

Introduction:

The COVID-19 pandemic has necessitated rapid and efficient access to up-to-date information for researchers, clinicians, policymakers, and the general public. In response, numerous information retrieval (IR) systems have been developed and evaluated to provide access to relevant scientific literature, legal information, and real-time updates regarding the pandemic. This literature review aims to summarize and evaluate the effectiveness of these systems in addressing the diverse information needs arising from the pandemic.

Overview of Literature:

TREC-COVID, rationale and structure of an information retrieval shared task for COVID-19:

This article introduces the TREC-COVID shared task, which uses the COVID-19 dataset for training information retrieval (IR) systems. The dataset includes a large number of scientific articles related to COVID-19, updated weekly. The intended users of the system are experts, such as researchers, clinicians, policy makers, and journalists. The article discusses the high level of participation in the task, indicating strong community interest. The article provides a comprehensive overview of the TREC-COVID shared task, highlighting the importance of effective IR systems during the COVID-19 pandemic. However, the lack of specific outcome details from the evaluation limits the assessment of the task's effectiveness.

Semantic matching based legal information retrieval system for COVID-19 pandemic:

This system, implemented on WeChat, utilizes a convolutional neural network (CNN) model trained on Chinese legal cases to provide legal information related to the COVID-19 pandemic. The CNN-based encoder yields the best results, and the introduction of contrastive learning leads to performance improvements.

The article presents a novel approach to legal information retrieval using neural network models. The system's performance improvements with contrastive learning demonstrate its effectiveness in enhancing representation learning, especially in scenarios with limited data.

Amazon IR system for pandemic:

This system uses Amazon Comprehend Medical (ACM) for processing unstructured medical text and achieved good performance, especially with the question+narrative variant. The study highlights the potential pitfalls of deploying deep learning-based systems without proper tuning. The article emphasizes the importance of tuning deep learning-based systems for optimal performance. The comparison with the sabir system underscores the advantages of machine learning approaches over manually tuned systems.

Google IR system for pandemic:

This system adapts BERT for biomedical articles using the BioASQ dataset for tuning. It employs synthetic query generation for data augmentation due to the smaller size of the datasets. The article demonstrates the effectiveness of adapting BERT for biomedical articles and the use of synthetic

query generation for data augmentation. These approaches address the challenges of domain-specific tuning and dataset limitations.

Information retrieval and question answering: A case study on COVID-19 scientific literature:

The IR module of CORD-19 involves preliminary retrieval and reranking, with a focus on passages over full documents for better answer extraction in question-answering systems. Passage retrieval offers a better starting point for answer extraction, but document retrieval performs better in QA metrics. The article provides insights into the IR module of CORD-19, highlighting the importance of passage retrieval for question-answering systems. The findings suggest the need for fine-tuning the QA module for optimal performance.

COVID-19 information retrieval with deep-learning based semantic search, question answering, and abstractive summarization:

This system combines the CORD-19 corpus with the TREC-COVID competition's evaluation dataset to provide a comprehensive information retrieval solution for COVID-19-related content. CO-Search achieves top performance in the TREC-COVID challenge. CO-Search demonstrates top performance in semantic search, question answering, and summarization tasks related to COVID-19. The limitations highlighted in the article, such as handling long documents and potential misinformation, indicate areas for future improvement.

Detecting Biomedical Named Entities in COVID-19 Texts:

The article introduces a ML pipeline, Biomedical Named Entity Recognition Pipeline (BNP), for recognizing and classifying biomedical entities in COVID-19 texts. The BNP approach shows superior performance across various benchmarks, including a dataset curated from COVID-19 case reports. The BNP approach demonstrates excellent performance in recognizing biomedical entities in COVID-19 texts. The use of domain-specific pre-trained language models and a stacked ML pipeline contributes to its success, although challenges remain in training new biomedical entities due to the absence of specific datasets.

Answering Questions on COVID-19 in Real-Time:

The COVID-19 Questions dataset is used to provide real-time answers to questions related to COVID-19 using a combination of biomedical text mining, QA techniques, and information retrieval approaches. The system aims to address the need for real-time information dissemination during the COVID-19 pandemic. However, specific details about the dataset, algorithms, or evaluation metrics are not provided, limiting the assessment of its effectiveness.

Efficient Self-Supervised Metric Information Retrieval: A Bibliography Based Method Applied to COVID Literature:

The SUBLIMER approach is a self-supervised method for metric information retrieval that leverages bibliography citations for semantic similarity. It outperforms state-of-the-art competitors in Precision@5 (P@5) and Bpref and can be applied to other domains, not just COVID-19 literature. The SUBLIMER approach demonstrates promising results in metric information retrieval, particularly in the context of COVID-19 literature. However, the scalability of the method to larger datasets or real-time scenarios is unclear, which could be a limitation.

An Evaluation of Two Commercial Deep Learning-Based Information Retrieval Systems for COVID-19 Literature:

The paper evaluates Amazon's COVID-19 Search and Google's COVID-19 Research Explorer, which use deep learning for information retrieval on COVID-19 literature. The evaluation suggests that these commercial systems under-performed compared to academic prototypes evaluated in the TREC-COVID task. The evaluation highlights potential limitations in the trustworthiness and performance of popular health search engines. However, the paper lacks specific details about the datasets, algorithms, or evaluation metrics used, limiting a comprehensive assessment of the systems.

CAiRE-COVID: A Question Answering and Query-focused Multi-Document Summarization System for COVID-19 Scholarly Information Management:

CAiRE-COVID is a system for question answering and query-focused multi-document summarization of COVID-19 scholarly information. It consists of a Document Retriever, Relevant Snippet Selector, and Query-focused Multi-Document Summarizer, utilizing pre-trained language models. The system addresses the need for efficient information extraction from COVID-19 scholarly articles. However, the lack of evaluation on specific datasets and a comprehensive analysis of its limitations limits a full assessment of its effectiveness.

Rapidly Deploying a Neural Search Engine for the COVID-19 Open Research Dataset:

The paper describes the development and deployment of the Neural Covidex search engine, which aims to provide improved information access to the COVID-19 Open Research Dataset. The system utilizes sequence-to-sequence models, specifically T5, for relevance classification and reranking of search results. While the paper highlights the development of a search engine for COVID-19 literature, it lacks experimental results or a systematic evaluation of the search engine's performance, limiting the assessment of its effectiveness.

Fact Retrieval and Matching System for COVID-19 Epidemic News Content:

The system utilizes COVID-19 epidemic news content as its dataset and employs a multi-model fusion approach, utilizing basic Word2vec, BERT model, and unsupervised SimCSE two-tower model for semantic analysis. The system demonstrates promising results in fact retrieval for COVID-19 epidemic news content. However, challenges in effectively combining different semantic models and ensuring accuracy of retrieved facts are noted, which could impact its effectiveness in providing accurate and relevant information.

Cognitive Search for COVID-19:

The system utilizes the COVID-19 dataset and employs multiple retrieval algorithms, including Term-Frequency, Semantic Neural Search, and Hybrid Term-Neural, to provide relevant search results for COVID-19-related queries. The system demonstrates an innovative approach to information retrieval for COVID-19 literature. However, challenges in integrating and optimizing multiple retrieval algorithms and managing the complexity of the system's interface for users are noted, which could impact its usability and effectiveness.

Developing A Clinical Evidence Retrieval Service in Response to the COVID-19 Pandemic:

The COVID-19 Evidence Retrieval Service (CERS) provides clinicians with updated evidence on COVID-19 using a structured approach involving support, evidence retrieval, and expert panel teams. It utilizes Population, Intervention, Control, and Outcome (PICO) format for query segmentation and employs search strategies based on key concepts and terms related to COVID-19. The service demonstrates high satisfaction among users and significant changes in clinical practice. However, challenges related to data processing and computational constraints are noted, which could impact the service's ability to provide timely and accurate information.

KnowCOVID-19:

KnowCOVID-19 utilizes the CORD-19 dataset and employs evidence-based filtering techniques, including topic modeling and category modeling, to provide relevant information related to COVID-19. It introduces a domain-specific Topic Model (DSTM) for discovering latent patterns among research topics, drugs, and genes related to COVID-19. The system demonstrates effectiveness in retrieving relevant information from COVID-19 literature. However, challenges related to data reliance, validity, and diffusion of innovations are noted, which could impact its widespread adoption and reliability.

INKAD COVID-19 IntelliSearch:

INKAD COVID-19 IntelliSearch utilizes the COVID-19 Open Research Dataset (CORD-19) and employs a combination of information retrieval (IR) and question answering (QA) techniques. It uses the BM25 algorithm for document retrieval and the RoBERTa Base model for question answering. The system demonstrates promising results in information retrieval and question answering for COVID-19 literature. However, challenges related to scalability, language coverage, and continuous updates are noted, which could impact its effectiveness in handling a wide range of queries and maintaining accuracy over time.

CoviBioBERT:

CoviBioBERT is a system that utilizes finely-tuned ensemble neural networks based on the BioBERT pre-trained model for answering COVID-19-related queries. It employs transfer learning to customize the BioBERT model specifically for COVID-19-related queries. The system demonstrates superior performance in addressing COVID-19-related queries compared to foundational models. However, challenges in fine-tuning the pre-trained model, scalability, and generalization of the system's performance to different datasets are noted, which could impact its effectiveness in handling a wide range of queries.

Semantic Malayalam Dialogue System for COVID-19 Question Answering:

The system utilizes a Malayalam dataset containing information related to COVID-19 and employs NLP techniques, Neural Network models, and cosine similarity measure for document processing, semantic modeling, and answer retrieval, respectively. The system demonstrates an innovative approach to question answering in the Malayalam language. However, challenges in accurately representing nuances of the language in the word embedding process and ensuring relevance and

accuracy of answers are noted, which could impact its effectiveness in handling a wide range of queries.

These articles and research papers we studied collectively highlight the importance of effective information retrieval and named entity recognition systems in addressing the challenges posed by the COVID-19 pandemic. They showcase innovative approaches and the need for domain-specific tuning and dataset augmentation to improve system performance.

Conclusion:

The literature reviewed underscores the importance of effective IR systems in addressing the information needs arising from the COVID-19 pandemic. While these systems have shown promising results in providing access to relevant information, challenges related to scalability, domain-specific tuning, and system complexity persist. Future research should focus on addressing these challenges to further enhance the effectiveness of IR systems in combating global health crises.

I have also attached my top 4 findings in the table below:

Name	Dataset Information	Dataset Language	Approach/Model	Test/Accuracy
TREC-COVID	CORD-19	English	BERT/RoBERTa	93%
CAiRE-COVID	CORD-19	English	Neural question answering and query-focused multi-document summarization system	Not mentioned
CoviBioBERT	The system is evaluated on a benchmark dataset of question-answer pairs related to COVID-19.	English	Utilizes finely-tuned ensemble neural networks based on the BioBERT pre-trained model	The system is evaluated on a benchmark dataset of question-answer pairs, and its performance is measured using metrics such as Exact Match (EM) Almost 89% accuracy
CERS	CERS primarily relies on retrieving and appraising relevant literature and evidence.	English	CERS utilized a structured approach involving three main teams: the support team, evidence retrieval team, and expert panel.	Formal feedback from users indicates high satisfaction with the service, with 11 out of 12 users reporting satisfaction.

References:

The information extracted from all the papers can be seen in the table form from the link:

<https://docs.google.com/spreadsheets/d/1i4ZJleq7O9LY35fhB25FYqWWcAn0dyaFtykXtIFqjXg/edit?usp=sharing>