COVID-19 Explorer Documentation

COVID-19 iKnow Content Navigator - Domain Explorer - Documentation for end users

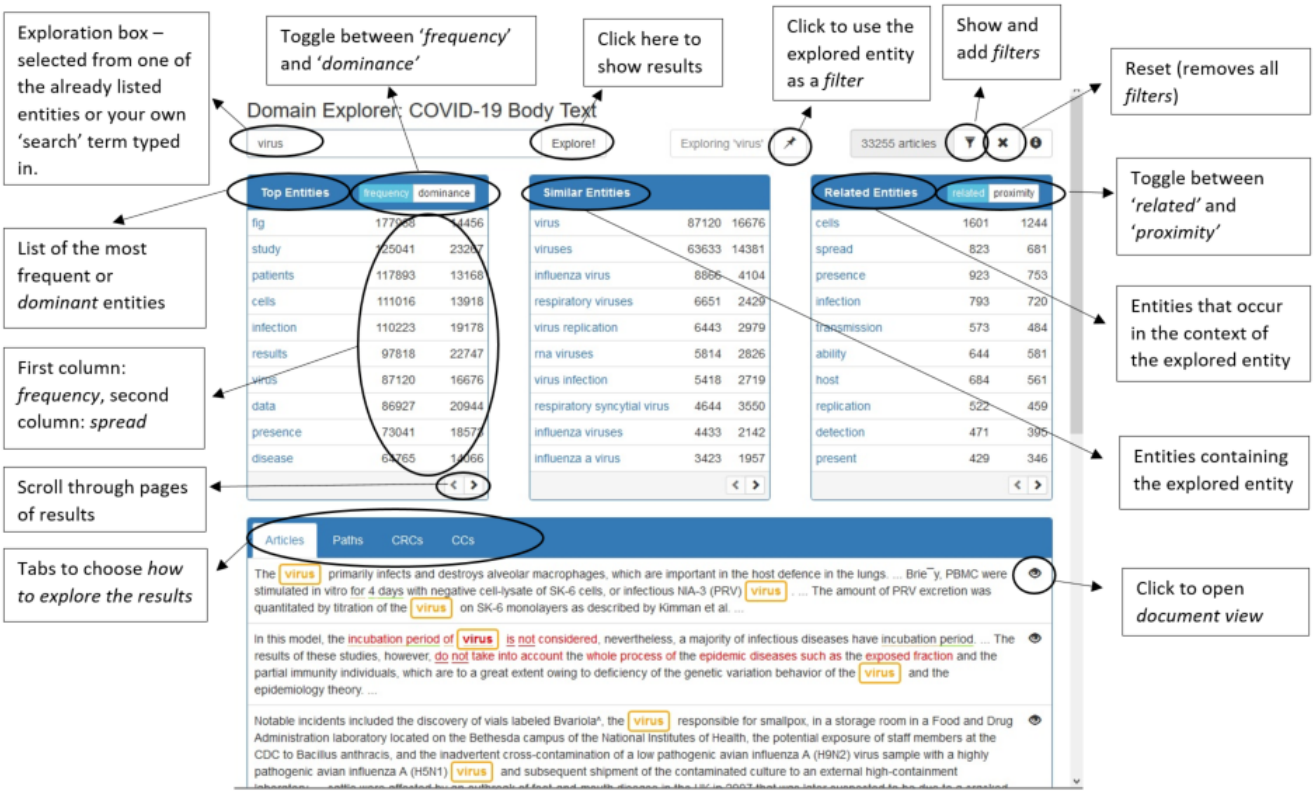
Background information

This data exploration application for the COVID-19 data set (CORD-19, 2020) is built on top of the InterSystems' iKnow NLP technology. The iKnow technology analyses written text and identifies the "Concepts" and the "Relations" between those Concepts. Concepts and Relations are called "Entities". More information about the technology can be found in the wiki section of our GitHub repository: https://github.com/intersystems/iknow/wiki

Using the Domain Explorer

Start the exploration

The iKnow Domain Explorer is not a traditional search tool. It enables you to explore what is in the data set even if you don't know what you are looking for. You can start diving into the data by selecting an entity from the Top Entities or - if you do know what you are looking for - you can start typing in the exploration box in the top left corner. The Explorer will start suggesting actual entities from the data set that match what you are typing in so that you can explore even faster. Click "Explore!" to initiate exploration.



Terms in italics are explained below.

Frequency: the number of occurrences of the entity in the data set

Dominance: a metric expressing entity relevance within the data set

Spread: the number of distinct documents in which the explored entity appears

Filter: a means to limit the number of articles that is taken into consideration. See below for details.

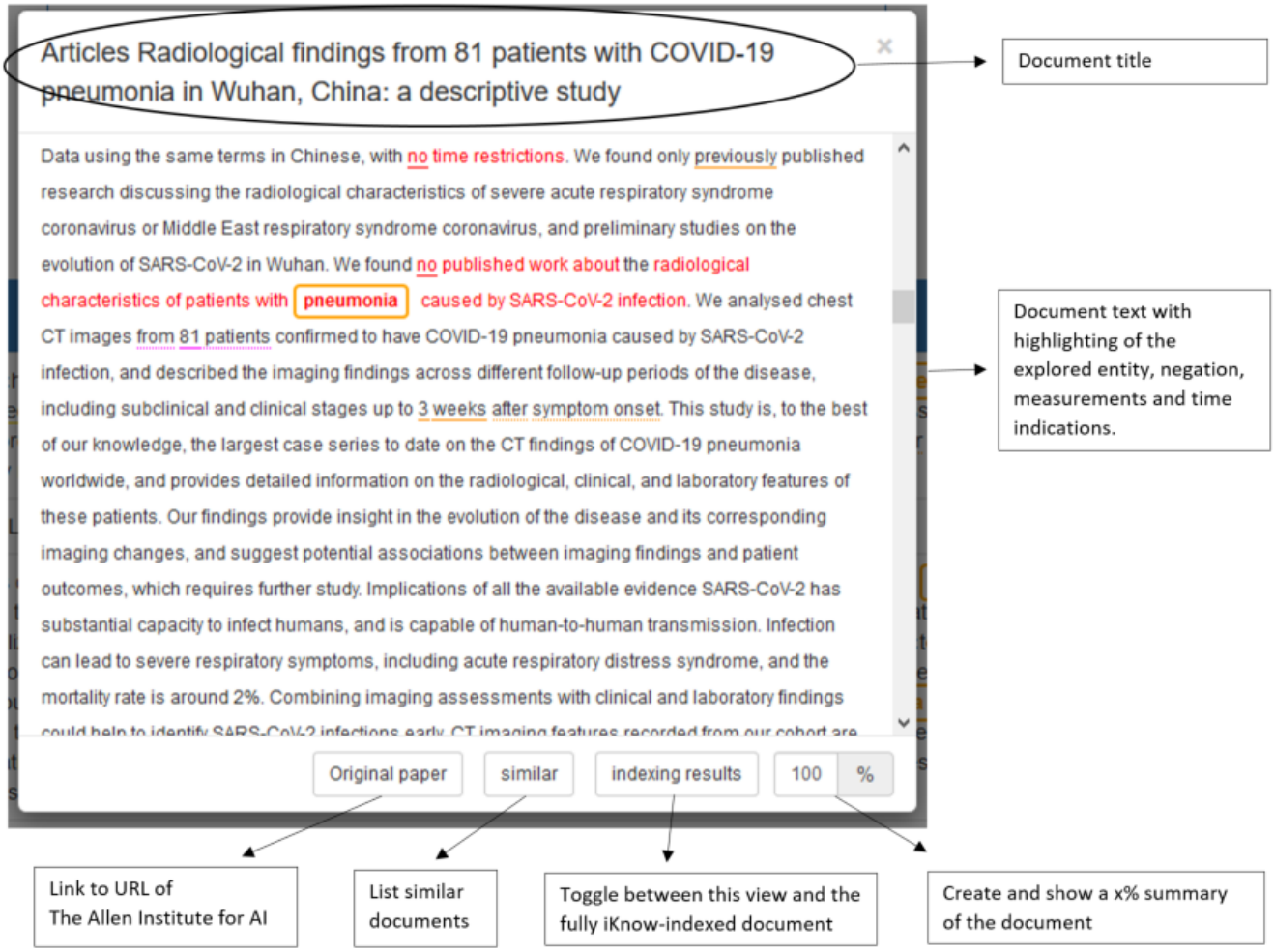
Related versus ***proximity***: "related" shows all other concepts the entity being explored is directly related to (with a single relational element in-between). When "related" is selected, the frequency and spread numbers express how often the concepts appear in a direct relationship. If you toggle to "proximity" in the title bar, the relationship can now also be an indirect one, with the concepts appearing at a longer distance within the same sentence.

How to explore the results: The full-width widget below will show the **articles** (documents), **paths** (sentence subsections), **CRCs** (concept-relation-concept triples) and **CCs** (concept-related concept) in which the entity appears, with the entity being explored highlighted in orange. In the Articles view, text fragments containing the entity being explored are displayed, with negative sections marked in red.

Document view: Pop-up window with additional functionality. See below for details.

Explore the context

When looking at sources or paths in the main bottom widget, you can click the eye icon to consult the complete source text and optionally highlight indexing results. Using the controls at the bottom, you can also look up the source's metadata and generate summaries of the text. In other words, after the initial exploration points you to a couple of documents, you can quickly find the original context and read a summary of the entire document.

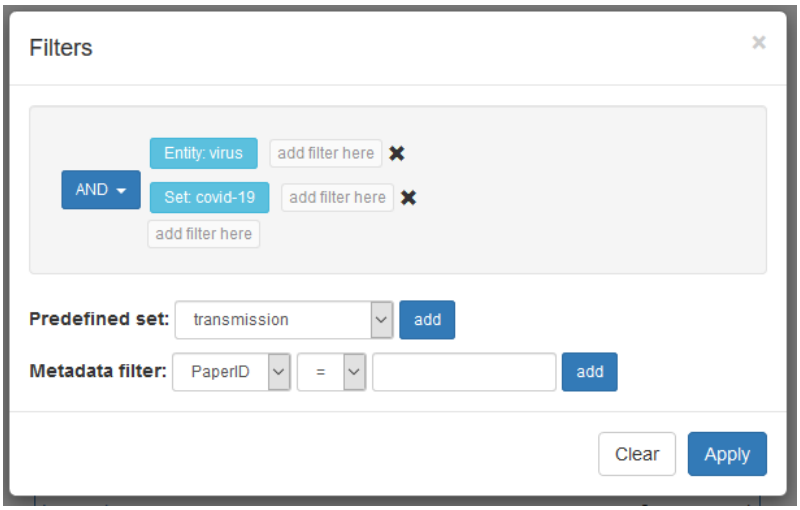


Focus by filtering

The navigator contains three types of filters.

1. Each entity, CRC or CC can be used to filter the results, i.e., to limit the exploration to the articles that contain the selected entity, CRC or CC. To define an entity, CRC or CC as a filter, just click the  button next to the element in the navigator. The filter button in the upper right corner will turn green:  .
2. The navigator also contains a number of predefined terms - called "sets" - that can be used as filters.
3. (Moreover, the metadata PaperID, Source and Title can be added as filters. - *not available yet*)

All types of filters can be managed through the dialogue box that opens when you click the filter button. The entities, CRCs and CCs you selected will appear automatically in the upper part. Use the dropdown menus and click the "add" buttons to define additional filters. Activate them by clicking "Apply".



The filter icon will remain highlighted as long as any filter conditions are active and the number of sources that are being explored is updated to reflect the filter conditions.

Reference for the COVID-19 dataset:

COVID-19 Open Research Dataset (CORD-19). 2020. Version 2020-03-20. Retrieved from https://pages.semanticscholar.org/coronavirus-research. Accessed 2020-03-27. doi:10.5281/zenodo.3715505