

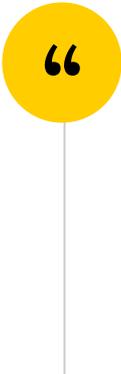
Investigating COVID-19 Analytics and Research with Clinical Knowledge Organization Systems



Julaine Clunis, PhD | jclunis@uwm.edu

“A concept is a brick. It can be used to build a courthouse of reason. Or it can be thrown through the window.”

Gilles Deleuze, *A Thousand Plateaus: Capitalism and Schizophrenia*



“

1

The playing field

Let's start with where we are

45,000,000+

The number of concepts related to medicine, biomedical molecules, genes, organisms, patients, conditions, populations, healthcare actions, technical methods, and social concepts (ISO, 2018).





What is the background?

- Data contains critical information – diagnostics, treatments, prevention methods, unknown knowledge
- Data sharing critical for leveraging research, testing drug effectiveness, therapeutic strategies, policies for control, intervention, and eradication
- Secondary use of data hidden in medical documentation and clinical highlights the need for integrated clinical coding schemes to create links between content.



What is **the issue?**

- Maximize the reuse of data
- Extract all potential points of data contained in those texts
- Limited data description impedes advancement in enabling semantic interoperability, health information exchange, analytics, and research
- Discoverability challenging due to variations in the amount of concept information represented in medical terminology or lack of applied standards describing the data.



Knowledge Organization Systems

Structured lists of terms and their associated definitions intended to describe a domain categorically



Clinical coding schemes

- Critical for defining and structuring concepts and terms in healthcare
- Support data sharing, link clinical evidence with administrative decisions, support evidence-based practice, enable population-based interventions, use electronic health records and decision support systems, and advance medical research
- Critical for creating insight and bridging the contextual differences across systems
- Needed for meaningful and accurate utilization of the information exchanged at the syntactic level of interoperability and further act as a method for information enrichment and facilitate better information analysis processes



Research Questions

Mapping

How can the data analytics and extract transform load (ETL) tool KNIME* support the task of clinical coding scheme mapping?

Annotation

How can the output produced from the mapping be used to annotate clinical trial documents?



Why mapping?

- Improve efficiency and promote better sharing, combining, and linking data sets from different sources and ensuring that the meaning of information coming from disparate systems is the same
- Allows comparisons between research studies which would otherwise be impossible because of confusion caused by lack of alignment (Gliklich et al., 2014).
- Support browsing and searching of unstructured data such as clinical trials through semantic annotation
- Integrates data from different resources into a single context to enhance understanding of complicated biomedical systems
- Description of concepts for new diseases and alignment of those terms with preexisting terminologies is a current and pressing issue



Mapping methods and tools?

- Lexical and morphological text matching algorithms (Barrows, Cimino, & Clayton, 1994)
- String matching algorithms (Saitwal et al, 2012)
- Natural Language Processing (Zhou et al, 2012)
- Association Rules Mining (Dias, Alves, Felipe, 2014)
- Structural and Disambiguation techniques (Allones, Martinexz, Taboada, 2014)
- Feature Engineering and Deep Learning (Kolyvakis et al, 2018)
- Tools – MeTMapS, BioPortal, YAM++



KNIME Mapping and Annotation workflow

Input Data
575 Clinical Trials,
Clinical Coding Schemes:
COVOC, COVID-19, CIDO, LOINC

1

Matching Segments

3

Performance Measures
and Benchmarking

5

Data Cleaning and
Transformation

2

Annotation Segment

4

Assess Knowledge
Contributions

6

KNIME Explorer

Overview of the available workflows.

Workflow Editor

Canvas for editing the currently active workflow.

KNIME Hub

Access to KNIME Hub to drag and drop nodes, components, workflows or extensions.

Description

Description of the active workflow or a selected node.

Workflow Coach

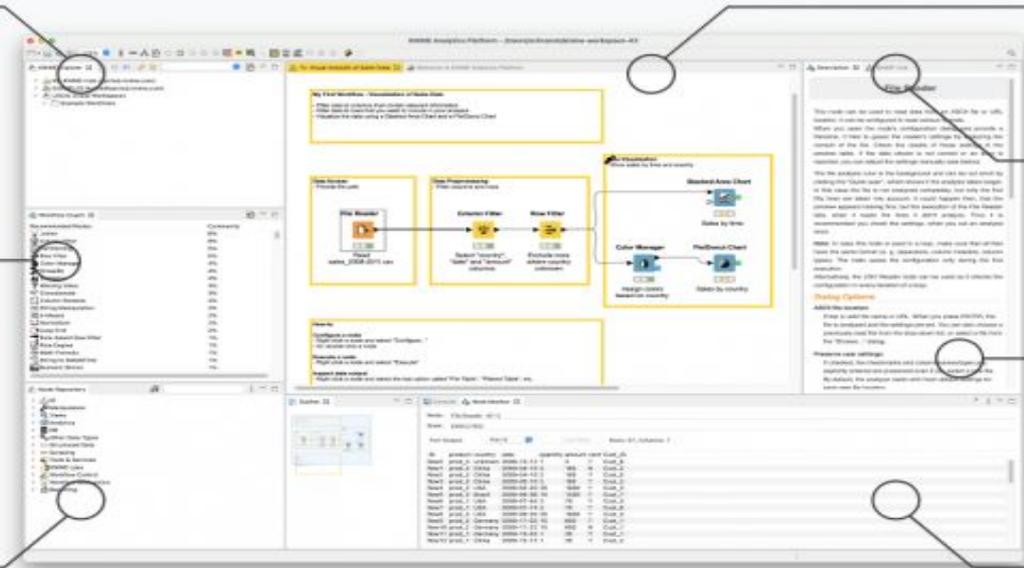
Lists node recommendations based on the workflows built by the wide community of KNIME users.

Node Repository

All available nodes in KNIME Analytics Platform to build your workflows.

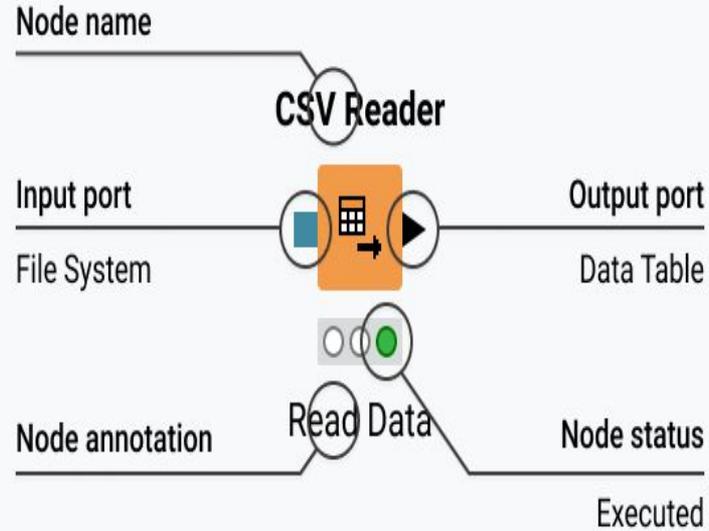
Node Monitor

Shows the current flow variable values or a preview of the output data of the selected node.



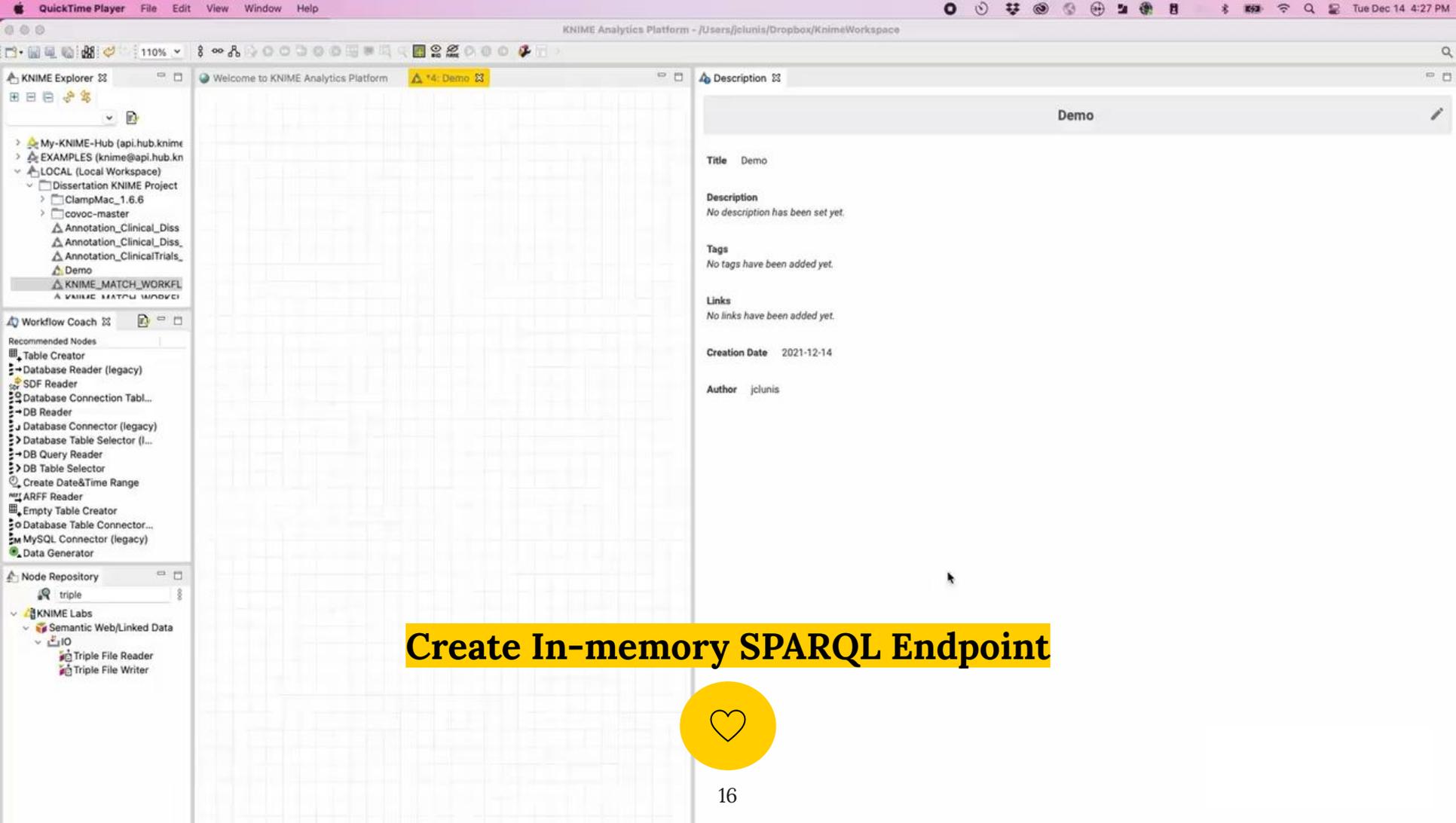
The KNIME Workbench





KNIME Nodes





Create In-memory SPARQL Endpoint





Connected workflows

- Lexical Series Matcher
 - Uses string distance algorithms to match concept labels
- Document Similarity Matcher
 - Uses cosine similarity to compare and match concept definitions
- Semantic Similarity Matcher
 - Uses a vector space model embedding of concepts
- Clinical Trial Annotation
 - Uses named entity recognition to identify dictionary terms in unstructured data

2

The outcome

What kind of results were obtained?

Output from Lexical Matchers

S COVOC_URI	S TERM	S CIDO_URI	S TERM (right)
http://purl.obolibrary.org/obo/HP_0031417	rhinorrhea	http://purl.obolibrary.org/obo/HP_0031417	rhinorrhea
http://purl.obolibrary.org/obo/NCBITaxon_9397	chiroptera	http://purl.obolibrary.org/obo/NCBITaxon_9397	chiroptera
http://purl.obolibrary.org/obo/PR_Q98YF1	angiotensin-converting enzyme 2 (human)	http://purl.obolibrary.org/obo/PR_Q98YF1	angiotensin-converting enzyme 2 (human)
http://purl.obolibrary.org/obo/NCBITaxon_11118	coronaviridae	http://purl.obolibrary.org/obo/NCBITaxon_11118	coronaviridae
http://purl.obolibrary.org/obo/NCBITaxon_40674	mammalia	http://purl.obolibrary.org/obo/NCBITaxon_40674	mammalia
http://purl.obolibrary.org/obo/PATO_0001422	dead	http://purl.obolibrary.org/obo/PATO_0001422	dead
http://purl.obolibrary.org/obo/PR_P0DTC7	orf7a protein (sars-cov-2)	http://purl.obolibrary.org/obo/PR_P0DTC7	orf7a protein (sars-cov-2)
http://purl.obolibrary.org/obo/HP_0003326	myalgia	http://purl.obolibrary.org/obo/HP_0003326	myalgia
http://purl.obolibrary.org/obo/NCBITaxon_2697...	severe acute respiratory syndrome coro...	http://purl.obolibrary.org/obo/NCBITaxon_2697049	sars-cov-2

Terms with Same URIs

S COVOC_URI	S TERM	J	S nearest neighbor - COVID19_URI	D distance	S COVID19_URI	S TERM (right)
http://purl.obolibrary.org/obo/NCIT_C15783	clinical data	0	http://purl.obolibrary.org/obo/OGMS_0000123	0.172	http://purl.obolibrary.org/obo/OGMS_0000123	clinical data item
http://purl.obolibrary.org/obo/NCIT_C155831	nasopharyngeal swab specimen	0	http://purl.obolibrary.org/obo/NCIT_C155835	0.098	http://purl.obolibrary.org/obo/NCIT_C155835	oropharyngeal swab specimen
http://purl.obolibrary.org/obo/NCIT_C82562	clinical significance	0	https://bio.scai.fraunhofer.de/ontology/COVID_0000081	0.172	https://bio.scai.fraunhofer.de/ontology/COVID_0000081	clinical sign
http://purl.obolibrary.org/obo/NCIT_C82562	clinical significance	0	https://bio.scai.fraunhofer.de/ontology/COVID_0000081	0.172	https://bio.scai.fraunhofer.de/ontology/COVID_0000081	clinical signs
http://purl.obolibrary.org/obo/NCIT_C15843	preventive intervention	0	http://purl.obolibrary.org/obo/ERO_0000347	0.231	http://purl.obolibrary.org/obo/ERO_0000347	intervention
http://purl.obolibrary.org/obo/HP_0033006	diffuse alveolar damage	0	https://bio.scai.fraunhofer.de/ontology/COVID_0000014	0.222	https://bio.scai.fraunhofer.de/ontology/COVID_0000014	alveolar damage
http://purl.obolibrary.org/obo/COVOC_0050018	epidemics	0	http://purl.obolibrary.org/obo/APOLLO_SV_00000298	0.067	http://purl.obolibrary.org/obo/APOLLO_SV_00000298	epidemic
http://purl.obolibrary.org/obo/CHEBI_17334	penicillin	0	http://purl.obolibrary.org/obo/CHEBI_18208	0.217	http://purl.obolibrary.org/obo/CHEBI_18208	benzylpenicillin
http://purl.obolibrary.org/obo/GO_0006810	transport	0	http://purl.obolibrary.org/obo/GO_0050658	0.2	http://purl.obolibrary.org/obo/GO_0050658	rna transport
http://purl.obolibrary.org/obo/MONDO_0006502	acute respiratory distress syndr...	0	http://purl.obolibrary.org/obo/DOID_11394	0.156	http://purl.obolibrary.org/obo/DOID_11394	acute respiratory distress syndr...
http://purl.obolibrary.org/obo/GO_0006412	translation	0	http://purl.obolibrary.org/obo/GO_0019081	0.2	http://purl.obolibrary.org/obo/GO_0019081	viral translation
http://purl.obolibrary.org/obo/NCIT_C53287	health care professional	0	http://purl.obolibrary.org/obo/OGMS_0000096	0.212	http://purl.obolibrary.org/obo/OGMS_0000096	health care process
http://purl.obolibrary.org/obo/COVOC_0030017	interferon alfacon-1	0	http://purl.obolibrary.org/obo/DRON_00017633	0.212	http://purl.obolibrary.org/obo/DRON_00017633	interferon alfa-2b
http://www.ebi.ac.uk/efo/EFO_0009727	shortness of breath	0	http://purl.obolibrary.org/obo/NCIT_C126929	0.122	http://purl.obolibrary.org/obo/NCIT_C126929	some shortness of breath
http://purl.obolibrary.org/obo/NCIT_C123937	distribution	0	https://bio.scai.fraunhofer.de/ontology/COVID_0000012	0.241	https://bio.scai.fraunhofer.de/ontology/COVID_0000012	posterior distribution
http://purl.obolibrary.org/obo/GO_0043657	cytokine	0	http://purl.obolibrary.org/obo/GO_0000910	0.176	http://purl.obolibrary.org/obo/GO_0000910	cytokinesis
http://purl.obolibrary.org/obo/MONDO_0005550	infectious disease	0	http://purl.obolibrary.org/obo/DOID_934	0.158	http://purl.obolibrary.org/obo/DOID_934	viral infectious disease
http://purl.obolibrary.org/obo/GO_0043657	host cell	0	http://purl.obolibrary.org/obo/CL_0000084	0.231	http://purl.obolibrary.org/obo/CL_0000084	t cell
http://purl.obolibrary.org/obo/OBI_0000552	reverse transcribed polymerase...	0	http://purl.obolibrary.org/obo/OBI_0001170	0.125	http://purl.obolibrary.org/obo/OBI_0001170	reverse transcription polymerase...
http://purl.obolibrary.org/obo/GO_0006955	immune response	0	http://purl.obolibrary.org/obo/GO_0006959	0.222	http://purl.obolibrary.org/obo/GO_0006959	humoral immune response
http://purl.obolibrary.org/obo/COVOC_0050043	masks	0	http://purl.obolibrary.org/obo/NCIT_C86570	0.143	http://purl.obolibrary.org/obo/NCIT_C86570	mask
http://purl.obolibrary.org/obo/MONDO_0005719	coronavirinae infectious disease	0	http://purl.obolibrary.org/obo/DOID_0080599	0.077	http://purl.obolibrary.org/obo/DOID_0080599	coronavirus infectious disease
http://purl.obolibrary.org/obo/GO_0009058	biosynthetic process	0	http://purl.obolibrary.org/obo/GO_0006754	0.095	http://purl.obolibrary.org/obo/GO_0006754	atp biosynthetic process
http://purl.obolibrary.org/obo/MONDO_0100142	severe covid-19 infection	0	http://purl.obolibrary.org/obo/HP_0032169	0.231	http://purl.obolibrary.org/obo/HP_0032169	severe infection
http://purl.obolibrary.org/obo/GO_0016310	phosphorylation	0	http://purl.obolibrary.org/obo/GO_0016572	0.226	http://purl.obolibrary.org/obo/GO_0016572	histone phosphorylation
http://purl.obolibrary.org/obo/GO_0030154	cell differentiation	0	http://purl.obolibrary.org/obo/GO_0042093	0.163	http://purl.obolibrary.org/obo/GO_0042093	t-helper cell differentiation

Terms with string distance < 0.25



Lexical Matcher results

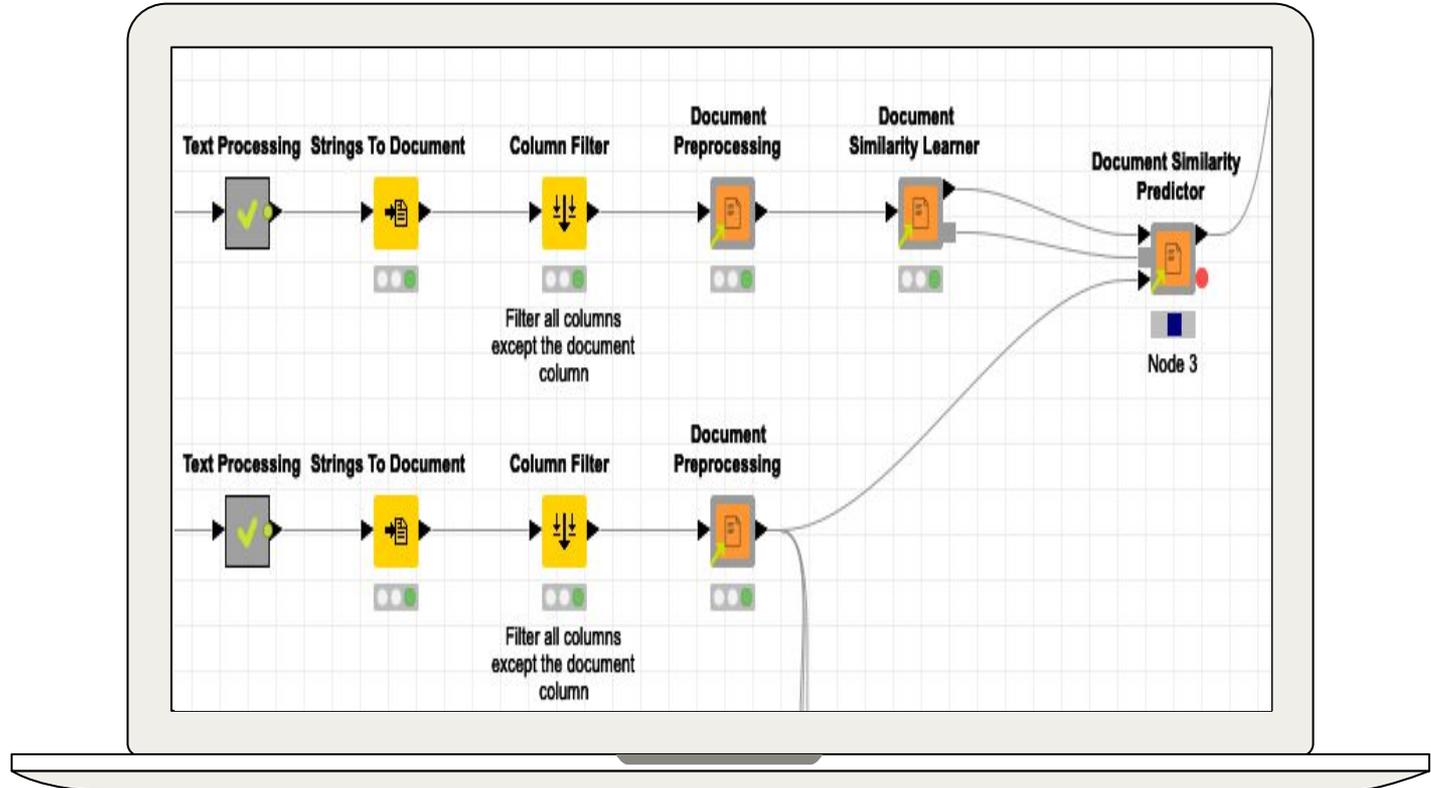
Match Type	COVOC/CIDO	COVOC/COVID19	COVOC/LOINC	CIDO/COVID-19	CIDO/LOINC	COVID-19/LOINC
skos:exactMatch	98	94	0	586	0	0
skos:closeMatch	48	53	121	28	424	245
Matched (Review)	346	38	23	289	323	153
Total Mappings <i>[String distance < 0.25]</i>	492	185	144	903	747	398
Gold Standard				666	871	489

Table 1. Mappings found with Lexical Algorithms



Document Similarity Workflow

Find similar concept definitions.



Using Cosine
Similarity –
Examples of
related
definitions in the
same subclass



```
<owl:Class rdf:about="http://purl.obolibrary.org/obo/GO_0044423">
  <rdfs:subClassOf rdf:resource="http://purl.obolibrary.org/obo/GO_0005575"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty
rdf:resource="https://bio.scai.fraunhofer.de/ontology/COVID_0000411"/>
      <owl:someValuesFrom rdf:resource="http://purl.obolibrary.org/obo/DOID_0080600"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <obo:IAO_0000115>Any constituent part of a virion, a complete fully infectious
extracellular virus particle.</obo:IAO_0000115>
```

```
<owl:Class rdf:about="http://purl.obolibrary.org/obo/GO_0019012">
  <rdfs:subClassOf rdf:resource="http://purl.obolibrary.org/obo/GO_0005575"/>
  <obo:IAO_0000115>The complete fully infectious extracellular virus particle.</obo:IAO_0000115>
```

Results from
document
similarity
matching of
definitions



nearest neighbor - Document: "complete infectious extracellular virus particle"

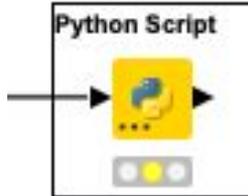
similarity: 0.9128709291752769

Document: "any constituent part of a virion, a complete fully infectious extracellular virus particle."

□



Semantic Similarity Matcher



```
input to output
_table_1 = input_table_1.copy()

pandas as pd
spacy
spacy.load('en_core_sci_lg')

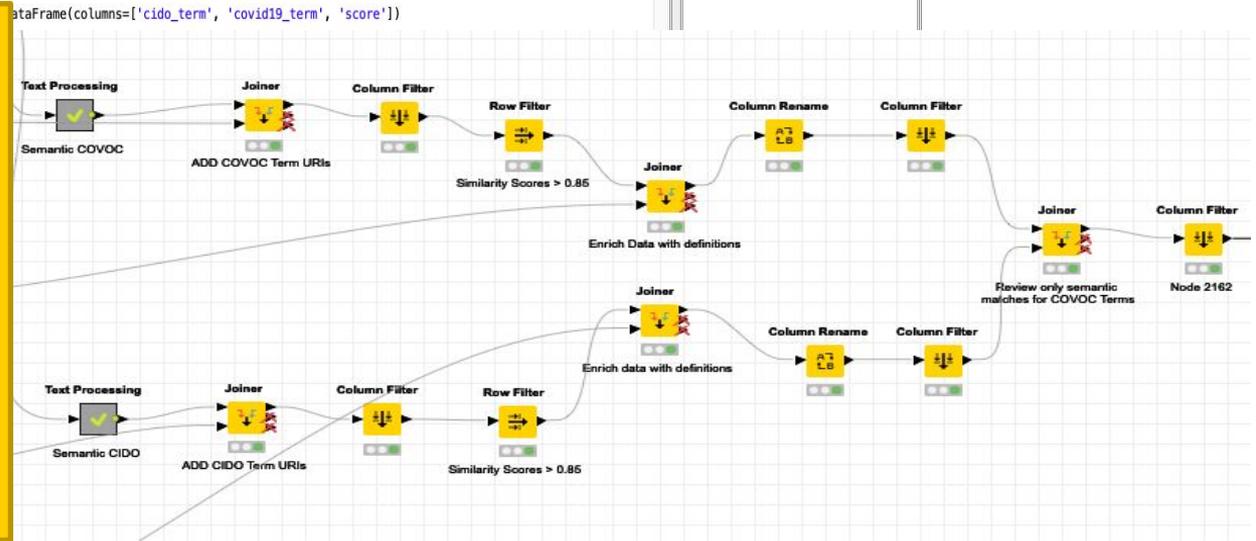
erms = input_table_1['CIDO_Term']
9_Terms = input_table_1['COVID19_Term']
COVID19_Terms
```

Name	Type	Value
knime...	module	
os	module	
sys	module	
INT_SE...	int	-21474...
LONG_S...	int	-92233...
flow_v...	Ordere...	Ordere...
input_...	DataFrame	...
python...	str	4
workspace	Python...	<pytho...

Use node for custom python script with SciSpacy library.

Pass output to semantic similarity matcher.

Segment shown





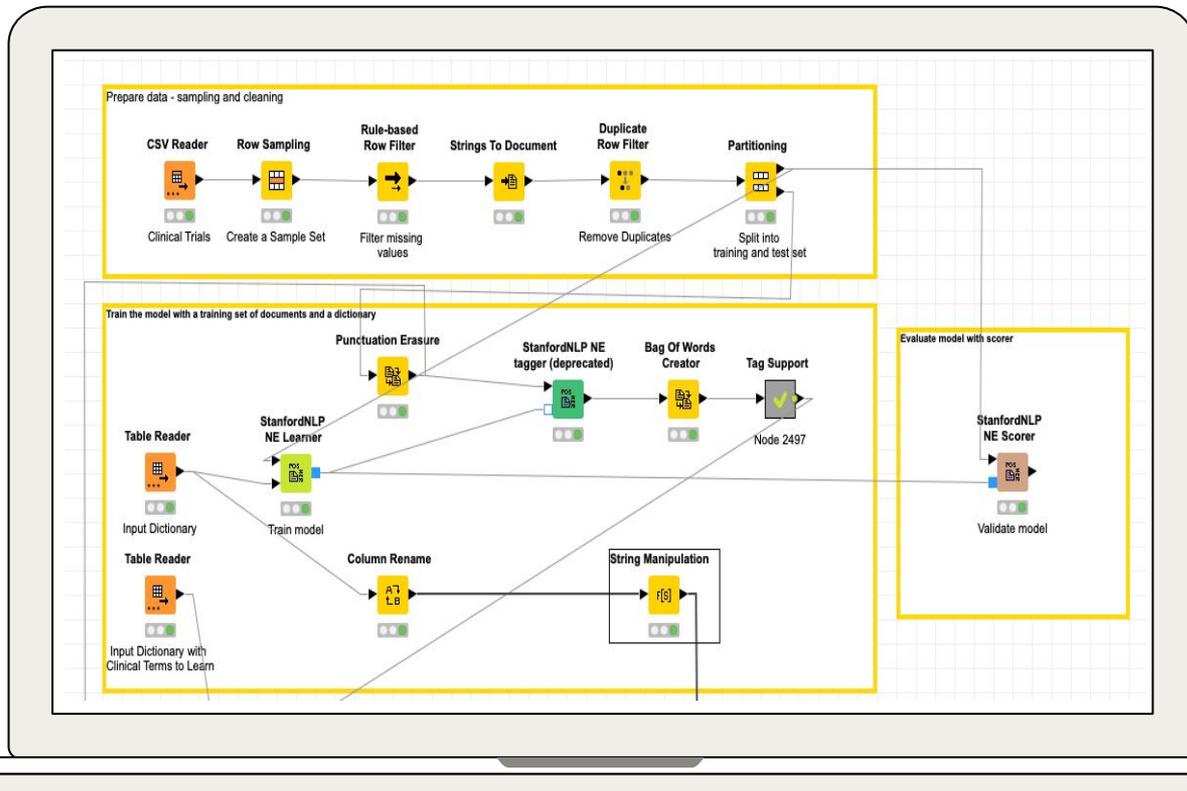
amino acid	A carboxylic acid containing one or more amino groups.	http://purl.obolibrary.org/obo/CHEBI_33709	amino acid	Any amino acid whose side chain is capable of forming one or more hydrogen bonds.	http://purl.obolibrary.org/obo/CHEBI_26167	1
amino acid	A carboxylic acid containing one or more amino groups.	http://purl.obolibrary.org/obo/CHEBI_33709	amino acid	Any monocarboxylic acid which also contains a separate (alcoholic or phenolic) hydroxy substituent.	http://purl.obolibrary.org/obo/CHEBI_35868	1
amino acid	A carboxylic acid containing one or more amino groups.	http://purl.obolibrary.org/obo/CHEBI_33709	amino acid	An oxoacid containing a single carboxy group.	http://purl.obolibrary.org/obo/CHEBI_25384	1
amino acid	A carboxylic acid containing one or more amino groups.	http://purl.obolibrary.org/obo/CHEBI_33709	amino acid	Any aromatic carboxylic acid that consists of benzene in which at least a single hydrogen has been substituted by a carboxy group.	http://purl.obolibrary.org/obo/CHEBI_22723	1
chest pain	An unpleasant sensation characterized by physical discomfort (such as pricking, throbbing, or aching) localized to the chest.	http://purl.obolibrary.org/obo/HP_0100749	Chest pain	An unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage.	http://purl.obolibrary.org/obo/HP_0012531	1
abdominal pain	An unpleasant sensation characterized by physical discomfort (such as pricking, throbbing, or aching) and perceived to originate in the abdomen.	http://purl.obolibrary.org/obo/HP_0002027	Abdominal pain	An unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage.	http://purl.obolibrary.org/obo/HP_0012531	1
anxiety disorder	A category of psychiatric disorders which are characterized by anxious feelings or fear often accompanied by physical symptoms associated with anxiety.	http://purl.obolibrary.org/obo/MONDO_0005618	anxiety disorder	An anxiety disorder that is characterized by unexpected and repeated episodes of intense fear accompanied by physical symptoms	http://purl.obolibrary.org/obo/DOID_594	1





Clinical Trial Annotation Workflow

Identify and enrich entities of interest with unique ids and URIs from clinical coding schemes.





Clinical Trial Annotation Output

- Show all documents and Mapped Terms.
- URI and assigned codes for selected terms shown.

The screenshot displays a web interface with two main sections: "Ontology Terms" and "Clinical Trial Annotation".

Ontology Terms Section:

- Header: "Ontology Terms" with a sub-header "Those ontology terms were tagged in the list of document for your selected ontology terms".
- Filter: "Show [dropdown] entries".
- Term Cards: Five cards for "Favipiravir", "Fever", "Genome", "Hospitalization", and "Host".
- Navigation: "Showing 31 to 35 of 94 entries" with a pagination bar (Previous, 1, ..., 6, 7, 8, ..., 19, Next).

Clinical Trial Annotation Section:

- Header: "Clinical Trial Annotation" with a sub-header "Show [dropdown] entries" and a "Search:" field.
- Table:

Ontology Term URI	Assigned Codes
<input checked="" type="checkbox"/> http://purl.obolibrary.org/obo/CHEBI_134722	CHEBI:134722
<input checked="" type="checkbox"/> http://purl.obolibrary.org/obo/CHEBI_134722	DG_34

Showing 1 to 2 of 2 entries (filtered from 215 total entries)

Navigation: Previous, 1, Next

Document Cards Section:

- Filter: "Show [dropdown] entries".
- Five document cards, each with a title, nct_id, and official_title.

Document Title	nct_id	official_title
Persuasion in Medicine: Experimental Evidence on Sender and Signal Effects	NCT04160975	Persuasion in Medicine: Experimental Evidence on Sender and Signal Effects
Glucocorticoid Therapy for COVID-19 Critically Ill Patients With Severe Acute Respiratory Failure	NCT04244591	Glucocorticoid Therapy for Critically Ill Patients With Severe Acute Respiratory Infections Caused by COVID-19: a Prospective, Randomized Controlled Trial
The Efficacy of Lopinavir Plus Ritonavir and Arbidol Against Novel Coronavirus Infection	NCT04252885	A Randomized, Open-label, Controlled Study of the Efficacy of Lopinavir Plus Ritonavir and Arbidol for Treating With Patients With Novel Coronavirus Infection
Clinical Study of Arbidol Hydrochloride Tablets in the Treatment of Pneumonia Caused by Novel Coronavirus	NCT04260594	Randomized, Open, Multicenter Study on the Efficacy and Safety of Arbidol Hydrochloride Tablets in Treating Pneumonia in Patients Infected With Novel Coronavirus
Efficacy and Safety of Hydroxychloroquine for Treatment of COVID-19	NCT04261517	Efficacy and Safety of Hydroxychloroquine for Treatment of COVID-19



Takeaways from Mapping

- Produces results comparable to the gold standard
- Allows for a combination of methods
- Facilitate easy loading and analysis of datasets
- Data Cleaning and Transformation
- Reductions in operating cost
- Supports Assessment and Improvement of Data Quality
 - Support for FIT Metric Impactful
 - Support for FIT Metric Transformable
- Easy Maintenance and Modification

Proportion of matches with the gold standard	Functionality Level Achieved
>50%	Partial
100%	Full
>100%	Met and exceeded

Table 2. Criteria for workflow functionality assessment



Takeaways from Annotation

- Support for highly specific annotation needs
 - Identifying concepts that are unique to the topic, makes it possible to perform intelligent knowledge extraction
- Easy refinement of results
 - quickly ascertain whether a clinical scheme is providing the type of annotations that will be considered ideal for a use context
- Connect annotation to mapping tasks
 - annotation is connected to and is a natural extension of the novel workflow, data preprocessing, dictionary creation, model building, training, tagging and visualization are all embedded as part of the novel workflow
- Extensible to other domains



Workflow Impact Analysis

STRENGTHS

Support for highly specific annotation tasks. | *Connect Annotation to Mapping Tasks* | Simple process for mapping | Tailoring terms for annotation with desired KOS

S

WEAKNESSES

Inherent weaknesses of the algorithms | Lack of definitions in clinical coding schemes

W

FAIRification and FITness of data | *Extensible to other domains* | Reductions in operating, product, and personnel-related costs

O

OPPORTUNITIES

Data unavailable in downloadable formats

T

THREATS



Thanks!

Any **questions** ?

You can find me at

- jclunis@kent.edu
- jclunis@uwm.edu



Credits

Special thanks to all the people who made and released these awesome resources for free:

- Open-Source analytics platform by [KNIME Software](#)
- Biomedical ontology repository [BioPortal](#)
- Presentation template by [SlidesCarnival](#)