Topic Model and Deep Reinforcement Learning Applied to the Extractive Query-Based Summarization Task

Abeer Hussien
Software Department, College of Information Technology of University of Babylon, Iraq
abirhusseinj.sw@student.uobabylon.edu.iq (corresponding author)

Wafaa Al Hameed
Software Department, College of Information Technology of University of Babylon, Iraq
it.wafaa.mohammed@uobabylon.edu.iq

Received: 9 September 2025 | Revised: 18 October 2025 and 19 October 2025 | Accepted: 21 October 2025

 DOI: https://doi.org/10.48084/etasr.14632

ABSTRACT

The rapid expansion of digital information has generated an increasing desire for intelligent systems that can produce abstract and relevance summaries from a large document corpus. Conventional summarization methods sometimes fail to address multiple documents effectively, particularly when the summaries should meet a specific user's query. Although significant advances have been made, several extractive summarization methods face challenges in preserving non-redundancy, coherence, and relevance, especially when dealing with different query information and multiple document inputs. Additionally, traditional methods lack mechanisms for balancing diversity and semantic similarity while creating summaries that align with the intention of queries. To address these challenges, this study suggests an extractive query-based summarization model that combines BERT embeddings, semantic clustering (K-means), topic modeling (LDA), and Deep Reinforcement Learning (DRL), identifying a sentence and choosing it or skipping it based on a reward function that is designed with multi-objective integration of BERT-based coherence scores with Maximal Marginal Relevance (MMR). The proposed system was trained on the QuerySum dataset and tested on the CNN/Daily Mail dataset. The experimental results show that the proposed system outperforms traditional approaches in various measures. Combining the BERT-based coherence score and MMR for designing a reward function helps to improve ROUGE scores [ROUGE-1 (50.03%), ROUGE-2 (27.30%), and ROUGE-L (39.86%)] and increases the BERT score (88.70%). Additionally, the generated summaries were relevant, coherent, concise, and less redundant compared to existing approaches.

Keywords-Maximal Marginal Relevance (MMR); Deep Reinforcement Learning (DRL); Recall-Oriented Understudy for Gisting Evaluation (ROUGE); topic modeling (LDA)

I. INTRODUCTION

The rapid expansion of information on the internet has made information overload a common issue. Capturing all information is impossible, especially with large text messages and massive news records. Therefore, access to prominent and critical information becomes a challenge [1]. Text Summarization (TS) represents a solution for these issues. TS is the process of generating a summary from a single or multiple documents, maintaining its meaning by effectively extracting the main information from the sources [2]. The summary provides a useful guide to draw attention to information, make a decision about whether a document is helpful or not, and help users reduce the time spent. As a result, it is necessary to improve automatic summary models.

Recently, many attempts have been made in summarization methods [3]. Extractive summarization methods aim to choose prominent phrases, sentences, or words from the original documents, and are commonly used [4]. Before deep learning, traditional summarization approaches utilized simple statistics and handcrafted features for specific tasks. Since the advent of deep learning, various attempts have focused on training automated end-to-end summarization models. Multi-Document Summarization (MDS) is a set of related documents with the same topic that is summarized in one summary and is more complex than Single Document Summarization (SDS) [5]. Although several approaches have been applied to both MDS and SDS, there are major challenges in multi-document summarization, which must be addressed to generate high-quality summaries [6].

The performance of general summarization approaches is constrained and does not meet the requirements. This is attributed to the mismatch between the evaluation criterion, Recall-Oriented Understudy for Gisting Evaluation (ROUGE), and learning. In query-based summarization, a summary is generated based on the user query, where the textual contents are investigated to match such a query. The learning aim is to maximize the probability of the ground-truth summary. ROUGE mostly depends on the lexical concordance between the candidate and ground-truth summaries. To address this problem, this study proposes a new hybrid model for extractive query-based summarization for MDS, using a topic model (LDA) [7] and a clustering algorithm (K-means) with Deep Reinforcement Learning (DRL), employing a Deep Q-Network (DQN) to train summarization models [8]. Using DRL, a reward function is designed using the Maximal Marginal Relevance (MMR) technique that guides learning toward goals beyond traditional reward functions. The proposed model improves coverage, coherence, relevance, diversity, and readability. The experimental results utilized the CNN/Daily Mail and Query Sum datasets, and the ROUGE measures were used to evaluate the performance of the model.

Text summarization has been an active research area for many years. Previous research in Automatic Text Summarization (ATS) has focused on different methods to summarize large-scale textual content into brief yet useful summaries. In [9], the query-based summarization problem was addressed by combining transformer-based models and deep reinforcement learning. The proposed extractive model was unsupervised, based on clustering mechanisms and sentence ranking using cosine similarity and sentence embedding to determine the relevance of sentences to the query. In the RL-based model, an agent chooses the main sentences based on the reward function. The proposed system was evaluated on the Debatepedia and DUC 2005-2007 benchmark datasets and evaluated using the ROUGE score. However, reinforcement learning was based on traditional reward functions, and this study did not use pretrained models, such as GPT or BERT, for contextual understanding and semantic representation. RL-MMR [10] is an RL model for extractive MDS that combines the MMR precept from neural SDS in a hierarchical and comprehensive neural sentence representation approach utilizing a soft attention mechanism. This system implements a hierarchical encoder (Bi-LSTM+CNN) for document and sentence representation that is enhanced by leveraging an actor-critic policy utilizing ROUGE-released final and intermediate rewards to support non-redundant and prominent summaries. Experiments on standard datasets (TAC 2011 and DUC 2004) demonstrated that RL-MMR outperformed both traditional neural (RNN-EXT+RL, PG-MMR) and baseline statistical approaches (e.g., Centroid, Lex Rank, DPP). However, this model was based on TF-IDF for embedding and was not generalized.

In [11], an extractive text summarization framework combined coreference resolution, dynamic clustering, and BERT to generate optimal-length summaries. The BERT model was used to create sentence embeddings. Then, the k-means algorithm was used to enhance representation, selecting the sentences that are nearest to the cluster's centroid. The number of chosen sentences from each cluster was dynamically satisfied by a regression model based on the Within Cluster Sum of Squares (WCSS), which represents the length of the created summary. The CNN/Daily Mail benchmark dataset was used to evaluate this model using ROUGE scores. However, this model's regression accuracy was limited by the fixed-length and short gold-standard summaries in CNN/Daily Mail. In [12], an unsupervised technique for multi-document query-based extractive summarization was based on a sense-oriented semantic relatedness measure, comparing the input text sentence with the query to enhance the accuracy of sentence choice. This approach suggested a Word Sense Disambiguation (WSD) approach to extract query-relevant sentences. The implementation involved the DUC 2005, DUC 2006, and DUC 2007 datasets and was evaluated using ROUGE measures. However, this system did not use the BERT model for contextual meaning, relying on static lexical resources such as WordNet, and had a lack of adaptability.

In [13], an interactive model was proposed for extractive query-based summarization to enhance the relevance and quality of summaries for scientific documents. This model applied Rapid Automatic Keyword Extraction (RAKE) to determine possible keywords. A Genetic Algorithm (GA) was used to expand the sentence group to ensure minimal redundancy and diversity by using the MMR method. This system was evaluated using different approaches, applying it to 72,000 scientific documents from the Web of Science (AI category), and testing it with 45 graduate students in Artificial Intelligence (AI). This system was based on explicit user feedback and domain-specific evaluation, could not capture semantic meaning since it did not use state-of-the-art models such as BERT and SBERT, and required more computational time for GA. In [14], an unsupervised query-focused summarization system combined redundancy reduction approaches with transfer learning from a pretrained sentence embedding paradigm and classical information retrieval. This system utilized USE-DAN, USE-Transformer, and uSIF to establish syntactic and contextual relationships. In addition, a hybrid technique was used to linearly integrate semantic similarity with BM25 term-based retrieval to compute sentence relevance, and the MR technique was used to balance novelty and relevance for reordering candidate sentences. This model was used on the DUC benchmark datasets (2007, 2006, 2005) and assessed using the ROUGE scores. In addition, there were suggestions for using recent models, such as GPT-3 and T5, for text summarization issues and investigating transfer learning capabilities from pre-trained models for summary creation. However, this system suffered from a lack of coherence because it did not use embedding models to capture contextual meaning.

In [15], a DRL approach was presented for interactive multi-document summarization, consisting of two models: MSumm for creating the initial summary and query-based expansions, and MSugg for generating dynamic lists of proposed queries to assist exploration. Based on an MMR reward strategy, MSumm selects the sentences that are relevant to the query to enhance semantic similarity based on ROUGE. MSugg works on the sentence level, utilizing RL to select the newly formed sentences that are salient and unexplored, depending on their novelty and relevance. The model was trained and assessed on the DUC 2006/2007 datasets. However, this framework occasionally falls short of providing enough attention to the user's query, and the system's training with synthetically created queries cannot precisely represent actual user input. In [16], an unsupervised extractive summarization approach integrated topic modeling and clustering to create coherent and non-redundant summaries. This model was trained and evaluated using ROUGE measures on two benchmark datasets, CNN/DailyMail and Opinosis, performing well on small datasets. However, this system has several limitations, since LDA is more effective with multiple documents and is based on WordNet for semantic similarity, leading to a lack of capturing contextual dependencies.

BERT is an effective deep learning model for NLP tasks, based on the transformer architecture [17], built to understand the context of words in a sentence bidirectionally, i.e., it takes the objective word and looks for the words that are before and after it, rather than looking in one direction as in traditional models. The pretraining processes that make BERT robust are Next Sentence Prediction (NSP)—where the model learns sentence-level coherence—and Masked Language Modeling (MLM)—where random tokens are masked and predicted. This dual-task setup supplies BERT with syntactic understanding and powerful semantic similarity, making it highly effective for downstream NLP tasks such as question answering, summarization, and text classification.

In [18], a comparative investigation on the performance of unsupervised extractive summarization approaches was presented, assessing Probabilistic LSA (PLSA), Latent Semantic Analysis (LSA), and MMR on a Persian broadcast news dataset consisting of over 7,000 sentences and 115,000 words. This system focused on query-based and generic summarization. Query Expansion (QE) was utilized for reducing the mismatch between the vocabulary issues in information retrieval, enriching ambiguous or short queries with extra terms.

Classical methods, such as Rocchio and RM3, depend on pseudo-relevance feedback or BOW and fail to obtain deeper semantic relationships. Recent research has advanced through contextualized expansion techniques, leveraging pretrained transformer models. For instance, in [19], concept-based expansions (adding semantically related terms) were integrated with structural paraphrasing (e.g., converting short queries into full natural language questions), resulting in more powerful improvements. In [20], a three-phase model used BERT to re-rank candidate documents, then select the most relevant tokens from the top-ranked documents, and finally combine them into an expanded query for a second re-ranking phase. BERT-based QE is most effective when expansion is not limited to additional keywords but instead includes semantic structure and contextual evidence. In [21], a supervised learning-to-rank system effectively combined global and local information for extractive MDS to enhance sentence choice. This method relied on a rich set of feature engineering to measure the significance of a sentence, and was evaluated on five benchmark datasets, both Vietnamese and English. In the Vietnamese datasets, ROUGE-1 results were 0.819 and 0.801, while on DUC 2004, ROUGE-1 was 0.3954. However, this system did not utilize contextual embedding models, was strongly based on manual feature engineering, and redundancy was addressed through constant similarity thresholds.

The key contributions of this study are:

- Builds a hybrid summarization model that integrates topic modeling, clustering, contextual embedding, and DLR, leveraging a cluster-specific query matching mechanism that ensures efficient and focused summarization by narrowing the search space to the most relevant cluster of documents to support short or ambiguous queries.
- Leverages the Bellman equation in DRL and implements a BERT-based coherence scoring method that integrates the MMR principle in building the reward function. This allows the agent to generate logically consistent and semantically rich summaries.
- Designs a DQN-based sentence selection model that uses BERT embeddings [22] to evaluate and optimize relevance, novelty, and coherence during summary generation.

II. PROPOSED SYSTEM

The proposed architecture, shown in Figure 1, incorporates several NLP techniques to generate a relevant and concise summary, each crucial in ensuring that the generated summary is query-focused and comprehensive while avoiding redundancy. The proposed system operates in six main steps.

A. Input Stage

The input for the training step of the system is a corpus of documents D with multiple queries Q and multiple reference summaries. Testing is performed in the same manner but with a new corpus.

B. Initialization Stage

This stage integrates a topic model (LDA) and clustering (K-means algorithm). LDA is used to identify the distribution of topics within the document and align them with the query. Thus, each sentence and document has a probability distribution on these topics. This creates a pure understanding of topics in documents and helps solve the challenge of choosing the sentences that are most relevant to the query to generate the summary. Filtering out irrelevant topics, condensing the content to the most pertinent information, matching the topic distribution with the query, and analyzing the distribution of these topics help address some of the challenges in summarization. Based on the probabilistic distribution of topics identified by the LDA (t_1, t_2,…, t_n), these documents are sorted using a k-means algorithm into a set of clusters (c_1, c_2,…, c_n). The idea is to group documents that are semantically similar and cover specific topics. The system tokenizes the documents in each cluster into sentences. The tokens are then transformed into vector numerical representations, i.e., embedding using BERT. Since BERT is bidirectional, it can efficiently learn the context of each word by examining its neighboring words, and utilizes an attention method for focusing on significant words. This helps in best understanding semantic relationships and generating context-aware embeddings for each token.

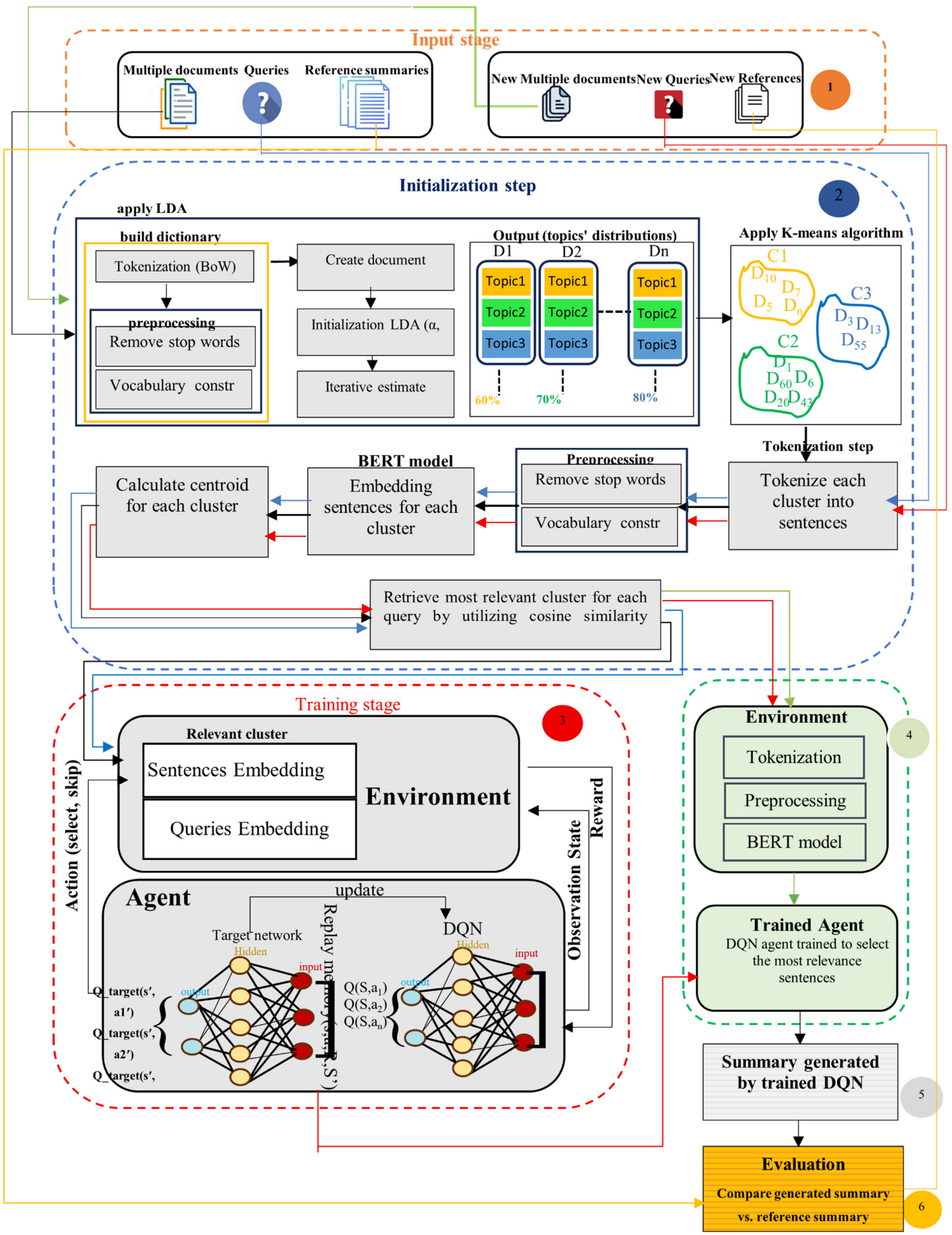

Fig. 1. The proposed system.

Each cluster contains vectors (sentence embeddings) created with BERT following the tokenization stage. Using cosine similarity between the query embedding and the centroid calculated for each cluster, rather than comparing with each sentence embedding, the proposed system reduces the time required to retrieve only the cluster that is most relevant to each query. Assume, for instance, that each cluster C has n vectors, v_1, v_2,v_n, where each vector v_i represents a sentence embedding. The following formula is used to determine the cluster's centroid c:

$$c = \frac{1}{n}\sum_{i=1}^{n} vi \quad (1)$$

Applying the k-means with LDA addresses the relevance challenge faced by the summarization issue after retrieving the most relevant cluster. The initialization stage also helps to narrow the search space.

C. Training Stage

This stage utilizes a DRL model and a construct agent for unit (sentence) selection to generate a summary by training on the most relevant clusters $[C_0\ (D_1, D_2, \ldots, D_n)$, $C_1\ (D_1, D_2, \ldots, D_n)$, ..., $Cn]$ with the corresponding queries $(q_1, q_2, \ldots q_n)$. The output of the initialization stage is used in the testing step to generate an Extractive Query-Based Summary (EQBS). The combination of (D and Q) represents an environment where each document includes m sentences, with $\{S_1, S_2 \ldots., S_m\}$ being the possibility of an action. The selected sentences and their relationships with the query represent the current state (S). An action (a) will either select or skip a sentence based on the current state. The reward function (R) considers the relevance to the query based on the cosine similarity measure, novelty to avoid redundancy, and coherence. The main goal of this system is to create an EQBS by selecting m of sentences from D , where $E = \{se_1, se_2, \ldots, se_m\}$ and $m < n$. This summary must be prominent, should not have repeated sentences, and should be relevant to the query.

First, preprocessing is applied, with a set of sentences $\{S_1, S_2, \ldots , S_n\}$ and a query Q as input. Preprocessing involves removing noise such as stop words and special characters, and then converting the text to lowercase to ensure uniformity.

The summarization environment is defined by employing DRL, where the summarization process is modeled as an MDP with the concepts State Representation, Action Space, and Reward Function. The combination of the selected sentence vector and query vector, resulting from the embedding stage, contributes to constructing an environment that an agent interacts with to generate a summary. The agent observes the current state representation (selected sentences and remaining candidates) and, from the action space, either selects the current sentence or skips it.

The agent is built using two deep neural networks, a Deep Q-Network (DQN) and a target network with experience replay. The DQN agent is used to learn an optimal policy for selecting sentences to predict the long-term cumulative reward for each possible action (2). DQN uses a neural network to approximate the Q-value function to generate informative features to represent the states of the reinforcement Bellman model. DQN is used when the action space or states are huge. During training, a reward is mapped to an agent, taking into account relevance, novelty, and coherence. The agent learns which sentences maximize the reward, balancing query relevance and information diversity, and utilizing experience replay and policies to balance the exploration of new sentence combinations and the exploitation of known good policies. At the training step, an agent periodically selects sentences, updates the state, and receives rewards:

$$CumulativeReward = r_0 + \gamma\, r_1 + \gamma\, r_2 + \cdots \quad (2)$$

where r_0 is the instant reward after the first action, r_1 represents the reward after the second action, and r_2 ,r_3, ... refer to rewards further in the future, and γ refers to the discount factor between 0 and 1 that controls how much future rewards are valued compared to instant rewards

A target network is used to stabilize learning by reducing oscillations in Q-value updates. At each step, Experience Replay is employed to store tuples [state (s), action (a), reward (r), next state(s')], which provide samples for training an agent. During the training stage, an agent periodically selects sentences, updates the state, and receives rewards. This allows the model to work in a sequential decision-making process to generate the summary.

The reward given to an agent during training takes into account many factors, as shown in Figure 2.

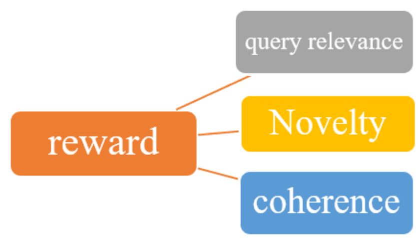

Fig. 2. Reward factors.

This model implements the MMR technique with a BERT-based coherence model to build the reward function:

$$r_t = \underbrace{\lambda_1 \times Relevance + \lambda_2 \times Novelty}_{MMR\ technique} + \lambda_3 \times Coherence \quad (3)$$

The $\lambda_1 \times Relevance + \lambda_2 \times Novelty$ represents the implementation of MMR, which aims to handle some challenges faced by generative extractive summarization tasks, such as relevance (measured using cosine similarity between the query embedding and sentence embedding) and diversity, by reducing duplicates and ensuring that the selected sentences are distinct from those previously selected for inclusion in the generated summary. MMR guarantees that the final summary is comprehensive and varied, capturing various aspects of the query without repeating information (4). This process will finish when the required length of the generated summary is reached.

$$MMR(s) = \underbrace{\lambda \times Similarity(s, Query)}_{Relevance(S,Q)} - \underbrace{(1-\lambda) \times \max_{s' \in S} Similarity(s, s')}_{Novelty(news, Sselected)} \quad (4)$$

where S is the selected sentence, Q is the query, $s'\colon \in S_{selected}$, and $\lambda_1, \lambda_2, \lambda_3$ refer to hyperparameters (weights).

By design, the reward function also takes into account coherence to measure the Semantic Relatedness Between Adjacent Sentences, which is measured using the BERT-based coherence model equation (5):

$$Coherence(s_{t-1}, s_t) = \frac{1}{n-1} \sum_{i=1}^{n-1} \cos\,(s_{t-1}, s_t) \quad (5)$$

D. Testing Stage

This stage involves the trained DQN agent forming a concise final extractive summary for each query by concatenating selected relevant sentences tailored to the query. After training, a new corpus with the same structure as that used in training (multiple documents, multiple queries, and multiple reference summaries) is initially used to apply the LDA model and K-means to narrow the research space by retrieving the most relevant cluster for each query. The trained DQN agent uses the query and sentence embeddings achieved using the BERT model. The combination of sentence embedding and query embedding represents the state for the trained DQN agent to interact with an environment and select the sentence to be added into the generated summary or skip it to generate the extractive summarization for each query.

E. Evaluation Stage

The performance of the proposed method was assessed and compared with different text summarization methods using the ROUGE standard to evaluate validity and dependability. ROUGE has metrics that are used to determine a summary's quality by contrasting it with other people's summaries [23]. ROUGE fundamentally measures the overlap of n-grams, i.e., counts the number of words that interfere between the system-created summaries and the summaries' references. ROUGE involves a variety of automatic metrics such as ROUGE-1, which measures the overlap of unigrams (single words), ROUGE-2, which measures the overlap of bigrams (two-word sequences), and ROUGE-L, which measures the Longest Common Subsequence (LCS) for two provided sequences. ROUGE-L estimates the proportion of the LCS of the generated summary to the LCS of the reference summary to determine how similar a reference summary and a candidate one are to each other. These metrics are typically used to evaluate how the proposed system captures important content from a reference summary.

In addition, the proposed model was evaluated using the BERT score to measure the relevance of selected sentences to the user query using contextual embeddings (BERT similarity) [24], which is a recently suggested metric for the text summarization task to detect how two tokens of text are semantically similar to each other. This metric is more robust in extractive summaries, as it is necessary that sentences are extracted from the document but may not exactly match the reference, and are highly precise and relevant to the query. Additionally, human evaluation can be used to evaluate the proposed framework, taking into account its readability, diversity, and relevance to the query.

III. DATASET AND QUERIES USED

Two datasets were used. For DRL training, the QuerySum dataset was used because it directly supports query-based summarization tasks. The CNN/DailyMail dataset was used in the testing and comparative validation stage. These datasets were used to ensure that the proposed model not only learns from a query-focused dataset but also demonstrates generalizability and robustness when applied to an expanded summarization benchmark.

The QuerySum dataset [25] is a large-scale benchmark for multi-document Query-Focused Summarization (QFS) tasks. It includes 27,041 high-quality samples, each including a natural-language query (Q), a human-validated summary (reference or gold summary), and up to 10 related multiple source documents, with an average of 5.5 per sample, retrieved primarily from Wikipedia. QFS simulates real-world requirements where the user introduces a question and the system retrieves the condensed and relevant information. Unlike classical summarization datasets, QuerySum has summaries dependent on user queries, which are complex questions across multiple documents. It focuses on non-factoid queries, such as "why," "what," and "how," that demand a deeper understanding and synthesis of information. This makes it more challenging and realistic compared to factoid-based Query Answering (QA) datasets. It is built using seed queries extracted from Answers.com and extended via Google Search to form query clusters, further annotated manually to define semantic relationships (synonymous, related, or unrelated). This structure supports deeper modeling of query intent and semantic generalization.

article data	query	summary
(ew.com) -- former `` Saturday night live " cast member Kristen wig is returning to her old stomping grounds to make her debut as in the next few months, beginning see the entertainment weekly and time inc. all rights reserved.	vampire weekend	vampire weekend will perform may 11 and kanye west will perform may 18
Austin, Texas (CNN) -- lady bird Johnson, who was first lady during the 1960s and in her later years became a national its mission is the research and preservation of native plants throughout the United States. e-mail to a friend	Texas	former first lady, widow of lyndon baines johnson, dies in texas

Fig. 3. Sample of Query Sum.

The CNN/Daily Mail dataset [26], used for testing, contains CNN and Daily Mail news articles and a human-written gold summary. The dataset has 286,817 training pairs, 13,368 validation pairs, and 11,487 test pairs. The training set has 766 words across 29.74 sentences, while the summaries include 53 words and 3.72 sentences.

The query is an essential part of the input stage, as it provides the context or intent that assists in selecting sentences from documents to generate an EQBS. This query is either already founded in the structure of the dataset, as in the QuerySum dataset (used in the training stage), extracted automatically using different approaches, such as Text Rank to extract keywords in short queries (as in the test stage with the CNN/Daily Mail dataset), or written by humans (i.e. written by users or experts in the domain).

IV. RESULTS AND DISCUSSION

The proposed extractive query-based summarization framework was evaluated using ROUGE and BERT scores. Initially, LDA was applied to obtain the topic distribution for each document, and Figures 4-5 show the results based on the distribution of topics for each document.

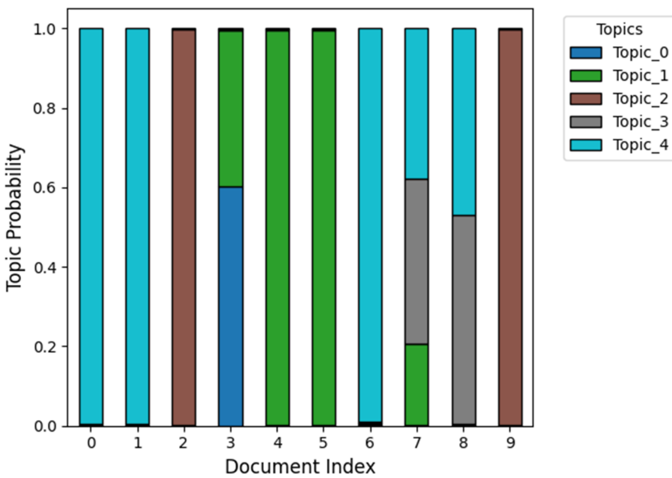

Fig. 4. Topic distribution per document (first 10).

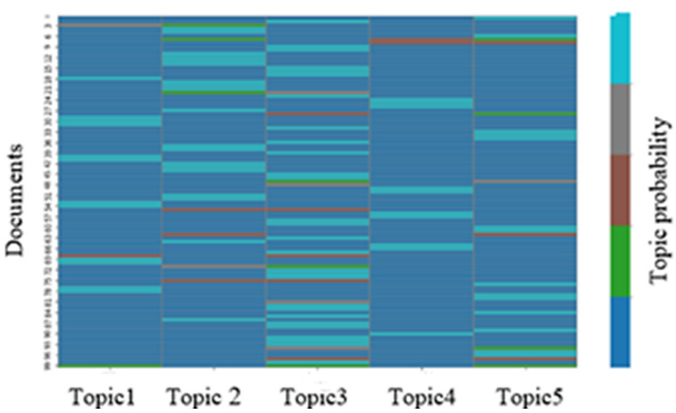

Fig. 5. Heatmap of topic distributions across documents.

Figure 6 shows a word cloud for the top words in each document. Based on these topic distributions, the documents were clustered using K-means, introducing a focused context for each query through the summarization process. Finally, Figure 7 shows documents in each cluster.

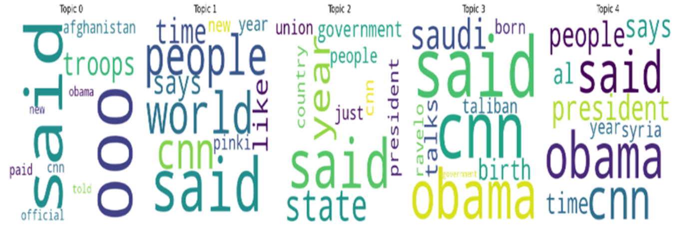

Fig. 6. Results of applying the LDA model.

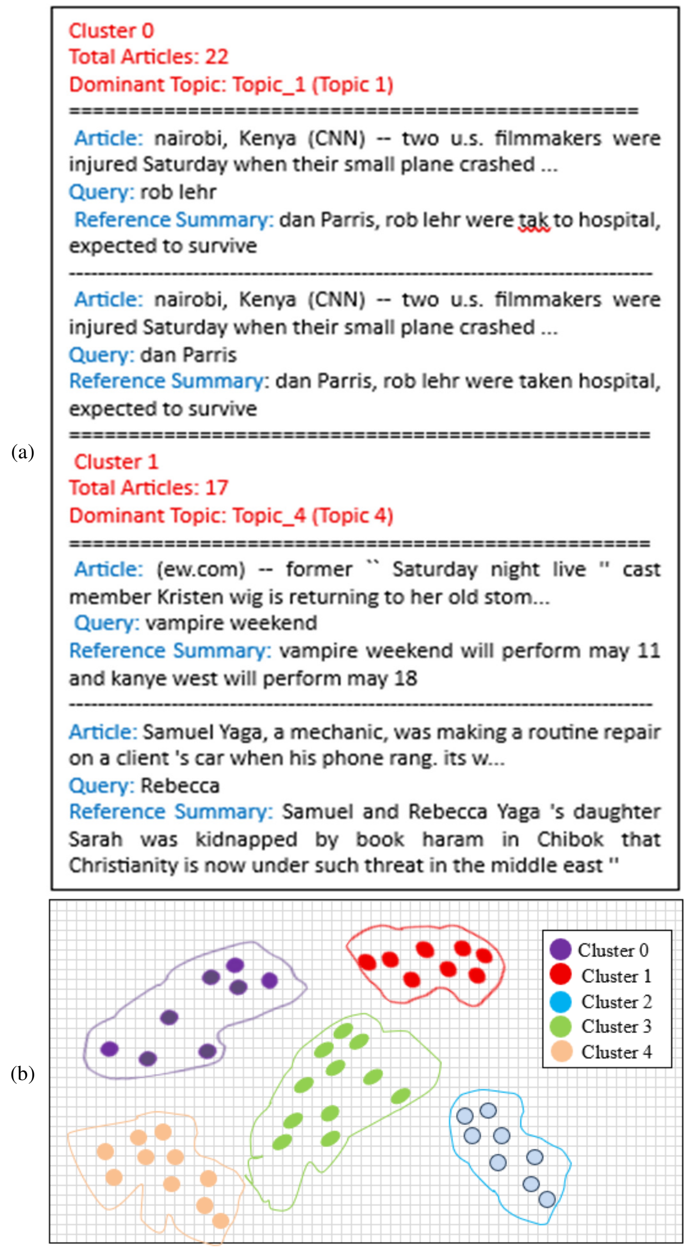

Fig. 7. Documents in each cluster.

A pretrained BERT model was used to tokenize the documents in each cluster, along with the queries. To determine the most relevant cluster for each query, the centroid for each cluster (representing the mean for all sentence embeddings) was calculated as shown in Figure 8.

```
Cluster Centroids:
Cluster 0 Centroid:
  Topic_0: 0.0009   Topic_1: 0.9361   Topic_2: 0.0331   Topic_3: 0.0081   Topic_4: 0.0218
------------------------------
Cluster 1 Centroid:
  Topic_0: 0.0009   Topic_1: 0.0030   Topic_2: 0.0416   Topic_3: 0.0085   Topic_4: 0.9460
------------------------------
Cluster 2 Centroid:
  Topic_0: 0.0294   Topic_1: 0.0270   Topic_2: 0.9041   Topic_3: 0.0060   Topic_4: 0.0335
------------------------------
Cluster 3 Centroid:
  Topic_0: 0.9662   Topic_1: 0.0310   Topic_2: 0.0009   Topic_3: 0.0009   Topic_4: 0.0009
------------------------------
Cluster 4 Centroid:
  Topic_0: 0.0008   Topic_1: 0.0176   Topic_2: 0.0008   Topic_3: 0.9098   Topic_4: 0.0710
```

Fig. 8. Calculation of the centroid for each cluster.

The cosine similarity between each centroid vector and each query embedding was then calculated, retrieving the cluster that has the highest similarity to represent the most relevant context per query, as shown in Figure 9(a). Figure 9(b) refers to each query with its relevant cluster and describes the total number of documents that are grouped according to the LDA distributions within this cluster.

(a)

```
Query 1: "dan parris"
Cluster 0: Cosine Similarity = 0.9891 (matched query)
-----------------------------------------------------------
Query 2: "rob lehr"
Cluster 0: Cosine Similarity = 0.9890 (matched query)
-----------------------------------------------------------
Query 3: "el salvador"
Cluster 1: Cosine Similarity = 0.9879 (matched query)
```

(b)

```
Query 1: "dan parris"
Cluster 0: Cosine Similarity = 0.9891 (matched query)
  Cluster 0 — Total Articles: 22
Article (preview): nairobi, kenya (cnn) -- two u.s. filmmakers were injured saturday ...
Article (preview): ( cnn ) -- while all eyes seem to be on `` slumdog millionaire "or ...
Article (preview): samuel yaga, a mechanic , was making a routine repair on a client ...
Article (preview): ( cnn ) -- world no.1 novak djokovic ended up on the losing side ...
Article (preview): ( cnn ) locked inside the gemstone there appears to be a human ..
Article (preview): ( cnn ) -- the northern pakistani school where a teenage boy ...
Article (preview): yelena isinbayeva , the golden girl of russian sport , has ...
Article (preview): ( cnn ) -- call david x. cohen a nerd all you like , but never ...
Article (preview): jerusalem ( cnn ) -- israeli soldiers and several dozen palestinians ...
Article (preview): ( cnn ) -- efforts to remove cats from macquarie island ,
```

Fig. 9. Retrieving the most relevant cluster for each query.

The DRL model, dependent on the DQN architecture with a target network and replay memory, was trained to choose or skip sentences from the relevance cluster. The reward function was built to balance various factors: sentence diversity (novelty), semantic relevance, and coherence calculated using a BERT-based coherence model, as shown in Figure 10. Table I shows a comparative evaluation of the proposed framework and various established benchmark approaches reported in the literature. The results of the proposed model were obtained by experimental implementation, while the baseline scores of other methods refer to their original publications. All models were evaluated using ROUGE-1, ROUGE-2, and ROUGE-L metrics, which calculate the overlap of unigrams, bigrams, and longest popular subsequences between the created and reference summaries. These results show that the proposed model consistently outperformed traditional approaches, where high ROUGE values indicate that the summaries generated by the proposed model preserve a readable structure, capture the main information (coverage), and align well with the reference or query (relevance) under comparable settings to ensure a fair performance comparison.

```
Training on Query: franklin Roosevelt... | Cluster: 2 | Sentences in cluster: 1374
Query: franklin Roosevelt
Action: Select sentence index (local to cluster): 706
Selected Sentence: e-mail to a friend
State vector (first 10): [-0.01636478 0.04401933 -0.05003017 -0.05332302 0.0104168
0.02764818
 0.04463525 0.0418681 -0.02328707 0.05508748]
Next state vector (first 10): [-0.01636478 0.04401933 -0.05003017 -0.05332302 0.0104168
0.02764818
 0.04463525 0.0418681 -0.02328707 0.05508748]
Reward: 0.5374359741757544
Accumulated Reward: 0.5374359741757544
Query: franklin Roosevelt
Action: Select sentence index (local to cluster): 1043
```

Fig. 10. Training sample.

TABLE I. COMPARISON OF THE PROPOSED SYSTEM WITH BASELINE APPROACHES

Approaches	R1	R2	R-L
Proposed model (DRL+LDA+K-means)	50.03	27.30	39.86
Dynamic Clustering and Co-Reference on BERT [11]	41.4	17.9	39.9
Reinforcement Learning and Transformer Model [9]	7.9	9.1	8.5
K-means, Centroid-based Method, MMR, and Sentence Position [14]	40.93	9.53	37.05
DRL(CNN-RNN) DRL(RNN-RNN) [27]	39.3 39.4	15.8 16.1	/

The proposed model utilizes the BERT score to assess coherence by calculating the semantic similarity between the generated and reference summaries based on the embedding. In the reward function, the implementation of the BERT-based coherence model contributes to satisfying the logical cohesion and sentence transitions within the generated summaries. Diversity was validated by utilizing MMR, which rewards novelty while penalizing redundancy in sentence selection. Furthermore, when comparing sentence similarity scores, redundancy reduction indicates a lower redundancy rate and higher diversity.

As shown in Table II, ROUGE-2, ROUGE-L, and ROUGE-1 scores were enhanced significantly when BERT-based coherence and MMR were integrated in the reward function, with a BERT score of 88%, ROUGE-1 of 50%, ROUGE-2 of 27%, and ROUGE-L of 39%. This indicates that the summaries created by the trained agent not only involve relevant information but also reduce redundancy and preserve semantic context. These results highlight the importance of combining both diversity-aware selection mechanisms and coherence modeling in the reward function. Visualizations of training progression, such as sentence chosen positions and the mean reward for each episode, show that an agent learned to prefer top sentences that are semantically contextual and coherent. In general, the proposed system successfully integrates BERT-based coherence, unsupervised topic modeling, and DRL in a reward function with multiple factors to create diverse, relevant, concise summaries for user queries. The improvement over classical methods in both BERT and ROUGE scores, along with readable, coherent summaries, demonstrates the effectiveness of the proposed framework.

TABLE II. RESULTS OF THE SUGGESTED SYSTEM COMPARED TO REFERENCE SUMMARIES

Queries	Generated summaries	Reference summaries	ROUGE			BERT similarity score
			R1	R2	R-L	
a Chinese prison	A high-level ministerial briefing exclusively obtained by Daily Mail Australia reveals an 'unprecedented' number of Australians could face the death penalty in China for drug smuggling. The revelation follows widespread public anger at ………..	The number and seriousness of Australians facing the death penalty in China is 'unprecedented', that's according to a high-level ministerial briefing obtained by Daily Mail Australia under freedom of information laws. Many Australians arrested were for the drug 'ice'. As many as ……….	39.7	19.5	26.8	86.4
a further 11 dentists	Britain's five highest-earning NHS dentists are paid average salaries of £690,000 a year. The sum is almost five times the Prime Minister’s £142,500 pay, thousands of …..	Highest-earning NHS dentists earn almost five times more than the Prime Minister’s £142,500 pay packet. A further 11 dentists were paid between £400,000 and £500,000 a year. Lay bare huge ……	60.50	42.74	47.06	90.13%
three women	Christopher Bridger, 25, from Stevenage, Hertfordshire, attacked three women after separate drinking sessions and was jailed for 12 ….	Christopher Bridger, 25, attacked three women after drinking sessions. An ambulance worker told women he was gay before assaulting them in bed. …….	48.65	21.92	39.19	88.76%
shooters	DARPA says the smart bullets will also help shooters who are trying, for example, to hit targets in high winds. The goals of the EXACTO are the safety of American troops, DARPA said.	50-caliber bullets equipped with optical sensors can follow moving targets. The "smart bullets" can help shooters shooters a greater range and make American troops safer.	51.22	25.00	46.34	89.44%

V. ABLATION STUDY

An ablation study was performed to assess the contribution of each technique in the proposed model. This approach systematically replaces or isolates one component at a time, maintaining the remainder unchanged, to measure the effect of each component (LDA, K-means, MMR, and BERT-based Coherence) on the performance of the summarization framework. The results in Table III show that all components contribute positively to the results of the proposed system. Specifically, MMR plays a major role in increasing diversity and reducing redundancy, and BERT-based coherence ensures logical sentence flow and overall readability. Removing topic modeling components (LDA and K-Means) negatively impacts the coverage and relevance of generated summaries, highlighting the significance of topic-aware clustering. These results emphasize that the combination of all components leads to the best summarization performance.

TABLE III. ABLATION STUDY RESULTS FOR DIFFERENT MODEL VARIANTS USING ROUGE AND BERT SCORE

Model variants	R1	R2	R-L	BERT similarity score
Proposed model (DRL+LDA+K-means)	50.03	27.30	39.86	86.4
Proposed model without LDA+K-means	48.22	20.9	38.9	85.5
Proposed model without MMR	44.9	17.1	28.5	83.5
Proposed model without BERT-based coherence	41.44	13.53	37.05	80.2

VI. CONCLUSION

This study presented a comprehensive extractive query-based summarization system that combines a clustering algorithm, topic modeling, and BERT for embedding. This system efficiently arranges documents based on topic distributions in coherent clusters by leveraging K-means and LDA. Representing both sentences and queries semantically by BERT embeddings helps to accurately retrieve the most relevant cluster for each query after calculating the cosine similarity with the centroid for each cluster. The reward function integrates BERT-based coherence with MMR, guiding the DQN-based summarization agent and training it to choose sentences that best match the query context. The results indicate that this model creates more coherent, less redundant, and more relevant summaries than baseline extractive approaches. The combination of multiple reward considerations and the use of topic-aware clustering highly improves the model's ability to focus on contextually appropriate content, resulting in enhanced performance during evaluation measures.

VII. LIMITATIONS AND FUTURE WORK

Although the proposed system demonstrated powerful performance, it suffers from several limitations, such as addressing ambiguous and very short queries, where the lack of query context makes it hard to identify the most relevant sentences or clusters, limiting scalability. Future work will investigate query expansion approaches to better support vague or short queries.

REFERENCES

[1] D. H. Nguyen *et al.*, "Robust Deep Reinforcement Learning for Extractive Legal Summarization," in *Neural Information Processing*, vol. 1517, T. Mantoro, M. Lee, M. A. Ayu, K. W. Wong, and A. N. Hidayanto, Eds. Springer International Publishing, 2021, pp. 597–604.

[2] M. Zhong, P. Liu, Y. Chen, D. Wang, X. Qiu, and X. Huang, "Extractive Summarization as Text Matching." arXiv, Apr. 19, 2020, https://doi.org/10.48550/arXiv.2004.08795.

[3] L. Abualigah, M. Q. Bashabsheh, H. Alabool, and M. Shehab, "Text Summarization: A Brief Review," in *Recent Advances in NLP: The Case of Arabic Language*, vol. 874, M. Abd Elaziz, M. A. A. Al-qaness, A. A. Ewees, and A. Dahou, Eds. Springer International Publishing, 2020, pp. 1–15.

[4] Z. H. Ali, A. K. Hussein, H. K. Abass, and E. Fadel, "Extractive multi document summarization using harmony search algorithm," *TELKOMNIKA (Telecommunication Computing Electronics and*

Control), vol. 19, no. 1, Feb. 2021, Art. no. 89, https://doi.org/10.12928/telkomnika.v19i1.15766.

[5] M. S. Bewoor and S. H. Patil, "Empirical Analysis of Single and Multi Document Summarization using Clustering Algorithms," *Engineering, Technology & Applied Science Research*, vol. 8, no. 1, pp. 2562–2567, Feb. 2018, https://doi.org/10.48084/etasr.1775.

[6] S. R. Basha, J. K. Rani, and J. J. C. P. Yadav, "A Novel Summarization-based Approach for Feature Reduction Enhancing Text Classification Accuracy," *Engineering, Technology & Applied Science Research*, vol. 9, no. 6, pp. 5001–5005, Dec. 2019, https://doi.org/10.48084/etasr.3173.

[7] N. Dutta, "Topic Modelling With LDA -A Hands-on Introduction," *Analytics Vidhya*, Jul. 08, 2021. https://www.analyticsvidhya.com/blog/2021/07/topic-modelling-with-lda-a-hands-on-introduction/.

[8] F. Tan, P. Yan, and X. Guan, "Deep Reinforcement Learning: From Q-Learning to Deep Q-Learning," in *Neural Information Processing*, vol. 10637, D. Liu, S. Xie, Y. Li, D. Zhao, and E. S. M. El-Alfy, Eds. Springer International Publishing, 2017, pp. 475–483.

[9] A. Mahmud, "Query-Based Summarization using Reinforcement Learning and Transformer Model," M.S. thesis, University of Lethbridge, Canada, 2020.

[10] Y. Mao, Y. Qu, Y. Xie, X. Ren, and J. Han, "Multi-document Summarization with Maximal Marginal Relevance-guided Reinforcement Learning." arXiv, Sep. 30, 2020, https://doi.org/10.48550/arXiv.2010.00117.

[11] A. Srikanth, A. S. Umasankar, S. Thanu, and S. J. Nirmala, "Extractive Text Summarization using Dynamic Clustering and Co-Reference on BERT," in *2020 5th International Conference on Computing, Communication and Security (ICCCS)*, Patna, India, Oct. 2020, pp. 1–5, https://doi.org/10.1109/ICCCS49678.2020.9277220.

[12] N. Rahman and B. Borah, "Improvement of query-based text summarization using word sense disambiguation," *Complex & Intelligent Systems*, vol. 6, no. 1, pp. 75–85, Apr. 2020, https://doi.org/10.1007/s40747-019-0115-2.

[13] F. Bayatmakou, A. Mohebi, and A. Ahmadi, "An interactive query-based approach for summarizing scientific documents," *Information Discovery and Delivery*, vol. 50, no. 2, pp. 176–191, Apr. 2022, https://doi.org/10.1108/IDD-10-2020-0124.

[14] S. Lamsiyah, A. El Mahdaouy, B. Espinasse, and S. El Alaoui Ouatik, "An unsupervised method for extractive multi-document summarization based on centroid approach and sentence embeddings," *Expert Systems with Applications*, vol. 167, Apr. 2021, Art. no. 114152, https://doi.org/10.1016/j.eswa.2020.114152.

[15] O. Shapira, R. Pasunuru, M. Bansal, I. Dagan, and Y. Amsterdamer, "Interactive Query-Assisted Summarization via Deep Reinforcement Learning," in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Seattle, WA, USA Apr. 2022, pp. 2551–2568, https://doi.org/10.18653/v1/2022.naacl-main.184.

[16] R. C. Belwal, S. Rai, and A. Gupta, "Extractive text summarization using clustering-based topic modeling," *Soft Computing*, vol. 27, no. 7, pp. 3965–3982, Apr. 2023, https://doi.org/10.1007/s00500-022-07534-6.

[17] C. Zhou *et al.*, "A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT." arXiv, May 01, 2023, https://doi.org/10.48550/arXiv.2302.09419.

[18] M. Ramezani, M. S. Shahryari, A. R. Feizi-Derakhshi, and M. R. Feizi-Derakhshi, "Unsupervised Broadcast News Summarization; a Comparative Study on Maximal Marginal Relevance (MMR) and Latent Semantic Analysis (LSA)," in *2023 28th International Computer Conference, Computer Society of Iran (CSICC)*, Tehran, Iran, Jan. 2023, pp. 1–7, https://doi.org/10.1109/CSICC58665.2023.10105403.

[19] R. Padaki, Z. Dai, and J. Callan, "Rethinking Query Expansion for BERT Reranking," in *Advances in Information Retrieval*, vol. 12036, J. M. Jose, E. Yilmaz, J. Magalhães, P. Castells, N. Ferro, M. J. Silva, and F. Martins, Eds. Springer International Publishing, 2020, pp. 297–304.

[20] Z. Zheng, K. Hui, B. He, X. Han, L. Sun, and A. Yates, "BERT-QE: Contextualized Query Expansion for Document Re-ranking." arXiv, Nov. 03, 2020, https://doi.org/10.48550/arXiv.2009.07258.

[21] V. H. Nguyen, S. T. Mai, and M. T. Nguyen, "Learning to summarize multi-documents with local and global information," *Progress in Artificial Intelligence*, vol. 12, no. 3, pp. 275–286, Sep. 2023, https://doi.org/10.1007/s13748-023-00302-z.

[22] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Minneapolis, MN, USA, Mar. 2019, pp. 4171–4186, https://doi.org/10.18653/v1/N19-1423.

[23] W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. Mohamed, "Automatic text summarization: A comprehensive survey," *Expert Systems with Applications*, vol. 165, Mar. 2021, Art. no. 113679, https://doi.org/10.1016/j.eswa.2020.113679.

[24] M. Akter, N. Bansal, and S. K. Karmaker, "Revisiting Automatic Evaluation of Extractive Summarization Task: Can We Do Better than ROUGE?," in *Findings of the Association for Computational Linguistics: ACL 2022*, Dublin, Ireland, Feb. 2022, pp. 1547–1560, https://doi.org/10.18653/v1/2022.findings-acl.122.

[25] P. Gupta, "Evaluating the BERTScore of synthetic text and its sentiment analysis." In Review, Aug. 16, 2023, https://doi.org/10.21203/rs.3.rs-3248507/v1.

[26] Y. Liu, Z. Wang, and R. Yuan, "QuerySum: A Multi-Document Query-Focused Summarization Dataset Augmented with Similar Query Clusters," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 17, pp. 18725–18732, Mar. 2024, https://doi.org/10.1609/aaai.v38i17.29836.

[27] K. Yao, L. Zhang, T. Luo, and Y. Wu, "Deep reinforcement learning for extractive document summarization," *Neurocomputing*, vol. 284, pp. 52–62, Apr. 2018, https://doi.org/10.1016/j.neucom.2018.01.020.