Informatics and Control Problems 45 Issue 1 (2025)
journal homepage: www.icp.az

Application of Retrieval-Augmented Generation technology to teaching the subject of computer science

Aynur Aliyeva

Institute of Control Systems, Baku, Azerbaijan

ARTICLE INFO

Article history:
Received 03.04.2025
Received in revised form 16.04.2025
Accepted 21.04.2025
Available online 04.06.2025

Keywords:
Computer science
Adaptive learning
Retrieval-Augmented Generation
Natural Language Processing (NLP)
Large Language Models (LLM)

ABSTRACT

Retrieval-Augmented Generation (RAG) technology plays an important role in the field of artificial intelligence, combining the strengths of information retrieval systems with the generative capabilities of large language models. This study explores the application of RAG technology to further develop adaptive learning in computer science teaching. The application of RAG technology to computer science teaching will serve to overcome the limitations of traditional teaching methods through a question-based intelligent learning environment. The study was conducted on five computer science topics. A structured set of 20 questions with precise answers was developed for each topic. The structure based on these questions allows RAG technology to retrieve relevant information and generate specific explanations, which promotes deeper understanding and connections of computer science knowledge. The results show that RAG technology can significantly support learners in computer science teaching.

1. Introduction

Teaching computer science is very important in today's digital world. It not only equips students with the technical skills needed in a rapidly changing job market, but also helps them develop critical thinking and problem-solving skills that are essential in all disciplines. Since computer science is the foundation of all technological innovations, acquiring effective knowledge and skills in this field is a top priority for universities.

Currently, traditional methods of teaching computer science are inadequate in the process of adaptive learning. Since the materials are not developed in accordance with the knowledge and skills of each student, they can limit students' learning and cannot adapt to individual learning styles. In addition, students often have difficulty understanding abstract concepts without interactive and contextual support, which can lead to frustration and gaps in knowledge.

Latest advances in artificial intelligence offer promising solutions to these problems. One such innovation is Retrieval-Augmented Generation (RAG) technology, a hybrid AI model that combines the strengths of information retrieval systems with the dynamic response capabilities of generative language models [1-4]. RAG systems first retrieve relevant information from a database or collection of documents, then use a language model to generate coherent, context-aware responses. This enables a highly adaptive and personalized learning process that goes beyond static content.

E-mail address: aynur.aliyeva8020@gmail.com (A.E. Aliyeva)

www.icp.az/2025/1-09.pdf https://doi.org/10.54381/icp.2025.2.09

This study proposes the use of RAG technology to improve the teaching and learning of computer science through a question-and-answer system. Five topics were selected from among the computer science topics taught at universities, and an experiment was conducted. RAG technology was applied based on structured questions prepared for the topics, and accurate and contextual answers were obtained that were tailored to the needs of the students.

2. Background

The integration of artificial intelligence (AI) into education has increased significantly in recent years. AI systems are increasingly being used to personalize learning, assess student knowledge, and provide real-time feedback. Holmes (2021) [5] and Zawacki-Richter (2019) [6] and other researchers emphasize that AI, especially when incorporated into intelligent learning systems, can bridge learning gaps and make education more comprehensive and effective.

Retrieval-Augmented Generation (RAG) is a new AI approach that combines two separate components: a retriever and a generator. First introduced by Lewis et al. (2020) [7], RAG models extract relevant contextual documents from a corpus and synthesize this extracted information with the input query to generate responses. This hybrid architecture addresses a key limitation of purely generative models by basing responses on factual source data.

RAG has been successfully applied in areas such as customer support, legal document processing, and academic research (Yao et al., 2023) [8]. Its growing use in education is reflected in studies showing that building on structured knowledge bases can improve factual accuracy and reduce misinformation in AI responses (Zhou et al., 2022) [9].

Structured questions, i.e., questions that are designed to be focused and systematic, based on a predetermined logical and pedagogical structure, have long been a key element in pedagogy. They stimulate active recall, reinforce understanding, and help educators identify misconceptions. According to Bransford et al. (2000) [10], structured inquiry promotes critical thinking and metacognitive development. When combined with artificial intelligence systems, structured questions can be used for automated, formative assessment and personalized content delivery.

Integrating RAG into computer science teaching allows for real-time, topic-specific answers to student questions based on a curated database of question-answer pairs. This structured approach leverages RAG’s adaptive capabilities to provide pedagogical adaptation. By incorporating question-based learning into RAG-enabled systems, higher education institutions can create intelligent learning environments tailored to individual student needs.

3. Application of RAG technology to support intelligent learning systems

3.1. The role of RAG technology in education

Retrieval-Augmented Generation (RAG) is an advanced artificial intelligence framework that combines two main components. Unlike traditional language models, which generate responses based solely on pre-learned knowledge, RAG improves the accuracy of the output by actively retrieving relevant information from external sources (such as document databases or knowledge bases) before generating a response. This approach allows RAG systems to provide more accurate, relevant, and contextually appropriate responses [11-13].

The RAG process typically includes two main components:

1. Retriever – Module that searches a large repository of documents to find links that best match the user's query.

2. Generator – Language model that uses both extracted documents and its internal understanding to generate coherent, informative responses.

RAG technology is particularly effective in the learning process. It allows systems to tailor responses to individual students' questions, providing explanations based on reliable data.

The result is a dynamic, adaptive learning process that surpasses static explanations of educational materials, in which students can explore topics through natural interaction, receive detailed explanations, and deepen their understanding with the help of artificial intelligence [14, 15].

3.2. Analysis of the dataset used

For the effective application of RAG technology in computer science education, we selected five computer science topics taught at universities. The following are the five topics we selected:

1. Computer science and its components.

2. Stages of development of computer science.

3. Modern computer software and its components.

4. Operating systems and Windows OS.

5. Algorithms and programming.

For each topic, a structured set of 20 questions was developed, reflecting various subtopics, levels of difficulty, and learning objectives. Each question in the data set includes the following:

- A clear name or identifier (e.g., "History of Third Generation Computers");
- Question formulation (e.g., "What characterizes third generation computer technology?");
- A correct and validated correct answer suitable for use in teaching by a subject expert;
- Related theoretical chunks of text;
- Title and distance values for those chunks.

This structured dataset forms the basis of the RAG system's retrieval component. When a question related to the subject matter is asked, the model searches the question pool and its accompanying collection of documents for relevant information before generating a personalized, understandable, and pedagogically sound response.

The methodology focuses not only on the accuracy of content, but also on student engagement and adaptability, making it a promising AI-based model for use in computer science education.

The dataset we use is divided into chunks as illustrated in Fig.1.

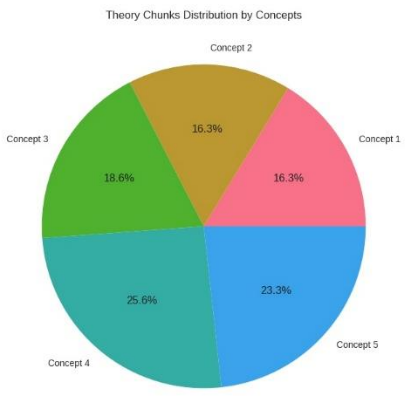

Fig. 1. Theory Chunks Distribution by Concept

Chunking in RAG is breaking up concepts (topics) into smaller, manageable fragments. These fragments are designed to help the model perform better in the search and generation stages. Each concept (topic) is divided into subheaders, and each section is entered into the RAG system as a "chunk". The question asked by the student is answered based on the corresponding chunks.

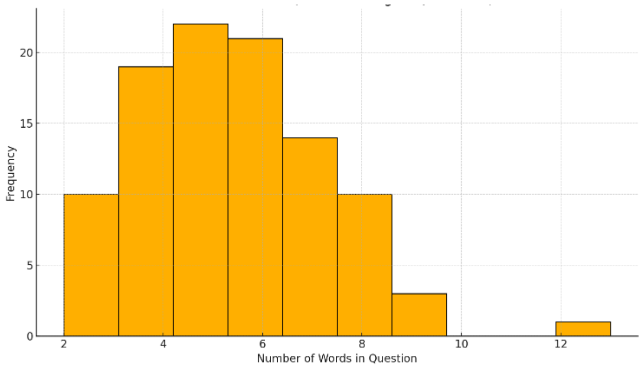

Fig. 2. Distribution of Question Lengths (in words)

Fig. 2 shows the distribution of question lengths in words. Most questions are within the range of 5–10 words. This indicates that a structured and short question model is used. The question asked by the student is processed by the system, and an answer based on the related concept is provided. After studying the compilation of the dataset, to examine the principle of the model's operation, we will describe the mathematical expression of the principle of the Retrieval-Augmented Generation model in the next section.

4. Model for applying RAG in computer science teaching

4.1. Mathematical formalization of the RAG model

In this section, we will demonstrate how RAG technology is applied to five topics in computer science education. Selected questions covering each topic constitute the input data for the RAG system, which finds and generates answers tailored to the student's needs. The RAG (Retrieval-Augmented Generation) model, unlike traditional generation models, is based not only on learned information but also on knowledge obtained from external databases. The structure of this model consists of two main components: the search block and the answer generation block [16-20].

The function of the retrieval component (search block) is to find matching documents from an external database (most often a vector database) according to a given query. Let us denote the question or learning query given by the student by Q il. $D = \{d_1, d_2, \dots, d_N\}$ is a database or collection of resources that contains the teaching topics of computer science. The search function $R(Q, D)$ that selects documents that match the query finds teaching materials or examples that match the student's question. The most relevant k number of topics or sources, $TopK\big(R(Q,D)\big) = \{d'_1, d'_2, \dots, d'_N\}$, are selected in response to the student's query.

The generation component (answer generation block) is the generation function $G(Q, \{d'_i\})$ that generates an answer to the query Q based on the selected documents $\{d'_i\}$, and the answer A, which is the output of the model, is the final answer obtained by the model.

For a query Q, the probability of generating an answer A, aggregated over all found documents, is $P(A|Q)$:

$$P(A|Q) = \sum_{i=1}^{k} P(d'_i|Q) * P(A|Q, d'_i).$$

Here, $P(d'_i|Q)$ is the probability of matching document d'_i to query Q, and $P(A|Q, d'_i)$ is the probability of generating response A based on the given query Q and document.

The application of Retrieval-Augmented Generation (RAG) technology to computer science teaching makes it possible to overcome the long-standing limitations of traditional teaching methods. By using artificial intelligence to obtain relevant information in real time, RAG technology creates a more adaptive and effective learning environment. RAG systems allow students to receive support on topics they find difficult according to their individual characteristics and to review topics as needed. Unlike textbooks, RAG offers a dynamic approach to learning that adapts to each student's learning abilities. Traditional learning is based on pre-written lectures, which can limit interactivity and engagement. RAG offers an interactive question-and-answer structure that allows students to ask questions in natural language and receive clear, informative answers immediately. This conversational model mimics human learning, promoting deeper understanding and sustained interest. Since the RAG system consists of a live or regularly updated knowledge base, the information provided to students is always up to date. This is particularly useful in the rapidly changing and evolving field of computer science. RAG ensures that students receive the most current and accurate information available.

4.2. General architecture of the system

The application of the RAG system in the teaching of computer science consists of two main stages: the preparation and processing of information (Fig. 3) and the RAG process chain (Fig. 4).

As shown in Fig. 4-d, the first stage of the system is the process of structuring and vectorizing information. This stage consists of three main components:

1. Input data: The system works with two types of data: theories and questions. The theories consist of teaching materials covering various topics of computer science, while the questions are a structured dataset of 20 questions prepared for each topic.
2. Indexing and chunking: At this stage, both the headers and content of the theoretical materials, as well as the questions, are processed separately. The theoretical materials are divided into header and content components, which enables more accurate searching.
3. Vector representation of texts: Using OpenAI's “text-embedding-3-large” model, all text components are converted into 3072-dimensional vectors. These vectors represent the semantic content of the texts in mathematical form.
4. Storage: Created vectors are stored in two separate collections in ChromaDB: the header collection and the content collection. Questions in JSON format are also stored together with their vectors.

In this section, teaching materials are analyzed and structured. The theoretical materials divided into topics are first divided into small text fragments called chunks. Each chunk is arranged in a way that is relevant in terms of concept and content. These pieces are then converted to vector format and indexed in a vector database. At the same time, question-answer pairs are formalized for each topic and used as a basis for evaluation.

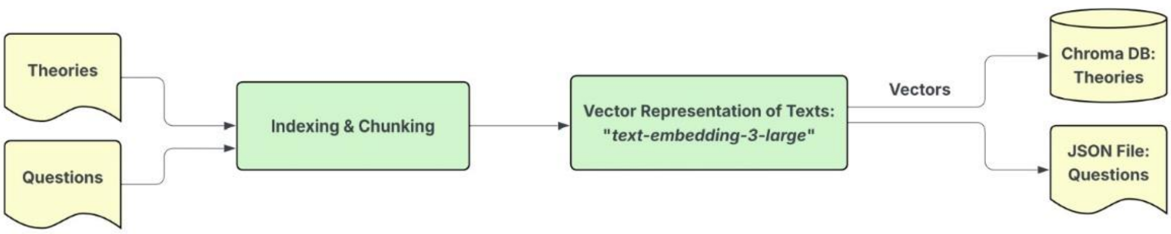

Fig. 3. Preparation and processing of information

Fig. 3. shows the basic operating principle of the RAG system which consists of five stages:

1. Starting stage: Student's query is accepted by the system from the JSON file and its pre-prepared vector is used.

2. Retrieval: Semantic search is performed in collections stored in ChromaDB. The system finds the conceptual segments that are most relevant to the query vector in both the header and content collections.

3. Prompting (Augmentation): The found theoretical materials are combined with the student's question to create a contextual prompt for the GPT model.

4. Generation: GPT-4o-mini model generates responses in the Azerbaijani language.

5. Storage: The generated answer is saved in Excel format for subsequent evaluation.

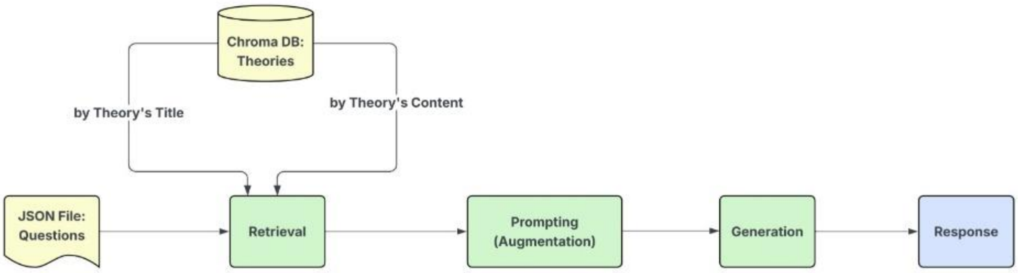

Fig. 4. RAG prosses chain

The found concept information is transferred to the generation module, where the answer is formed. As a result, a context-based, accurate, and pedagogically appropriate answer is presented to the student.

5. Evaluation criteria

To evaluate the application of RAG technology in the teaching of computer science, a dataset-based experiment was conducted and the results were analyzed using a structured question-answer database. Each question in the dataset was compared with the answers generated by the model. During the evaluation, the correctness of the answer (content relevance) and the accuracy of the terminology were taken into account.

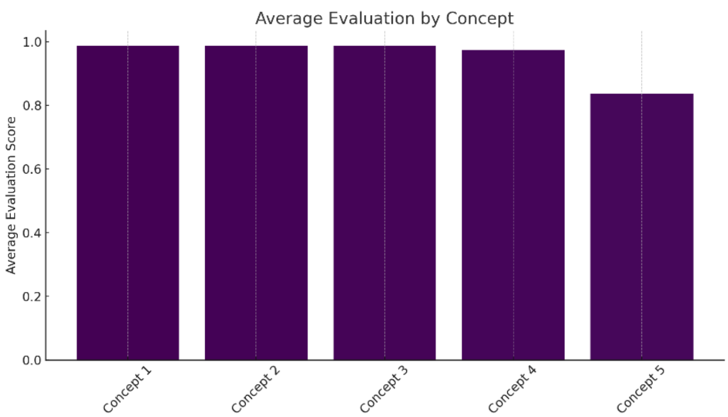

Fig. 5. Average evaluation by concept

Fig. 5 illustrates the results of the average scores given by students and experts. This diagram provides a visual comparison of the RAG system's performance in different topics (chunks). The chunks are rated on a scale of [0,1] (0, 0.25, 0.5, 0.75, 1).

Based on the overall analysis, approximately 90–92% of the answers were found to be correct or highly relevant. The dataset showed that the RAG model generates answers consistent with the materials previously entered into the system and provides accurate information without deviating from the context. Although there are some differences in terminology and synonyms in some cases, these answers are acceptable from a pedagogical point of view.

6. Conclusion

The article demonstrates the significant impact of applying Retrieval-Augmented Generation technology to the teaching process of computer science on the development of artificial intelligence-based learning. By combining the capabilities of information retrieval with generative language models, RAG creates a personalized, interactive, and reliable content-based teaching process.

This study demonstrates how the structured, question-based methodology can be effectively applied to the teaching of computer science using RAG. The system supports a deeper comprehension of complex concepts and allows students to monitor their learning process.

It reduces the workload of teachers and educators in educational institutions and creates an interactive and adaptive learning environment for students. It also helps students develop their knowledge and skills in a more accessible way.

Furthermore, it supports the development of students' logical thinking, information retrieval, and problem-solving skills.

It should be noted that the future application of RAG technology in other fields, the expansion of multilingual support, and its integration with learning management systems (LMS) will open up new horizons in this area.

References

[1] P. Lewis, E. Perez, A. Piktus, et al., Retrieval-Augmented generation for knowledge-intensive NLP tasks, Advances in Neural Information Processing Systems (NeurIPS 2020). https://arxiv.org/abs/2005.11401

[2] G. Chalkiadakis, N. Sidiropoulos, G. Tzagkarakis, AI in education: current trends and future directions, International journal of artificial intelligence in education, (2021).

[3] R. Luckin, W. Holmes, M. Griffiths, L.B. Forcier, Intelligence unleashed: An argument for AI in education, Pearson education white paper, (2016).

[4] S. Grover, R. Pea, Computational thinking in K–12: A review of the state of the field, Educational researcher. 42 No.1 (2013) pp.38-43.

[5] W. Holmes, M. Bialik, C. Fadel, Artificial intelligence in education: Promises and implications for teaching and learning, Center for curriculum redesign, (2021).

[6] O. Zawacki-Richter, V.I. Marín, M. Bond, F. Gouverneur, Systematic review of research on artificial intelligence applications in higher education – where are the educators? International journal of educational technology in higher education, 16 No.1 (2019) p.39. https://doi.org/10.1186/s41239-019-0171-0

[7] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, S. Riedel, Retrieval-Augmented generation for knowledge-intensive NLP tasks, Advances in neural information processing systems. 33 (2020) pp.9459-9474. https://arxiv.org/abs/2005.11401

[8] L. Yao, D. Deng, Y. Chen, Z. Lin, N. Duan, M. Zhou, A survey on retrieval-augmented generation, (2023). arXiv preprint arXiv:2303.16896. https://arxiv.org/abs/2303.16896.

[9] K. Zhou, S. Prabhumoye, G. Neubig, (2022). Learning to reduce hallucination from external knowledge for open-domain conversations, Proceedings of the 60th annual meeting of the association for computational linguistics (ACL 2022), 5287-5300. https://aclanthology.org/2022.acl-long.362

[10] J.D. Bransford, A.L. Brown, R.R. Cocking, (Eds.), (2000). How people learn: Brain, mind, experience, and school (Expanded ed.). Washington, DC: National academy press. https://nap.nationalacademies.org/catalog/9853/how-people-learn-brain-mind-experience-and-school-expanded-edition

[11] R.S. Baker, Stupid tutoring systems, intelligent humans, International journal of artificial intelligence in education. 26 No.2 (2016) pp.600-614.

[12] A. Oleynik, A.M. Fedorov et al., Using rag technology to design an intelligent information system for support exploratory search, Transaction Kola Science Centre. 15 No.3 (2025) pp.5-26. doi:10.37614/2949.1215.2024.15.3.001.

[13] P. Lomov, M. Malozemova, M. Shishaev, Extracting relations from NER-tagged sentences for ontology learning, Artificial intelligence trends in systems: Proceedings of 11th computer science on-line conference, Cham: Springer. 2 (2022) pp.337-344.

[14] L. Wang et al., Genre: Generative multi-turn question answering with contrastive learning for entity–relation extraction, Complex & Intelligent Systems. 10 (2024) pp.3429-3443.

[15] Y. Gao et al., (2023) Retrieval-Augmented generation for large language models: A Survey, arXiv:2312.10997 [cs].

[16] P. Zhao, H. Zhang, Q. Yu, Z. Wang, Y. Geng, F. Fu, L. Yang, W. Zhang, J. Jiang, B. Cui, Retrieval-Augmented generation for AI-Generated content: A Survey. arXiv 2024, arXiv:2402.19473.

[17] A. Bora, H. Cuayáhuitl, Systematic analysis of retrieval-augmented generation-based LLMs for medical chatbot applications, Mach. Learn. Knowl. Extr. 6 (2024) pp.2355-2374.

[18] J.Q. Pérez, T. Daradoumis, J.M.M. Puig, Rediscovering the use of chatbots in education: A systematic literature review, Comput. Appl. Eng. Educ. 28 (2020) pp.1549-1565.

[19] S. Wollny, J. Schneider, D. Di Mitri, J. Weidlich, M. Rittberger, H. Drachsler, Are We There Yet? – A systematic literature review on chatbots in educ ation, Front. Artif. Intell. (2021).

[20] K. Pichai, (2023). A retrieval-augmented generation based large language model benchmarked on a novel dataset, Journal of student research. 12 No.4. https://doi.org/10.47611/jsrhs.v12i4.6213