THE UNIVERSITY OF MELBOURNE

ARC TRAINING CENTRE IN COGNITIVE COMPUTING FOR MEDICAL TECHNOLOGIES

IBM

COVID-SEE

Scientific Evidence Explorer for COVID-19

Simon Šuster

Stream 4: Real-time clinical decision support

17/7/2020

Team

Led by **Karin Verspoor**

Core team: **Simon Suster**, **Yulia Otmakhova**, **Zenan Zhai**, **Biaoyan Fang**

Help from colleagues **Jey-Han Lau** and **Tim Baldwin**, and external collaborators **Antonio Jimeno Yepes** and **David Martinez** (IBM Research Melbourne)

Web developer: **Shevon Mendis**

Timeline

13 March
CORD Challenge and Dataset
by Kaggle, AI2 and others

27 March
We have a team!
Covidex.ai released

16 April
Kaggle Round 1 deadline
500 submissions

Missed the deadline :(

16 June
Kaggle Round 2 deadline
+100 submissions

Our first prototype COVID-SEE

Today
A working system :)

And a submission to EMNLP System Demonstrations

Goals

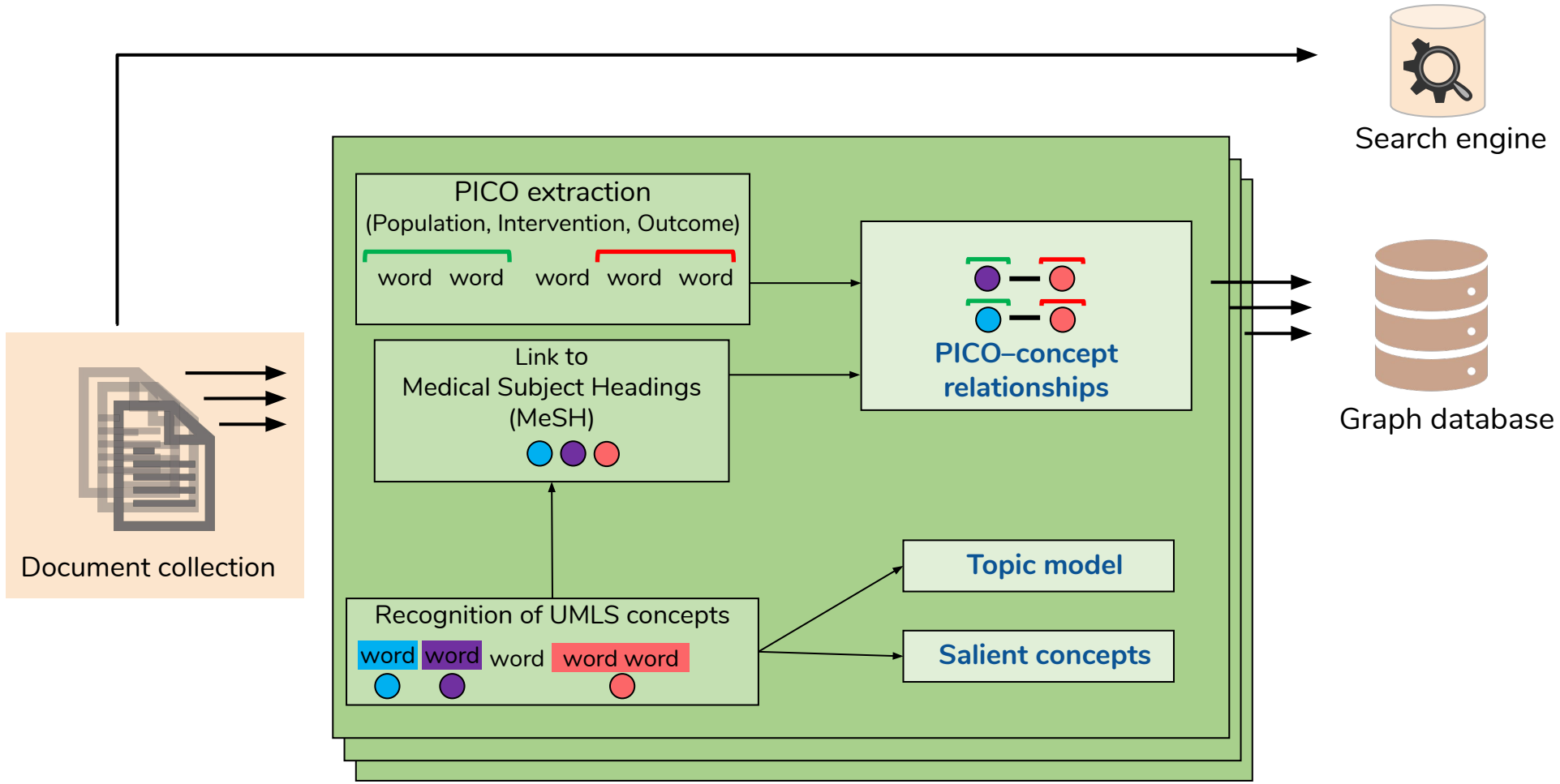
A large number of systems perform lookup

- information retrieval and QA with free-text queries (cord19.vespa.ai, discovid.ai, covid19.mendel.ai)
- semantic search (IBM COVID-19 Navigator, DOC Search, Trialstreamer)

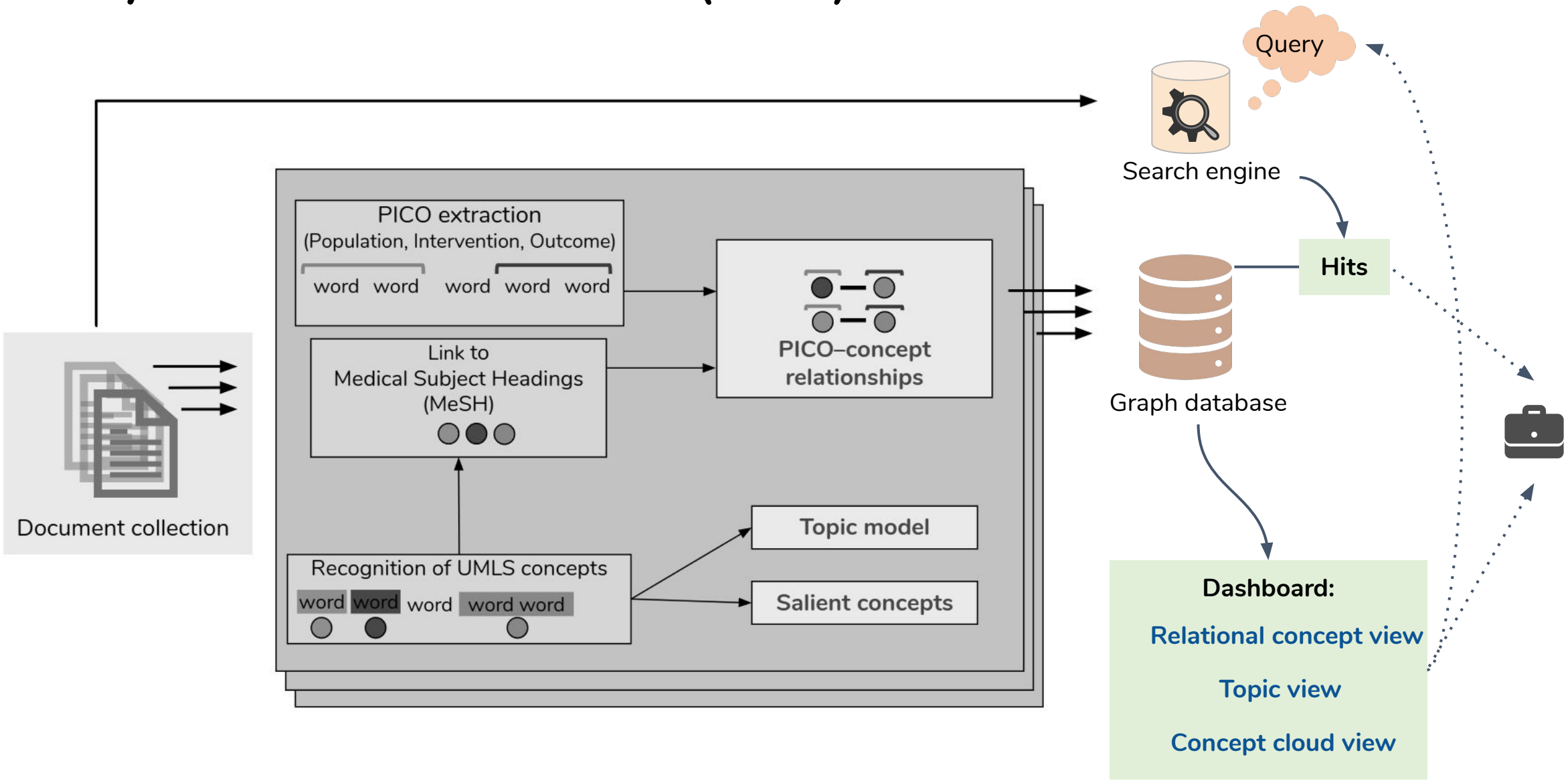
We focus on interactive exploration of literature (White and Roth, 2009; Pang et al., 2015)

- initiate search with a tentative query in natural language
- explore different views of the retrieved documents and obtain cues about next steps
- high-level (current collection) and low-level (document) views
- progressively build up your briefcase

System overview (1/2)



System overview (2/2)



Information retrieval

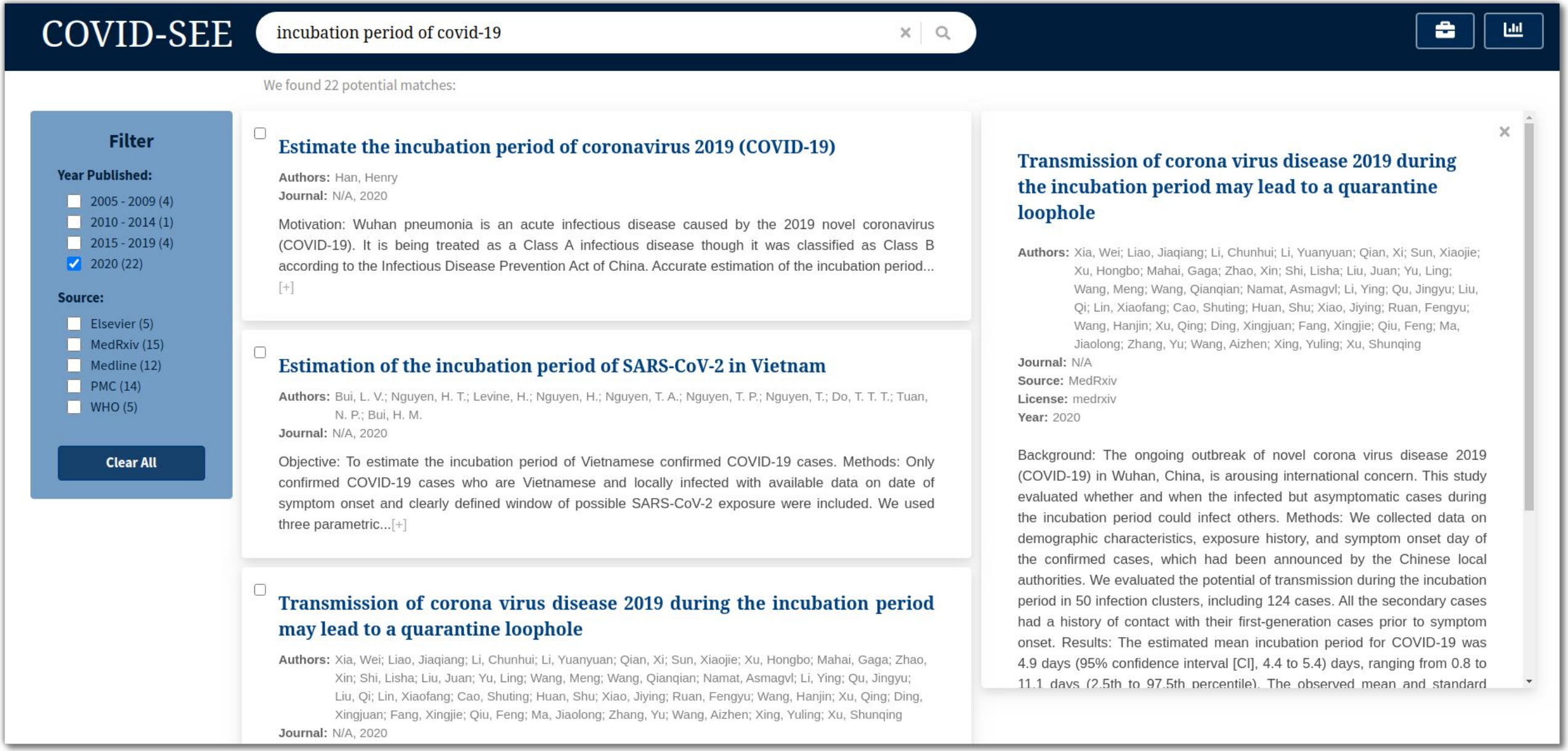
Use Anserini (Zhang et al., 2020) in Python, builds on Lucene search

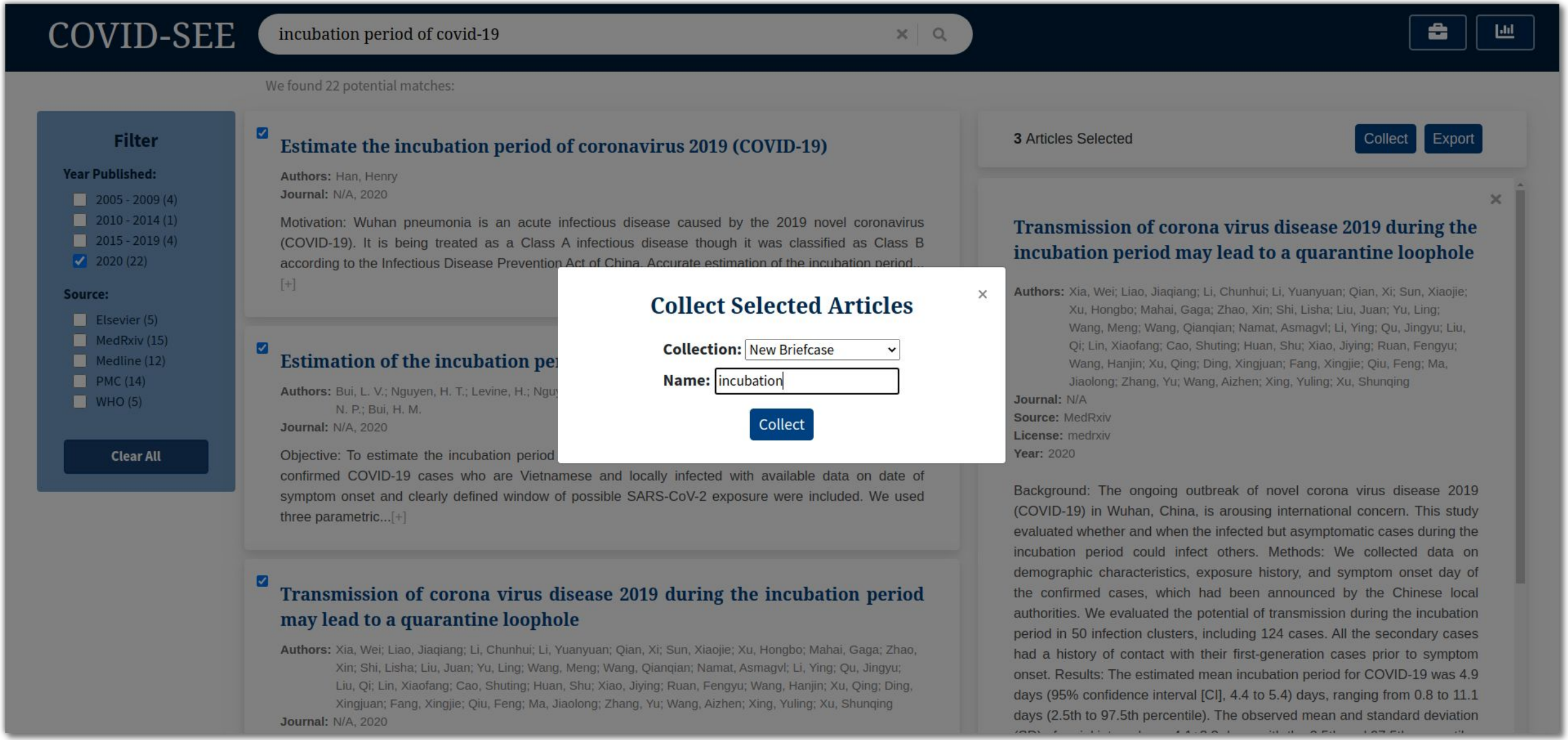
- index from CORD-19 abstracts
- retrieved documents ranked with BM25 scoring function

Simple search, but in top 30% of submissions to TREC-COVID shared task (Roberts et al., 2020)

Possible extensions with neural re-ranking


?
COVID-SEE
Search literature...
Search ›


COVID-SEE
incubation period of covid-19
We found 22 potential matches:
Filter
Year Published:
2005 - 2009 (4)
2010 - 2014 (1)
2015 - 2019 (4)
2020 (22)
Source:
Elsevier (5)
MedRxiv (15)
Medline (12)
PMC (14)
WHO (5)
Clear All
Estimate the incubation period of coronavirus 2019 (COVID-19)
Authors: Han, Henry
Journal: N/A, 2020
Motivation: Wuhan pneumonia is an acute infectious disease caused by the 2019 novel coronavirus (COVID-19). It is being treated as a Class A infectious disease though it was classified as Class B according to the Infectious Disease Prevention Act of China. Accurate estimation of the incubation period... [+]
Estimation of the incubation period of SARS-CoV-2 in Vietnam
Authors: Bui, L. V.; Nguyen, H. T.; Levine, H.; Nguyen, H.; Nguyen, T. A.; Nguyen, T. P.; Nguyen, T.; Do, T. T. T.; Tuan, N. P.; Bui, H. M.
Journal: N/A, 2020
Objective: To estimate the incubation period of Vietnamese confirmed COVID-19 cases. Methods: Only confirmed COVID-19 cases who are Vietnamese and locally infected with available data on date of symptom onset and clearly defined window of possible SARS-CoV-2 exposure were included. We used three parametric...[+]
Transmission of corona virus disease 2019 during the incubation period may lead to a quarantine loophole
Authors: Xia, Wei; Liao, Jiaqiang; Li, Chunhui; Li, Yuanyuan; Qian, Xi; Sun, Xiaojie; Xu, Hongbo; Mahai, Gaga; Zhao, Xin; Shi, Lisha; Liu, Juan; Yu, Ling; Wang, Meng; Wang, Qianqian; Namat, Asmagvl; Li, Ying; Qu, Jingyu; Liu, Qi; Lin, Xiaofang; Cao, Shuting; Huan, Shu; Xiao, Jiying; Ruan, Fengyu; Wang, Hanjin; Xu, Qing; Ding, Xingjuan; Fang, Xingjie; Qiu, Feng; Ma, Jiaolong; Zhang, Yu; Wang, Aizhen; Xing, Yuling; Xu, Shunqing
Journal: N/A, 2020
Transmission of corona virus disease 2019 during the incubation period may lead to a quarantine loophole
Authors: Xia, Wei; Liao, Jiaqiang; Li, Chunhui; Li, Yuanyuan; Qian, Xi; Sun, Xiaojie; Xu, Hongbo; Mahai, Gaga; Zhao, Xin; Shi, Lisha; Liu, Juan; Yu, Ling; Wang, Meng; Wang, Qianqian; Namat, Asmagvl; Li, Ying; Qu, Jingyu; Liu, Qi; Lin, Xiaofang; Cao, Shuting; Huan, Shu; Xiao, Jiying; Ruan, Fengyu; Wang, Hanjin; Xu, Qing; Ding, Xingjuan; Fang, Xingjie; Qiu, Feng; Ma, Jiaolong; Zhang, Yu; Wang, Aizhen; Xing, Yuling; Xu, Shunqing
Journal: N/A
Source: MedRxiv
License: medrxiv
Year: 2020
Background: The ongoing outbreak of novel corona virus disease 2019 (COVID-19) in Wuhan, China, is arousing international concern. This study evaluated whether and when the infected but asymptomatic cases during the incubation period could infect others. Methods: We collected data on demographic characteristics, exposure history, and symptom onset day of the confirmed cases, which had been announced by the Chinese local authorities. We evaluated the potential of transmission during the incubation period in 50 infection clusters, including 124 cases. All the secondary cases had a history of contact with their first-generation cases prior to symptom onset. Results: The estimated mean incubation period for COVID-19 was 4.9 days (95% confidence interval [CI], 4.4 to 5.4) days, ranging from 0.8 to 11.1 days (2.5th to 97.5th percentile). The observed mean and standard


COVID-SEE
incubation period of covid-19
We found 22 potential matches:
Filter
Year Published:
2005 - 2009 (4)
2010 - 2014 (1)
2015 - 2019 (4)
2020 (22)
Source:
Elsevier (5)
MedRxiv (15)
Medline (12)
PMC (14)
WHO (5)
Clear All
Estimate the incubation period of coronavirus 2019 (COVID-19)
Authors: Han, Henry
Journal: N/A, 2020
Motivation: Wuhan pneumonia is an acute infectious disease caused by the 2019 novel coronavirus (COVID-19). It is being treated as a Class A infectious disease though it was classified as Class B according to the Infectious Disease Prevention Act of China. Accurate estimation of the incubation period...
[+]
Estimation of the incubation pe
Authors: Bui, L. V.; Nguyen, H. T.; Levine, H.; Ngu
N. P.; Bui, H. M.
Journal: N/A, 2020
Objective: To estimate the incubation period
confirmed COVID-19 cases who are Vietnamese and locally infected with available data on date of symptom onset and clearly defined window of possible SARS-CoV-2 exposure were included. We used three parametric...[+]
Transmission of corona virus disease 2019 during the incubation period may lead to a quarantine loophole
Authors: Xia, Wei; Liao, Jiaqiang; Li, Chunhui; Li, Yuanyuan; Qian, Xi; Sun, Xiaojie; Xu, Hongbo; Mahai, Gaga; Zhao, Xin; Shi, Lisha; Liu, Juan; Yu, Ling; Wang, Meng; Wang, Qianqian; Namat, Asmagvl; Li, Ying; Qu, Jingyu; Liu, Qi; Lin, Xiaofang; Cao, Shuting; Huan, Shu; Xiao, Jiying; Ruan, Fengyu; Wang, Hanjin; Xu, Qing; Ding, Xingjuan; Fang, Xingjie; Qiu, Feng; Ma, Jiaolong; Zhang, Yu; Wang, Aizhen; Xing, Yuling; Xu, Shunqing
Journal: N/A, 2020
Collect Selected Articles
Collection: New Briefcase
Name: incubation
Collect
3 Articles Selected
Collect
Export
Transmission of corona virus disease 2019 during the incubation period may lead to a quarantine loophole
Authors: Xia, Wei; Liao, Jiaqiang; Li, Chunhui; Li, Yuanyuan; Qian, Xi; Sun, Xiaojie; Xu, Hongbo; Mahai, Gaga; Zhao, Xin; Shi, Lisha; Liu, Juan; Yu, Ling; Wang, Meng; Wang, Qianqian; Namat, Asmagvl; Li, Ying; Qu, Jingyu; Liu, Qi; Lin, Xiaofang; Cao, Shuting; Huan, Shu; Xiao, Jiying; Ruan, Fengyu; Wang, Hanjin; Xu, Qing; Ding, Xingjuan; Fang, Xingjie; Qiu, Feng; Ma, Jiaolong; Zhang, Yu; Wang, Aizhen; Xing, Yuling; Xu, Shunqing
Journal: N/A
Source: MedRxiv
License: medrxiv
Year: 2020
Background: The ongoing outbreak of novel corona virus disease 2019 (COVID-19) in Wuhan, China, is arousing international concern. This study evaluated whether and when the infected but asymptomatic cases during the incubation period could infect others. Methods: We collected data on demographic characteristics, exposure history, and symptom onset day of the confirmed cases, which had been announced by the Chinese local authorities. We evaluated the potential of transmission during the incubation period in 50 infection clusters, including 124 cases. All the secondary cases had a history of contact with their first-generation cases prior to symptom onset. Results: The estimated mean incubation period for COVID-19 was 4.9 days (95% confidence interval [CI], 4.4 to 5.4) days, ranging from 0.8 to 11.1 days (2.5th to 97.5th percentile). The observed mean and standard deviation

Collections

Selection Manager

Selected Collection: incubation
No. of Articles(Total): 3
No. of Articles(Selected): 0

Move Delete Export

Explorer

Active Collection

incubation

New Briefcase

Collection Viewer

Estimate the incubation period of coronavirus 2019 (COVID-19)

Authors: Han, Henry
Journal: N/A, 2020

Motivation: Wuhan pneumonia is an acute infectious disease caused by the 2019 novel coronavirus (COVID-19). It is being treated as a Class A infectious disease though it was classified as Class B according to the Infectious Disease Prevention Act of China. Accurate estimation of the incubation period...[+]

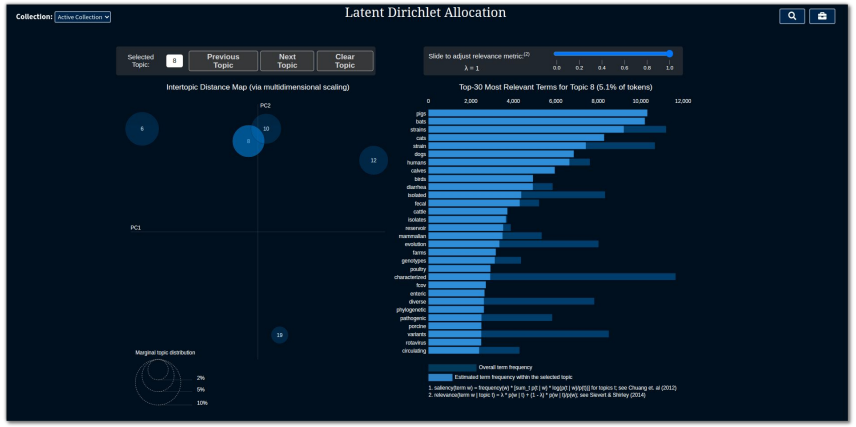
Transmission of corona virus disease 2019 during the incubation period may lead to a quarantine loophole

Authors: Xia, Wei; Liao, Jiaqiang; Li, Chunhui; Li, Yuanyuan; Qian, Xi; Sun, Xiaojie; Xu, Hongbo; Mahai, Gaga; Zhao, Xin; Shi, Lisha; Liu, Juan; Yu, Ling; Wang, Meng; Wang, Qianqian; Namat, Asmagvl; Li, Ying; Qu, Jingyu; Liu, Qi; Lin, Xiaofang; Cao, Shuting; Huan, Shu; Xiao, Jiying; Ruan, Fengyu; Wang, Hanjin; Xu, Qing; Ding, Xingjuan; Fang, Xingjie; Qiu, Feng; Ma, Jiaolong; Zhang, Yu; Wang, Aizhen; Xing, Yuling; Xu, Shunqing
Journal: N/A, 2020

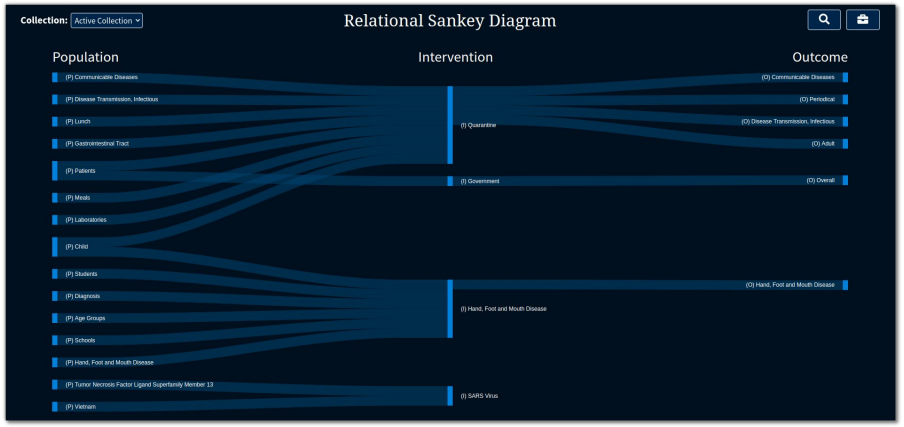
Background: The ongoing outbreak of novel corona virus disease 2019 (COVID-19) in Wuhan, China, is arousing international concern. This study evaluated whether and when the infected but asymptomatic cases during the incubation period could infect others. Methods: We collected data on demographic characteristics,...[+]

Dashboard

1. Relational concept view



2. Topic view



3. Concept cloud view

1. Relational concept view

Identify spans describing populations, interventions/comparators and outcomes (PICO) & recognise MeSH terms

Hospitalization Mortality

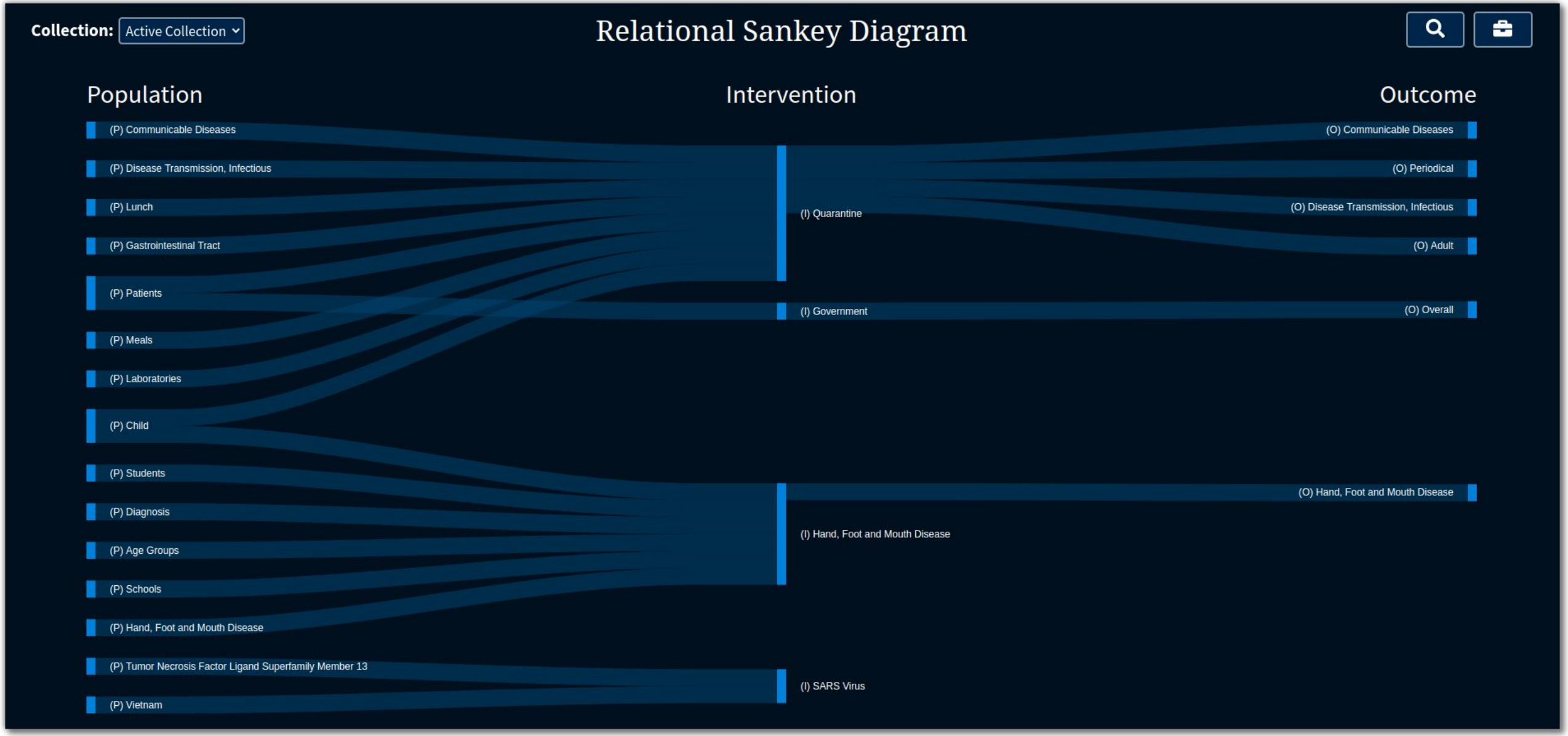
"*cumulative COVID-19-related hospitalization and death rates*"

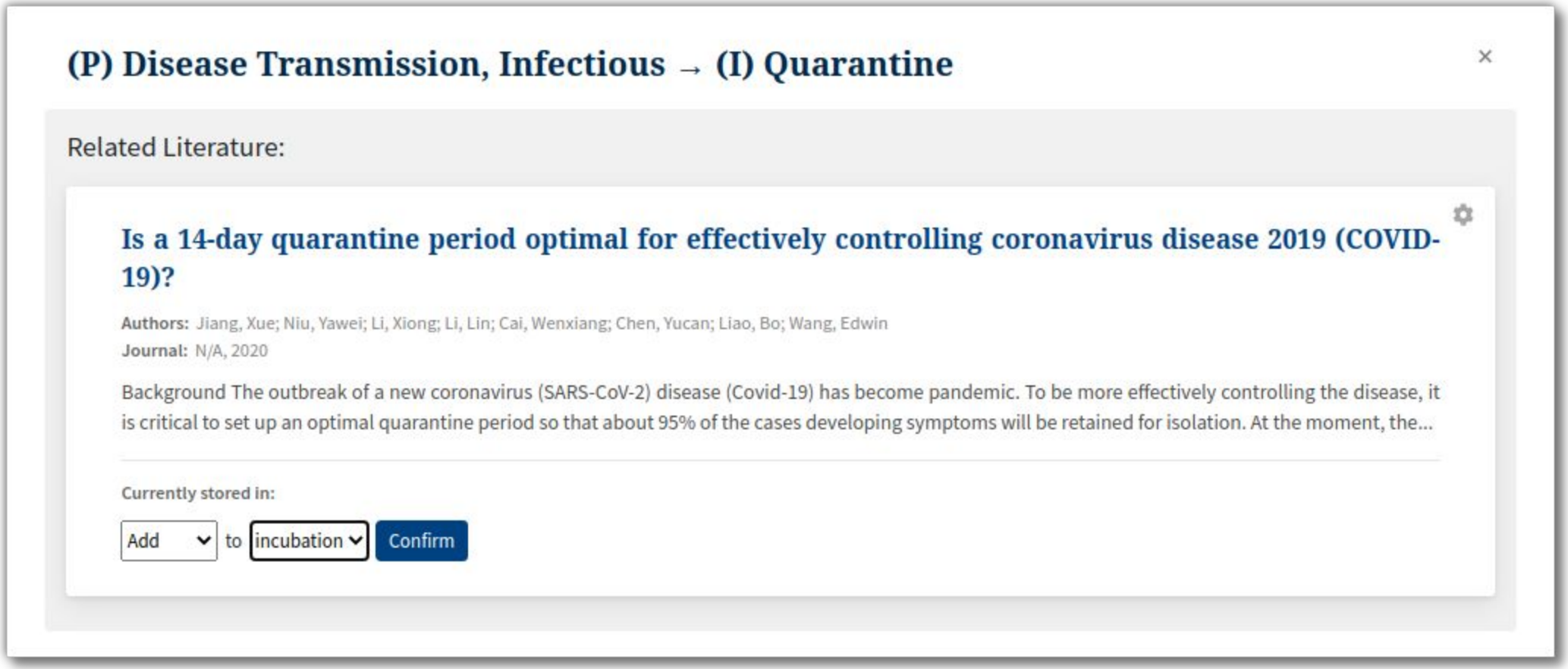
OUTCOME

PICO concepts: *Hospitalization*$_{OUTCOME}$, *Mortality*$_{OUTCOME}$

Relate Population-Intervention (or Intervention-Outcome) concepts if they occur in the same abstract

- BiLSTM-CRF labeler from 5,000 annotated PubMed abstracts (EBM-NLP, Nye et al., 2018)
- Test set F1: 0.69 for all PICO labels
- MetaMap (metamap.nlm.nih.gov) to extract MeSH terms

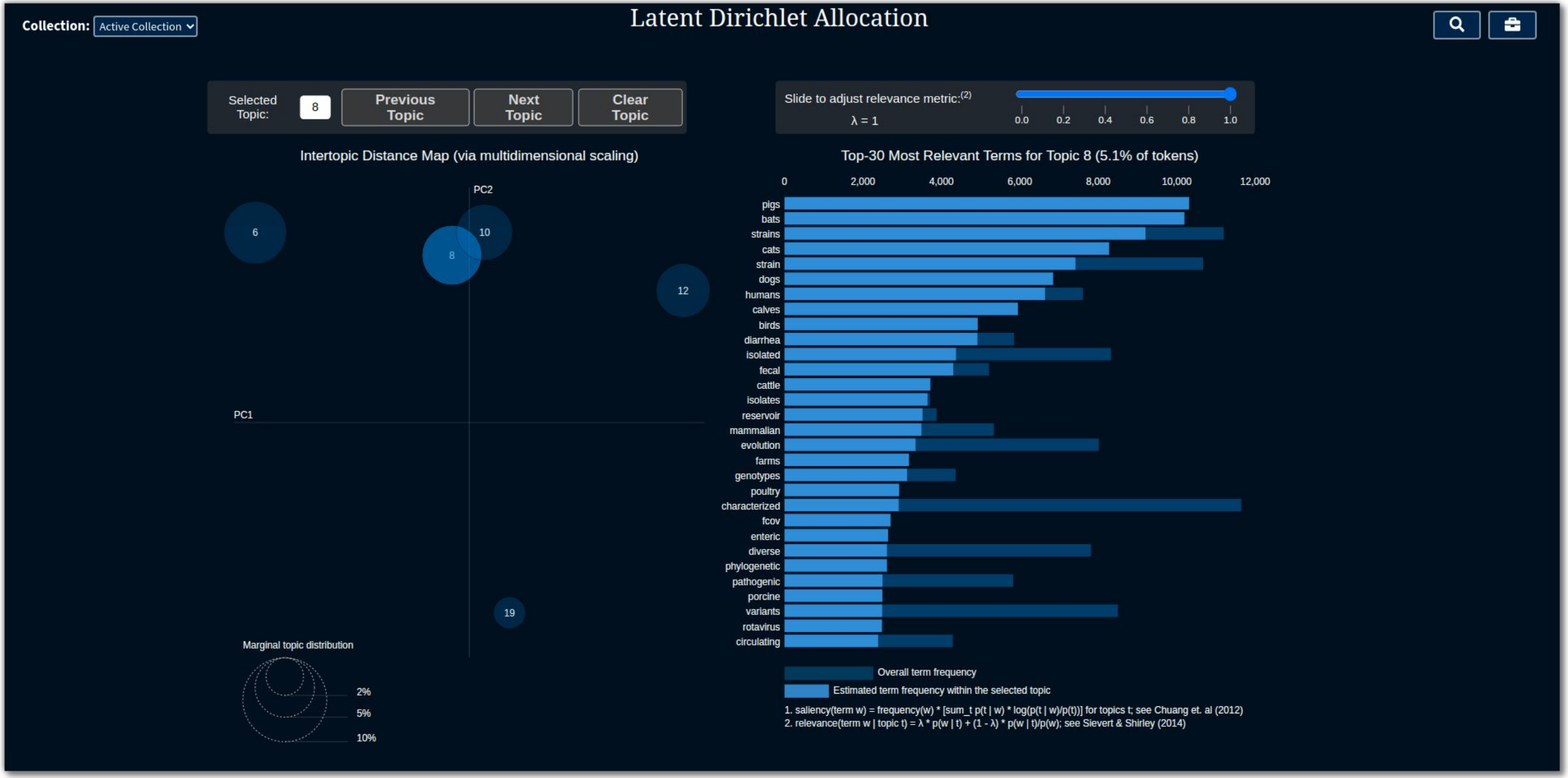

Collection:
Active Collection
Relational Sankey Diagram
Population
Intervention
Outcome
(P) Communicable Diseases
(P) Disease Transmission, Infectious
(P) Lunch
(P) Gastrointestinal Tract
(P) Patients
(P) Meals
(P) Laboratories
(P) Child
(P) Students
(P) Diagnosis
(P) Age Groups
(P) Schools
(P) Hand, Foot and Mouth Disease
(P) Tumor Necrosis Factor Ligand Superfamily Member 13
(P) Vietnam
(I) Quarantine
(I) Government
(I) Hand, Foot and Mouth Disease
(I) SARS Virus
(O) Communicable Diseases
(O) Periodical
(O) Disease Transmission, Infectious
(O) Adult
(O) Overall
(O) Hand, Foot and Mouth Disease


(P) Disease Transmission, Infectious → (I) Quarantine
Related Literature:
Is a 14-day quarantine period optimal for effectively controlling coronavirus disease 2019 (COVID-19)?
Authors: Jiang, Xue; Niu, Yawei; Li, Xiong; Li, Lin; Cai, Wenxiang; Chen, Yucan; Liao, Bo; Wang, Edwin
Journal: N/A, 2020
Background The outbreak of a new coronavirus (SARS-CoV-2) disease (Covid-19) has become pandemic. To be more effectively controlling the disease, it is critical to set up an optimal quarantine period so that about 95% of the cases developing symptoms will be retained for isolation. At the moment, the...
Currently stored in:
Add
to
incubation
Confirm

2. Topic view

Provide a more thematic summary of the current collection

- LDA (Blei et al., 2013) with 20 topics
- Topic model as a mixture over documents, topics and terms
- Shown as a two-dimensional map (pyLDAvis; Sivert and Shirley, 2004)
- We use concept strings as terms


Collection: Active Collection
Latent Dirichlet Allocation
Selected Topic: 8
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:(2)
λ = 1
0.0 0.2 0.4 0.6 0.8 1.0
Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
6
10
8
12
19
Marginal topic distribution
2%
5%
10%
Top-30 Most Relevant Terms for Topic 8 (5.1% of tokens)
0 2,000 4,000 6,000 8,000 10,000 12,000
pigs
bats
strains
cats
strain
dogs
humans
calves
birds
diarrhea
isolated
fecal
cattle
isolates
reservoir
mammalian
evolution
farms
genotypes
poultry
characterized
fcov
enteric
diverse
phylogenetic
pathogenic
porcine
variants
rotavirus
circulating
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

3. Concept cloud view

- 20 most representative concepts per document
- Concept distributions compared with a log-likelihood test (Rayson and Garside, 2000)

Where to go from here

- Implement semantic search using concepts and PICO elements
- Recommend articles which have similar topic distributions
- Allow varying concept granularity, visualise how concepts are embedded in the hierarchy
- Evaluate with a user study to better understand exploration and design

COVID-SEE: http://covid-see.cis.unimelb.edu.au

- Designed a tool to help scientists navigate the literature on COVID-19
- Used NLP methods (ontology concept recognition, PICO element detection, topic modelling) to structure the key biomedical information
- To facilitate exploration and discovery of novel information, COVID-SEE exposes a user to visual overviews of the content of a document collection

References

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent Dirichlet allocation. Journal of Machine Learning Research, 3(Jan):993–1022.

Patrick Cheong-Iao Pang, Karin Verspoor, Shanton Chang and Jon Pearce. 2015. Better Health Information Exploration. Proceedings of the Asia Pacific HCI and UX Design Symposium (APCHIUX '15), pp. 10-15, Melbourne, Australia, 7 Dec 2015, ACM.

Paul Rayson and Roger Garside. 2000. Comparing corpora using frequency profiling. The workshop on comparing corpora, pages 1–6.

Kirk Roberts, Tasmeer Alam, Steven Bedrick, Dina Demner-Fushman, Kyle Lo, Ian Soboroff, Ellen Voorhees, Lucy Lu Wang, and William R Hersh. 2020. TREC-COVID: Rationale and Structure of an Information Retrieval Shared Task for COVID-19. Journal of the American Medical Informatics Association.

Sievert, Carson, and Kenneth Shirley. "LDAvis: A method for visualizing and interpreting topics." Proceedings of the workshop on interactive language learning, visualization, and interfaces. 2014.

Edwin Zhang, Nikhil Gupta, Rodrigo Nogueira, Kyunghyun Cho, and Jimmy Lin. 2020. Rapidly Deploying a Neural Search Engine for the COVID-19 Open Research Dataset: Preliminary Thoughts and Lessons Learned. arXiv:2004.05125.

Ryen W. White and Resa A. Roth (2009). Exploratory Search: Beyond the Query-Response Paradigm, San Rafael, CA: Morgan and Claypool.