I.J. Modern Education and Computer Science, **2024, 4, 58-70**
Published Online on August 8, 2024 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijmecs.2024.04.05

Document Summarization Based on Information Retrieval Using Query Search Ranking Method

Surya S.*
Department of Computer Science and Computer Application, Vivekanandha College of Arts and Sciences for Women (Autonomous), India
E-mail: suryaamca88@gmail.com
ORCID iD: https://orcid.org/0009-0008-5898-343x
*Corresponding author

Sumitra P.
Department of Computer Science and Computer Application, Vivekanandha College of Arts and Sciences for Women (Autonomous), India
E-mail: sumitraavaradharajan@gmail.com
ORCID iD: https://orcid.org/0009-0008-9991-0165

Received: 29 January, 2024; Revised: 20 February, 2024; Accepted: 08 March, 2024; Published: 08 August, 2024

Abstract: This work proposes a Query Searching Based Ranking Summarization Data Retrieval (QS-RSDR) method for document summarization based on information retrieval. QS-RSDR ranks query-based retrieval of important information, enabling the creation of a detailed report of information requirements using generated sections of sample documents. Relating Keyword Query Search Summarization (RKQSS) generates the main summary from the most relevant document in the query and then the summary from the other documents. The method resolves similarity terms related to the query using the Word Frequency (WF) method. Sentence ranking weights and sentence frequency improve the accuracy of the retrieved documents. Simulation results show improved accuracy in information retrieval. The proposed method can help address unclear and short queries and understand the nature of the required information behind the query. The paper concludes that QS-RSDR is an effective solution for document summarization based on information retrieval using the query search ranking method.

Index Terms: Information Retrieval (IR), Query, Relating Keyword Query Search Summarization (RKQSS), Word Frequency (WF), Query Searching Based Ranking Summarization Data Retrieval (QS-RSDR)

1. Introduction

It can be difficult for users to analyze and efficiently extract the most pertinent information from the large amount of information available on the internet [1]. The goal of query-based document summarizing is to create a succinct and logical summary of a given document that is pertinent to the user's search query [2]. Therefore, it is essential that search engines enhance user experience and search efficiency [3].

This area has been examined from a number of angles, including automatic keyword extraction, text databases, summarizing methods, summarization methodologies, and evaluation matrices [4]. Sentence ranking continues to be a key difficulty in document summary [5] as a result of individuals having to cope with enormous amounts of information and needing strategies to save time and effort by summarizing the most crucial and pertinent information [6,7]. Furthermore, the queries sent to search engines can be ambiguous and short, and their meanings may change over time, making it crucial to understand the nature of the required information behind the query [8].

This study aims to propose a QS-RSDR approach to rank query-based retrieval of important information and enable the creation of a detailed report of the information requirements using the generated sections of the sample documents [9, 10]. Furthermore, this proposed model aims to improve the performance of query-based document sentence extraction models [11].

Query-based document summarization can also serve as an essential upstream task for machine reading comprehension, which generates an answer to a question based on a passage [12]. The significance of this study lies in its event-oriented information extraction, which produces Query Retrieval-based Ranking Summary Data Retrieval [13].

To evaluate the performance of the proposed model, objective evaluations such as recall, precision, F-measure, the false rate for information retrieval, query retrieval, and reduced time complexity for searching are used, as are subjective evaluations such as informativeness and coherence [14, 15, 16]. As a result, the proposed model demonstrates improved performance compared to traditional methods and encourages future work in this field [17].

In conclusion, this study proposes a QS-RSDR approach to improve the performance [18] of query-based document summarization and reduce the computational overhead while preserving all the essential features of real-world events [19]. This method mainly studies three problems: (1) using common sense knowledge to elaborate queries, (2) finding the appropriate meaning of words associated with sentences, and (3) computing the relationship between words in documents. Finally, redundancy is checked by looking for common words between the documents and the proposed method can help address unclear and short queries and understand the nature of the required information behind the query.

The structure of this paper is as follows: In Section 2, we provide an overview of related work to contextualize our approach. Section 3 elaborates on the materials and methodology employed in our study. The experimental results are presented in Section 4. Finally, in Section 5, we provide concluding remarks and discuss avenues for future research.

2. Literature Review

The literature review provides an overview of key research papers in information retrieval, covering topics like spatial-keyword range queries, query expansion techniques, ranking models, text retrieval privacy protection, multi-keyword ranked search, and AI methodologies for IR ranking.

Tampakis et al. (2021) proposed an indexing scheme for efficient support of spatial-keyword range queries [20]. Their approach involves mapping spatio-textual data to a 2D space, generating compact partitions, and using bounds to design a filter-and-refine algorithm. The experimental evaluation demonstrated the effectiveness of their algorithm implemented in PostgreSQL with PostGIS.

Nasir et al. (2019) focused on query expansion techniques for improving information retrieval performance [21]. They surveyed existing techniques, highlighted strengths and limitations, and introduced a method that combines knowledge-based or corpus-based techniques with relevant feedback. Experimental evaluation of benchmark datasets showed improved information retrieval performance.

Guo et al. (2020) discussed the importance of ranking models in information retrieval [22]. They explored techniques from traditional heuristic methods to modern machine-learning approaches. They emphasized the power of neural ranking models, which can learn from raw text inputs and handle the complexity of relevance estimation. They also identified areas for future research in the field.

A dummy-based method for text retrieval privacy protection was put forth by Wu et al. (2020) [23]. Their approach uses well-designed dummy queries to protect user privacy without compromising retrieval accuracy. They presented a client-based system framework and demonstrated the effectiveness of their technique through theoretical analysis and experimental evaluation.

Handa et al. (2019) proposed a technique for clustering documents based on keyword relationships and efficient ranking [24]. Their method involves searching documents within relevant clusters, reducing communication overhead. In addition, they incorporated a ranking method using Term Frequency-Inverse Document Frequency values and proposed an efficient query randomization approach. Experimental results demonstrated significant reductions in search time while maintaining high recall and precision.

Ganesh and Chithambarathanu (2022) investigated AI methodologies for determining IR ranking positions [25]. They focused on SVM and PSO as significant IR positioning methods. They developed a novel PSO-SVM model, combining PSO and SVM, to improve ranking accuracy with a limited feature subset. The hybrid PSO-SVM framework was designed to decrease computational time using a distributed architecture.

The literature review encompasses a range of crucial topics in information retrieval, such as efficient support for spatial-keyword range queries, query expansion techniques, ranking models, text retrieval privacy protection, multi-keyword ranked search, and AI methodologies for IR ranking. The reviewed papers present innovative approaches, conduct experimental evaluations, and offer insights into these areas. By highlighting the importance of the reviewed research, the literature review contributes to the advancement of information retrieval. The findings shed light on retrieval techniques, privacy protection, ranking models, and query expansion, providing valuable knowledge to the research community. Moreover, identifying research gaps and future directions paves the way for further exploration and investigation in these domains.

3. Materials and Method

A QS-RSDR method generates document summaries for information retrieval. To retrieve information, users need a topic to query. The IR system provides a list of retrieved documents containing queries. The user should select the file containing the relevant information. The approach is that users can offer this task by providing document summaries. Some key aspects of relevant information users may rank documents in their queries search, and following searches will succeed.

3.1. Problem Statement

Problem: The problem at hand is document summarization, which arises due to the overwhelming amount of information available to users. It becomes challenging and time-consuming for users to extract the most relevant and important content from lengthy documents.

3.1.1 Inefficiencies and information overload

Users often face difficulties in sifting through extensive documents to identify key information. This leads to inefficiencies and information overload, making it difficult to find the relevant content quickly.

3.1.2 Limitations of traditional methods

Traditional document summarization methods, such as predefined templates or extraction rules, may not adequately address user needs. They do not consider the specific information requirements of users, resulting in summaries that may not be tailored to their queries.

3.1.3 Need for an effective approach

There is a need for an effective approach that combines information retrieval techniques and query search ranking to generate concise and informative summaries. This approach should consider user queries and retrieve relevant documents while ranking the extracted sentences or sections based on their importance and relevance.

3.1.4 Addressing limitations

The proposed approach aims to overcome the limitations of existing document summarization methods by leveraging information retrieval and query search ranking. It seeks to develop a system that can efficiently retrieve relevant documents based on user queries and rank the extracted content to provide concise and tailored summaries.

3.1.5 Improved efficiency and effectiveness

By using query search ranking, the proposed approach can prioritize and present the most important information in a concise and relevant manner. This saves time and effort for users by enabling them to quickly grasp the main points of a document and determine its relevance without having to go through the entire text.

3.1.6 Meeting user information needs

The ultimate goal of the proposed approach is to provide users with summaries that meet their information needs effectively. By leveraging information retrieval techniques and query search ranking, the system can generate summaries that are highly relevant, tailored to user queries, and offer a comprehensive overview of the document.

3.2. Research objective

3.2.1 Develop an efficient document summarization system

The primary objective is to design and develop a document summarization system that can effectively extract and present the most important information from documents. The system should be capable of generating concise summaries that capture the key points of the original text.

3.2.2 Improve retrieval accuracy

Enhance the accuracy of document retrieval by incorporating query search ranking techniques. The objective is to retrieve documents that are highly relevant to the user's query, ensuring that the extracted content aligns with the user's information needs.

3.2.3 Tailor summaries to user queries

Develop a methodology that considers user queries as input and generates summaries specifically tailored to address those queries. The objective is to align the summary content with the user's information requirements, increasing the overall relevance and usefulness of the generated summaries.

3.2.4 Enhance ranking methodology

Improve the ranking methodology used in the document summarization process. The objective is to develop innovative ranking techniques that effectively prioritize and order the extracted sentences or sections based on their importance and relevance to the user's query.

3.2.5 Evaluate system performance

Conduct a comprehensive evaluation of the proposed methodology to assess its performance and effectiveness. This involves measuring the efficiency of the summarization system, the accuracy of document retrieval, and the quality of the generated summaries in meeting user information needs.

3.2.6 Compare against existing methods

Compare the proposed methodology against existing document summarization approaches, including traditional methods and other information retrieval-based techniques. The objective is to demonstrate the superiority of the proposed methodology in terms of efficiency, retrieval accuracy, and summary quality.

3.3. Data Collection

Describe the process of collecting relevant documents for the research, including the selection criteria, sources, and any necessary pre-processing steps.

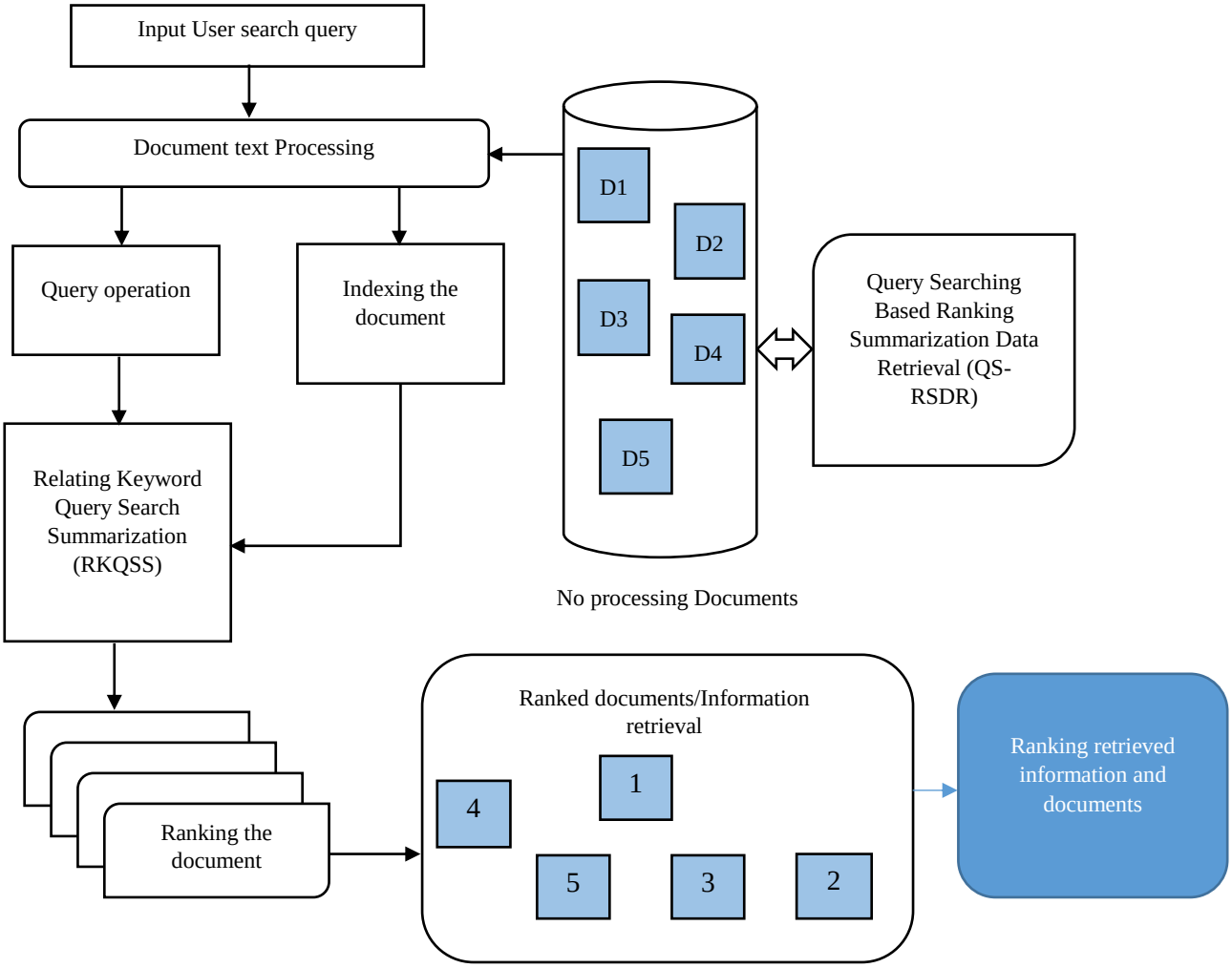

Fig. 1. Proposed block diagram of Query Searching Based Ranking Summarization Data Retrieval (QS-RSDR)

Information retrieval is the process of retrieving relevant documents or information from a collection based on user queries. The following steps and techniques are commonly used in the information retrieval process:

3.3.1 User Query

The process begins with the user inputting a query, which is a set of keywords or a natural language expression representing their information need.

3.3.2 Query Preprocessing

The user query undergoes preprocessing steps to enhance its effectiveness in retrieving relevant documents. This may involve removing stop words (common words with little semantic value), stemming (reducing words to their base or root form), or applying other linguistic transformations.

3.3.3 Query Expansion

Query expansion techniques may be employed to broaden the query's scope and increase the chances of retrieving relevant documents. This can involve adding synonyms, related terms, or concepts to the original query, thereby improving recall and addressing potential vocabulary mismatches between the query and the documents.

3.3.4 Keyword Matching

The pre-processed or expanded query is then matched against the indexed documents. Keyword matching involves comparing the query terms with the indexed terms in the document collection. Various matching algorithms can be employed, such as exact matching, fuzzy matching, or probabilistic models.

3.3.5 Ranking and Scoring

Once the documents are retrieved, they are ranked and scored based on their relevance to the query. Ranking algorithms assign a numerical score to each document, considering factors like term frequency, inverse document frequency, document length normalization, and other relevance metrics. To put the most pertinent documents at the front of the sorted list is the objective.

3.3.6 Search Results Presentation

The final step is to present the search results to the user. Typically, the user receives a sorted list of documents that best match their query. The presentation can include snippets or summaries of the documents to provide a preview of their content.

Fig. 1 illustrates a block diagram of the proposed Query Searching-based Information Retrieval system framework, highlighting the key stages of document retrieval. This framework involves selectively collecting relevant documents, preprocessing them, performing indexing, executing the search process, and ultimately presenting the user with a sorted list of matching documents.

3.4. Exploring User Query Patterns and Improving Search Relevance

Search interfaces play a crucial role in understanding user queries, even in the presence of incorrect grammatical syntax. Gaining insights into user query patterns is essential for search engines to comprehend the intended meaning behind queries and provide relevant search results. By leveraging key searching-based query elements, search systems can enhance their understanding of user queries, ultimately leading to more accurate and contextually appropriate search results.

Pseudo code

Algorithm: Explore User Query Patterns and Improve Search Relevance

```
Input: Query count
Output: Query-based matching documents
Step 1:    Begin the process
Step 2:    Input query
Step 3:    Get terms (S1, S2...Sn) using the document list
            Count term weights (w1, w2...wn) based
            on word frequency
             For each term Q_seq in S:
                Stage 1: Number of indexed documents
                 Stage 2: Find the shortest terms using
                          query length
                 Stage 3: Identify each query term
                 Stage 4: Match documents with terms and
                          add them (D) to Term with
                          Query document

              If (TermWithQuerydocument is not empty and
                  TermWithoutQuerydocument is not empty):
                 For each term weight w in
                       TermWithQuerydocument:
                  If (MatchingDocuments1 is empty):
                     QueryExpansion(S1, S2...Sn)
                      Get retrieval for similar documents
                       Return MatchingDocuments1
                  For each term weight w in Term Without
                      Query document:
                     If (MatchingDocuments2 is empty):
                        Calculate similarity to similar documents
                        based on term weights
                        Return MatchingDocuments2

Step 4:    Stop
```

The system performs semantic analysis of user query requests during the search process. The query starts the search with the stop words and checks if the query keyword belongs to the document database.

3.5. Indexing the document

Indexing the document is a pre-processed stage. At this stage, a file is created for each document that contains words without stop words (at, the, is, an, etc.). Also, stemming is done to get concepts in their original form (e.g., used, used, available stemming). Each word's frequency is calculated in the next step, and a threshold (based on the formula) is used as code words. All the times in the collection form, create a table for that document. In IR systems, indexing is a technique that makes information retrieval more accurate, faster, and more relevant. Indexes can be generated from keywords in stored documents. The first step is to elaborate the query based on the stored database.

All indexing term is a measurement. This indexing term is usually a word of each document.

All documents is a vector of term or weight, and similarity approaches are evaluated as cosine measures by Equation (1) – (6),

$$Dx = \left(t_{x1}, t_{x2}, t_{x3}, t_{x4} \ldots\ldots t_{xn}\right) \tag{1}$$

$$Dy = \left(t_{y1}, t_{y2}, t_{y3}, t_{y4} \ldots\ldots t_{yn}\right) \tag{2}$$

Document similarity is described as cosine similarity,

$$S(Dx, Dy) = \frac{\sum_{i=1}^{n} t_{xi} * t_{yi}}{\sqrt{\sum_{i=1}^{n} t_{xi}^{\ 2}\, X} \sqrt{\sum_{i=1}^{n} t_{yj}^{\ 2}}} \tag{3}$$

The similarity-based Tf-df is used for weighting approaches,

$$wtf - df_{t,d} = wtf_{t,d} \bullet df_t \tag{4}$$

$$wtf_{t,d} = \begin{cases} 1 + log_{10} tf_{t,d} & if \quad _{TFT,d}, otherwise \\ 0 \end{cases}$$

$$df = \log_{10}\left(K/dft\right) \tag{5}$$

Finally evaluated the weight,

$$w_{t,d} = \left(1 + \log_{10} tf_{t,d}\right) * \log_{10}\left(K/_{dft}\right) \tag{6}$$

Where, $T_{TFT,d}$- Term frequency of term (t), d-document, dft-document frequency, calculating the terms and document frequency is analysis for similarity approached. The cosine measurements are used for document similarity frequency for terms weights and document indexing values.

3.6. Relating Keyword Query Search Summarization (RKQSS)

Keyword searches have always been the most popular and accessible technique to use. It is developed to integrate multiple documents retrieved by information retrieval systems based on key queries like document databases. To perform the text summarization task with the proposed method, extract keyword-based queries from the given information and rank the documents.

First, in the pre-processing structure, all documents can be viewed as tokenized, and weighted terms are called documents. Then, given a set of keywords at query time, it performs a keyword search of the keywords in the document graph to find the association of the keywords in the document. Each document summary is a minimum weight for the corresponding document containing all keywords.

Document Ranking

- Selecting the most relevant document D_1, which queries frequently evaluated document weighs.
- Ranking the other document frequencies of low-weight terms in the document set

Calculating the summary length in Equation (7)

$$Le_{Dx} = \left(Le_T - Le_{sq}\right) * \left(0.2 + 0.6^{x-3} / N - 1\right) \tag{7}$$

To generate a summary, be given the summary length of each term (Le_T), Length (Le_{Dx}) of sequence D_x is calculated based on the terms (sentence) weights, and N is the number of retrieved documents using keywords. Collect keywords used in the searching field and save them in a separate document file for analysing the terms and sentence length.

3.7. Query Searching Based Ranking Summarization Data Retrieval (QS-RSDR)

Utilizing query-based ranking to retrieve and summarize relevant information from unstructured documents employing a simple similarity measure to calculate the similarity between queries, a separate ranking model is created for each training query and its corresponding document. The sentences are ranked based on their scores, determined by tokenizing the sentence and identifying the token with the highest frequency. To eliminate redundancy, a cosine similarity matrix is employed. The rank function prioritizes the removal of more sentences than non-extracted sentences in each training document rather than training a classifier for sentence extraction.

Steps for QS-RSDR

Step 1: Collection of Indexing Documents
For each document:
Perform user searching sentence using a query.
Remove irrelevant words and tokenize the remaining words.

Step 2: Extracting Sentences from the Query
Extract features from the query, such as term length, position, similarity, nouns, weight, date, and numerical data.

Step 3: Sentence Score Calculation
Calculate a score for each sentence based on the extracted features.

Step 4: Sentence Grouping Using Query Searching
Group sentences based on their similarity to the query terms.
Remove similar terms to avoid redundancy.

Step 5: Ranking
Rank the sentences based on their position within the document.

Step 6: QS-RSDR
Select sentences based on the calculation of term weights and generate scores for each sentence.
For each user search, u_s
Retrieve the user's queries (Qs=u_r.getQuerySet()).
Segment the query set by function values.
For each query set, Q_(sr)
If the document contains Q_sr
If the score values for Q_sr in the document < dataset for Q_sr
Update the values for Q_sr in the dataset.
Else
Add Q_sr to the document.
Return the document.

Step 7: Summarization
Generate a summary based on the selected sentences from the document.

Step 8: Stop

The proposed work extracts information features and uses similarity measurement techniques to generate sentence scores. Once the sentence score is generated, the sentence is generated. Sort the set of queries and sentences in each group and include the sentences with relatively high scores in each set in the final summary. A summary of a text document is created by identifying the basic sentences in the document.

4. Results and Discussion

The purpose of the findings and discussion section is to assess the value of the rankings attained and to assess how well different methods perform inside our ranking system. Additionally, the section presents the results obtained from using test data on some documents. The evaluation of summaries generated by the collector involves calculating a score measure to determine the similarity value of the query. In this study, we compare the proposed method, QS-RSDR, with several existing methods, including Combined Query Search Ranking Optimization (CQSRO), Devise Optimal Quality Threshold Clustering (DOQTC), Sequenced User Search Pattern Query Optimization (SUSPQO), and Fuzzy-Based Web Search Privacy Evaluation (FB-WSPES). The simulations conducted in this study allow us to assess the effectiveness and performance of the dissimilar approaches. We evaluate the quality of the rankings achieved by each technique and analyse how well they meet the information retrieval requirements. The comparisons provide insights into the strengths and weaknesses of each method in terms of summarization and retrieval accuracy.

The findings indicate that QS-RSDR outperforms the existing methods regarding query search ranking and document summarization. It demonstrates superior performance regarding the search results' accuracy, relevance, and contextual appropriateness. The proposed method effectively addresses the challenges associated with information retrieval and significantly improves the existing approaches. The comparison with CQSRO, DOQTC, SUSPQO, and FB-WSPES highlights the advantages of QS-RSDR in terms of its effectiveness in retrieving relevant information, optimizing query results, and ensuring privacy protection in web searches. Furthermore, the results validate the significance and potential of the proposed method for improving search performance and enhancing the user experience. These findings contribute to advancing the field of information retrieval and provide valuable insights for researchers and practitioners. In addition, the effectiveness of QS-RSDR opens up opportunities for further research and development in query-based ranking and document summarization.

Table 1. Simulation Parameters

Parameters	Values
Simulation tool	Anaconda
Language	Python
Number of Data	4000
Train data	3000
Test data	1000
Dataset Name	Kaggle query dataset

In Table 1, we provide an overview of the simulation parameters used in the study. The simulation tool employed was Anaconda, a popular platform for data science and machine learning tasks. The programming language utilized for the simulations was Python. The dataset used in the study consisted of a total of 4000 data points. Among these, 3000 data points were utilized for training the models, while the remaining 1000 data points were utilized for testing and evaluating the performance of the methods. The dataset used in the study was sourced from the Kaggle query dataset, which provides a diverse set of queries for information retrieval and analysis purposes. These simulations Parameters were chosen to ensure a comprehensive evaluation of the proposed method and facilitate a meaningful comparison with existing approaches by Equation (8) – (11). Tp – True positive, Tn – True negative, Fp – False positive, Fn – False negative.

$$\text{Precision} = \frac{Tp}{Tp + Fp} \tag{8}$$

$$\text{Recall} = \frac{Tp}{Tp + Fn} \tag{9}$$

$$Accuracy = \frac{Tp + Tn}{Tp + Tn + Fp + Fn} \tag{10}$$

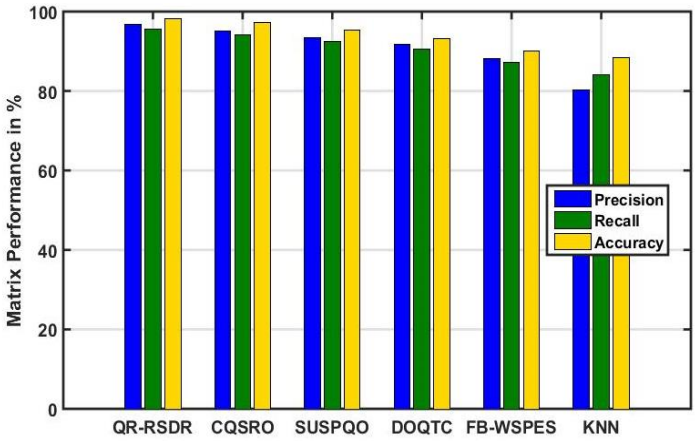

Fig. 2. Analysis of Performance Methods Matrix of Query Searching Based Ranking Summarization Data Retrieval (QS-RSDR)

Accuracy measures the ratio of true positive predictions to all positive predictions made by the model. Recall measures the proportion of true positive predictions among all true positive rates in the dataset. Precision measures the overall accuracy of the model across all categories.

Fig. 2 displays the performance matrix, illustrating the calculation of precision, recall, and accuracy values based on the true and false values in the confusion matrix. The evaluation compares the test data results obtained using different methods. In our proposed method, QS-RSDR, we achieved a precision of 96.8%, recall of 95.5%, and accuracy of 98.1%. These results outperform the previous methods, demonstrating the effectiveness of our approach.

Comparatively, Combined Query Search Ranking Optimization (CQSRO) achieved a precision of 95.2%, recall of 94.2%, and accuracy of 97.2%. Devise Optimal Quality Threshold Clustering (DOQTC) attained a precision of 91.8%, recall of 90.5%, and accuracy of 92.3%. Sequenced User Search Pattern Query Optimization (SUSPQO) yielded a precision of 93.4%, recall of 92.5%, and accuracy of 95.4%. Fuzzy-Based Web Search Privacy Evaluation (FB-WSPES) resulted in a precision of 88.2%, recall of 87.2%, and accuracy of 90.1%. Lastly, K-Nearest Neighbour (KNN) achieved a precision of 80.2%, recall of 84.2%, and accuracy of 88.3%.

These findings highlight the superior performance of our proposed method, QS-RSDR, in terms of precision, recall, and accuracy, indicating its effectiveness in document summarization and information retrieval tasks. The results underscore the significance of incorporating query-based ranking and summarization techniques to enhance the retrieval accuracy and relevance of the retrieved documents.

$$F-Measure=\left(\frac{2*\Pr ecision+\mathrm{Re}\,call}{\Pr ecision+\mathrm{Re}\,call}\right) \tag{11}$$

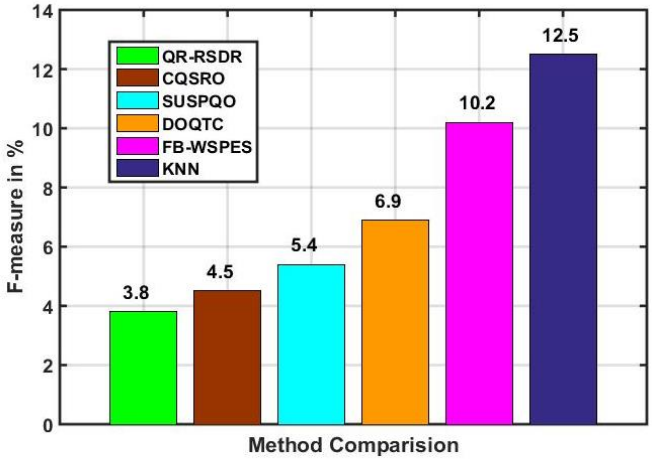

Fig. 3. Analysis of F-measure of Query Searching Based Ranking Summarization Data Retrieval (QS-RSDR)

Fig. 3 presents the false measure based on the evaluation of precision and recall scores. It provides insights into the performance of different methods. In the context of ranked summary data retrieval based on query search, the F-measure can be used to evaluate how well a system returns relevant information while minimizing irrelevant results.

In our proposed method, QS-RSDR, the inaccurate measure is only 3.8%. Comparatively, Combined Query Search Ranking Optimization (CQSRO) yielded an inaccurate measure of 4.5%, Sequenced User Search Pattern Query Optimization (SUSPQO) achieved 5.4%, Devise Optimal Quality Threshold Clustering (DOQTC) resulted in 6.9%

inaccuracy, Fuzzy-Based Web Search Privacy Evaluation (FB-WSPES) showed 10.2% inaccuracy, and K-Nearest Neighbour (KNN) had an inaccurate measure of 12.5%.

These results demonstrate that our proposed technique, QS-RSDR, outperforms the previous methods in terms of minimizing inaccuracies. The lower inaccurate measure in QS-RSDR indicates its ability to provide more precise and relevant results compared to the other classification methods. By reducing the false measures, QS-RSDR enhances the overall performance of document summarization and information retrieval. These findings validate the effectiveness of our proposed method in improving search relevance and accuracy, thereby providing valuable insights for enhancing search engine functionality and user experience.

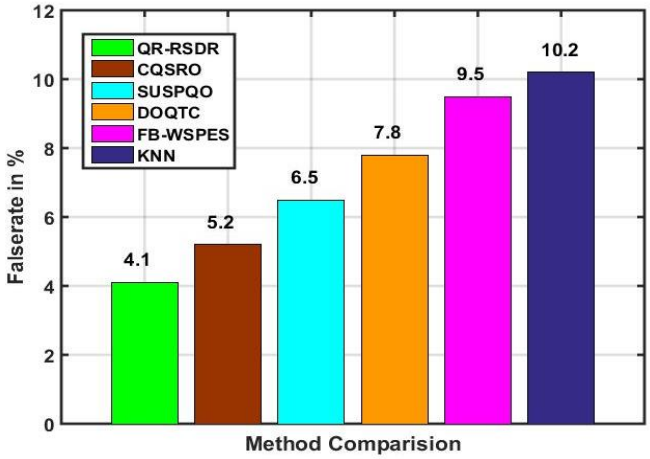

Fig. 4. Information Retrieval false rate Query Searching Based Ranking Summarization Data Retrieval (QS-RSDR)

Fig. 4 presents the false rate analysis classification for information retrieval. It provides a comparison of different methods based on their false measures. Based on a summary of query search rankings, the false positive rate measures the percentage of irrelevant documents that are incorrectly classified as relevant out of all documents that are actually irrelevant. In our proposed method, QS-RSDR, the false rate is calculated to be 4.1%. In comparison, Combined Query Search Ranking Optimization (CQSRO) yielded a false rate of 5.2%, Sequenced User Search Pattern Query Optimization (SUSPQO) achieved 6.5%, Devise Optimal Quality Threshold Clustering (DOQTC) resulted in 7.8% false rate, and Fuzzy-Based Web Search Privacy Evaluation (FB-WSPES) showed 9.5% false rate. K-nearest Neighbour (KNN) had a false rate of 10.2%.

These results indicate that our proposed method, QS-RSDR, performs better than the existing approaches in terms of minimizing false classifications. With a lower false rate, QS-RSDR demonstrates its effectiveness in providing more accurate and reliable information retrieval compared to the other methods. By reducing the false rate, QS-RSDR enhances the precision and reliability of the document summarization process. These findings support the notion that our proposed method improves the quality of search results and contributes to a more effective and efficient information retrieval system.

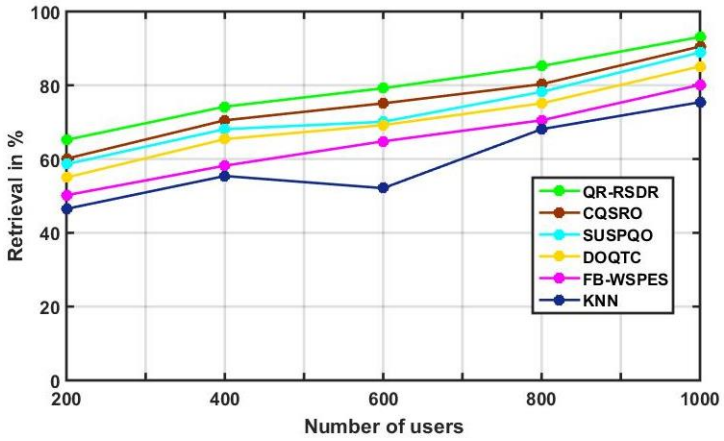

Fig. 5. Analysis of Query retrieval of Query Searching Based Ranking Summarization Data Retrieval (QS-RSDR)

Fig. 5 illustrates a comprehensive evaluation of query performance, comparing the query and retrieval performance of both existing and proposed methods. In our proposed method, QS-RSDR, the query retrieval performance was measured at 93.1%. In comparison, Combined Query Search Ranking Optimization (CQSRO) achieved a performance of 90.5%, Sequenced User Search Pattern Query Optimization (SUSPQO) demonstrated 88.9% performance, Devise

Optimal Quality Threshold Clustering (DOQTC) resulted in 85.1% performance, and Fuzzy-Based Web Search Privacy Evaluation (FB-WSPES) showed 80.1% performance. Finally, k-nearest Neighbour (KNN) yielded a performance of 75.4%.

These findings demonstrate how our suggested method, QS-RSDR, outperforms the current methods in terms of query performance. Furthermore, with a higher query retrieval rate, QS-RSDR effectively retrieves relevant information based on user queries. By improving query performance, QS-RSDR enhances the overall search experience and efficiency of the information retrieval system. These findings support the conclusion that our proposed method outperforms the existing methods and contributes to more precise and reliable query-based information retrieval.

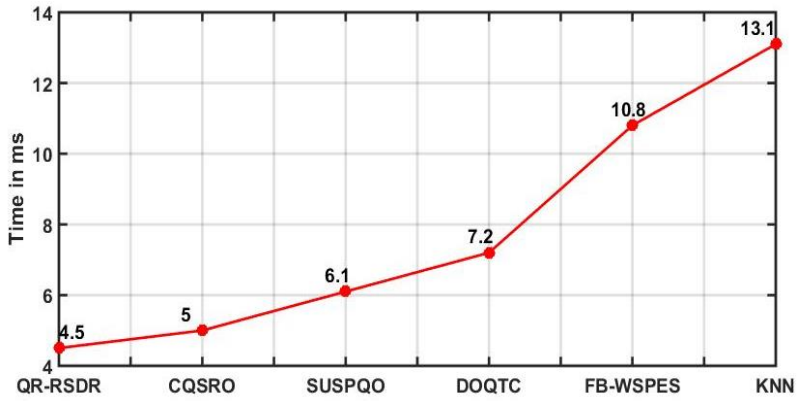

Fig. 6. Time complexity of Query Searching Based Ranking Summarization Data Retrieval (QS-RSDR)

Fig. 6 presents a comparison of Query Execution Time between the existing methods and our proposed technique, QS-RSDR. In QS-RSDR, the time complexity for executing a query was measured at 4.5ms. In contrast, Combined Query Search Ranking Optimization (CQSRO) required 5ms for query execution, Sequenced User Search Pattern Query Optimization (SUSPQO) took 6.1ms, Devise Optimal Quality Threshold Clustering (DOQTC) required 7.2ms, Fuzzy-Based Web Search Privacy Evaluation (FB-WSPES) took 10.8ms, and K-Nearest Neighbour (KNN) took 13.1ms. These results demonstrate that our proposed method, QS-RSDR, achieves a faster query execution time compared to the existing methods. The reduced query execution time enhances the efficiency and responsiveness of the information retrieval system, allowing users to obtain search results more quickly.

By optimizing query execution time, QS-RSDR improves the overall performance and usability of the information retrieval system. These findings highlight the effectiveness of our proposed technique in providing timely and efficient query-based information retrieval. It is important to note that the execution time may vary based on the specific system configuration and dataset size. However, the consistently lower execution time observed in QS-RSDR suggests its potential for practical implementation in real-world information retrieval systems.

5. Conclusion

In this study, we proposed a novel method, QS-RSDR, for document-based query searching using a keyword search in information retrieval. The study contributes to the ongoing efforts in query-based information retrieval, providing valuable insights and a promising model for further advancements in the field. Through the evaluation of summaries and automated analysis, we demonstrated the success of QS-RSDR in achieving high precision (96.8%), recall (95.5%), F-measure (3.8%), and accuracy (98.1%). Additionally, the false rate for information retrieval was found to be 4.1%, and the query retrieval rate reached 93.1%. Moreover, QS-RSDR significantly reduced the time complexity for searching to 4.5 ms.

The results of our study highlight the importance of continuous research and development in the field of information retrieval. The proposed QS-RSDR model offers a promising approach for improving the efficiency and effectiveness of query-based information retrieval. By allowing users to extract named entities, detect multi-word phrases, and perform keyword-based searches, QS-RSDR enhances the overall search experience.

Future work will further refine and optimise the QS-RSDR model to achieve even better processing results. This may involve exploring additional features, refining scoring mechanisms, and evaluating the model on larger datasets. With continued advancements in information retrieval techniques, we can anticipate improved search relevance, faster query execution, and enhanced user satisfaction.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

[1] R. Das, A. Godbole, D. Kavarthapu, Z. Gong, A. Singhal, M. Yu, and A. McCallum, "Multi-step entity-centric information retrieval for multi-hop question answering," *In Proceedings of the 2nd Workshop on Machine Reading for Question Answering* pp. 113-118, 2019. DOI:10.18653/v1/D19-5816

[2] K. Roitero, E. Maddalena, S. Mizzaro, and F. Scholer, "On the effect of relevance scales in crowdsourcing relevance assessments for Information Retrieval evaluation," *Inf Process Manag,* vol. 58(6), pp. 102688, 2021. https://doi.org/10.1016/j.ipm.2021.102688

[3] M. Zhao, S. Yan, B. Liu, X. Zhong, Q. Hao, H. Chen, and W. Guo, "QBSUM: A large-scale query-based document summarization dataset from real-world applications," Comput Speech Lang, vol. 66, pp. 101166, 2021. https://doi.org/10.1016/j.csl.2020.101166

[4] M. Mohd, R. Jan, and M. Shah, "Text document summarization using word embedding," *Expert Syst. Appl,* vol. 143, pp. 112958, 2020. https://doi.org/10.1016/j.eswa.2019.112958

[5] R. C. Belwal, S. Rai, and A. Gupta, "A new graph-based extractive text summarization using keywords or topic modeling," Ambient Intell Humaniz Comput, vol. 12(10), pp. 8975-8990, 2021. https://doi.org/10.1007/s12652-020-02591-x

[6] A.Al-Saleh, and M. E. B. Menai, "Solving multi-document summarization as an orienteering problem," Algorithms, vol. 11(7), pp. 96, 2018. https://doi.org/10.3390/a11070096

[7] K. Manju, S. David Peter, and S. M. Idicula, "A framework for generating extractive summary from multiple Malayalam documents." Information, vol. 12(1), pp. 41, 2021. https://dx.doi.org/10.3390/info12010041

[8] H. Van Lierde, and T. W. Chow, "Query-oriented text summarization based on hypergraph transversals," *Inf Process Manag,* vol. 56(4), pp. 1317-1338, 2019. https://doi.org/10.1016/j.ipm.2019.03.003

[9] K. Kumar, "Text query based summarized event searching interface system using deep learning over cloud," *Multimed. Tools Appl,* vol. 80(7), pp. 11079-11094. 2021. https://doi.org/10.1007/s11042-020-10157-4

[10] D. Patel, S. Shah, and H. Chhinkaniwala, "Fuzzy logic based multi document summarization with improved sentence scoring and redundancy removal technique," *Expert Syst. Appl,* vol. 134, pp. 167-177, 2019. https://doi.org/10.1016/j.eswa.2019.05.045

[11] A.P.Widyassari, S. Rustad, G. F. Shidik, E. Noersasongko, A. Syukur, and A. Affandy, "Review of automatic text summarization techniques & methods." *J. King Saud Univ., Comp,inf. Sci,* vol. 34(4), pp.1029-1046, 2022. https://doi.org/10.1016/j.jksuci.2020.05.006

[12] F. Bayatmakou, A. Mohebi, and A. Ahmadi, "An interactive query-based approach for summarizing scientific documents," Inf. Discov, vol. 50(2), pp. 176-191, 2022. https://doi.org/10.1108/IDD-10-2020-0124

[13] J. A. Jilo, A. T. Alemu, F. Z. Abera, and F. Rashid, "An integrated development of a query-based document summarization for Afaan Oromo using morphological analysis," *Indian J Sci Technol,* vol. 14(38), pp. 2946-2952, 2021. https://doi.org/10.17485/IJST/v14i38.1182

[14] Y. Li, J. Ma, and Y. Zhang, "Image retrieval from remote sensing big data: A survey." Inf Fusion, vol. 67, pp.94-115, 2021.https://doi.org/10.1016/j.inffus.2020.10.008

[15] A.Qaroush, I. A. Farha, W. Ghanem, M. Washaha, and E. Maali, "An efficient single document Arabic text summarization using a combination of statistical and semantic features," *Journal of King Saud University-Computer and Information Sciences,* vol. 33(6), pp. 677-692, 2021. https://doi.org/10.1016/j.jksuci.2019.03.010

[16] 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, pp. 113679, 2021. https://doi.org/10.1016/j.eswa.2020.113679

[17] N. Rahman, and B. Borah, "Improvement of query-based text summarization using word sense disambiguation," COMPLEX INTELL SYST, vol. 6, pp. 75-85, 2020. https://doi.org/10.1007/s40747-019-0115-2

[18] A.Sutton, M. Clowes, L. Preston, and A. Booth, "Meeting the review family: exploring review types and associated information retrieval requirements," *COMPLEX INTELL SYST*, vol. 36(3), pp. 202-222, 2019. https://doi.org/10.1111/hir.12276

[19] 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),* pp. 330-335, 2020. IEEE. DOI: 10.1109/ICACCM50413.2020.9212923

[20] P. Tampakis, D. Spyrellis, C. Doulkeridis, N. Pelekis, C. Kalyvas, and A. Vlachou, "A novel indexing method for spatial-keyword range queries," *In 17th International Symposium on Spatial and Temporal Databases,* pp. 54-63, 2021. https://doi.org/10.1145/3469830.3470897

[21] J. A. Nasir, I. Varlamis, and S. Ishfaq, "A knowledge-based semantic framework for query expansion," Inf Process Manag, vol. 56(5), pp. 1605-1617, 2019. https://doi.org/10.1016/j.ipm.2019.04.007

[22] J. Guo, Y. Fan, L. Pang, L. Yang, Q. Ai, H. Zamani, and X. Cheng, "A deep look into neural ranking models for information retrieval," Inf Process Manag, vol. 57(6), pp. 102067, 2020. https://doi.org/10.1016/j.ipm.2019.102067

[23] Z. Wu, S. Shen, X. Lian, X. Su, and E. Chen, "A dummy-based user privacy protection approach for text information retrieval," *Knowledge-Based Systems,* vol. 195, pp. 105679, 2020. https://doi.org/10.1016/j.knosys.2020.105679

[24] R. Handa, C. R. Krishna, and N. Aggarwal, "Document clustering for efficient and secure information retrieval from cloud," *Concurr. Comput. Pract. Exp*, vol. 31(15), pp. e5127, 2019. https://doi.org/10.1002/cpe.5127

[25] D. R. Ganesh, and M. Chithambarathanu, "A Survey on Hybrid PSO and SVM Algorithm for Information Retrieval." *In Data Intelligence and Cognitive Informatics: Proceedings of ICDICI 2022* pp. 121-130, 2022. Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-19-6004-8_11

Authors' Profiles

Surya S. is an accomplished scholar and researcher, holds a Master's degree in Computer Applications (MCA) and an M. Phil. Currently pursuing a Ph.D. at the PG and Research Department of Computer Science and Computer Applications in Vivekanandha College of Arts and Sciences for Women (Autonomous), situated in Elayampalayam, Tiruchengode, Namakkal District, Tamil Nadu.

Sumitra P. is an esteemed academic figure and a proficient researcher, boasting a remarkable academic journey with degrees including M.Sc., M.Phil., M.C.A., and a Ph.D. Currently serving as a Research Supervisor at the PG and Research Department of Computer Science and Computer Applications at Vivekanandha College of Arts and Sciences for Women (Autonomous) in Elayampalayam, Tiruchengode, Namakkal District, Tamil Nadu.

How to cite this paper: Surya S., Sumitra P., "Document Summarization Based on Information Retrieval Using Query Search Ranking Method", International Journal of Modern Education and Computer Science(IJMECS), Vol.16, No.4, pp. 58-70, 2024. DOI:10.5815/ijmecs.2024.04.05