
SEMANTIC TECHNOLOGY AND THE COVID-19 PANDEMIC

Michael DeBellis

mdebellsf@gmail.com

<https://www.michaeldebellis.com/>

Biswanath Dutta

bisu@isibang.ac.in

Indian Statistical Institute, Bangalore, India

CHAPTER FOR PENDING BOOK ON SEMANTIC WEB TECHNOLOGY

- Our chapter focused on applications of Semantic Web technology to the Covid-19 pandemic.
- Initial Methodology: Searched Google Scholar for the keywords: “Ontology Covid-19” and “Knowledge graph Covid-19”. We used the first 3 pages of each search.
- Followed references from the initial papers that looked to be especially of value. We also included papers from colleagues that seemed of high value. We define high value to be systems that met both of the following criteria:
 - 1. The primary technology in the research was semantic technology, i.e., ontologies or knowledge graphs. We defined knowledge graphs to include OWL, RDF, and property graphs.
 - 2. The system utilized real world data and/or was utilized by actual users (health professionals or the general public)

ALLEN INSTITUTE CORD-19 DATASET

- Allen Institute (2020).
- All systems (except where explicitly noted) that utilized a corpus of papers on the Covid-19 pandemic and related topics utilized the CORD-19 dataset from the Allen Institute.
- Over 500,000 research articles about COVID-19, SARS-CoV-2, and related coronaviruses.
- Data not indexed with knowledge graphs, utilized JSON. Emphasis was on providing the corpus for Machine Learning technology (focus of Allen Institute) but Semantic researchers utilized it as well.

WE IDENTIFIED EIGHT TYPES OF SYSTEMS

1. Clinical Trials: Tools to search and analyze papers and data on Covid-19 clinical trials
2. Data Analysis: Papers that utilized semantic technology to integrate data from various data sources and to provide additional understanding of the data.
3. Drug Repurposing: The process of taking a medication initially used to treat one disease and utilize it to treat a different disease. Example: Vyvanse is an ADHD medication but many Multiple Sclerosis patients take it to combat symptoms of MS such as chronic fatigue and “brain fog”.
4. Harmonization. One of the biggest problems in medical research are the disparate taxonomies and vocabularies utilized to describe symptoms, diseases, treatments, lab results, etc. Whereas some industries suffer from a lack of standards, healthcare has the opposite problem. Semantic technology can be used to provide integrated models for medical domain models such as SNOMED CT, LOINC, ICD, UMLS, HL7 FYRE, etc.

WE IDENTIFIED EIGHT TYPES OF SYSTEMS

5. Semantic Search: One of the first and most common knowledge graph applications in Industry (e.g., see Pool Party case studies). Semantic search allows researchers to search for papers related to the pandemic using ontologies rather than keywords.
6. Statistics: Statistical models take data and represent it in a graph model. The data is provided by governments and public health organizations on the number of people who have contracted the virus in a specific area over given time intervals. In addition, they record data such as the number of patients who recover, die, have been vaccinated, etc.
7. Surveillance: Surveillance systems aid in monitoring patients. These papers focus on an Internet of Things (IoT) architecture. They take input from mobile devices as well as stationary medical devices. This includes the domain of risk detection: systems that identify people in specific communities (e.g., older people in residential care facilities) who are most at risk from the pandemic.
8. Vocabularies: Ontologies designed to model specific domains such as the Covid-19 virus (e.g., CIDO) in order to be reused by other researchers.

SUMMARY OF FINDINGS

Article Reference	Domain	Usage	Used ML?	Knowledge Representation
COVID-NMA 2021	Clinical Trials	Released		PG
Dutta 2020	Data Analysis	Data		OWL
Wang 2020	Data Analysis	Tested	√	PG
Groza 2020	Data Analysis	Data		OWL
Fries 2021	Data Analysis	Tested	√	OWL
Stebbing 2020	Drug Repurposing	Tested	√	PG
Domingo-Fernandez 20	Drug Repurposing	Tested	√	PG
Wang 2021	Drug Repurposing	Tested	√	PG
Hsieh 2020	Drug Repurposing	Tested	√	PG
Visweswaran 2021	Harmonization	Tested	N/A	PG
Bauer 2021	Harmonization	Tested	N/A	PG
Lin 2021	Harmonization	Reuse	N/A	OWL
Steenwinckel 2020	Semantic Search	Reuse	√	RDF
CORD-19 on FYRE 201	Semantic Search	Reuse	√	RDF
Wise 2020	Semantic Search	Released	√	PG
HealthECCO 2021	Semantic Search	Released	√	PG
Cernile 2021	Semantic Search	Released	√	PG
Michel 2020	Semantic Search	Released		RDF
Dong 2020	Statistics	Released	N/A	RDF
Almukhtar 2021	Statistics	Released	N/A	RDF
Nonita 2021	Surveillance	Data	√	PG
Bettencourt-Silva 2021	Surveillance	Data	√	RDF
de Lusignan 2020	Surveillance	Tested		OWL
He 2020	Vocabulary	Reuse	N/A	OWL

Key

Usage:

Reuse: Vocabularies reused by at least one other system in the list

Data: Used real data

Tested: Used in test settings with users

Released: Released for regular usage on Internet or an Intranet

Knowledge Representation:

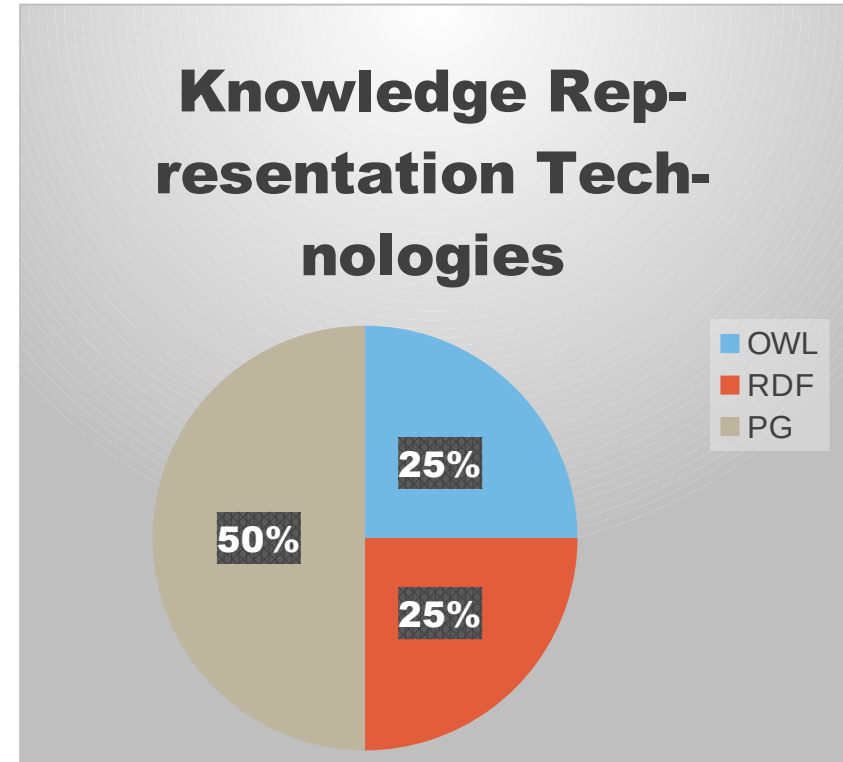
RDF: Resource Description Framework

OWL: Web Ontology Language

PG: Property Graphs (no standard so either custom technology or vendor products such as Neo4J)

SUMMARY OF FINDINGS

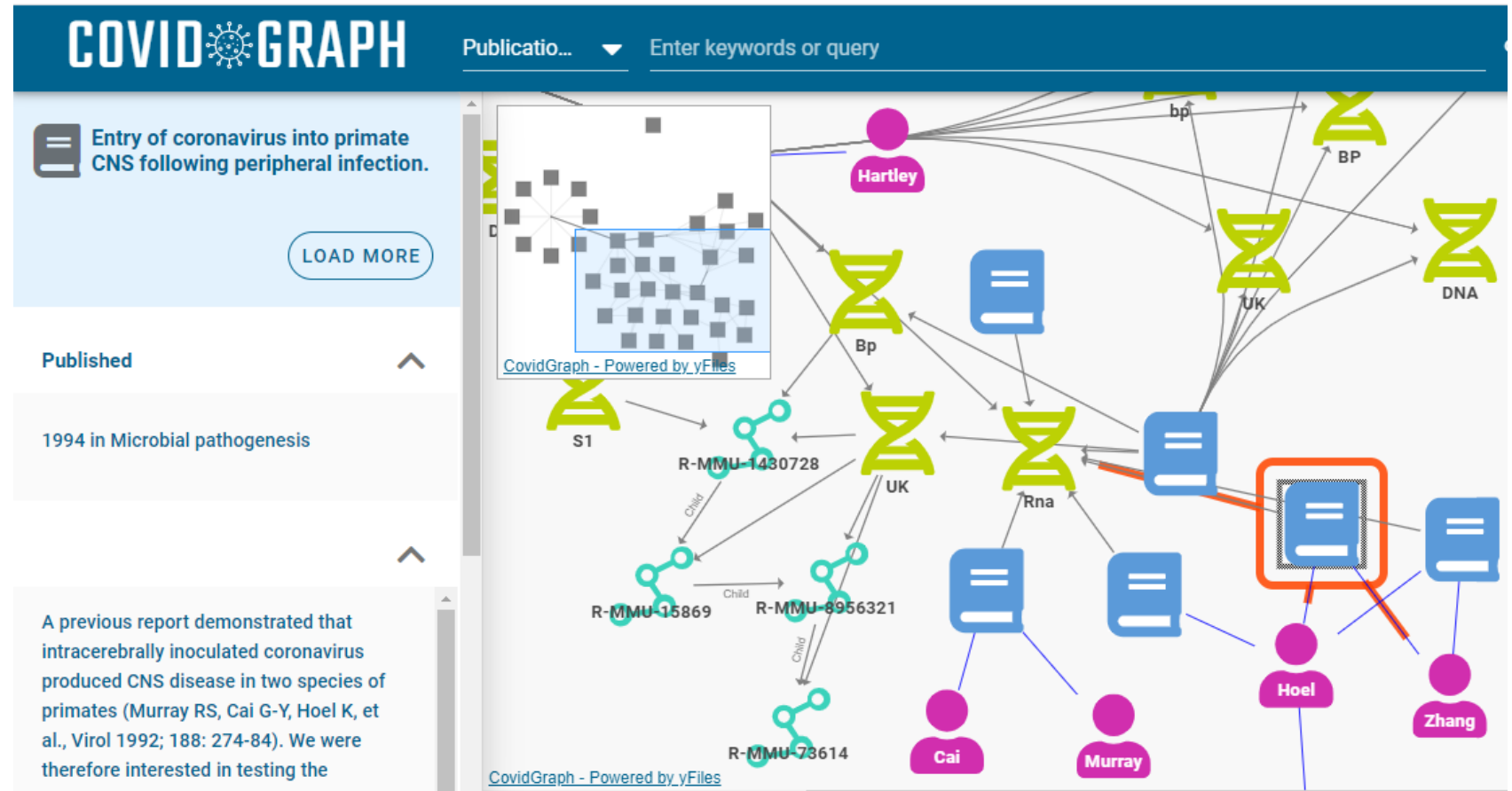
- Half of the systems utilized non-standard knowledge graph models (Property Graphs) and 75% did not use OWL
- 72% of the systems where Machine Learning (ML) was relevant (excluding Statistics, Vocabularies, and Harmonization) used ML.



SEMANTIC SEARCH

COVID*GRAPH

- HealthECCO (2021)
- <https://healthecco.org/covidgraph/>
- A group of volunteers added semantics to CORD-19 as well as other data sources.
- Utilizes Neo4J graph database and the Cypher query language (Property Graphs).
- Designed primarily for healthcare researchers



SEMANTIC SEARCH

Semantic Search With Amazon Web Services (Wise 2020).

- <https://www.cord19.aws/>
- The knowledge graph contains 5 top level classes: Paper, Author, Institution, Concept, and Topic. Utilizes NLP to recognize medical concepts and relations.
- Utilized SciBERT for semantic embedding
- Intuitive GUI that can be used by consumers or professionals

The screenshot displays the AWS CORD-19 Search web application. At the top, the AWS logo and 'CORD-19 Search' are visible. A search bar contains the query 'Are vaccine booster shots a good idea?'. Below the search bar, filters for 'Publication date' (Anytime, Since 2020, Since 2019, Custom) and 'Topic' (Public-health-policies, Healthcare-industry, Clinical-treatment, Epidemiology, Lab-trials-human, Pulmonary-infections, Influenza, Vaccines-immunology, Genomics, Virology) are shown. The results section indicates '1-10 of about 115,000' items, sorted by 'Best match'. The top suggested answer is 'A Guide to COVID-19: a global pandemic caused by the novel coronavirus SARS-CoV-2' by Atzrodt, Cassandra L. et. al. from FEBS J. (2020-05-23) at Vanderbilt University. The answer text discusses vaccination methods and booster shots. Below this, 'More suggested answers (2)' are listed, including 'Pros and Cons of Adenovirus-Based SARS-CoV-2 Vaccines' by Kremer, Eric J. from Mol Ther. (2020-10-09). A 'Learn more' section at the bottom provides links to the 'CORD-19 Search FAQ' and 'CORD-19 challenge'.

aws CORD-19 Search

Are vaccine booster shots a good idea?

Filter search results by

1-10 of about 115,000

Sort by Best match

Publication date

- ☒ Anytime
- ☐ Since 2020
- ☐ Since 2019
- ☐ Custom

Topic

Select up to 5 topics to filter your results

- ☒ Public-health-policies (115465)
- ☐ Healthcare-industry (75693)
- ☐ Clinical-treatment (37223)
- ☐ Epidemiology (32047)
- ☐ Lab-trials-human (12760)
- ☐ Pulmonary-infections (5274)
- ☐ Influenza (4964)
- ☐ Vaccines-immunology (4460)
- ☐ Genomics (4022)
- ☐ Virology (1927)

Top suggested answer

A Guide to COVID-19: a global pandemic caused by the novel coronavirus SARS-CoV-2

Atzrodt, Cassandra L. et. al.
FEBS J. 2020-05-23.
Vanderbilt University

Another method of **vaccination** is to use a dead or inactivated virus, like in the polio **vaccine**. One upside to these **vaccines** is that there is no concern about the virus regaining pathogenicity and causing disease. However, **booster shots are needed for continued protection because these vaccines trigger smaller immune responses** [147]. Rather than include the entire virus, subunit **vaccines** work by including only parts of the virus' structure, or subunits [148]. For example, a subunit **vaccine** for a coronavirus strain may only include the S protein or parts of the S protein (Fig).

Public-health-policies | Epidemiology

More suggested answers (2)

Pros and Cons of Adenovirus-Based SARS-CoV-2 Vaccines

Kremer, Eric J.
Mol Ther. 2020-10-09.
Cited by 5

...But the **idea** of 4 doses/person for billions of people during the first round of **vaccinations**-perhaps followed by **booster shots** every year-is difficult to imagine. While at least one phase 1 trial suggested that a 1-shot Ad-based **vaccine** could be sufficient, it is my bet...

Learn more

[CORD-19 Search FAQ](#)

[CORD-19 challenge](#)



THANK YOU

- Bibliography and Extra Slides after this slide

BIBLIOGRAPH

- Allen Institute for AI. (2020) COVID-19 Open Research Dataset Challenge. May 29, 2020.
<https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge>
- Bauer, DC, (2021) with Metke-Jimenez, A, Maurer-Stroh, S, et al. Interoperable medical data: The missing link for understanding COVID-19. Transbound Emerg Dis, 2021; 68: 1753– 1760.
<https://doi.org/10.1111/tbed.13892>
- CORD-19 Github site (2020). <https://github.com/fhircat/CORD-19-on-FHIR/wiki> Apr 8, 2020.
- Dutta (2020). Dutta, B. and DeBellis, M. (2020). CODO: an ontology for collection and analysis of COVID-19 data. Accepted for publication in the Proc. of 12th Int. Conf. on Knowledge Engineering and Ontology Development (KEOD), 2-4 November 2020.
- Domingo-Fernandez, D. (2020) with Baksi, S.; Schultz, B.; Gadiya, Y.; Karki, R.; Raschka, T.; Ebeling, C.; Hofmann-Apitius, M. COVID-19 Knowledge Graph: A computable, multi-modal, cause-and-effect knowledge model of COVID-19 pathophysiology. bioRxiv, 2020.
- Dong, Ensheng (2020); Hongru Du; Lauren Gardner. An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases, ISSN: 1473-3099, Vol: 20, Issue: 5, Page: 533-534. 2020.

BIBLIOGRAPH

- Gangemi Aldo (2017) with Valentina Presutti, Diego Reforgiato Recupero, Andrea Giovanni Nuzzolese, Francesco Draicchio, and Misael Mongiov`i. Semantic web machine reading with fred. Semantic Web, 8(6):873–893, 2017.
- Groza, A. (2020). Detecting fake news for the new coronavirus by reasoning on the Covid-19 ontology. ArXiv, abs/2004.12330
- HealthECCO (2021). <https://healthecco.org/about/> Retrieved August 17, 2021.
- HL7-FHIR Release 4 (2019). FHIR Release 4 (Technical Correction #1) (v4.0.1) generated on Fri, Nov 1, 2019. <https://www.hl7.org/fhir/>
- Lin, Asiyah Yu. Et. al. (2021) A community effort for COVID-19 Ontology Harmonization. The 12th International Conference on Biomedical Ontologies, September 2021.
- Stebbing, Justin et al. (2020). “COVID-19: combining antiviral and anti-inflammatory treatments.” The Lancet. Infectious diseases, vol. 20,4 (2020): 400-402. doi:10.1016/S1473-3099(20)30132-8
- Steenwinckel, B. (2020) with Vandewiele, G.; Rausch, I.; Heyvaert, P.; Taelman, R.; Colpaert, P.; Simoens, P.; Dimou, A.; de Turck, F.; Ongenaes, F. Facilitating the analysis of covid-19 literature through a knowledge graph. In International Semantic Web Conference, Springer: Cham, Switzerland, 2020; pp. 344–357.

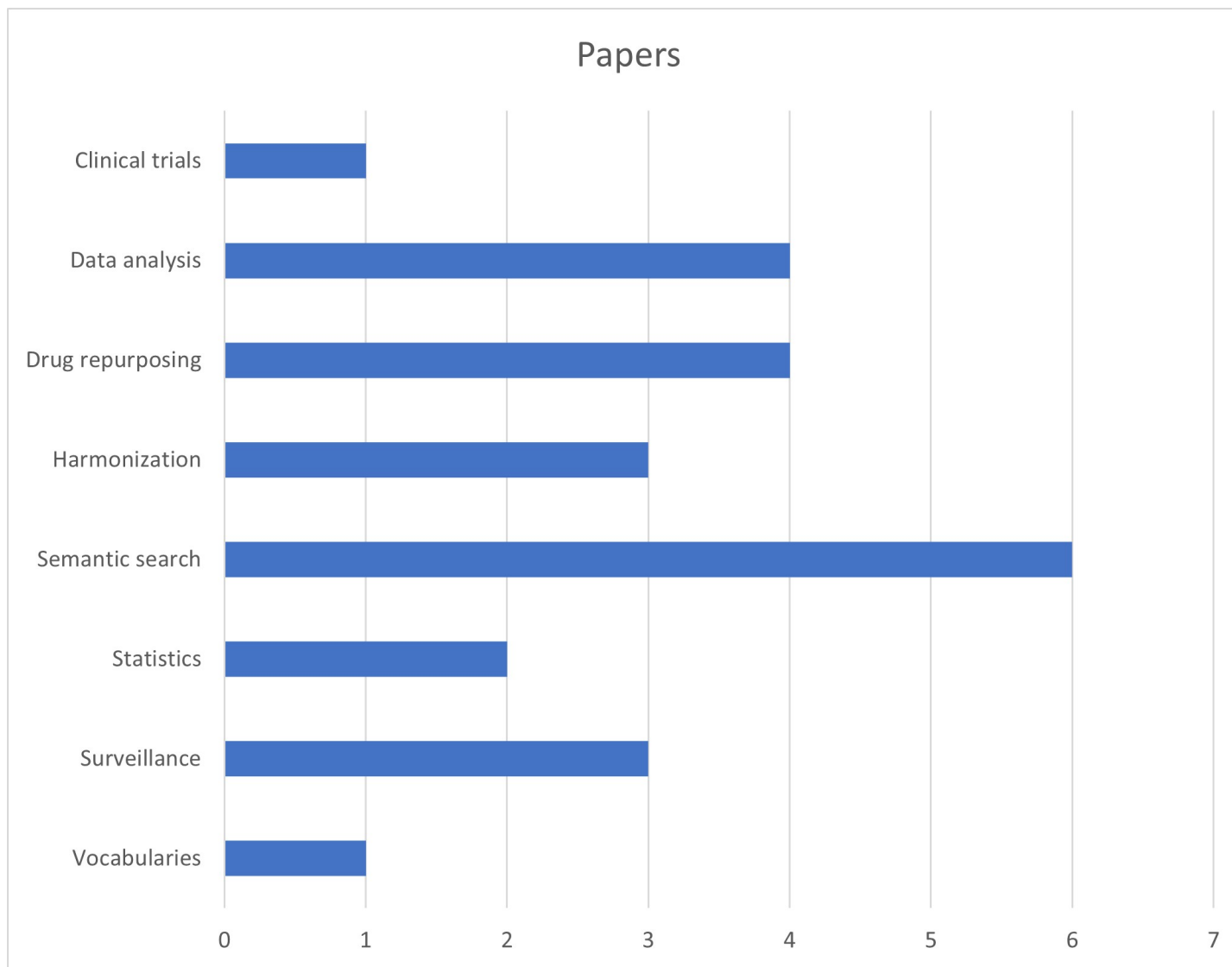
BIBLIOGRAPH

- Visweswaran, S. et al. (2021). Development of a Coronavirus Disease 2019 (COVID-19) Application Ontology for the Accrual to Clinical Trials (ACT) network. JAMIA, Open 4 (2021).
- Wang, Qingyun et al. (2021). “COVID-19 Literature Knowledge Graph Construction and Drug Repurposing Report Generation.” NAACL.
- Wang, Yang (2020) with Fengwei Zhang, Hong Yu, Xianwei Ye and Y. He. Ontology-based annotation and analysis of COVID-19 phenotypes. arXiv: Other Quantitative Biology,
- Wang, Qingyun et al. (2021). “COVID-19 Literature Knowledge Graph Construction and Drug Repurposing Report Generation.” NAACL.
- Wise, C. (2020); Ioannidis, V.N.; Calvo, M.R.; Song, X.; Price, G.; Kulkarni, N.; Brand, R.; Bhatia, P.; Karypis, G. COVID-19 knowledge graph: Accelerating information retrieval and discovery for scientific literature. arXiv 2020, preprint. arXiv:2007.12731.



EXTRA SLIDES

DISTRIBUTION OF PAPERS



DRUG REPURPOSING

- BenevolentAI Stebbing (2020): A proprietary knowledge graph to enable identification of drugs that could be repurposed against Covid-19. The BenevolentAI system is a combination of machine learning algorithms and knowledge graph data. The system identified potential candidates based on their anti-viral and anti-inflammatory properties.
 - Resulted in a clinical trial of baricitinib plus remdesivir that showed that the combination was superior to remdesivir alone in reducing recovery time.
- Domingo-Fernandez (2020): Cause and effect knowledge graph. Ontology includes 10 classes (e.g., proteins, genes, and biological processes) and over 10K relations (e.g., increases, decreases, association).
 - Used this knowledge graph to identify over 300 candidate drugs for repurposing which (circa Fall 2021) are under various states of evaluation including several clinical trials.

DATA ANALYSIS

- CODO: Biswanath's presentation
- Covid-19 Phenotypes Wang (2020) : collected data about COVID-19 phenotypes (e.g., loss of smell, nausea, abdominal pain) and utilized the Human Phenotype Ontology (HPO) with Protégé to classify and analyze them.
 - Analysis yielded new insights into the geographic distribution of various phenotypes. E.g., abdominal and nervous pain phenotypes were more common in European and USA patients than in Asian patients.
 - In addition, 23 comorbidities (e.g., diabetes and kidney failure) were identified which significantly worsened the probable outcomes for patients with those comorbidities.

STATISTICS

Johns Hopkins

- Dong (2020)
- <https://coronavirus.jhu.edu/map.html>
- One of the most widely used tools
- Utilized by general public as well as healthcare professionals

