MS(CS) Research Thesis-II

Interim Report

Name	**Maaz Rafiq**
Student Id	**FA21-MSCS-0003**
Thesis Title	**EVIDENCE RETRIEVAL FOR EBM**
Supervisor Name	**Dr. Shaukat Wasi**

List of Meetings

S. No.	Date	Topic
1	13th Feb'2023	Discuss methodology
2	25th Feb'2023	Discuss experiment results of evidence finding
3	11th March'2023	Discuss different retrieval approaches
4	22nd March'2023	Discuss data labeling for evidence finding
5	31st March'2023	Discuss experiments of Relevance Retrieval results

List of Research Papers for Literature Review/ Gap Analysis (at least 5)

S.No.	Paper Title	Journal/ Conference	Year
1.	A relevance and quality-based ranking algorithm applied to evidence-based medicine	Computer Methods and Programs in Biomedicine ***Elsevier***	**2020**
2.	COVID-19 information retrieval with deep-learning based semantic search, question answering, and abstractive summarization	npj Digital Medicine ***Nature***	**2021**
3.	EBM+: Advancing Evidence-Based Medicine via two level automatic identification of Populations, Interventions, Outcomes in medical literature	Artificial Intelligence in Medicine ***Elsevier***	**2022**
4.	TREC-COVID: constructing a pandemic information retrieval test collection	Special Interest Group on Information Retrieval (SIGIR) ***ACM***	**2021**

5.	TransforMED: End-to-End Transformers for Evidence-Based Medicine and Argument Mining in medical literature	Journal of Biomedical Informatics ***Elsevier***	**2021**

Abstract

Evidence-Based Medicine (EBM) is a vital approach to clinical decision-making that emphasizes the use of the best available evidence. Retrieval of relevant and high-quality evidence is a critical step in EBM, and this research proposes a methodology for evidence retrieval and finding that combines information retrieval techniques and deep learning models.

The proposed methodology involves retrieving research documents based on their relevance to a user's query and their quality. The quality assessment is based on various factors such as journal impact factor, author reputation, publication date, and citation count.

The deep learning model used in the methodology is designed to extract PICO (Patient, Intervention, Comparison, and Outcome) entities from the retrieved documents to identify evidence that is relevant to the user's query. The PICO entities are identified using natural language processing techniques and machine learning algorithms.

Method

The methodology we are proposing consists of three major components: **1)** Relevance Ranking, **2)** Quality Ranking, and **3)** Evidence Finding

1. **Relevance Ranking:**

In this component, we extract the documents from the collection that is relevant to the user query.

To measure the relevancy there could be multiple retrieval models like vector space models, probabilistic models, language models, etc. Right now, we are experimenting with Vector Space Models that establish the ranking based on the frequency of words.

2. **Quality Ranking:**

After getting the relevant documents along with their relevancy score, we ranked the documents based on their quality. The quality of the document will be assessed based on two things **1)** quality of the paper, **2)** quality of the author**.** The quality of the paper includes publication year, Citations count, where the paper is published, impact factor, etc. Author quality includes h-index, citation counts, etc.

When we rank our documents on both relevancy and quality-wise, we combine the properties of both rankings and as a result, we will have a list that contains the documents ranked with the fused properties of quality and relevancy.

3. **Evidence Finding:**

When we will have the final, we will find the evidence from these documents. To find the evidence we will find the PICO entities in the documents and to find PICO entities we are using deep learning models like NER, classification, etc.

Outputs

- Relevant documents against the query
- Quality of documents against the query

- List of Evidence against each document

Achievements

- List of relevant documents **(completed)**
- List of quality documents **(working)**
- Evidence finding **(experimenting)**

Project Plan

TASK NAME	START DATE	END DATE	STATUS
Experiments on available evidence-finding models	13th Feb'2023	5th March'2023	Not Completed
Dataset annotations	5th March'2023	30th April'2023	working
Retrieval of relevant documents	1st March'2023	20th March'2023	Completed
Clustering on Relevant results	20th March'2023	25th March'2023	Completed
Topic labeling on clusters	25th March'2023	31st March'2023	Completed
Implementation Quality of author algorithm	1st April'2023	15th April'2023	Working
Implementation Quality of paper algorithm	16th April'2023	30th April'2023	Pending
Train model on an annotated dataset	1st May'2023	15th May'2023	Pending
Evaluating the Results of the Test dataset	16th May'2023	31st May'2023	Pending
Creating end-to-end pipeline	1st June'2023	10th June'2023	Pending
Comparing Results with the latest research	11th June'2023	20th June'2023	Pending