

# COVID-19 Search System

Sophia Ananiadou, Chryssa Zerva, Axel Soto, Sam Taylor

# Motivation

- COVID-19 led to an overload of textual resources
- Overlook of information, difficulties in discovering knowledge, navigation through differences sources
- **Aim: improve navigation, facilitate discovery, make associations**

## Scientific reports:

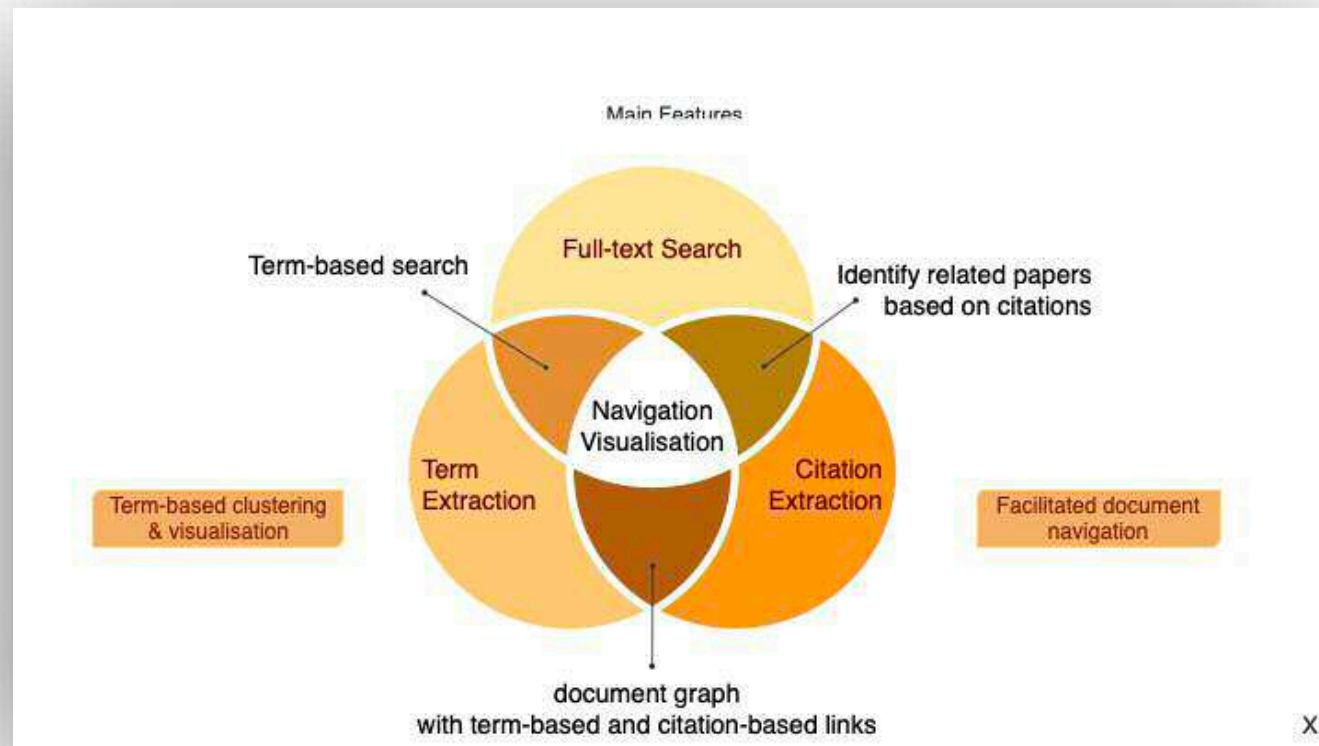
- CORD-19
  - PubMed, BioRxiv, MedRxiv
  - Limited WHO reports
  - Current size: ~500K, ~200 full papers
  - Open-source

Allen Institute for AI in partnership with the Chan Zuckerberg Initiative, Georgetown University's Center for Security and Emerging Technology, Microsoft Research, IBM, and the National Library of Medicine - National Institutes of Health, in coordination with The White House Office of Science and Technology Policy.

<https://allenai.org/data/cord-19>

# Main facets

<http://www.nactem.ac.uk/cord/>



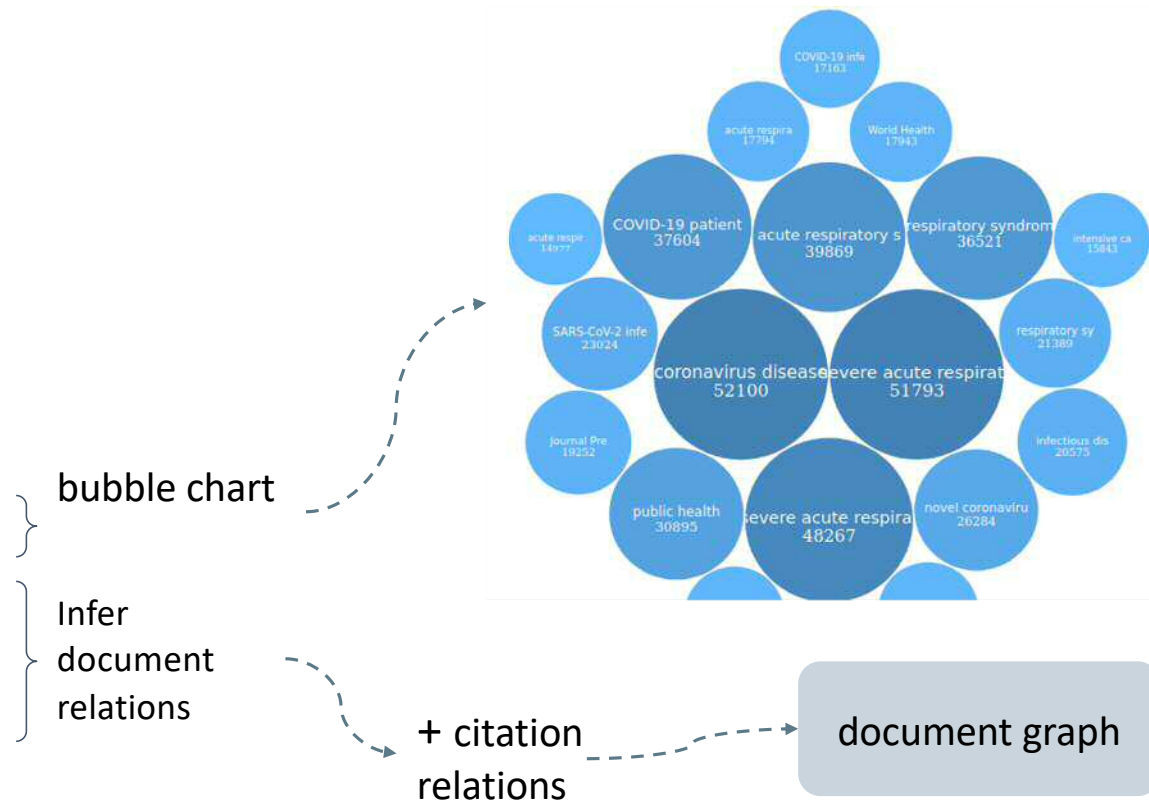
# Term extraction

## Term extraction:

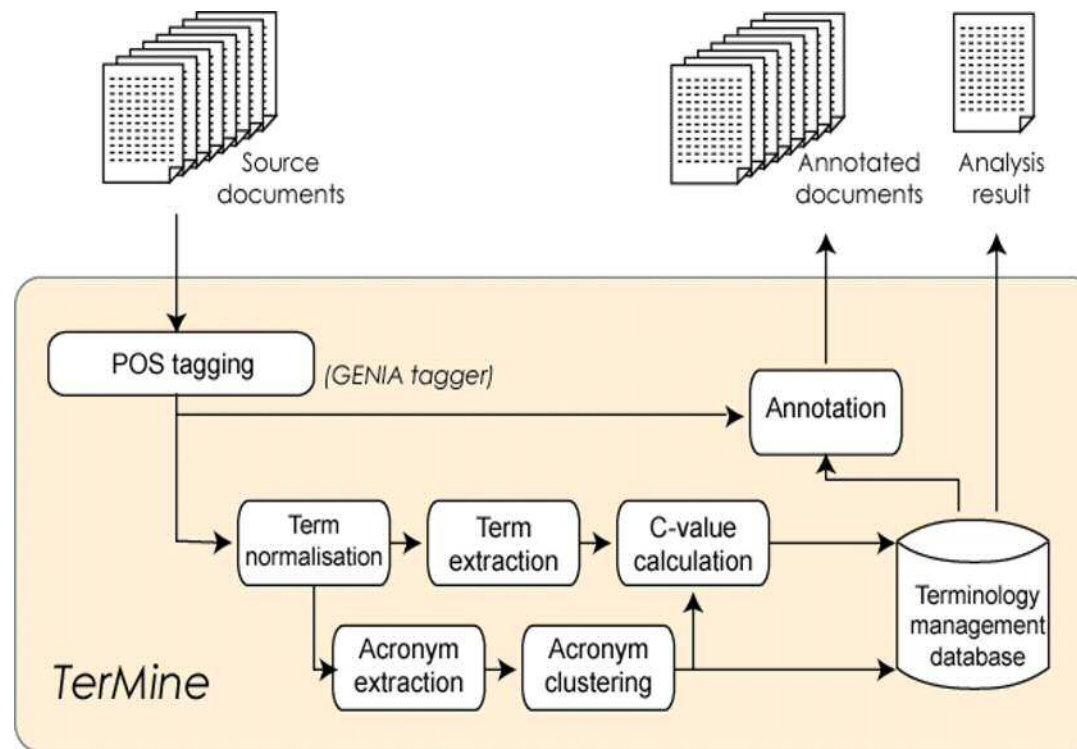
- Termine API
- Extract multi-word terms
- Rank by significance
  - Document-level
  - Query-level

## Term usage:

- Weighted facet search
  - Filter down results
- Topic inference
  - Term co-occurrence
  - Term hierarchy



# TerMine: A Term Extraction System



Frantzi, K. T., & Ananiadou, S. (1999). The C-value/NC-value domain-independent method for multi-word term extraction. *Journal of Natural Language Processing*, 6(3), 145-179.

# Term bubbles and queries

The screenshot displays the NaCTeM (The National Centre for Text Mining) interface. At the top, the search query is "transmission control". Below the search bar, there are tabs for "Bubble Chart", "Terms", "Entities", "Year", and "Journal". The "Bubble Chart" tab is active, showing a cluster of blue bubbles representing related terms and their frequencies. The largest bubbles are labeled "World Health Organization" (77), "Public Health" (88), "World Health" (88), "infectious disease" (92), "severe acute respiratory..." (92), "acute respiratory..." (94), "high level" (78), "previous study" (74), "disease control" (62), "important role" (46), "viral infection" (46), "health care" (36), "severe acute..." (31), "other hand" (69), "recent study" (62), "control measure" (61), "acute respiratory..." (27), "Middle East..." (18), and "lar" (partially visible).

On the right side, the search results are displayed. The first result is "Mask Wearing and Control of SARS-CoV-2 Transmission in the United States" with a score of 3.0427597 and ID 176132. The second result is "Characteristic changes of traumatic dental injuries in a teaching hospital of Wuhan under transmission control measures during the COVID-19 epidemic" with a score of 3.021086 and ID 165114. The third result is "Characteristic changes of traumatic dental injuries in a teaching hospital of Wuhan under transmission control measures during the COVID-19 epidemic Running title: Traumatic dental injuries during the COVID-19 epidemic Characteristic changes of traumatic dental injuries in a teaching hospital of Wuhan under transmission control measures during the COVID-19 epidemic" with a score of 2.949275 and ID 243121. The fourth result is "The focus and timing of COVID-19 pandemic control measures under healthcare resource constraints" with a score of 2.946024 and ID 220754.



# CORD-19

"transmission control"



Step-by-step guide

Feedback

Bubble Chart Terms Entities Year Journal

92 articles found in 0.444 seconds.

List Network

## Emerging infections in Asia and its possible global effects

Score: 2.057749

id: 176658

... **transmission control** ... [More snippets...](#)

## Potential implications of SARS-CoV-2 on pregnancy

Score: 2.0461283

id: 203236

... **transmission control** measure ...

## Viral shedding patterns of coronavirus in patients with probable severe acute respiratory syndrome

Score: 2.03865

id: 200596

... development of an appropriate testing strategy and **transmission control** measures for SARS. ... [More snippets...](#)

## Bayesian modeling of COVID-19 cases with a correction to account for under-reported cases

Score: 2.032786

id: 235307

... quarantine for the strategy of the COVID-19 **transmission control** . O. N. (2018) . Epidemics: models and ... [More snippets...](#)

## Children and Fecal SARS-CoV-2 shedding: Just the Tip of the Iceberg of Italian COVID-19 outbreak?

Score: 2.0186636

id: 177605

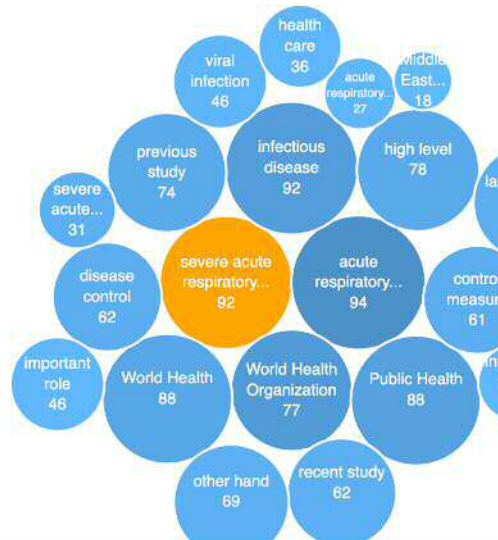
... in children is of relevant importance for **transmission control** and appropriate healthcare policies. Recently ... [More snippets...](#)

## Review of economic evaluations of mask and respirator use for protection against respiratory infection transmission

Score: 2.006468

id: 223835

... other two studies focussed on tuberculosis **transmission control** interventions. The settings and methodologies ... [More snippets...](#)



"copd exacerbation"



Step-by-step guide

Feedback

Bubble Chart Terms Entitles Year Journal

add

Termine	494633	c-value	count
x chronic obstructive pulmonary disease	634	228	
x viral infection	446	302	
x severe acute respiratory syndrome	376	171	
x COPD exacerbation	363	276	
x intensive care unit	328	171	
x pulmonary disease	308	252	
x recent study	304	280	
x high level	300	256	
x acute respiratory syndrome	298	174	
x respiratory tract infection	288	158	
x respiratory tract	264	206	
x previous study	263	205	
x acute exacerbation	258	179	
x T cell	249	141	
x Risk factor	244	189	
x immune response	241	193	
x Respiratory syncytial virus	235	136	

2089 articles found in 0.645 seconds.

List Network

## Factors associated with chronic obstructive pulmonary disease exacerbation, based on big data analysis

Score: 13.095466 id: 186336  
 ... reported. To identify factors associated with **COPD exacerbation**, we collected data, including smoking status ... [More snippets...](#)  
 ... viral infection, are known to contribute to **COPD exacerbation** [7] [8] [9], few studies have analysed the ...  
 ... COPD-related web queries were associated with **COPD exacerbation**. Multivariate analysis revealed that the number ...  
 ... reported. To identify factors associated with **COPD exacerbation**, we collected data, including smoking status ...  
 ... COPD-related web queries were associated with **COPD exacerbation**. Multivariate analysis revealed that the number ...

## Association of COPD exacerbation frequency with gene expression of pattern recognition receptors in inflammatory cells in induced sputum.

Score: 13.050671 id: 83995  
 ... Association of **COPD exacerbation** frequency with gene expression of pattern recognition receptors in inflammatory ...  
[More snippets...](#)

## Serum inflammatory biomarkers and clinical outcomes of COPD exacerbation caused by different pathogens.

Score: 12.955318 id: 38512  
 ... outcomes of **COPD exacerbation** caused by different pathogens. BACKGROUND AND OBJECTIVE **COPD exacerbation** is characterized ... [More snippets...](#)

## Models of infection and exacerbations in COPD This review comes from a themed issue on Respiratory pharmacology Edited by Brendan Canning and Stephen Farmer

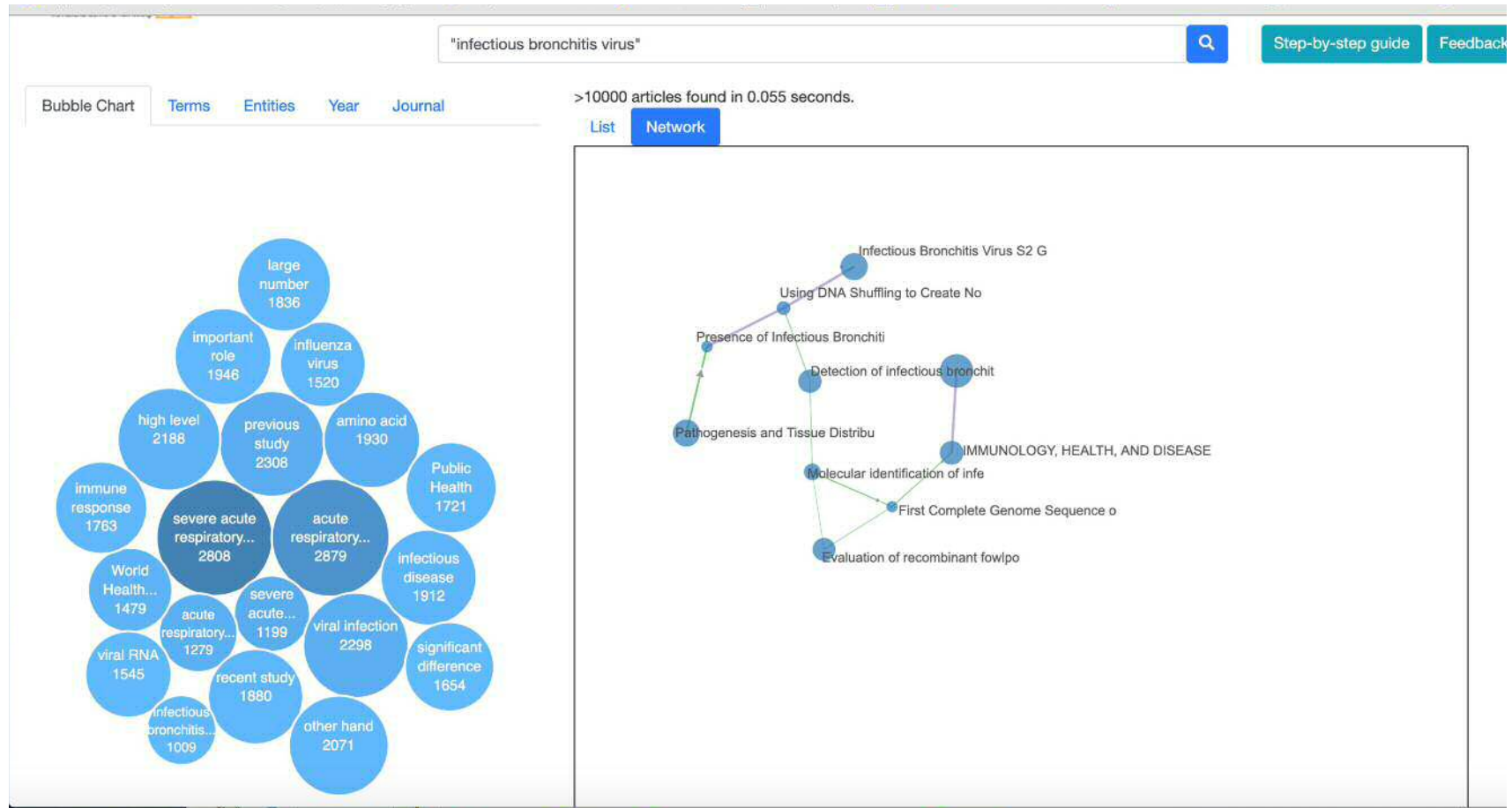
Score: 12.874373 id: 204165  
 ... development of the first human rhinovirus-induced **COPD exacerbation** model represents an invaluable tool towards ...  
[More snippets...](#)



# Type of queries

- Autocomplete
- OR operator: Using free text, or consecutive terms in quotes will behave as an OR operator between queries. E.g. try “coronavirus transmission risk” “public transport”, and “coronavirus transmission risk in public transport”
- AND operator: Using the + symbol between quoted queries will behave as an AND operator between queries. E.g. try “coronavirus transmission risk” + “public transport”

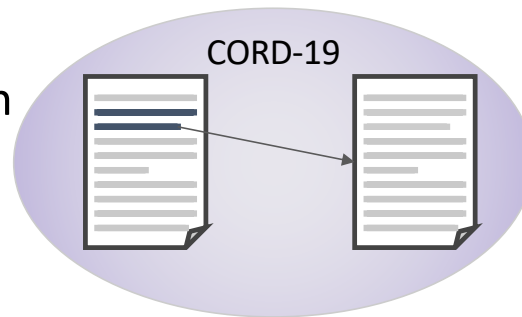
# Document Network



# Citations and network visualisation

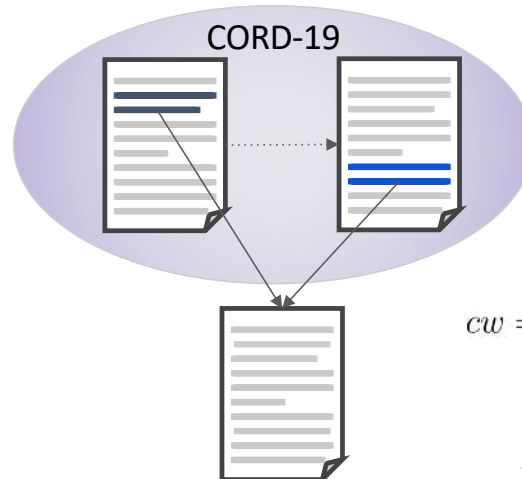
## Direct citation links:

Refers to citations between two documents within the dataset



## Indirect citation links:

Inferred link between two documents citing the same publication outside the dataset



## Weighting of citations:

- **Zoning information**  
(zone\_w)  
Intuition: citations mentioned in Methods and Discussion sections more important \*
- **Repetition of citation**  
(repeat\_w)  
Intuition: increased mentions of the same citation signify important
- **Relative publication date**  
Intuition: older citations tend to be less important

$$cw = zone\_w \cdot repeat\_w \cdot \frac{1}{\max(doc_{Ayear} - doc_{Byear}, 1)}$$

\*Taşkın, Z., & Al, U. (2018). A content-based citation analysis study based on text categorization. *Scientometrics*, 114(1), 335-357.

# Evaluation

TREC CORD task performance

	NDCG	BPREF	MAP@10	P@5	P@10	<b>50Q time</b>
<b>Full text</b>	0.391	0.251	0.170	0.549	0.433	03:30
<b>Abstract/title</b>	0.403	0.241	0.148	0.620	0.400	03:03
<b>Paragraph</b>	0.582	0.258	0.165	0.680	0.648	04:23
<b>Sentence selection</b>	0.625	0.232	0.155	0.704	0.700	04:35
<b>Term-based reranking</b>	0.689	0.230	0.190	0.715	0.745	12:56

# User evaluation

- Please fill in our user evaluation after the demo

[USER EVALUATION](#)

Thank you