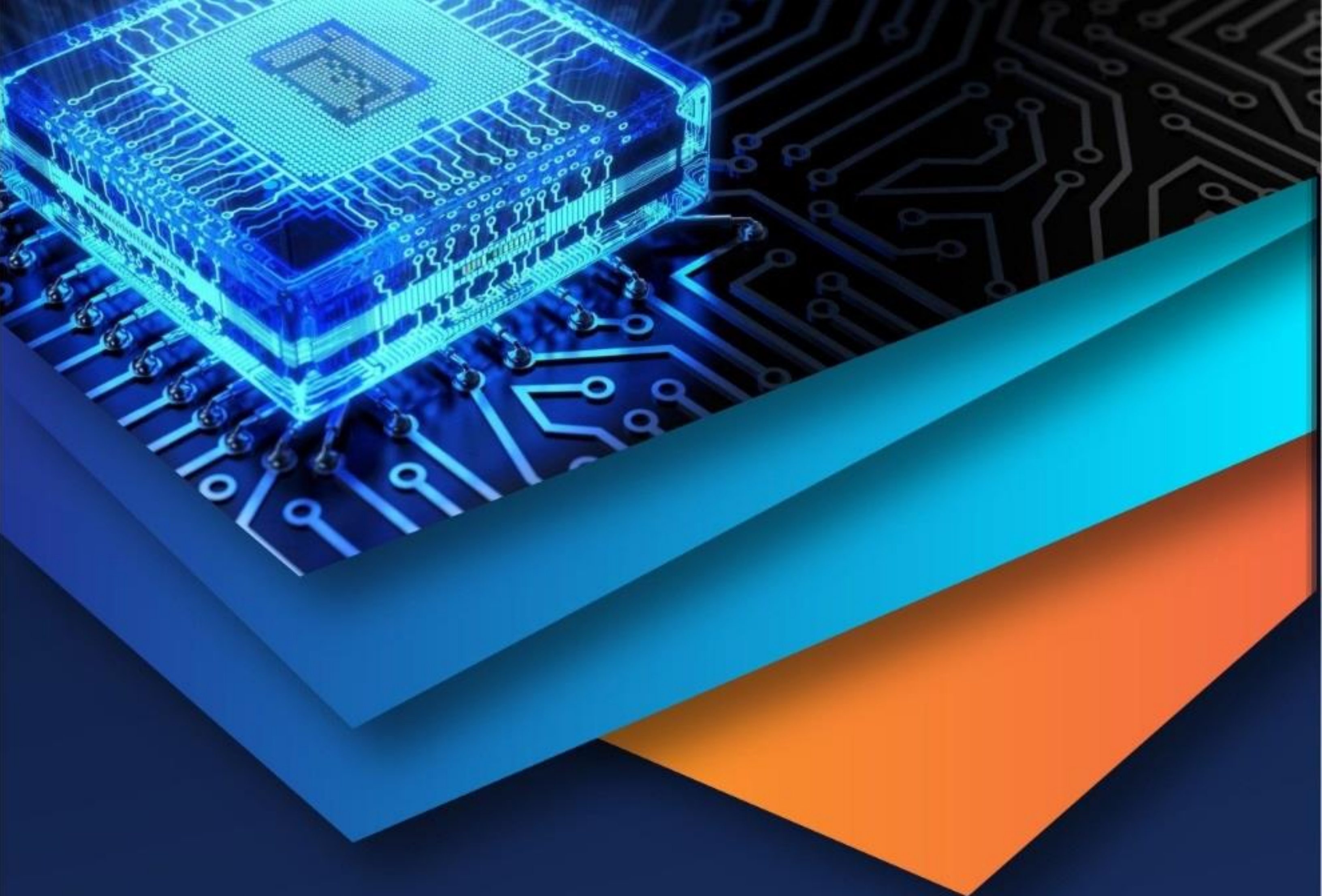
[6] S. Mohammad. 2011. From once upon a time to happily ever after: Tracking emotions in novels and fairy tales. In Proceedings of the 5th ACL- HLT Workshop on Language Technology for Cul- tural Heritage, Social Sciences, and Humanities, pages 105–114. Association for Computational Lin- guistics.

[7] David K Elson and Kathleen R McKeown. 2009. Ex- tending and evaluating a platform for story under- standing. In Proceedings of the AAAI 2009 Spring Symposium on Intelligent Narrative Technologies II.

[8] Nathanael Chambers and Dan Jurafsky. 2009. Un- supervised learning of narrative schemas and their participants. In Proceedings of the Joint Confer- ence of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Lan- guage Processing of the AFNLP: Volume 2-Volume 2, pages 602–610. Association for Computational Linguistics.

[9] P. Mutton. 2004. Inferring and visualizing social net- works on internet relay chat. In Information Visuali- sation, 2004. IV 2004. Proceedings. Eighth Interna- tional Conference on, pages 35–43. IEEE.

[10] Agarwal, A. Corvalan, J. Jensen, and O. Rambow. 2012. Social network analysis of alice in wonder- land. NAACL-HLT 2012, page 88.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY