Query-based Text Summarization: A Comparative Investigation and TextRank Implementation

Varad Unhale[1], Suyash Udchan[2], Mohak Shah[3] and Bhakti D. Kadam[4]

Department of Electronics and Telecommunication Engineering,
SCTR's Pune Institute of Computer Technology, Pune, India.
varadunhale.vu@gmail.com[1], suyash.udchan@gmail.com[2],
shahmohak2311@gmail.com[3], bdkadam@pict.edu[4]

Abstract. Owing to the proliferation of internet and multimedia applications, there has been a huge increase in digital content in the form of text, images, audio and video worldwide. The various applications of text summarization has attracted the computer vision researchers to generate the optimal text summaries. Several methodologies have been reported in literature to generate text summaries. However, the key challenge is incorporating user's preference as text summarization is a subjective task. User generated queries act as a guiding beacon in the summarization process. Instead of producing generic text summaries, query-based text summarization techniques offer user preferred responses. This paper begins by discussing the fundamental concepts and objectives of text summarization, emphasizing on the role of user queries in the summarization process. It proceeds to categorize existing query-based summarization techniques into distinct paradigms, including extractive, abstractive, and hybrid approaches, highlighting the advantages and limitations of each. This paper presents the evolution of query-based text summarization with developed techniques, available datasets, evaluation metrics, and performance comparison. After conducting a survey on query-based text summarization, a text summarization using the TextRank algorithm is implemented and Recall-Oriented Understudy for Gisting Evaluation (ROUGE) scores are calculated for the generated summaries.

Keywords: Query-based Text Summarization, Extractive Summarization, Abstractive Summarization, Deep Learning, Evaluation Metrics

1 Introduction

Text summarization is an essential task in a world where information is abundant but time and attention are limited. The ability to distill the most crucial information from lengthy documents or vast datasets enables users to quickly access the information that matters most to them. However, generic summarization methods may fall short of addressing the specific information needs of users. This is where query-based text summarization comes into play. Query-based text summarization, also known as topic-based, user-focused, or query-focused summarization, integrates user-provided query information to craft the text summaries. In contrast to generic text summarization, which aims to provide broad document summaries, query-based summarization extracts or generates text that handles query-specific points and condenses the documents. In essence, text summarization can be seen as a higher-level category comprising three forms: generic summarization, extractive query-based summarization, and abstractive query-based summarization. It often results in a summary of content related to the user's question.

There have been surveys, like the one conducted by Afantenos et al. [1], that addressed text summarization in the context of medical documents. Nevertheless, these surveys limited their focus to medical-related queries. In 2010, Damova et al. [2] conducted a survey on query-based text summarization, but the field has evolved significantly in the following decade, with the introduction of technologies such as Word2Vec, end-to-end sequence generation models, and various attention mechanisms.The challenge lies in developing an efficient and effective query-based text summarization method that can distill essential information from extensive textual datasets in response to user-generated queries, considering the diversity of summarization paradigms and their unique advantages and limitations. Therefore, this survey aims to provide a comprehensive overview of the specific techniques in query-based text summarization and bridge the gap in the literature, considering the recent developments in this field.

This paper presents a survey of generic and query-based text summarization techniques including existing techniques, performance comparison, limitations and challenges in the same. The paper also discusses an implementation of TextRank algorithm with obtained results.

2 Problem Statement

Given an input text, the aim of query-based text summarization model is to generate a meaningful text summary focusing on user inputted query. Figure 1 shows the systematic block diagram of query-based text summarization. Text pre-processing, Text extraction, Query pre-processing, training of the deep learning model and generating a text summary are the major steps in query-focused text summarization. From the perspective of text summarization, query-based text summarization can be categorized into two main approaches: query-based extractive and query-based abstractive text summarization. This type of summarization takes into account user-provided queries, which can be in the form of individual words or complete sentences. Additionally, research in this area can be divided into Single-Document Summarization (SDS) and Multi-Document Summarization (MDS). Figure 2 presents the broad categorization of Text Summarization techniques.

The survey's structure overview begins by examining a key division: extractive techniques and abstractive techniques in text summarization. These two

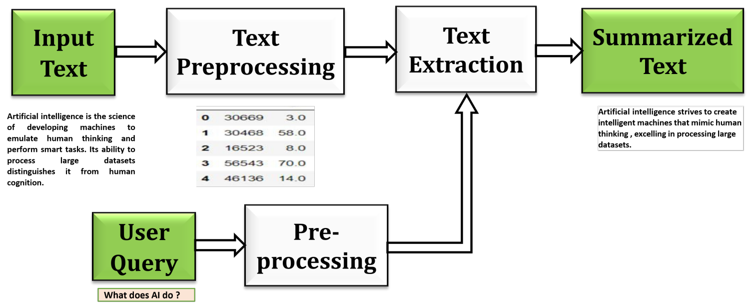

Fig. 1. Block Diagram of Query-based Text Summarization

approaches differ fundamentally, with extractive techniques focusing on selecting prominent sections from a document and directly extracting them. Consequently, query-based extractive summarization primarily evaluates the relevance of the content concerning the query. In contrast, abstractive text summarization aims to identify the essential information in the text and generate new text that encapsulates the identified information, which is not present in the original document. While most research in text summarization revolves around either extractive or abstractive techniques, there are also efforts to combine these two approaches using hybrid methods.

3 Existing Techniques

A crucial element in the development of a successful deep learning or machine learning algorithm is the establishment of a validated evaluation process that guarantees the generalization of the developed method. Evaluation comprises two key components: evaluation datasets and evaluation metrics. Choosing the appropriate evaluation datasets and metrics is task-dependent and essential for accurate generalization. This section provides a brief overview of existing approaches reported in the literature for both generic and query-based text summarization. Table 1 presents a comparison of existing text summarization techniques in terms of methodology used, datasets experimented, achieved results and limitations.

In the paper authored by Aakash Sinha et al. [3], feedforward neural networks are utilized in conjunction with Word2Vec and Fasttext for word and sentence vectorization, respectively, all while employing the softmax activation function. Their work is evaluated using the DUC 2002 dataset and ROUGE metrics, demonstrating that optimal performance is achieved when processing 40 sentences at a time. The paper authored by Ángel Herandez-Castenda et al. [4] employs clustering, Genetic Algorithm, LDA, TF-IDF, and N-grams techniques. They have evaluated their method using the DUC02 and TAC11 datasets, achieving an impressive 86% accuracy rate for their generated summaries and suggesting its potential.

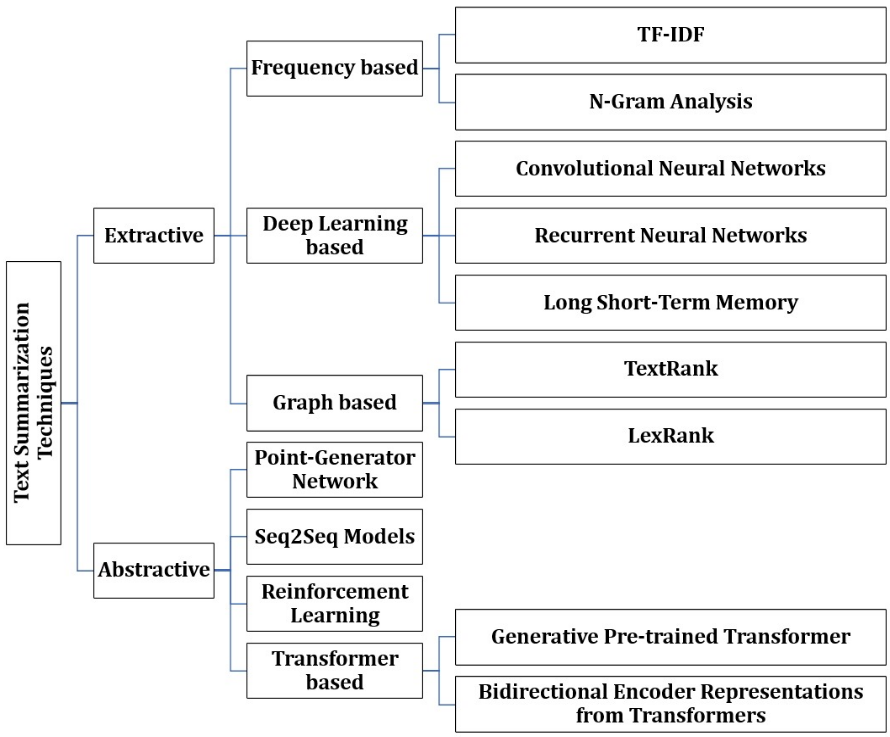

Fig. 2. Categorization of Text Summarization techniques

The research work presented in [5] utilizes cosine similarity, word-order similarity, semantic similarity, a hybrid similarity measure, and clustering algorithms for query-based summarization. The authors Ahmed A. Mohamed et al. introduced multiple summarization methods in [6], with the Q-Summarizer method standing out as particularly effective for generating query-based summaries with high relevance to the query. The paper authored by Samridhi Murarka et al.[7] presents a hybrid approach for query-relevant text summarization. Their methodology involves tokenization, normalization, stop-word removal, part-of-speech tagging, lemmatization, and context modeling using Latent Semantic Indexing. They also incorporate a Text Rank Algorithm.

A neural attention model is designed by the authors in [8] generate concise summaries from input sentences. Their approach combines a neural language model with various encoders, with a focus on minimizing the negative log-likelihood of input-summary pairs during training, utilizing stochastic gradient descent for parameter estimation. In a separate study by Johan Hasselqvist et al. [9] , they present a sequence-to-sequence model with attention and a pointer mechanism tailored for query-based abstractive summarization. This model considers both a document and a query as input, processing them using bidirectional Recurrent Neural Network (RNN) encoders. It strategically focuses on relevant segments of the document by using attention mechanism and employs a generator mechanism to craft summaries. Information about the evaluation metrics used in text summarization is provided in Table 2.

Table 1: Comparative study of Video Activity Classification techniques

Research Work	Methodology	Dataset	Results	Limitations
[3]	Feedforward neural networks, Word2Vec for word vectorization Fasttext for sentence, vectorization, Softmax activation function.	DUC 2002	The best performance in terms of ROUGE scores was observed when using a ‘page-len’ parameter set to 40.	Enhancing the model’s capability to generate summaries exceeding the ‘page-len’.
[10]	BERT, a bidirectional transformer model, boosts NLP tasks through pre-training and efficient fine-tuning, achieving top results.	BooksCorpus and English Wikipedia.	Unsupervised pre-training improves language understanding, benefits low-resource tasks.	Current language models are unidirectional, limiting context and pre-training options, impacting their performance.
[5]	Cosine similarity, word-order similarity,semantic similarity,a hybrid similarity measure, and clustering algorithms.	Various online websites of Amrita School of Engineering	These results suggest that setting the ‘page-len’ parameter to 65	Summarization methods can excel or struggle in specific domains due to specialized terminology.
[6]	DGS- Summarizer, Q-Summarizer & QInc- Summarizer.	DUC02 dataset.	Effective in generating query-based summaries.	Further experiments on larger text corpora are needed to evaluate the performance of the method.

[4]	A text summarization method utilizing Doc2Vec, LDA, and a genetic algorithm to cluster sentences based on semantic and lexical features.	DUC02 dataset.	Excellent results, outperforming prior methods on DUC02, displaying competitiveness on TAC11.	The paper lacks explicit limitations but potential concerns include scalability and GA parameter tuning.
[11]	Term Sense, Extraction Specificity Power,Informativeness Power, Sentence Selection, Maximal Marginal Relevance (MMR).	News corpus built.	Optimal summarization results were achieved with a compression rate of 20 for documents with many closely related sub-topics.	Further experiments on larger text corpora are needed to evaluate the performance of the method.
[9]	Extractive text summarization using a statistical novel approach based on sentence ranking	DUC 2002.	The proposed model improves the accuracy compared to traditional approaches.	The need for labeled data, anaphora and cataphora problems, and the ongoing.
[12]	A text summarization model is built using a sequence-to-sequence architecture with attention and a pointer mechanism.	The dataset used in the paper is created from CNN and Daily Mail news articles.	The experiment results show that the model performs better when queries are incorporated, as indicated by ROUGE scores.	The model's performance relies on queries, posing a limitation when they are absent or less informative.
[13]	MeanSum is a model for summarizing multiple documents without human-labeled data. It tests different architectures and pre-trained models to enhance summarization quality.	The document mentions using Yelp dataset.	The MeanSum model performs well in evaluations, surpassing extractive models in ROUGE scores and sentiment accuracy.	Designed for multi-document summarization and doesn't work well for single-document summarization due to the lack of redundancy cues.

[14]	A combination of neural language models and different encoders for context capture. The model is trained using negative log-likelihood on input-summary pairs, and beam search decoding is utilized for efficient summary generation.	DUC-2003 and DUC-2004, The training data is sourced from the Gigaword dataset, with training pairs.	ABS and MOSES+ models outperformed TOPIARY.	Pretrained language models like BERT can inherit biases from the internet data they are trained on, which can result in biased or objectionable outputs.

Table 2. Evaluation Metrics for Text Summarization

Metric	Information	Papers Cited In
ROUGE	Recall-Oriented Understudy for Gisting Evaluation (ROUGE) is a widely-used metric for checking how well text summaries perform. It measures the overlap between machine-generated summaries and reference summaries, considering precision, recall, and the F1 score. It's a tool that helps us assess how accurately our summaries capture the essence of the original text.	[5], [7], [20], [11], [15], [3], [16], [17], [8], [18], [13], [19]
F1-Score	The F1-Score is a metric that combines precision and recall into a single value, making it useful for summarization evaluation.	[20], [11]
MMR-MD Parameter	Maximal Marginal Relevance for Multi-Domain (MMR-MD) parameter is used to control redundancy in multi-document summaries. It affects the trade-off between removing redundancy and maintaining relevance in summaries.	[21]

4 Implementation

In this research, we dive into two main ways of summarizing text: extraction-based and abstraction-based approaches. The spotlight is on extraction-based summarization, which involves pulling out keyphrases from source documents without changing the original text. This method proves particularly handy for handling diverse types of text resources like books, news articles, blog posts, research papers, emails, and tweets. This work is essentially driven by an algorithm called TextRank, which takes its cues from the PageRank algorithm, a concept initially explored by Jinghua Wang et al. [22]. You might be familiar with PageRank—it's the algorithm commonly used to rank web pages in online search results. In our work, we applied TextRank to develop a text summarization system. What TextRank does is identify similarities between sentences in articles, helping us generate concise and informative summaries. The proposed approach included preprocessing a dataset containing articles, including attributes such as article ID, title, text, and source. We imported the data into a Pandas dataframe, extracted the article text, and performed necessary preprocessing steps. It included converting to lowercase, removing punctuation, numbers, and special characters, as well as eliminating stopwords. These clean sentences were then represented as vectors using GloVe word vectors which is one of the various word embeddings provided by Jeffrey Pennington et al. [23]. In the implementation phase, a systematic process is followed as:

1. **Sentence Representation:** Sentences were represented as vectors using word embeddings.
2. **Similarity Matrix Preparation:** A similarity matrix was constructed based on sentence similarities.
3. **Graph/Network Conversion:** The similarity matrix was transformed into a graph/network.
4. **TextRank Algorithm Application:** The TextRank algorithm was applied to rank sentences.
5. **Summary Extraction:** A summary was extracted by selecting the top N sentences based on their rankings.

4.1 Results

Figure 3 shows the result of text summarization using TextRank algorithm in the form of Generated Summary and ROUGE scores. Table 3 presents the comparison of calculted ROUGE sores between GPT 3.5 and Bard. This demonstrates the effectiveness of TextRank-based summarization system in generating bullet-point summaries from multiple articles. This work contributes to the broader field of NLP and addresses the practical need for efficient information condensation in the era of information overload.

Table 3. Comparison of ROUGE scores

Parameters	GPT 3.5	Bard
Rouge-1	0.8889	0.8275
Rouge-2	0.8824	0.8055
Rouge-L	0.8889	0.8275

```
GENERATED SUMMARY:
Automatic summarization is the process of reducing a text document with a computer program in order to create a summary that retai
ns the most important points of the original document.
ROUGE Scores: [{'rouge-1': {'r': 0.7058823529411765, 'p': 1.0, 'f': 0.8275862020451843}, 'rouge-2': {'r': 0.6744186046511628, 'p':
 1.0, 'f': 0.8055555507445988}, 'rouge-l': {'r': 0.7058823529411765, 'p': 1.0, 'f': 0.8275862020451843}}]
```

Fig. 3. Result of Text Summarization using TextRank algorithm in the form of Generated Summary and ROUGE scores

5 Conclusion

This comprehensive survey provides as a definitive exploration of the intricacies surrounding query-based text summarization. Through the meticulous categorization of research papers, it has illuminated four distinct taxonomies, artfully differentiating between machine learning methods, namely supervised and unsupervised, and the summary types, whether extractive or abstractive. By doing so, this survey not only highlights the remarkable developments in query-based text summarization but also underscores the notable gaps and scarcities in certain taxonomies, offering a concise summary of the current state of research. Furthermore, it emphasizes the potential for cross-pollination of methodologies from the broader domain of generic text summarization into the specialized field of query-based summarization. A text summarization based on TextRank algorithm is implemented and results are reported for the same.

References

1. Afantenos S, Karkaletsis V, Stamatopoulos P. Summarization from medical documents: a survey. Artif Intell Med. 2005 Feb;33(2):157-77. doi: 10.1016/j.artmed.2004.07.017. PMID: 15811783.
2. Mariana Damova and Ivan Koychev. 2010. Query-based summarization: A survey. (2010).
3. Aakash Sinha,Abhishek Yadav, Akshay Gahlot "Extractive Text Summarization using Neural Networks" arXiv:1802.10137 2018
4. Ángel Herandez-Castenda , Rene Arnulfo Garcia-Hernandez, Yulia Ledenza, And Christian Eduardo Millan-Hernandez "Extractive Automatic Text Summarization Based on Lexical-Semantic Keywords"IEEE 2020
5. Gayathri Venu,Madhuri Chandu "Extractive Approach For Query Based Text Summarization" , IEEE,2020
6. Ahmed A. Mohamed, Sanguthevar Rajasekaran "Improving Query-Based Summarization Using Document Graphs" IEEE 2006
7. Samridhi Murarka, Akshat Singhal "Query-based Single Document Summarization using Hybrid Semantic and Graph-based Approach ", IEEE,2020
8. Stergos Afantenos, Vangelis Karkaletsis, and Panagiotis Stamatopoulos "Summarization from medical documents: a survey" Artificial intelligence in medicine 33, 2 (2005), 157–177.

9. Johan Hasselqvist,Niklas Helmertz,Mikael Kågebäck "Query based abstractive summarization using Neural Networks" arXiv:1712.06100 2017
10. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova.
11. Xinghuo Ye; Hai Wei "Query-Based Summarization for Search Lists" IEEE,2008
12. Jon M Kleinberg. 1999. Authoritative sources in a hyperlinked environment. Journal of the ACM (JACM) 46, 5 (1999), 604–632.
13. Eric Chu and Peter Liu. 2019. MeanSum: a neural model for unsupervised multi-document abstractive summarization. In International Conference on Machine Learning. PMLR, 1223–1232.
14. Nazreena Rahman, Bhogeswar Borah "A survey on existing extractive techniques for query-based text summarization" 2015 International Symposiwn on Advanced Computing and Communication (ISACC)
15. Yan Du And Hua Huo "News Text Summarization Based on Multi-Feature and Fuzzy Logic" IEEE 2020
16. Gupta, Virendra Siddiqui, T.J.. (2012). Multi-document summarization using sentence clustering. 4th International Conference on Intelligent Human Computer Interaction: Advancing Technology for Humanity, IHCI 2012. 1-5. 10.1109/IHCI.2012.6481826.
17. Y. K. Meena, A. Jai and D. Gopalani "Survey on graph and cluster based approaches in multi-document text summarization" , IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014), IEEE, Jaipur, India, 2014
18. Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014).
19. John Conroy, Judith D Schlesinger, and Dianne P O'leary. 2006. Topicfocused multi-document summarization using an approximate oracle score. In Proceedings of the COLING/ACL 2006 main conference poster sessions. 152–159. 2018. Bert: Pre-training of deep bidirectional transformers for language
20. Nazreena Rahman, Bhogeswar Borah "A survey on existing extractive techniques for query-based text summarization", IEEE , 2016
21. A.A.Mohamed, "Generating User Focused Content Based Summaries for multi-Documents Using Document Graphs", 5th IEEE symposium on signal processing and information technology (ISSPIT), pp.675-679, 2005.
understanding. arXiv preprint arXiv:1810.04805 (2018).
22. Wang, J., Liu, J., Wang, C."Keyword Extraction Based on PageRank. In: Zhou, ZH., Li, H., Yang, Q. (eds) Advances in Knowledge Discovery and Data Mining". PAKDD 2007. Lecture Notes in Computer Science(), vol 4426. Springer, Berlin, Heidelberg.
23. Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. GloVe: Global Vectors for Word Representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.