# TEXT SUMMARIZATION AND ITS EVALUATION TECHNIQUE

**Rimpi Rani[1], Baljinder Kaur[2]**

[1,2]Guru Kashi University, Talwandi Sabo

## ABSTRAC

Text data from a number of sources has grown rapidly in recent years. Automated text summarising creates a concise summary of the original content that incorporates all essential information. In the last few years, a massive amount of text data from a wide range of sources has been coming out. In order to be helpful, this vast amount of information and expertise must be adequately summarised. As a result, this study makes a double contribution. There are several strategies for extracting text summarization that are discussed in this work. Even more challenging are extraction issues for single and multiple document summarization. Textual evaluation and textual similarity measurement issues are at the heart of this project. Every text summarising situation can benefit from addressing the issues raised here. Extractive summarization strategies that address the highlighted issues are then reviewed.

**Keywords**: Summarization, Text Summarization, Evaluation techniques, Graph based summarization, Meta-heuristic based summarization, Maximal marginal relevance based summarization

### I. INTRODUCTION

People are becoming overwhelmed by the vast amounts of data and documents available online as the internet and big data continue to expand. Many scholars are motivated by this to devise a method for automatically summarising texts. Automated text summarization provides summaries that incorporate the most significant sentences from the original content, as well as all pertinent information (Gambhir and Gupta, 2017). In this way, the data is delivered promptly while maintaining the original objective of the paper (Murad and Martin, 2007). Using a statistical approach called word frequency diagrams, Lun (1958) was the first to openly disclose text summarising research in the mid-20th century. To date, a wide range of strategies have been developed. There are single and multi-document summaries depending on the amount of documents. Meanwhile, the extractive and abstractive outcomes are based on the summary results.

Text summarising is the process of creating a succinct and readable summary of a lengthy piece of writing while retaining its general meaning. Numerous techniques to automatic text summarization have been developed in recent years and widely used in a variety of fields. Snippets, for example, are generated by search engines as previews of content. Some more examples are news sites which reduce news stories into headlines in order to make surfing easier, or information extraction techniques in a variety of different fields (Pouriyeh et al, 2017).

Text summarization is a difficult task for humans since we often read a piece of text in order to get a comprehension of it before writing a summary that focuses on its most important parts. Automated text summarization is challenging because computers lack human understanding and linguistic capabilities.

Since the 1950s, text summary has been increasingly popular. For summarising scientific documents, was a key research of these days. Using variables like word and phrase frequency, Luhn (1958) developed a technique for extracting significant phrases from the text. High frequency terms should be used to weight sentences, while highly common words should be ignored. There are three techniques for determining the weight of sentences, as outlined by Edmundson (1969), in addition to the traditional frequency-based method.

**1) Cue Method:** There are certain cue terms in the cue dictionary that may be used to determine whether or not a statement is relevant.

**2) Title Method:** It is possible to calculate a sentence's weight by adding all of the content words found in a text's title and headers.

**3) Location Method:** It is assumed that statements at the beginning of a paragraph are more likely to be relevant than those at the beginning of a text as a whole.

Since then, a slew of papers tackling the issue of text summarization have been released. Extraction and abstraction are the two most common techniques to automated summarization. Rather than relying on the original text, extractive synopsis depends just on the extraction of expressions from the source material. Strategies for abstractive synopsis, then again, are intended to create critical data in another manner. To put it another way, they utilize strong normal language methods to interpret and break down the first text to create another more limited text that conveys the main data from it. There has been a great deal of interest on extractaive summing up, in spite of the way that human synopses are frequently non-unequivocal. With regards to rundowns, extractive outlines for the most part beat robotized synopses (Binwahlan et al, 2009). To put it another way, abstractive rundown strategies manage issues like derivation and normal language creation that are more troublesome than information driven approaches like sentence extraction can deal with. In all actuality there is at present no strategy for dynamic synopsis. It is normal for existing abstractive summarizers to depend on an extractive preprocessing part to build the text's theoretical.

It is a difficult subject in Natural Language Processing (NLP) as a result of the trouble in grasping each place of the text in a record. With a great deal of word information, it is feasible to do a definite examination of the text, including semantic investigation, lexical relations, and named substance ID. The trouble in getting word information in different aspects, like the significance of a word comparable to other data, related terms, inferential understanding, sentence creation, etc, has made it hard to produce abstracts as outlines.Abstractive summarization is the NLP term for this form of summary. Extractive summarization is a type of approximation that is more forgiving. Additionally, the system must select the text's most important/relevant material, extract it, organise it, and provide it to the user. When it comes to the extraction job of summarising, which has been studied since 1958 (Luhn, 1958), there is still more work to be done in the area of computerised summarization.

## II. PHASES OF TEXT SUMMARIZATION

One way to find the information contained in an entire document is using automated text summarising (ATS). A text summarising system, according to Elrefaiy et al (2018), extracts just the most important information from a document in order to produce a condensed version. In general, there are three stages to the summarising process:

- Text representation is obtained by analysing the document's text.
- Translation of the text into a simplified summary
- Summaries can be generated by transforming summaries into text

Phases one, two, and three require the following processes, components, and resources.

**Sentence Seperation:** Identifying the particular sentences in a document that are used in summarising is the goal of this technique.

**Stop words removal:** Most common words, such as articles, prepositions, conjunctions and interrogative questions and answers are removed during the stop-word removal procedure. The stop words are omitted from the sentence extraction procedure because of their insignificance.

**Stemming:** Converting semantically derived terms into their morpheme equivalents is the goal. The Porter stemmer is used to stem English language content. For systems dealing with semantic analysis, the stemming process may have a negative or inconsequential influence, according to Toman et al (2006). So, we've tried out the recommended method using both stemming and pre-processing methods (with and without).

**Part-of-Speech Tagging:** It's a method for figuring out which words in a phrase are nouns, adverbs, verbs, and so on. POS tags like 'nounplural,' on the other hand, are more commonly used in computational applications. The Stanford Log-linear POS tagger was used in this example.

**Keywords extraction:** In this stage, the keywords in a document are retrieved. All words except stop words are treated as keywords in this context.

## III. APPROACHES FOR EVALUATION

According to Goularte et al (2019), evaluations are conducted in three stages, each of which is briefly explained below: co-selection assessment (with a reference summary), content evaluation (without a reference summary), and original document review.

**1) Co-selection based Evaluations:** According to a co-selection based evaluation, it is necessary to have a reference summation for comparison. It is done by picking common phrases from the system summary and the reference summary. Recall, precision, and F-score are all linked measures for co-selection-based assessment.

**Recall:** It is the overall number of correct sentences recovered from a text divided by the total number of correct sentences retrieved and incorrect sentences retrieved. The following is a rough estimate:

$$Recall = \frac{\sum_{s\in sys}\sum_{gram_N\in s} Count_{match}(gram_N)}{\sum_{s\in ref}\sum_{gram_N\in s} Count(gram_N)} \qquad (1)$$

Countmatch(gramN) is the maximum number of N-grams that appear in both the system summary and the reference summary, where re f stands for reference summary. The number of N-grams in the reference summary is counted.

**Precision:** The total number of correct sentences recovered from the document divided by the total number of correct and wrong sentences retrieved from the document. The following formula can be utilised to arrive to this result:

$$Precision = \frac{\sum_{s\in sys}\sum_{gram_N\in s} Count_{match}(gram_N)}{\sum_{s\in sys}\sum_{gram_N\in s} Count(gram_N)} \qquad (2)$$

Count(gramN) is the number of N-grams in the system summary, and sys is a member of that summary.

**F-score:** A user's recall is given times more weight than accuracy in this test, which examines the effectiveness of retrieval. Non-negative real (0 ) has the following F-score:

$$F_\beta = \frac{(1+\beta^2)(Precision * Recall)}{(\beta^2 * Precision + Recall)} \qquad (3)$$

**Improved Rates:** On the basis of the above-mentioned parameters, the proposed approaches perform better than other methods in terms of improved rates (IR).

$$IR = \frac{(PM - OM)}{OM} \qquad (4)$$

where, PM is the proposed strategy, OM is the other technique, and IR are the better rates.

**2) Content based Evaluations:** A co-choice strategy takes a gander at an outline framework utilizing similar words. it can't associate thoughts, stream of sentences, relatedness of words, and non-overt repetitiveness in a synopsis on the grounds that these things are impractical in an outline. Assuming you utilize the substance based strategy, you can manage these things. Coming up next are a few substance based strategies for assessing a text. They consider various things about the text. The substance based assessment just requirements an outline of the framework. The measurements for content-based assessment resemble this:

**Cohesion:** It's critical to understand how a piece of writing connects to the rest of the text. It has been shown that there are five sorts of coherent relationships: conjunction, reference, ellipsis/substitution, and lexical/lexical. The following are the primary categories: (Halliday and Hassan, 2014) Two sentences can be linked via anaphoric and cataphoric referential links. "and" is the most basic and least coherent connection between sentences. An anaphoric link occurs when a phrase refers back to something that has been previously discussed. To put it another way, cataphoric is the complete polar opposite of anaphoric. The words are eliminated from a repeated sentence after a more extensive explanation. The ellipsis connection is the term for this. Instead of ellipses, the words are substituted with more broad ones. This is distinct from ellipsis.

**Non-redundancy:** A summary's non-redundancy relates to its originality. The non-redundant summary has been explored by a number of researchers (Oufaida et al., 2014). A non-redundant summary, they say, can better capture the full breadth of information contained inside a given document. So, calculating non-redundancy in the produced summary would be interesting.

**Readability:** The readability of a summarizer's work is also an important metric to consider. Text readability shows us how quickly and readily a piece of material may be comprehended. There are two ways to evaluate a text's readability: content and how closely connected a sentence is to its preceding sentence. When it comes to content readability, vocabulary and grammatical complexity are key, but the degree to which a sentence is connected to what came before it is a sign of reading fluency, as well.

## IV. ISSUES

According to Gambhir and Gupta (2017), summarization of texts raises a number of challenges, including the following:

**Redundancy:**

In summarising a document, redundancy is usually a bad thing. As long as it doesn't include information that has already been covered, a summary is more useful. The majority of existing techniques focus on extracting useful text from a document and generating a summary of that information. As a result, similarity assessment is critical in identifying repetitive material in a text. Redundancy in the summary can be minimised if the similarity between the contents of a document can be properly measured.

**Irrelevancy:**

If you're looking to get a rapid overview of a whole text, summarization systems are your best bet. These qualities are often used to evaluate the sentences or sections of a document. To avoid unnecessary information, some features may be omitted from summaries because it is not possible to include them all. Complexity and insignificance are both increased by considering all potential text elements for sentence evaluation. Because of this, understanding which characteristics are in charge of generating a high-quality summary from the provided data is essential. Furthermore, in a reference summary, all of the qualities studied are not presented in the same proportion.

**Loss of coverage:**

Generic text summarising relies heavily on the summary's ability to cover the document's main points. The material in a decent generic summary should include all of the information that is provided in the text. However, query-based summarization may not necessarily necessitate this. Topics aren't given much attention in the summaries provided by existing summarising methods. They therefore fail to create excellent summaries when summarising generically. This is especially true when summarising many publications, as the number of themes in each document is significantly more than in a single document.

**Non-readability and less cohesive content:**

Readability and cohesion are essential for a successful summary. The terms "readable" and "cohesive" refer to the fact that the summary's contents should have a common conceptual framework.

## V. CLASSIFICATION OF SUMMARIZATIOM TECHNIQUES

**Graph based methods:**

In these approaches, graphs are built up of edges, which indicate the connections between two points on the graph, which in this case represents a node for each word included inside the text. Nodes are scored based on their structured and non-structured text properties, and the similarity between phrases is an important factor in navigating the network. It takes use of WordNet-based semantic similarity in order to find similarity scores and to compute sentence relevance, among other things. It also takes use of ten text properties for extraction-based summarization, which are described in detail below (Sripada et al, 2005). It is necessary to apply the affinity matrix of sentence similarity to identify the locally and internationally vital sentences in a range of sources. A unified technique based on four assumptions is employed to locate the locally and internationally key phrases in the various sources (Wan, 2010). G-FLOW (Christensen et al., 2013) is a method that measures coherence and salience components based on the estimated discourse graph. The G-FLOW system produces an ordered summary by optimising the components of coherence and salience to the greatest extent possible. Glavas and Snajder (2014) developed an event-based summarising strategy for event-oriented document gathering that makes use of the ability of machine learning rule-based methodologies to achieve high levels of accuracy. This strategy is intended to be used in conjunction with event-oriented document gathering.

An extractive summarizer based on the multilayer technique by Tohalino and Amancio (2018) weights the edges in a network of documents using a variety of metrics such as degree and strength as well as page rank and accessibility. These metrics include symmetry, shortest path, absorption time, and so on. In their research, they found that creating a clear distinction between the intra- and inter-layer edges may significantly improve the performance of a summarizer. In a survey conducted by Liu et al (2018), summarization methods that employ graphs as input are classified based on the input and associated procedures used in the approach. The findings of this survey lead to a number of areas of research that have still to be investigated, including document temporal graph summarization and developments in standardising and generalising algorithms, among other things. When it comes to dealing with huge volumes of data, these solutions have a significant flaw. Consequently, it's probable that their talents are limited to summarising a single page as a consequence. The development of meta-heuristic summarising approaches based on these meta-heuristic methodologies is recommended as a solution to this challenge.

**Maximal Marginal Relevance based methods:**

Methods that use maximum marginal relevance (MMR) for summarising aim to ensure that the summaries they create contain only important information, while also ensuring that there is minimum overlap across the summaries they produce. Goldstein and Carbonell (1998) provide a multi-document summarising approach based on maximal marginal relevance. This approach strikes a balance between comprehensiveness and relevance in the summary by using query factors. 0.7 is the weight given to relevance, while 0.3 is the weight given to nonredundancy. Compression, speed, redundancy, and passage selection are all major concerns in multi-document summarising, according to Goldstein et al. (2000). According to Wang et al. (2009), a way of enhancing accuracy scores by examining the link between the MMR model and email content cohesiveness may be found in their summary of e-mail data. An MMR model with a naive tone biassing model is presented by Chaudhari and Mattukoyya (2018). In the Naive biassing model, the material created by the MMR model is employed as an input by means of polarity tags. Meta-heuristic approaches that look at MMR as an optimization issue explain some of the MMR-based strategies. As a rule of thumb, the performance of these approaches cannot be guaranteed to include coverage and avoid redundant information. Clustering-based summarising algorithms are proposed in order to overcome this restriction.

**Meta-heuristic based methods:**

The last several years have seen a flurry of activity in the fields of single and multi-document summarization optimization methods for example, Genetic Algorithms, Particle Swarm Optimization, Harmony Search, and Differential Evolution. (He et al., 2016) used genetic algorithms to retrieve significant phrases based on four summary factors: fulfilled length, high coverage, high informativeness, and minimal redundancy. This approach employs WordNet to find phrase frequency, which considers term similarity.

They offer a cuckoo search strategy for summarising that employs three features: coverage, coherence, and readability. Alguliev et al. (2013) suggest optimising content coverage, diversity, and summary length through differential evolution. Cosine functions measure similarity. Alguliev et al. (2011) created a PSO-based summarising model that eliminates repetition. With this method, you may maximise relevance, redundancy, and summary length at the same time (NGD). Metaheuristic methodologies employed for summarization tend to become stuck at local optima, which is a shortcoming of these methods. On top of that, these methods don't reveal anything about the search space behaviour of a function like steepness or extremes. As a result, the suggested method employs a gradient-based optimization strategy in order to significantly speed up the convergence of the algorithm.

**Other methods:**

Techniques for summarization that are based on Reinforcement Learning have been discussed (RL). Real-time optimization (RL) is generally used to improve nondifferentiable functions, such as ROUGE, in the real world. In this regard, Narayan et al. (2018) describe an extractive summary learned by maximising the ROUGE function. The sentence encoder uses convolutional neural networks (CNNs) for continuous sentence representation, whereas the document encoder uses recurrent neural networks (RNNs) Plus long short-term memory (LSTM) to overcome the vanishing gradient problem. The policy gradient RL's cross-entropy loss and rewards are also used as objective functions in this model's optimization.The abstractive summarization model developed by Paulus et al. (2017) includes an attention mechanism as well as a learning objective for dealing with redundancy. Building the encoder and decoder for this sentence involves the usage of RNNs. It has been proved empirically that ROUGE is not the only metric to optimise in summarization, as demonstrated by the success of this approach in summarising big documents. Lee and Lee (2017) offer a single document summarising strategy based on the Q-Network that utilises both content-based embedding features and position-based embedding characteristics to select the phrases that should be summarised.

Regarding sentence ranking, phrase similarity, and text representation, Mehta and Majumder (2018) undertake a comparative evaluation of existing text summarizers in terms of their performance in these areas. Their research has led them to assume that summary systems can benefit from the systematic application of a variety of summarising methodologies. According to Goularte et al. (2019), automated text evaluation may be accomplished by applying fuzzy rules for text summarization in conjunction with a text evaluation algorithm. As illustrated by these data, fuzzy summarising can help to enhance the quality of the summaries that are provided. According to Hu et al. (2017), one technique to summarising opinion data (Opinosis) shows that the clustering of reviews might play a significant role in the coverage of reviews associated to each occurrence. Wang et al (2017) go into great length about heuristic-based techniques for extracting phrases from enormous amounts of textual data. In their opinion, deleting superfluous material from a text can increase the speed with which a summarising algorithm can discover relevant words and summaries. Tayal et al. (2017) developed a summarising technique that employs clustering to find related sentences in a document and combines them based on their similarities. They do so using soft computing methodologies. Feature Symmetry with Heterogeneous Characteristics Developed by Wei et al. (2016), summarization is a way of condensing diverse characteristics into a single summary (HFSS). The sentiment analysis and summarising phases of Abdi et al. (2018)'s QMOS query-based summarization approach are separated into two parts. An extensive number of various sentiment lexicons are integrated in order to do the semantic sentiment analysis. The sentences that are utilised in the summary have been subjected to a syntactic and semantic examination. For Arabic text documents, Abstractive summaries can be created using a sentence reduction strategy and a rhetorical structure theory-based sentence extraction approach. In a study of swarm intelligence (SI)-based summaries, Mosa et al. (2018) found that SI systems for summarising are rather constrained in their use cases. They have built their own framework for summarising a multi-objective optimization problem using SI. As demonstrated by Sanchez-Gomez et al. (2018), an ABC-based summarising technique may be used to effectively summarise large datasets

## VI. CONCLUSION

The fast expansion of the Internet has resulted in a vast volume of information becoming freely available to the public. Large amounts of text are difficult for people to comprehend and summarise. Therefore, in today's world of information overload, there is a crucial need for automatic summarization tools to help people make sense of it all. In this work, the topic of document summarization has been explored. The researchers addressed a number of challenges as well as evaluations of summarising methodologies in this research. There have been numerous concerns raised about various techniques as a result of these discussions, including the increase in data size, the requirement for prior knowledge about the number of clusters present in a dataset, the uncertainty of coverage, and the inclusion of non-redundancy features in the summary, amongst other factors. Furthermore, further study is necessary in order to develop more efficient techniques of summarising information.

**REFERENCES**

1. Abdi, A., Shamsuddin, S. M., and Aliguliyev, R. M. (2018). Qmos: Query-based multi-documents opinion-oriented summarization. *Information Processing & Management*, 54(2), 318–338.
2. Alguliev, R. M., Aliguliyev, R. M., and Isazade, N. R. (2013). Multiple documents summarization based on evolutionary optimization algorithm. *Expert Systems with Applications*, 40(5), 1675–1689.
3. Alguliev, R. M., Aliguliyev, R. M., Hajirahimova, M. S., and Mehdiyev, C. A. (2011). Mcmr: Maximum coverage and minimum redundant text summarization model. *Expert Systems with Applications*, 38(12), 14514–14522.
4. Binwahlan, M., Salim, N., Suanmali, L. (2009). Fuzzy swarm diversity hybrid model for text summarization. *Inf. Process. Manag*., 3, 23-33.
5. Chaudhari, M. and Mattukoyya, A. N. (2018). Tone biased mmr text summarization. arXiv preprint arXiv:1802.09426.
6. Christensen, J., Soderland, S., Etzioni, O. (2013). Towards coherent multi-document summarization. In Proceedings of the 2013 conference of the North American chapter of the association for computational linguistics: Human language technologies, 1163–1173.
7. Edmundson, H. (1969). New methods in automatic extracting. *Journal of the ACM (JACM)*, 16(2), 264–285.
8. Elrefaiy, A., Abas, A., Elhenawy, I. (2018). Review of recent techniques for extractive text summarization. *J. Theor. Appl. Inf. Technol*., 96, 7739-7759.
9. Gambhir, M., Gupta, M. (2017). Recent automatic text summarization techniques: A survey *Artif. Intell. Rev*., 47, 1-66.
10. Glavas, G. and Snajder, J. (2014). Event graphs for information retrieval and multi-document summarization. *Expert Systems with Applications*, 41(15), 6904–6916.
11. Goldstein, J. and Carbonell, J. (1998). Summarization:(1) using mmr for diversity-based reranking and (2) evaluating summaries. In Proceedings of a workshop on held at Baltimore, Maryland: October 13-15, 1998, pages 181–195. Association for Computational Linguistics.
12. Goldstein, J., Mittal, V., Carbonell, J., and Kantrowitz, M. (2000). Multi-document summarization by sentence extraction. In Proceedings of the 2000 NAACL-ANLP Workshop on Automatic summarization, 40–48. Association for Computational Linguistics.
13. Goularte, F., Modesto, N., Fileto, R., Saggion, H. (2019). A text summarization method based on fuzzy rules and applicable to automated assessment. *Expert Syst. Appl*., 115, 264-275.
14. Halliday, M. A. K. and Hasan, R. (2014). *Cohesion in english*. Routledge
15. He, R., Tang, J., Gong, P., Hu, Q., and Wang, B. (2016). Multi-document summarization via group sparse learning. *Information Sciences*, 349:12–24
16. Hu, Y.-H., Chen, Y.-L., and Chou, H.-L. (2017). Opinion mining from online hotel reviews–a text summarization approach. *Information Processing & Management*, 53(2), 436– 449.

17. Lee, G. H. and Lee, K. J. (2017). Automatic text summarization using reinforcement learning with embedding features. In Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers), volume 2, 193–197.
18. Liu, Y., Safavi, T., Dighe, A., and Koutra, D. (2018). Graph summarization methods and applications: A survey. *ACM Computing Surveys (CSUR)*, 51(3), 62.
19. Luhn, H. (1958). The automatic creation of literature abstracts. *IBM Journal of Research and Development*, 2(2), 159–165.
20. Lun, H. (1958). The automatic creation of literature abstracts. *IBM J*, 159-165.
21. Mehta, P. and Majumder, P. (2018). Effective aggregation of various summarization techniques. *Information Processing & Management*, 54(2), 145–158.
22. Mosa, M. A., Anwar, A. S., and Hamouda, A. (2018). A survey of multiple types of text summarization with their satellite contents based on swarm intelligence optimization algorithms. Knowledge-Based Systems.
23. Murad, M., Martin, T. (2007). Similarity-based estimation for document summarization using fuzzy sets. *Int. J. Comput. Sci. Secur*., 1, 1.
24. Narayan, S., Cohen, S. B., and Lapata, M. (2018). Ranking sentences for extractive summarization with reinforcement learning. arXiv preprint arXiv:1802.08636.
25. Oufaida, H., Nouali, O., and Blache, P. (2014). Minimum redundancy and maximum relevance for single and multidocument arabic text summarization. *Journal of King Saud University-Computer and Information Sciences*, 26(4), 450–461.
26. Paulus, R., Xiong, C., and Socher, R. (2017). A deep reinforced model for abstractive summarization. arXiv preprint arXiv:1705.04304.
27. Pouriyeh, S., Vahid, S., Sannino, G., De Pietro, G., Arabnia, H. and Gutierrez, J. (2017). A comprehensive investigation and comparison of machine learning techniques in the domain of heart disease, in Computers and Communications (ISCC), 2017 IEEE Symposium on. IEEE, 204–207.
28. Sanchez-Gomez, J. M., Vega-Rodr´ıguez, M. A., and Perez, C. J. (2018). Extractive multi-document text summarization using a multi-objective artificial bee colony optimization approach. *Knowledge-Based Systems*, 159, 1–8.
29. Sripada, S., Kasturi, V. G., and Parai, G. K. (2005). Multidocument extraction based summarization. CS 224N, Final Project.
30. Tayal, M. A., Raghuwanshi, M. M., and Malik, L. G. (2017). Atssc: Development of an approach based on soft computing for text summarization. *Computer Speech & Language*, 41, 214–235.
31. Tohalino, J. V. and Amancio, D. R. (2018). Extractive multidocument summarization using multilayer networks. *Physica A: Statistical Mechanics and its Applications*, 503, 526–539.
32. Wan, X. (2010). Towards a unified approach to simultaneous single-document and multi-document summarizations. In Proceedings of the 23rd international conference on computational linguistics, pages 1137–1145. Association for Computational Linguistics.
33. Wang, B., Liu, B., Sun, C., Wang, X., and Li, B. (2009). Adaptive maximum marginal relevance based multi-email summarization. In International Conference on Artificial Intelligence and Computational Intelligence, pages 417–424. Springer.
34. Wang, W., Li, Z., Wang, J., and Zheng, Z. (2017). How far we can go with extractive text summarization? heuristic methods to obtain near upper bounds. *Expert Systems with Applications*, 90, 439–463.
35. Wei, W., Ming, Z., Nie, L., Li, G., Li, J., Zhu, F., Shang, T., and Luo, C. (2016). Exploring heterogeneous features for query-focused summarization of categorized community answers. *Information Sciences*, 330, 403–423.