

A Novel Approach for Multi-Document Summarization to Produce Readable Text using K-Means and Flesch Score

J.Tamilselvan

Ph.D., Research Scholar, Department of Computer Science, Alagappa University
Karaikudi, Tamilnadu, India
tamilselvan@ymail.com

Dr. A.Senthilrajan

Professor and Director, Department of Computational Logistics, Alagappa University
Karaikudi, Tamilnadu, India
Agni_senthil@yahoo.com

Abstract

Text summarization plays a vital role in minimizing the reading time. Instead of going through a long document, which may or may not be in the interest of the person. This kind of summarized text help to get the gist of the document based on the information present in the summary. The person may opt for reading the whole document or he/she may terminate the reading with the summary alone based on the interest/information present in the summary. To attract and to make the reader, read the full document the summary must be in a readable format. As the extractive summary extracts the important sentences from the given document based on the ranking of sentences the summary may miss the coherence that can be diffuse and sometimes even misleading, which will make the reader lose interest in reading the summary as well as the base document from the summary extracted. To enhance the readability, we evaluate the summary for its readability with various readability scores. Readability scores are a set of grade levels on a scale that determines how difficult a text is to read. The extractive summary produces the summary based on the parameters set. This helps researchers to improve the language by setting different parameters so that it is more understandable to a wider audience, making the content more engaging and it supports the reader in identifying appropriate stuff to read. To enhance the readability, we evaluate the summary generated by k-means, for its readability with Flesch readability score and set up a technique "SumRe" for summarizing the text with elevated readable value.

Keywords: *Text Summarization; K-Means; Flesch Score; Latent dirichlet allocation; Latent semantic analysis.*

1. Introduction

In recent years, a vast amount of data has been increasingly stored digitally, facilitating interpretation and analysis by a computer. In the text document type of data, when the number of documents or the content in the documents is large, manually reconstructing the data through human efforts is an expensive procedure. As a result, numerous automated approaches for automatically synthesizing documents and providing a summary version to the user have been developed. The automated text summarizing task chooses the most important ideas in a text so that the reader may comprehend the target material. Ideas of [Kagebäck M, et. al. (2014)][Ziqiang C, et. al. (2015)] [Lea F, et. al. (2014)] are related to the 'importance of specific keywords in documents', 'importance of sentence placement in documents', 'similarity between sentences in a document', and so on.

There are two types of automatic summarization techniques: extractive and abstractive. The output summary of extractive approaches contains text units from the source text such as words, sentence segments, or whole sentences, and is based on a rudimentary assessment of the text that covers only the syntactic level [Pattnaik, S. and Nayak, A. K. (2020)]. Abstractive techniques, on the other hand, provide a more in-depth analysis; for example, they may include semantic analysis, and the output summary may include new entities not found in the original text. As a result, there is a chance that sentences in "abstractive" summaries will be rewritten with a different interpretation than the original author's.

Extraction is the process of concatenating extracts from a corpus into a summary, whereas abstraction is the process of creating new phrases from the extracted data. Methods for abstracting text summarization provide highly coherent, cohesive, information-rich, and less redundant summaries with good readability. The extractive

summary, on the other hand, provides assertions that are less coherent, cohesive, and redundant, thus being difficult to read [Luu, M. (2021)]. There are indeed automated approaches for evaluating model summaries, such as ROUGE, which assigns a score based on the similarity of word sequences between a human-written model summary and a machine summary. While ROUGE scores have been demonstrated to correlate well with inference, they do not provide information about the summary's readability characteristics. To evaluate and analyze the readability of the summarized documents different testing methods such as Flesch Kincaid Reading Ease, Flesch Kincaid Grade Level, Gunning Fog Score, Coleman Liau Index, Automated Readability Index (ARI), and SMOG Index are used.

$$\text{FKRE} = 206.835 - 1.015 \times (\text{words/sentences}) - 84.6 \times (\text{syllables/words}) \quad (1)$$

$$\text{FKGL} = 0.39 \times (\text{words/sentence}) + 11.8 \times (\text{syllable/words}) - 15.59 \quad (2)$$

$$\text{FOG} = 0.4 \times (\text{average sentence length} + \text{percentage of hard words}) \quad (3)$$

$$\text{SMOG} = 1.0430 \times \sqrt{30 \times \text{complex words} / \text{sentences}} + 3.1291 \quad (4)$$

$$\text{CL} = 5.89 \times (\text{characters} / \text{words}) - 0.3 \times (\text{sentences} / \text{words}) - 15.8 \quad (5)$$

$$\text{ARI} = 4.71 \times (\text{characters} / \text{words}) + 0.5 \times (\text{words/sentences}) - 21.43 \quad (6)$$

A problem in extractive multi-document summarization is the understanding of the summarized content when it lacks readability. The summary produced by the k-means algorithm alone does not solve this problem. To deal with the problem of text readability, we propose an extractive multi-document approach using the k-means clustering algorithm combined with the readability score calculation and matching. Experimental results show that our system regarding readability is significantly efficient compared to existing methods on multi-document summarization. The architecture is represented in Fig. 1.

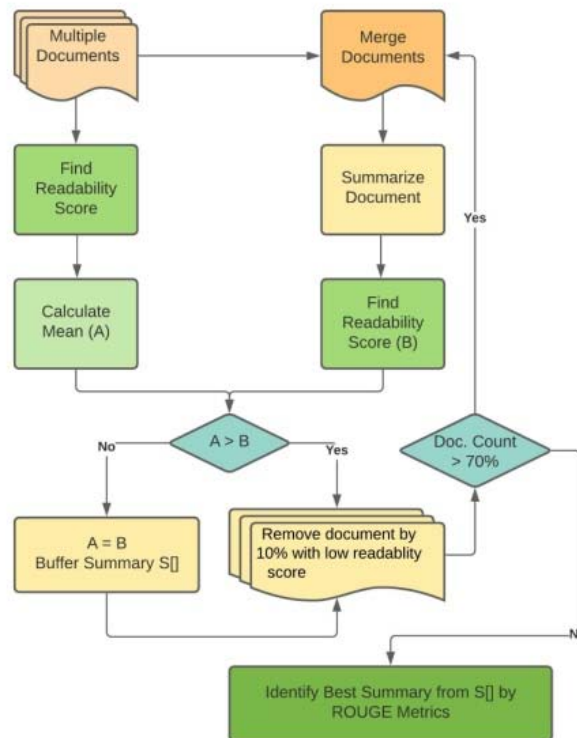


Fig. 1: Architecture of extractive summarization with readability check

2. Related Works

This section discusses some of the document summarization techniques that use different methods to get the summary. The authors [Hernández-castañeda, Á., et. al.(2020)] formulated a method for Extractive Automatic Text Summarization based on Lexical-semantic Keywords; it enhances precision by finding the keywords in the clusters, as well as increasing coverage by clustering sentences to identify the key concepts from the source document. The authors had a problem while generalizing the contents. The supervised and unsupervised techniques are used to evaluate the features retrieved from documents to construct models that allow the key parts of essential concepts to be discovered. [Ma, Y., and Wu, J. (2014)] defined combined n -gram and dependency word pairs for multi-document summarization. The syntactic relationships between the words are defined by the dependency word pair. Each feature reproduces the co-occurrence of the common topics of multi documents in

diverse perspectives. Finally, using the higher significance score model, the summary was created by retrieving the most significant sentences.

2.1. Text Summarization Strategies

Textual summarization extracts and concatenates significant sentences from the source text to make a customized summary. A list of features is discussed below that can be applied to identify the critical sentences for the summary. The author [Idris, N. (1986)] proposes several summarizing approaches for creating relevant summaries. The following are detailed explanations of these strategies:

- (1) Selection and Deletion: trivial or redundant information is eliminated from the sentence.

Minimizing redundant information from the final summary because it is a greater chance to get multiple sentences having similar contents from documents. The significant step in removing redundant information is to identify similar contents from the summary using appropriate similarity measures [Patel, D. B (2019)]. By using the cosine similarity measures, the similarity between the sentences is calculated and the most similar sentences are removed.

- (2) Sentence combination/clustering: two or more original sentences are combined to construct a summary sentence [Han, K., and Chen, G. (2019)].

Sentences Clustering, determines the number of sentences clusters, we can use the K-means method to cluster the sentences of the document [Pei-ying, Z. (2009)].

- (3) Generalization: A list of words or elements is substituted with a more generic word from the same class.

The existence of a taxonomy of concepts that may be derived from text is the basis of textual inference. More specifically, the said taxonomy contains concepts and their hypernyms in a hierarchical structure. Once the concepts have been extracted, the taxonomy path, containing the ordered sequence of concepts according to their taxonomy depth, is used for generalizing text [Stafylopatis, A. (2019)].

- (4) Syntactic transformation: The words in a sentence are rearranged in a different sequence.

The authors [McKeown, R] present a “cut and paste based text summarization technique”, aiming at narrowing the space between automatic system-generated summaries and human-written summaries. The summarizer edits extracted sentences by removing in essential terms and merging the resulting phrases into logical sentences using reduction and combination. A statistically-based sentence decomposition software that detects where the phrases of a summary originate in the original document, resulting in an aligned corpus of summaries and articles, is one representation of their work. The authors [McKeown, K. R., et al (1999)] create a multi-document summarizing system that identifies and contextualizes similarities across a series of linked papers to automatically construct a summary. By intersecting comparable terms inside paragraphs, the authors employed statistical techniques with machine learning to find related paragraphs, and language generation to reformulate the summary's text. The author [Jing, H. (2000)] implemented an automatic sentence reduction system. The extracted sentences, as well as the original document, are sent into the reduction mechanism as input. The output of reduction is a reduced forms of the extracted sentences, which can either be used to produce summaries directly or be merged with other sentences. To make reduction assessments, the reduction system consults a variety of sources of information, including syntactic knowledge, context, and statistics computed from a training corpus.

- (5) Sentence reordering: The original text's sentence sequence has been modified.

The author [Deng, Z. (2021)] used the BERT-based ensemble model to select the most salient sentences in reference papers and then uses a similar model to reorder these sentences for better readability.

- (6) Invention: The original sentences are replaced with new, more concise ones.

The invention rule is used to construct a summary sentence when one creates explicit subject sentences by stating the implicit major concept of the paragraphs in one's own words. As a result, rather than simply removing, selecting, or editing sentences already provided for them, the invention rule compels them to contribute the information. The author [Idris, N. (1986)] formulated the heuristic rules based on the expert summarizing strategies which are acquired mainly by studying the experts' summaries. The study was conducted to identify the experts' summarizing strategies and how they used the strategies to produce their summary sentences.

- (7) Paraphrase: The original phrase or word is replaced with a related phrase or word.

In the paraphrasing process, a term from the source phrase is substituted in the summary sentence with a synonym (a different word with the same meaning).

- (8) Topic Sentence Selection:

The topic sentence selection approach is used to extract a significant sentence from the original text to reflect the core concepts of a paragraph and generate a summary sentence. There are four ways to figure out which statement is the most important:

- (i) Key method. The most often occurring words in a document are the most representative of their topic, therefore a section of text that contains them is more relevant. Word frequency is a technique for identifying non-stop words that appear frequently in a document. Keywords or content words in sentences have a higher possibility of being included in the summary.

- (ii) Location method: Important sentences are usually found at the start and end of a page or paragraph, as well as just beneath section heads. The first and last sentences of paragraphs at the beginning and the end of a text are more likely to contain information that can be used in a summary.
- (iii) Title method: Important sentences usually contain words that appear in the document's title and main headings. As a result, words in the title can be used to represent document-specific features.
- (iv) Cue method: Cue phrases are words or phrases that directly indicate a discourse's structure. Sentences with cue phrases like "conclusion" or "in particular" are seen as more essential than sentences without them.

2.2. Readability Testing Methods

The simplicity with which a reader may interpret written content is known as readability [Derguech, W. (2018)]. The readability of text in natural language is determined by its content (the complexity of its vocabulary and syntax). It focuses on the words the summarizer selects and how it structures the sentences into phrases and paragraphs so that readers may understand them more readily [Klare, G. R., and Stolurow, L. M. (1972)]. The author [Kouamé B. Julien. (2010)] stated that "Readability statistics are simple, but they are often neglected. They can not only provide information on the level of difficulty of a document's readability, but they can also boost an evaluator's credibility." The author [Hu N. (2012)] describes that the writing style plays a major role in readability enhancement; in turn, it defines the reading ease that improves the comprehension as well as the retention of the textual material. Our primary goal is to double-check and relay information to ensure that both the source document and the summary content are relevant and readable. Summarization efforts are lost if the summary fails to convey that information. It is vital to give information to the reader that they will gladly keep reading and grasp well to interest them. As a result, the text must be as easy to read and comprehend as feasible.

3. Implementation and Computational Results

We applied two methods for automatic extraction of system-generated summaries following readability Flesch score like 1) Vector method by building cosine similarity matrix (Cosine similarity is the measure of similarity between two non-zero vectors of an inner product space that measures the cosine of the angle between them). 2) Vector method using K-Means clustering (Cluster of sentence embeddings can be interpreted with the sentence vector which has the minimum Euclidean distance from the cluster centroid forms the whole group. These sentences from each cluster are ordered to form a meaningful summary). To evaluate the effectiveness of k-means i) ranking by sentences and ii) ranking sentences from clustered documents. We compared them with other clustering techniques like Latent Semantic Analysis (LSA) [Gong, Y. (2001)], Latent Dirichlet Allocation (LDA) [Ravindran, B. (2014)] for the multi-document summarization task.

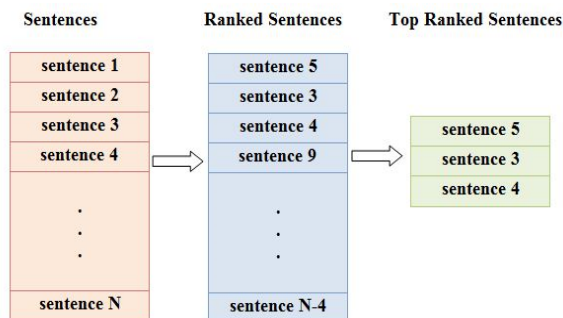


Fig. 2. Summary by ranking sentences

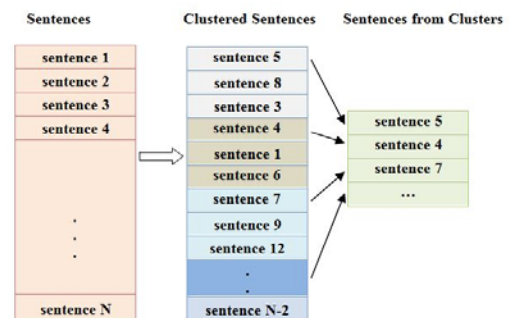


Fig. 3. Summary by clustering and ranking sentences

3.1. Dataset

We used the DUC 2006 dataset for document summarization, it consists of 50 topics with 25 documents representing each topic, we selected four topics consisting of 25 documents released from different press. The details of the documents used are given in table 1.

DUC Doc. ID	Topic	Documents from	Nos.	Total	Description
D0603C	Wetlands Protection	Associated Press	07	25	Highlights the significance of wetlands and how they can be beneficial in a variety of ways.
		New York Times	08		
		Xinhua newswire	10		
D0605E	Herbal Remedy	Associated Press	10	25	The use of herbal medicine in the treatment of rheumatic illnesses is documented.
		New York Times	14		
		Xinhua newswire	01		
D0608H	Vehicle Safety	Associated Press	13	25	Explains the necessity of vehicle safety procedures and provides various safety solutions.
		New York Times	09		
		Xinhua newswire	03		
D0614E	Quebec Province	Associated Press	05	25	The news narrates the story of Quebec's history and conflicts.
		New York Times	09		
		Xinhua newswire	11		

Table 1. Description of DUC2006 Dataset

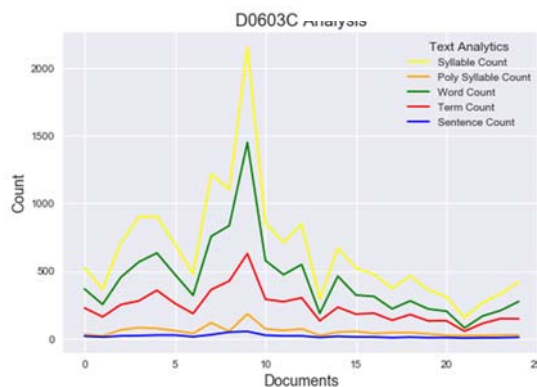


Fig. 4. Text Analytics of D0603C

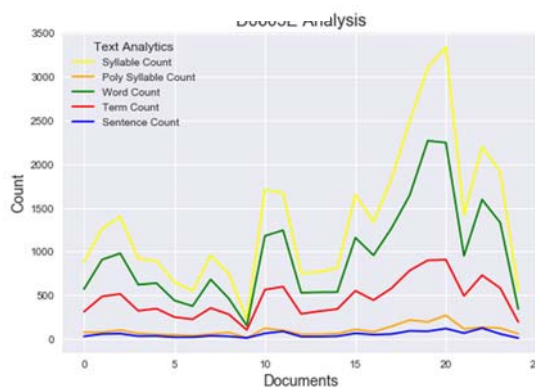


Fig. 5. Text Analytics of D0605E

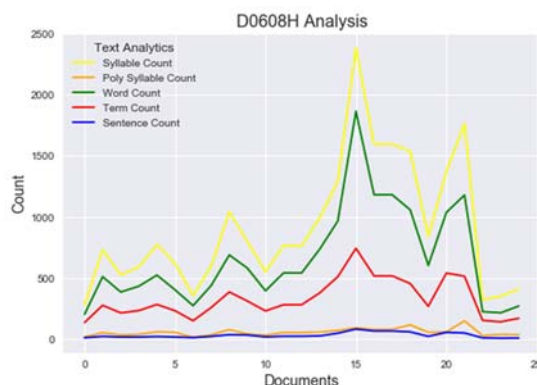


Fig. 6. Text Analytics of D0608H

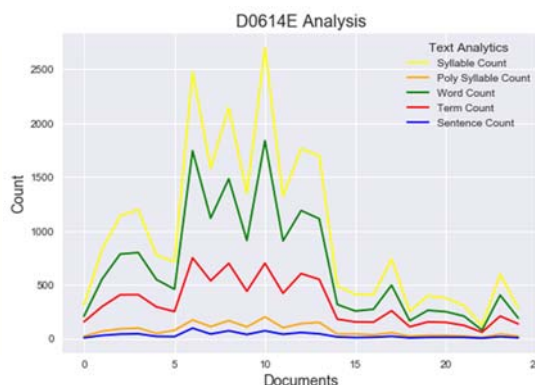


Fig. 7. Text Analytics of D0614E

The DUC 2006 dataset consists of pyramid data with four manual summaries with 250 words that are used as the gold standard summary to compare with the system-generated summaries. The detailed analysis of the documents is represented in the above-given figures 4,5,6 and 7 with word count, term count, sentence count, syllable count, and polysyllable count that plays a key role in text readability. The readability scores of system-generated summaries and summaries provided by people using various readability metrics are compared. It is necessary to examine words and phrases in Natural Language Processing to determine the text's difficulty. Readability Scores are a set of grade levels on a scale that determines how difficult a text is to read. It assists the system to generate a better summary by improving the text to make it understandable for a larger audience, thus making content engaging.

3.2. ROUGE Evaluation

The ROUGE metrics are used for comparing the automatically generated summaries with the set of reference summaries; we use ROUGE 1 metrics to evaluate the summaries. The reason to use ROUGE 1 in conjunction with ROUGE 2 and ROUGE 3 is to show the fluency of the summaries. The Precision (7), Recall (8), and F-Score (9) are the three criteria used to investigate the performance comparisons which are generated by ROUGE metrics. The idea is that it is more fluent to closely follow the word orderings of the reference summary.

Precision = number of n-grams found in model and reference/number of n-grams in a model (7)

Recall = number of n-grams found in model and reference/number of n-grams in reference (8)

F-Score = $2 \times (\text{Precision} \times \text{recall}) / (\text{Precision} + \text{Recall})$ (9)

4. Experimental Results

K-Means (Sentence Ranking from Cluster)	Documents Collection	Flesch_score	ROUGE1		
			Precision	Recall	F-Measure
	100%	58.194935	0.396	0.410788	0.403259
	90%	53.283921	0.456	0.454183	0.45509
	80%	67.526493	0.4824903	0.4862745	0.484375

Table 2. D0603C – Summary Readability Scores

K-Means (Sentence Ranking from Cluster)	Documents Collection	Flesch_score	ROUGE1		
			Precision	Recall	F-Measure
	100%	53.283921	0.406639	0.379845	0.392786
	90%	55.194935	0.454183	0.456	0.45509
	80%	58.661962	0.398406	0.414938	0.406504

Table 3. D0605E – Summary Readability Scores

K-Means (Sentence Ranking from Cluster)	Documents Collection	Flesch_score	ROUGE1		
			Precision	Recall	F-Measure
	100%	45.452583	0.4722222	0.463035	0.467583
	90%	51.375769	0.4722222	0.4666667	0.469428
	80%	47.884091	0.484127	0.4784314	0.481262

Table 4. D0608H – Summary Readability Scores

K-Means (Sentence Ranking from Cluster)	Documents Collection	Flesch_score	ROUGE1		
			Precision	Recall	F-Measure
	100%	53.283921	0.5019608	0.4980545	0.5
	90%	55.194935	0.4392157	0.4392157	0.439216
	80%	58.661962	0.4784314	0.484127	0.481262

Table 5. D0614E – Summary Readability Scores

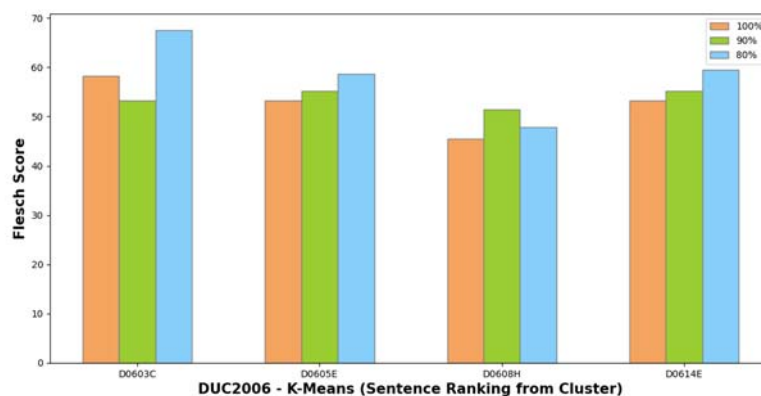


Fig. 8. Flesch Score Comparison of K-Means SRFC

From the results shown in tables 2, 3, 4 and 5, it is clear that when the documents with low readability scores are removed, then the overall readability score of the summary improves. Along with this ROUGE score, is also taken care to produce a good summary with most of the contents covered from the original corpus. Three of the documents collection produce the best readability Flesch score when the size of the documents is reduced to 80% and one document set we got a comparatively better Flesch score with 90% of documents. The D0603C documents collection produced a decent result compared with other documents set D0605E, D608H, and D0614C.

K-Means (Sentence Ranking)	Documents Collection	Flesch_score	ROUGE1		
			Precision	Recall	F-Measure
	100%	63.049646	0.4824903	0.4862745	0.484375
	90%	65.348997	0.396	0.410788	0.403259
	80%	61.622733	0.4980545	0.5019608	0.5

Table 6. D0603C – Summary Readability Scores

The K-Means (Sentence The table 6 describes the

K-Means (Sentence Ranking)	Documents Collection	Flesch_score	ROUGE1		
			Precision	Recall	F-Measure
	100%	43.832071	0.4862745	0.4824903	0.484375
	90%	44.047115	0.4666667	0.4722222	0.469428
	80%	41.715227	0.4392157	0.4392157	0.439216

Table 7. D0605E – Summary Readability Scores

K-Means (Sentence Ranking)	Documents Collection	Flesch_score	ROUGE1		
			Precision	Recall	F-Measure
	100%	40.697361	0.334661	0.325581	0.330059
	90%	35.625205	0.395349	0.408	0.401575
	80%	41.304006	0.379845	0.406639	0.392786

Table 8. D0608H – Summary Readability Scores

K-Means (Sentence Ranking)	Documents Collection	Flesch_score	ROUGE1		
			Precision	Recall	F-Measure
	100%	0.488281	0.5	0.494071	0.488281
	90%	0.47012	0.472	0.471058	0.47012
	80%	0.49004	0.480469	0.485207	0.49004

Table 9. D0614E – Summary Readability Scores

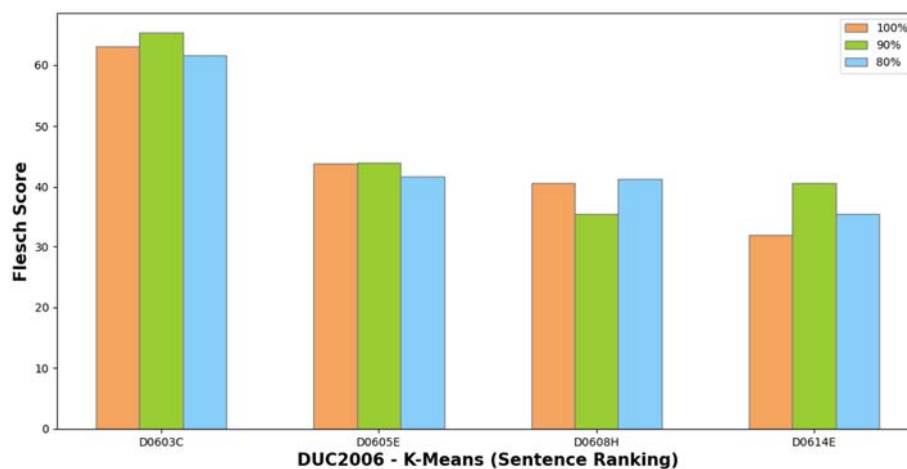


Fig. 9. Flesch Score Comparison of K-Means SR

The summary acquired from K-Means Sentence Ranking produces a low readable score compared with the summary generated from K-Means Sentence Ranking From Cluster. In this method, two documents set produced a good readability score with 90% of documents and two with 80% of documents. When the documents are reduced it gives a low readability score.

Method	Flesch Score	ROUGE1		
		Precision	Recall	F-Measure
K-Means (Sentence Ranking from Cluster)	67.52649264	0.482490272	0.48627451	0.484375
K-Means (Sentence Ranking)	65.34899712	0.396	0.410788	0.403259
LDA	50.45678224	0.484690222	0.49346524	0.489038
LSA	52.64834442	0.468824332	0.48888654	0.478645

Table 10. D0603C - Comparison Results with Basic Methods

Method	Flesch Score	ROUGE1		
		Precision	Recall	F-Measure
K-Means (Sentence Ranking from Cluster)	58.66196247	0.398406	0.414938	0.406504
K-Means (Sentence Ranking)	44.04711538	0.466666667	0.472222222	0.469428
LDA	42.46344462	0.424888242	0.466248882	0.444609
LSA	40.64244242	0.463345664	0.486462242	0.474623

Table 11. D0605E - Comparison Results with Basic Methods

Method	Flesch Score	ROUGE1		
		Precision	Recall	F-Measure
K-Means (Sentence Ranking from Cluster)	51.37576923	0.472222222	0.466666667	0.469428
K-Means (Sentence Ranking)	41.30400554	0.379845	0.406639	0.392786
LDA	40.63898473	0.488626266	0.484242444	0.486424
LSA	39.64242224	0.478902235	0.462456653	0.470536

Table 12. D0608H - Comparison Results with Basic Methods

Method	Flesch Score	ROUGE1		
		Precision	Recall	F-Measure
K-Means (Sentence Ranking from Cluster)	59.44166232	0.478431373	0.484126984	0.481262
K-Means (Sentence Ranking)	40.6973607	0.47012	0.472	0.471058
LDA	30.64224442	0.460606622	0.48452222	0.472262
LSA	38.84388283	0.480464402	0.42464224	0.450832

Table 13. D0614E - Comparison Results with Basic Methods

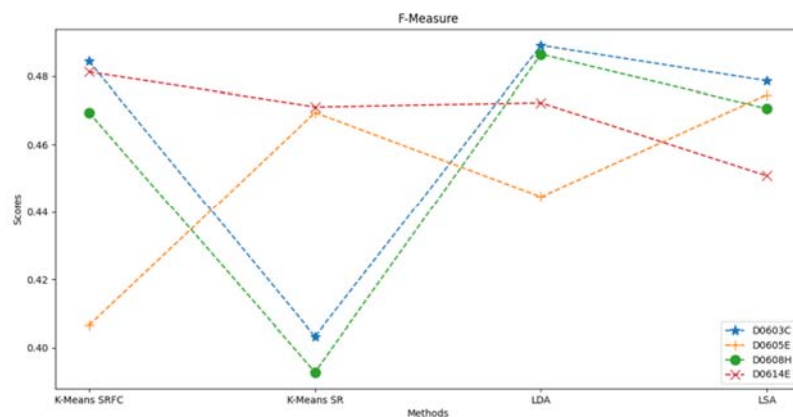


Fig. 10. ROUGE1 Score Comparison

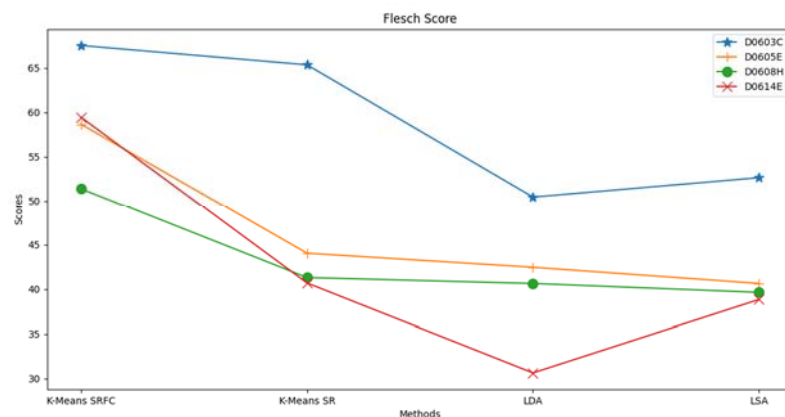


Fig. 11. Flesch Score Comparison with various methods

It is evident that the readability Flesch score K-Means (Sentence Ranking from Cluster) is comparatively high when the low readable score documents are eliminated from the corpus and also we found that in the same K-Means (Sentence Ranking) it does not give the same result as the K-Means (SRFC) here the readability scores are not improved that much. As the elimination is based on readability score by removing the whole documents itself, instead if it is done by eliminating sentence level from the documents, it may give the result with a high readable score. The ROUGE1 score compared with other methods like LDA and LSA are nearly similar and the Flesch readability score is better in K-Means (Sentence Ranking from Cluster) and K-Means (Sentence Ranking).

Word clouds are useful for any kind of text data in exploratory qualitative analysis, including essays, short answers, or written responses to survey or opinion questions [Depaolo C. and Wilkinson K. (2014)].

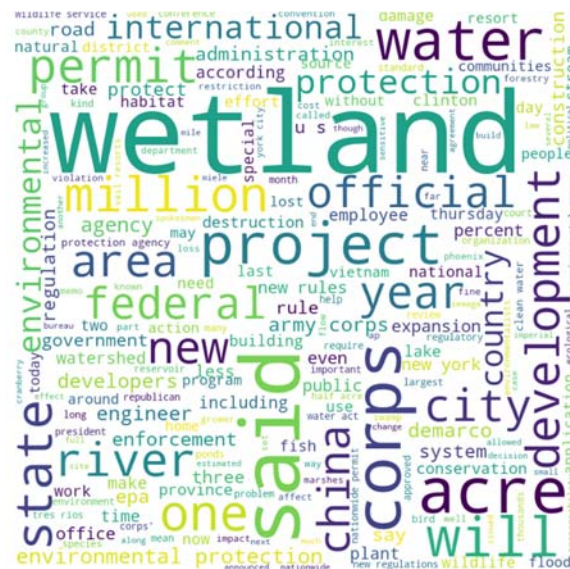


Fig. 12. D0603C Documents Collection Word Cloud

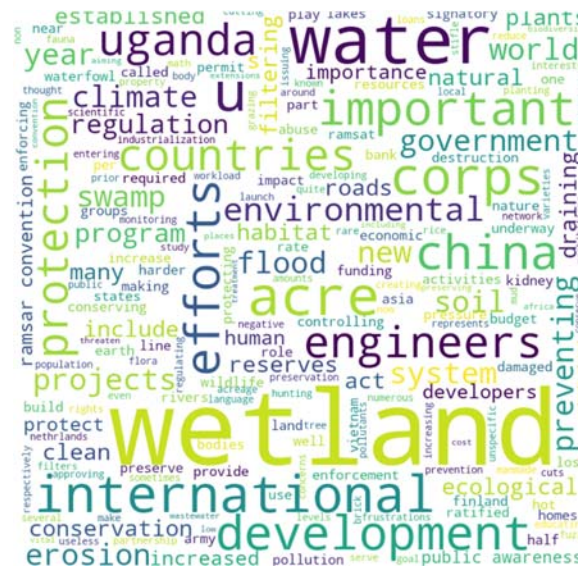


Fig. 13. D0603C Summary Word Cloud

The word cloud for the content and its summary are given for the D0603C documents collection, which has a good Flesch score along with ROUGE1's F-measure is given. By analyzing the resulting word clouds in Figures 12 and 13, we can determine that the keywords such as “wetland”, “water”, “environmental”, “international”, and “development” are seen in both.

5. Conclusions and Future work

In this paper, we have proposed an efficient extractive multi-document summarization system using the K-means clustering algorithm matching with the Flesch readability score. Experiments with DUC 2006 dataset show that our system is better than several modern systems in this field regarding readability and nearly similar to ROUGE metrics. It proves that our method of combining readability scores techniques is efficient in improving the readability of the multi-document summarization system. In the future, we will investigate the sentence level readability measures with other sources of knowledge to improve further the system performance and explore other interesting applications to produce the best readable summary.

References

- [1] Deng, Z. (2021). *Automatic Related Work Section Generation by Sentence Extraction and Reordering*. 1–10.
- [2] Depaolo, C., & Wilkinson, K. (2014). *Get Your Head into the Clouds : Using Word Clouds for Analyzing Qualitative Assessment Data*. December 2016.
- [3] Derguech, W. (2018). *Assessing the Readability of Policy Documents : The Case of Terms of Use of Online Services*.
- [4] Gong, Y. (2001). *Generic Text Summarization Using Relevance Measure and Latent Semantic Analysis*.
- [5] Han, K., & Chen, G. (2019). *An Approach to Sentence-Selection-Based Text Summarization*.
- [6] Hernández-castañeda, Á., García-hernández, R. A., & Millán-hernández, C. E. (2020). *Extractive Automatic Text Summarization based on Lexical-semantic Keywords*. 1–12.
- [7] Hu, N., Bose, I., Sian, N., & Liu, L. (2012). *Manipulation of online reviews : An analysis of ratings , readability , and sentiments*. 674–684.
- [8] Idris, N. (1986). *A Summary Sentence Decomposition Algorithm for Summarizing Strategies Identification*. 200–210.
- [9] Jing, H. (2000). *Sentence Reduction for Automatic Text Summarization*. 310–315.
- [10] Kagebäck M, Mogren O, Tahmasebi N, Dubhashi D, (2014). *Extractive Summarization using Continuous Vector Space Models*, Proceedings of the 2nd Workshop on Continuous Vector Space Models and their Compositionality (CVSC), pp.31-39.
- [11] Klare, G. R., and Stolurow, L. M. (1972). *The Cloze Procedure: A Convenient Readability Test for Training Materials and Translations*. 21(2).
- [12] Lea F, Ivan T, Manfred P, (2014). *A Hierarchical Bayesian Model for Unsupervised Induction of Script Knowledge*, 14th Conference of the European Chapter of the Association for Computational Linguistics.
- [13] Kouamé B. Julien. (2010). *Using Readability Tests to Improve the Accuracy of Evaluation Documents Intended for Low-Literate Participants*. 132–139.
- [14] Luu, M. (2021). *An Effective Deep Learning Approach for Extractive Text Summarization*. 434–444.
- [15] Ma, Y., & Wu, J. (2014). *Combining N-gram and Dependency Word Pair for Multi-Documnet Summarization*.
- [16] McKeown, K. R., et al (1999). *Towards Multidocumnet Summarization by Reformulation : Progress and Prospects*.
- [17] McKeown, R. *Cut and Paste Based Text Summarization*.
- [18] Patel, D. B., Shah, S., & Chhinkaniwala, H. R. (2019). *Fuzzy logic based multi Document Summarization with improved sentence scoring and redundancy removal technique*.
- [19] Pattnaik, S., & Nayak, A. K. (2020). *A Simple and Efficient Text Summarization Model for Odia Text Documents*. 11(6), 825–834.
- [20] Pei-ying, Z. (2009). *Automatic text summarization based on sentences clustering and extraction*. 1(1), 167–170.
- [21] Ravindran, B. (2014). *Latent Dirichlet Allocation Based Multi-Documnet Summarization*.
- [22] Stafylopatis, A. (2019). *Abstractive text summarization based on deep learning and semantic content generalization*. 5082–5092.
- [23] Ziqiang C, Furu W, Li D, Sujian L, Ming Z, (2015). *Ranking with Recursive Neural Networks and Its Application to Multi-documnet Summarization*, AAAI.

Authors Profile



J.TAMILSELVAN, pursuing his Ph.D. in Computer Science, Department of Computer Science, Alagappa University, Karaikudi, Tamilnadu, India and working as an Assistant Professor in the Department of Computer Science, K.S.Rangasamy College of Arts and Science (Autonomous). He has 18 years of teaching experience. He has published research papers in refereed journals and presented papers at various National / International conferences and seminars. His area of specialization includes Machine Learning, Text Mining, Deep Learning and Web Technologies.



Dr. A.SENTHILRAJAN, B.E., MBA., M.Sc. (IT)., M.Phil., Ph.D., is working as a Professor and Director in the Department of Computational Logistics, Alagappa University, Karaikudi Tamilnadu, India. He has 20 years of teaching experience at various positions and around 08 years of research experience. He is the member of various professional bodies, Editorial board, and reviewer of various journals. He has organized many Training programs, National and International conferences. He acted as a Chair Person and given Keynote Speeches in International/National Conference. He has published many research papers in various refereed journals and presented papers at various National / International conferences and seminars. His area of specialization includes Image Processing, Networks and Artificial Intelligence.