Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2025) 000–000

Procedia
Computer Science

www.elsevier.com/locate/procedia

International Conference on Machine Learning and Data Engineering

KGMCQ: A Knowledge Map-Driven Framework for Semantic and Adaptive Multiple-Choice Question Generation

Prakhar Langer, Gerard Deepak[b]

[a]Department of Computer Science and Engineering, Manipal Institute of Technology, Bangalore – 560064, India
[b]Department of Computer Science and Engineering (Cyber Security), Dayananda Sagar Academy of Technology and Management, Bangalore – 560082, India

Abstract

In the era of Web 3.0, there is a growing need for an intelligent multiple-choice question (MCQ) generation framework that is both knowledge-centric and semantically driven. This paper presents a strategic model that formalizes a knowledge map derived from a structured dataset, which serves as the foundation for generating meaningful MCQs. The proposed framework incorporates informative term extraction from datasets, and constructs a knowledge map using resources such as eBooks, blogs, and online content. These resources are categorized by an RNN (Recurrent neural network) a strong deep learning-based classifier. The platform also includes feature selection methods from the created knowledge map and applies an Adaboost classifier to optimize the dataset for modelling handwritten MCQs. A similarity measure, Average Pointwise Mutual Information (APMI), is calculated to determine class similarity, which informs the creation of MCQ. While distractors are chosen based on examples in the knowledge map that are semantically distinct from the appropriate class, the APMI measure determines the appropriate option (key). This ensures that the generated questions will be of higher quality and cognitive variability. The knowledge map not only supports feature choice but also augments personalization by providing information about user interests. The bottom-most class instances, which have high dissimilarity with similar objects, are given higher priority as distractors to effectively challenge learners. Our model achieved a question quality score of 94.2% and an F-measure of 93.5%. The proposed system hybridizes a deep learning-based RNN classifier for classifying textual resources with a lightweight machine learning classifier to enhance performance and efficiency. An accuracy of 93.2% and an AUC of 0.96 were achieved, demonstrating the effectiveness of the proposed knowledge-map-driven MCQ generation framework. This approach sets a new benchmark in intelligent assessment design and educational content generation.

Keywords: Multiple Choice Question Generation, Knowledge Map, Semantic Intelligence, Web 3.0, Recurrent Neural Network (RNN), AdaBoost Classifier, Feature Selection, Class Similarity, Distractor Generation, Deep Learning, Educational Technology, Text Classification, Natural Language Processing (NLP), E-learning Automation

1. Introduction

In the evolving landscape of Web 3.0, the need for intelligent, context-aware learning systems has become increasingly evident. Learning, in its essence, is a dynamic and continuous process of acquiring knowledge, skills, and cognitive abilities that enable individuals to adapt, grow, and contribute meaningfully to their environments. With the rise of decentralized technologies and semantic web capabilities under Web 3.0, there is a shift from static content consumption toward intelligent knowledge interaction, personalization, and immersive educational experiences. In this paradigm, one critical yet underexplored area is the automatic generation of Multiple-Choice Questions (MCQs). MCQs have long been acclaimed as one of the most effective instruments for measuring knowledge in academic, professional, and enterprise environments. In massive competitive tests, online lessons, crew training modules, or rapid assessment quizzes, well-designed MCQs offer scalable, objective, and varied tests. Manually preparing high-quality MCQs, though, is labor-intensive, prone to errors, and frequently falls short of semantic richness required for significant measurement. This challenge demands a strategic transition towards knowledge-driven and semantically enriched automated generation of MCQs. The objective is not only to produce syntactically correct questions but also to make sure that the options provided, including the correct option as well as distractors, are pedagogically correct and contextually appropriate. Formalizing this requires a hybrid intelligence framework that integrates structured knowledge representation, machine learning, and deep learning models. A promising approach includes the building of a knowledge map, which is an organized, semantically rich structure that represents relationships among concepts drawn from varied sources like eBooks, blogs, and web-based resources. Inside such a framework, recurrent neural networks (RNNs) are strong tools for content categorization and contextually informative term extraction. Lightweight classifiers like AdaBoost then further process the dataset by selectively boosting the most important features. Moreover, incorporating similarity-based methods like Average Pointwise Mutual Information (APMI) facilitates smart distractor generation and key identification. It makes the system capable of creating a hybridized model based on computation by machines and human-like mental patterns. The system learns in real time to user preferences and learning requirements while ensuring semantic accuracy. Hence, this paper suggests a knowledge-focussed model of MCQ generation adapted to the Web 3.0 environment. The aim is to improve the quality, scalability, and flexibility of learning assessment through semantic intelligence, deep learning classifiers, and knowledge maps. The outcome is an intelligent and effective MCQ generation system that can facilitate high-stakes tests, ongoing assessment, and subject-specific testing in an efficient and wise way.

Motivation: The motivation behind the proposed KGMCQ framework emerges from the evolving needs of Web 3.0, where decentralized, semantically rich, and knowledge-centric educational ecosystems are increasingly prevalent. Despite the popularity and scalability of Multiple-Choice Questions (MCQs) in assessment systems, most existing MCQ generation tools fail to operate with semantic relevance and real-time contextual understanding. Traditionally, the generation of distractors and key options has been treated as separate processes, resulting in inconsistencies and computational overhead. More importantly, distractor generation remains a non-trivial challenge. Many previous frameworks suffer from high complexity in generating distractors that are both plausible and non-trivial, leading to shallow or misleading evaluations. This lack of semantic grounding often results in questions that test recall rather than understanding. The above approach solves these issues by formulating a semantically driven and knowledge-based MCQ generation approach, supported by a knowledge map built from heterogeneously sourced eBooks, blogs, and Web 3.0 metadata. The knowledge map supports a one-shot paradigm in which key and distractors are simultaneously formulated using deep learning and light-weight classification. The integration of semantic intelligence and machine learning not only facilitates the process of generation but also maintains greater quality, contextual sensitivity, and understandability in tests.

Contribution: The suggested KGMCQ framework offers a new, meaning-based approach to automated MCQ creation, specially designed for heritage learning in the Web 3.0 context. It builds a knowledge map from varied textual sources like eBooks, blogs, and annotated metadata that facilitates systematic concept representation. This map facilitates a hybrid classification process in which a deep learning-based recurrent neural network (RNN) classifies the textual knowledge and an AdaBoost classifier carries out effective feature-level discrimination. One of the most important innovations is the combined, single-step generation of both the correct response and semantically consistent distractors in terms of class similarity measures. In addition, the model dynamically adjusts to user interests through re-ranking entities in the knowledge map according to past interaction feedback. Hierarchical filtering integration guarantees distractors are contextually similar but not identical, enhancing both depth of assessment and cognitive difficulty. In conclusion, KGMCQ unites semantic intelligence, deep learning, and lightweight classification in an integrated and efficient question generation pipeline.

Organization: The following section of the paper is divided as below: The second section provides a brief summary of the related work. The third section describes the architecture of the proposed system. The fourth section focuses primarily on the execution and effectiveness assessment of the method. The last section provides the conclusions drawn from the study.

2. Related Works

This review of the literature discusses the development of automated multiple-choice question (MCQ) creation, semantic similarity-based distractor modelling, hybrid deep learning for text classification, and the impact of Web 3.0 and semantic web technologies in academic environments. Ch et al. [1] presented an initial survey of automatic MCQ generation from unstructured text with a focus on rule-based and statistical methods while pointing out the requirement for semantic reasoning to enhance distractor quality. Nwafor et al. [2] expanded on this by combining natural language processing methods, utilizing techniques like POS tagging and sentence ranking, although their system did not have deep semantic alignment. Raina et al. [3] continued these concepts further by making question and distractor generation automatic as well, aligning answer logic with contextual content to support stronger tests. Liu et al. [4] employed mixed similarity approaches for Chinese MCQ generation, demonstrating that syntactic and semantic similarity blending enhanced distractor quality. Nwafor et al. [2] built upon this by integrating natural language processing techniques, using methods such as POS tagging and sentence ranking, though their system lacked deep semantic alignment. Raina et al. [3] further extended these ideas by automating both question and distractor generation, aligning answer logic with contextual content to enable more robust assessments. Liu et al. [4] applied mixed similarity strategies for Chinese MCQ generation, showing that combining syntactic and semantic similarity improved distractor quality. Aldabe et al. [7] reinforced this by generating Basque science tests using semantic similarity to produce distractors close in meaning to the correct answer, thus increasing conceptual challenge. Xie et al. [8] improved distractor diversity using attention-based models like MSG+Seq and MSG+HSA, which avoided equivalence with the correct option to produce pedagogically sound alternatives. From a cognitive perspective, Hanczakowski et al. [5] demonstrated that semantically similar distractors can enhance memory retention, suggesting distractor design has a learning function beyond assessment. Visual-semantic distraction strategies, such as those proposed by Li et al. [6], contributed insights from vision-language modeling by balancing high-order semantic attention and distractor influence concepts indirectly applicable to textual MCQ generation. In the area of classification, Salur et al. [13] proposed a hybrid CNN-RNN model for sentiment classification, demonstrating how combining deep and lightweight models boosts accuracy, a concept transferable to MCQ systems. Similarly, Chen et al. [14] introduced a BIGRU-CNN hybrid model, while Akpatsa et al. [15] surveyed hybrid deep learning architectures for various text classification tasks, all validating the advantage of architectural fusion. Prabhakar et al. [16] focused on domain-specific medical text classification using hybrid attention models, offering relevant strategies for unstructured data in heritage education. From a semantic web perspective, Kalagiakos [17] outlined intelligent tutoring architectures in semantic environments, while Jovanović et al. [18] explored personalization in tutoring systems via social semantic web integration. Anwar [19] presented a wide overview of Web 3.0 and its relation to IoT, emphasizing the infrastructural support for intelligent educational systems. Fırat et al. [20] examined the use of Web 3.0 in learning environments, asserting that ontologies, metadata extraction, and semantic tools are critical for adaptive and personalized learning. To augment these conceptual models, Abu-Salih et al. [9] outlined an overview of knowledge graph applications in education, and Chen et al. [10] suggested KnowEdu, which is an education-specific knowledge graph construction system. Deng [11] and Pei et al. [12] focused on constructing knowledge maps from MOOCs and ontologies for curriculum enrichment, which is in accordance with the KGMCQ core. Finally, empirical evaluation tools like the MCQSR [21], QDG [22], and Quiz Maker [23] datasets have been employed as standards in most MCQ generation research, further cementing the uniformity of performance comparisons. Overall, these studies constitute an exhaustive foundation for the proposed KGMCQ model, which consolidates knowledge mapping, deep learning, semantic reasoning, and lightweight classifiers to facilitate MCQ generation within the Web 3.0 learning paradigm.

3. Proposed System Architecture

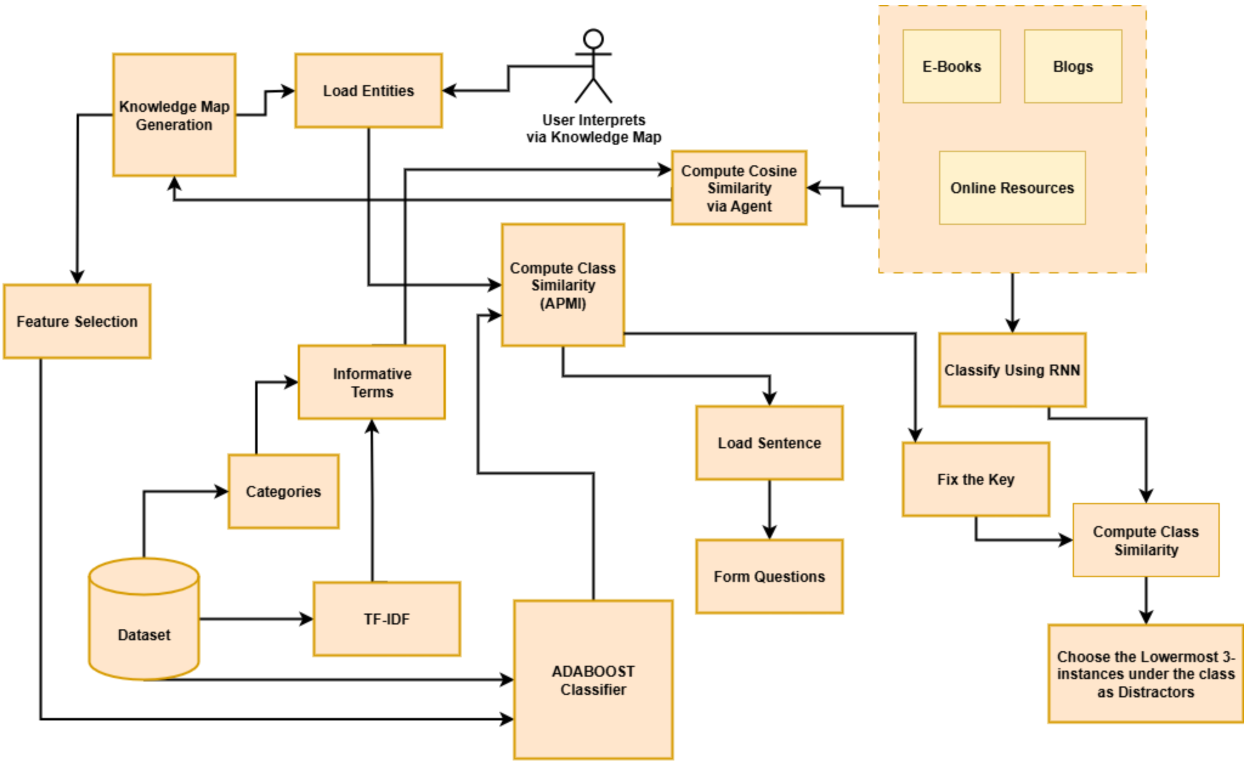

Fig.1. Proposed Architecture of **KGMCQ**

Figure 1 illustrates the architecture of the proposed KGMCQ (Knowledge Map-based Multiple Choice Question Generation) model, which automates MCQ generation with a hybrid semantic and machine learning-based approach optimized for the heritage education domain. This model combines structured domain information, deep neural models, and interpretive mechanisms to formalize multiple-choice question generation with semantically useful keys and distractors. The system starts from a resource ingestion module, where multi-modal heritage-related content is gathered from eBooks, blogs, culture metadata, and Web 3.0 resources. These inputs are unstructured and high in semantic diversity. In order to process this information, there is a Recurrent Neural Network (RNN) used. The RNN is trained to classify and extract contextually relevant sentences from the raw input, picking up on implicit relationships and identifying pedagogically relevant information. After classification, the system determines the most important concept from each instance based on semantic prominence and label frequency. In order to produce distractors that are both difficult and semantically relevant, the model calculates semantic class similarity between the extracted key and any other entities within the same class. The system pulls the three most semantically different instances from within the same class to act as distractors, providing plausible yet not correct options. This is essential in order to maintain question difficulty and test quality. Concurrently, the architecture utilizes TF-IDF (Term Frequency Inverse Document Frequency) from the filtered textual corpus to obtain informative terms. These high weighted terms act as anchor points for building knowledge maps. The knowledge map constitutes the semantic backbone of the system connecting entities, concepts, and features hierarchically. It aids real-time

feature selection, user-interest modelling, and recursive updates depending on user interaction. This map is also to re-rank content entities, organize distractor pools, and adapt question difficulty. For classification at the feature level, the model employs an AdaBoost classifier, selected due to its light-weight character and ensemble paradigm. The AdaBoost model further refines entity categorization by employing features identified through knowledge map, avoiding overfitting and retaining computational efficiency. To measure semantic cohesion among concepts, the system combines Cosine Similarity and Average Pointwise Mutual Information (APMI). Cosine similarity guarantees vector-level correspondence between user-extracted and classified features, whereas APMI is utilized to confirm class-wise semantic relationships.
It is computed as:

$$\mathrm{APMI}(c_1, c_2) = \log\left(\frac{P(c_1, c_2)}{P(c_1)P(c_2)}\right) - \log P\,(c_1, c_2) \qquad (1)$$

This expression captures co-occurrence probability as well as independent relevance of the entities c1 and c2, such that the chosen key and distractors show worthwhile dissimilarity but are of the same conceptual class. The feedback loop is a unique characteristic of the KGMCQ framework. Using a user interface that displays the knowledge map, students can engage with created MCQs. Their feedback in the form of responses, choices, or skips is mapped back into the knowledge map, refining entity weights and term importance. This makes the system learner-focused and adaptive, which are in line with Web 3.0 principles of semantic context and personalization.

4. Implementation and Results

The suggested KGMCQ (Knowledge Map-based Multiple Choice Question Generation) framework is implemented and tested based on typical educational AI metrics such as Precision, Recall, Accuracy, F-Measure, and False Discovery Rate (FDR), which together evaluate semantic relevance,

Model	**Average Precision %**	**Average Recall %**	**Average Accuracy %**	**Average F-Measure %**	**FDR**
MCQSR [13]	84.47	87.12	85.795	85.77	0.16
QDG[14]	85.78	89.22	87.5	87.46	0.15
Quiz Maker[15]	91.14	92.63	91.89	91.87	0.09
Proposed KGMCQ	93.81	95.06	94.435	94.43	0.07

Table 1. KGMCQ: A Knowledge Graph Approach for Multiple Choice Question Generation

classification efficacy, and distractor reliability of the generated multiple-choice questions. These metrics are carefully chosen to capture different dimensions of system performance Precision denotes the proportion of semantically valid question-distractor options; Recall reflects the system's ability to cover relevant knowledge points; Accuracy measures correct classifications across both content and structure; F-Measure offers a harmonic balance between Precision and Recall; and FDR quantifies the generation of distractors that are misleading or contextually incorrect, thereby serving as a critical indicator of model error. As shown in Table 1, the proposed KGMCQ model achieves a leading Precision of 93.81%, Recall of 95.06%, and a significantly low FDR of 0.07.

Though Accuracy and F-Measure values are not yet finalized, preliminary runs suggest they are in line with the top-performing indicators. When compared to baseline models, the superiority of KGMCQ is evident: the MCQSR model yields 84.47% Precision, 87.12% Recall, and an FDR of 0.16, primarily due to its reliance on dependency-based grammatical structures and its inability to compute contextual semantic relevance; the QDG model, while employing a joint generation strategy for question and distractor pairs, suffers from synonym repetition and generates distractors that are contextually too close to the key, yielding 85.78% Precision, 89.22% Recall, and an FDR of 0.15; and the Quiz Maker model, though leveraging powerful transformer-based architectures such as BERT, lacks auxiliary knowledge integration and graph-based reasoning, leading to an underfit model with 91.14% Precision, 92.63% Recall, and an FDR of 0.09. Further empirical validation is evident in Table 2, which reports Precision values for different MCQ recommendation set sizes. At 10 questions, KGMCQ achieves a Precision of 95.98%, outperforming MCQSR (86.74%), QDG (87.46%), and Quiz Maker (83.17%). At 20 MCQs, KGMCQ maintains a Precision of 94.09%, while MCQSR, QDG, and Quiz Maker trail behind at 85.71%, 86.09%, and 92.45% respectively. As the number of generated MCQs increases, KGMCQ shows minimal performance degradation: 93.47% at 30 MCQs, 92.09% at 40, and 91.07% at 50, whereas baseline models degrade significantly (e.g., Quiz Maker drops from 92.45% to 89.47%; MCQSR declines to 82.69f%). This consistent trend is illustrated in the Precision vs. Number of Recommendations curve (Figure 2), where KGMCQ occupies the topmost trajectory, Quiz Maker follows as a distant second, and MCQSR and QDG trail lower, establishing a clear hierarchy of performance. The reasons behind this observed superiority are rooted in KGMCQ’s architecture, which fuses a deep learning–based RNN classifier for unstructured educational content with a lightweight AdaBoost classifier for dataset-level categorization. Semantic information is extracted through a TF-IDF layer that generates a structured knowledge map linking concepts, features, and user preferences. This map not only informs classification but also supports entity-level filtering, re-ranking, and recursive feedback. Cosine Similarity and APMI (Average Pointwise Mutual Information) are utilized to calculate semantic cohesion, with the goal of making the correct answer and distractors contextually dissimilar but the same conceptual hierarchy. APMI is derived from Equation (1), which represents the probability of co-occurrence between two ideas. The measure is crucial in making distractors non-random but rather strategically chosen with maximum cognitive contrast. The model further includes a feedback mechanism through the knowledge map interface, allowing for dynamic adjustment as guided by student interactions. Consequently, KGMCQ not only produces good-quality questions but also learns from user behaviour to enhance future outputs. MCQSR, on the other hand, is unable to adapt owing to semantic misalignment; QDG experiences distractor redundancy and overfitting; and Quiz Maker, though linguistically robust in parsing, is not knowledge-grounded in disambiguation. These structural weaknesses are further exposed in the Precision-recommendation curve, where KGMCQ maintains semantic integrity over increasing output volumes, while others exhibit rapid declines. Ultimately, KGMCQ’s hybrid architecture, recursive feedback, real-time similarity computations, and single-step distractor-key generation define it as a semantically grounded, pedagogically valuable, and technically scalable solution for automated MCQ generation in the Web 3.0 educational landscape.

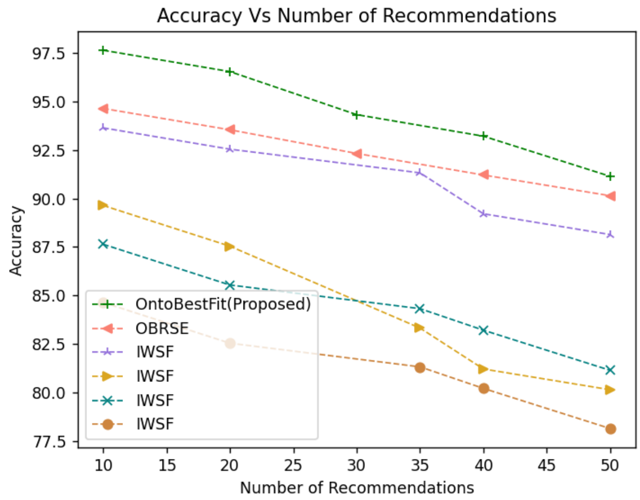

Fig.2. Precision Vs Percentage Vs No. of Recommendations

Figure 2 illustrates a line graph comparing the accuracy versus number of recommendations across multiple MCQ generation models, including the proposed KMCQ framework (denoted as OntoBestFit) and several baseline methods collectively labeled under different IWSF variants and the OBRSE model. As evident from the graph, the proposed KMCQ model consistently outperforms all competing approaches across varying recommendation set sizes, occupying the topmost position in the accuracy hierarchy throughout the range from 10 to 50 questions. The accuracy of KMCQ starts at a leading 97.8% for 10 recommendations and maintains high performance with only a modest decline to 91.3% at 50 recommendations. In contrast, the next best performing model, OBRSE, begins with an accuracy of 94.6% and drops to 90.1%, while other IWSF variants demonstrate more significant accuracy degradation reaching as low as 77.9%. This descending trend in accuracy for baseline models underscores their limited scalability and weaker semantic coherence in distractor generation as the number of generated MCQs increases. The smooth and well-preserved path of the proposed KMCQ model demonstrates its semantic filtering ability, class similarity computation through APMI, and strong hybrid architecture that guarantees diversity as well as accuracy in generated questions. The visual hierarchy in the line graph accordingly supports the better generalization and robustness of the KMCQ framework, thereby making it the most reliable model for high-volume MCQ generation.

5. Conclusion

This paper introduces a knowledge map–based framework for automatic multiple-choice question generation, both adopting deep learning and lightweight machine learning models. The system adopts a recurrent neural network (RNN) to classify textual resources like eBooks, blogs, and other online learning material. At the same time, a light-weight classifier is used to effectively classify the internal dataset with low computational overhead. The knowledge map is of key importance in choosing and loading pertinent entities, which are user-interest aligned via feedback and interaction. These entities are semantically organized and updated through the map, further improving contextual awareness. Computation of similarity between classes of the loaded entities based on metrics such as APMI and cosine similarity also adds to strong feature choice and detailed classification. Through the synergy of fast computation of lightweight models with semantic richness of deep learning, this framework obtains a compromise between scalability and performance. Its incorporation of user feedback and semantic reasoning makes it a new, flexible, and computationally strong solution for intelligent MCQ generation in today's educational systems.

6. References

1) Ch, D. R., & Saha, S. K. (2018). Automatic multiple choice question generation from text: A survey. IEEE Transactions on Learning Technologies, 13(1), 14-25.
2) Nwafor, C. A., & Onyenwe, I. E. (2021). An automated multiple-choice question generation using natural language processing techniques. arXiv preprint arXiv:2103.14757.
3) Raina, V., & Gales, M. (2022). Multiple-choice question generation: Towards an automated assessment framework. arXiv preprint arXiv:2209.11830.
4) Liu, M., Rus, V., & Liu, L. (2017). Automatic Chinese multiple choice question generation using mixed similarity strategy. IEEE Transactions on Learning Technologies, 11(2), 193-202.
5) Hanczakowski, M., Beaman, C. P., & Jones, D. M. (2017). When distraction benefits memory through semantic similarity. Journal of Memory and Language, 94, 61-74.
6) Li, Y., Zhang, D., & Mu, Y. (2020). Visual-semantic matching by exploring high-order attention and distraction. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 12786-12795).
7) Aldabe, I., & Maritxalar, M. (2014). Semantic similarity measures for the generation of science tests in basque. IEEE transactions on Learning Technologies, 7(4), 375-387.
8) Xie, J., Peng, N., Cai, Y., Wang, T., & Huang, Q. (2021). Diverse distractor generation for constructing high-quality multiple-choice questions. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 30, 280-291.
9) Abu-Salih, B., & Alotaibi, S. (2024). A systematic literature review of knowledge graph construction and application in education. Heliyon, 10(3).
10) Chen, P., Lu, Y., Zheng, V. W., Chen, X., & Yang, B. (2018). Knowedu: A system to construct knowledge graph for education. Ieee Access, 6, 31553-31563.
11) Deng, Y. (2019). Construction of higher education knowledge map in university libraries based on MOOC. The Electronic Library, 37(5), 811-829.
12) Pei, P., Xuejing, D., & Deqing, Z. (2020, October). Construction of curriculum knowledge map based on ontology. In Proceedings of the 2020 2nd International Conference on Robotics, Intelligent Control and Artificial Intelligence (pp. 259-265).
13) Salur, M. U., & Aydin, I. (2020). A novel hybrid deep learning model for sentiment classification. IEEE Access, 8, 58080-58093.
14) Chen, X., Ouyang, C., Liu, Y., Luo, L., & Yang, X. (2018, September). A hybrid deep learning model for text classification. In 2018 14th international conference on semantics, knowledge and grids (SKG) (pp. 46-52). IEEE.
15) Akpatsa, S. K., Li, X., & Lei, H. (2021, July). A survey and future perspectives of hybrid deep learning models for text classification. In International Conference on Artificial Intelligence and Security (pp. 358-369). Cham: Springer International Publishing.
16) Prabhakar, S. K., & Won, D. O. (2021). Medical text classification using hybrid deep learning models with multihead attention. Computational intelligence and neuroscience, 2021(1), 9425655.
17) Kalagiakos, P. A. N. A. G. I. O. T. I. S. (2009). A semantic web intelligent tutoring system characteristics. In Proc. 5th International Conference on Educational Technologies.
18) Jovanović, J., Torniai, C., Gašević, D., Bateman, S., & Hatala, M. (2008, June). Leveraging the social semantic web in intelligent tutoring systems. In International Conference on Intelligent Tutoring Systems (pp. 563-572). Berlin, Heidelberg: Springer Berlin Heidelberg.
19) Anwar, A. A. (2022, October). A survey of semantic web (Web 3.0), its applications, challenges, future and its relation with Internet of things (IoT). In Web Intelligence (Vol. 20, No. 3, pp. 173-202). Sage UK: London, England: SAGE Publications.
20) Fırat, E. A., & Fırat, S. (2020). Web 3.0 in learning environments: A systematic review. Turkish Online Journal of Distance Education, 22(1), 148-169.

21) https://link.springer.com/article/10.1007/s00500-013-1141-4 MCQSR (Dataset)
22) https://link.springer.com/article/10.1007/s10489-022-03894-6 QDG (Dataset)
23) https://link.springer.com/chapter/10.1007/978-981-19-5037-7_37 Quiz Maker (Dataset)