

preVIEW: A Service for Central Access to COVID-19 Preprints

Lisa Langnickel *

Roman Baum**

Johannes Darms***

Juliane Fluck****

Abstract

Introduction

During the current COVID-19 pandemic, many researchers decide to publish their work on preprint servers. Since the findability of profound information is crucial during the pandemic, information provider should provide access to both preprints and peer-reviewed publications, but these resources should be clearly distinguishable for service users. Various preprint servers exist, which all have different access politics and interfaces, different search functionalities and content. Thereby, new research items are hard to find even though they are published as preprints to allow faster communication of findings. Therefore, ZB MED – Information Centre for Life Sciences developed the search engine preVIEW COVID-19 that enables a central access point to COVID-19 related preprints offering open API access.

Methods

Metadata is retrieved from different preprint servers or via Crossref and harmonized into a defined data format. A text mining workflow is integrated that applies state-of-the-art machine learning methods to annotate MeSH disease concepts, human and SARS-CoV-2 gene/protein entities in addition to SARS-CoV-2 variants of interest. Preprint properties such as title, abstract, authors, Digital Object Identifier (DOI) and biomedical annotations are indexed to allow semantic search. A web-based user interface and an API have been developed.

* ZB MED Information Centre for Life Sciences , Germany, langnickel@zbmed.de

** ZB MED Information Centre for Life Sciences , Germany, baum@zbmed.de

*** ZB MED Information Centre for Life Sciences , Germany, darms@zbmed.de

**** ZB MED Information Centre for Life Sciences , Germany, fluck@zbmed.de

Results

In March 2021, the service already integrates more than 26,000 COVID-19 related preprints from currently seven different preprint servers. The web interface displays preprint publications with their details like title, abstract, authors, DOI, date of publication and a link to the original preprint server. A search interface is provided that allows both simple and complex search queries. Through the integration of text mining, semantic search capacities are enhanced. Furthermore, preVIEW offers facet searching, export functionality and an API access. The service is publicly available under <https://preview.zbmed.de>.

Discussion

preVIEW has been developed based on the acute need during the current pandemic. Within the short development time, other systems with similar goals emerged and existing systems started to integrate preprints. Nevertheless, the feature of semantic search provided by text mining tools is still unique. Additionally, there is a potential risk in the aggregation of peer-reviewed and non-reviewed articles and therefore, it is reasonable to handle those domains by different systems.

Keywords: COVID-19, Information Retrieval, Biomedical Text Mining, Semantic Search Engine

Bibliography

1. Langnickel L, Baum R, Darms J, Madan S, Fluck J. COVID-19 preVIEW: Semantic Search to Explore COVID-19 Research Preprints. In: Studies in Health Technology and Informatics. IOS Press; 2021. doi:10.3233/SHTI210124