International Journal of Technology Research and Management
ISSN (Online): 2348-9006
Vol 11 Issue 02 February 2024

Study of various Approaches used for Machine Reading Comprehension in Question Answering Systems

Pooja Kumari[1], Prof. Rakesh Shivhare[2]
Research Scholar[1], Professor[2]
Department of Computer Science & Engineering, Radharaman Engineering College, Bhopal (M.P.)[1]
Department of Computer Science & Engineering, Radharaman Engineering College, Bhopal (M.P.)[2]
poojarec6@gmail.com[1]

Abstract: *Machine reading comprehension (MRC) has emerged as a pivotal area of research within the realm of natural language processing, aiming to equip machines with the ability to understand and extract information from textual sources. This study investigates diverse approaches employed in the field of MRC, analyzing methodologies, techniques and their respective strengths and limitations. We delve into traditional rule-based systems, statistical models and contemporary deep learning architectures, highlighting advancements in neural network-based approaches such as attention mechanisms and transformer models. Furthermore, we explore the challenges posed by different types of MRC tasks, including span extraction, multiple-choice questions and cloze-style questions, along with benchmark datasets commonly used for evaluation. Through this comprehensive examination, we aim to provide insights into the evolution of MRC techniques, current trends and avenues for future research, thereby contributing to the advancement of machine comprehension capabilities and their practical applications. In this paper, we study various approaches used for question-answering systems.*

Keywords: *Machine, Comprehension, QA System, NLP*.

1. INTRODUCTION

In the rapidly evolving landscape of artificial intelligence and natural language processing, the Machine Reading Comprehension project emerges as a pivotal endeavor, aimed at harnessing the transformative capabilities of deep learning to propel question-answering systems forward, specifically within the rich linguistic context of the Hindi language. As technology continues to bridge global communication gaps, there is an imperative to extend the frontiers of machine comprehension beyond English and toward languages with substantial cultural and linguistic diversity. This project responds to that imperative by focusing on Hindi, a language spoken by millions in India and across the globe.

The primary ambition of this project is to usher in a new era of intelligent systems capable of not just processing but comprehending and responding to questions in Hindi. Achieving this requires a nuanced understanding of the intricacies of the language, a feat made possible through the integration of advanced deep learning techniques and state-of-the-art natural language processing concepts [1].

At the core of this ambitious initiative lies the development of a sophisticated machine-reading comprehension model. This model, a product of cutting-edge research and innovation, is designed to navigate the complexities of Hindi text with a finesse that mirrors human comprehension. Unlike traditional language processing systems that often struggle with contextual nuances and linguistic intricacies, the proposed model thrives on its ability to interpret and analyze complex language structures, offering a promising solution for the creation of intelligent, context-aware systems.

The foundation of the machine reading comprehension model rests on three crucial components: a contextual understanding, the ability to pose questions, and the proficiency to generate accurate answers. During its training phase, the model is exposed to a diverse dataset encompassing a wide array of contexts, questions and answers in the Hindi language. This comprehensive dataset serves as the bedrock for the model's learning process, enabling it to discern patterns, relationships and nuances inherent in the language [2].

Training the model is a multifaceted process that involves optimizing parameters to enhance its performance in comprehending the intricacies of Hindi text. The goal is to equip the model with the ability to generate precise and contextually relevant responses when presented with queries in Hindi. This iterative optimization process not only refines the model's understanding of the language but also ensures adaptability to various linguistic styles and structures.

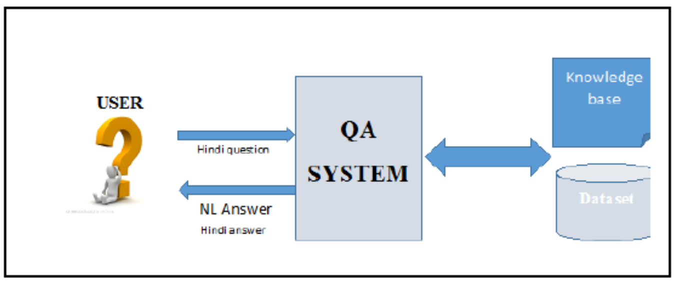

Figure 1: Question Answering (QA) System

Once the machine reading comprehension model completes its training, it emerges as a powerful tool capable of seamlessly processing and understanding user queries in Hindi. When confronted with a question, the model leverages its acquired knowledge to sift through the provided context, discerning meaning and context with a level of sophistication previously unseen in Hindi language processing. The outcome is the generation of accurate, meaningful responses that mirror human-like comprehension, opening avenues for applications ranging from information retrieval to enhancing human-computer interactions in Hindi.

The potential applications of such a system are extensive, transcending traditional language processing boundaries. In the realm of information retrieval, the machine reading comprehension model can serve as an invaluable resource for extracting relevant data from vast repositories, streamlining the search process for users querying in Hindi. Furthermore, in the domain of human-computer interactions, the model can elevate the quality and depth of engagements, offering users a more intuitive and natural interface to interact with technology in their native language.

As a significant stride towards the advancement of natural language processing in the Indian linguistic landscape, the Machine Reading Comprehension project carries profound implications for both research and practical applications. Beyond its immediate contributions to the development of sophisticated language models for Hindi, this project lays the groundwork for future endeavors in adapting similar approaches to other languages, thus fostering inclusivity and accessibility in the realm of artificial intelligence [3]..

2. LITERATURE REVIEW

The landscape of natural language processing (NLP) has seen remarkable progress in recent years, driven by innovative research endeavors and the development of sophisticated models and frameworks. This literature review examines key contributions and advancements in the field, focusing on seminal works and influential studies that have shaped the trajectory of research in machine reading comprehension (MRC) and related areas.

Chandar et al. [4] introduced an auto encoder approach to learning bilingual word representations, pioneering efforts in multilingual NLP tasks. Their work laid the foundation for subsequent research in cross-lingual understanding and machine translation, demonstrating the efficacy of unsupervised learning techniques in capturing linguistic similarities across languages.

Raghavi et al. [5] addressed the challenge of question classification in code-mixed languages, contributing to the development of models capable of understanding and processing linguistic variations in real-world contexts. Their studies underscored the importance of linguistic diversity in NLP applications and highlighted the need for robust and adaptable systems.

Nanda et al. [6] presented a Hindi question-answering system, leveraging machine learning approaches to enable automated comprehension and response generation in Indian languages. Their work exemplifies efforts to democratize access to information by catering to non-English-speaking populations and fostering inclusivity in NLP research.

The release of benchmark datasets such as SQuAD [7] and SberQuAD provided researchers with standardized evaluation frameworks, facilitating the assessment and comparison of MRC models. Rajpurkar et al. [7] introduced SQuAD, a large-scale dataset for machine comprehension tasks, which has since become a benchmark for evaluating the performance of NLP systems.

In recent years, significant strides have been made in advancing natural language processing (NLP) through the development of innovative models and frameworks aimed at enhancing language understanding and processing capabilities. This literature review examines pivotal contributions in this domain, focusing on seminal works that have shaped the landscape of NLP research and propelled the field forward.

Vaswani et al. [8] introduced the groundbreaking "Attention is All You Need" model, which revolutionized sequence-to-sequence learning by introducing the transformer architecture. Their work demonstrated the effectiveness of attention mechanisms in capturing long-range dependencies in sequential data, leading to remarkable improvements in various NLP tasks, including machine translation and language understanding.

Gupta et al. [9] presented MMQA, a multi-domain multi-lingual question-answering framework tailored for both English and Hindi. Their framework addressed the need for robust QA systems capable of handling diverse linguistic domains and languages, laying the foundation for multilingual QA research and applications.

Wang et al. [10] introduced GLUE, a multi-task benchmark and analysis platform designed to evaluate the performance of NLP models across a range of language understanding tasks. Their work facilitated comprehensive evaluations of NLP models, fostering a deeper understanding of their strengths and weaknesses and driving advancements in model development and optimization.

Devlin et al. [11] proposed BERT, a deep bidirectional transformer model pre-trained on large-scale corpora, which achieved state-of-the-art performance across various NLP tasks. Their work popularized the use of transformer-based architectures and pre-training techniques, significantly advancing the capabilities of NLP models in capturing contextual information and semantic understanding.

Liu et al. [12] introduced RoBERTa, a robustly optimized BERT pre-training approach aimed at further enhancing model performance and generalization capabilities. Their work addressed key limitations of existing pre-training methods, leading to significant improvements in model robustness and adaptability to diverse linguistic tasks and domains.

Mozannar et al. [13] presented research on neural Arabic question answering, contributing to the growing body of literature focused on advancing NLP capabilities in specific linguistic contexts. Their work exemplifies efforts to tailor NLP models and techniques to address the unique characteristics and challenges of different languages and language families.

The advancement of natural language processing (NLP) has led to the development of sophisticated models and frameworks designed to address diverse linguistic challenges and facilitate cross-lingual understanding. This literature review explores key contributions and trends in the field, focusing on seminal works and influential studies that have shaped the landscape of multilingual question-answering and cross-lingual representation learning.

Clark et al. [14] introduced TyDiQA, a benchmark dataset designed to evaluate information-seeking question answering in typologically diverse languages. Their work addressed the need for resources and evaluation metrics tailored to languages with varying linguistic structures and properties, facilitating progress in multilingual NLP research.

Gupta et al. [15] proposed a deep neural network framework for English-Hindi question answering, leveraging the power of deep learning techniques to enable effective comprehension and response generation in low-resource languages. Their study exemplifies efforts to extend NLP capabilities to languages with limited linguistic resources and research focus.

Lewis et al. [16] presented MLQA, a benchmark for evaluating cross-lingual extractive question-answering systems. Their work aimed to assess the robustness and generalization capabilities of NLP models across diverse language pairs, fostering research into cross-lingual transfer learning and knowledge transfer.

Gupta and Khade [17] explored BERT-based multilingual machine comprehension in English and Hindi, demonstrating the effectiveness of transformer-based models in capturing cross-lingual semantic similarities and facilitating knowledge transfer across languages.

Artetxe et al. [18] investigated the cross-lingual transferability of monolingual representations, shedding light on the potential of pre-trained embeddings to capture linguistic similarities across languages and enable cross-lingual understanding.

Conneau et al. [19] proposed unsupervised cross-lingual representation learning at scale, presenting a framework for learning language-agnostic embeddings from large-scale multilingual corpora. Their work laid the groundwork for unsupervised approaches to cross-lingual NLP tasks, offering insights into the challenges and opportunities of knowledge transfer across languages.

Hanuja et al. [20] introduced MuRIL, a multilingual representation model tailored to Indian languages. Their study addressed the need for language-specific representations and resources in linguistically diverse regions, advancing the goal of linguistic inclusivity in NLP research.

The integration of cutting-edge technologies like fog computing and the Internet of Things (IoT) into healthcare systems has garnered significant attention due to its potential to enhance disease prediction, healthcare delivery and patient outcomes. This literature review examines key studies focusing on the application of fog computing, IoT and predictive analytics in healthcare settings, along with recent advancements in machine reading comprehension (MRC) and question-answering (QA) systems.

Singh and Kaur [21] presented a framework leveraging fog computing to predict COVID-19 disease at an early stage. Their study showcased the effectiveness of fog computing in processing and analyzing real-time healthcare data, thereby enabling timely intervention and management of the pandemic.

Singh et al. [22] explored the role of IoT in sustaining smart and secure healthcare systems, emphasizing its potential to improve patient monitoring, medication adherence and overall healthcare service delivery. Their work underscores the importance of IoT-enabled solutions in enhancing healthcare accessibility and efficiency.

Singh et al. [23] proposed an ensemble-based approach to predict health insurance premiums at an early stage, showcasing the utility of predictive analytics in optimizing insurance risk assessment and pricing strategies. Their study highlights the value of data-driven insights in informing decision-making processes within the healthcare industry.

Singh and Kaur [24] introduced a software-based framework for the development of smart healthcare systems using fog computing, providing a comprehensive platform for healthcare data management, analysis and decision support. Their framework contributes to the advancement of fog computing applications in healthcare, enabling seamless integration with existing infrastructure and systems.

Beyond healthcare-focused research, recent studies have explored advancements in NLP, MRC, and QA systems. Ramesh et al. [25] introduced Samanantar, the largest publicly available parallel corpora collection for 11 Indic languages, facilitating research in multilingual NLP and cross-lingual understanding.

Recent years have witnessed a surge of interest and research focus on machine reading comprehension (MRC) and question-answering (QA) systems, driven by advancements in natural language processing (NLP) and deep learning techniques. This literature review delves into seminal studies and surveys that provide valuable insights into semantic matching in MRC, the proliferation of QA datasets, the state of open-domain complex question answering, and the evolution of deep learning-based QA systems.

Liu et al. [26] conducted an empirical study on semantic matching in MRC, shedding light on the effectiveness of various semantic matching techniques in enhancing comprehension and answer retrieval accuracy. Their work contributes to the understanding of semantic modeling approaches and their impact on MRC performance, paving the way for more nuanced and effective QA systems.

Rogers et al. [27] presented a comprehensive taxonomy of NLP resources for question answering and reading comprehension, highlighting the diverse array of datasets available for training and evaluating QA models. Their taxonomy provides researchers and practitioners with a systematic framework for navigating the landscape of QA datasets, facilitating informed decision-making and benchmarking efforts.

Etezadi and Shamsfard [28] surveyed the state of the art in open-domain complex question answering, offering insights into the challenges, approaches and advancements in addressing the complexity of real-world questions. Their survey encompasses a wide range of techniques and methodologies employed in open-domain QA systems, providing a holistic overview of the field.

Abdel-Nabi et al. [29] presented a survey focusing on deep learning-based question-answering systems, analyzing the evolution of deep learning techniques and their applications in QA tasks. Their survey synthesizes key research findings and trends in deep learning-based QA systems, offering valuable perspectives on the state of the art and future directions in the field.

Collectively, these studies and surveys contribute to our understanding of semantic matching in MRC, the landscape of QA datasets, the challenges and advancements in open-domain complex question answering, and the role of deep learning in QA systems. They provide researchers and practitioners with valuable insights and resources for advancing the state of the art in MRC and QA, fostering innovation and progress in the field of NLP.

Legal question-answering systems have gained significant attention due to the growing need for efficient access to legal information and expertise. Martinez-Gil [30] conducted a comprehensive survey on legal question-answering systems, providing insights into the various methodologies, techniques and challenges associated with these systems. The survey highlights the importance of machine learning and deep learning approaches in addressing legal question-answering tasks and discusses the potential applications and future directions in this domain.

Community question-answering (CQA) platforms play a crucial role in facilitating knowledge sharing and collaboration within communities. Roy et al. [31] conducted a state-of-the-art review of CQA issues, focusing on the application of machine learning and deep learning techniques. Their analysis provides valuable insights into the challenges and opportunities for improving CQA systems, including question understanding, answer quality assessment and community engagement.

Mahbub et al. [32] introduced CPGQA, a benchmark dataset tailored for machine reading comprehension tasks on clinical practice guidelines. Their work addresses the need for specialized datasets in the healthcare domain, facilitating research and development efforts in leveraging transfer learning for CQA tasks related to clinical practice.

Manjunath et al. [33] proposed a smart question-answering system utilizing a vectorization approach and statistical scoring method. Their system enhances question understanding and answers retrieval by leveraging advanced vectorization techniques and statistical modeling, improving the overall effectiveness of CQA systems.

Darvishi et al. [34] introduced PQuAD, a Persian question-answering dataset, catering to the needs of Persian-speaking communities. Their work contributes to the diversity of available datasets, enabling research and development in Persian language processing and CQA systems tailored for Persian speakers.

Qiu et al. [35] presented a question-answering system based on mineral exploration ontology generation, utilizing a deep learning methodology. Their system leverages domain-specific knowledge to enhance question-answering accuracy in the context of mineral exploration, showcasing the applicability of deep learning techniques in specialized domains.

Abedissa, Usbeck, and Assabie [36] introduced AMQA, an Amharic question-answering dataset, addressing the need for resources in underrepresented languages. Their work contributes to the advancement of natural language processing research by providing a valuable resource for developing question-answering systems tailored for the Amharic language.

Zhang and Zhang [37] proposed FinBERT–MRC, a financial named entity recognition system using BERT under the machine reading comprehension paradigm. Their work focuses on enhancing the extraction of financial entities from textual data, leveraging state-of-the-art deep learning techniques to improve the accuracy and efficiency of financial information extraction tasks.

Wu et al. [38] presented a memory-aware attentive control mechanism for community question answering, incorporating knowledge-based dual refinement techniques. Their work aims to improve the performance of community question-answering systems by enhancing attention mechanisms and incorporating domain-specific knowledge for more accurate and contextually relevant answers.

Ezzini et al. [39] introduced an AI-based question-answering assistance system for analyzing natural-language requirements. Their work focuses on leveraging artificial intelligence techniques to assist in the analysis of natural-language requirements, aiming to improve the efficiency and accuracy of requirement analysis processes in software engineering.

Suissa et al. [40] proposed a question-answering system utilizing deep neural networks for semi-structured heterogeneous genealogical knowledge graphs. Their work addresses the challenges of question answering in complex knowledge graphs, leveraging deep learning techniques to enhance the retrieval and comprehension of information from genealogical data.

3. CONCLUSION

In this paper, we are inclined to mention numerous techniques that are enforced for machine reading Comprehension. Overall, the reviewed literature highlights the diverse applications and methodologies in the field of NLP, ranging from fundamental research on language understanding to practical applications in healthcare, finance, legal domains and beyond. The integration of deep learning techniques, coupled with the availability of large-scale datasets and evaluation platforms, continues to drive innovation and progress in natural language processing research. Ultimately, the overall research is the narrowing of the identified research gap in the literature, fostering advancements in natural language processing for Hindi language comprehension, question answering and contributing to the broader landscape of linguistic diversity in AI applications.

REFERENCES

[1] Xu, Hu, Bing Liu, Lei Shu and Philip S. Yu. "BERT post-training for review reading comprehension and aspect-based sentiment analysis". arXiv preprint arXiv:1904.02232 (2019).

[2] Zappone, Alessio, Marco Di Renzo, and Mérouane Debbah. "Wireless networks design in the era of deep learning: Model-based, AI-based, or both?". IEEE Transactions on Communications 67, no. 10 (2019): 7331-7376.

[3] Chowdhary, KR1442 and K. R. Chowdhary. "Natural language processing". Fundamentals of artificial intelligence (2020): 603-649.

[4] Chandar A P, Sarath, Stanislas Lauly, Hugo Larochelle, Mitesh Khapra, Balaraman Ravindran, Vikas C. Raykar, and Amrita Saha. (2014) "An Auto encoder Approach to Learning Bilingual Word Representations". Advances in Neural Information Processing Systems 27.

[5] Raghavi, K.C., Chinnakotla, M., Shrivastava, M.: Answer ka type kya he? Learning to classify questions in code-mixed language. In: The Proceedings of the International World Wide Web Conference Committee (IW3C2), pp. 853–858 (2015)

[6] Nanda, G., Dua, M., Singla, K.: A Hindi question answering system using machine learning approach. In: 2016 International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT) (2016)

[7] Rajpurkar, Pranav, Jian Zhang, Konstantin Lopyrev and Percy Liang. (2016) "SQuAD: 100,000+Questions for Machine Comprehension of Text., "Proceedings of the 2016 Conferenceon Empirical Methodsin Natural Language Processing: 2383–2392.

[8] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser and Illia Polosukhin. (2017) "Attention is All you Need." Advances in Neural Information Processing Systems 30.

[9] Gupta, Deepak, Surabhi Kumari, Asif Ekbal and Pushpak Bhattacharyya. (2018) "MMQA: A Multi-domain Multi-lingual Question-Answering Framework for English and Hindi. "Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC2018)

[10] Wang, Alex, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy and Samuel Bowman. (2018) "GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding." Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP: 353–355.

[11] Devlin, Jacob, Ming-Wei Chang, Kenton Lee and Kristina Toutanova. (2019) "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics:Human Language Technologies 1: 4171–4186.

[12] Liu, Yinhan, Myle Ott, Na The creationl, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer and Veselin Stoyanov. (2019) "RoBERTa: A Robustly Optimized BERT Pretraining Approach." CoRRabs/1907.11692

[13] Mozannar, Hussein, Karl El Hajal, Elie Maamary and Hazem Hajj. (2019) "Neural Arabic question answering". Proceedings of the Fourth Arabic Natural Language Processing Workshop: 108–118.

[14] Clark, Jonathan H., Eunsol Choi, Michael Collins, Dan Garrette, Tom Kwiatkowski, Vitaly Nikolaev and Jennimaria Palomaki. (2019) "TyDiQA: A Benchmark for Information-Seeking Question Answering in Typologically Diverse Languages".Transactions of the Association for Computational Linguistics 8:454–470.

[15] Gupta, Deepak, Asif Ekbal and Pushpak Bhattacharyya. (2019) "A Deep Neural Network Framework for English Hindi Question Answering". ACM Transactionson Asian and Low-Resource Language Information Processing 19(2):1–22.

[16] Lewis, Patrick, Barlas Oguz, Ruty Rinott, Sebastian Riedel and Holger Schwenk. (2020) "MLQA: Evaluating Cross-lingual Extractive Ques-tion Answering". Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: 7315–7330.

[17] Gupta, Somil, and Nilesh Khade. (2020) "BERT Based Multilingual Machine Comprehension in English and Hindi". CoRRabs/2006.01432.

[18] Artetxe, Mikel, Sebastian Ruder and Dani Yogatama. (2020) "On the Cross-lingual Transfer ability of Monolingual Representations". Proceed-ings of the 58th Annual Meeting of the Association for Computational Linguistics: 4623–4637.

[19] Conneau, Alexis, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzm´an, Edouard Grave, Myle Ott, Luke Zettlemoyer and Veselin Stoyanov. (2020) "Unsupervised Cross-lingual Representation Learning at Scale". Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: 8440–8451.

[20] Hanuja, Simran, Diksha Bansal, Sarvesh Mehtani, Savya Khosla, Atreyee Dey, Balaji Gopalan, Dilip Kumar Margam, Pooja Aggarwal, Rajiv Teja Nagipogu, Shachi Dave, Shruti Gupta, Subhash Chandra Bose Gali, Vish Subramanian and Partha Talukdar. (2021) "MuRIL: Multilingual Representations for Indian Languages". CoRRabs/2103.10730.

[21] P. Singh and R. Kaur, "Implementation of the QoS framework using fog computing to predict COVID-19 disease at early stage", World Journal of Engineering, vol. 12, pp. 345-356, 2021.

[22] P. D. Singh, G. Dhiman and R. Sharma, "Internet of things for sustaining a smart and secure healthcare system", Sustainable Computing: Informatics And Systems, vol. 33, pp. 622-634, 2022.

[23] P. Singh, K. D. Singh, V. Tripathi and V. Chaudhari, "Use of ensemble based approach to predict health insurance premium at early stage", International Conference on Computational Intelligence and Sustainable Engineering Solutions, pp. 566-569, IEEE, 2022.

[24] P. Singh, and R. Kaur, "A software-based framework for the development of smart healthcare systems using fog computing", IET Software, vol. 34, pp. 145-159, 2022.

[25] Ramesh, Gowtham, Sumanth Doddapaneni, Aravinth Bheemaraj, Mayank Jobanputra, Raghavan AK, Ajitesh Sharma, Sujit Sahoo, Harshita Diddee, MahalakshmiJ, Divyanshu Kakwani, Navneet Kumar, Aswin Pradeep, Srihari Nagaraj, Kumar Deepak, Vivek Raghavan, Anoop Kunchukuttan, Pratyush Kumar and Mitesh Shantadevi Khapra. (2022) "Samanantar: The Largest Publicly Available Parallel Corpora Collec-tion for 11 Indic Languages". Transactions of the Association for Computational Linguistics 10:145–162.

[26] Liu, Qian, Rui Mao, Xiubo Geng and Erik Cambria. "Semantic matching in machine reading comprehension: An empirical study". Information Processing & Management 60, no. 2 (2023): 103145.

[27] Rogers, Anna, Matt Gardner and Isabelle Augenstein. "Qa dataset explosion: A taxonomy of nlp resources for question answering and reading comprehension". ACM Computing Surveys 55, no. 10 (2023): 1-45.

[28] Etezadi, Romina and Mehrnoush Shamsfard. "The state of the art in open domain complex question answering: a survey". Applied Intelligence 53, no. 4 (2023): 4124-4144.

[29] Abdel-Nabi, Heba, Arafat Awajan and Mostafa Z. Ali. "Deep learning-based question answering: a survey". Knowledge and Information Systems 65, no. 4 (2023): 1399-1485.

[30] Martinez-Gil, Jorge. "A survey on legal question–answering systems". *Computer Science Review* 48 (2023): 100552.

[31] Roy, Pradeep Kumar, Sunil Saumya, Jyoti Prakash Singh, Snehasish Banerjee and Adnan Gutub. "Analysis of community question-answering issues via machine learning and deep learning: State-of-the-art review". CAAI Transactions on Intelligence Technology 8, no. 1 (2023): 95-117.

[32] Mahbub, Maria, Edmon Begoli, Susana Martins, Alina Peluso, Suzanne Tamang and Gregory Peterson. "cpgqa: A benchmark dataset for machine reading comprehension tasks on clinical practice guidelines and a case study using transfer learning". IEEE Access 11 (2023): 3691-3705.

[33] Manjunath, T. N., Deepa Yogish, S. Mahalakshmi and H. K. Yogish. "Smart question answering system using vectorization approach and statistical scoring method." Materials Today: Proceedings 80 (2023): 3719-3725.

[34] Darvishi, Kasra, Newsha Shahbodaghkhan, Zahra Abbasiantaeb and Saeedeh Momtazi. "PQuAD: A Persian question answering dataset". Computer Speech & Language 80 (2023): 101486.

[35] Qiu, Qinjun, Miao Tian, Kai Ma, Yong Jian Tan, Liufeng Tao, and Zhong Xie. "A question answering system based on mineral exploration ontology generation: A deep learning methodology". Ore Geology Reviews (2023): 105294.

[36] Abedissa, Tilahun, Ricardo Usbeck and Yaregal Assabie. "Amqa: Amharic question answering dataset". arXiv preprint arXiv:2303.03290 (2023).

[37] Zhang, Yuzhe, and Hong Zhang. "FinBERT–MRC: Financial Named Entity Recognition Using BERT Under the Machine Reading Comprehension Paradigm". Neural Processing Letters (2023): 1-21.

[38] Wu, Jinmeng, Tingting Mu, Jeyan Thiyagalingam, and John Y. Goulermas. "Memory-Aware Attentive Control for Community Question Answering With Knowledge-Based Dual Refinement". IEEE Transactions on Systems, Man, and Cybernetics: Systems (2023).

[39] Ezzini, Saad, Sallam Abualhaija, Chetan Arora, and Mehrdad Sabetzadeh. "Ai-based question answering assistance for analyzing natural-language requirements". arXiv preprint arXiv:2302.04793 (2023).

[40] Suissa, Omri, Maayan Zhitomirsky-Geffet and Avshalom Elmalech. "Question answering with deep neural networks for semi-structured heterogeneous genealogical knowledge graphs". Semantic Web 14, no. 2 (2023): 209-237.