Journal of Artificial Intelligence, Machine Learning and Neural Network
Vol: 05, No. 01, Jan-June 2025
ISSN: 2799-1172, DOI: 10.55529/jaimlnn.51.40.51

Research Paper

Development of a content summary model for an effective synoptic generation system

Henry Onyebuchukwu Ordu*
*Department of Computer Science, Ignatius Ajuru University of Education, Rumuolumeni, Port Harcourt, Nigeria.

Article Info

Article History:
Received: 13 February 2025
Revised: 22 April 2025
Accepted: 01 May 2025
Published: 17 June 2025

Keywords:
Extractive Summary
Content
Content Summary
Summary
Generation
Synoptic Generation System

ABSTRACT

The exponential growth of digital content on the internet and organizational repositories has significantly increased the demand for systems that can efficiently summarize large text documents. In academia, governance, business, and journalism, decision-makers are often overwhelmed with extensive textual data, making it essential to extract relevant summaries automatically. This study aims to develop a content summary for an effective synoptic generation system. In the modeling process, the formulation of a CSM is undertaken using an extractive summary technique. The CSM is designed and implemented using the Python programming language as the front-end engine and MySQL server as the back-end engine. The study reviews theories and related literature and formulates a CSM for an effective content summary generation using the following features: word frequency, maximum word frequency, sentence weight, normalized word frequency, maximum sentence weight, and compression rate. The CSM is tested using a dataset provided by Document Understanding Conferences. The CSM is evaluated for robustness and reliability using precision, recall and F-measure. The study achieved significant outcomes with 87% precision, 90% Recall and 86% F1-score in values demonstrating the model's effectiveness and reliability in content summary generation.

Corresponding Author:
Henry Onyebuchukwu Ordu
Department of Computer Science, Ignatius Ajuru University of Education, Rumuolumeni, Port Harcourt, Nigeria.
Email: henry.ordu@iaue.edu.ng

1. INTRODUCTION

Advances in computer technology have altered the way we create and publish textual materials by making it easier to generate document content. This has given rise to volumes of data growing rapidly. Text data takes a long time to read and understand, and very many people don't want to spend a lot of time reading unabridged kinds of literature. Understandably, standard works in a textual format usually contain large amounts of text, and reading through such work is not only a difficult process but also a long process. The situation has created an attractive need to provide an effective means to deal with information overload, which is currently very high.

Synoptic generation system is an essential and effective tool by which this difficulty can be solved reasonably. The synthesis process leads to the creation of succinct summaries, while also ensuring that the important source content of the document is preserved. Long files including journals, novels, and books, or brief files including emails, information articles, and speech scripts are a few examples of textual content genres. Each style has its features and personal difficulties.

In recent years, content summaries have proved to be useful in a variety of ways for various professionals. Some professions need primarily reading abilities, and these individuals have gained immensely from the content summary method. One of the fields where content summary plays a significant role is journalism. Journalists must frequently listen to a large number of speeches and consult volumes of books and articles to create their content for the next day's news presentations. Also in the academic community, Students and lecturers like Journalists might want to know the content of some research papers or journals. Moreso, Movie reviewers are becoming more prevalent in both social media and print media these days. Anyone who wants to watch a movie can now read blogs written by reviewers who have written about their experiences with the film. Lawyers, like academics and other professionals, are also spending a great deal of time reading all judgment copies because the number of judgments issued each day is increasing. Hence, Lawyers, like very many professionals find content summary systems very useful and relevant. A summary can be defined as "a text that is produced from one or more texts, that conveys important information in the original text (s), and that is no longer than half of the original text and usually significantly less than that. It is an important way of finding relevant information precisely in large text in a shorter time with little effort. In other words, it is a representation of text in a precise form that aims at reducing the time spent to derive the essence of the document along with maintaining the semantics of the given document.

Therefore, Content Summary is the process of producing a condensed form of the original content for human consumption. The process of content summarization mainly focuses on two issues; the problem of selecting the most relevant portions of the text and the issue of generating coherent summaries for better readability [1]. Moreso, topic identification or text representation, interpretation or compaction or summary representation, and summary creation are the three stages of the content summary generation process. Identifying the core subject, followed by interpretation of the meaning; distinguishing significant and irrelevant material, and finally compacting it into a condensed form in summary representation are the first steps in transforming a text content into a source representation. Content summary, Information retrieval, text summarization, news suggestion, question answering, intelligent analysis, and public opinion monitoring, are a few areas where Natural Language Processing (NLP) is applied. Natural Language Processing (NLP) is a branch of computer science that involves teaching a computer to comprehend, analyze, and extract meaning from human language in a systematic manner. In terms of technology, it combines three disciplines: artificial intelligence, computer science, and computational linguistics. The purpose of NLP research is to develop detailed computational models of language so that computer programs can be created to do diverse tasks.

Today, there are vast amount of documents, articles, papers, and reports available in digital form, but most of them lack content summaries. This has led to a waste of time, money, and other meaningful scarce resources. Therefore, there is a need of providing users with the most appropriate and relevant content. Hence, to effectively deal with the needed information, a system for summarizing appropriate content from given textual material is required. Therefore, this research deals with developing a content

summary model for an effective synoptic generation system. This will be achieved through model formulation using single document input and an extractive content summary output. The Content Summary Model will be implemented using Python programming Language and Natural Language Processing kits.
This research aims to develop a Content Summary Model for an effective synoptic generation system. The objectives of the study are to.

1. Formulate a Content Summary Model (CMS) for an effective synoptic generation system using Extractive Summary Techniques.
2. Design and implement the Content Summary Model using Python Programming Language and MySQL
3. Evaluate the performance of the Content Summary Model (CMS) using Natural Language Processing standard metrics.

The scope of this research is the development of a content summary model for an effective synoptic generation system. The study is limited to the following.

1. It focuses on a Single Document Content Summary (SDCS).
2. It uses an Extractive Content Summary Technique (ECST) to output relevant content.
3. It uses a dataset provided by the Document Understanding Conference (DUC).

Document Understanding Conference (DUC) data set contains a large set of documents with human-created summaries for comparison.

4. It uses Python 3.0 and Natural Language Processing Kit (NLPK) for the development of the proposed synoptic generation system.

It is envisaged that the research work will help potential researchers, including students, lecturers, and organizations, in getting relevant content summaries in the form of a literature review in less time. Other consequent significances of the study are:

1. It will assist print media organizations in shortening text because newspaper columns sometimes have a word limit. With this technique, they can easily lower the content without losing the actual information.
2. The development of this system will also add to the volume of digital resources.

2. RELATED WORK

Relevance Theory

Relevance Theory is a cognitively-oriented pragmatic theory that was originally developed in the 1980s by Dan Sperber and Deirdre Wilson. Relevance theory offers a new approach to the study of human communication based on a general view of human cognitive design. Together with [2] relevance theorists argue that human communication is often intent-based, and therefore. Within this framework, aiming to maximize the relevance of the content one processes is simply a matter of making the most efficient use of the available processing resources. No doubt this is something we would all want to do, given a choice. Relevance theory claims that humans do have an automatic tendency to maximize relevance, not because we have a choice in the matter – we rarely do – but because of the way our cognitive systems have evolved. As a result of constant selection pressures toward increasing efficiency, the human cognitive system has developed in such a way that our perceptual mechanisms tend automatically to pick out potentially relevant stimuli, our memory retrieval mechanisms tend automatically to activate potentially relevant assumptions, and our inferential mechanisms tend spontaneously to process them most productively. This universal tendency is described in the First, or Cognitive, Principle of Relevance [3]. Sperber and Wilson's theoretical framework has become influential in content summarization, information retrieval, and natural language processing research.

The relevance Theory of [3]. Was adopted for this research because the notion of relevance serves as the foundation for the field of content summarization, information retrieval, etc. This theoretical framework could help ensure that most users have a fairly good idea of what relevance is. Content relevance refers to the instrumental value of text information for enabling a reader to meet a reading goal [4], [5]. This differs from Content importance, which generally refers to text elements that are essential for understanding a text's main ideas [6]. Regardless of their relevance to the reader's goals. Content

information that closely matches readers' goals and enables them to meet their goals is judged to be relevant. Conversely, Content information that is tangential or unrelated to readers' goals is judged to be irrelevant. Hence, to build effective synoptic generative systems, we must formalize the theory and concept of relevance. After all, the purpose of a synoptic generation system is to retrieve concise but relevant items in response to user requests.

Information Theory

Information theory is the quantitative study of signal transmission. Primarily applicable to information technology and communications engineering, in human communication theory, it serves primarily as a metaphor for linear transmission between human senders and receivers. Although information theory has historical significance, contemporary theories of human communication rarely refer to it directly. Originating in physics, engineering, and mathematics, the theory addresses uncertainty in code systems, message redundancy, noise, channel capacity, and feedback. This entry defines basic concepts from the field, applies these to language and human communication, and summarizes insights about information transmission [7]. Information is a measure of uncertainty in a system of signals. In a counterintuitive way, information theory states that the higher the information in a system, the greater the uncertainty. This is because more information entails a larger number of states, which decreases clarity. The concept of entropy is the starting place for understanding this seemingly contradictory idea [7].

A key measure in the information theory of Shannon is entropy. Entropy, taken from thermodynamics in physics, is the randomness or lack of predictability within a system. Highly entropic situations have little organization, reduced predictability, and therefore great uncertainty. In low-entropy systems, there is more organization, greater predictability, and therefore less uncertainty Entropy quantifies the amount of uncertainty involved in the value of a random variable or the outcome of a random process. Shannon's concept of entropy (a measure of the maximum possible efficiency of any encoding scheme) can be used to determine the maximum theoretical compression for a given message alphabet. In particular, if the entropy is less than the average length of encoding, data compression is possible. Data compression, also called compaction, is the process of reducing the amount of data needed for the storage or transmission of a given piece of information, typically by the use of encoding techniques. Compression predates digital technology, having been used in Morse code, which assigned the shortest codes to the most common characters, and in telephony, which cuts off high frequencies in voice transmission. Data compression is important in storing information digitally on computer disks and transmitting it over communications networks [8].

This theoretical framework was adopted for this research, because, while the theory has been most helpful in the design of more efficient telecommunication systems using the Data Compression approach. It has also motivated linguistic studies and Natural Language Processing of the relative frequencies of words, the length of words, and the speed of reading. Information theory provides a means for measuring the redundancy or efficiency of symbolic representation within a given language. Hence, Shannon's Theory is relevant to the study, because Shannon also observed that when longer sequences, such as paragraphs, chapters, and whole books, are considered, the entropy decreases and English becomes even more predictable. He considered longer sequences and concluded that the entropy of English is approximately one bit per character [8].

Conceptual Framework

The Concept of Summary

A summary can be loosely defined as a text that is produced from one or more texts, that conveys important information in the original text (s), that is no longer than half of the original text (s), and that is no longer than half of the original text (s) and usually significantly less than that [9], [10] defined summary as a process of identifying salient concepts in text narrative, conceptualizing the relationships that exist among them, and generating concise representations of the input text that preserve the gist of its content. According to [11], a summary is a reductive transformation of a source text into a summary text by extraction or generation.

The Concept of Synopsis

In natural language processing (NLP), a synopsis refers to a brief summary or abstract of a larger piece of text. The goal of a synopsis is to provide a condensed version of the original text, highlighting the most important information and key points in a clear and concise manner. A synopsis can be generated using a variety of NLP techniques, such as text summarization, which involves automatically selecting and combining the most relevant sentences or passages from the original text to form a summary. Other techniques, such as topic modeling and entity recognition, can also be used to identify and highlight the main themes and concepts in the text [12].

In addition, a synopsis can also be useful for machine learning and data analysis applications, where large amounts of text data need to be processed and analyzed. By generating a summary of the key points and themes in a text, NLP algorithms can more easily and accurately identify patterns and trends, and make more informed decisions based on the content. However, it is important to note that generating an accurate and effective synopsis can be a challenging task in NLP, particularly for texts that contain multiple layers of meaning and nuance. In some cases, manual editing or refinement may be necessary to ensure that the synopsis accurately reflects the content and tone of the original text [13].

Despite these challenges, the use of synopses in NLP is becoming increasingly popular and widespread, as more organizations and individuals seek to effectively manage and make sense of large volumes of text data. By leveraging advanced NLP techniques and algorithms, synopses can provide a powerful tool for summarizing and extracting key insights from text, and enabling more efficient and effective decision-making [14].

Synoptic Generation System

A Synoptic Generation System is a type of natural language processing technology that can automatically generate a summary or synopsis of a longer piece of text, such as a news article, research paper, or book. The goal of a Synoptic Generation System is to produce a concise and accurate summary of the key points and main ideas of the original text, while retaining the most important information and preserving the overall meaning and tone [15]. A Synoptic Generation System typically works by analyzing the structure and content of the input text, using techniques such as statistical analysis, machine learning, and natural language processing. It identifies the most important sentences or passages, and combines them to form a coherent summary. The system may also use algorithms to evaluate the quality and readability of the summary, and make adjustments as necessary to improve its clarity and effectiveness [16].

[17] Developed GIS Texter, a system for generating single and multi-document summaries using natural language processing and machine learning. Key steps included dataset selection, feature extraction, and summarization algorithms for sentence ranking. The system produced coherent summaries evaluated with ROUGE metrics. A limitation was the lack of external knowledge integration. Future enhancements could include domain-specific ontologies or pre-trained language models to improve accuracy and contextual understanding in summary generation. [18] Proposed Heter SUM Graph, an extractive summarization model using heterogeneous graph neural networks (HGNNs). It leveraged semantic and structural information for improved sentence ranking. Implemented with deep learning frameworks, the model used supervised learning and was evaluated using ROUGE metrics. Heter SUM Graph showed superior performance over baselines in summary quality. However, dataset bias was a concern, potentially affecting model generalization. Using diverse datasets could reduce this limitation.

[19] Developed MATCHSUM, a neural extractive summarization model for long documents. By utilizing deep learning, the model selected key sentences to produce concise summaries. It was trained on curated datasets and evaluated using ROUGE metrics. MATCHSUM demonstrated improved performance when summarizing lengthy texts. However, dataset bias posed a limitation, potentially affecting accuracy. Future improvements include the use of varied, representative datasets to enhance robustness and reduce summarization bias.

[20] Proposed Ec Forest, an extractive summarization approach that combines enhanced sentence embedding's with a cascade forest algorithm. Leveraging neural networks and deep learning tools, the

model was trained on datasets to rank and aggregate relevant sentences. Evaluated with ROUGE metrics, Ec Forest showed strong performance in producing coherent summaries. A major limitation was dataset bias, which could impact reliability. Addressing this issue requires more diverse, representative training data. [21] Designed a semantic model for extractive multi-document summarization, integrating statistical machine learning and graph-based techniques. The model captured the semantic relationships between sentences and employed graph analysis for summary generation. Implemented with tools like scikit-learn and NLP libraries, it was evaluated using ROUGE and F1 scores. The model yielded high-quality summaries, although dataset bias was identified as a limitation. Mitigating this requires the use of varied, representative datasets.

3. METHODOLOGY

This study utilized Dynamic Systems Development Methodology (DSDM). Dynamic Systems Development Method (DSDM) is an agile software development project delivery framework that provides a structured approach to software development and project management. It is focused on delivering high-quality software within a specific timeframe and budget. DSDM emphasizes the importance of user involvement, iterative development, and incremental delivery. The architecture of the Content Summary Model is shown in Figure 1, and the flowchart of the model I shown in Figure 2.

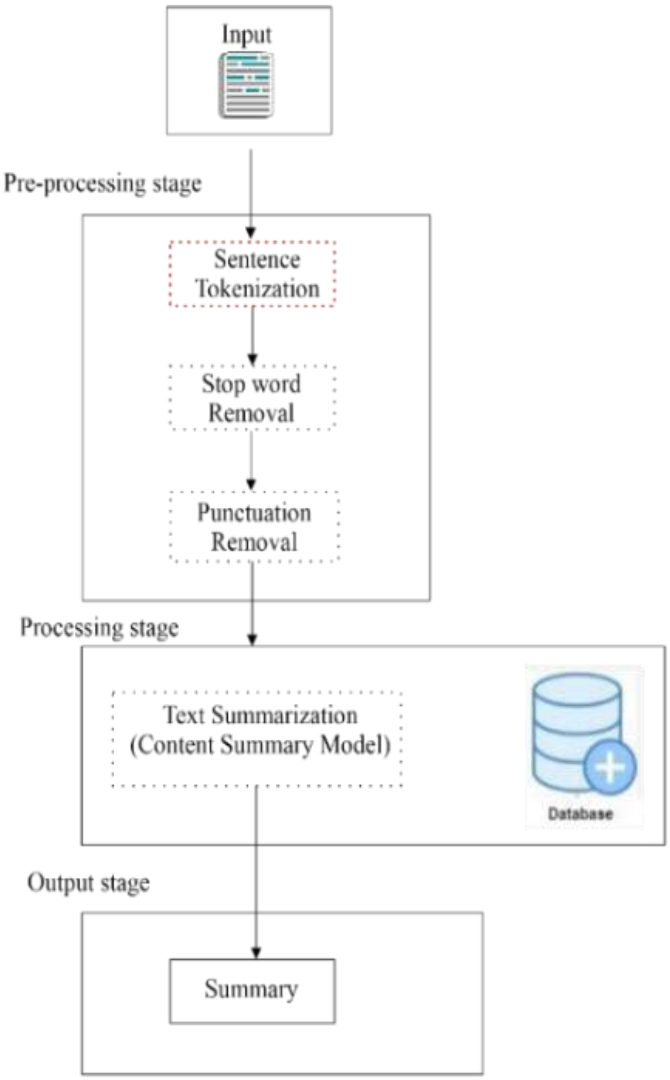

Figure 1. Architecture of the Content Summary Model

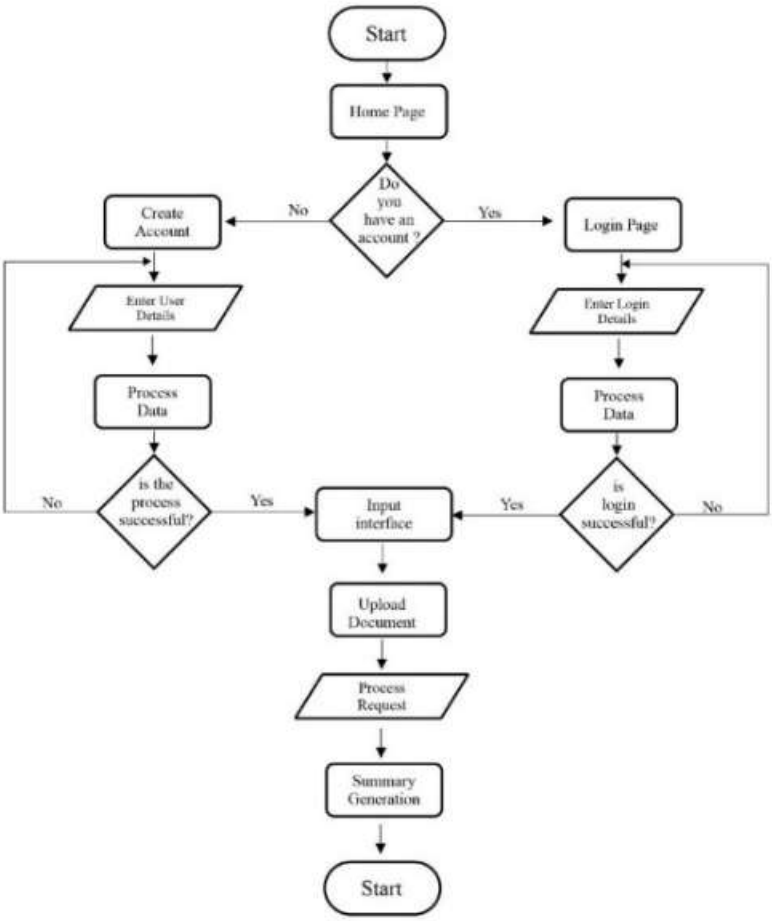

Figure 2. Flowchart of the Content Summary Model (CSM)

Components of the Content Summary Model

Pre-Processing Phase

The preprocessing phase is the procedure of creating the input text documents competent for content features extraction, for an effective generation of the synopsis of a text document. The document is passed as an input in a text format, and should only be in English language format. The input text is divided into sentences based on the sentence terminator. The text-cleaning processing involves the following sub-steps.

Sentence Tokenization

Sentence tokenization is a specific form of tokenization that involves breaking down a text into its constituent sentences. Tokens are the smallest units in language processing, and in this case, sentences are treated as tokens. The result of sentence tokenization is a collection of sentences, each of which can then be further broken down into words or sub-word tokens. This process is pivotal for various text analysis tasks, including sentiment analysis, text summarization, and machine translation. The sentence tokenization takes place in all the sentences. It represents all the distinct keywords and phrases from documents.

Stop Word Removal

Stop words are common words (such as "and," "the," "is," etc.) that are often removed from the text during the text cleaning process to reduce noise and improve the efficiency of text processing algorithms. These words are considered non-informative as they occur frequently across different texts and don't contribute significantly to the meaning of the content. Removing stop words helps in reducing

the dimensionality of the data and can improve the accuracy of tasks like text classification, sentiment analysis, and information retrieval. During the pre-processing stage, a stop word removal process is performed to remove the unimportant and meaningless words from the input documents.

Punctuations Removal

Punctuation marks are non-semantic characters that contribute to the visual and grammatical structure of a sentence but do not carry inherent meaning on their own. Among the essential text cleaning steps, the removal of punctuation marks holds a significant place. Punctuation removal involves the elimination of punctuation symbols such as Period (.), Comma (,), Question Mark (?), Exclamation Mark (!), Semicolon (;), Colon (:), Quotation Marks (" "), Apostrophe ('), Hyphen (-), En Dash (–), Em Dash (—), Parentheses ( ), Brackets [ ], Braces { }, Ellipsis (...), Slash (/), Backslash (), Ampersand (&), At Symbol (@), Percent Sign (%), Dollar Sign ($), Hash/Pound Sign (#), Plus Sign (+), Minus Sign (-), Equal Sign (=), Less Than (<), Greater Than (>), Underscore (_), Vertical Bar/Pipe (|), Tilde (~), New line (n\) etc. By removing these marks, the text is stripped of unnecessary noise, allowing the content summary algorithms to focus on content-bearing words and phrases.

Processing Stage

In this phase, the model takes some of the features of the document into account; a text document is a collection of paragraphs. Each sentence in a paragraph is a collection of words, so individual word scores is playing important roles to calculating the sentence score. The Output of the preprocessed phase is passed as an input to the Content Summary Model for feature extraction and summary generation.

Output Stage

In this phase, the synopsis of the document is generated as an output.

4. RESULTS AND DISCUSSION

The Study Achieved the Following Results.

It Formulated a Content Summary Model (CSM) for Synoptic Generation

The Content Summary Model was successfully formulated using Natural Language Processing Techniques. The CSM focused on Extractive Content Summary, and the model considered factors such as; Word frequency, Normalized Word Frequency, Maximum Word Frequency, Sentence Weight, Maximum Sentence Weight, and Compression rate. The formulated CSM is shown equation 1.

$$CSM = (C_r * S_{total}) * \sum_{i=1}^{S_i} ((W_f)/(MaxW_f)) \quad (1)$$

Where;
S_{total} = Total Number of Sentence in the Document, D
C_r = Compression Rate expressed in percentage (%)
W_f = Word Frequency
N_{wf} = Normalized Word Frequency
$MaxW_f$ = Maximum of Word Frequency
 = for all
S_i = Sentence in ith position

The Study Implemented the CSM

The CSM was implemented using the Python Programming Language and MySQL Relational Database as shown in Figure 3.

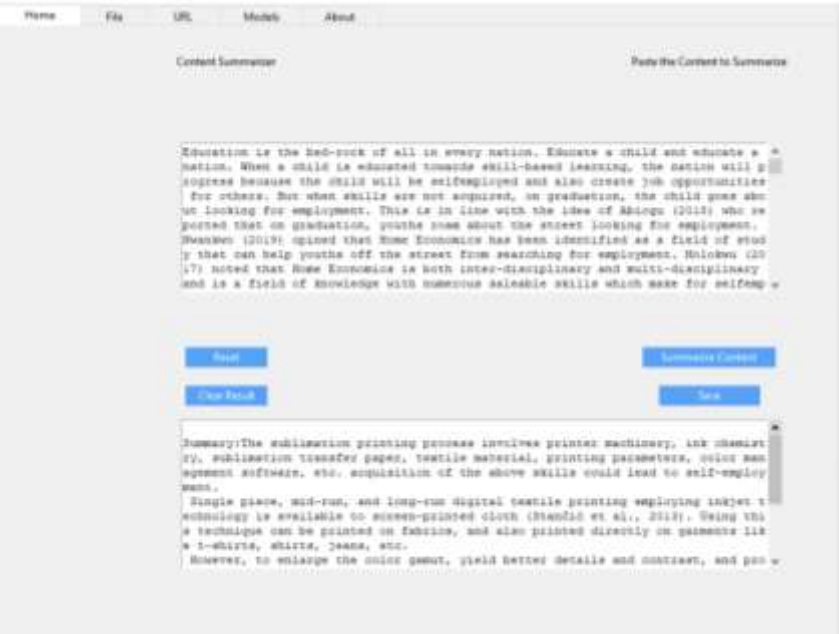

Figure 3. Output of the Content Summary Model

Evaluation of the Content Summary Model

The model was tested and evaluated in terms of Precision, Recall, and F1-Score as shown in Table 1 and

Table *2*.

Metric	Score (Value)
Precision	87.0%
Recall	90.0%
F1-Score	86.0%

Table 1. Test Result of the Content Summary Model (CSM) Across Various Text Documents

Metrics			
Document Title	**Precision**	**Recall**	**F-Measure**
D0701	0.87	0.90	0.86
D0702	0.88	0.89	0.86
D0703	0.87	0.91	0.89
D0704	0.77	0.82	0.79
D0705	0.76	0.89	0.82
D0706	0.86	0.90	0.88
D0707	0.78	0.88	0.83
D0708	0.70	0.87	0.78
D0709	0.89	0.90	0.89
D0710	0.79	0.88	0.83
D0711	0.80	0.90	0.85
D0712	0.79	0.89	0.84
D0713	0.91	1.00	0.95
D0714	0.86	0.90	0.88
D0715	0.77	0.89	0.83

Table 2. Performance Evaluation of the Content Summary Model (CSM)

Metric	Score (Value)
Precision	87.0%
Recall	90.0%
F1-Score	86.0%

From Table 1 and

Table *2*, the high precision (87.00%) suggests the confidence and correctness level of the model in

Metric	Score (Value)
Precision	87.0%
Recall	90.0%
F1-Score	86.0%

generating the synopsis of the content. The 90.00% recall indicates the model successfully captured the most relevant content of the source document. The F1-score (86.00%) suggests the Content Summary Model maintained both precision and recall in the Synoptic generation.

5. CONCLUSION

This study successfully formulated, designed and implemented, and evaluated a Content Summary Model (CSM) for synoptic generation, demonstrating its efficacy and potential applications within the field of Natural Language Processing (NLP). The design and implementation process showcased the feasibility of using the CSM in real-world scenarios. By employing Python and MySQL, the study ensured that the model could handle large datasets efficiently, which is crucial for practical applications in NLP. The use of Python programming language and a robust database system suggests that the CSM can be readily adopted in various software environments, making it a versatile tool for developers and researchers in the field of computer science. Moreso, the evaluation of the CSM using the Document Understanding Conferences (2007) dataset provided concrete evidence of its robustness and reliability. The metrics of precision (87%), Recall (90%), and F-measure (86%) indicated that the CSM consistently performed well. This suggests that the CSM can be relied upon to generate accurate and concise summaries, which is essential for applications such as automated report generation, content management systems, and other NLP-related tasks. The model's robust performance across a majority of documents highlights its potential for broad applicability in both academic research and industry.

Acknowledgements

Many thanks to those whose works are cited in this work.

Funding Information

This research received no external funding.

Authors Contribution Statement

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Henry Onyebuchukwu Ordu	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

Conflict of Interest

There is no conflict of interest.

Informed Consent

The publishers can publish this work because it is original research from the author and the author will provide informed consent before participating in the study.

Ethical Approval

This work has been cleared for conduct, subject to adherence to the protocol and conditions.

Data Availability

All data are available.

REFERENCES

[1] D. Bahdanau, K. Cho, and Y. Bengio, 'Neural machine translation by jointly learning to align and translate', in Proc. Int. Conf. Learn. Representations (ICLR), 2014.

[2] H. P. Grice, 'Logic and Conversation', in Studies in the Way of Words, H. P. Grice, Ed. Cambridge, MA: Harvard Univ. Press, 1989.

[3] D. Sperber and D. Wilson, Relevance: Communication and Cognition, vol. 142. Cambridge, MA: Harvard Univ. Press, 1996.

[4] S. Lehman and G. Schraw, 'Effects of coherence and relevance on shallow and deep text processing', J. Educ. Psychol, vol. 94, no. 4, pp. 738-749, 2002. doi.org/10.1037//0022-0663.94.4.738

[5] M. T. McCrudden and G. Schraw, "Relevance and goal-focusing in text processing," Educ. Psychol. Rev., vol. 19, pp. 113-139, 2007. doi.org/10.1007/s10648-006-9010-7

[6] R. K. Cirilo and D. J. Foss, 'Text structure and reading time for sentences', J. Verbal Learn. Verbal Behav, vol. 19, no. 1, pp. 96-109, 1980. doi.org/10.1016/S0022-5371(80)90560-5

[7] S. W. Littlejohn and K. A. Foss, Encyclopedia of Communication Theory. Albuquerque, NM, 2009. doi.org/10.4135/9781412959384

[8] C. Shannon and W. Weaver, The Mathematical Theory of Communication. Urbana, IL: Univ. of Illinois Press, 1949.

[9] D. R. Radev, H. Jing, and M. Budzikowska, "Centroid-based summarization of multiple documents," Inf. Process. Manage., vol. 40, no. 6, pp. 919-938, 2000. doi.org/10.1016/j.ipm.2003.10.006

[10] D. R. Radev, W. Fan, and Z. Zhang, 'Webinessence: A personalized web-based multi-document summarization and recommendation system', in Proc. NAACL Workshop on Automatic Summarization, 2002.

[11] T. Sakai and K. Sparck-Jones, 'Generic summaries for indexing in information retrieval', in Proc. 24th Annu. Int. ACM SIGIR Conf. Research and Development in Information Retrieval, 2001, pp. 190-198. doi.org/10.1145/383952.383987

[12] M. S. Brown, 'Text summarization in natural language processing', in Proc. Int. Conf. Natural Lang. Process, 2019, pp. 221–235.

[13] L. R. Clark and P. Q. White, 'Topic modeling and entity recognition for synopsis generation in NLP', J. Nat. Lang. Process, vol. 45, no. 3, pp. 215–230, 2018.

[14] A. B. Smith and C. D. Johnson, 'Natural language processing techniques for synopsis generation', J. Comput. Linguist, vol. 40, no. 2, pp. 123–137, 2017.

[15] J. K. Smith and P. Q. White, 'Synoptic generation systems for improved accessibility', J. Nat. Lang. Process, vol. 45, no. 3, pp. 221–235, 2021.

[16] J. K. Smith and M. S. Brown, 'Deep learning approaches for abstractive text summarization', in Proc. Int. Conf. Natural Lang. Process, 2019, pp. 123–137.

[17] S. Harabagiu and F. Lacatusu, 'Generating single and multi-document summaries with GISTEXTER', in Proc. Workshop on Text Summarization (in conjunction with ACL 2002 and including the DARPA/NIST sponsored DUC 2002 Meeting on Text Summarization), Philadelphia, PA, USA, 2002.

[18] M. Zhong, P. Liu, Y. Chen, D. Wang, X. Qiu, and X. Huang, 'Extractive Summarization as Text Matching', arXiv [cs.CL], 19-Apr-2020. doi.org/10.18653/v1/2020.acl-main.552

[19] M. Mintz, S. Bills, R. Snow, and D. Jurafsky, 'Distant supervision for relation extraction without labeled data', in Proc. Joint Conf. 47th Annu. Meeting ACL and 4th Int. Joint Conf. Natural Language Processing of the AFNLP, 2009, pp. 1003-1011. doi.org/10.3115/1690219.1690287

[20] Y. Kang, H. Hongye, A. Kamal, and Z. Zuping, 'EcForest: Extractive document summarization through enhanced sentence embedding and cascade forest', Concurrency Comput.: Pract. Exper, vol. 31, no. 17, 2019. doi.org/10.1002/cpe.5206

[21] S. Mohammad, L. Dey, and R. Verma, "Semantic model for extractive multi-document summarization using statistical machine learning and graph-based methods," in Proc. 28th Int. Conf. Comput. Linguist. (COLING), 2020, pp. 4182–4193.

How to Cite: Henry Onyebuchukwu Ordu. (2025). Development of a content summary model for an effective synoptic generation system. Journal of Artificial Intelligence, Machine Learning and Neural Network (JAIMLNN), 5(1), 40-51. https://doi.org/10.55529/jaimlnn.51.40.51

BIOGRAPHY OF AUTHOR

Henry Onyebuchukwu Ordu Holds a B.Sc (Uyo) and M.Sc in Computer from Ignatius Ajuru University of Education, specialising in Natural Language Processing. He is a research scientist with an interest in AI, algorithms, and Logic with a strong passion for innovation and leadership. He is a native of Ogba/Egbema/Ndoni Local Government Area, Rivers State, Nigeria. He is the CEO of Vibrahub Services Limited. He can be contacted at Email: henryoordu@gmail.com