

How to Cite:

Aggarwal, R., Srivastava, R., & Singh, P. (2022). Opinion mining based on partial textual entailment. *International Journal of Health Sciences*, 6(S5), 3825–3835.
<https://doi.org/10.53730/ijhs.v6nS5.9452>

Opinion mining based on partial textual entailment

Ruchika Aggarwal

Associate Professor, Echelon Institute of Technology, Faridabad
Corresponding author email: ruchikaaggarwalech@gmail.com

Rachana Srivastava

Assistant Professor, Echelon Institute of Technology, Faridabad
Email: rachmini2610@gmail.com

Priya Singh

Assistant Professor, Echelon Institute of Technology, Faridabad
Email: priyasishodiya@gmail.com

Abstract--The net is a heap of uncooked facts within the form of multimedia messages such as tweets, blogs and many others. Unordered and unconnected. The information that turns into records is most effective after it is processed, organized and structured in the form of online textual content snippets. According to related research, we determined the current sentiment analysis tools to be sluggish, cumbersome and arithmetically heavy which causes the venture handy to be inefficient. Consequently, to bypass the above-mentioned trouble of reading opinions proficiently, a latest procedure is put-forward in the current tabloid utilizes the conception of Partial Textual Entailment (PTE) to handle this job of Opinion Mining. Among the tweets the method for calculating Semantic resemblance as a way to organize homogeneous tweets simultaneously the utilization of PTE is recommended. To minimize the load on Sentiment analyzer this approach is expected as it also helps to make the processing speed quicker. Additionally, for making countless Natural Language Processing applications we advise an amendment in the current procedure of PTE which can also be additionally accepted.

Keywords--NLP, Opinion Mining, Textual Entailment etc.

1. Introduction

In the current generation of continuing numerical upheaval, you could discover a excess of text on multimedia web sites, webcasting sites, and so on. The bite sized

multimedia messages permit operators to express thoughts, trade sentiments on modern-day problems and put up thoughts - positive or negative - of mechanically used merchandise/resources. Currently, organizations additionally look on social media websites, microblogging websites, and many others as a major network to measure public sentimentality for their outcomes. Numerous organizations have teams committed to addressing consumer response. This has spread out the area of integrated era-driven sentimentality evaluation.

Levy et al [1] helps in figuring out the polarity of a sentimentality with the help of the study of sentiment analysis. The challenge of Sentiment evaluation consists of locating the emotion, mindset, opinion and view of a individual towards any other individual, item, outcome or an event [2][3] turns out to be beneficial in the scope of academics, business and production. Those extracted feelings, standpoints or sentiments provide the ability to offer precise views on numerous packages in extraordinary areas. Social media websites like Facebook or Twitter have turn out to be a widespread network to carry out evaluation on shared views. Consequently, it is able to be stated that Sentiment analysis classifies the sentimentality or emotional state of an individual in a field. It enables to check if an individual is contented or dis-contented while Opinion Mining [4] is used to classify view or opinions of an individual in a field. It allows to evaluate if someone is contented or dis-contented. The foundation of diverse classes can be categorized as a job of Sentiment evaluation: -

Opinion can also be analysed based on classes such as positive, negative and neutral [5].

- *Positive Opinion Mining:* If the opinions expressed by the document about an entity are positive, it can be said that the polarity of the data is positive. In sentences or aspect of sentences Polarity can also be determined. As a Sample: The scent of the flora is ideal. in the above sentence “scent” can be taken into consideration as a characteristic of the “flora” and “ideal” suggests the advantageous opinion about the scent of the flora.
- *Negative Opinion Mining:* If the data indicates bad opinions about an object, it can be said that the the polarity of the record is negative. As a Sample: The coatings on the pie have been unpleasant. On this statement the “coatings” is deemed as a certain characteristic of the pie and “unpleasant” designates the negative opinion about the pie coatings.
- *Neutral Opinion Mining:* If we cannot determine the positive or negative opinion from a report, its polarity is considered as neutral. As an example: Ram is going to home.

Study about emotional data can be similarly categorised on the foundation of one-of-a-kind granularity of textual content. It can also be categorised at a stage known as document stage, sentence degree or aspect degree [5] that is in brief conversed beneath

- *Document Based Opinion Mining:* To find the polarity of the data the file is analysed as a whole.
- *Sentence Based Opinion Mining:* Various sentences are taken into account of a report with the aid of similarly disintegrating them into miniature pieces of terms for determining its polarity.

- *Word Based Opinion Mining*: To find the polarity of the various arguments of a given sentence it is determined with connection to a certain item or occasion.

Due to the ever-growing variety of multimedia posts and tweets, the task of performing the assignment of examining all evaluations inside affordable time lines has become a challenging issue. It has come to be extraordinarily critical to classify similar opinions to be able to decrease the ciphering burden on the corporation. To diminish the above stated problems, In the technique of Opinion Mining the technique of the Textual Entailment is proposed in the given paper. Textual Entailment can be explained [6] as a guiding courting among two textual fragments which can be represented as textual content (TC) and as hypothesis (HT) It can be defined as a procedure of examining such that HT should be determined from TC or not. From the Literature review the data mined indicates, the authors have located for excess detection by means of Lloret et al. [7] TE has been used. With the help of the mined data the researchers in this paper have been encouraged to apply it in venture of Opinion Mining too.

A belief which is used to verify semantic inference among textual content terminologies with the help of syntax-derived, semantic-derived through diverse strategies [8] is known as TE. The work of TE is passed ahead the usage of numerous processes such as the container of words, vectors and directional tree strategies.

Through the previous couple of ages TE has been growing to be a widespread subject and matter of curiosity in the region of Natural Language Processing (NLP) which includes scholar reaction study, report summary, machine translation assessment, paraphrase identification and many others [9]. Preferably, the process of accurately gauging the inference with the help of TE cannot be accurately measured by the authors due to many reasons such as the difficulty of texts and barriers of prevailing language origin. Consequently, to identify whether the data absolutely involves the hypothesis or not a necessity of the Partial Textual Entailment (PTE) is felt such that it determines the degree of entailment. To enhance the venture of Sentiment analysis in this research the authors have proposed to appoint Partial Textual Entailment (PTE). it's far anticipated to diminish the quantity of labor performed in studying the sentimentality of the text thus presenting an improved answer for endeavor of Opinion Mining. The upcoming research is drafted as follows:- the associated task in the discipline of Textual Entailment and Opinion Mining has been mentioned at Part-2, In Part-3 Partial Textual Entailment is brought and debated, In Part- 4 the intended technique for Opinion Mining is proposed, deduction and future tasks are assigned in phase 5.

2. Literature Review

Opinion Mining that is achieved through the ages and its exclusive models are hooked into our lifestyles. The use of system mastering has been advanced to produce polarity on a record degree for a sentiment [10]. A procedure to determine the semantic categorization of terms the usage of factor-smart Mutual records (PMI) rating have been advanced [11]. Numerous techniques have been utilized

and developed feature over the years on records, video and audio for sentiment evaluation [12][13]. NLP founded totally techniques were meant for purchaser evaluations to deliver characteristic-based synopses [14]. NLP methods were used to determine the views for an object from the internet and produce the expression of polarity [15]. To calculate the polarity of expression a device for getting to know the set of rules is proposed for the use of sentiment evaluation on word level [16]. An unmanaged tool is developed for the elimination of functional-derived synopsis for client opinions [14][17]. PMI rating is derived to finish if a noun is used as a characteristic of a material review. the 2 grading techniques PMI and the Latent Opinion Mining (LSA) are verified through the use of two distinct amounts to finish LSA technique that are greater correct in semantic orientation class [18].

The flexibility of text in Partial Textual Entailment (PTE) and its polarity is endorsed through this paper by the authors. To decrease computations in Sentiment evaluation the authors advocate the amendment over the present side's era approach within the proposed outline. Through the following segment we will speak about the problem of TE and present PTE.

3. Utilizing PTE

A binary decision problem consists of Recognizing Textual Entailment (RTE) i.e., if Text (T) involves or does not involve the Hypothesis (H) is indicated by the technique. The process is determined to remain inflexible and frail [19] as we are incapable of identifying the equivalent at which T involves H [20] with the help of the processes.

Nielsen et al. [1] presented the notion of PTE with the help of facets to overcome the aforementioned limitations, Substantial development over threshold approaches like bag of words are achieved by the concept of PTE with the help of paper reports. The additional improvement over the proposed concept is performed by Levy et al.[1] with the help of aspect PTE. With the help of resulting work performed the systemized pair facets which are help in achieving improved outcomes as they are unrestricted from partaking semantic relationship.

PTE resolved the trouble of negative Textual Entailment according to Levy et al. [1]. In Hypothesis (HT), Text (TC) is partially involved which is indicated by the fact that a portion of hypothesis is involved through text. From the given Sample:

- a) TC: BHU is founded by Madan Mohan Malviya.
- b) HT: Madan Mohan Malviya founded BHU.
- c) TC: Madan Mohan Malviya founded BHU in Banaras near Ganga River.
- d) HT: Madan Mohan Malviya founded BHU.

With the help of given example of Partial Textual Entailment is provided by Sentence (c) and (d) due to the fact that some data like Banaras, Ganga River is absent. Following, such as to acquaint PTE and indorse a technique to convey the process of distinguishing entailment among text with the help of PTE.

Study of PTE

The basis of our endeavour is founded on the belief of identifying Partial Textual Entailment. It is requisite to present Faceted PTE [1] to advance an arrangement for the same. A completely programmed machine of Faceted Partial Textual Entailment is planned via the Authors. Nielsen et al.[21] proposed the notion of Partial Textual Entailment. Levy et al. [1] further elaborated the concept in the method of the faceted PTE [1]. As a explanation to the inflexibility and frailness of the process of Textual Entailment the concept of PTE has been planned.

TC: The actual work of doctors is to save lives.

HT: The doctor saves life.

Among the text(TC) and hypothesis(HT) a relationship could be seen between the pair of sentences so to specify that there is entailment amongst the both of them is not right. The relationship among them still exists and is labelled as PTE such that the text is involved as a part of Hypothesis.

The ordered pair w_1, w_2 where w_1 and w_2 characterize 53 words that harmonize to nouns, pronouns, verbs or adjectives is defined by Neilsen et al. [21] in the form of idea of facets. The relationship between the ordered pair is said to be semantic and also belong to the hypothesis. The notion of applying faceted Partial Textual Entailment- on the whole the ordered pair is extended by Levy et al. [1] by depriving any semantic relationship between them. The resulting three phase structure for PTE is suggested by Levy et al. [1].

A. The Hypothesis is degraded into Facets

From the hypothesis the aspects are produced in the process of well-organized pairs.

B. Determining the Entailment of Facets

To determine the entailment between them the facets of the hypothesis are produced in the form of well-organized pairs.

C. Combine the Distinct Facet Outcomes and Selecting the Complete Entailment Consequently

The accumulated result of complete matching of aspects of Hypothesis and text conclude in comprehensive entailment.

By means of exact match technique, the notion of faceted PTE is put forward, the lexical inference technique and the semantic technique in related work [7]. Following, the given approaches are succinctly discussed by the authors. Exact Match Method: In the sack of words all the tokens or lemma of H are confined in. The decision is said to be positive if the lemma or tokens of T both existing in container. When TC involves HT a positive decision is implied. In a quantity of textual entailment challenges this approach is frequently used.

Lexical Inference Techniques: Using the Resnik Comparison Measure the resemblance between well-ordered pair of TC and HT is projected and calculated.

A model created on word intersection amongst the text and the hypothesis [22] is represented by a lexical based system. We know that the system is not efficient as it has been reported by the Authors and a web-established word resemblance technique is given that uses lexical edit distance for evaluation [23].

Semantic Inference Techniques: If the mental for of a given sentence requires that of alternative process of Semantic entailment used [24]. The generally used device for identifying entailment is BIUTEE [6] in which the dependence tree consisting of facets for H and T is formed. The rate of producing entailment between the two H from T is determined with the help of path among the ordered pair. Our approaches will be verified for results using the given methods and their different configurations.

By generating pairs of the words PTE uses facets generation. Due to this a great quantity of faceted pairs and calculations are generated. This major limitation is faced by existing methods. To solve the following problems a new belief is anticipated which will gradually be transformed into a employed procedure:

- 1) A huge quantity of facets is generated which intensifies the quantity of computing energy. The Authors were inspired by this causing them to suggest a answer using facets.
- 2) A lot of computing is used to generate a sentiment with the use of existing methods of sentiment analysis because the reviews are cumulating daily. The processing and examination of text/terminologies further decreased due to the incorporation of hefty linguistic reserves such as entity acknowledgment, stemmers and many more. SA is done using PTE to tackle this problem.

In the present paper the main aim is:

- 1) For the formation of facets in PTE a new modification is proposed. This method is performed over prevailing technique of facets formation by Levy et al. [1] to deliver effective results and decreased computation and also to reduce the number of facets.
- 2) As a outcome of Partial Textual Entailment, number of text/expressions is reduced, the sentiment analyser is also given with lesser expressions. The examination of a minor set of text reduces the burden on Sentiment Analyser instead of the model huge set of terminologies. The processing time is considerably reduced due to this.
- 3) Finally, for each kind of sentiment the total percentage can be determined.

To the finest of our awareness, for sentiment analysis Partial Textual Entailment is not utilized till now. Therefore, the idea is:

- 1) It is also used for implementing the problem of faceted Partial Textual Entailment.
- 2) For the assembling tweets with comparable version Partial Textual Entailment is used.

- 3) From the tweets derived from Partial Textual Entailment it is used for applying a theoretical method.
- 4) With the help of existing methods, the proposed method will also be checked.

The projected alteration for faceted PTE is conferred in the next session. The stages for formation of factor and the method of faceted PTE is also discussed.

4. Proposed Methodology

With the help of the method proposed by the Authors the task of carrying around Sentiment Analysis can be fragmented into two sub-tasks:- (a) a technique is established for identifying PTE and (b) A technique of Opinion Mining with the help of PTE has been established.

A. Establishing a technique to Recognize PTE

The anticipated proposal is to advance a prototype to diminish computing with the help of huge number of facets creation and to produce effective outcomes. This is performed with the help of eliminating accommodating words. With the help of the theme of sentence as a main term in the formation of facets the facets will be created. As the text sort out primarily around the theme of the sentence this technique is anticipated to predict improved Textual Entailment. Consider,

A: Data

B: Set of words in the data A

C: Subject of the data

D: Characteristic of the data

1) Choosing Text and Hypothesis:

With the help of the reviews the text and hypothesis are selected.

Example,

Text: Rajesh goes to college daily. He studies diligently.

Hypothesis: Rajesh gets excellent grades in his test.

2) Modification of the Text and Hypothesis through eliminating assisting words such as "to", "gets", "his".

The assisting words such as halting words, prepositions are detached.

Text and hypothesis are altered as

X (T): {Rajesh, goes, college, daily, He, studies, diligently}

X (H): {Rajesh, gets, excellent, grades, his, test}

3) Discerning of the Subject where Q signifies the subject and R signifies the residual words

The residual words are listed down distinctly and the subjects is designated from the sentences.

Q (T): {Rajesh, studies}

R (T): {goes, college, daily, He, diligently}

Q (H): {Rajesh}

R (H): {gets, excellent, grades, his, test}

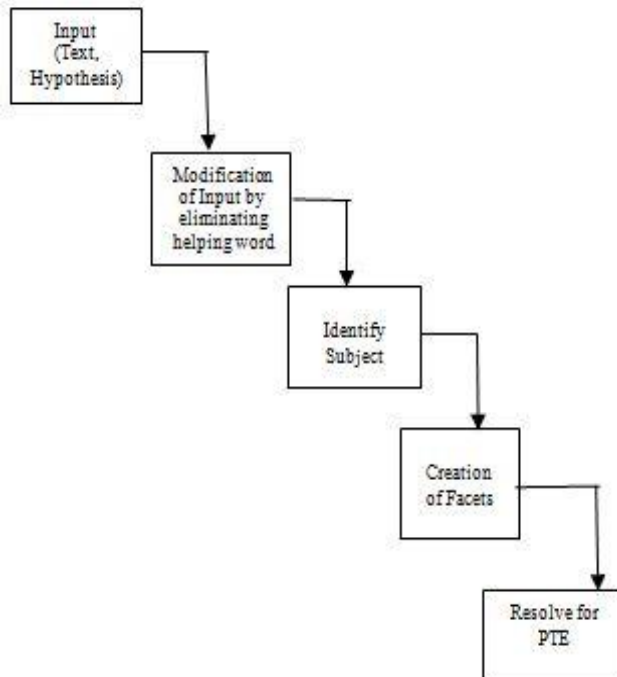


Fig. 1- Partial Textual Entailment Flow Diagram

4) Creation of the facets:

The residual words are paired with the subjects for the generation of facets:

Facet (TC): $Q^*R = \{ \text{Rajesh, goes} \}, \{ \text{Rajesh, college} \}, \{ \text{studies, diligently} \} \dots$ Total 12 facets

Facet (HT): $Q^*R = \{ \text{Rajesh, gets} \}, \{ \text{Rajesh, excellent} \}, \{ \text{Rajesh, grades} \} \dots$ Total 5 facets

5) Calculation of PTE:

The number of facets gets diminished greatly in the last step as it is detected by the authors. Overall number of facets produced is equivalent to $60(12 \times 5 = 60)$. This will help in reducing the number of assessments and thus helps in plummeting the vital computation. The entire procedure is applied to the tweets due to which the outcome will be numerable dis-similar sentences which could be simply passed through the Opinion Mining to regulate the opinion of the group of tweets, evaluations or terminologies. How strongly people approve or dis-approve to a subject will be determined by the amount of text inputs linked to a sentiment. The flow of work is discussed in the next section in other words with

the help of Partial Textual Entailment a complete architecture of Sentiment Analysis.

B. Creating a technique for Sentiment Analysis with the help of Partial Textual Entailment

The complete framework for OM using PTE below is discussed by the authors

- 1) Distinguishing innovative prototype for PTE:
This way will achieve the task of assembling like tweets so as to decrease the burden of the Opinion Mining ensuing in quicker processing of evaluations as conversed in Section 4.1. The entailment acknowledgement is done with the help of BIUTEE tool [6].
- 2) The outcome found from phase 1 is passed to the opinion miner to identify the opinions of the text/terminologies.
- 3) The overall quantity of sentences which will involve the hypothesis is then delivered as a figure of how sturdy the opinion is.

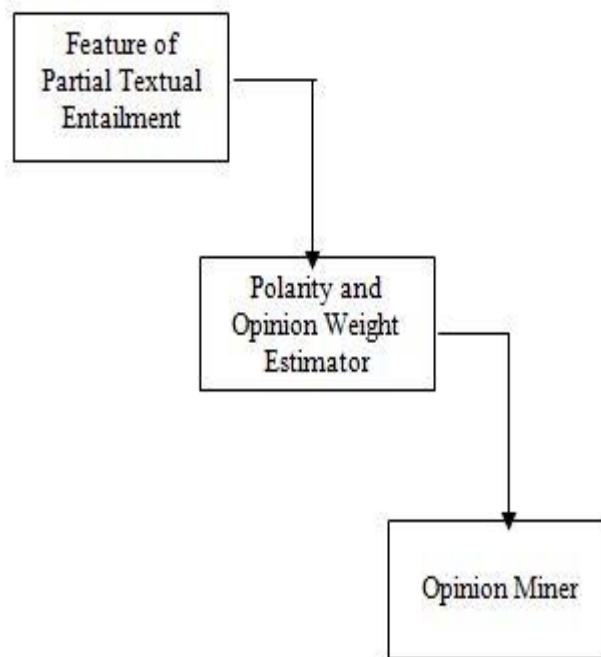


Fig. 2. Processing step of Opinion Miner

In this architecture define a procedure of sentiment analysis of partial textual entailment first we will perform the job of feature of partial textual entailment and after that set the polarity and sentiment weight estimator for sentiment analysis purpose . The Fig. 2 representing the complete model working step for better understanding of researchers and programmers.

5. Conclusion

This technique of Opinion Mining using Partial Textual Entailment is proposed by the authors within the current paper. To the high-quality of our understanding, it is used for lowering the computational overhead for the first time in this research. The researchers proposed to put in force the anticipated approach of Partial Textual Entailment that's mentioned and described Section IV. In destiny, the method of Sentiment evaluation the usage of the anticipated approach of Partial Textual Entailment is put into effect by the Authors. The authors additionally proposed to apply the same old facts units to calculate the effectiveness in their anticipated approach in destiny. BIUTEE tool is used for the mission of entailment popularity by the researcher. To spread this work for diverse territorial languages of the country is the aim of the Researchers.

References

- [1] O. Levy, T. Zesch, I. Dagan, I. Gurevych. "Recognizing Partial Textual En-tailment" Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Vol 2, pp. 451-455, 2013
- [2] <https://www.lexalytics.com/technology/sentiment>
- [3] H. Kaur, V. Mangat, Nidhi "A survey of sentiment analysis techniques" I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (ISMAL), In-ternational Conference on IEEE, 2017.
- [4] I. Guellil, K. Boukhalifa, "Social big data mining: A survey focused on opinion mining and sentiments analysis" Programming and Systems (ISPS), 12th International Symposium on IEEE, 2015
- [5] K. Mouthami, K. Nirmala Devi, V. Murali Bhaskaran. "Sentiment analysis and classification based on textual reviews" Information communication and embedded systems (ICICES), international conference on IEEE, 2013
- [6] A. Stern, and I. Dagan "The BIUTEE research platform for transformation-based Textual Entailment recognition" LiLT (Linguistic Issues in Language Technology), Vol 9, 2014.
- [7] E. Lloret, O. Ferr´andez, R. Mu˜noz and M. Palomar, "A Text Summarization Approach under the Influence of Textual Entailment" Natural Language Processing and Information Systems Group, pp: 22-31, June 2008
- [8] M H. Haggag, Marwa M.A. ELFattah, A M. Ahmed "Different Models and Approaches of Textual Entailment Recognition" International Journal of Computer Applications. Vol 142(1), 2016.
- [9] I. Dagan, B. Dolan, B. Magnini And D. Roth, "Recognizing Textual Entailment: Rational, evaluation and approaches-erratum" Natural Language Engineering, Vol 16(1) pp: 105-105, 2010
- [10] B. Pang, L. Lee, S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques" In Proceedings of the ACL-02 conference on Empirical methods in natural language processing, Vol 10 pp:79-86, 2004.
- [11] P. D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews" In Proceedings of the 40th annual meeting on association for computational linguistics, pp: 417-424.

- [12] E. Cambria, B. Schuller, Y. Xia, Catherine, “New avenues in opinion mining and sentiment analysis” *IEEE Intelligent Systems*, Vol 28(2) pp: 15–21, 2013.
- [13] M. Wöllmer, F. Wenginger, T. Knaup, and B. Schuller, C. Sun, K. Sagae and L-P Morency. 2013. YouTube Movie
- [14] M. Hu and B. Liu, “Mining and summarizing customer reviews” In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp: 168–177, 2004.
- [15] K. Dave, S. Lawrence, David, “Mining the peanut gallery: Opinion extraction and semantic classification of product reviews” In *Proceedings of the 12th international conference on World Wide Web*, pp: 519–528, 2003.
- [16] T. Wilson , J. Wiebe , P. Hoffmann, “Recognizing contextual polarity in phrase-level sentiment analysis” In *Proceedings of the conference on human language technology and empirical methods in natural language processing*, pp: 347–354.
- [17] A-M Popescu, O. Etzioni, “Extracting product features and opinions from reviews” In *Natural language processing and text mining*, Springer, pp: 9–28.
- [18] P. D. Turney, M. L. Littmanp, “Measuring praise and criticism: Inference of semantic orientation from association” *ACM Transactions on Information Systems (TOIS)*, Vol 21(4) pp: 315–346.
- [19] M. Vita. “Computing Semantic Textual Similarity Based on Partial Textual Entailment”.
- [20] <http://ama.liglab.fr/GreenSeminars/2015/Speechs/Diana.pdf>
- [21] R. D Nielsen, W. Ward, and J. H Martin, “Recognizing entailment in intelligent tutoring systems”. *Natural Language Engineering*, Vol 15(4) pp:479–501, 2009
- [22] D. Majumdar, P. Bhattacharyya, “Lexical Based Text Entailment System for Main Task of RTE6” *Text Analysis Conference (TAC) Proceedings*, 2010.
- [23] R. Adams “Textual Entailment through extended lexical overlap” In *Proceedings of the Second PASCAL Challenges Workshop on Recognising Textual Entailment*, pp: 128-133, 2006.
- [24] R. S. Braz, V. Punyakanok, D. Roth, M. Sammons. “An inference model for semantic entailment in natural language” *Machine Learning Challenges. Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Textual Entailment*. Springer, pp: 261-286, 2006.
- [25] Rinatha, K., Suryasa, W., & Kartika, L. G. S. (2018). Comparative Analysis of String Similarity on Dynamic Query Suggestions. In *2018 Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS)* (pp. 399-404). IEEE.
- [26] Rahma, D. Y., & Atmaja, M. H. S. (2022). Peritoneal carcinomatosis and mimicking on CT scan findings. *International Journal of Health & Medical Sciences*, 5(1), 154-162. <https://doi.org/10.21744/ijhms.v5n1.1864>