

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 Stolorow, 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-Documents Summarization*.
- [16] Mckeown, K. R., et al (1999). *Towards Multidocument 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-Documents 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-document 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.