Design and Implementation of Medical QA System using Machine Learning Techniques

M/s. Nisha Patil*, Dr. Kuldeep Sambrekar

Department of Computer Science and Engineering, KLS GIT, Belagavi, Karnataka, India

ABSTRACT

Nowadays, people rely on Traditional books or Google for every answer to their questions on a day-to-day basis, from basic information to medical queries. Till now, people are facing problems and are unable to find the accurate answer to their questions or fetch relevant results. Also, this technique is time-consuming as people have to go through many books to obtain one relevant answer or search various websites, which is a tedious task and not an efficient way in today's world where time is the top priority, yet the majority of people follow these techniques. So, to overcome this technique and solve the current problems, we have implemented a new technique in this paper. The BERT model, pre-trains deep bidirectional representations from the unlabeled text which conditions on both left and right, as a result, provides accurate answers to the user's query when compared to the state-of-the-art model. This same model can be further implemented in other domains to obtain accurate results.

Keywords : Question Answering, Natural Language Processing, Semantic Search, Medical Informatics, Machine Learning, Neural Networks, Artificial Intelligence, Language Processing.

I. INTRODUCTION

Natural language processing is a type of artificial intelligence, which consists of analyzing, understanding, and generating natural human spoken and written languages so that computers can process natural language without using computer-driven language. Natural language processing also called "computational linguistics," uses both semantics and syntax to help computers understand written and spoken human language and its meaning from what they say. It incorporates the ability of artificial intelligence and computer programming into an understanding that programs can even translate one language into another reasonably accurately. NLP also consists of voice recognition, which enables the computer to understand and respond appropriately to human language.

The Question Answering task aims to provide accurate and quick answers to users' queries from a collection of documents or databases. This kind of IR system is severely needed with the drastic growth of

---

digital information. Here we have proposed a semantic approach to QA based on

a) Natural Language Processing Techniques, which allows a deep analysis of medical questions and documents.
b) Semantic Web Technologies at both representation and interrogation levels.

We have come up with the solution as Semantic Question Answering System and propose methods "Answer Search" based on semantic search and query relaxation.

We have evaluated the overall system performance on real questions and answers extracted from MEDLINE articles [9].

Our experiments show promising results and suggest that the query-relaxation strategy can further improve the overall performance of the system as well as patients.

We had extensive data collected by preprocessing and storing it in the document store database moving further, training the model using transfer learning on document store data, and then running the server to connect to the model. Lastly, asking questions and returning the answers and the context to it.

This paper addresses the problems and its solution for QA in the medical domain where several specific conditions are satisfied.

II. LITERATURE SURVEY

This is just a yes or no question. But the clinical outcomes of the therapies could be complicated. Answers to these questions can be obtained in the book Clinical Evidence (CE) [5]. Natural language processing is involved in two places:1) Queries to the system 2) Searching the text of CE for answers. (NFQA) Non-factoid question-answering which we frame as a problem is insufficient in simple keyword searching.

Analysis of Semantic Classes: TOWARD NON-FACTOID QUESTION ANSWERING. By Yun Niu,

this research was a component of the University of Toronto, Canada EPoCare project at the Centre for Evidence-Based Medicine [4]. Niu put forward an approach in her thesis to the problem that uses semantic class analysis as the organizing principle to answer non-factoid questions. This approach contains four major components:

1. Detection of semantic classes in QA sources;
2. Recognizing properties of semantic classes;
3. Question-answer matching: investigating semantic class properties to find relevant information;
4. Constructing answers by merging relevant information using relations between semantic classes

Niu investigated NFQA within the context of clinical Question responsive and targeted 3 linguistics categories that correspond to roles within the normally accepted PICO format of describing clinical eventualities. An important property of semantic classes, their cores was recognized by Niu, which showed how cores of interventions, problems, and outcomes in a sentence can be extracted automatically by developing an approach exploring semi-supervised learning techniques. Another property that she analyzes is polarity, an inherent property of clinical outcomes. She developed a way of employing a supervised learning model to detect the polarity of clinical outcomes. She has shown that text summarization is closely related to answer construction in NFQA [4]

Moreover, the research was done by Asma Ben Abacha and Pierre Zweigenbaum which states that the Question Answering task aims to provide precise and quick answers to user questions from a collection of documents or a database. With the dramatic growth of digital information, the Information-Retrieval system is needed. This paper addresses the problem of QA in the medical domain where several specific conditions are met. We have proposed a linguistic approach, based on a) Natural language

processing techniques, which permit a deeper understanding of medical queries and documents and

b) Semantic Web technologies at both representation and interrogation levels. We have examined the system performance on real questions and answers extracted from MEDLINE articles. Our experiments show promising results and suggest that a query-relaxation strategy can further improve the overall performance [3]. The information provided by MedlinePlus also provided us with more insights and data regarding the research work.[9]

This works focuses on the following objectives:

1. Collecting the data, preprocessing, and storing it in the document store.
2. Training the model using transfer learning on document store data.
3. Running the server to connect to the model.
4. Asking questions and returning the answer and context.

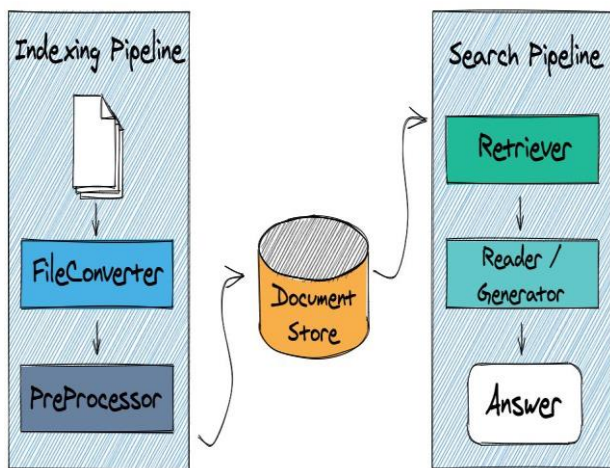## III. SYSTEM ARCHITECTURE

Figure 1. System Architecture

Graphics processing unit is used in the proposed model, as the processing unit and haystack framework from the GitHub repository. Given a pool of articles or documents related to the medical field can answer and context the answer from those pool of articles

using Natural Language Processing. The system used for development is by using Haystack Framework, which is an open-source, end-to-end framework that enables to build production-ready pipelines for different search use cases, it is built in a modular fashion, which uses all the latest transformer-based systems like BERT, Roberta, MiniLM for extractive QA, Generative QA, and document retrieval, you can use the state-of-the-art NLP models in Haystack to provide a unique search experience and allow your users to query in natural language. Haystack also allows for custom routing of queries to only the relevant components due to its node and pipeline design. Millions of documents are scaled over retrievers, production-ready backends like Elasticsearch or FAISS and a fast API REST API, and many more. It collects new training data via user feedback in production and improves the model continuously.

File Converter: Extracts pure text from files (pdf, Docx, pptx, HTML, and many more).

Pre-Processor: Pure texts are extracted from various types of files using PDF To Text Converter imported from the haystack, which passes the English language from the document and performs split by word. This is done by using NLP libraries and glob function that helps in providing a list of files along with their name into smaller chunks and writing them to a Document store.

Document Store: Database stores the documents, metadata, and vectors for our search. Elasticsearch is recommended as it comes preloaded with features or FAISS, but also has more lightweight options like SQL or In-Memory for fast prototyping.

Search Pipeline: This contains both the reader and the retriever into a single unit thus creating a pipeline.

Retriever: For a given query, from a large collection of databases the applicable documents are identified using a fast algorithm. Retrievers then narrow down the search space significantly for scalable QA. TF-IDF

retriever is used to retrieve the answers from a document that is based on Pandas data frames.

Reader: Neural network that reads through texts in detail to find answers to the query. FARM reader is used, which uses ROBERTA-based SQuAD which is an NLP model. The pre-trained model can be loaded from Hugging Face's model hub or fine-tune it on your domain data.

Answers: We have used Gradio for GUI, which provides a questions-answers interface that runs user queries and provides answers that the reader provides. BERT model is used, to pre-train deep bidirectional representations from the unlabeled text which conditions on both left and the right, as a result, with just one additional output layer the pre-trained BERT model can be fine-tuned to create state-of-the-art models for a broad range of NLP tasks. The pre-training model is half the work done, this is because we train a model on a large text corpus, and our model starts to pick up a deeper understanding of how the language works. One of the biggest challenges in NLP is the lack of training data because NLP is a diversified field with many well-defined tasks, most datasets contain only a few hundred thousand humans labeled training examples.
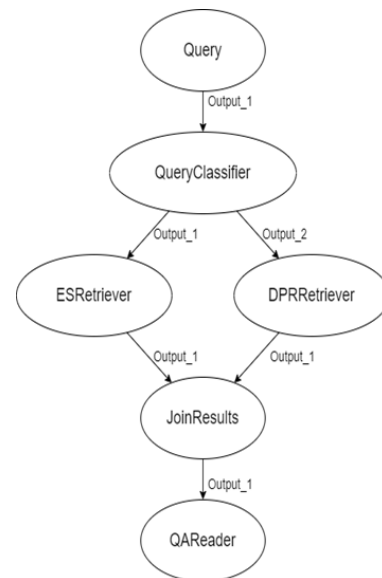

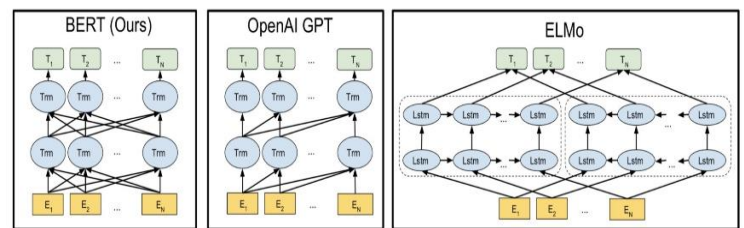**Figure 2**. Shows the Actual Flowchart of the System Architecture [Figure 1]

IV. COMPARISON

Figure 3. Comparison between BERT, OPEN-AI GPT, ELMo

	Base model	Pre-training	Downstream tasks	Fine-tuning	Training
BERT	Transformer encoder	Bi-directional	BooksCorpus (800M words) and Wikipedia (2,500M words)	Task-specific	1M steps with a batch size of 128,000 words.
OPEN-AI GPT	Transformer decoder	Unidirectional	BooksCorpus (800M words)	Supervised	1M steps with a batch size of 32,000 words
EMLo	Two-layer LSTM	Semi-bidirectional	Feature-based	--	--

Table:1 – Comparison of various models

Among three, BERT is more efficient and gives accurate results to users' queries.

V. RESULTS

The proposed model was implemented on these hardware interfaces, Processor: Dual-core with 2.4 GHZ or more, RAM: Min 1GB or more, Memory Space: Minimum 5GB 3.1.2 Software Interfaces, Language: Python, Development IDE: Colab, Cloud Platform: Colab and the following results were obtained. The results show that the proposed model gives accurate results for the given input.
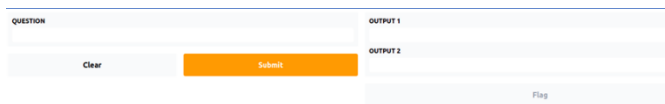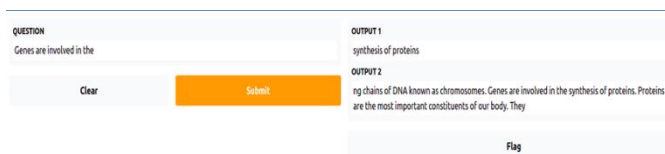

Figure 4. User Interface for entering queries.

Figure 5. Testing the QA system.

Figure 6. Case 1: Output for the given query-1.

Figure 7. Case 2: Output for given query-2.

VI. CONCLUSION

In this paper, we have discussed various flaws and solutions to the problems faced currently. We have also discussed test cases in various techniques and have compared various models where the BERT model efficiently outperforms other models and gives accurate results.

This model helps to extract the correct answer and context so that users can gain the knowledge they want from the pool of articles without spending time reading every article they have. Context feature will help in giving the scope of the article so that users can directly go and read the article that has the answer. This same model can be further implemented in other domains to obtain accurate results.

VII. REFERENCES

[1]. Niu, Yun, and Hirst, Graeme. "Analyzing the text of clinical literature for question answering." In: Prince, Violaine and Roche, Mathieu (editors), Information Retrieval in Biomedicine, IGI Global, 2009, 190-220 DOI: 10.4018/978-1-60566-274-9

[2]. Niu, Yun, and Hirst, Graeme. "Identifying cores of semantic classes in unstructured text with a semi-supervised learning approach." Proceedings, International Conference on Recent Advances in Natural Language Processing, September 2007, Borovets, Bulgaria, 418–424. http://ftp.cs.toronto.edu/pub/gh/Niu+Hirst-RANLP-2007.pdf

[3]. Asma Ben Abacha, Pierre Zweigenbaum, MEANS A medical question-answering system combining NLP techniques and Semantic Web technologies, Information Processing & Management, Volume 51, Issue 5, 2015, Pages 570-594, ISSN 0306-4573, https://doi.org/10.1016/j.ipm.2015.04.006.

[4]. Niu, Yun. Analysis of Semantic Classes: Toward non-factoid question answering. Ph.D. Thesis. Department of Computer Science, University of Toronto. March 2007. https://ftp.cs.toronto.edu/pub/gh/Niu-thesis.pdf

[5]. Niu, Yun; Zhu, Xiaodan; and Hirst, Graeme. "Using outcome polarity in sentence extraction

for medical question-answering." Proceedings of the American Medical Informatics Association 2006 Annual Symposium, Washington, D.C., November 2006, 599-603. http://ftp.cs.toronto.edu/pub/gh/Niu-etal-2006.pdf

[6]. Niu, Yun; Zhu, Xiaodan; Li, Jianhua; and Hirst, Graeme. "Analysis of polarity information in medical text." Proceedings of the American Medical Informatics Association 2005 Annual Symposium, Washington, D.C., October 2005, 570-574. http://ftp.cs.toronto.edu/pub/gh/Niu-etal-2005.pdf

[7]. Niu, Yun, and Hirst, Graeme. "Analysis of semantic classes in medical text for question answering." Workshop on Question Answering in Restricted Domains at the 42nd Annual Meeting of the Association for Computational Linguistics, Barcelona, July 2004, 54-61. https://aclanthology.org/W04-0509.pdf

[8]. Niu, Yun; Hirst, Graeme; McArthur, Gregory; and Rodriguez-Gianolli, Patricia. "Answering clinical questions with role identification." Proceedings, Workshop on Natural Language Processing in Biomedicine, 41st Annual Meeting of the Association for Computational Linguistics, Sapporo, Japan, July 2003, 73-80 https://aclanthology.org/W03-1310.pdf

[9]. MEDLINE PORTAL, 29 June 2021https://medlineplus.gov/about/using/usingcontent/

[10]. Sarrouti, M., Ouatik, S.E.A., A Passage Retrieval Method based on Probabilistic Information Retrieval Model and UMLS Concepts in Biomedical Question Answering, Journal of Biomedical Informatics (2017), DOI: http://dx.doi.org/10.1016/j.jbi.2017.03.001.

[11]. Hristovski D, Dinevski D, Kastrin A, Rindflesch TC. Biomedical question answering using semantic relations. BMC bioinformatics 2015;16(1):6. doi:10.1186/s12859-014-0365-3.

[12]. Bauer MA, Berleant D. Usability survey of biomedical question answering systems. Human Genomics 2012;6(1):17. doi:10.1186/1479-7364-6-17.

[13]. Sarker A, Mollá D, Paris C. Query-oriented evidence extraction to support evidence-based medical practice. Journal of Biomedical Informatics 2016;59:169–84. doi:10.1016/j.jbi.2015.11.010.

[14]. ullivan DO, Wilk S, Kuziemsky C, Michalowski W, Farion K, Kukawka B. Is there a consensus when physicians evaluate the relevance of retrieved systematic reviews? Methods of Information in Medicine 2016;55(3):292–8. doi:10.3414/me15-01-0131.

Cite this article as :

Nisha Patil, Dr. Kuldeep Sambrekar, "Design and Implementation of Medical QA System using Machine Learning Techniques", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 7 Issue 6, pp. 129-134, November-December 2021. Available at doi : https://doi.org/10.32628/CSEIT217635
Journal URL : https://ijsrcseit.com/CSEIT217635