COVID-SEE: The Scientific Evidence Explorer for COVID-19 Related Research

We present COVID-SEE\(^1\), a system for medical literature discovery, which augments search through a structured visual overview of a collection enabling exploration. While many search and question answering tools emerged in response to the COVID-19 outbreak, relatively few leverage domain knowledge to organise and present information found within the literature \([5]\). To fill this gap, we developed a web application that combines a search engine for COVID-19 medical literature with summary visualisations of document content, such as concepts, relations, and topics.

A typical usage scenario in COVID-SEE begins with a textual query over the COVID-19 literature, supporting semantic search and providing: (i) a list of retrieved documents together with their metadata, and (ii) a visualisation dashboard with three distinct interactive views which highlight the relations between entities and concepts detected in the documents. As a user reviews and interacts with the information in these views, documents of interest can be saved into a collection for later export or targeted visualisation. Our objective is to combine learning and investigation with direct retrieval to support the known health information seeking behaviour of a user reviews and interacts with the information in these views. As a user reviews and interacts with the information in these views, documents of interest can be saved into a collection for later export or targeted visualisation. Our objective is to combine learning and investigation with direct retrieval to support the known health information seeking behaviour of a user [3]. We facilitate exploration by providing views of document content that provide a user with deeper insight into retrieved articles.

The first view is a relational concept view – a Sankey diagram frame, in which we organise the medical concepts found in the retrieved articles according to key entities for clinical queries, known as PICO [4] (Population, Intervention, Comparator, Outcome). In this view, more salient relations – based on the number of supporting abstracts – carry more weight, and once a relation is clicked, the corresponding articles are retrieved using MetaMap rather than tokens, as it helps to preserve multi-word concepts such as intensive care unit and map different variants of a given term into a single concept, thus reducing noise, and highlighting important keywords [2].

Our third component is a concept cloud view (Fig. 2, Inset), showing the 20 most representative concepts for each active document. Concepts here again correspond to UMLS terms extracted using MetaMap. To select discriminative concepts, concept distributions in the selected article are compared to those in the collection using the log-likelihood test.

In conclusion, COVID-SEE facilitates more interactive exploration of the COVID-19 literature, through integration of sub-collection thematic analysis, document-level salient concept summaries, and PICO-structured concept relations.

REFERENCES


\[1\] https://covid-see.com

\[2\] http://metamap.nlm.nih.gov

Fig. 1. Visualisation of PICO concepts and relations in articles retrieved for query incubation period of COVID-19. Links between concepts can be selected to reveal papers with those relations.

Fig. 2. Topic visualisation for articles retrieved for query incubation period of COVID-19. Inset: Word cloud view of an individual document showing 20 key concepts, including multi-word terms.
