SPIKE-CORD: Extractive Search Over CORD-19

This a power-tool for performing extractive search, using various query modes. It allows a level of query expressivity and control that is substantially more powerful than existing search solutions. This tool searches over the COVID-19 Open Research Dataset (CORD-19), a free resource of over 50K scholarly articles about COVID-19 and related coronaviruses provided by AI2's Semantic Scholar project.

Example queries:

- Boolean: coronavirus lemma=treatment (run) (Sentences with specific terms.)
- Token: incubation period ... from:* -|to to:* days (run) (Extract incubation periods.)
- Structured: <>v:virus $infection $causes a <>c:condition . (run) (Things caused by kinds of infections.)

Contents

Allen Institute for Artificial Intelligence (AI2)

Vision

More powerful searches

Mission

To enable extractive searching

Values

Discovery: There are many search systems over the CORD-19 resource. What is special about this one?

Complexity: Ours is more complex and harder to use!

Power: However, it is also much more powerful.

Cooperation: We are not competing with other systems – we aim to be complementary to them.

Relationships: Why is this useful? The various query modes can be used to powerfully enhance search (by looking for related words, related terms, etc)

Interpretation: They can be used to extract information to be aggregated and interpreted by users

Training: They can be used to extract examples for data annotation, which can then be used to train machine learning models.

Searching

Enable sentence-level, context-aware, and linguistically informed extractive search

SPIKE Extractive Search is a sentence-level, context-aware, and linguistically informed extractive search system. That is a mouthful, so let's unpack:

1. Sentences

Enable searches by sentences, not paragraphs or documents.

Sentence-level: our main objects of search are sentences, not paragraphs or documents.

2. Context

Enable searches for sentences that appear within paragraphs or within papers based upon words in their paragraphs.

Context-aware: we do not lose sight of the document. We can search for sentences that appear within a paragraph with certain words, or within papers that have certain words in their paragraphs, etc.

3. Linguistic Analysis

Perform linguistic analysis over documents.

Linguistically-informed: we performed linguistic analysis over the documents, and the queries can make use of this information.

4. Words & Spans

Extract words or spans from a sentence into tables.

Extractive: we can extract specific words or spans from a sentence into tables.

Administrative Information
Start Date:
End Date:

Publication Date: 2020-04-28
Source: https://spike.covid-19.apps.allenai.org/search/covid19

Submitter:
Given Name: Owen
Surname: Ambur
Email: Owen.Ambur@verizon.net
Phone: