COVIDExplorer: Exploring the Universe of COVID-19 Knowledge

Heer Ambavi, heer.ambavi@iitgn.ac.in
Kavita Vaishnaw, kavita.vaishnaw@iitgn.ac.in
Udit Vyas, udit.18110176@iitgn.ac.in
Abhisht Tiwari, abhisht.tiwari@iitgn.ac.in
Mayank Singh, singh.mayank@iitgn.ac.in

Abstract

The entire world is engulfed in the fight against the COVID-19 pandemic, leading to an information surge on the pandemic through research experiments, government organizations, and social media platforms. A multi-modal information access and data visualization platform can play a critical role in supporting research aimed at understanding and developing preventive measures for the pandemic.

We present such a multi-faceted AI-based search and visualization system, COVIDEXPLORER. Our system aims to aid researchers in understanding current state-of-the-art COVID-19 research, identify research articles relevant to their domain, and visualize real-time trends and statistics of COVID-19 cases. In contrast to other existing systems, COVIDEXPLORER also brings in the topical discussions going on in social media to study different aspects of COVID-19. That is, we seamlessly integrate three different aspects of the pandemic into our system:

- **Search and recommendation**: We support keyword-based full-text search on the CORD-19 dataset \cite{cord19} of 63k research articles on coronaviruses. We show top relevant papers with authors, journal of publication, the date of publication, and bio-entity mentions. We use a NER system fine-tuned from SciBERT \cite{sciBERT}, the current state of the art language model for scientific and biomedical text to extract the bio-entities. We currently identify DNA, RNA, proteins, cell types, cell lines, diseases, and chemical entities.

- **Statistics**: We provide interesting insights on the recognized bio-entities including timelines of first mentions, a visualization of popularly co-mentioned entities, a year-wise distribution of mention frequencies. We also keep track of the daily evolving pandemic situation, displaying the total cases, active cases, deaths, and recovery statistics visually.

- **Social media discussions**: Currently, we process COVID-19 Tweet IDs dataset \cite{tweetIDs} and extract the most common hashtags, mentions, tweet locations and URLs displayed along with a timeline of the twitter activity in the context of the pandemic.

While our system primarily focuses on the Indian-subcontinent, similar systems can be built for other regions too. The system is accessible at: http://covidexplorer.in.
References

