Towards More Effective Automatic Question Generation: A Hybrid Approach for Extracting Informative Sentences

Engy Yehia[1], Neama Hassan[2], Sayed AbdelGaber[3]
Department of Information Systems-Faculty of Commerce and Business Administration, Helwan University, Cairo, Egypt[1, 2]
Department of Information Systems-Faculty of Computers and Artificial Intelligence, Helwan University, Cairo, Egypt[3]

***Abstract*—Informative Sentence Extraction (ISE) is one of the crucial components in Automatic Question Generation (AQG) and directly influences the quality and relevancy of the generated questions. Instructional texts often contain not only informative but also irrelevant sentences. This results in the creation of poor-quality or distorted questions when irrelevant, non-informative sentences have been used as input. Therefore, the basic problem discussed in this paper is how to provide a systematic method for filtering out such sentences and retaining those that are pedagogically valuable. The purpose of ISE is to filter out irrelevant, low-quality information and retain only the factually dense sentences, express key concepts and are contextually significant. This paper proposes a hybrid approach for extracting informative sentences that combines lexical, statistical, and semantic criteria to identify informative sentences suitable for generating educational questions. The proposed approach consists of two modules: the first module employs four techniques in order to evaluate the informativeness of sentences, which are the keyword-based scoring, Named Entity Recognition (NER), information gain (IG) and Sentence-BERT (SBERT). The second module utilizes multiple fusion strategies to integrate the results derived from the informative sentence extraction techniques. The preprocessed sentences extracted from educational materials were ranked and filtered based on their informativeness coverage. The evaluation results indicate that the hybrid approach can improve the extraction of informative sentences rather than using individual methods. Such a contribution is important for enhancing the performance of downstream tasks in AQG systems, such as distractor generation and question formulation.**

Keywords*—Automatic Question Generation (AQG); informative sentence extraction; NER; SBERT; question answering; information gain; fusion strategies

I. Introduction

Automatic Question Generation (AQG) has attracted considerable interest in the field of educational technology and has provided scalable assessment, intelligent tutoring and personalized learning environment applications [1] [2]. An important part of any functional AQG system is the task of Informative Sentence Extraction (ISE). It entails the process that determines textual parts that are semantically rich, pedagogically valuable, and contextually relevant [3]. These sentences are the main component upon which meaningful and high-quality questions are constructed. Improper or inaccurate selection of sentences may decrease the level of coherence and relevancy of generated questions [4], [5].

The main issue addressed in this paper is that the sentence extraction step remains a relatively unexplored bottleneck in AQG pipelines, despite significant advances in natural language generation and deep learning methods. While recent progress in neural language models has substantially improved the fluency, grammaticality, and contextual relevance of generated questions, much less attention has been given to identifying which sentences within instructional material are most informative and pedagogically valuable as question sources. Most current systems use heuristic rules [6] or deep neural networks in an end-to-end methodology where sentence selection is an implicit or black-box procedure. There are possibilities of including redundant, indefinite, and insignificant sentences that adversely influence the importance and quality of the generated questions [4], [7].

The current paper attempts to mitigate this problem by introducing a hybrid approach of informative sentence retrieval that combines shallow linguistic features with deeper semantics. Our approach combines: lexical relevance scoring based on the coverage of educationally relevant keywords and key phrases, Named Entity Recognition (NER) to identify domain-relevant concepts, proper nouns, and entities at the center of instructional content, Information Gain (IG) computations quantifying the discriminative power of each sentence within the informational landscape of a document, and semantic encoding with sentence-Bert [8] that captures word context and relatedness. We implement our methodology on educational texts within information system academic disciplines. The findings demonstrate that the system outperforms existing heuristic or embedding-only baselines in both the quality of sentence selection and downstream performance in question generation. The proposed approach has been optimized to perform effectively with low-resource educational materials, such as PDF textbooks or lecture notes.

The structure of this paper is organized as follows: Section II summarizes the background and relevant studies on informative sentence extraction. Section III presents a hybrid framework for extracting informative sentences in automatic question generation systems. Section IV discusses the experimental setup, results, and evaluation, while Section V discusses the evaluation results. Section VI concludes the paper and outlines potential directions for future research.

II. Background and Related Work

A. Background

The extraction of the informative sentences is one of the most applicable processes in the context of natural language processing that could apply in the education field in an automatic question generator (AQG), answer selection, and educational content summarization. Traditional summarization strategies focus only on conciseness and identifying key components [9], [10], which are insufficient for educational question generation, where these sentences must be both informative and question-worthy. In AQG, it is essential to select sentences that are both semantically dense and academically rich. Traditional methodologies, including term frequency-inverse document frequency (TF-IDF) and Latent Dirichlet Allocation (LDA), were employed to extract significant textual components [11].

Named entity recognition (NER) has been used to provide a semantic aspect to text, enabling the extraction of sentences that may contain information worthy of questioning [12]. However, traditional approaches are frequently unable to provide sufficient information regarding required contextual knowledge of the learning resources. Recent developments in transformer-based models, such as BERT and Sentence-BERT, significantly enhance semantically sensitive representations, enabling reliable meaning extraction across sentences and facilitating deeper filtering, clustering, and classification [8].

Researchers have proven that phrases defining and explaining causal relationships and structured processes form the basis of high-quality educational questions [2], [13]. Effective sentence extraction has to be consistent with learning objectives, as outlined by Bloom's Taxonomy, which assesses students across many cognitive skill levels, from simple recall to analytical reasoning [14]. Consequently, sentence extraction systems have to satisfy cognitive requirements to select not only syntactically perfect but also semantically valuable sentences. Traditional approaches employed shallow principles, such as sentence location or keyword presence [15], whereas current methods use pretrained transformers to evaluate sentence significance [16]. However, neither of these approaches is sufficient independently, especially for domain-specific contexts where linguistic indicators can greatly vary.

B. Related Works

The extraction of informative sentences is a concern addressed from various perspectives, including general text summarization [17], [18], [19], [20], [21] and specifically opinion summarization, where users need concise yet detailed insights [22], [23]. In information retrieval, the selection of informative sentences is crucial for improving information access efficiency and enhancing the retrieval process, particularly for domain-specific contexts, by applying models that improve relevance and reduce redundancy [24], [25]. Customer requirements analysis is the most recent domain under investigation in informative sentence extraction research.

In [26], researchers employed Transformer networks to identify informative sentences that mirror customer needs in user-generated content. However, it does not clearly address Automatic Question Generation (AQG) in its study or findings.

In the context of automatic question generation systems (AQG), various methodologies for the selection of informative sentences have been proposed. Early methods proposed rule-based and stylized filtering heuristics, such as sentence length, the presence of proper names, and POS tag sequences, to recognized sentences that are worthy of being turned into questions in early AQG systems [27]. Although effective to a certain degree, these techniques are characterized by limited generality and insufficient semantic knowledge.

In this study [28], sentences are considered informative if they contain at least one term derived from specified important concepts for generating multiple-choice test items from electronic documents.

In [29], the author combined a set of criteria, including the number of tokens, the number of clauses, the probabilistic context-free grammar score, and well-defined context. Then manually calculate the score of sentences depending on the occurrence of these criteria; if the score is greater than a threshold, select the sentence as informative.

Another study employed candidate rules to identify phrases with a certain form, such as sentences containing definitions, to create a specific type of question [30]. Additional research used machine learning approaches such as the Support Vector Machine (SVM) classifier based on some features such as sentence length, verb domain, named entity, parts-of-speech, chunk, word position, known-unknown word and acronym [31].

Majumder and Kumar suggested an approach based on parse structure similarity and a rule-based technique to identify informative sentences for multiple-choice questions. The algorithm compared the testing content to a pre-compiled reference set of parse structures that are associated with existing multiple-choice questions (MCQs) and then selected the final informative sentences through rule-based post-processing. This method is considered domain-specific, as it requires a set of existing MCQs within the same domain to create the reference set, which is not easily available [32].

In [33], the author proposed a method to identify informative micro aspects in news texts which can be used to generate relevant questions based on the extracted informative aspects by applying semantic role labelling, named-entity recognition, handcrafted rules and machine learning techniques.

A number of investigations focus on employing sentence structure analysis, dependency parsing, and named entity recognition to discern informative target sentences and concepts for question generation from provided sentences [34], [35], [36], [37].

Several studies employed neural models to discover informative phrases, such as [38], which implemented a two-stage system: neural key-phrase identification followed by sequence-to-sequence question creation.

Previous research on informative sentence extraction has investigated various approaches, each with specific methodological limitations as presented in Table I. Rule-based and heuristic approaches, dependent on basic features like

sentence length, keyword presence, or part-of-speech (POS) patterns, indicate restricted semantic depth, inadequate abstraction, and limited domain relevance [27] [30].

TABLE I. SUMMARY OF RELATED WORK ON INFORMATIVE SENTENCE EXTRACTION

Study	Approach	Technique	Key Limitations
[27], [30]	Rule-based and Heuristic	Sentence length, keyword presence, and POS patterns	Limited semantic depth, poor generalization, narrow coverage
[28], [29]	Concept-based Scoring	Predefined concept lists, token count, clause count, and PCFG scores	Manual curation, brittle to unseen terms
[31]	Feature-based ML	SVM classifier with lexical, and syntactic features (length, POS, position…)	Requires annotated training data, extensive feature engineering, limited context capture
[32]	Parse-Structure Matching	Matching parse trees to MCQ reference set	Needs pre-compiled MCQs, Domain-specific, not widely applicable
[33]–[37]	Hybrid Linguistic and ML	SRL, NER, handcrafted rules, and ML models	Computationally intensive, rule-heavy, Complex pipeline
[38]	Neural Models	Transformer-based key phrase, and sequence-to-sequence generation	Requires large training datasets, less interpretable

Concept-based scoring approaches, such as those employing predefined concept lists, token frequencies, or probabilistic context-free grammar (PCFG) scores, involve considerable manual curation and exhibit instability when faced with unique or domain-specific terminology [28] [29]. Feature-based machine learning methodologies, such as support vector machines (SVMs), which use lexical and syntactic features, require annotated datasets and extensive feature engineering while still capturing only a limited amount of contextual information [31]. Parse-structure matching techniques, which correlate sentence parse trees with multiple-choice question (MCQ) references, reveal significant domain dependency, requiring pre-compiled MCQs and presenting constrained scalability [32].

Hybrid linguistic and machine learning pipelines that incorporate semantic role labelling (SRL), NER, manually generated rules, and classifiers frequently demonstrate high processing costs, a large number of rules, and overly complicated structures [33], [34], [35], [36], [37]. Recently, neural methodologies, such as transformer-based key phrase extraction and sequence-to-sequence creation, have demonstrated potential but depend on extensive training datasets and encounter interpretability challenges [38].

Our work provides a hybrid framework, described as domain-adaptive, that combines keyword scoring, named entity recognition (NER), information gain, and SBERT similarity with alternative fusion strategies to ensure robust sentence selection, even when using low-resource educational material. It bridges the gap between shallow heuristics and complex neural systems while keeping interpretability and practical deployability, which addresses the needs of Automatic Question Generation (AQG) systems in the educational environment.

III. METHODOLOGY

The extraction of informative sentences is a significant NLP technique, with particular application in Automatic Question Generation (AQG), answer selection, and content summarization. Informative sentence extraction (ISE) is an essential part of our comprehensive AQG smart model, which is currently under development. During the development of our AQG smart model, we found that the output of informative sentence extraction significantly impacts the quality of the generated questions; therefore, we conducted this research to provide hybrid approaches aimed at optimizing the quality of informative sentence extraction (ISE). This section outlines the suggested methodology, a multi-step hybrid technique designed for the identification of informative sentences specifically for automatic question generation (AQG). This methodology incorporates various supplemental techniques: keyword-based scoring, named entity recognition (NER), information gain (IG) scoring, and contextual semantic filtering using Sentence-BERT (SBERT). Each technique assists in evaluating sentence informativeness from different linguistic and semantic perspectives. Following that, several fusion scoring methods have been used to combine the informativeness scores derived from the four implemented methods, therefore generating an aggregated score to rank sentence informativeness. The hybrid ISE approach includes two modules: Sentence Informativeness Scoring and Fusion Scoring Strategies, as seen in Fig. 1.

The primary input for the hybrid ISE approach includes a database of simple phrases derived from unstructured instructional resources that were preprocessed by the preprocessing module of our AQG smart model before being used in this research. The subsequent subsections present a detailed description of the hybrid ISE approach.

A. Sentence Informativness Scoring

The Sentence Informativeness Scoring Module is designed to evaluate and filter informative content based on its instructional value. This module takes as its input a learning material database composed of preprocessed simple sentences. The output is a refined Sentence Knowledge Database, consisting exclusively of sentences identified as informative, conceptually substantial, and suitable for AQG downstream educational application.

1) Keyword-based scoring: This approach evaluates sentence informativeness through a keyword-based implication to align the text with the instructional domain. A set of domain-specific terms such as DSS, information systems, software, and hardware has been manually selected to cover critical areas of the learning material. Every sentence is evaluated by phrase-level matching to detect both single-word keywords and multi-word phrases in this list. Sentences containing one or more matching words are designated as informative and assigned a keyword score, which corresponds to the quantity of matched terms. The key concepts of this methodology originate from prior studies on automatic question generation, wherein

manually generated keywords effectively assist the selection of key content [39]. Keyword existence is considered one of the heuristic feature-based sentence selections that have been frequently used in educational question generation systems [40], [41].

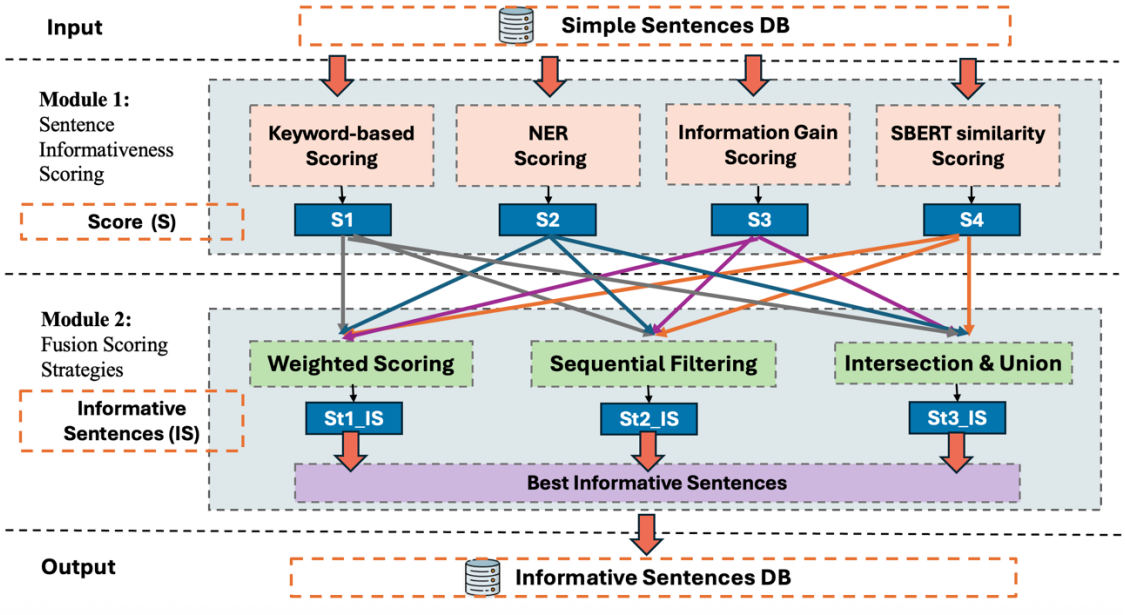

Fig. 1. The proposed hybrid ISE approach.

2) Named Entity Recognition (NER) scoring: Named entity recognition (NER) is a subset of computer science and natural language processing (NLP) that is more directly related to the identification and classification of entities in unstructured text into categories established in advance, such as persons, geographical regions, and organizations [42]. Entity recognition is the process of extraction, disambiguation and connectivity of an entity in a raw text to some useful and structured knowledge bases [43]. Such emphasis allows the generation of more relevant and logical questions adapted to the content [44]. In this approach, the NER technique is used to identify semantically rich sentences using factual and contextual entities. This implies that sentences that refer to individuals, companies, places, or times are most likely to express evaluative information and may be used in the question generation task [45], [46].

SpaCy's lightweight English language model (en_core_web_sm) is used to extract named entities from each sentence in the input corpus. Then the model parses every sentence to identify the entities that exist in it and assign them to the existing types of entities (e.g., PERSON, ORG, DATE, GPE, etc.) [47]. All sentences including at least one named entity are retained. The NER score for each retained sentence is the sum of the number of discovered entities found within it.

3) Information gain scoring: In data and text mining, Information Gain (IG) is a statistical tool that evaluates a term's or word's importance inside a given text. This metric assesses a term's prominence inside a document as well as the likelihood that it falls into a particular category [48]. The provided method is based on the information-theoretic measure of the informativeness of sentences calculated as IG scores. The basic idea is that statements containing statistically significant or relevant terms will serve as more effective question generators, as they are likely to capture fundamental concepts and differences [49]. Text processing starts with tokenizing the simplified sentences corpus into a single word and the identification of the global frequency distribution of all information. In contrast to frequency-based scoring, the IG method utilized here depends on entropy. The reason for the distinction is that words able to reduce the majority of ambiguity, according to the corpus, are considered to represent higher informative quality. Information gain is computed as shown in Eq. (1), (2), and (3).

$$IG(w) = H(C) - H(C|w) \quad (1)$$

$$H(C) = -\sum_{x \in V} P(x)\ log_2 P\ (x) \quad (2)$$

$$IG(s) = \frac{1}{|s|}\sum_{w \in s} IG(w) \quad (3)$$

where, *H(C)* denotes the entropy of the full corpus of sentences C. It is computed as the negative sum over all possible vocabulary labels $x \in V$, where V is the set of vocabulary and $P(x)$ is the probability of observing vocabulary x. The conditional entropy, $H(C|w)$, represents the remaining uncertainty in corpus given the presence/absence of word w. $IG(w)$ is the reduction in entropy achieved by conditioning on that word. A sentence s is considered as a sequence of words $\{w1,w2,...,w|s|\}$, where $|s|$ denotes the number of words in the sentence. The information gain of the sentence, $IG(s)$, is calculated as the average of the information gain values of its constituent words. Algorithm 1 presents the IG scoring steps. This approach was implemented by the tokenisation with the Natural Language Toolkit (NLTK) and an entropy-based

model applied to obtain the IG scores [50]. Sentences are then ranked in a descending order according to their IG scores.

Algorithm 1: Information Gain (IG) Scoring for Sentence Informativeness

Input: A corpus C of sentences

Output: A list of sentence-level information gain scores $IG(s_i)$ for each sentence $si \in C$

Step 1: Tokenize all sentences and compute global word frequencies $f(w_j)$.

Step 2: Calculate corpus entropy $H(C)$.

Step 3: For each word w_j, compute conditional entropy $H(C|w_j)$.

Step 4: Compute word-level information gain $IG(w_j)$.

Step 5: For each sentence s_i, compute sentence-level score $IG(s_i)$ as the average $IG(w_j)$ of its words.

Step 6: Return all sentence scores $\{IG(s_i)\}$.

4) SBERT-based semantic relevance scoring: Sentence-BERT (SBERT) is a variant of BERT, which represents semantically informed sentence embeddings that enable efficient similarity calculations and thus dramatically lower the computation costs [8]. Large pretrained language models, such as BERT and Sentence BERT, provide effective sentence embeddings that are highly correlated with human similarity ratings [51]. SBERT was suggested to learn sentence representation only by performing the computation on a single query question rather than on pairs of sentences. This model is effective for semantic textual similarity (STS) [52]. To improve the informativeness rank, we used Sentence-BERT (SBERT) as a semantic representation method.

SBERT, unlike other shallow linguistic features such as word frequency (count) or keyword matching, SBERT enables us to encode once and for all our sentences into a dense vector that encapsulates the context in which it appears. This enables semantic comparison, beyond simple lexical overlap, which proves particularly useful in educational texts [8]. The algorithm is performed in multiple steps. First, a pre-trained SBERT was used to encode the fixed-length embedding of each sentence of the input corpus. Second, the average of all embeddings was used to compute a semantic centroid vector, which would represent a general semantic space of the instructional material. Third, cosine similarity was used to estimate a scalar informativeness score between each sentence embedding and the centroid as shown in Eq. (4), (5), and (6).

$$\mathrm{sim}(s_i, s_j) = \frac{v_i \cdot v_j}{|v_i||v_j|} \quad (4)$$

$$c_i^* = \frac{1}{n-1} \sum_{j \neq i} \mathrm{sim}(s_i, s_j) \quad (5)$$

$$c_i = \frac{c_i^* - \min(C^*)}{\max(C^*) - \min(C^*)} \quad (6)$$

Where the raw centrality score of a sentence, denoted by c_i^*, is the average semantic similarity of sentence s_i. c_i is the normalized centrality score. The collection of all raw centrality scores across the corpus is represented by $C^*=\{c_1^*, c_2^*, \ldots, c_n^*\}$, with $\min(C^*)$ and $\max(C^*)$ denoting the minimum and maximum values within this set, respectively, for use in the normalization step. The pairwise similarity between two sentences, $sim(s_i, s_j)$, is computed as the cosine similarity between their embeddings. v_i and v_j represent the dense vector embeddings of sentences s_i and s_j generated using a pre-trained SBERT model. The Euclidean norms of these vectors, $|v_i|$ and $|v_j|$, are used to scale the cosine similarity, while the dot product $v_i \cdot v_j$ constitutes its numerator. Algorithm 2 presents the SBERT-Based Semantic Centrality Scoring steps.

Algorithm 2: SBERT-Based Semantic Centrality Scoring

Input: A set of candidate sentences $S=\{s_1, s_2, \ldots, s_n\}$.

Output: A vector of normalized semantic centrality scores $C=\{c_1, c_2, \ldots, c_n\}$.

Step 1: Encode each sentence $s_i \in S$ into a dense vector representation v_i using a pre-trained Sentence-BERT (SBERT) model.

Step 2: Compute the pairwise semantic similarity $sim(s_i, s_j)$ between all sentence embeddings.

Step 3: For each sentence s_i, calculate its raw centrality score c_i^*, as the average similarity with all other sentences in S.

Step 4: Normalize all raw centrality scores to the range $[0,1]$.

Step 5: Return the normalized semantic centrality scores $C=\{c_1, c_2, \ldots, c_n\}$.

Sentences having higher scores are expected to have more central and representative information, while sentences with lower scores might be less informative. The result of this step is a ranked list of sentences, and each has a numerical value corresponding to the degree of semantic proximity to the distribution of the overall content. This ranked output is further utilized in the downstream level of the fusion scoring strategies to choose informative sentences.

B. Fusion Scoring Strategies for Informative Sentence

To exploit the complementary properties of the individual scoring techniques (NER, keyword-based scoring, IG, and SBERT), we suggested and evaluated three distinct fusion strategies. Each strategy calculates the final informativeness score by integrating the outputs of the primary methods in different ways, thereby providing varied trade-offs among precision, recall, and semantic coverage.

1) Weighted scoring: In this strategy, each of the four methods is used as one complementary signal within a unified scoring model. Each score is normalised in the range [0, 1]. Both NER and keyword scores are taken as binary, while IG and SBERT scores are taken as continuous. The final score is calculated as:

$$\mathrm{FinalScore}(s) = w_{\mathrm{NER}} \cdot \mathrm{NER}(s) + w_{\mathrm{KW}} \cdot \mathrm{Keywords}(s) + w_{\mathrm{IG}} \cdot \mathrm{IG}(s) + w_{\mathrm{SBERT}} \cdot \mathrm{SBERT}(s) \quad (7)$$

where, $w_{\mathrm{NER}} + w_{\mathrm{KW}} + w_{\mathrm{IG}} + w_{\mathrm{SBERT}} = 1$

This strategy depends on the distribution of weight to make a balance between precision and recall. For example, a higher weight on SBERT prioritizes semantic richness, while increasing the weight on NER ensures factual bases.

2) Sequential filtering: The sequential filtering approach sequentially filters the results of different selection processes; at each stage, it eliminates a portion of possible sentences. Such an approach reduces informative sentences to the most

informative and pedagogically salient ones. As presented in Fig. 2, sequential filtering strategy works as follows:

a) Stage 1:Boolean filtering: This initial stage applies a logical filter to retain only sentences with high potential for informativeness:

- Filtering with NER: The process first discards those sentences that do not have named entities. Sentences with numerous entities (e.g. persons, places, organizations, or whatever is relevant to the domain) are most likely to carry information that can be evaluated as factual and conceptually significant [53]. Entity sentences are also quite useful in providing domain-specific and factual questions [1].
- Filtering with Keyword: In parallel, domain-specific keywords are used to retain sentences that directly reference core course concepts. This step ensures that essential and pedagogically valuable content is preserved early in the process.

b) Stage 2: Ranking_based filtering: The reduced candidate subset from stage1 is then used in:

- IG, SBERT ranking: The reduced candidate subset from stage1 is then ranked using a combined score derived from Information Gain (IG) and SBERT semantic similarity scores after normalization. IG highlights sentences that introduce statistically distinctive information within the corpus [54], while SBERT measures their semantic density and relevance. The fusion of these metrics prioritizes sentences that are relatively novel, semantically rich, and aligned with the instructional objectives.
- Last Selection: Top informative sentences are selected from previously ranked scoring as the final informative set.
- In this strategy, Sentences that match many criteria are given priority at each of the filtering stages. This multilevel approach has made the selection process precise by ensuring that the chosen sentences contain many entities and are informative regarding context and domain-specific knowledge. Sequential filtering is also used to reduce the noise and improve downstream AQG tasks.

3) Set-based combination (intersection/union): This strategy addresses sentence selection as a task of set-based combinations instead of a cross-methods approach. This technique outlines two mechanisms: Intersection and Union as shown in Fig. 3.

- Intersection (I): All those sentences that have failed to correspond to at least 2 methods or more are removed. The accuracy of this practice has been enhanced because only sentences that have been agreed upon by multiple methods are retained where precision has been observed from many perspectives. Intersection-based models have been employed in multi-criteria decision-making and have delivered positive outcomes in terms of inspiring false positives in text mining processes [55]. The purple central overlap in Fig. 3 highlights the intersection region.

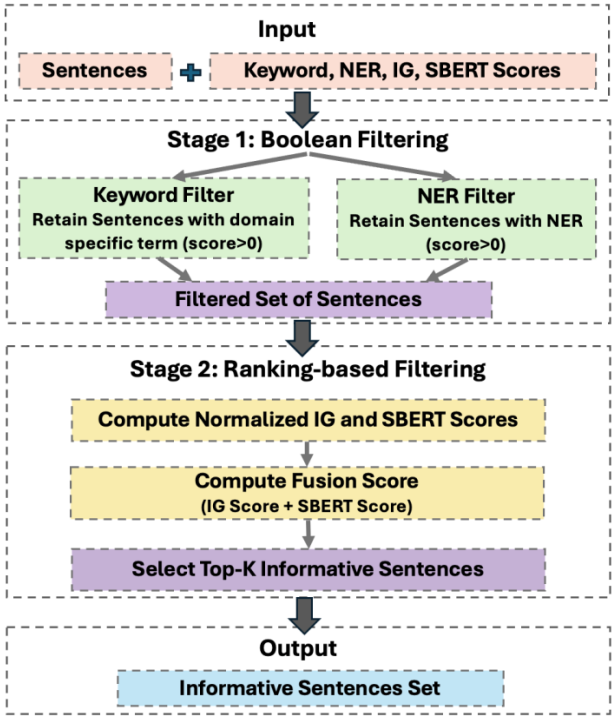

Fig. 2. Sequential filtering strategy steps.

- Union (U): The sentences identified by at least one method are considered and kept in the final selection. This strategy gives the greatest emphasis to recall (maximization of coverage), which is desired when answering a large number of different types of questions, as one may want to cover a large set of sentence candidates in an education setting [2]. The dashed rectangle in Fig. 3 outlines the Union region.

Typically, the set-based composition also involves a type of flexibility to apply an intersection- or union-based approach that is ranged between precision and recall.

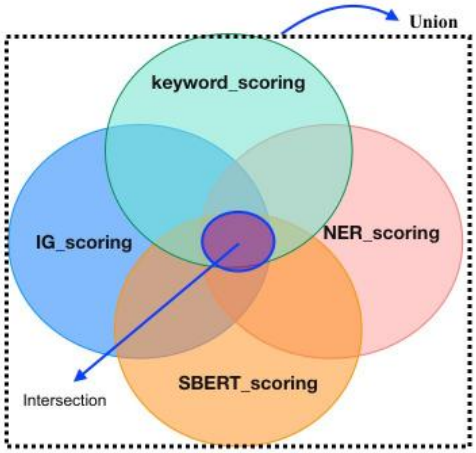

Fig. 3. Intersection and union strategies.

IV. Experimental Setup and Results

This section outlines the experimental setup, results, and evaluation applied to measure the effectiveness of the proposed hybrid ISE approach.

A. Tools and Dataset

The experimental implementation was conducted in Python 3.10 using open-source NLP libraries. Text preprocessing and NER were performed with NLTK and spaCy, while word-level IG scores were computed through a custom entropy-based implementation to estimate sentence informativeness. For semantic modelling, SBERT embeddings were obtained using the Sentence-Transformers library with the all-MiniLM-L6-v2 model, which provides efficient contextual sentence-level representations. All development was carried out in the PyCharm IDE [56], and experiments were executed on a macOS Monterey machine with an Apple M1 8-core processor and 16 GB RAM, eliminating the need for external high-performance computing resources.

The experimental dataset used to test the hybrid approach was generated from instructional materials for an information systems course taught in the business information systems program at Helwan University in Egypt. The dataset sentences are extracted from a book chapter covering core concepts such as hardware, software, and processes. After applying the preprocessing pipeline, which is part of our smart AQG model mentioned before, the dataset was transformed into a sentence-level corpus consisting of approximately 460 sentences. A subset of 58 candidate sentences was selected for evaluation, and a gold-standard reference set was manually annotated by two domain experts, labelling each sentence either informative or non-informative. A sample of the sentences dataset is represented in the tables that located in the following sections.

B. Sentence Informativness Scoring Results

In this module, each sentence in the preprocessed educational dataset was evaluated using the four informative sentence scoring techniques outlined in the proposed approach, to capture different dimensions of informativeness. The keyword-based scoring method assigned a score to each sentence based on the number of matched keywords it contained. Table II provides an example result of the keyword-based scoring method. Sentences such as S1, received higher scores (score = 3) because they contain several domain-specific words (e.g. software, hardware, and information system). Likewise, statements describing system components such as S3 and S5 received moderate scores (score = 2), whereas those using only a single keyword (e.g. “agile”) attained a score of 1. In contrast, phrases without any keywords, such as S40 received a score of 0, indicating poor informativeness according to this criterion.

The second technique applied in the suggested approach is NER, which detects factual entities in text and acts as a strong indicator of informative content. Table III displays a sample of the output produced by this method. As illustrated, sentences involving multiple entities earned higher scores, such as in S58, which earned a score of 3 due to the identification of entities. Sentences containing two entities, such as S8, received a score of 2, whereas those with a single defined entity, such as S19, were assigned a value of 1. On the other hand, sentences without any of the identifiable items, such as S42, received a score of 0.

TABLE II. Sample of the Keyword Based Scoring Output

Index	Sentence	Matched Keywords	Keyword Score
S1	Information systems are combinations of hardware, software, and telecommunications networks.	software, hardware, information system	3
S3	Computers, keyboards, disk drives, iPads, and flash drives are all examples of information systems hardware.	hardware, information system	2
S5	Software is a set of instructions that tells the hardware what to do.	software, hardware	2
S36	Agile methods emphasize flexibility in IS development.	agile	1
S40	This chapter served as an introduction to key IS themes.	-------	0

TABLE III. Sample of the Output Produced by NER-Based Scoring Method

Index	Sentence	Entities with Labels	NER Score
S58	In 1995, Jeff Bezos expanded Amazon from an online bookstore into a global e-commerce and cloud computing leader.	1995 (DATE), Jeff Bezos (PERSON), Amazon (ORG)	3
S8	Examples of operating systems include Microsoft Windows on a personal computer and Google’s Android on a mobile phone.	Microsoft Windows (ORG), Google (ORG)	2
S19	We will discuss processes in Chapter 8.	Chapter 8 (LAW)	1
S42	Big data technologies enable real-time analysis.	--------	0

The third technique applied is IG scoring, which evaluates the informativeness of sentences through examining the contribution of individual words inside every sentence, as illustrated in Eq. (1), (2), and (3). Table IV displays sample outcomes of IG-based scoring. S14 received the highest score (0.327), indicating that it has words with significant discriminatory power within the dataset. Sentences such as S41 and S47 received comparatively high scores; however, S47 did not clarify the basic concepts of the instructional material. In contrast, lines such as S40 received lower scores (0.235), indicating a reduced concentration on high-IG words.

The final technique applied in the proposed approach is SBERT-based semantic centrality scoring, which evaluates informativeness at the semantic level by using the pre-trained all-MiniLM-L6-v2 SBERT model. Semantic centrality has been evaluated, and the centrality score for each sentence was calculated as illustrated in Eq. (4), (5), and (6). Sentences with higher centrality scores are considered more informative, as they are semantically important to the main topics of the dataset.

Table V displays an example of the outcomes of the SBERT-based testing. The top-ranked sentences are S2 and S1, with each sentence scoring 0.332. Both statements clearly provide fundamental definitions and applications of

information systems, highlighting their significance within the corpus.

TABLE IV. SAMPLE OF THE OUTPUT PRODUCED BY IG-BASED SCORING METHOD

Index	Sentence	IG score
S14	Data is stored and processed by the Information System.	0.327321791
S41	Readers should now recognize the importance of IS in daily and organizational life.	0.271537004
S47	The next chapter will discuss system development in detail.	0.263299162
S5	Software is a set of instructions that tells the hardware what to do.	0.256487202
S40	This chapter served as an introduction to key IS themes.	0.235468327

TABLE V. SAMPLE OF THE OUTPUT PRODUCED BY SBERT-BASED SCORING METHOD

Index	Sentence	SBERT Score
S2	Information systems are essential to modern organizations, supporting daily operations, decision-making, and long-term strategies.	0.332139224
S1	Information systems are combinations of hardware, software, and telecommunications networks.	0.332084447
S16	People include all the individuals who interact with the Information System, from users who input data to IT professionals who maintain and manage the system.	0.293755561
S43	User-centered design ensures systems meet user needs.	0.286940813
S5	Software is a set of instructions that tells the hardware what to do.	0.252976418

Sentences such as S16 received relatively high scores 0.294, indicating their contribution to describing the vital roles and practical relevance of Information Systems. In contrast, sentences such as S5 achieved slightly lower scores 0.253, indicating limited relevance to the overall semantic context of the dataset.

C. Fusion Scoring Strategies Results

This section illustrates the results that were achieved through the fusion scoring strategies module.

1) ST1_weighted scoring: The input for this strategy contains four different scores resulting from the applied methods in module 1. The results have been normalized between 0 and 1 to facilitate integrating them into a final score. We assigned a weighted contribution of (0.4, 0.1, 0.2, 0.3) to each score, respectively, based on the nature of the questions that needed to be extracted from the course material. The highest weight has been assigned to keyword-based scoring, which indicates that in educational text, explicit domain terminology is the most reliable marker of informativeness. NER has the lowest weight, indicating that entity presence alone is not a strong enough criterion in the course domain, such as dates and names. The combined weights must equal one, as presented in Eq. (7). Table VI illustrates how ST1 provides the distinct differences between well-informed and poorly informed information.

TABLE VI. SAMPLE OF THE OUTPUT PRODUCED BY ST1

Index	Sentence	Combined Score
S1	Information systems are combinations of hardware, software, and telecommunications networks.	0.863
S23	Management Information Systems (MIS): This system is used to collect data from different sources, process the information, and present it to the management team for decision making.	0.811
S7	Most models describe information systems as having five main components: hardware, software, data, people, and processes.	0.783
S14	Data is stored and processed by the Information System.	0.659
S3	Computers, keyboards, disk drives, iPads, and flash drives are all examples of information systems hardware.	0.623
S17	Without people, even the most advanced system cannot generate meaningful results.	0.159
S15	For example, your street address, the city you live in, and your phone number are all pieces of data.	0.158
S32	Later sections will revisit ethics in greater detail.	0.107
S35	Sustainability is a new focus, considering environmental impacts.	0.099
S24	Readers should not worry if some terms seem unfamiliar, as they will be explained later.	0.054

The first five sentences (e.g. S1, S23, S7, S14, S2) involve definitions, functional roles, and system structures, which are essential to knowledge acquisition and therefore effective in formulating questions. In contrast, the last five sentences (i.e. S34, S35, 0S24, S27, S30) notably involve historical notes, general insights, and commentary that decrease their informative significance within the context. This indicates that this method effectively prioritizes statements based on their relevance to questions by emphasizing conceptually and semantically rich content.

2) ST2_sequential filtering: In sequential filtering strategy (ST2), the four different scores resulting from the applied methods in module one were used in sequential order.

Sentences were initially reviewed using keyword-based scoring and NER and then fine-tuned using IG and SBERT similarity scores. The filter applied at the first stage was Boolean filtering applied based on the existence of a keyword or NER in sentences to ensure that relevant sentences are kept. This step of filtering reduced the size of the data from 58 to 41 possible sentences, thereby discarding less informative data at an early stage. The second stage used the candidate list from stage 1 to apply a ranking process based on their combined IG and SBERT semantic similarity scores. The combination has enabled the system to test statistical relevance and semantic density, creating a more balanced measure of informativeness. The ranking process also further refined the results in order to end up with the 30 high-quality sentences. Table VII presents samples of these sentences.

3) Set-based combination (intersection/union): The union strategy was firstly applied to keyword and NER methods, which produced a total of 41 candidate sentences. Subsequently, 20 additional sentences were integrated into the existing 41 candidate sentences, utilizing the highest IG+SBERT scores, resulting in a total of 44 unique sentences.

(41+20 → 44 final) indicating that only 4 of the 20 semantically selected sentences didn't selected with the rule-based set and other 16 common was removed to avoid redundancy.

TABLE VII. SAMPLE OF THE OUTPUT PRODUCED BY ST2

Index	Filtered Sentence
S14	Data is stored and processed by the Information System.
S1	Information systems are combinations of hardware, software, and telecommunications networks.
S2	Information systems are essential to modern organizations, supporting daily operations, decision-making, and long-term strategies.
S51	Today, most IS are computer-based, but people and processes remain critical components.
S47	The next chapter will discuss system development in detail.
S3	Computers, keyboards, disk drives, iPads, and flash drives are all examples of information systems hardware.
S23	Management Information Systems (MIS): This system is used to collect data from different sources, process the information, and present it to the management team for decision making.
S16	People include all the individuals who interact with the Information System, from users who input data to IT professionals who maintain and manage the system.
S18	Processes refers to the procedures and protocols that guide the use of the system, including security measures, backup procedures, and data management policies.

This evidence indicates that rule-based filters neglected to recognize certain semantically selected statements, consequently underscoring the complementary nature of semantic scoring. An example would be sentences such as [S54], "During the input stage, data are collected from different sources, such as sensors, transactions, or user input," which would not usually be captured using keyword or Named Entity Recognition but are obtained through semantic scoring. The union strategy generally produces a wider and more informative final set of sentences, resulting from a balanced combination of explicit lexical structure and semantic similarity.

The Intersection strategy produced only 18 sentences, as it required the previous identification of "candidate sentences" by both keyword and NER techniques. The resulting sentences were then ranked using a combination of IG and SBERT scores to balance statistical and semantic relevance in the final ranking. Fig. 4 shows the intersection of the 18 sentences across scoring methods.

When attempting to retrieve the top 20 sentences common to the four applied methods, only 18 sentences were provided due to the constrained size of the intersection set, with a maximum possible of 18. This indicates that intersection procedures prioritise precision, albeit at the expense of recall, since multiple sentences selected by one individual method are excluded.

D. Evaluation Results

The results comparing the performance of the various strategies used to extract informative sentences demonstrate that there are disparities in performance in relation to three measures, namely, precision, recall, and overall F1-score as shown in Eq. (8), (9), and (10).

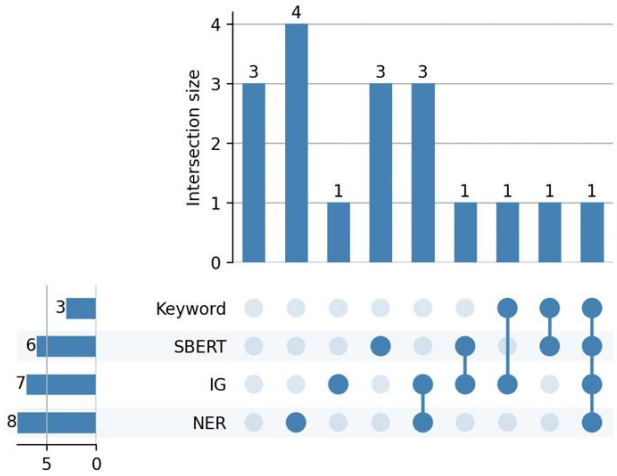

Fig. 4. Intersection of sentences across scoring methods.

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (10)$$

According to Dalianis (2018), these metrics are defined as follows: Precision is the proportion of correctly identified informative sentences among all those classified as informative. Recall is the proportion of informative sentences successfully retrieved by the system, and the F_1-score is their harmonic mean [57]. In this context:

- True Positives (TP): Candidate sentences correctly classified as informative.
- False Positives (FP): Non-informative sentences incorrectly selected by the system.
- False Negatives (FN): Informative sentences that were missed by the system.

This metric choice balances correctness (precision) and coverage (recall): selecting too few sentences risks missing important question-worthy content, while selecting too many introduces noise into the generated questions.

The findings suggest that the combination of methods usually scores better than single methods. Table VIII and Fig. 5 present the evaluation results of the four individual methods and the fusion scoring strategies. The Sequential method was the most precise (0.967), which can be interpreted as its capacity to identify sentences with the highest correctness, despite the relatively low precision (0.527). In parallel to that, the Weighted Score solution had one of the highest precisions (0.925) yet a more moderate recall (0.673) when compared to the rest of the solutions, leading it to a moderately high F1-score (0.779).

The Union strategy showed the highest overall trade-off, with both a high recall (0.745) and high precision (0.932) and, therefore, the highest F1-score (0.828). Single-component approaches, like NER (F1 = 0.500) and Information Gain (F1 = 0.627) were not effective, too either due to weak recall or

average-level precision. The worst in overall performance was the Intersection strategy (F1 = 0.411), whose restrictive filtering resulted in only a very few numbers of selected sentences and poor recall (0.273). The number of sentences selected also influenced performance, with broader selections, as in the Union method (44 sentences) and Keyword method (40 sentences), support high recall but restrict recall in other methods, such as Intersection (18 sentences) or NER (19 sentences) due to limited coverage. Overall, the results indicate that although high-precision strategies will guarantee the correctness, strategies as Union and Weighted Score that balance between precision and recall are still better to use for enhancing quality and diversity of questions in automatic question generation systems.

TABLE VIII. THE EVALUATION RESULTS OF THE FOUR INDIVIDUAL METHODS AND THE FUSION SCORING STRATEGIES

Strategy	Precision	Recall	F1_Score	Selected Sentences
Keyword	0.700	0.757	0.727	40
NER	0.737	0.378	0.500	19
IG	0.700	0.568	0.627	30
SBERT	0.767	0.622	0.687	30
Weighted_Score	0.925	0.673	0.779	40
Sequential	0.967	0.527	0.682	30
Union	0.932	0.745	0.828	44
Intersection	0.833	0.273	0.411	18

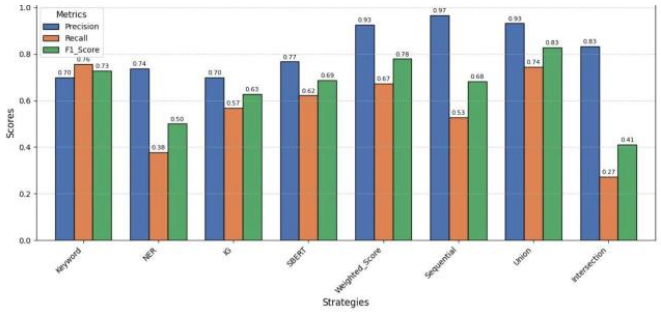

Fig. 5. Comparison of evaluation results.

V. DISCUSSION

This research proposed a multi-step hybrid framework that combines complementary methodologies for evaluating informativeness from linguistic and semantic perspectives. Keyword-based scoring identifies domain-specific lexical factors, named entity recognition (NER) identifies factual and entity-filled content, information gain (IG) scoring determines word-level contributions to sentence informativeness using entropy-based metrics, and Sentence-BERT (SBERT) provides contextual semantic representations to capture deeper meanings.

To overcome the limitations of individual methodologies, different fusion methods have been used to combine scores from these four methods, providing aggregated measures of informativeness for sentence ranking. The findings highlight a significant research shortcoming in existing AQG pipelines: despite major effort to enhance question generation models, the sentence extraction process is inadequately examined and often simplistic, depending only on individual features like keywords or entity presence. The evaluation results provide significant insights into the performance of different methods for extracting informative sentences in automatic question generation.

A significant finding is that individual methodologies, such as Named Entity Recognition (NER) and Information Gain (IG), have constrained efficiency. NER achieved the lowest F1-score (0.500), largely due to its poor recall (0.378), which indicates that while NER can capture entity-rich sentences with relatively satisfactory precision, it misses a substantial proportion of informative content. Likewise, IG (F1 = 0.627) shows decreased recall, indicating that scoring entirely on frequency and entropy fails to adequately cover semantically diverse informative sentences. These limitations highlight that relying solely on isolated linguistic or statistical indicators is insufficient for identifying a broad and reliable set of candidate sentences. The SBERT-based scoring method achieved a reasonably high F1-score of 0.687, surpassing both NER and IG by utilising contextual embeddings that incorporate sentence-level semantics. However, the absence of complementary structural or domain-specific signals continues to limit its effectiveness. In contrast, the examination of fusion-based methods indicates that hybrid approaches can significantly reduce these limitations. Strategies like Weighted Score and Union integrate lexical, statistical, and semantic attributes into a comprehensive metric of informativeness. The Weighted Score technique achieved an F1-score of 0.779 by balancing high accuracy (0.925) with satisfactory recall (0.673), whereas the Union strategy obtained superior overall performance (F1 = 0.828), utilising both strong precision (0.932) and recall (0.745). The results indicate that hybrid evaluation methods not only highlight highly accurate sentences but also offer extensive coverage of potentially informative content, which is essential for supporting diversity in AQG.

Another significant result is that hybrid approaches provide a practical solution to the trade-off between precision and recall, a continuous challenge in information extraction tasks. Whereas highly precise methods such as the Sequential approach minimise noise but at the cost of excluding valuable content, more balanced strategies like Union and Weighted Score ensure that extracted sentences are both accurate and academically rich. This finding highlights the necessity of overcoming rigid, single-dimensional extraction criteria in preference for adaptive frameworks that can incorporate several dimensions of informativeness.

Overall, these outcomes reinforce the key point of this work: the extraction of informative sentences should be treated as a core research problem in AQG, rather than as a secondary or optional task. The demonstrated improvements achieved by hybrid strategies highlight their potential to bridge the gap between technical advances in natural language generation and the practical needs of educational applications, where both the quality and diversity of generated questions are essential.

VI. CONCLUSION

The paper presented a novel hybrid model of Informative Sentence Extraction (ISE) in Automatic Question Generation (AQG) with keyword scoring, Named Entity Recognition, Information Gain, and Sentence-BERT embedding. The integration of complementary methods with fusion techniques substantially enhanced the process of selecting contextually precise and rich informative sentences. Experimental evidence showed that both the Union and Weighted Score strategies would be effective when a balance in precision and recall is required. The results demonstrate the success of multi-method combination in the improvement of question quality and diversity, which will serve as a strong basis for downstream AQG tasks. Generally, the presented framework contributes to the evolution of intelligent learning tools because it optimizes sentence selection based on the unstructured instructional material. Future work should extend these hybrid models by incorporating large language models (LLMs), adaptive weighting mechanisms, and domain-specific knowledge graphs to further improve generalizability and pedagogical validity. Additionally, evaluating the generated questions with learners and educators would provide practical evidence of the impact of hybrid extraction strategies on learning outcomes, thus bridging technical innovation with educational practice.

REFERENCES

[1] Y. Chali and S. A. Hasan, "Towards Topic-to-Question Generation," Computational Linguistics, vol. 41, no. 1, pp. 1–20, Mar. 2015, doi: 10.1162/COLI_a_00206.

[2] G. Kurdi, J. Leo, B. Parsia, U. Sattler, and S. Al-Emari, "A Systematic Review of Automatic Question Generation for Educational Purposes," Int J Artif Intell Educ, vol. 30, no. 1, pp. 121–204, Mar. 2020, doi: 10.1007/s40593-019-00186-y.

[3] G. Chen, J. Yang, and D. Gasevic, "A comparative study on question-worthy sentence selection strategies for educational question generation," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Verlag, 2019, pp. 59–70. doi: 10.1007/978-3-030-23204-7_6.

[4] X. Du, J. Shao, and C. Cardie, "Learning to ask: Neural question generation for reading comprehension," in ACL 2017 - 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers), Association for Computational Linguistics (ACL), 2017, pp. 1342–1352. doi: 10.18653/v1/P17-1123.

[5] E. M. Perkoff, A. Bhattacharyya, J. Cai, and J. Cao, "Comparing Neural Question Generation Architectures for Reading Comprehension," in Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023), Stroudsburg, PA, USA: Association for Computational Linguistics, 2023, pp. 556–566. doi: 10.18653/v1/2023.bea-1.47.

[6] D. Lindberg, F. Popowich, J. Nesbit, and P. Winne, "Generating Natural Language Questions to Support Learning On-Line," in Proceedings of the 14th European Workshop on Natural Language Generation (ENLG), Sofia, Bulgaria: Association for Computational Linguistics, 2013, pp. 105–114. doi: 10.3115/v1/W13-2114.

[7] S. Guo, Y. Guan, R. Li, X. Li, and H. Tan, "Incorporating Syntax and Frame Semantics in Neural Network for Machine Reading Comprehension," in Proceedings of the 28th International Conference on Computational Linguistics, Stroudsburg, PA, USA: International Committee on Computational Linguistics, 2020, pp. 2635–2641. doi: 10.18653/v1/2020.coling-main.237.

[8] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using siamese BERT-networks," in EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference, 2019. doi: 10.18653/v1/d19-1410.

[9] H. P. Luhn, "The Automatic Creation of Literature Abstracts," IBM J Res Dev, vol. 2, no. April, pp. 159–165, 1958.

[10] P. Watanangura, S. Vanichrudee, O. Minteer, T. Sringamdee, N. Thanngam, and T. Siriborvornratanakul, "A Comparative Survey of Text Summarization Techniques," SN Comput Sci, vol. 5, no. 1, 2024, doi: 10.1007/s42979-023-02343-6.

[11] W. A. N. G. Zhuohao, W. A. N. G. Dong, and L. I. Qing, "Keyword Extraction from Scientific Research Projects Based on SRP-TF-IDF," Chinese Journal of Electronics, vol. 30, no. 4, 2021, doi: 10.1049/cje.2021.05.007.

[12] V. Harrison and M. Walker, "Neural generation of diverse questions using answer focus, contextual and linguistic features," in INLG 2018 - 11th International Natural Language Generation Conference, Proceedings of the Conference, 2018. doi: 10.18653/v1/w18-6536.

[13] Y. Lu and S.-E. Lu, "A survey of Approaches to Automatic Question Generation: from 2019 to Early 2021," in Proceedings of the 33rd Conference on Computational Linguistics and Speech Processing (ROCLING 2021), The Association for Computational Linguistics and Chinese Language Processing, Oct. 2021, p. 151.

[14] R. Heer, "A Model of Learning Objectives based on a Taxonomy for Learning, Teaching, and Assessing: A revision of Bloom's Taxonomy of Educational Objectives," Jul. 2012. Accessed: Sep. 15, 2025. [Online]. Available: https://www.celt.iastate.edu/wp-content/uploads/2015/09/RevisedBloomsHandout-1.pdf

[15] M. Heilman and N. A. Smith, "Good question! Statistical ranking for question generation," in NAACL HLT 2010 - Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Proceedings of the Main Conference, 2010.

[16] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using siamese BERT-networks," in EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference, 2019. doi: 10.18653/v1/d19-1410.

[17] A. Nenkova and K. McKeown, "A Survey of Text Summarization Techniques," in Mining Text Data, Boston, MA: Springer US, 2012, pp. 43–76. doi: 10.1007/978-1-4614-3223-4_3.

[18] Y. J. Kumar, O. S. Goh, H. Basiron, N. H. Choon, and P. C. Suppiah, "A review on automatic text summarization approaches," 2016. doi: 10.3844/jcssp.2016.178.190.

[19] H. N. Fejer and N. Omar, "Automatic multi-document Arabic text summarization using clustering and keyphrase extraction," Journal of Artificial Intelligence, vol. 8, no. 1, 2015, doi: 10.3923/jai.2015.1.9.

[20] Z. Edress and Y. Ortakci, "Optimizing Text Summarization with Sentence Clustering and Natural Language Processing," International Journal of Advanced Computer Science and Applications, vol. 15, no. 10, 2024, doi: 10.14569/IJACSA.2024.01510115.

[21] A. P. Widyassari et al., "Review of automatic text summarization techniques & methods," 2022. doi: 10.1016/j.jksuci.2020.05.006.

[22] L. Zhu, S. Gao, S. J. Pan, H. Li, D. Deng, and C. Shahabi, "Graph-based informative-sentence selection for opinion summarization," in 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013), 2013, pp. 408–412. doi: 10.1109/ASONAM.2013.6785738.

[23] S. and P. S. J. and L. H. and D. D. and S. C. Zhu Linhong and Gao, "The Pareto Principle Is Everywhere: Finding Informative Sentences for Opinion Summarization Through Leader Detection," in Recommendation and Search in Social Networks, A. U. and A. E. Ulusoy Özgür and Tansel, Ed., Cham: Springer International Publishing, 2015, pp. 165–187. doi: 10.1007/978-3-319-14379-8_9.

[24] J.-L. Koh and C.-W. Cho, "Informative Sentence Retrieval for Domain Specific Terminologies," 2011, pp. 242–252. doi: 10.1007/978-3-642-21822-4_25.

[25] K. P. Snehalkumar Nanubhai Patel, "Informative Term Selection And Novel Sentence Extraction," Journal of Engineering Computers & Applied Sciences, vol. 4, no. 6, pp. 165–167, Jun. 2015.

[26] M. Kashi, S. Lahmiri, and O. A. Mohamed, "Comprehensive analysis of Transformer networks in identifying informative sentences containing customer needs," Expert Syst Appl, vol. 273, p. 126785, May 2025, doi: 10.1016/J.ESWA.2025.126785.

[27] M. Heilman and N. A. Smith, "Human Good Question! Statistical Ranking for Question Generation," in Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL, Los Angeles, California: Association for Computational Linguistics, Jun. 2010, pp. 609–617. Accessed: Sep. 16, 2025. [Online]. Available: https://aclanthology.org/N10-1086/

[28] R. Mitkov, L. A. Ha, and N. Karamanis, "A computer-aided environment for generating multiple-choice test items," Nat Lang Eng, vol. 12, no. 2, 2006, doi: 10.1017/S1351324906004177.

[29] J. Pino, M. J. Heilman, and M. Eskenazi, "A selection strategy to improve cloze question quality," Proceedings of the Workshop on Intelligent Tutoring Systems for Ill-Defined Domains, 2008.

[30] A. Kurtasov, "A system for generating cloze test items from Russian-language text," in International Conference Recent Advances in Natural Language Processing, RANLP, 2013.

[31] R. Correia, J. Baptista, M. Eskenazi, and N. Mamede, "Automatic generation of Cloze question stems," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2012. doi: 10.1007/978-3-642-28885-2_19.

[32] M. Majumder Sujan Kumar Saha, M. Majumder, and S. Kumar Saha, "Knowledge Management & E-Learning Automatic selection of informative sentences: The sentences that can generate multiple choice questions Automatic selection of informative sentences: The sentences that can generate multiple choice questions," Knowledge Management & E-Learning, vol. 6, no. 4, pp. 377–391, 2014.

[33] A. Y. Bokan and T. A. S. Pardo, "Automatic Aspect Identification: The Case of Informative Microaspects in News Texts," Research in Computing Science, vol. 90, no. 1, 2015, doi: 10.13053/rcs-90-1-18.

[34] O. Keklik, T. Tuglular, and S. Tekir, "Rule-based automatic question generation using semantic role labeling," IEICE Trans Inf Syst, vol. E102D, no. 7, 2019, doi: 10.1587/transinf.2018EDP7199.

[35] W. T. Sewunetie and L. Kovacs, "Automatic Question Generation based on Sentence Structure Analysis," in Proceedings of the 2023 24th International Carpathian Control Conference, ICCC 2023, 2023. doi: 10.1109/ICCC57093.2023.10178946.

[36] M. Blšták and V. Rozinajová, "Automatic question generation based on sentence structure analysis using machine learning approach," Nat Lang Eng, vol. 28, no. 4, 2022, doi: 10.1017/S1351324921000139.

[37] M. Blšták and V. Rozinajová, "Automatic question generation based on analysis of sentence structure," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9924 LNCS, 2016. doi: 10.1007/978-3-319-45510-5_26.

[38] S. Subramanian, T. Wang, X. Yuan, S. Zhang, A. Trischler, and Y. Bengio, "Neural Models for Key Phrase Detection and Question Generation," 2017, doi: 10.48550/arXiv.1706.04560.

[39] C. A. Nwafor and I. E. Onyenwe, "An Automated Multiple-Choice Question Generation using Natural Language Processing Techniques," International Journal on Natural Language Computing, vol. 10, no. 02, pp. 1–10, Apr. 2021, doi: 10.5121/ijnlc.2021.10201.

[40] I. Pilán, E. Volodina, and L. Borin, "Candidate sentence selection for language learning exercises: from a comprehensive framework to an empirical evaluation," ArXiv, vol. abs/1706.03530, 2017, [Online]. Available: https://api.semanticscholar.org/CorpusID:896837

[41] X. Ma, Q. Zhu, Y. Zhou, and X. Li, "Improving Question Generation with Sentence-Level Semantic Matching and Answer Position Inferring," Dec. 2020. doi: 10.48550/arXiv.1912.00879.

[42] I. Keraghel, S. Morbieu, and M. Nadif, "Recent Advances in Named Entity Recognition: A Comprehensive Survey and Comparative Study," arXiv preprint, Dec. 2024, doi: 10.48550/arXiv.2401.10825.

[43] M. Munnangi, "A Brief History of Named Entity Recognition," Nov. 2024. doi: 10.48550/arXiv.2411.05057.

[44] C. Lopez et al., "Recursive Named Entity Recognition," in Studies in Computational Intelligence, 2022. doi: 10.1007/978-3-030-90287-2_2.

[45] H. Kim, J. Yoo, S. Yoon, and J. Kang, "Automatic Creation of Named Entity Recognition Datasets by Querying Phrase Representations," in Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Toronto, Canada: Association for Computational Linguistics, Jul. 2023, pp. 7148–7163. doi: 10.18653/v1/2023.acl-long.394.

[46] I. E. Fattoh, "Semantic Based Automatic Question Generation using Artificial Immune System," Issn, vol. 5, no. 8, 2014.

[47] Explosion AI, "spaCy 101: Everything you need to know," Explosion AI, Berlin, Germany. Accessed: Sep. 15, 2025. [Online]. Available: https://spacy.io/usage/spacy-101

[48] R. ÇEKİK and M. KAYA, "A New Feature Selection Metric Based on Rough Sets and Information Gain in Text Classification," Gazi University Journal of Science Part A: Engineering and Innovation, vol. 10, no. 4, pp. 472–486, Dec. 2023, doi: 10.54287/gujsa.1379024.

[49] A. Mazyad, F. Teytaud, and C. Fonlupt, "Information Gain Based Term Weighting Method for Multi-label Text Classification Task," 2018. [Online]. Available: https://hal.science/hal-01859697v1

[50] S. , K. E. , & L. E. Bird, "NLTK: The Natural Language Toolkit – nltk.probability.FreqDist," Association for Computational Linguistics. Accessed: Sep. 15, 2025. [Online]. Available: https://www.nltk.org/

[51] J. Opitz and A. Frank, "SBERT studies Meaning Representations: Decomposing Sentence Embeddings into Explainable Semantic Features," Oct. 2022, [Online]. Available: http://arxiv.org/abs/2206.07023

[52] T. T. Ha, V. N. Nguyen, K. H. Nguyen, K. A. Nguyen, and Q. K. Than, "Utilizing SBERT For Finding Similar Questions in Community Question Answering," in Proceedings - International Conference on Knowledge and Systems Engineering, KSE, 2021. doi: 10.1109/KSE53942.2021.9648830.

[53] J. Li, A. Sun, J. Han, and C. Li, "A Survey on Deep Learning for Named Entity Recognition," Mar. 2020, doi: 10.1109/TKDE.2020.2981314.

[54] T. M. Cover and J. A. Thomas, Elements of Information Theory. 2005. doi: 10.1002/047174882X.

[55] G. Kou, P. Yang, Y. Peng, F. Xiao, Y. Chen, and F. E. Alsaadi, "Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods," Applied Soft Computing Journal, vol. 86, Jan. 2020, doi: 10.1016/j.asoc.2019.105836.

[56] JetBrains, "PyCharm IDE (Version 2023.3) [Computer software]," JetBrains s.r.o. (or JetBrains s.r.o., Prague, Czech Republic). Accessed: Sep. 15, 2025. [Online]. Available: https://www.jetbrains.com/pycharm/, 2023.

[57] H. Dalianis, "Evaluation Metrics and Evaluation," in Clinical Text Mining, Springer International Publishing, 2018, pp. 45–53. doi: 10.1007/978-3-319-78503-5_6.