

# Efficient Notes Generation through Information Extraction

C. Nalini, Shwetambari Kharabe, Sangeetha S



**Abstract:** For a specific topic we find several reference books and textbooks to refer and the content is much more than one needs, therefore at the time of revision or quick study we need the summary and short version of the whole content for revision, especially at the time of examinations.

There we expect to build up a semi-mechanized method to produce notes from English content records like Reference Books and Text books. The method talked about is viewed as a spearheading endeavor in the field of NLP (Natural Language Processing). This strategy has a wide scope in the instructive space. The procedure when executed as an application can be utilized by both employees and understudies.

**Keywords :**NLP, Segmentation ,Parsing, Ontology

## I. INTRODUCTION

Here we talk about condensing a content report into progressively reasonable archive with less substance which covers a large portion of the vital focuses and gives a reasonable thought regarding the entire record, similar to a synopsis[19],[21],[23].

For the most part, in the vast majority of the course readings, just 20-22% of the words contain the data you have to comprehend or require while overhauling the theme. They are known as watchwords. Furthermore, the staying 80% contains near no fundamental data in that capacity, which comprises of pronouns, connectives like "of", "has", "for", and so forth. The main motivation behind these words is to connect the catchphrases together to frame significant and reasonable sentences[1]. They are valuable for first time perusing, yet for amendments, simply the watchwords can take the necessary steps effectively, so overhauling only a little measure of content is much useful as perusing the entire content once more.

We achieve this by using various methodologies and tools like Parsing, ontology creation, segmentation etc which are discussed further[20],[22],[24]. For a given text document like a reference book we find too much information on a topic which is more than one needs or can handle, to tackle this problem we can make notes and study a topic, here we try to

**Manuscript published on 30 August 2019.**

\* Correspondence Author (s)

**C.Nalini,** Department of Computer science and Engineering, Bharath Institute of Higher Education and Research, Chennai, Tamilnadu, India.  
Email: nalinicha2002@gmail.com

**Shwetambari Kharabe,** Department of Computer science and Engineering, Bharath Institute of Higher Education and Research, Chennai, Tamilnadu, India. Email: shweta.chiwhane@gmail.com

**Sangeetha S,** Department of Computer science and Engineering, Bharath Institute of Higher Education and Research, Chennai, Tamilnadu, India.Email: sangeethasathy01@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](#) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

accomplish the generation of notes automatically with abstractive approach[2]. This application can be used to get a gist of a given document and can be used for quick study like revision and omit the information from a document which is not much useful.

To accomplish larger amount of precision in [3] archive order increasingly educational highlights of reports are considered. For this reason, for example, weight is doled out to HTML labels, which influences the proficiency of data recovery; these are characterized utilizing hereditary calculations. Archives arrangement happens at the degree of

Separate words, unlike old style works, the importance of each word here is characterized in connection to their useful highlights, which are the events of a word in the title, accentuation is given on words by methods for italic, intense text styles or its underlining and position of a word on the page. A DIG (Document Index Graph) calculation depends on diagram hypothesis and expressions and their loads are considered for making proposals in work [3].

In this methodology for content segmentation,[5] we utilize a productive direct message division calculation (called TSHAC). It thinks about both computational intricacy and division precision. The procedure of TSHAC has 4 stages. To begin with, Preprocessing of long message; tokenization, stopwords are evacuated, and stemming are led to build the jargon of the content. Content is then spoken to as vector after content preprocessing, every one of which speaks to a sentence inside the content. A piece of sentence similitudes are then processed to build the sentence-likeness lattice. At long last, a the ideal point limits are distinguished by the proposed algorithm.[5]

The strategy for extraction of significant sentences from the content to synopsis dependent on meaning of score of pertinence for every one of the sentences and called "sentence by sentence" is recommended in this segment[14],[16],[18].

The procedure to be utilized is space autonomous which makes it special from different strategies. From the assessment estimates applied to the method, we can say that the strategy helps in semi-computerized age of notes. The presentation thoroughly relies upon the cosmology gave as information. The more exact the philosophy creation, the more precise the yield will be[25],[27],[29].



Published By:

Blue Eyes Intelligence Engineering

and Sciences Publication (BEIESP)

© Copyright: All rights reserved.

## II. METHODOLOGY

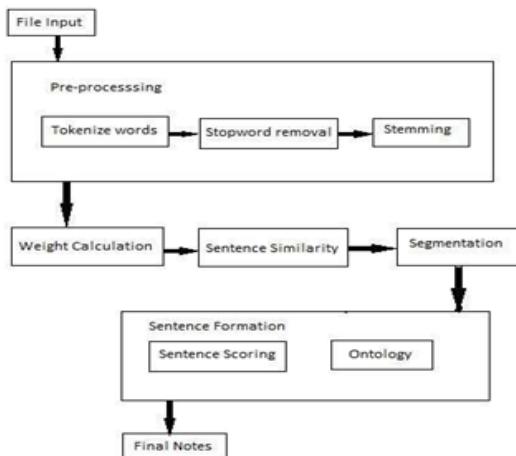


Figure 1 – Architecture

### • File Input

Input the file that is to be read and analyzed.

### • Pre-processing

In pre-processing module there three main activities that take place namely (i) Word Tokenization (ii) Stopword Removal (iii) Stemming. In Word Tokenization we separate the input documents into separate unique individual words and sentences which are formerly known as tokens. For this we have used the NLTK .tokenize library.

In the Stopword Removal phase we segregate the stopwords from the tokens using the NLTK. corpus library. And in the stemming phase we bring down the words to their root phase. Eg: using will be stemmed as use[8],[10],[12].

### A. Weight Calculation

$$w_{ij} = \frac{\sqrt{termFreq_{ij}}}{\sqrt{tokenCount_j}} \left( \log \left( \frac{docCount}{docFreq_i + 1} \right) + 1 \right)$$

Using the above we calculate the weights of each individual words and store the value in the database.

### B. Sentence Similarity

In this module we compare every sentence with its consecutive three sentences. We calculate the similarity function of the sentences using the Euclidian distance formula. We have used the SK learn library for this.

### C. Segmentation

In the segmentation module we club the similar sentences together within an array. This is done by using the value of the similarity function of the sentences.

### D. Sentence Formation

In sentence formation module we score the sentences based on the value of similarity function and the weights. The sentences above the threshold value are then selected. And then using ontology we make the sentence meaningful by connecting the sentences. This whole process makes the sentences extractive[13],[15],[17].

### E. Notes

This is the final output of the system. it is the summary of the input document.

## III. LITERATURE REVIEW

Here we discussed the literature review of existing techniques:

Okul Prasad, Harish Mathivanan, Madan, Jayaprakasam, T.V.Geetha [1] they propose a semi automated technique that generates slide presentation from a text document which is given as an input. A Model is proposed which contains 13 modules after implementing which we get a slide presentation using the bullet points extracted from the input document.

Alguliev, R.M, Aliguliyev, R.M, [2] propose a text summarization method that creates text summary by assigning relevant score to each sentence and extracting sentences from the original documents. While summarization this method considers weight of each sentence in the document. The relevance score of a sentence is calculated through its comparison with all the other sentences and with the document title by cosine measure.

Ji-Wei Wu, Judy C.R. Tseng, Wen-Nung Tsai [3] propose an efficient linear text segmentation algorithm based on hierarchical agglomerative clustering. The proposed linear text segmentation algorithm is implemented without any auxiliary knowledge base, parameter setting, and user involvement. This technique provides comparable segmentation accuracy with several well know linear text segmentation algorithms.

## IV. COMPARITIVE ANALYSIS

Sr. No	Paper Name	Author	Method Proposed	Limitations
1.	Document Summarization and Information Extraction for Generation of Presentation Slides	K.Okul Prasad, Harish Mathivanan, Madan, Jayaprakasam, T.V.Geetha	A semi robotized procedure that produces slide introduction from a context archive which is given as an info.	The proposed system uses tokenizing algorithm for text segmentation which is not the optimum way for text segmentation.
2.	Effective summarization method of text documents	R.M. Alguliev, R.M. Aliguliyev	A cosine-type strategy is adopted to generate summaries by assigning significant scores to each sentence and removing sentences from the first reports.	The process of <del>summarization</del> identification could be <del>unreliable</del> efficient <del>and fast</del> .
3.	An Efficient Linear Text Segmentation Algorithm Using Hierarchical Agglomerative Clustering	Ji-Wei Wu, Judy C.R. Tseng, Wen-Nung Tsai	The proposed direct message division can be achieved with no auxiliary learning base, parameter setting, and client contribution. This uses an effective direct message division calculation utilizing Hierarchical Agglomerative Clustering.	The algorithm can be applied in a much better way so that it can tackle more real life problems during difficult situations and perform efficiently.
4.	Creation of Ontology in EducationDomain	Ayesha Arora, Kishore, Rabena Khan, B.PadmaKali	In this paper they have created course ontology based on <del>various</del> courses offered by the university and added appropriate properties and relations. They have proposed various properties based on which ontologies can be prepared and how <del>these properties</del> can be chosen.	A generic approach can be proposed to ontology creation. Using the semantics and the linguistics of the data.
5.	Text Summarization using Sentence Scoring Method	T. Sri Rama Saja, Deekshith Aluru	This paper discusses the basic and extractive techniques of text summarization. It also proposes various methods for sentence scoring.	The paper proposes a linguistic approach for sentence scoring. Higher accuracy can be obtained by taking into consideration the context of the document.

## V. CONCLUSION

The proposed procedure is area autonomous, so we can concentrate notes from any report with no assistant information base which makes it remarkable from different strategies[7],[9],[11]. From all the assessment estimates applied to the method, we presume that the procedure produces notes utilizing a semi robotized notes age framework. The presentation absolutely subject to the cosmology gave as info. The more precise the metaphysics, the more exact the yield or the produced notes will be.



## REFERENCES

1. Gowri Sankaran, B., Karthik, B. & Vijayaragavan, S.P. 2019, "Weight ward change region plummeting change for square based image huffman coding", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 10, pp. 4313-4316.
2. Gowri Sankaran, B., Karthik, B. & Vijayaragavan, S.P. 2019, "Image compression utilizing wavelet transform", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 10, pp. 4305-4308.
3. Kandavel, N. & Kumaravel, A. 2019, "Offloading computation for efficient energy in mobile cloud computing", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 10, pp. 4317-4320.
4. Vinoth, V.V. & Kanniga, E. 2019, "Reversible data hiding in encrypting images-an system", International Journal of Engineering and Advanced Technology, vol. 8, no. 6, pp. 3051-3053.
5. Selvapriya, B. & Raghu, B. 2019, "Pseudocoloring of medical images: A research", International Journal of Engineering and Advanced Technology, vol. 8, no. 6, pp. 3712-3716.
6. Senthil Kumar, K. & Muthukumaravel, A. 2019, "Bi-objective constraint and hybrid optimizer for the test case prioritization", International Journal of Engineering and Advanced Technology, vol. 8, no. 6, pp. 3436-3448.
7. Kavitha, G., Priya, N., Anuradha, C. & Pothumani, S. 2019, "Read-write, peer-to-peer algorithms for the location-identity split", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9 Special Issue 3, pp. 445-447.
8. Kaliyamurthie, K.P., Michael, G., Anuratha, C. & Sundaraj, B. 2019, "Certain improvements in alzheimer disease classification using novel fuzzy c means clustering for image segmentation", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9 Special Issue 3, pp. 599-604.
9. Kaliyamurthie, K.P., Sundaraj, B., Geo, A.V.A. & Michael, G. 2019, "RIB: Analysis of I/O automata", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9 Special Issue 3, pp. 1019-1022.
10. Velvizhi, R., Rajabhusanam, C. & Vidhya, S.R.S. 2019, "Opinion mining for travel route recommendation using Social Media Networks (Twitter)", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9 Special Issue 3, pp. 508-512.
11. Kavitha, R., Sangeetha, S. & Varghese, A.G. 2019, "Human activity patterns in big data for healthcare applications", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9 Special Issue 3, pp. 1101-1103.
12. Pothumani, S., Anandam, A.K., Sharma, N. & Franklin, S. 2019, "Extended VEOT framework - Implemented in a smart boutique", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9 Special Issue 3, pp. 762-767.
13. Kaliyamurthie, K.P., Michael, G., Krishnan, R.M.V. & Sundaraj, B. 2019, "Pseudorandom techniques for the internet", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9 Special Issue 3, pp. 915-918.
14. Aravindasamy, R., Jeffrin Rajan, M., Rama, A. & Kavitha, P. 2019, "Deep learning provisions in the matlab: Focus on CNN facility", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9 Special Issue 3, pp. 990-994.
15. Theivasigamani, S., Linda, M. & Amudha, S. 2019, "Object sensing and its identification & motion sensing", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9 Special Issue 3, pp. 545-549.
16. Mary Linda, I., Vimala, D. & Shanmuga Priya, K. 2019, "A methodology for the emulation of IPv4", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9 Special Issue 3, pp. 848-852.
17. Velvizhi, R., Priya, D.J., Vimala, D. & Linda, I.M. 2019, "Increased routing algorithm for mobile adhoc networks", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9 Special Issue 3, pp. 1606-1608.
18. Sangeetha, S., Anuradha, C. & Priya, N. 2019, "DNS in real world", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9 Special Issue 3, pp. 937-940.
19. Geetha, C., Vimala, D. & Priya, K.S. 2019, "Constructing multi-processors and spreadsheets with SKIVE", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9 Special Issue 3, pp. 516-519.
20. Yugendar, K., Sugumar, V. & Kavitha, P. 2019, "A novel method of univac using fuzzy logic", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9 Special Issue 3, pp. 435-437.
21. Kaliyamurthie, K.P., Michael, G., Elankavi, R. & Jijo, S.A. 2019, "Implementing aggregate-key for sharing data in cloud environment using cryptographic encryption", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9 Special Issue 3, pp. 957-959.
22. Jeffrin Rajan, M., Aravindasamy, R., Kavitha, P. & Rama, A. 2019, "A novel method of object orientation variation in C++ and java", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9 Special Issue 3, pp. 708-710.
23. Nayak, R., Dinesh, S. & Thirunavukkarasu, S. 2019, "A novel method improvement of rapid miner for the data mining applications", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9 Special Issue 3, pp. 457-460.
24. Sivaraman, K., Krishnan, R.M.V., Sundarraj, B. & Sri Gowtham, S. 2019, "Network failure detection and diagnosis by analyzing syslog and SNS data: Applying big data analysis to network operations", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9 Special Issue 3, pp. 883-887.
25. Vimala, D., Linda, I.M. & Priya, K.S. 2019, "Decoupling online algorithms from erasure coding in DNS", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9 Special Issue 3, pp. 950-953.
26. Rama, A., Kumaravel, A. & Nalini, C. 2019, "Preprocessing medical images for classification using deep learning techniques", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9 Special Issue 3, pp. 711-716.
27. Sangeetha, S., Srividhya, S.R., Anita Davamani, K. & Amudha, S. 2019, "A procedure for avoid overrun error in universal synchronous asynchronous receiver transmitter (uart) by utilizing dummy join and interrupt latency method", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9 Special Issue 3, pp. 657-660.
28. Aravindasamy, R., Jeyapriya, D., Sundarajan, B. & Sangeetha, S. 2019, "Data duplication in cloud for optimal performance and security", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9 Special Issue 3, pp. 1156-1158.
29. Aravindasamy, R., Jeffrin Rajan, M., Sugumar, V. & Kavitha, P. 2019, "A novel method on developing superblocks and the transistor using apodryal", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9 Special Issue 3, pp. 982-985.
30. Sasikumar, C.S. & Kumaravel, A. 2019, "E-learning attributes selection through rough set theory and data mining", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 10, pp. 3920-3924.

## AUTHORS PROFILE



**C.Nalini** Professor, Department of Computer science and Engineering, Bharath Institute of Higher Education and Research, Chennai, India



**Shwetambari Kharabe** Research Scholar, Department of Computer science and Engineering, Bharath Institute of Higher Education and Research, Chennai, India



**Sangeetha S** Assistant Professor, Department of Computer science and Engineering, Bharath Institute of Higher Education and Research, Chennai,



Published By:

Blue Eyes Intelligence Engineering

and Sciences Publication (BEIESP)

© Copyright: All rights reserved.