Title: Text Mining for Automated Search and Exploration of COVID-19 Literature

Abstract

Since Dec 2019, more than 50,000 scientific articles have been published on COVID-19. The exponential growth in the number of publications makes it difficult for clinicians, public health officials, biomedical researchers, and the general public to stay current with the latest findings. Text-mining tools are needed to facilitate rapid information extraction and summarization from experimental/clinical studies and answer high-priority questions. The main goal of this project is to use text-mining to extract and summarize the rapidly evolving information on risk factors, biomarkers, and drug targets in the COVID-19 scientific literature in PubMed abstracts. Co-occurrence based networks are used to summarize multiple articles and identify key biomedical terms (e.g. genes, cytokines, chemicals, mutations) that are related to the search query (e.g. COVID-19 delta variant). As a case study, we searched for top genes that are mentioned in the literature related to the query term: "(SARS-COV-2 or COVID-19) and diabetes" The top 5 genes included: angiotensin converting enzyme 2, cardiac troponin-1, plasmin, interleukin 6, and C-reactive protein. Automated summarization of biomedical text will enhance access to information and allow biomedical researchers and general public to find information related to risk factors of COVID-19 including pregnancy, smoking, and comorbidities and identify potential drug targets.