

# Query-Focused Multi-document Summarization Survey

Entesar Alanzi, Safa Alballaa

Computer Science Department, Imam Mohammad Ibn Saud Islamic University, Riyadh, Saudi Arabia

**Abstract**—With the exponential growth of textual information on the web and in multimedia, query-focused multi-document summarization (QFMS) has emerged as a critical research area. QFMS aims to generate concise summaries that address user queries and satisfy their information needs. This paper provides a comprehensive survey of state-of-the-art approaches in QFMS, focusing specifically on graph-based and clustering-based methods. Each approach is examined in detail, highlighting its advantages and disadvantages. The survey covers ranking algorithms, sentence selection techniques, redundancy removal methods, evaluation metrics, and available datasets. The principal aim of this paper is to present a thorough analysis of QFMS approaches, providing researchers and practitioners with valuable insights into the field. By surveying existing techniques, the paper identifies the challenges and issues faced in QFMS and discusses potential future directions. Moreover, the paper emphasizes the importance of addressing coherency, ambiguity, vague references, evaluation methods, redundancy, and diversity in QFMS. Performance standards and competing approaches are also discussed, showcasing the advancements made in QFMS. The paper acknowledges the need for improving summarization coherence, readability, and semantic efficiency, while balancing compression ratios and summarizing quality. Additionally, it highlights the potential of hybrid methods and the integration of extractive and abstractive techniques to achieve more human-like summaries.

**Keywords**—Text summarization; query-based extractive text summarization; multi-document; graph-based approach; clustering-based approach

## I. INTRODUCTION

Computers and human interaction have significantly impacted on Natural Language Processing (NLP). The evolution of NLP has a direct influence on numerous fields and serves a wide range of applications. Most importantly, it assists computers in recognizing and understanding human language. Recently, there has been a rapid growth of available documents online. As a result, retrieving helpful information from the vast amount of electronically accessible documents available has become a big challenge. Text summarization can be effectively used to reduce this issue.

Automatic summarization can be categorized based on different factors. Extractive and abstractive summaries are the two major categories [1]. An extractive summary is created by linking some chosen sentences from the input document to form a summary. The output summary presents these selected sentences precisely as they appear in the original text without any changes. In contrast, language-generation algorithms are used to produce an abstractive summary automatically. This

usually requires the system to do sentence compression, paraphrasing, and reformulation to make the summary look more human-like.

Moreover, depending on the number of input documents, summarization can be generally classified as either a single-document or multi-document system [2]. Early studies dealt with a single document where the system presented that document in a shorter form while retaining the most essential information. The use of multi-document summarization has become more critical with the growth of the internet. Given the massive volume of redundant content on the web, summarization can be more beneficial if they offer a concise summary of numerous papers on the same subject.

Summarization can be either a generic or query-focused [2]. A general summary gives an overview of the critical information from the input document to help the reader to understand its contents quickly. In this regard, the content of the entire input document determines the significance of the information. When the summary is generated based to a query, instead, the query itself chooses what data is essential and ought to be included in the output summary.

Query-focused multi-document text summarization (QFMS) is a relatively active automatic summarization subfield with many applications. It is a quick and efficient approach to navigating and grasping web texts, including news, articles, blogs, and data analysis. Search engines use query-focused text summarization methods to produce a summary of retrieved data, which helps the users to grasp the critical content quickly. These techniques save consumers' time, improving the search engine's service.

Moreover, in many sectors nowadays, chatbots automatically respond to users' inquiries and requests made through chat interfaces. A natural-language search query, such as "Return an item," is often the first step in using a chatbot on a shopping website, for example. The chatbot answers by displaying a summary of the results. Additionally, it is crucial for marketing since, through user inquiries, companies may learn more about what attracts customers and save this information through some data collection techniques. As a result, essential business choices may be made by examining and analyzing these unstructured data. Organizations may thus be proactive with their strategies and enhance their poise and confidence.

While existing text summarization surveys have mainly focused on generic summarization, query-focused summarization has received limited attention. To the best of

our knowledge, there are no state-of-the-art surveys investigating query-based summarization problems, such as extracting query-relevant sentences and reducing redundancy. Therefore, the primary motivation of this study is to provide a comprehensive review of existing studies on query-focused multi-document summarization, aiming to assist researchers in improving query-based summaries. This survey delves into two main QFMS approaches, their algorithms, sentence extraction techniques, similarity scoring methods, redundancy removal techniques, evaluation metrics, standard datasets, existing challenges, and future research directions.

The remainder of this paper is structured as follows: Section II presents a detailed description of the QFMS problem. The third section introduces the QFMS approaches and reviews the different methods proposed in the literature for each approach. The discussion of findings, open research problems, and future directions are presented in Section IV. Finally, Section V concludes the paper.

## II. PROBLEM STATEMENT

QFMS resolves the problem of extracting useful information from an extensive amount of data, which improves the effectiveness of obtaining and utilizing information. However, automatic text summarization has many challenges, particularly query-focused ones. Relevancy, diversity, and redundancy are the three main bottlenecks.

A summary must be relevant and provide information based on a given query. Query-based summarization is complicated because the user's query must evaluate the relevance of sentences and choose the ones suitable for inclusion in summary. In addition to selecting the most important information from all document sets, the system is supposed to ensure that the information is based on the specified query. As a result, the query's specific features should be included throughout the summarizing process by calculating the similarity measure between the query and each sentence from the input text and, then sorting the sentences based on the generated scores [3]. Taking into account only the precise match between the query's terms and the terms in the sentences does not provide the best measure. Additionally, doing that to multi-documents sets makes the process potentially more complicated since we have to deal with the variations and similarities across document sets.

Another crucial factor of a robust query-based summary is the diversity [4]. To ensure user satisfaction, different aspects of the question should be considered in the summary to the greatest extent possible. This task is difficult because it demands recognizing and modeling the connections between the sentences and how they relate to the query. Therefore, it is founded that diversity is the most essential and challenging task in QFMS that has interested many researchers lately [4].

Additionally, redundancy is a typical issue that practically all methods of multi-documents automatic summarization encounter [5]. A good summary should be more informative and less repetitive. In single-document summarization, every phrase is distinct and doesn't often contain redundant data. In contrast, information from multi-documents will overlap. In

fact, the information can be repetitive, or it can represent the same concepts in multiple ways without adding any new data. This issue makes the automatic summarization more difficult, as it involves finding and analyzing the connections among the sentences in all texts to remove redundant and repeated data.

The diversity of automatic summarization methodologies comes from different ways of tackling the ranking and selection issues. A ranking problem is a process of ranking all sentences in the input documents. This needs an algorithm that evaluates the importance of each sentence in accordance with the input inquiry. The selection problem is how to choose some of those ranked sentences to create the summary [6]. This requires a model that increases the diversity and decreases the redundancy to form an informative summary under a limited length. Fig. 1 shows a general architecture of a QFMS, which consists generally of the following steps:

1) *Pre-Processing*: This step is done for both input documents and the query. The objective is to reduce noisy and unfiltered text, decrease calculation time, and allow diverse term variants to be treated equally. This can be done using several NLP methods. The following techniques are used in the surveyed literature:

a) *Normalization*: Extends acronyms, lowercase all words, eliminates digits, or changes them to terms, etc.

b) *Tokenization*: Converts each sentence into a list of individual words [2].

c) *Stop-Word Removal*: Stopwords are the commonly used terms in a language, such as: how, are, to, etc. Removing such words drives attention to the important ones. They are irrelevant for search purposes, and they can disturb the result [7].

d) *Stemming*: From a text summarization aspect, stemming is the process of returning the words to their root form [8].

e) *Part-of-Speech Tagging*: Categorizes the terms into nouns, verbs, adjectives, and adverbs [9].

f) *Named Entity Recognition*: Classifies words as item names such as person name, location name, etc.

2) *Processing*: this includes the following:

a) *Creating text representation*: Generating a proper representation of the input documents to simplify the subsequent ranking process and selection. This representation can be graphs, clusters, topic models, etc.

b) *Ranking algorithm*: The input sentences are ranked according to relevance to the query, and they are then arranged from highest to lowest. This step varies depending on the approach being used.

c) *Selecting algorithm*: Choosing the best-ranked sentences considering the limited length. That length can be computed as a number of sentences, number of terms, or a ratio.

3) *Post-Processing*: This can include reordering the sentences, redundant sentence reduction, etc.

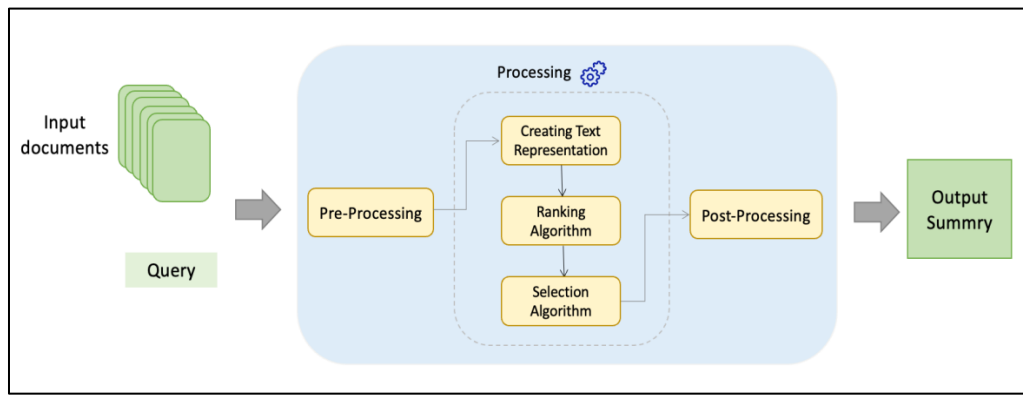


Fig. 1. General architecture of a query-focused multi-document summarizer.

### III. LITERATURE REVIEW

#### A. QFMS Approaches

Different approaches have been applied to QFMS, such as graph-based, clustering-based, machine learning-based, statistical-based, semantic-based, optimization-based, etc. [8]. In this survey, two main approaches will be reviewed separately in respective subsections. Fig. 2 presents a classification of the reviewed QFMS approaches.

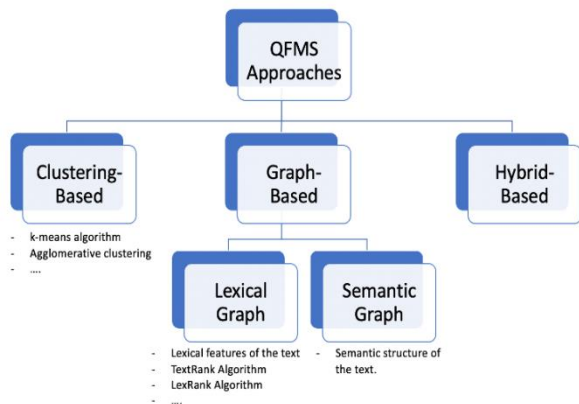


Fig. 2. A classification of the reviewed QFMS approaches.

1) *Graph-based approach*: The graph-based approach has been a commonly used approach for extractive text summarization because of its capacity to create sentence linkages within one document and linkages with sentences on other documents [10]. Particularly, it is best suited to extract a list of the essential query-related sentences from the documents [11]. A graph is a data structure that contains a set of nodes connected with edges. It is a domain-independent [2] and language-independent and can enhance the coherency [8], and reduce the redundancy [5], [12]. In QFMS, to display the input, a graph is employed such that each sentence from the input text is a node, and the weighted edges between nodes and the query measure the similarity between each respective pair. Then, the nodes are ranked using ranking algorithms such as PageRank [13], TextRank [14], and LexRank [15] algorithms to assess each sentence's significance. Finally, the summary is produced by selecting high-ranked sentences [8].

Lei and Zeng [16] used a manifold-ranking algorithm [17] to create a graph where edges represented the degree of similarity to the query and the vertices represented the sentences and the query. According to their study, the manifold ranking system, which used a graph-based placement algorithm, can quickly determine the most relevant and prestigious phrases to answer the query. Their model produced unique top-ranked sentences by modifying the iteration operation, and these sentences were utilized to build a review without the need for extra procedures. In comparison to earlier query-focused summarizing techniques, their method produced high-quality summaries. Wan and Xia [18] also created a multi-modality system that considered both intra-document relevance and other documents' similarities to improve the manifold-ranking algorithm. Their method outperformed the standard manifold-ranking algorithm.

Similarly, Wei et al. [19] argued that the inter-document links (i.e., the edges that link sentences from different documents) are more significant than the intra-document links (i.e., the edges that connect sentences from the same documents). They are supposed to be more comprehensive and able to capture more relevant information from the whole document set. Hence, they assigned the inter-document edges an additional weight compared to the intra-document edges. The results of the proposed approach outperform the best-performing systems in DUC2005. As aforementioned, redundancies are challenging in multi-document summarizations. Balaji et al. [5] presented a graph-based technique for QFMS that effectively minimized redundancies. In this study, graph matching was used to create a global semantic graph, which decreased the number of repeated sentences appearing in the summary.

Mohamed and Rajasekaran [20] created a document graph to represent the text document which has two forms of relations, "is a" and "related to." They made three tries, each time slightly altering the centric graph to include a generic summary. Then they expanded their work to include query-based summaries and finally introduced the query modification technique to incorporate additional query information. Although their solution outperforms many baseline systems in terms of performance, it is ineffective when the query contains a variety of hidden subtopics.

Due to the insufficient information that a query can express, query expansion is proposed [21]. Abdi et al. [22] expanded the terms in the query and the sentences using the Content Word Expansion (CWE) approach. The CWE is based on semantic similarity. Also, Jia et al. [21] proposed a query expansion technique that used -including the query itself- some external resources, which are: WordNet, mean, variance, and TexRank algorithm to expand the query. Better performance was found using their approach.

QFMS has effectively utilized the hypergraph-based concept. It can provide more precise similarity calculations [27]. Xiong and Ji in [26] developed a vertex-enhanced hypergraph approach. Using the cosine similarity metric, they used a topic model to group sentences according to the probability distribution of their topic. Then they expressed these distributions and the relationship between phrases using the hypergraph. A random walk algorithm determined the score of the sentences on the graph. Experiments on datasets showed an improvement.

Similarly, two summarization approaches are integrated by Akhtar et al. in [28] to benefit from both topic model-based and graph-based approaches. Their scoring technique used only common words for sentence ranking. They would like to enhance their work by considering semantic similarities in future work. Notably, Lierde and Chow [23] pointed out the two critical issues in graph-based summarization. First, the fact that each sentence covers multiple topics. Second, the joint relevance of sentences can't be measured by each sentence's individual relevance score, and this scoring tends to produce redundant summaries. To address these issues, they proposed a new summarizer based on hypergraph transversals, in which the nodes are sentences and the hyperedges are themes (topics). The hyperedge weights reflect both its importance and its relevance to the query. Hence, each hyperedge is associated with a specific topic, each node should belong to multiple hyperedges, and the themes may overlap. A summary is produced by generating a transversal of nodes in the hypergraph. Experiments on DUC 2007 dataset showed that their method outperforms the related graph and hypergraph-based approaches by at least 6% of the ROUGE-SU4 score. However, this approach is restricted to topical similarities between sentences. Authors would like to involve some linguistic features and discourse relations to enhance their model. Similarly, the same authors, Lierde and Chow [24] extended their work to develop a fuzzy hypergraph model where each node represents a sentence and fuzzy hyperedge is a topic. Sentences are scored according to how closely they relate to the query and how central they are to the hypergraph. In future work, the authors would like to improve the readability of their generated summaries.

One of the main drawbacks of the graph-based approach is that it doesn't consider the semantic structure of the sentence [29]. However, some researchers tried to handle this issue by taking into account additional language-dependent parallels, like semantic similarity [30]. Abdi et al. [31] also proposed a query-based summary method that combines sentiment analysis and summarization approaches. The proposed method has two main phases: 1) sentiment analysis, which calculates the sentiment score of each sentence and selects sentences that

have the same sentiment orientation of the opinion of the correlated query and passes them to the next phase 2) summarizer phase, it calculates the total score that combines the query-relevant score with the sentence sentiment score and rank them using a graph-based ranking algorithm. Although using a semantic graph gave good scoring, it required external knowledge sources. In the same manner, several statistical and semantic scoring techniques have been used by Krishna et al. [32] to assess how closely the user query matches the document's sentences, which are Word form similarity, N-gram similarity, Word Order Similarity, and Semantic similarity. Instead of applying a weighted scoring method (where each value has a predefined weight) to determine the overall score, they base it on the average (mean) value obtained using the abovementioned techniques. To avoid redundancy, an iterative clustering process is employed.

Conversely, three statistical features were proposed by He et al. [33] for sentence scoring. Similarity and Skip-Bigram co-occurrence are the first two query-dependent features. The third feature is a text graph's query-independent feature that is used to extract sentences with high information density. However, their approach might not be able to identify semantically equivalent sentences.

To support the assertion that multiple approaches in combination can improve text summarization, Murarka and Singhal [34] developed a hybrid system that combines the Latent Semantic Analysis (LSA) technique [35] and an enhanced PageRank algorithm to address the challenges of QFMS. The results showed a better performance of their hybrid model than many graph-based and semantic-based methods. In fact, PageRank [13] is a well-known graph-based technique to assess the relevancy of web pages by evaluating their related keywords, sentences, and reputable links. It commonly appears in text-summarization methods due to its capacity to extract meaning from texts. The PageRank method is suitable for giving significance to any collection of units with mutual references [36]. PageRank [37] and graph-based relevancy methods are used widely [38]. These graph-based methods emphasize global relevance and PageRank-inspired recursive scoring for phrase relevance. Generally, the model has a simple implementation, fast computation, and is language independent. However, they have reduced readability. Nastase [39] applied a summarization of Wikipedia and heavily relied on PageRank as a mechanism for measuring significance in the process. By utilizing the spreading activation technique in a graph, he visualized the relationship between the query and the documents. To identify the most crucial sentences, topic-expanded terms and activated nodes in the graph were used. Comparing the outcome of this experiment to 30 DUC systems was positive. Thakkar et al. [40] used TextRank in their PageRank-based system. They created a tightly connected graph for the text and applied the TextRank method to extract the relevant terms and assess their importance throughout the entire manuscript. Then they used the shortest path technique to get the sentences for the summary. They claimed that this provided the most diversified summary possible.

2) *Clustering-based approach*: Clustering-based or Sentence Centrality methods for QFMS are used in several

systems [2]. These methods use predefined features to assign scores for all sentence in the input text. After that, the sentences with similar contents will be grouped together in one cluster. In the end, the summary is generated by choosing representative sentences from each cluster. Various methods were developed to define the similarity measure between two objects in the text clustering [41]. The clustering-based approach is suitable for multi-document summarization since it groups different sentences by their topic. However, it requires prior specification of the clusters' amount, and top-ranked sentences may be similar. Hence, redundancy removal techniques are required [42].

Wang et al. [25] defined a clustering-based hypergraph where sentences are nodes and hyperedges are clusters. The sentences are scored using a semi-supervised ranking algorithm. To avoid redundancy, each extracted sentence is compared to previously selected ones before adding it to the generated summary. Chali and Joty [43] used k-means and expectation maximization techniques to determine the relevancy of sentences. Different features were used for weighting, such as lexical, lexical-semantic, statistical, and cosine similarity. Their work showed promising results that could be extended to consider more features, such as a fundamental element, tree kernel-based syntactic features, and shallow-semantic features [44]. Likewise, Naveen and Nedungadi [45] combined the Potential-based Hierarchical Agglomerative clustering algorithm and the k-means algorithm. Cosine similarity was employed to calculate the query-relevant score of each sentence, and the TextRank algorithm was used for ranking the sentences.

For the purpose of enhancing the similarity score between input text and user query, Chandu et al. [46] developed a hierarchical hybrid similarity measure with two tiers to check the similarity between input text and user query. The first tier uses cosine similarity with a threshold of 0.7. Then, for all the sentences passing this threshold, semantic and word order similarities are combined and applied to score the sentences. Redundant sentences are removed by using the DBSCAN and Agglomerative clustering algorithms. Similarly, Rahman and Borah [9] proposed a word sense disambiguation (WSD) method to improve the accuracy of the score for the sense-oriented sentence semantic between the input sentences and the query. The general method used to calculate this score between two sentences includes 1) calculating the Semantic Relatedness score. 2) calculating the Sense Relatedness score, 3) calculating the Word Order Similarity score. 4) finding the final Sense-Oriented Sentence Semantic relatedness score. Furthermore, they measured the informativeness of any sentence based on the presence of five features listed proper noun, numerical data, sentence length, thematic word, and cue phrase. Thus, any sentence that carries these features must be informative. The k-Mean clustering algorithm is employed to create clusters depending on the frequency of the five abovementioned features. Each cluster contains query-relevant sentences. To extract redundancy-free sentences, they established a cutoff point at which one of the sentences will be eliminated if the sense-oriented semantic relatedness score between the two sentences is higher than this cutoff point.

This algorithm achieved competitive results for all best participating models on DUC datasets as well as the current state-of-art QFMS systems.

Integrating various approaches to improve the final summary has received attention from many scholars. Bhaskar and Bandyopadhyay [11] used both graph-based and clustering-based approaches. The graph was reduced to include only seed nodes, which had a total score of all outgoing edges above a threshold. Such a reduction in a dense graph led to an effective execution time. The new graph was clustered to identify shared topical nodes. Each sentence was given a weight that represents the number of query words and keywords covered by that sentence. After using sentence compression, top-scored sentences in each cluster were selected for the summary. The approach gave commendable experimental results on a standard dataset. The performance of their method mainly relied on the selection of seed nodes, and since it is a query-based approach, their method could be enhanced if they consider the query during this reduction. Also, their ranking method is simplified; they just considered the exact match between each sentence and the query in terms of words and keywords. Likewise, clustering and a graph were merged by Canhasi and Kononenko for the summarization [47]. To combine the needed information from the query context and broaden the result options, an archetypal analysis was used. The sentences are grouped into various criteria depending on the type of analysis. The sentence that needs to be evaluated is plotted out on a graph, a score is given based on the relation to the query, which reflects its significance to the query. The weighted method, hence, weighted archetypal analysis, was designed to advance earlier archetypal analysis techniques.

For instance, a system that combines a topic model with graph-based semi-supervised learning was proposed by Li et al. [48]. The topic and sentence layers were the basis for the created graph. The relationship between the topic and sentence vectors was normalized after computing the cosine similarity between them. Sentence clustering was accomplished using a topic modeling technique. They took into account various data, including background and document-specific data. After evaluating this method, the summarization was greatly enhanced. Many scholars found that considering topic-level information might greatly enhance the output summary's quality [49]. He et al. [50] advocated a learning-based strategy that used content terms to rank sentences. They worked with both richness and relevant features, which resulted in a suitable choice of content terms. They used relevance features to give a relevance score to the query. And information richness feature was used to determine the significance of the phrase in the document collection. The scores from the aforementioned features were accumulated to determine the quality of the content term. On test data, their methodology produced promising results. Markedly, many methods have been developed for determining the significance of a sentence based on a query by considering other sentences' features.

These techniques integrated a variety of sentence features with the query's information to rate sentences. These features include term frequency of query, the log-likelihood ratio [6], the term overlap feature, sentence location, and the length of

sentence [26]. Wu et al. 's [1] query-focused summarization method was produced using an unsupervised two pattern-enhanced model. Using LDA topic modeling, the first pattern indicated the topic relevancy of the sentence while the relevance of the query to sentences was presented using the second. The sum of the two patterns for each sentence results was used to determine its importance score to the query. Moreover, to control redundancy during selection, they included a diversity penalty technique named maximal marginal relevance. They claimed that their results outperform state-of-the-art approaches.

The difficulty of scoring several sentences according to a query motivates the development of an interactive learning-to-rank technique to address it by Zhu et al. [53]. The model was initially defined as a sentence ranking issue. The ranking process then considers the connections between the previously chosen sentences and the current sentence in addition to the pertinent context of that specific sentence. The Plackett-Luce model was applied to minimize the likelihood of loss in the ranking function. The sentences in the summary are then chosen using the greedy selection technique based on the defined ranking function. Results from this approach are remarkably positive. An interesting method was proposed by

Woodsend and Lapata [54]. They used particular predictors to construct a model that learnt crucial summary elements independently from training data and then combined them optimally using integer linear programming. The system modeled less redundant content, content's critical and poor places, and stylistic norms, using bigram and positional information along with language modeling. The assessments of the expert learners were then combined using hard and soft constraints by ILP. A considerable improvement in text summarization was obtained using the approach. On the hand, the text-summarization method proposed by Yasuda et al. [55] adds the requirement that at least some terms from the query must appear in the summary. As a result of including that constraint, this optimization challenge was resolved via Lagrangian relaxation. By adding the constraints on the inclusion of query words, both ROUGE-1 and ROUGE-2 scores were increased and thereby increasing the relevance of summaries.

Table I provides an overview of the surveyed QFMS papers, highlighting their main characteristics, approach for selecting query-relevant sentences, redundancy removal techniques, and performance measures on datasets.

TABLE I. OVERVIEW OF A SET OF QFMS PAPERS

Ref/ Year	Approach	Selecting of query-relevant sentences	Redundancy removal technique	Performance
[9],2021	Clustering-based	Sense-oriented sentence semantic relatedness score to score the sentences. Extracted features: proper noun, numerical data, sentence length, thematic word, and cue phrase. Then, K-Means clustering algorithm is applied.	Cluster-based method: If the similarity between two sentences is >60%, then one of them will be removed.	+ The proposed algorithm obtained highest ROUGE score for all three datasets: DUC2005: Rouge1: 0.3951, Rouge2: 0.0893, RougeSU4: 0.1563. DUC2006: Rouge1: 0.5679, Rouge2: 0.1242, RougeSU4: 0.2181. DUC2007: Rouge1: 0.5735, Rouge2: 0.1367, RougeSU4: 0.2371.
[21], 2021	Semantic-graph-based	Sentence scoring= TF-ISF cosine similarity+ word overlap + proximity similarity Manifold Ranking algorithm	Cosine similarity threshold	+ The performance of using expansion of query is better than that using the original query DUC2006: Rouge1: 0.41674, Rouge2: 0.09202 , RougeSU4: 0.15071, Rouge-W: 0.14279. DUC2007: Rouge1: 0.43982, Rouge2: 0.11185, RougeSU4: 0.16870, Rouge-W: 0.15159.
[34], 2020	Semantic-graph-based	1) Jiang-Conrath measure is used to measure the semantic similarity between sentences. 2) Each sentence's cosine similarity to the query is determined. 3) the sentences are ranked using TextRank Algorithm	A similarity function is used to reduce similar sentences.	- Two evaluation metrics were used: ROUGE-N/L and human-based metrics. - They do not cover multi-documents summarization.
[23],2019	Hypergraph-based and topic model-based	Topics are modeled as hyperedges of a hypergraph. Hyperedge weights reflect their relevance to the query.	Hypergraph transversal is generated to capture the information jointly covered by a group of sentences.	+ Ability to produce non-redundant summaries that better cover the relevant topics of a corpus. DUC2005: Rouge2: 0.077392, RougeSU4: 0.12869. DUC2006: Rouge2: 0.10779, RougeSU4: 0.15854. DUC2007: Rouge2: 0.12997, RougeSU4: 0.17995 + It has low time complexity $O(N^2)$ , N is the total number of sentences. - Limitation: Restriction to purely topical similarities. - Authors would like to take the polysemy of terms into account.
[1], 2019	Topic model-based	- Query expansion technique is applied. - Sentence scoring = query-relevance score+ topical- relevance score + sentence position+ sentence length.	Maximal marginal relevance (MMR)[51] + Greedy algorithm [52]	DUC2006: Rouge1: 0.40551, Rouge2: 0.09228 , RougeSU4: 0.14966 . DUC2007: Rouge1: 0.43404, Rouge2: 0.11683, RougeSU4: 0.17026.

[46],2019	Clustering-based, semantic-based	A hierarchical hybrid similarity measure: 1. cosine similarity 2. a weighted sum of word order and semantic similarities. Then, DBSCAN and Agglomerative clustering algorithms are used.	Clustering methods	+ The generated summaries have shown 86% accuracy. - After the redundancy removal phase, their model removes some relevant sentences in their displayed examples. - Collected data from Amrita School of Engineering websites is used as a dataset.
[28],2019	Graph-Based and Topic model-Based	Similarity score: Normalized common words. 1. TextRank algorithm is applied to rank the most important sentences. 2. Two-Tiered topic used to sample query-relevant sentences.	NA	- The proposed method uses only common words for sentence scoring. - The scoring method should be enhanced to include both topical information and linguistic features.
[31], 2018	Semantic-graph-based	1) Sentence Scoring for each sentence = sentiment score + query relevant score . 2) graph-based ranking algorithm is applied to select top ranked sentences.	A Greedy algorithm in [52] is used.	- QMOS approach outperforms other existing methods. - DUC2006: ROUGE1: 0.4123 , ROUGE2: 0.0985, Average ROUGE Score:0.2554 - Used datasets: TAC 2008 and DUC 2006. -The authors would like to distinguish between active and passive sentences before comparing them.
[22],2017	Semantic-graph-based	1- Content word expansion (CWE) method to expand the words in the query and the sentence. 2. Combine: semantic similarity + Word order. 3. a graph-based ranking model is used to ranking the sentences.	A Greedy algorithm in [52] is used.	+ According to the experimental results, the proposed method performs well when compared to other methods. DUC2006: Rouge1: 0.4287, Rouge2: 0.0968, RougeSU4: 0.1673. - The authors intend to enhance their approach by considering recognizing passive and active sentences as well as increasing the semantic knowledge base.

**B. Text Summarization Datasets and Evaluation Metrics**

This is an overview of the basic resources used to evaluate and compare QFMS systems presented in the literature review. These resources include standard datasets besides evaluation tools.

1) *Datasets*: Several conferences and workshops have been organized for automatic summarization. To enable progress in this field, these conferences have made available datasets used in extensive research experiments. These datasets have undergone extensive work to prepare them to act as a standard text for summarizing while evaluating various methodologies.

- The Document Understanding Conferences (DUC): is a series of conferences for automatic summarization that are held by the National Institute of Standards and Technology (NIST) [56]. DUC-2005, DUC-2006, and DUC-2007 datasets are designed for extractive QFMS testing. Each data set contains several topics, including various related documents. Reference summaries are available for each topic for evaluation. Filling out some application forms found on the DUC website is needed to access these datasets. Table II shows a summary of these datasets.
- Text Analysis Conference (TAC) [57] : TAC is a group of evaluation workshops that aim to advance research in Natural Language Processing and related applications. It gives access to a massive test collection, standardized evaluation methods, and a platform to share their findings. The tasks from TAC-8 until TAC-15 support the query-based models.

2) *Evaluation metrics*: A system's produced summary can be accurately assessed for readability, succinctness, consistency, and compliance with information requirements using human assessments. Manual examination, however, is infeasible and takes much time. Consequently, it is necessary

to evaluate a summary automatically. Automatic evaluation of a system's generated summary can be done using ROUGE scoring [58]. This acronym means Recall-Oriented Understudy for Gisting Evaluation. It is a set of performance measures used to automatically calculate the quality of a summary. ROUGE compares the output summary to a set of summaries that were manually constructed by counting the number of intersecting units [59]. The intersecting units can be computed using n-grams, word pairs, or word sequences, which correspond to the ROUGE model.

TABLE II. SUMMARY OF USED DATASETS FOR QFMS

	DUC-2005	DUC-2006	DUC-2007
# of Topics	50	50	50
# of documents related to each topic	32-43	25	25
Data source	TIPSTER-TREC	AQUAINT	AQUAINT
Reference summaries for each topic	4 for each of 30 of the topics and 10 for each of 20 of the topic	4 summaries written by 4 different NIST assessors	4 summaries written by 4 different NIST assessors
Size of summary	250 words	250 words	250

**IV. DISCUSSIONS**

This section discusses the finding and current challenges that can lead the researcher in future directions.

**A. Findings**

This paper surveyed literature related to QFMS techniques. It is an active and attractive variant of automatic summarization due to its wide applications. Furthermore, they are less complicated, more affordable, and typically produce grammatically and semantically accurate summaries.

1) *QFMS approaches*: According to the different approaches for QFMS discussed previously, each approach

has distinct advantages and drawbacks. Most surveyed studies use the graph-based approach, which has shown effectiveness in QFMS due to its ability to enhance coherency and language-independent approach. However, it does not pay attention to the importance of the words in the document, as it assumes that the weights of the words are similar. As well as may be unable to identify semantically equivalent sentences. Consequently, the resulting summary can be less relevant and more redundant. However, some selected studies made different improvements to this approach by considering more language-dependent similarities like semantic similarity that enhanced the caliber of the summaries that were produced.

The clustering-based approach was successfully used to enhance the summaries' diversity and guarantee that all aspects of the needed information from the query were captured. It is appropriate for multi-document summarization because it groups several sentences about the same topic in the documents. Hence, each cluster contains highly related sentences. However, the highly scored sentences may be similar and thus have high redundancy. Therefore, there should be a mechanism for choosing sentences from each cluster that balances diversity, relevancy, and redundancy. Besides, it requires prior specification of the number of clusters. Another issue, some sentences may express more than one topic, but each sentence has to be assigned to only one cluster. The hypergraph-based approach is proposed in [23] to alleviate this issue. Each hyperedge is connected to a particular topic, and each sentence may be tagged with several different topics. Then, each sentence can be a member of various hyperedge.

According to the semantic-based approach, considerable performance has been provided from hybrid approaches that combine semantic-based and graph-based approaches. Some studies argued that QFMS couldn't be solved completely using only one method of the two methods, namely semantic-based and graph-based approaches.

Moreover, some of the overviewed articles have shown the effectiveness of combining topic model-based and graph-based approaches for QFMS to balance the three characteristics of summarization relevance, significance, and diversity. We found that most of the studies that were done concentrated on relevancy to the query by analysing the content of individual sentences depending on the query. Some research employed clustering to diversify the summary, although these algorithms only considered basic lexical similarity clustering [11], [47]. Other scholars are attentive to the diverse selection of the sentences while utilizing a straightforward Manifold method to assign relevancy score [60].

Generally, combining the previously mentioned approaches (graph-based, statistical-based, semantic-based, and clustering-based) would generate better summaries that benefit from their advantages and overcome the drawbacks of every single method.

2) *Extracting query-relevant sentences*: Many extraction techniques have been proposed in the surveyed systems. Most of them used statistical techniques such as TF-IDF and cosine similarity. However, these methods failed to capture semantic

similarities, thus decreasing the relevancy of the generated summary. At the same time, some studies boosted the statistical methods by introducing linguistic methods. The Word sense disambiguation (WSD) technique is proposed in [9] to determine each content word's appropriate meaning in a sentence. Their algorithm gained the highest ROUGE score for all three DUC query-based summarization datasets (DUC 2005, DUC 2006, and DUC 2007). However, the proposed WSD can only accurately determine a word sense if presented in WordNet.

Moreover, expanding query words has effectively solved mismatch problems in sentence comparison by extracting more relevant and essential sentences based on user demand. Hence, enhance the summaries' quality.

Furthermore, it is evident from the previous research that although they all seek QFMS, query-dependent features are given less attention by most of them. Without these query-dependent properties, their variation in speed is minimal.

3) *Redundancy removal techniques*: There are several redundancy removal techniques used in the surveyed literature. Most studies used a greedy algorithm in [52] as a post-processing step to force a diversity penalty on the sentences, which decreases the score of the less informative sentence before adding it to the final summary. Other approaches used the Maximum Marginal Relevance (MMR) [51] method to control redundancy. Moreover, the cosine similarity is calculated between each top-ranked sentence and the previously selected sentences to avoid redundancy. Sentence clustering algorithms are also used to prevent redundant information. In general, most of the redundancy-removal techniques under study rely on lexical similarity across sentences, which leaves semantic redundancies in the resulting summary unaffected. The Maximum Relevance and Coverage (MRC) [24] is suggested as a solution to this problem to maximize the relevancy and joint topical coverage. It showed enhancement compared to other redundancy removal techniques.

## B. Open Research Problems

Any QFMS system should be able to produce summaries of texts based on a query that are as near as possible to those produced by humans. Although many various strategies have been employed for the goal of QFMS, several concerns remain unresolved:

1) *Coherency*: The majority of summary techniques work by selecting the most important sentences to the query and presenting them verbatim. The reader must sense the flow of ideas rather than simply moving from one to the next. It can be beneficial to transfer one sentence to another due to their similarities. Therefore, sentence reordering is crucial, especially in multi-document summarization, since the sentences are from many sources with different flows of ideas. More post-processing techniques can be developed to tackle this issue, making it an active research problem.



2) *Ambiguity*: When determining the degree to which two content terms are semantically connected, sense plays a significant impact. Therefore, ambiguity in terms matters when summarizing statements. The quality of QFMS is indeed diminished by ambiguous words since they can reduce the number of sentences that can be found relevant to the query. Various kinds of ambiguity are known, such as word sense, local ambiguity, form class, structural, syntactical, and form factors [61]. In QFMS, removing ambiguity senses is a rising challenge that can interest many researchers.

3) *Vague reference*: In the multi-document summarization [62], a proper noun may appear in one sentence, and a pronoun may appear in the following sentence to refer to the proper noun. The summarizer will provide an ambiguous reference if it chooses the sentence with the pronoun but not the proper noun. This can open ideas for creating more pre-processing steps to resolve this issue and similar ones to enhance the overall generated summary.

4) *Evaluation*: Another critical challenge is the evaluation procedure. As aforementioned, existing evaluation in automatic summarization works by comparing the automatically generated summary to a human-generated one. This is admirable yet insufficient. Although reference summaries are created by expert humans, we cannot declare with certainty that this is the best summary due to the individual variation in writing and evaluation of the summary. Proposed techniques for QFMS are affected by the evaluation methods and the datasets available. It requires a lot of tools and corpora resources to create a powerful automatic text summarizer. DUCs, for example, produce a summary with 250 terms only. This is challenging for a system to generate a summary of just 250 words that is accurate and consistent with man-made summaries. More efficient ways to evaluate the summary would greatly help the researchers. Moreover, automated quality evaluations for grammar, reference clarity, readability, and coherence are still missing in this field [9].

5) *Redundancy*: Since the input is a multi-document that can share the same sentences and ideas, redundancy has been the main bottleneck when extracting query-relevant sentences. Although many techniques have been developed to reduce this issue, it is still an active and scalable domain.

6) *Diversity*: Designing a QFMS system not only requires extracting the essential sentences from the input but also demands diverse sentences to cover all aspects of the query. Only the sentences directly related to the query's primary request will be selected by a summarizer, leaving out any sub-request. As a result, the summary will concentrate more on the main point and neglect any supporting points that might be equally significant. Employing semantic analysis can help since it takes into account the meaning of every sentence and word. More techniques to handle this problem are open in this area.

### C. Future Research Directions

One of the most interesting issues is how a model can mimic a human's ability to summarize. The sentence information should be coherent, along with being concatenated. The summary's coherence has been a long-standing problem. Existing methods primarily seek to produce informative summaries; nevertheless, future research will be needed to improve the summary's readability by developing coherence scores between pairs of sentences and enhancing the order of sentences in the summary.

In addition, developing novel QFMS methods to generate query-related, higher-quality, and robust summaries under human criteria is a priority. More research needs to be done to improve and discover semantic, linguistic, and statistical features for terms in sentences. This will help systems to process natural language most effectively and to remove redundancy [7]. Additionally, choosing the appropriate query-relevant weights for various features is crucial because the final summary's quality depends on it. Studying new features and their effect on performance can be an eminent research domain.

A higher quality summary can be generated by making the system more intelligent by combining it with hybrid methods and other techniques. Important sentences can be chosen, combined, or compacted, or some information can be removed to provide better quality. A hybrid approach can be developed by combining extractive and abstractive techniques. Research can go on to generate a hybrid approach that combines extractive and abstractive methods to produce a more human-like generated summary.

Moreover, an effective balance between readability, compression ratio, and summarizing quality must be achieved. For QFMS of lengthy materials such as novels and books, larger compression ratios are needed. However, current systems struggle to meet this need [63]. Therefore, it is imperative to provide a more persuading and balanced method.

Automatically evaluating summaries is difficult since it is challenging to develop and apply a good criterion to determine whether the summaries produced by the systems are sufficient and satisfy the query [2]. Additionally, it is challenging to define the optimal summary since systems might produce effective summaries that differ from those produced by humans. Research can be conducted in automatic evaluation, creating new approaches and solutions to assess the query-related summary based on the data it includes, user satisfaction, how it is presented, and the level of readability and coherences.

An interesting future direction was suggested by [7]. The majority of QFMS systems work with text for both input and output. It will be beneficial if new summarizers can accept the input in a format of meetings, videos, audio, etc. Although this kind of input data is a valuable source for information extraction and knowledge discovery, users find it very challenging to track down or identify its occurrences due to their quantity and diversity. Moreover, the output can be in the form of statistics, graphs, tables, visual score measures, etc.

Users will obtain the necessary content faster with the aid of such summarizer systems that enable the summaries' visualization. There have been few works in the video summarization [64]; however, development in this important area is slow and requires more research efforts in the future.

The English language material is the main focus of most QFMS systems. There is a need to dedicate some future efforts to other languages. It is necessary to create and enhance NLP tools such as POS tagging, syntactic and semantic parsing, stemming, and NER that can be used for non-English languages [65]. Moreover, the absence of resources, such as annotated corpora and evaluation tools, is one of the most complicated issues that these types of summarizers must overcome.

Finally, developing a semi-supervised model for QFMS can be a potential future direction. This model can incorporate the user-required phrases to improve the semantic efficiency of the summary while incorporating a higher level of data's feature set. Thus, it may provide a query-based, more intelligent, and useful summary.

## V. CONCLUSIONS

In the last few years, there has been a huge expansion in the volume of text material on the internet. The research area of QFMS is intriguing and has many potential uses. It is a task of returning a concise and coherent response to a query entered by a user from multi-documents input.

This paper reviewed studies based on the main QFMS approaches: graph-based and clustering-based. It discusses their summarization process, advantages, and disadvantages. The findings show that hybrid approaches have been receiving increasing attention due to the satisfactory level of advanced performance. However, the currently generated summaries require further enhancements as they are still far from the quality of the human-generated summaries. Simultaneously, increasing research interest and rapid technological advancements could evolve QFMS and make summaries more relevant, significant, and less redundant.

The paper also underlined multiple open research problems and current challenges in QFMS. Furthermore, it presents future directions that may assist researchers in identifying crucial aspects that require deep investigation and more development.

## REFERENCES

- [1] Y. Wu, Y. Li, and Y. Xu, "Dual pattern-enhanced representations model for query-focused multi-document summarisation," *Knowl.-Based Syst.*, vol. 163, pp. 736–748, Jan. 2019, doi: 10.1016/j.knosys.2018.09.035.
- [2] W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. Mohamed, "Automatic text summarization: A comprehensive survey," *Expert Syst. Appl.*, vol. 165, p. 113679, Mar. 2021, doi: 10.1016/j.eswa.2020.113679.
- [3] S. Gupta, A. Nenkova, and D. Jurafsky, "Measuring Importance and Query Relevance in Topic-focused Multi-document Summarization," in *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, Prague, Czech Republic: Association for Computational Linguistics, Jun. 2007, pp. 193–196. Accessed: Oct. 28, 2022. [Online]. Available: <https://aclanthology.org/P07-2049>.
- [4] P. Du, J. Guo, and X. Cheng, "Supervised Lazy Random Walk for Topic-Focused Multi-document Summarization," in *2011 IEEE 11th International Conference on Data Mining*, Dec. 2011, pp. 1026–1031. doi: 10.1109/ICDM.2011.140.
- [5] J. Balaji, T. V. Geetha, and R. Parthasarathi, "A Graph Based Query Focused Multi-Documen Summarization," *Int. J. Intell. Inf. Technol. IJIT*, vol. 10, no. 1, pp. 16–41, Jan. 2014, doi: 10.4018/ijit.2014010102.
- [6] R. Ferreira et al., "Assessing sentence scoring techniques for extractive text summarization," *Expert Syst. Appl.*, vol. 40, no. 14, pp. 5755–5764, Oct. 2013, doi: 10.1016/j.eswa.2013.04.023.
- [7] M. Gambhir and V. Gupta, "Recent automatic text summarization techniques: a survey," *Artif. Intell. Rev.*, vol. 47, no. 1, pp. 1–66, Jan. 2017, doi: 10.1007/s10462-016-9475-9.
- [8] W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. Mohamed, "EdgeSumm: Graph-based framework for automatic text summarization," *Inf. Process. Manag.*, vol. 57, no. 6, p. 102264, Nov. 2020, doi: 10.1016/j.ipm.2020.102264.
- [9] N. Rahman and B. Borah, "Query-Based Extractive Text Summarization Using Sense-Oriented Semantic Relatedness Measure," In Review, preprint, Nov. 2021. doi: 10.21203/rs.3.rs-1102477/v1.
- [10] M. Zhao et al., "QBSUM: a Large-Scale Query-Based Document Summarization Dataset from Real-world Applications," *Comput. Speech Lang.*, vol. 66, p. 101166, Mar. 2021, doi: 10.1016/j.csl.2020.101166.
- [11] P. Bhaskar and S. Bandyopadhyay, "A Query Focused Multi Document Automatic Summarization," in *Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation*, Tohoku University, Sendai, Japan: Institute of Digital Enhancement of Cognitive Processing, Waseda University, Nov. 2010, pp. 545–554. Accessed: Oct. 31, 2022. [Online]. Available: <https://aclanthology.org/Y10-1063>.
- [12] Z. Nasar, S. W. Jaffry, and M. K. Malik, "Textual keyword extraction and summarization: State-of-the-art," *Inf. Process. Manag. Int. J.*, vol. 56, no. 6, Nov. 2019, doi: 10.1016/j.ipm.2019.102088.
- [13] D. F. Gleich, "PageRank beyond the Web," *arXiv*, Jul. 18, 2014. doi: 10.48550/arXiv.1407.5107.
- [14] R. Mihalcea and P. Tarau, "TextRank: Bringing Order into Text," in *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, Barcelona, Spain: Association for Computational Linguistics, Jul. 2004, pp. 404–411. Accessed: Oct. 29, 2022. [Online]. Available: <https://aclanthology.org/W04-3252>.
- [15] G. Erkan and D. R. Radev, "LexRank: Graph-based Lexical Centrality as Saliency in Text Summarization," *J. Artif. Intell. Res.*, 2004, doi: 10.1613/jair.1523.
- [16] K. Lei and Y. F. Zeng, "A Novel Biased Diversity Ranking Model for Query-Oriented Multi-Documen Summarization," *Appl. Mech. Mater.*, vol. 380–384, pp. 2811–2816, Aug. 2013, doi: 10.4028/www.scientific.net/AMM.380-384.2811.
- [17] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Schölkopf, "Ranking on data manifolds," in *Proceedings of the 16th International Conference on Neural Information Processing Systems*, in *NIPS'03*. Cambridge, MA, USA: MIT Press, Dec. 2003, pp. 169–176.
- [18] X. Wan and J. Xiao, "Graph-based multi-modality learning for topic-focused multi-document summarization," in *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, in *IJCAI'09*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., Jul. 2009, pp. 1586–1591.
- [19] F. Wei, Y. He, W. Li, and Q. Lu, "A Query-Sensitive Graph-Based Sentence Ranking Algorithm for Query-Oriented Multi-document Summarization," in *2008 International Symposiums on Information Processing*, May 2008, pp. 9–13. doi: 10.1109/ISIP.2008.21.
- [20] A. A. Mohamed and S. Rajasekaran, "Improving Query-Based Summarization Using Document Graphs," in *2006 IEEE International Symposium on Signal Processing and Information Technology*, Aug. 2006, pp. 408–410. doi: 10.1109/ISSPIT.2006.270835.
- [21] Q. Jia, R. Liu, and J. Lin, "Using Query Expansion in Manifold Ranking for Query-Oriented Multi-Documen Summarization," 2021, pp. 97–111. doi: 10.1007/978-3-030-84186-7\_7.
- [22] A. Abdi, N. Idris, R. M. Alguliyev, and R. M. Aliguliyev, "Query-based multi-documents summarization using linguistic knowledge and content

- word expansion,” *Soft Comput.*, vol. 21, no. 7, pp. 1785–1801, Apr. 2017, doi: 10.1007/s00500-015-1881-4.
- [23] H. Van Lierde and T. W. S. Chow, “Query-oriented text summarization based on hypergraph transversals,” *Inf. Process. Manag.*, vol. 56, no. 4, pp. 1317–1338, Jul. 2019, doi: 10.1016/j.ipm.2019.03.003.
- [24] H. Van Lierde and T. W. S. Chow, “Learning with fuzzy hypergraphs: A topical approach to query-oriented text summarization,” *Inf. Sci. Int. J.*, vol. 496, no. C, pp. 212–224, Sep. 2019, doi: 10.1016/j.ins.2019.05.020.
- [25] W. Wang, S. Li, J. Li, W. Li, and F. Wei, “Exploring hypergraph-based semi-supervised ranking for query-oriented summarization,” *Inf. Sci.*, vol. 237, pp. 271–286, Jul. 2013, doi: 10.1016/j.ins.2013.03.012.
- [26] S. Xiong and D. Ji, “Query-focused multi-document summarization using hypergraph-based ranking,” *Inf. Process. Manag.*, vol. 52, no. 4, pp. 670–681, Jul. 2016, doi: 10.1016/j.ipm.2015.12.012.
- [27] A. A. Bichi, P. Keikhosrokiani, R. Hassan, and K. Almekhlafi, “Graph-Based Extractive Text Summarization Models: A Systematic Review,” *J. Inf. Technol. Manag.*, vol. 14, no. Special Issue: 5th International Conference of Reliable Information and Communication Technology (IRICT 2020), pp. 184–202, Feb. 2022, doi: 10.22059/jitm.2022.84899.
- [28] N. Akhtar, M. M. S. Beg, and H. Javed, “TextRank enhanced Topic Model for Query focussed Text Summarization,” in 2019 Twelfth International Conference on Contemporary Computing (IC3), Aug. 2019, pp. 1–6. doi: 10.1109/IC3.2019.8844939.
- [29] A. Khan et al., “Abstractive Text Summarization based on Improved Semantic Graph Approach,” *Int. J. Parallel Program.*, vol. 46, no. 5, pp. 992–1016, Oct. 2018, doi: 10.1007/s10766-018-0560-3.
- [30] A. Aries, D. eddine Zegour, and W. K. Hidouci, “Automatic text summarization: What has been done and what has to be done.” arXiv, Apr. 01, 2019. doi: 10.48550/arXiv.1904.00688.
- [31] A. Abdi, S. M. Shamsuddin, and R. M. Aliguliyev, “QMOS: Query-based multi-documents opinion-oriented summarization,” *Inf. Process. Manag.*, vol. 54, no. 2, pp. 318–338, Mar. 2018, doi: 10.1016/j.ipm.2017.12.002.
- [32] R. V. V. M. Krishna, S. Y. P. Kumar, and C. S. Reddy, “A Hybrid Method for Query based Automatic Summarization System,” *Int. J. Comput. Appl.*, vol. 68, no. 6, pp. 39–43, Apr. 2013.
- [33] T. He, F. Li, W. Shao, J. Chen, and L. Ma, “A New Feature-Fusion Sentence Selecting Strategy for Query-Focused Multi-document Summarization,” in 2008 International Conference on Advanced Language Processing and Web Information Technology, Jul. 2008, pp. 81–86. doi: 10.1109/ALPIT.2008.45.
- [34] S. Murarka and A. Singhal, “Query-based Single Document Summarization using Hybrid Semantic and Graph-based Approach,” in 2020 International Conference on Advances in Computing, Communication & Materials (ICACCM), Aug. 2020, pp. 330–335. doi: 10.1109/ICACCM50413.2020.9212923.
- [35] J.-Y. Yeh, H.-R. Ke, W.-P. Yang, and I.-H. Meng, “Text summarization using a trainable summarizer and latent semantic analysis,” *Inf. Process. Manag.*, vol. 41, no. 1, pp. 75–95, Jan. 2005, doi: 10.1016/j.ipm.2004.04.003.
- [36] V. Phung and L. De Vine, “A Study on the Use of Word Embeddings and PageRank for Vietnamese Text Summarization,” in Proceedings of the 20th Australasian Document Computing Symposium, in ADCS ’15. New York, NY, USA: Association for Computing Machinery, Dec. 2015, pp. 1–8. doi: 10.1145/2838931.2838935.
- [37] L. Page, S. Brin, R. Motwani, and T. Winograd, “The PageRank Citation Ranking: Bringing Order to the Web,” Nov. 11, 1999. <http://ilpubs.stanford.edu:8090/422/> (accessed Oct. 30, 2022).
- [38] A. Nenkova and K. McKeown, “Automatic Summarization,” *Found. Trends® Inf. Retr.*, vol. 5, no. 2–3, pp. 103–233, Jun. 2011, doi: 10.1561/1500000015.
- [39] V. Nastase, “Topic-driven multi-document summarization with encyclopedic knowledge and spreading activation,” in Proceedings of the Conference on Empirical Methods in Natural Language Processing, in EMNLP ’08. USA: Association for Computational Linguistics, Oct. 2008, pp. 763–772.
- [40] K. S. Thakkar, R. V. Dharaskar, and M. B. Chandak, “Graph-Based Algorithms for Text Summarization,” in 2010 3rd International Conference on Emerging Trends in Engineering and Technology, Nov. 2010, pp. 516–519. doi: 10.1109/ICETET.2010.104.
- [41] R. Aliguliyev, R. Aliguliyev, N. Isazade, A. Abdi, and N. Idris, “A Model for Text Summarization,” *Int. J. Intell. Inf. Technol.*, vol. 13, pp. 67–85, Jan. 2017, doi: 10.4018/IJIT.2017010104.
- [42] M. F. Mridha, A. A. Lima, K. Nur, S. C. Das, M. Hasan, and M. M. Kabir, “A Survey of Automatic Text Summarization: Progress, Process and Challenges,” *IEEE Access*, vol. 9, pp. 156043–156070, 2021, doi: 10.1109/ACCESS.2021.3129786.
- [43] Y. Chali and S. R. Joty, “Unsupervised Approach for Selecting Sentences in Query-based Summarization.” <https://www.aaai.org/Library/FLAIRS/2008/flairs08-019.php> (accessed Oct. 27, 2022).
- [44] Y. Chali and S. R. Joty, “Answering Complex Questions Using Query-Focused Summarization Technique,” in 2008 20th IEEE International Conference on Tools with Artificial Intelligence, Nov. 2008, pp. 131–134. doi: 10.1109/ICTAL.2008.84.
- [45] G. K. R. Naveen and P. Nedungadi, “Query-based Multi-Document Summarization by Clustering of Documents,” in Proceedings of the 2014 International Conference on Interdisciplinary Advances in Applied Computing, in ICONIAAC ’14. New York, NY, USA: Association for Computing Machinery, Oct. 2014, pp. 1–8. doi: 10.1145/2660859.2660972.
- [46] G. V. Madhuri Chandu, A. Premkumar, S. S. K. and N. Sampath, “Extractive Approach For Query Based Text Summarization,” in 2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), Sep. 2019, pp. 1–5. doi: 10.1109/ICICT46931.2019.8977708.
- [47] CanhasiErcan and Kononenkolgor, “Weighted archetypal analysis of the multi-element graph for query-focused multi-document summarization,” *Expert Syst. Appl. Int. J.*, Feb. 2014, doi: 10.1016/j.eswa.2013.07.079.
- [48] J. Li and S. Li, “Query-focused Multi-document Summarization: Combining a Novel Topic Model with Graph-based Semi-supervised Learning,” Dec. 2012, doi: 10.48550/arXiv.1212.2036.
- [49] N. Liu, X.-J. Tang, Y. Lu, M.-X. Li, H.-W. Wang, and P. Xiao, “Topic-Sensitive Multi-document Summarization Algorithm,” in 2014 Sixth International Symposium on Parallel Architectures, Algorithms and Programming, Jul. 2014, pp. 69–74. doi: 10.1109/PAAP.2014.22.
- [50] T. He, W. Shao, F. Li, Z. Yang, and L. Ma, “The Automated Estimation of Content-Terms for Query-Focused Multi-document Summarization,” in 2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery, Oct. 2008, pp. 580–584. doi: 10.1109/FSKD.2008.260.
- [51] J. Carbonell and J. Goldstein, “The use of MMR, diversity-based reranking for reordering documents and producing summaries,” in Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval, in SIGIR ’98. New York, NY, USA: Association for Computing Machinery, Aug. 1998, pp. 335–336. doi: 10.1145/290941.291025.
- [52] X. Wan, J. Yang, and J. Xiao, “Manifold-ranking based topic-focused multi-document summarization,” in Proceedings of the 20th international joint conference on Artificial intelligence, in IJCAI’07. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., Jan. 2007, pp. 2903–2908.
- [53] Y. Zhu, Y. Lan, J. Guo, P. Du, and X. Cheng, “A Novel Relational Learning-to-Rank Approach for Topic-Focused Multi-document Summarization,” in 2013 IEEE 13th International Conference on Data Mining, Dec. 2013, pp. 927–936. doi: 10.1109/ICDM.2013.38.
- [54] K. Woodsend and M. Lapata, “Multiple Aspect Summarization Using Integer Linear Programming,” in Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Jeju Island, Korea: Association for Computational Linguistics, Jul. 2012, pp. 233–243. Accessed: Oct. 28, 2022. [Online]. Available: <https://aclanthology.org/D12-1022>.
- [55] N. Yasuda, M. Nishino, T. Hirao, J. Suzuki, and R. Kataoka, “A Query-Focused Summarization Method that Guarantees the Inclusion of Query Words,” in Proceedings of the 2012 23rd International Workshop on Database and Expert Systems Applications, in DEXA ’12. USA: IEEE

- Computer Society, Sep. 2012, pp. 126–130. doi: 10.1109/DEXA.2012.59.
- [56] “Document Understanding Conferences.” <https://duc.nist.gov/> (accessed Oct. 28, 2022).
- [57] Y. Fors-Isalguez, J. Hermosillo-Valadez, and M. Montes-y-Gómez, “Query-oriented text summarization based on multiobjective evolutionary algorithms and word embeddings,” *J. Intell. Fuzzy Syst.*, vol. 34, no. 5, pp. 3235–3244, Jan. 2018, doi: 10.3233/JIFS-169506.
- [58] C.-Y. Lin, “ROUGE: A Package for Automatic Evaluation of Summaries,” in *Text Summarization Branches Out*, Barcelona, Spain: Association for Computational Linguistics, Jul. 2004, pp. 74–81. Accessed: Oct. 28, 2022. [Online]. Available: <https://aclanthology.org/W04-1013>.
- [59] J. Steinberger and K. Ježek, “Evaluation Measures for Text Summarization,” *Comput. Inform.*, vol. 28, no. 2, Art. no. 2, 2009.
- [60] X.-Q. Cheng, P. Du, J. Guo, X. Zhu, and Y. Chen, “Ranking on Data Manifold with Sink Points,” *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 1, pp. 177–191, Jan. 2013, doi: 10.1109/TKDE.2011.190.
- [61] S. Jusoh, “A study on nlp applications and ambiguity problems,” *J. Theor. Appl. Inf. Technol.*, vol. 96, pp. 1486–1499, Mar. 2018.
- [62] D. Sahoo, R. Balabantaray, M. Phukon, and S. Saikia, “Aspect based multi-document summarization,” in *2016 International Conference on Computing, Communication and Automation (ICCCA)*, Apr. 2016, pp. 873–877. doi: 10.1109/CCAA.2016.7813838.
- [63] Z. Wu et al., “A topic modeling based approach to novel document automatic summarization,” *Expert Syst. Appl.*, vol. 84, pp. 12–23, Oct. 2017, doi: 10.1016/j.eswa.2017.04.054.
- [64] A. Sharghi, J. S. Laurel, and B. Gong, “Query-Focused Video Summarization: Dataset, Evaluation, and a Memory Network Based Approach,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 2127–2136. doi: 10.1109/CVPR.2017.229.
- [65] R. Belkebir and A. Guessoum, “TALAA-ATSF: A Global Operation-Based Arabic Text Summarization Framework,” in *Intelligent Natural Language Processing: Trends and Applications*, K. Shaalan, A. E. Hassanien, and F. Tolba, Eds., in *Studies in Computational Intelligence*. Cham: Springer International Publishing, 2018, pp. 435–459. doi: 10.1007/978-3-319-67056-0\_21.