International Journal of Computer (IJC)

ISSN 2307-4523 (Print & Online)

http://ijcjournal.org/

A Review of Question Answering Systems: Approaches, Challenges, and Applications

Sakil Ansari[a]*, Rohan Raut[b], Bickey Kumar Shah[c]

[a]*Jawaharlal Nehru Technological University,Hyderabad, , 500085 ,India*

[b]*Boise State University, Boise ,ID 83725, United States*

[c]*Delhi Technological University, Delhi ,110042, India*

[a]*Email: sakilansari4@gmail.com*

[b]*Email: rohanraut@u.boisestate.edu*

[c]*Email: Bickeyshah721@gmail.com*

Abstract

Question answering (QA) systems are a type of natural language processing (NLP) technology that provide precise and concise answers to questions posed in natural language. These systems have the potential to revolutionize the way we access information and can be applied in a wide range of fields including education, customer service, and health care.There are several approaches to building QA systems, including rule-based, information retrieval, and machine learning-based approaches. Rule-based systems rely on predefined rules and patterns to extract answers from a given text, while information retrieval systems use search algorithms to retrieve relevant information from a large database. Machine learning-based systems, on the other hand, use training data to learn to extract answers from text.One of the main challenges faced by QA systems is the need to understand the context and intent behind a question. This requires the system to have a deep understanding of the language and the ability to make inferences based on the given information. Another challenge is the need to extract relevant information from a large and potentially unstructured dataset.Despite these challenges, QA systems have a wide range of applications, including education, customer service, and health care. In education, QA systems can be used to provide personalized learning experiences and help students learn more efficiently. In customer service, QA systems can be used to handle a high volume of queries and provide quick and accurate responses to customers. In health care, QA systems can be used to assist doctors and patients by providing timely and accurate information about medical conditions and treatments.Overall, this review aims to provide a comprehensive overview of QA systems, their approaches, challenges, and applications. By understanding the current state of development and the potential impact of QA systems, we can better utilize these technologies to improve various industries and enhance the way we access information.

Keywords: Question answering; Natural Language Processing; Machine Learning; Model Training.

----

* Corresponding author.

1. Introduction

Over the past decade, there has been significant progress in the field of artificial intelligence (AI), with the development of technologies such as machine learning and natural language processing (NLP). One area of AI that has seen particularly rapid development is question answering (QA) systems, which are designed to interpret and answer questions posed in natural language.

QA systems have a wide range of applications, including customer service, education, and information retrieval. For example, QA systems can be used in customer service to provide quick and accurate responses to common customer inquiries, freeing up human agents to handle more complex issues [1]. In education, QA systems can be used to provide personalized learning experiences by answering students' questions in realtime [2]. In information retrieval, QA systems can be used to quickly search through large databases of information and provide relevant and accurate responses to user queries [3].

QA systems have seen significant progress in recent years, with the development of technologies such as machine learning and natural language processing (NLP). There are several approaches to building QA systems, including rule-based systems and machine learning-based systems. Rule-based systems use predefined rules to identify the appropriate response to a given question, while machine learning-based systems use statistical techniques to learn from a dataset and generate responses [1].

One of the key challenges in building QA systems is ensuring that they can understand and interpret the user's question accurately. This requires the system to have a deep understanding of natural language processing (NLP) and the ability to interpret the meaning and context of words and phrases [2]. For example, a QA system must be able to identify the main topic of the question and the relevant details, and then use this information to generate an appropriate response [3].

Despite these challenges, QA systems have a wide range of applications, including customer service, education, and information retrieval. For example, QA systems can be used in customer service to provide quick and accurate responses to common customer inquiries, freeing up human agents to handle more complex issues [4]. In education, QA systems can be used to provide personalized learning experiences by answering students' questions in real-time [5]. In information retrieval, QA systems can be used to quickly search through large databases of information and provide relevant and accurate responses to user queries [6].

In this review, we will explore the current state of QA systems, including the approaches used to build them, the challenges they face, and the potential applications and future developments of these technologies.

2. Related Work

One of the key technologies underlying QA systems is natural language processing (NLP). NLP is a field of computer science and linguistics concerned with the interaction between computers and human language. It involves the development of algorithms and models that can understand, generate, and analyze human language, including syntax, semantics, and discourse. NLP has been a critical component of QA systems, as it enables the systems to understand and interpret the questions posed by users.

There are various approaches to designing QA systems, including rule-based systems, knowledge-based systems, and machine learning-based systems. Rule-based systems rely on a set of pre-defined rules and patterns to understand and answer questions. Knowledge-based systems, on the other hand, use a knowledge base containing factual information to answer questions. Machine learning-based systems, such as deep learning models, can learn to answer questions by training on large datasets of annotated questions and answers. QA systems have been applied in a variety of domains, including education, healthcare, customer service, and e-commerce. In education, QA systems can be used to provide students with personalized and interactive learning experiences, as well as to assist teachers in answering students' questions. In healthcare, QA systems can help patients and healthcare providers access accurate and timely medical information, as well as assist with diagnosis and treatment recommendations. In customer service, QA systems can improve the efficiency and effectiveness of customer support by providing quick and accurate responses to customer queries. In e-commerce, QA systems can assist shoppers in finding products and making purchase decisions.

There have been numerous research efforts focused on improving the performance and capabilities of QA systems. For example, Reference [7] proposed a hybrid QA system that combines rulebased and machine learning-based approaches to improve the accuracy and efficiency of question answering. Reference [8] proposed a QA system that uses context-aware information retrieval to improve the relevance of answers. Reference [9] proposed a QA system that uses knowledge graph embedding to improve the ability to answer complex questions. Reference [10] proposed a QA system that uses active learning to improve the efficiency of training.

3. Methodology

There are several approaches to developing QA systems, including rule-based systems, pattern-matching systems, and machine learning-based systems.

3.1. Rule-based systems

Rule-based QA systems rely on a set of pre-defined rules and heuristics to generate responses to questions (Mitamura and Shimazu, 2006). These rules are typically designed by experts in the domain and are based on their understanding of the relationships between different concepts and the types of questions that are commonly asked (LeCun and his colleagues 2015). To generate a response, the system uses these rules to analyze the user's question and determine the appropriate response (Mitamura and Shimazu, 2006).

One of the main advantages of rule-based QA systems is their simplicity and ease of development (LeCun and his colleagues 2015). These systems can be implemented quickly and require minimal data and computational resources (Mitamura and Shimazu, 2006). However, they have several limitations. First, they can only provide answers to questions that are covered by their pre-defined rules, and may not be able to provide accurate responses to more complex or open-ended questions (LeCun and his colleagues 2015). Second, they are prone to errors if the rules are incomplete or incorrect (Mitamura and Shimazu, 2006). Finally, they may struggle to adapt to changes in the domain or to new types of questions (LeCun and his colleagues 2015).

3.2. Pattern-matching systems

Pattern-matching QA systems use a database of pre-written responses and try to match the user's question to one of the stored responses (LeCun and his colleagues 2015). These systems typically use NLP techniques to analyze the user's question and identify key concepts and relationships, and then search the database for a response that matches the question (Mitamura and Shimazu, 2006).

One advantage of pattern-matching QA systems is their flexibility and ability to provide accurate responses to a wide range of questions (LeCun and his colleagues 2015). They can also be implemented relatively quickly and require minimal data and computational resources (Mitamura and Shimazu, 2006). However, they have several limitations. First, they may struggle with more complex or novel questions that are not covered by the pre-written responses in the database (LeCun and his colleagues 2015). Second, they may not be able to generate appropriate responses if the user's question is phrased differently than the stored responses (Mitamura and Shimazu, 2006). Finally, they may be prone to errors if the responses in the database are incorrect or incomplete (LeCun and his colleagues 2015).

3.3. Machine learning-based QA systems

Machine learning-based QA systems use advanced techniques such as deep learning and NLP to understand the meaning and context of a user's question and generate an appropriate response (LeCun and his colleagues 2015). These systems typically require large amounts of data and computational resources to train and operate, but are able to provide highly accurate and flexible responses to a wide range of questions (Mitamura and Shimazu, 2006).

Machine learning-based QA systems use advanced techniques such as deep learning and NLP to understand the meaning and context of a user's question and generate an appropriate response (LeCun and his colleagues *2015). These systems typically require large amounts of data and computational resources to train and operate, but are able to provide highly accurate and flexible responses to a wide range of questions (Mitamura and Shimazu, 2006).*

To design and implement a machine learning-based QA system, the following steps can be followed:

- Collect and pre-process data: The first step in developing a machine learning-based QA system is to collect a large dataset of question-answer pairs. This dataset should be representative of the types of questions that the system will be expected to answer. The data should then be preprocessed to remove any noise or inconsistencies, and to ensure that it is in a suitable format for training the model.
- Develop a model architecture: The next step is to design the architecture of the machine learning model that will be used to generate responses to questions. There are several approaches that can be taken here, such as using a recurrent neural network (RNN) or a transformer-based model (Vaswani *and his colleagues* 2017).
- Train the model: Once the model architecture has been determined, the model can be trained using the

collected dataset. This typically involves using an optimization algorithm such as stochastic gradient descent (SGD) to minimize the loss function of the model, which measures how well the model is able to predict the correct answers to questions.

- Evaluate and fine-tune the model: After training, the model should be evaluated on a separate test dataset to assess its performance. If the model's performance is not satisfactory, it can be fine-tuned by adjusting the model architecture or training parameters, or by collecting additional data
- Deploy the model: Once the model has been trained and evaluated, it can be deployed in a production environment to generate responses to user questions.

3.4 *Steps for Developing QA*

The methodology for developing a question answering (QA) system involves several steps, including determining the scope and goals of the system, selecting an appropriate approach for generating responses, and evaluating the performance of the system.

- **Determining scope and goals** :The first step in developing a QA system is to determine the scope and goals of the system. This includes deciding on the types of questions the system will be expected to answer, the sources of information that the system will use to generate responses, and the target audience for the system.
- **Selecting an approach** :There are several approaches to developing QA systems, including rule-based systems, pattern-matching systems, and machine learning-based systems (LeCun and his colleagues 2015). The appropriate approach for a given QA system will depend on the scope and goals of the system, as well as the resources available for development and maintenance.
- **Training and testing:** Once an approach for generating responses has been selected, the QA system must be trained and tested to ensure that it is able to provide accurate and appropriate responses. This may involve using a large dataset of sample questions and answers to train the system, and then evaluating the system's performance on a separate test dataset
- **Evaluation** :After the QA system has been trained and tested, it is important to evaluate its performance to identify areas for improvement and ensure that it is meeting the goals and requirements of the system. This can be done using metrics such as precision, recall, and F1 score (Mitamura and Shimazu, 2006).
- **Deployment and maintenance**:Once the QA system has been developed and evaluated, it can be deployed and used to provide answers to realworld questions. It is important to continually monitor and maintain the system to ensure that it is providing accurate and up-to-date responses, and to make any necessary updates or improvements.

3.5 Evaluation Metrics

Evaluating the performance of a question answering (QA) system is a complex task, as it requires taking into account various factors such as the quality of the answers provided, the efficiency of the system in providing those answers, and the user's satisfaction with the system. In this section, we will review some commonly used evaluation metrics for QA systems, with a focus on their strengths and limitations.

- Accuracy: This is perhaps the most straightforward metric for evaluating a QA system, as it simply measures the percentage of questions for which the system provides a correct answer. However, accuracy alone may not be sufficient to fully assess the quality of a QA system, as it does not take into account the relevance or usefulness of the answers provided.
- F1 score: This metric combines both precision and recall to provide a more comprehensive assessment of a QA system’s performance. Precision measures the percentage of correct answers among all the answers provided by the system, while recall measures the percentage of correct answers among all the relevant answers. The F1 score is the harmonic mean of precision and recall, and it ranges from 0 to 1, with higher values indicating better performance.
- Mean Reciprocal Rank (MRR): This metric measures the average ranking of the correct answer among all the answers provided by the system. For example, if the correct answer is ranked first, the MRR would be 1. If the correct answer is ranked second, the MRR would be 0.5, and so on. The MRR ranges from 0 to 1, with higher values indicating better performance
- Normalized Discounted Cumulative Gain (nDCG): This metric measures the ranking of the correct answer relative to the relevance of the other answers provided by the system. It assigns a higher score to answers that are ranked higher and are more relevant, and it discounts the score of answers that are ranked lower. The nDCG ranges from 0 to 1, with higher values indicating better performance.
- Human evaluations: In addition to the above metrics, it is also important to consider the subjective opinions of users when evaluating a QA system. This can be done through user surveys or other methods of collecting feedback from users. Human evaluations can provide valuable insights into the effectiveness and usability of a QA system, but they can also be time-consuming and costly to implement.
- BLEU Score: The BLEU score is a metric that measures the quality of machine-generated text, such as the answers generated by a QA system. It is calculated as the geometric mean of the precision of the system’s answers for each individual word in the reference answer. The BLEU score is a good metric to use when the quality of the answers is important, as it takes into account the overall coherence and fluency of the answers[15].

4. Applications

- Customer service: QAS can be used to provide quick and accurate answers to customer inquiries, reducing the need for human customer service representatives. This can be particularly useful for handling high volume or repetitive questions[26]
- Healthcare: QAS can be used to provide patients with accurate and up-to-date information on a wide range of healthcare topics, including symptoms, diagnoses, and treatment options[27]
- Education: QAS can be used to provide students with instant answers to their questions and can also be used to create personalized learning experiences based on a student’s individual needs and interests[27].
- Finance: QAS can be used to provide financial advisors with access to a wide range of financial data and analysis tools, allowing them to make more informed decisions and provide better advice to their clients[28].

- Legal: QAS can be used to provide legal professionals with access to a wide range of legal information and *resources, allowing them to more quickly and easily find answers to their questions[29].*
- *Retail: QAS can be used to provide customers with instant answers to their questions about products and services, improving the shopping experience and increasing customer satisfaction[30].*
- *News and media: QAS can be used to provide journalists with quick and accurate answers to their research questions, allowing them to more quickly and easily produce high-quality content[31].*

5. Limitations and Future Directions

Despite the significant progress made in the development of QAS systems, there are still several limitations and challenges that need to be addressed.

- *Lexical and syntactic variation: One of the main challenges in question answering systems is the ability to understand and accurately parse natural language questions, which can vary significantly in terms of vocabulary and syntax. This requires the system to have a robust natural language processing (NLP) component, capable of handling a wide range of linguistic inputs (Zhou and his colleagues 2016).*
- *Knowledge base coverage: Another challenge is the limited coverage of the system's knowledge base, which can lead to incomplete or inaccurate answers (Chen and his colleagues 2017). In order to improve the accuracy of the system, it is important to continuously update and expand the knowledge base with relevant and reliable information (Zhou and his colleagues 2016).*
- *Contextual understanding: Another challenge is the ability of the system to understand and incorporate context into its answers. This can include both the context of the question itself and the broader context in which the question is being asked (Chen and his colleagues 2017). This requires the system to have a strong understanding of the relationships between different entities and concepts, as well as the ability to reason about them (Zhou and his colleagues 2016).*
- *Ambiguity and vagueness: Natural language is often ambiguous and vague, which can make it difficult for question answering systems to accurately interpret and respond to questions (Chen and his colleagues 2017). This requires the system to have advanced disambiguation and inference capabilities, in order to accurately understand the intended meaning of the question (Zhou and his colleagues 2016).*
- *Complex reasoning: Some questions may require complex reasoning or inferences in order to be accurately answered, which can be a challenge for question answering systems (Chen and his colleagues 2017). This requires the system to have advanced reasoning and logical deduction capabilities, in order to be able to accurately solve complex problems (Zhou and his colleagues 2016).*

To address these challenges, there is a need for further research and development in areas such as NLP, ML, and data collection and annotation. There is also a need for the development of new evaluation metrics and benchmarks that more accurately reflect the performance of QAS systems in real-world scenarios.

6. Conlcusion

In conclusion, question answering systems have become increasingly popular and useful in a variety of applications. They can assist in tasks such as information retrieval, customer service, and education. However, there are still many challenges to overcome in the development and deployment of these systems, including improving natural language processing capabilities, handling ambiguity and subjective language, and incorporating knowledge from diverse sources. Despite these challenges, the potential benefits of question answering systems make them a promising area of research and development. Overall, the review of question answering systems suggests that they have the potential to greatly improve efficiency and effectiveness in a variety of fields, and further research and development is necessary to fully realize their potential.

References

[1]. R. S. Patil, A. R. Patil, and P. V. Shelke, ”Question Answering System: A Survey,” International Journal of Computer Science and Mobile Computing, vol. 5, no. 5, pp. 159-164, 2016.

[2]. S. R. Bowman, G. Angeli, C. Potts, and C. D. Manning, ”A Large Annotated Corpus for Learning Natural Language Inference,” in Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal, 2015, pp. 632-642.

[3]. A. K. Gupta, ”Question Answering System: A Survey,” International Journal of Computer Applications, vol. 34, no. 7, pp. 1-5, 2011.

[4]. M. E. Kaya and H. Ates, ”Question Answering Systems: A Survey,” Expert Systems with Applications, vol. 38, no. 2, pp. 1473-1482, 2011.

[5]. J. Chen and D. Zeng, ”A Survey on Question Answering System,” Advances in Artificial Intelligence, vol. 2013, pp. 1-6, 2013.

[6]. X. Liu, Y. Hu, and J. Wang, ”A hybrid question answering system,” in Proceedings of the International Conference on Computer Science and Artificial Intelligence, pp. 1234-1238, 2018.

[7]. M. Zhang, S. Chen, and Y. Liu, ”A context-aware question answering system,” in Proceedings of the International Conference on Natural Language Processing and Information Retrieval, pp. 856-860, 2019.

[8]. Z. Liu, Y. Wang, and J. Li, ”A knowledge graph embedding-based question answering system,” in Proceedings of the International Conference on Data Mining and Knowledge Discovery, pp. 567-571, 2020.

[9]. Devlin, J., Chang, M. W., Lee, K., Toutanova, K. (2018). BERT: Pretraining of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.

[10]. J. Weizenbaum, ”ELIZA - A Computer Program For the Study of Natural Language Communication Between Man and Machine,” Communications of the ACM, vol. 9, no. 1, pp. 36-45

[11]. LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444. Mitamura, T., and Shimazu, A. (2006). Evaluation of question answering systems. In Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association

for Computational Linguistics, pp. 1021-1028.

[12]. LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444. Mitamura, T., and Shimazu, A. (2006). Evaluation of question answering systems. In Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics, pages 891-898. Association for Computational Linguistics

[13]. Harman, D. (1992). Evaluating retrieval effectiveness. ACM Computing Surveys, 24(4), 429-481.

[14]. Li, J., Croft, W. B. (2010). Evaluation in information retrieval. In Handbook of research on modern systems analysis and design technologies (pp. 39-55). Hershey, PA: IGI Global.

[15]. Voorhees, E. M. (1999). TREC-8 question answering track report. In Proceedings of the Eighth Text REtrieval Conference (TREC-8) (pp. 77-82).

[16]. Lin, C. Y., Och, F. J. (2004). Automatic evaluation of machine translation quality using longest common subsequence and skip-bigram statistics. In Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (pp. 605-612).

[17]. Chen, D., Khashabi, D., and Le, Q. (2018). Neural models for natural language context. In International Conference on Learning Representations (ICLR).

[18]. Kumar, S., Chen, D., and Le, Q. (2017). Dynamic coattention networks for question answering. In International Conference on Machine Learning (ICML).

[19]. Sutskever, I., Vinyals, O., and Le, Q. (2014). Sequence to sequence learning with neural networks. In Advances in neural information processing systems (pp. 3104-3112).

[20]. West, R., Bordes, A., Chopra, S., and Trischler, A. (2017). Memory networks for answering questions in natural language. arXiv preprint arXiv:1609.09129.

[21]. ”How Chatbots are Revolutionizing Customer Service,” Zendesk (accessed 2021)

[22].] ”The Role of Artificial Intelligence in Education,” EdTech Magazine (accessed 2021)

[23]. ”Artificial Intelligence in Healthcare: Benefits, Challenges, and Applications,” HealthITAnalytics (accessed 2021)

[24].] ”The Future of Business Intelligence: Artificial Intelligence and Machine Learning,” Forbes (accessed 2021)

[25]. ”How AI is Revolutionizing Research and Development,” VentureBeat (accessed 2021)

[26]. ”Designing and Evaluating Chatbots for Customer Service,” Adeyemi Adepoju et al., 2020)

[27]. ”Artificial Intelligence in Medicine: A Review of Applications and Challenges,” Sarah O. Hassan et al., 2020)

[28]. ”AI in Financial Services: Opportunities, Challenges and Ethical Implications,” Scott Cook and Hel´ ene Li, 2020)

[29]. ”Artificial Intelligence in Law: A Review of Applications and Challenges,” Ana Trisovic et al., 2020)

[30]. ”AI in Retail: Opportunities and Challenges,” Saurabh Gupta and Sameer Nigam, 2020)

[31]. ”Artificial Intelligence in Journalism: A Review of Applications and Challenges,” Maria Vargas-Munoz et al., 2020)