Coronavirus knowledge graph(CovKG) is a knowledge graph we have built focusing on coronaviruses. Including Severe Acute Respiratory Syndrome(SARS), Middle East Respiratory Syndrome(MERS), Coronavirus Disease 2019(COVID-19) and other coronaviruses. We used CovKG to predict potential targets and drugs for coronaviruses.

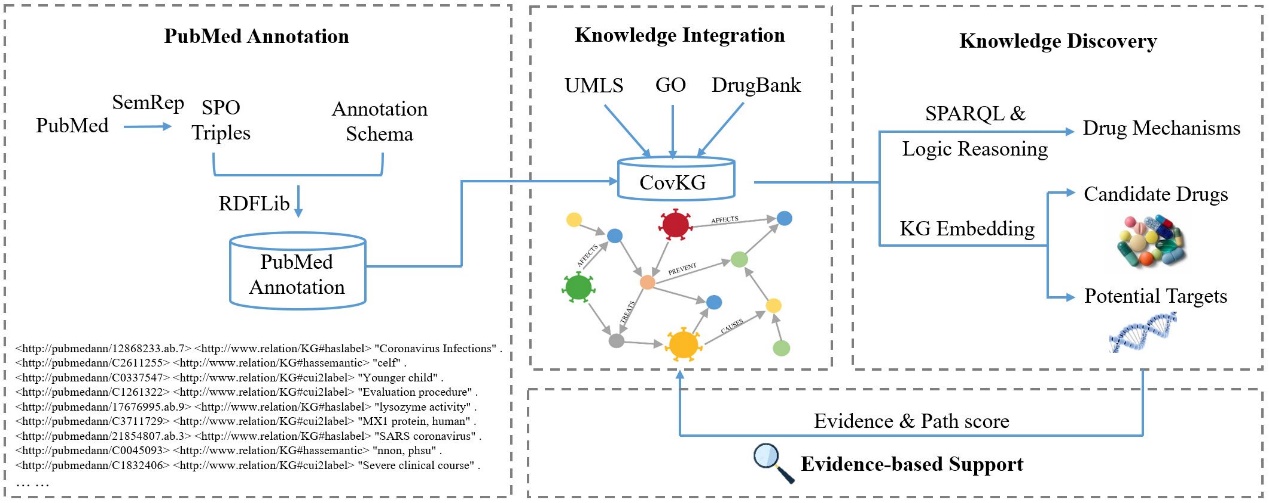


Fig. The workflow of our approach.

* Data

Our constructed PubMedAnn contains 18,687 coronavirus articles from January 2000 to June 2022. The transformed data was stored in a graph database, and we named it PubMedAnn with the namespace “http://www.imicams.ac.cn/pmd-cov-kg” (pubmedann202206rdf.zip).

In this study, we adopted the DrugBank data in RDF format, and defined a namespace “http://www.drugbank.ca/drugs” for it to organize semantic information across multiple knowledge bases.

The GO data is provided in OWL format, and we defined namespace “http://purl.obolibrary.org/obo/go” for it correspondingly.

The entity, relation, and triplet data we used for the experiment are in the file experiment\_data.zip.

* Model&Result

PubMedAnn construction

The triples extracted from the PubMed literature are organized through the semantic conversion model, and the knowledge is organized in an orderly manner, and the knowledge is stored in the RDF format. The semantic conversion model code is in file semantic conversion model.txt.

KG embedding：

We use the python implementation KG embedding method, the specific code we use <https://github.com/awslabs/dgl-ke>.

The link prediction results of the 6 models are in the files result\_linkpredict1.xlsx, result\_linkpredict2.xlsx.

The results of drug similarity calculation are in the file result\_sim.xlsx.

Semantic reasoning：

SPARQL code for the graph inference method we use:



