Data Engineering and Semantics
هندسة البيانات و دلالاتها

BIBLIOMETRIC-ENHANCED INFORMATION RETRIEVAL AS A TOOL FOR ENRICHING AND VALIDATING WIKIDATA

HOUCEMEDDINE TURKI

RESEARCH ASSISTANT, DATA ENGINEERING AND SEMANTICS RESEARCH UNIT, UNIVERSITY OF SFAX, TUNISIA
VICE-CHAIR, WIKIMEDIA TUNISIA, TUNISIA

DATA ENGINEERING AND SEMANTICS

Created in 2021, it is the first research structure in Tunisia specialized in Wikimedia projects. It is affiliated at the Faculty of Sciences of Sfax, University of Sfax, Tunisia. Its main objective is to develop novel applications of Wikimedia Projects based on Knowledge Engineering, Machine Learning, and Big Data Technologies.

Data Engineering and Semantics
هندسة البيانات و دلالاتها

TEAM

HOUCEMEDDINE TURKI

Medical Student
University of Sfax, Tunisia

MOHAMED ALI HADJ TAIEB

Associate Professor
University of Sfax, Tunisia

MOHAMED BEN AOUICHA

Assistant Professor
University of Sfax, Tunisia

KHALIL CHEBIL

Assistant Professor
University of Carthage, Tunisia

PRIMARY COLLABORATORS

BONAVENTURE DOSSOU

Ph.D. Student
Jacobs University Bremen, Germany

CHRIS EMEZUE

Ph.D. Student
Technische Universität München, Germany

LANE RASBERRY

Wikimedian-in-Residence
University of Virginia, United States of America

DANIEL MIETCHEN

Senior Researcher
Leibniz Institute of Freshwater Ecology and Inland Fisheries, Germany

ADVISORS

ANASTASSIOS POURIS

Professor
University of Pretoria, South Africa

ABRAHAM OWODUNNI

Research Assistant
University of Ilorin, Nigeria

CHRIS FOURIE

Research Engineer
Sisonkebiotik, South Africa

THOMAS SHAFEE

Data Specialist
Swinburne University of Technology, Australia

New Research Project Launched

This work is within the framework of a project funded by the Wikimedia Research Fund to be launched in August 2022 for one year.

This project is entitled *Adapting Wikidata to support clinical practice using Data Science, Semantic Web and Machine Learning*.

WIKIMEDIA
FOUNDATION

Data Engineering and Semantics
هندسة البيانات و دلالاتها

INTRODUCTION

COVERAGE OF WIKIDATA IN THE BIOMEDICAL CONTEXT – AS OF MARCH 2019

WIKIDATA

Created in October 2012.

Represents structured knowledge in the form of RDF triples (Subject – Predicate – Object).

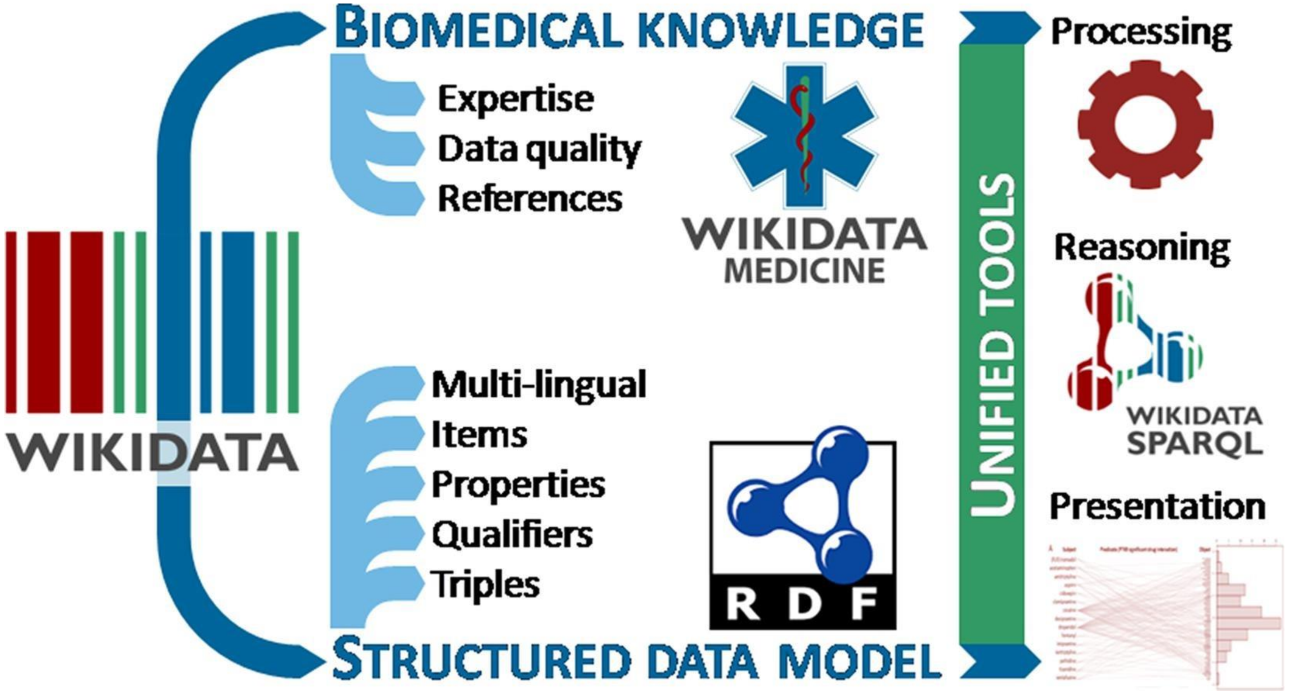
Findable – **A**ccessible – **I**nteroperable – **R**eusable.

Easily available at https://www.wikidata.org.

Items covering a significant subset of the human knowledge ranging from cultural heritage to biomedicine.

Items aligned to external biomedical resources such as *PubMed*, *Medical Subject Headings*, and *UMLS Metathesaurus*.

Wikidata statements are supported by references.



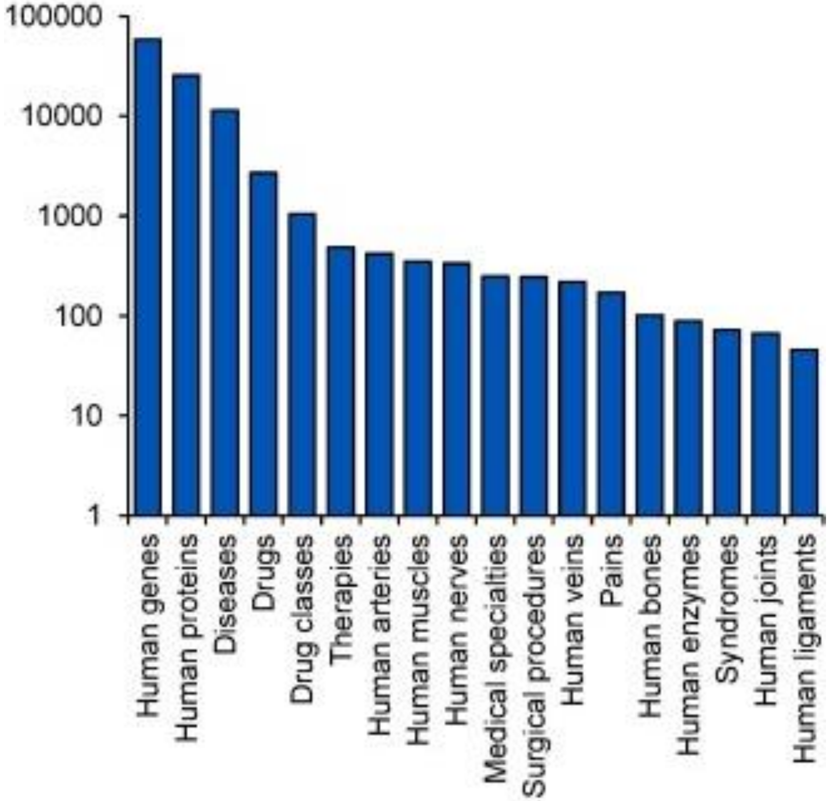
BIOMEDICAL KNOWLEDGE IN WIKIDATA

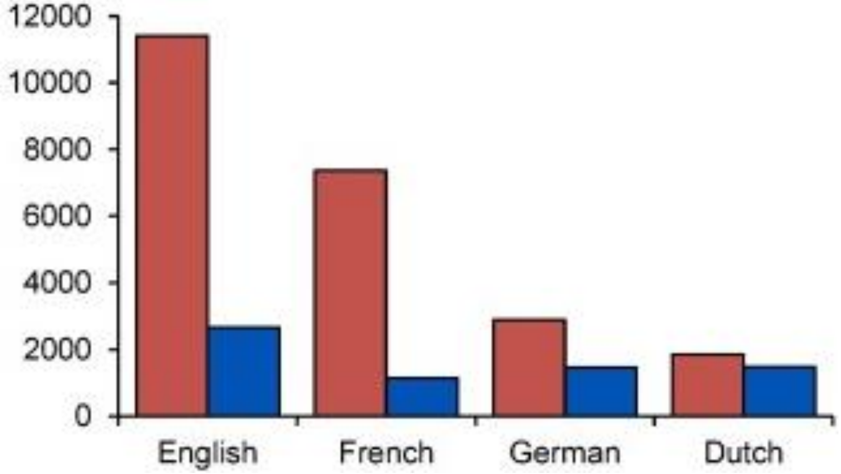
Various types of biomedical items represented: Human genes and proteins, diseases, drugs, therapies, anatomical structures, and symptoms.

Multiple languages are represented in Wikidata: +50 are significantly covered, mostly European and Asian languages.

Uneven coverage of natural languages for biomedical entities in Wikidata: English, French, German and Dutch are the main languages.

Uneven distribution of the types of biomedical entities in Wikidata: Human genes and proteins, diseases, and drugs.





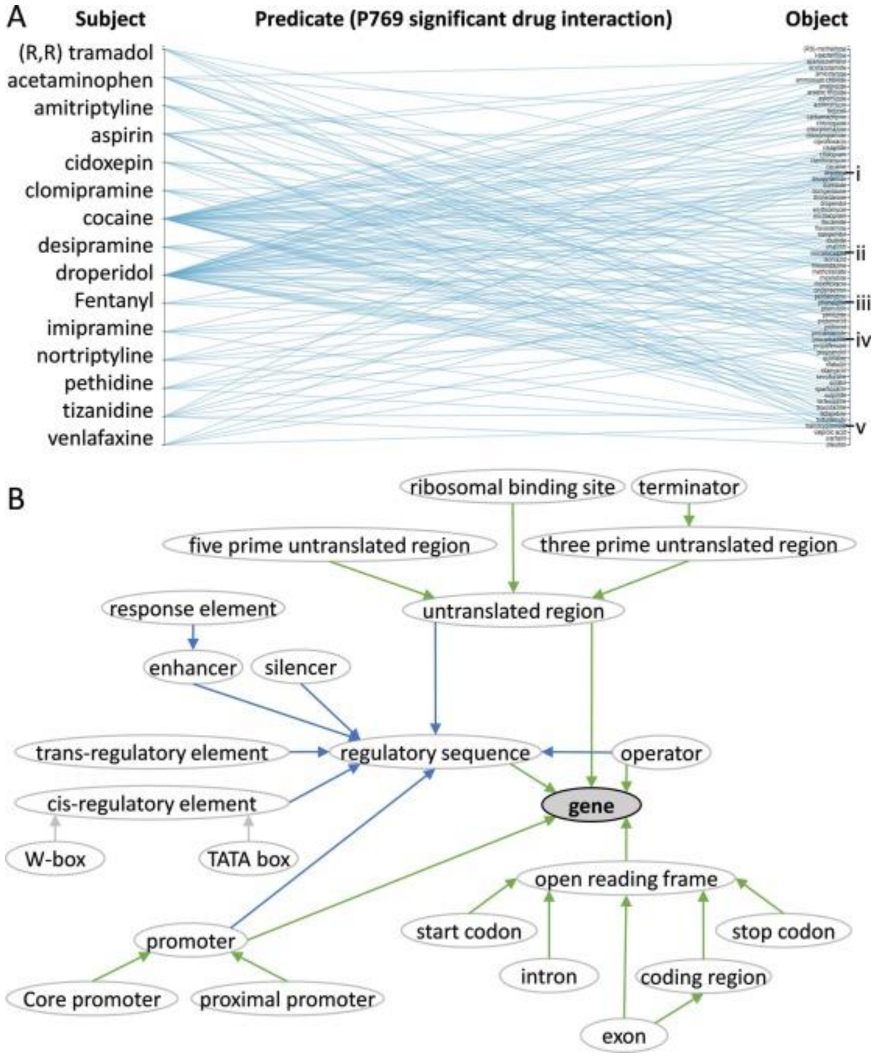
PARSING WIKIDATA

WIKIDATA QUERY SERVICE, MEDIAWIKI API

FINDING INSIGHTS

Synthetizing data based on integrating information about different aspects of information.

Extracting a specific piece of knowledge about a particular topic of interest.



VALIDATING DATA

Finding inconsistencies based on predefined rules using ShEx, SHACL, property constraints, and other tools.

Comparing data with their equivalents in external knowledge graphs.

EASILY EXTENSIBLE

Everyone can create new items and statements – *Special:NewItem*.

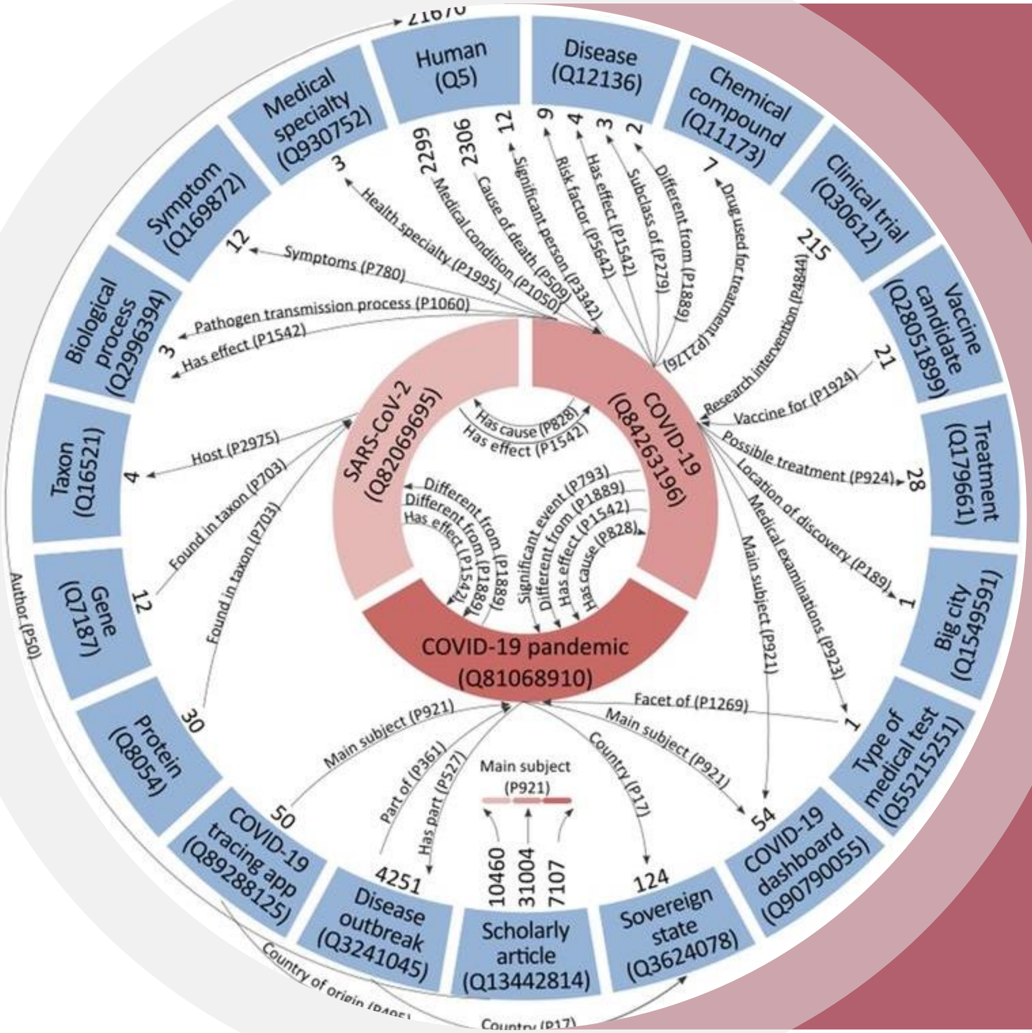
Everyone can apply for new property to support novel types of items – *Wikidata:Property proposal*.

Easy creation of data models and property constraints to ensure the data consistency – *Wikidata:WikiProject COVID-19/Data models*.

Easy alignment to new external resources – *Wikidata:Mix'n'match*.

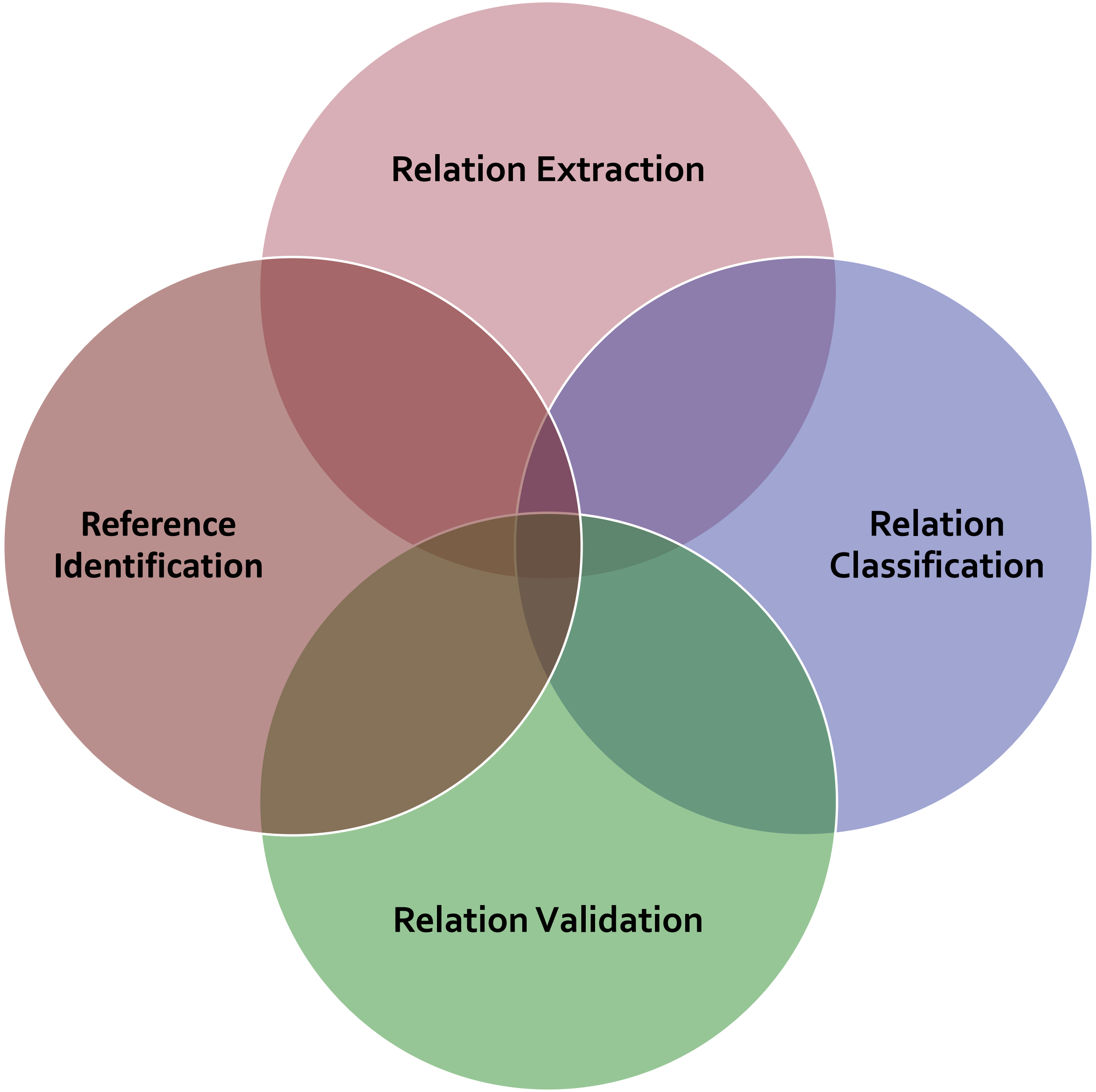
Intuitive embedding in bots for the automatic enrichment of Wikidata – *Wikibase Integrator*.

Possible change of data models upon community consent – *Project:Community portal*.



Biomedical entity (P31)	Number of items	Number of properties		Number of properties per item		Percentage of referenced data
		With references	Without references	With references	Without references	
Drugs	2713	75,259	35,302	27.7	13.0	68.1%
Drug classes	1043	16,855	10,537	16.2	10.1	61.5%
Human enzymes	89	1234	386	13.9	4.3	76.2%
Diseases	11,447	152,622	57,689	13.3	5.0	72.6%
Human genes	**58,691**	**671,282**	**12,949**	**11.4**	**0.2**	**98.1%**
Human proteins	**25,482**	**265,684**	**27,825**	**10.4**	**1.1**	**90.5%**
Human muscles	351	1690	2136	4.8	6.1	44.2%
Pains	171	725	858	4.2	5.0	45.8%
Syndromes	72	173	350	2.4	4.9	33.1%
Human arteries	418	964	2383	2.3	5.7	28.8%
Human joints	67	151	535	2.3	8.0	22.0%
Human bones	102	233	1119	2.3	11.0	17.2%
Human nerves	335	738	1738	2.2	5.2	29.8%
Human veins	220	478	1081	2.2	4.9	30.7%
Medical specialties	248	512	2069	2.1	8.3	19.8%
Therapies	487	931	2312	1.9	4.7	28.7%
Human ligaments	46	56	201	1.2	4.4	21.8%
Surgical procedures	244	261	1099	1.1	4.5	19.2%
Overall	102,226	1,189,848	160,569	11.6	1.6	88.1%

WHAT WIKIDATA REALLY NEEDS



Data Engineering and Semantics
هندسة البيانات و دلالاتها

RESEARCH OUTPUTS

SCHOLARLY PUBLICATIONS IN CONTEXT –
BIBLIOGRAPHIC METADATA, FULL TEXTS

LOADS OF SCHOLARLY PAPERS ARE PUBLISHED EVERY YEAR

2020 STATISTICS – AS OF JULY 7, 2022

PubMed.gov
1,633,069

Clarivate
Web of Science™
3,500,587

PubMed Central
730,997

DataCite
FIND, ACCESS, AND REUSE DATA
3,871,995

RESEARCH PUBLICATIONS IN BRIEF

FULL TEXTS

Detailed texts in a natural language involving insights about study contexts, results and outcomes.

Large size, requires extensive use of advanced techniques of natural language processing and machine learning.

Includes tables, images and diagrams that increase the complexity of their management.

BIBLIOGRAPHIC METADATA

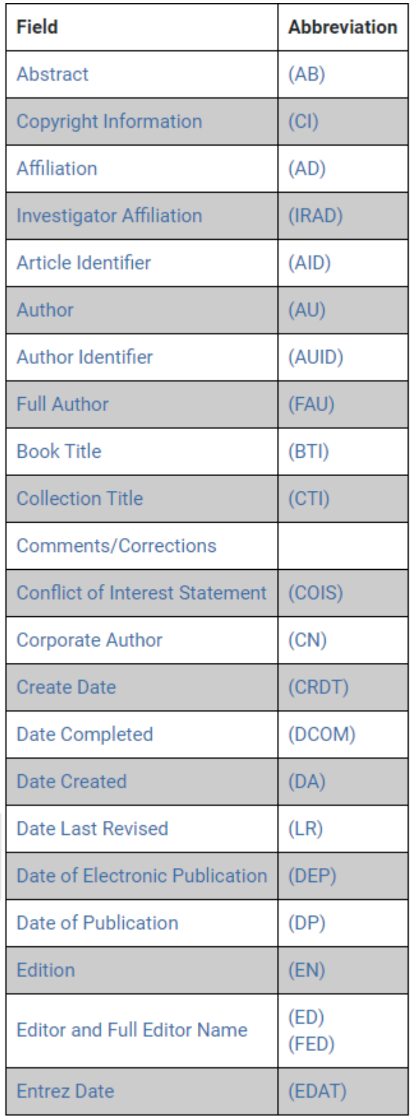
Semi-structured texts providing information about the research venue, the paper, and the authors.

Limited size, pre-processed and requires minor use of information retrieval and machine learning techniques.

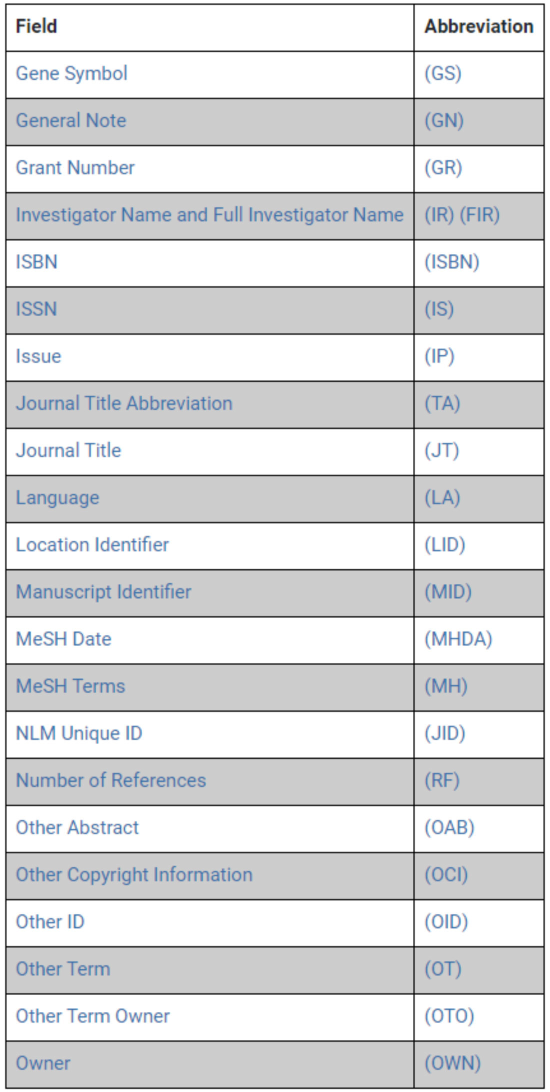
Formatted and annotated by design.

PUBMED SEARCH TAGS

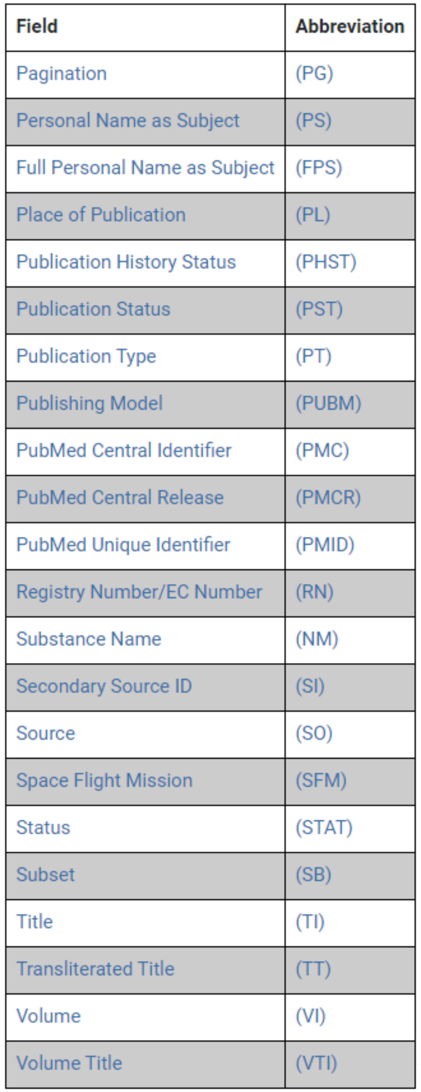
- Many types of bibliographic metadata are assigned abbreviations known as *PubMed Search Tags* or *PubMed Namespaces*.
- This database can be used to enrich bibliographic metadata in Wikidata despite several legal concerns.
- Processing this data can be used to enrich scientific knowledge in Wikidata.



Field	Abbreviation
Abstract	(AB)
Copyright Information	(CI)
Affiliation	(AD)
Investigator Affiliation	(IRAD)
Article Identifier	(AID)
Author	(AU)
Author Identifier	(AUID)
Full Author	(FAU)
Book Title	(BTI)
Collection Title	(CTI)
Comments/Corrections	
Conflict of Interest Statement	(COIS)
Corporate Author	(CN)
Create Date	(CRDT)
Date Completed	(DCOM)
Date Created	(DA)
Date Last Revised	(LR)
Date of Electronic Publication	(DEP)
Date of Publication	(DP)
Edition	(EN)
Editor and Full Editor Name	(ED) (FED)
Entrez Date	(EDAT)



Field	Abbreviation
Gene Symbol	(GS)
General Note	(GN)
Grant Number	(GR)
Investigator Name and Full Investigator Name	(IR) (FIR)
ISBN	(ISBN)
ISSN	(IS)
Issue	(IP)
Journal Title Abbreviation	(TA)
Journal Title	(JT)
Language	(LA)
Location Identifier	(LID)
Manuscript Identifier	(MID)
MeSH Date	(MHDA)
MeSH Terms	(MH)
NLM Unique ID	(JID)
Number of References	(RF)
Other Abstract	(OAB)
Other Copyright Information	(OCI)
Other ID	(OID)
Other Term	(OT)
Other Term Owner	(OTO)
Owner	(OWN)



Field	Abbreviation
Pagination	(PG)
Personal Name as Subject	(PS)
Full Personal Name as Subject	(FPS)
Place of Publication	(PL)
Publication History Status	(PHST)
Publication Status	(PST)
Publication Type	(PT)
Publishing Model	(PUBM)
PubMed Central Identifier	(PMC)
PubMed Central Release	(PMCR)
PubMed Unique Identifier	(PMID)
Registry Number/EC Number	(RN)
Substance Name	(NM)
Secondary Source ID	(SI)
Source	(SO)
Space Flight Mission	(SFM)
Status	(STAT)
Subset	(SB)
Title	(TI)
Transliterated Title	(TT)
Volume	(VI)
Volume Title	(VTI)

MESH KEYWORDS

Controlled keywords assigned to PubMed Records by the curators of NCBI databases

Easier to process: Have a particular layout (Heading/Qualifier):

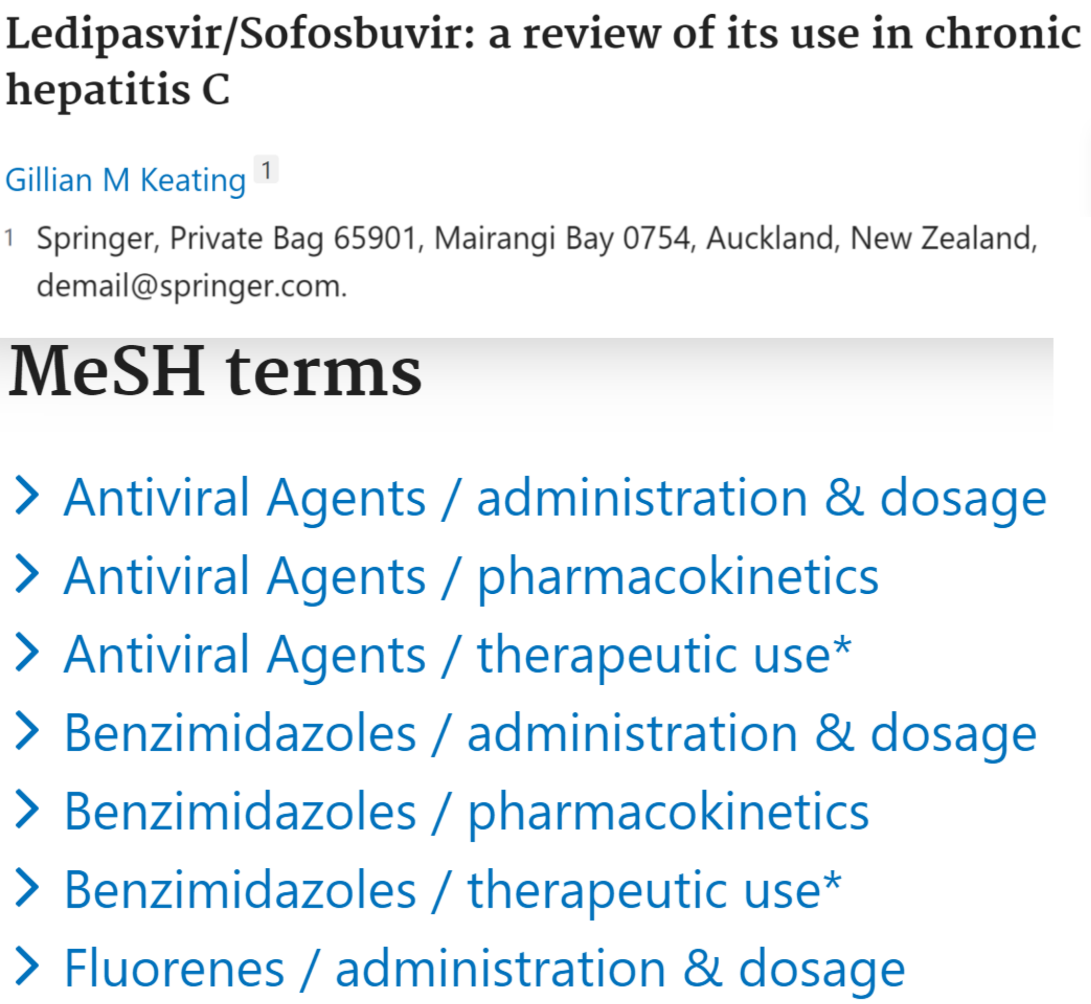
- MeSH qualifiers are predefined: 89 qualifiers
- MeSH headings are assigned from the *Medical Subject Headings* Taxonomy

Shorter than full texts and abstracts of scholarly publications

Reflect the output of scholarly publications

Can be retrieved thanks to:

- NCBI Entrez API
- Biopython Python Library



Ledipasvir/Sofosbuvir: a review of its use in chronic hepatitis C

Gillian M Keating [1]

[1] Springer, Private Bag 65901, Mairangi Bay 0754, Auckland, New Zealand, demail@springer.com.

MeSH terms

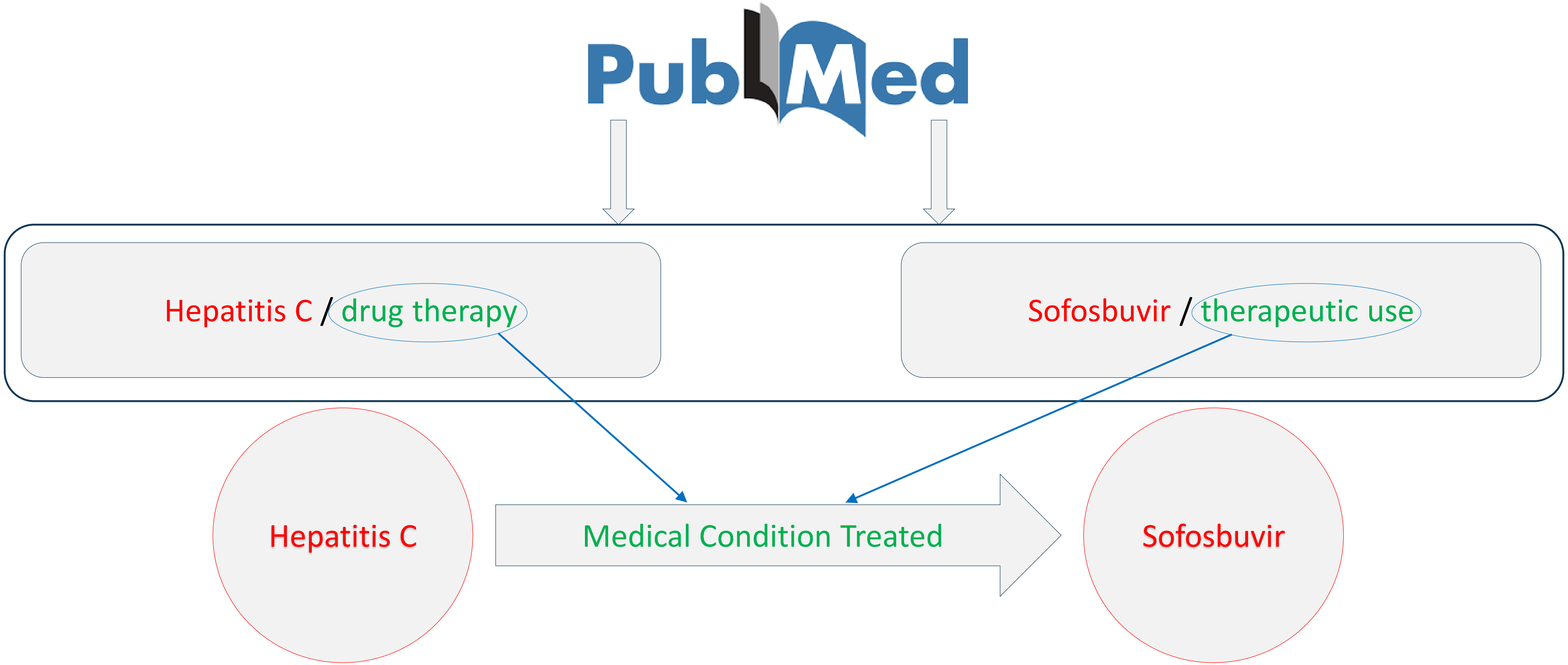
- Antiviral Agents / administration & dosage
- Antiviral Agents / pharmacokinetics
- Antiviral Agents / therapeutic use*
- Benzimidazoles / administration & dosage
- Benzimidazoles / pharmacokinetics
- Benzimidazoles / therapeutic use*
- Fluorenes / administration & dosage

Data Engineering and Semantics
هندسة البيانات و دلالاتها

RELATION CLASSIFICATION

MESH2MATRIX

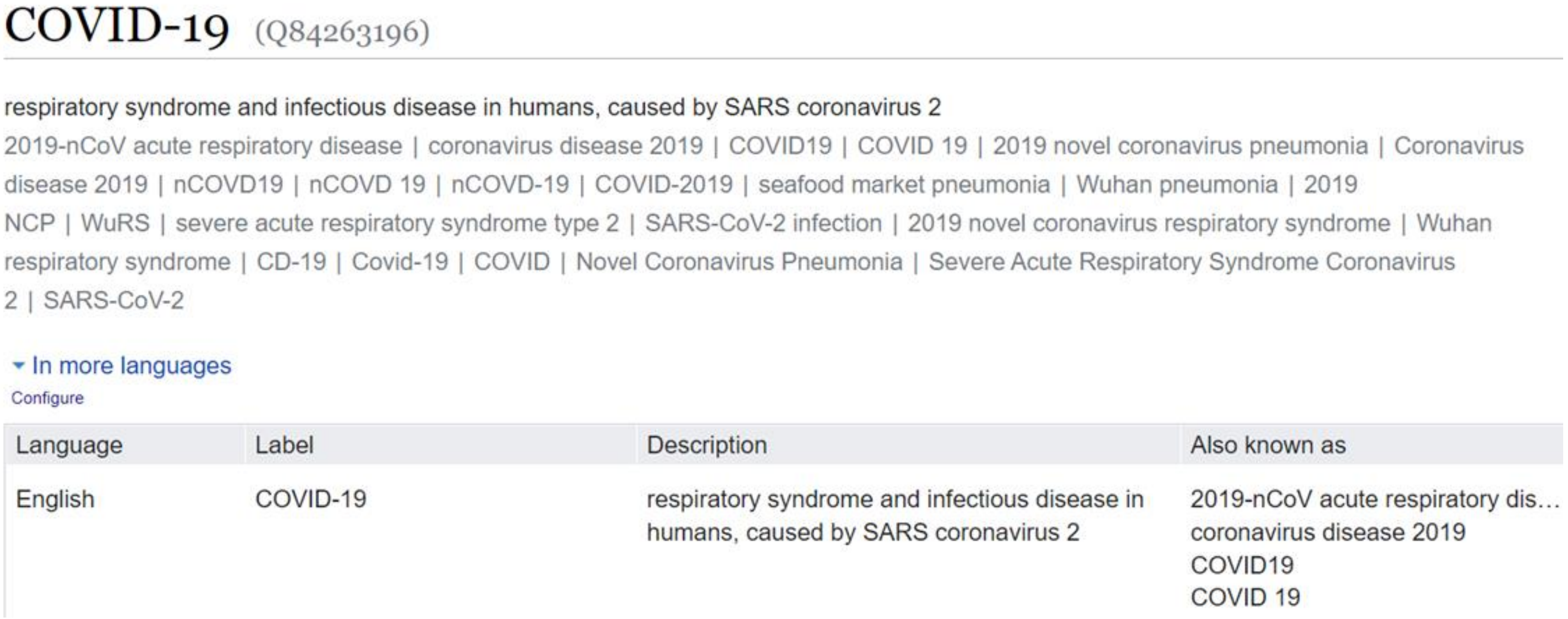
Principles


PubMed
Hepatitis C / drug therapy
Sofosbuvir / therapeutic use
Hepatitis C
Medical Condition Treated
Sofosbuvir

We need a dataset of biomedical relations

« Wikidata can provide such relations as a multidisciplinary open knowledge graph »

Wikidata



COVID-19 (Q84263196)

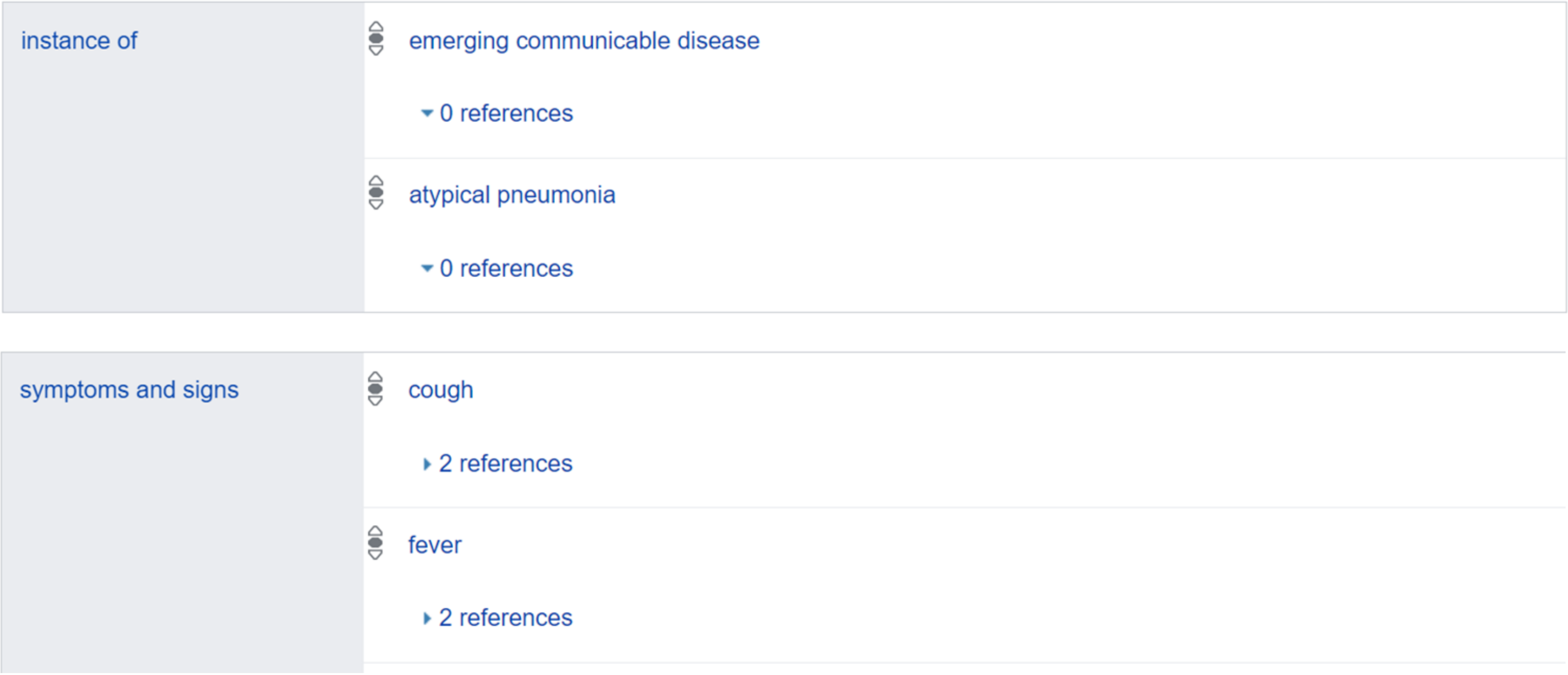
respiratory syndrome and infectious disease in humans, caused by SARS coronavirus 2

2019-nCoV acute respiratory disease | coronavirus disease 2019 | COVID19 | COVID 19 | 2019 novel coronavirus pneumonia | Coronavirus disease 2019 | nCOVD19 | nCOVD 19 | nCOVD-19 | COVID-2019 | seafood market pneumonia | Wuhan pneumonia | 2019 NCP | WuRS | severe acute respiratory syndrome type 2 | SARS-CoV-2 infection | 2019 novel coronavirus respiratory syndrome | Wuhan respiratory syndrome | CD-19 | Covid-19 | COVID | Novel Coronavirus Pneumonia | Severe Acute Respiratory Syndrome Coronavirus 2 | SARS-CoV-2

Language	Label	Description	Also known as
English	COVID-19	respiratory syndrome and infectious disease in humans, caused by SARS coronavirus 2	2019-nCoV acute respiratory dis… coronavirus disease 2019 COVID19 COVID 19

- Concepts assigned labels, descriptions and aliases in multiple languages
- Taxonomic relations (e.g., instance of)
- Non-Taxonomic relations (e.g., Symptoms and signs)
- Property constraints
- Aligned to MeSH Terms

Wikidata



instance of

emerging communicable disease

0 references

atypical pneumonia

0 references

symptoms and signs

cough

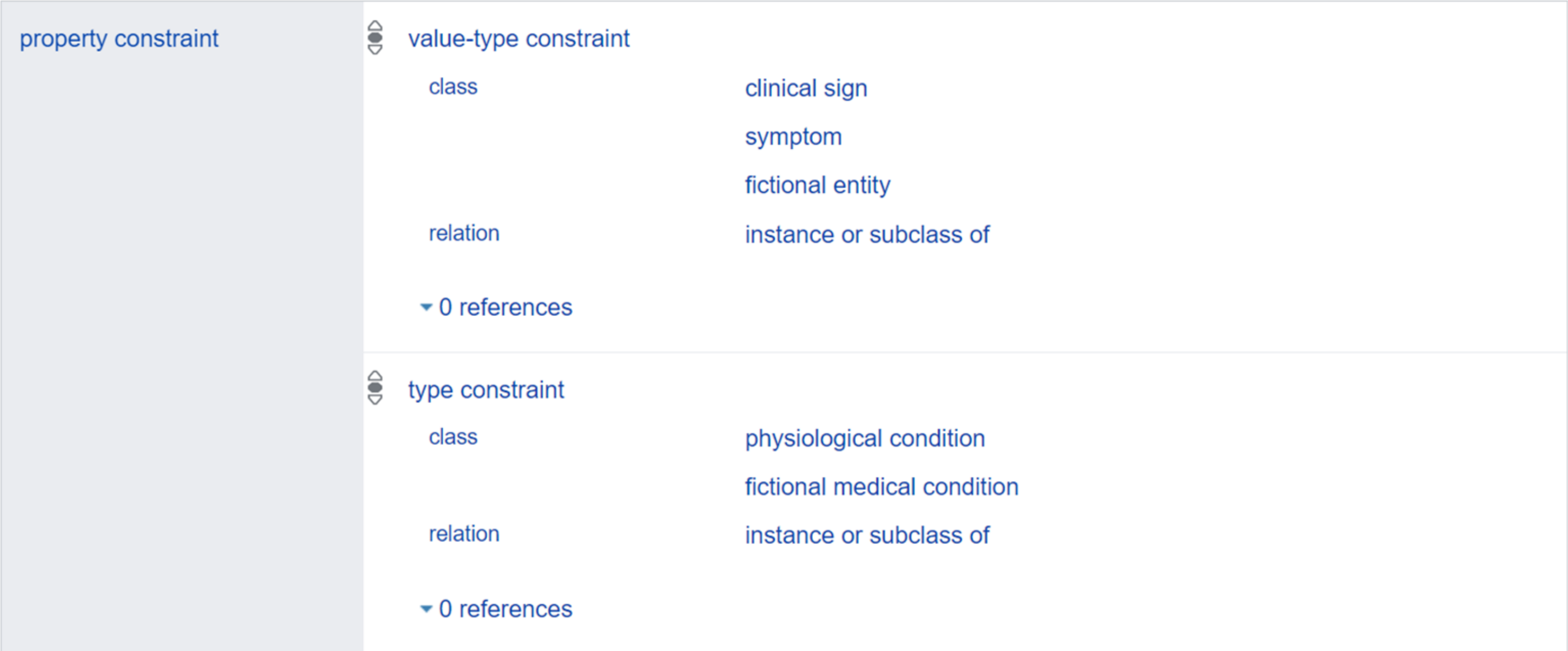
2 references

fever

2 references

- Concepts assigned labels, descriptions and aliases in multiple languages
- Taxonomic relations (e.g., instance of)
- Non-Taxonomic relations (e.g., Symptoms and signs)
- Property constraints
- Aligned to MeSH Terms

Wikidata



property constraint

value-type constraint

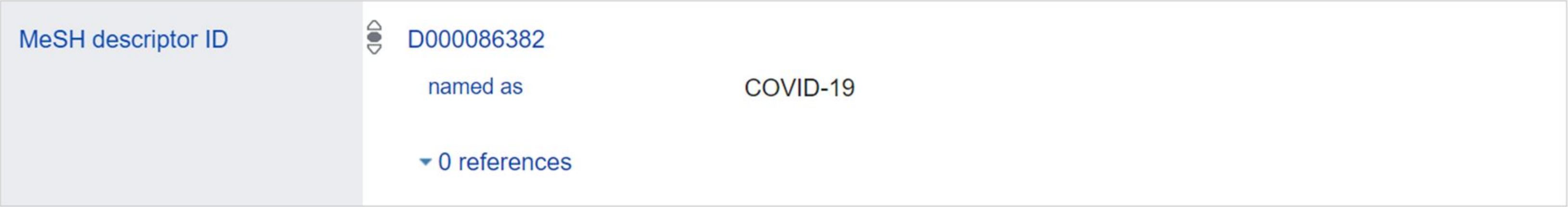
class	clinical sign symptom fictional entity
relation	instance or subclass of

0 references

type constraint

class	physiological condition fictional medical condition
relation	instance or subclass of

0 references



MeSH descriptor ID

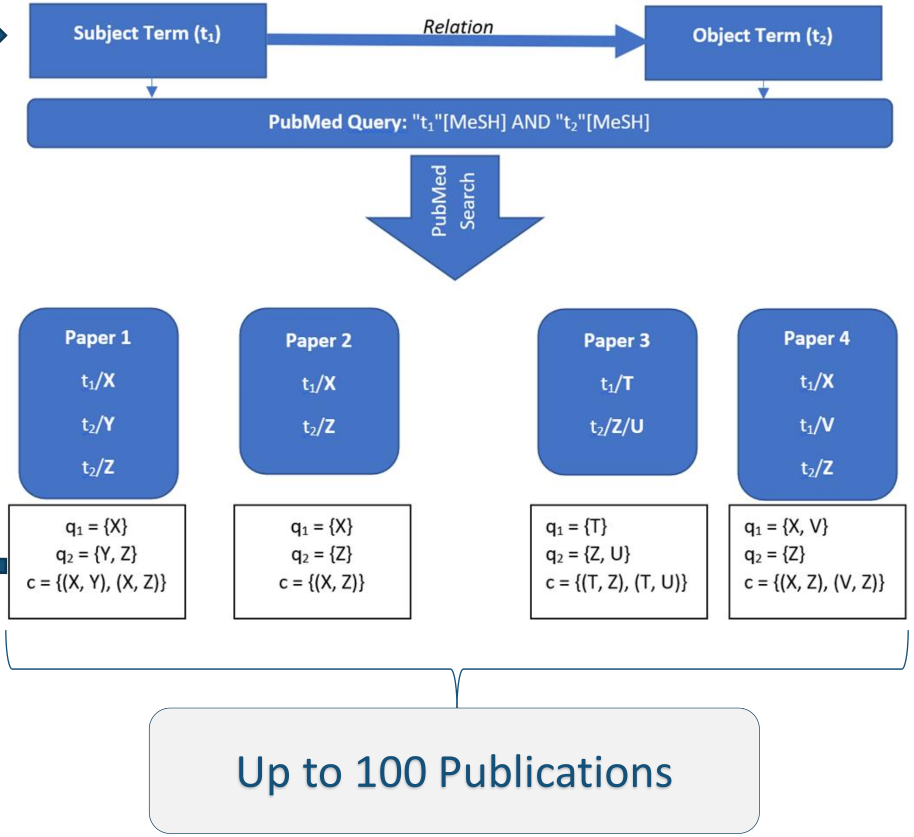
D000086382

named as	COVID-19

0 references

- Concepts assigned labels, descriptions and aliases in multiple languages
- Taxonomic relations (e.g., instance of)
- Non-Taxonomic relations (e.g., Symptoms and signs)
- Property constraints
- Aligned to MeSH Terms

```
SELECT ?subject ?reltype ?object WITH {
  SELECT * WHERE {
    ?item wdt:P486 ?subject.
    }
  }
AS %item
WHERE {
  INCLUDE %item.
  ?item ?reltype ?item1.
  ?item1 wdt:P486 ?object.
}
LIMIT 81000
```



Subject Term (t_1)

Relation

Object Term (t_2)

PubMed Query: "t_1"[MeSH] AND "t_2"[MeSH]

PubMed Search

Paper 1
t_1/X
t_2/Y
t_2/Z

$q_1 = \{X\}$
$q_2 = \{Y, Z\}$
$c = \{(X, Y), (X, Z)\}$

Paper 2
t_1/X
t_2/Z

$q_1 = \{X\}$
$q_2 = \{Z\}$
$c = \{(X, Z)\}$

Paper 3
t_1/T
t_2/Z/U

$q_1 = \{T\}$
$q_2 = \{Z, U\}$
$c = \{(T, Z), (T, U)\}$

Paper 4
t_1/X
t_1/V
t_2/Z

$q_1 = \{X, V\}$
$q_2 = \{Z\}$
$c = \{(X, Z), (V, Z)\}$

Up to 100 Publications

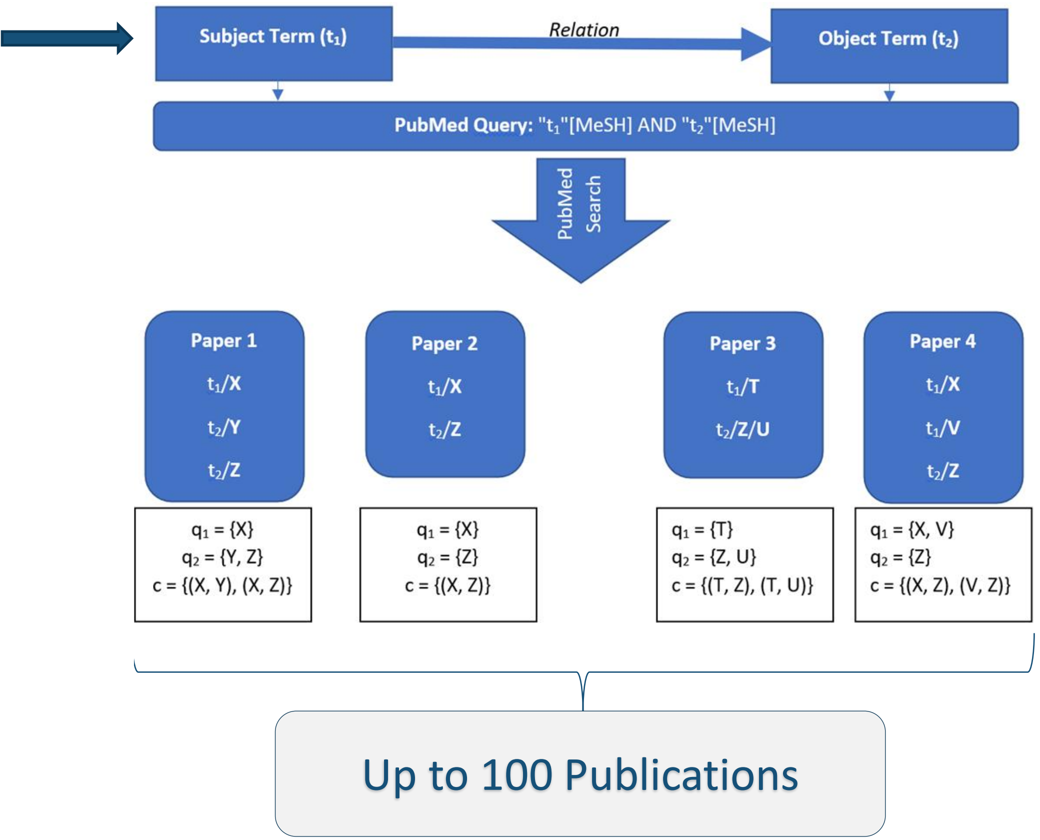
	T	U	V	X	Y	Z
T	0	0	0	0	0	0
U	0.25	0	0	0	0	0
V	0	0	0	0	0	0
X	0	0	0	0	0	0
Y	0	0	0	0.25	0	0
Z	0.5	0	0	0.75	0	0

Relation

Storage in *MeSH2Matrix* Dataset

```
SELECT ?subject ?reltype ?object WITH {
  SELECT * WHERE {
    ?item wdt:P486 ?subject.
    }
  }
AS %item
WHERE {
  INCLUDE %item.
  ?item ?reltype ?item1.
  ?item1 wdt:P486 ?object.
}
LIMIT 81000
```

```
SELECT ?subject ?reltype ?object WITH {
  SELECT * WHERE {
    ?item wdt:P486 ?subject.
    }
  }
AS %item
WHERE {
  INCLUDE %item.
  ?item ?reltype ?item1.
  ?item1 wdt:P486 ?object.
}
LIMIT 81000
```



Subject Term (t_1)

Relation

Object Term (t_2)

PubMed Query: "t_1"[MeSH] AND "t_2"[MeSH]

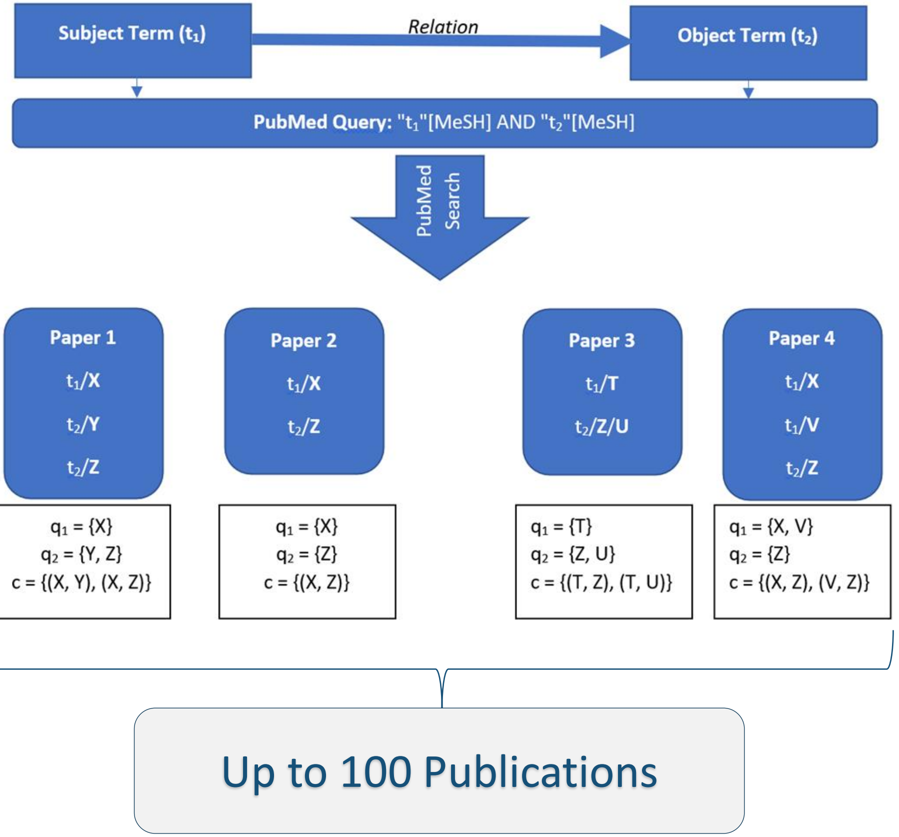
PubMed Search

Paper 1	Paper 2	Paper 3	Paper 4
t_1/X t_2/Y t_2/Z	t_1/X t_2/Z	t_1/T t_2/Z/U	t_1/X t_1/V t_2/Z
q_1 = {X} q_2 = {Y, Z} c = {(X, Y), (X, Z)}	q_1 = {X} q_2 = {Z} c = {(X, Z)}	q_1 = {T} q_2 = {Z, U} c = {(T, Z), (T, U)}	q_1 = {X, V} q_2 = {Z} c = {(X, Z), (V, Z)}

Up to 100 Publications

```
SELECT ?subject ?reltype ?object WITH {
  SELECT * WHERE {
    ?item wdt:P486 ?subject.
    }
  }
AS %item
WHERE {
  INCLUDE %item.
  ?item ?reltype ?item1.
  ?item1 wdt:P486 ?object.
}
LIMIT 81000
```

WIKIDATA



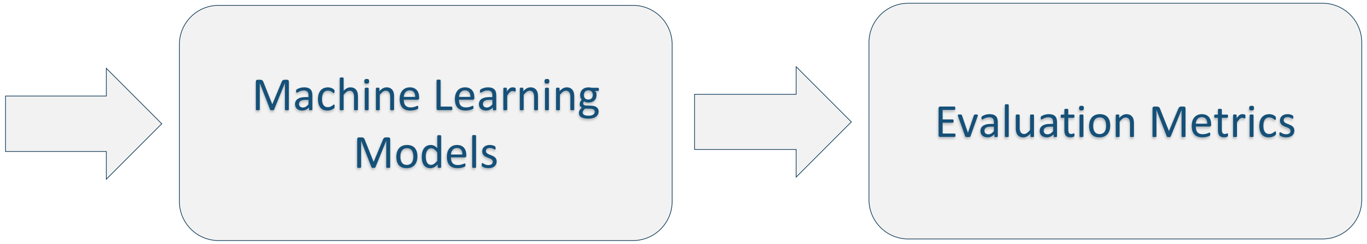
	T	U	V	X	Y	Z
T	0	0	0	0	0	0
U	0.25	0	0	0	0	0
V	0	0	0	0	0	0
X	0	0	0	0	0	0
Y	0	0	0	0.25	0	0
Z	0.5	0	0	0.75	0	0

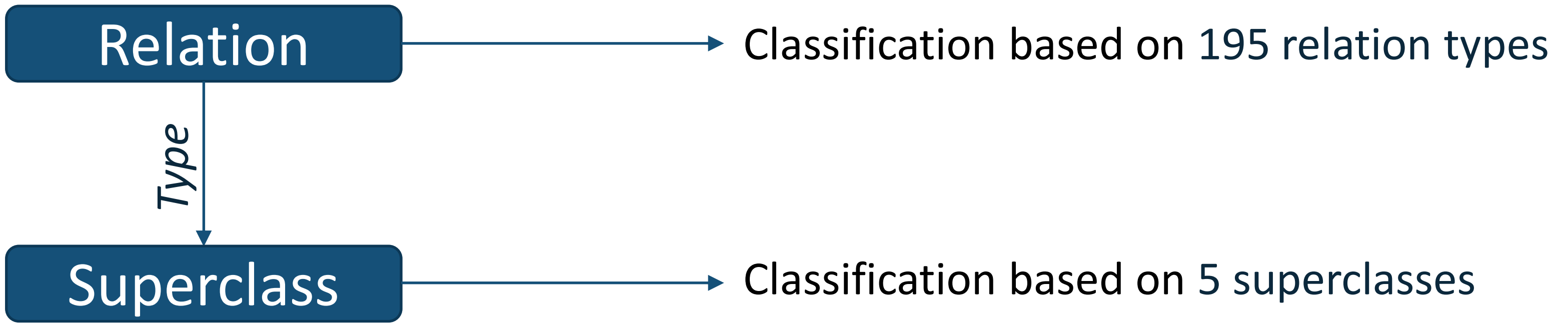
Relation

Storage in *MeSH2Matrix* Dataset

Biomedical Relation Classification

	T	U	V	X	Y	Z
T	0	0	0	0	0	0
U	0.25	0	0	0	0	0
V	0	0	0	0	0	0
X	0	0	0	0	0	0
Y	0	0	0	0.25	0	0
Z	0.5	0	0	0.75	0	0



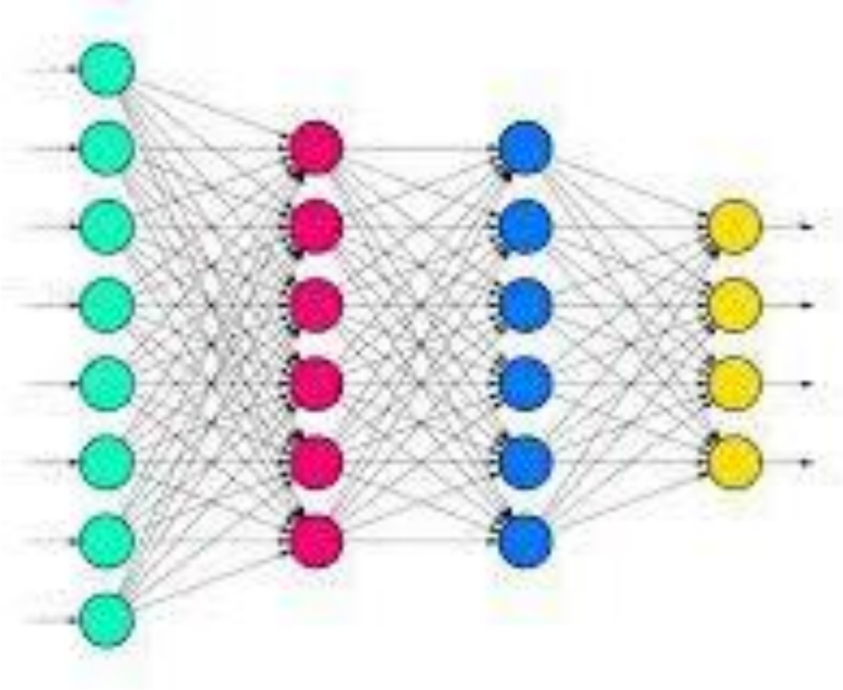


Classification based on 195 relation types

Classification based on 5 superclasses

Machine Learning Models

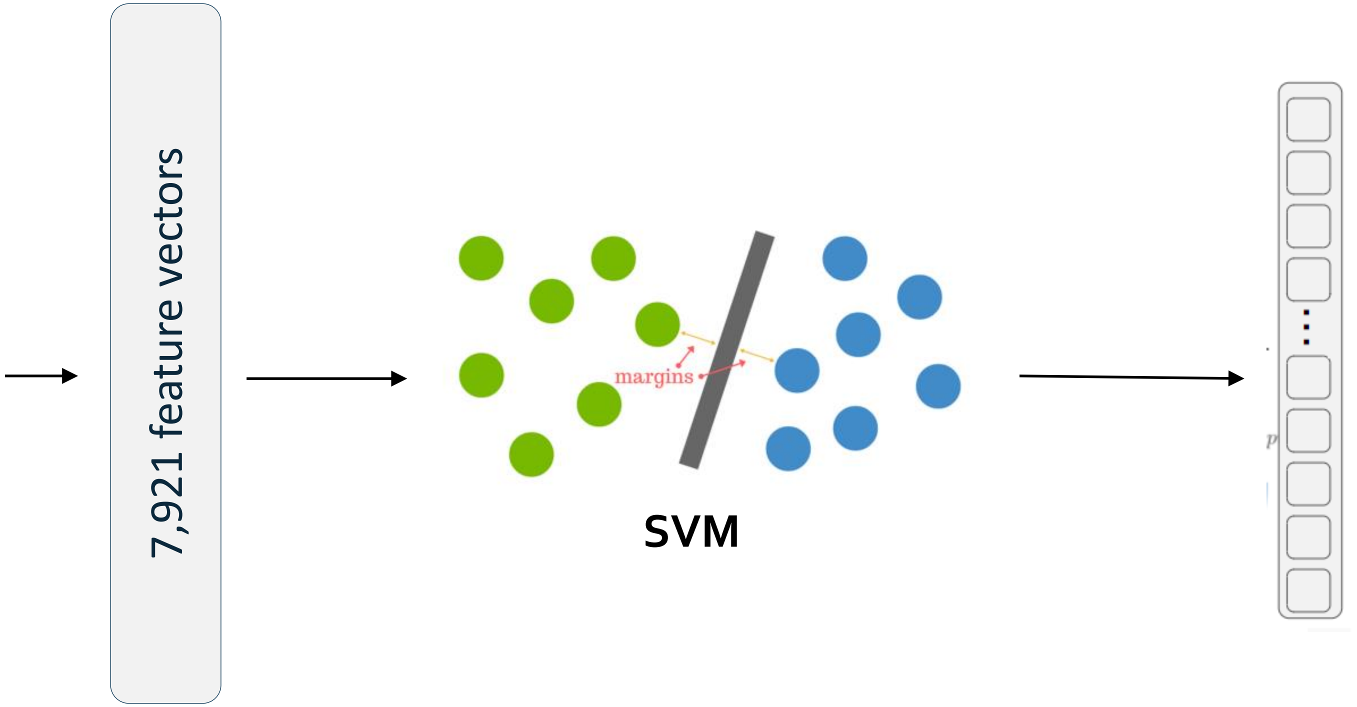
- **Output Size:** Number of classes (195, 5)
- ***D-Net:*** Fully Connected or Dense Model
  - **Feature Size:** (3, 960)
  - **Hidden Layer Size:** (1, 980)
  - **Regularization Method:** Dropout
  - **Activation Function between Input and Hidden Layers:** ReLU (introduces non-linearity)
  - **Activation Function on the Output Layer:** Softmax (computes the probability of the input to belong to each class)



Machine Learning Models

- ***SVM:*** Support vector machines (SVMs) are best suited for samples with many features because their ability to learn is independent of the features space

	T	U	V	X	Y	Z
T	0	0	0	0	0	0
U	0.25	0	0	0	0	0
V	0	0	0	0	0	0
X	0	0	0	0	0	0
Y	0	0	0	0.25	0	0
Z	0.5	0	0	0.75	0	0

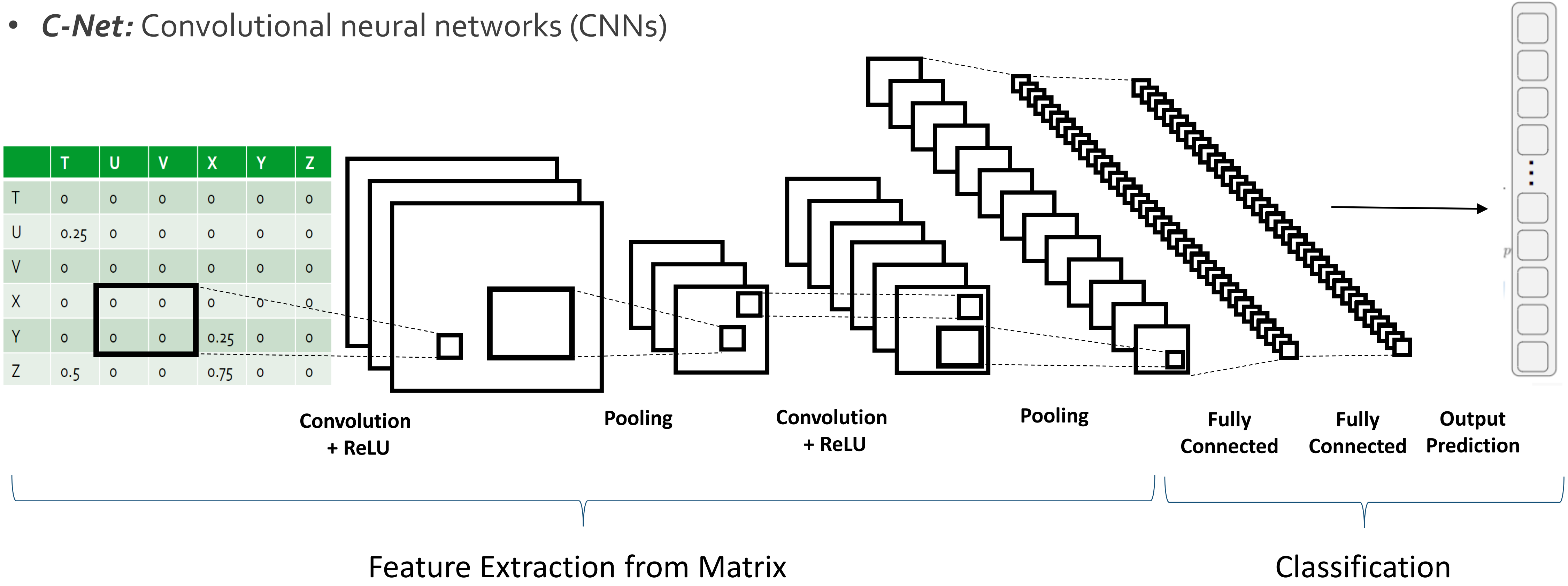


7,921 feature vectors

SVM

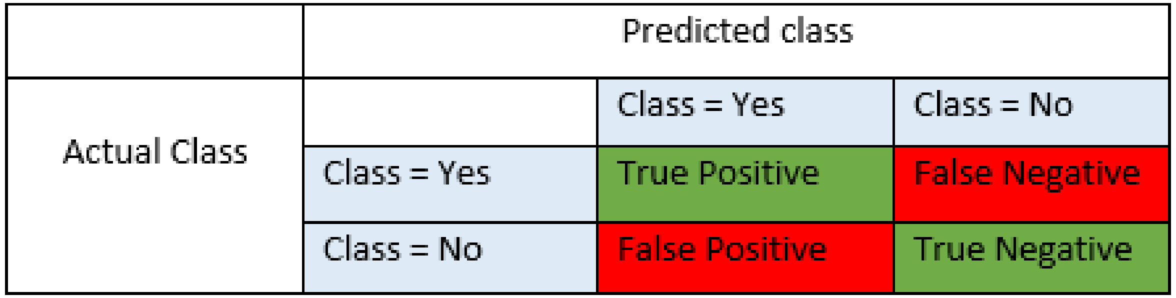
Machine Learning Models

- ***C-Net:*** Convolutional neural networks (CNNs)



	T	U	V	X	Y	Z
T	0	0	0	0	0	0
U	0.25	0	0	0	0	0
V	0	0	0	0	0	0
X	0	0	0	0	0	0
Y	0	0	0	0.25	0	0
Z	0.5	0	0	0.75	0	0

Evaluation Metrics



		Predicted class	
Actual Class		Class = Yes	Class = No
	Class = Yes	True Positive	False Negative
	Class = No	False Positive	True Negative

- True Positives (TP)
- True Negatives (TN)
- False Positives (FP)
- False Negatives (FN)

$$\boxed{Accuracy = \frac{TP + TN}{TP + TN + FP + FN}}$$

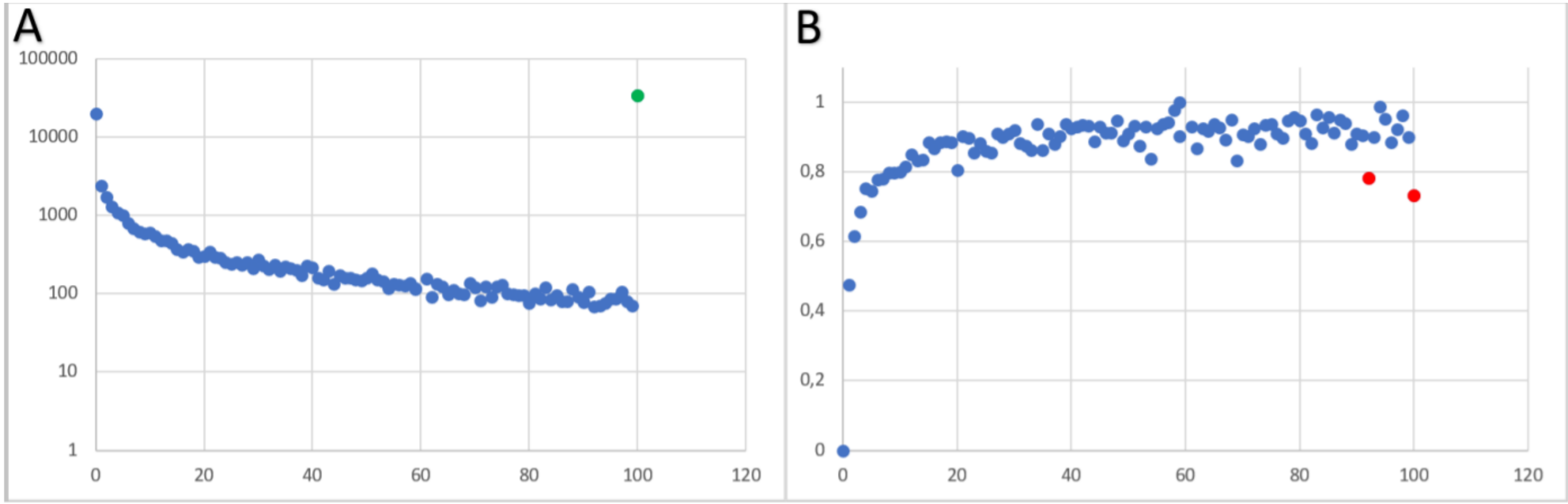
$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$\boxed{F1 = \frac{2 * (Recall * Precision)}{Recall + Precision}}$$

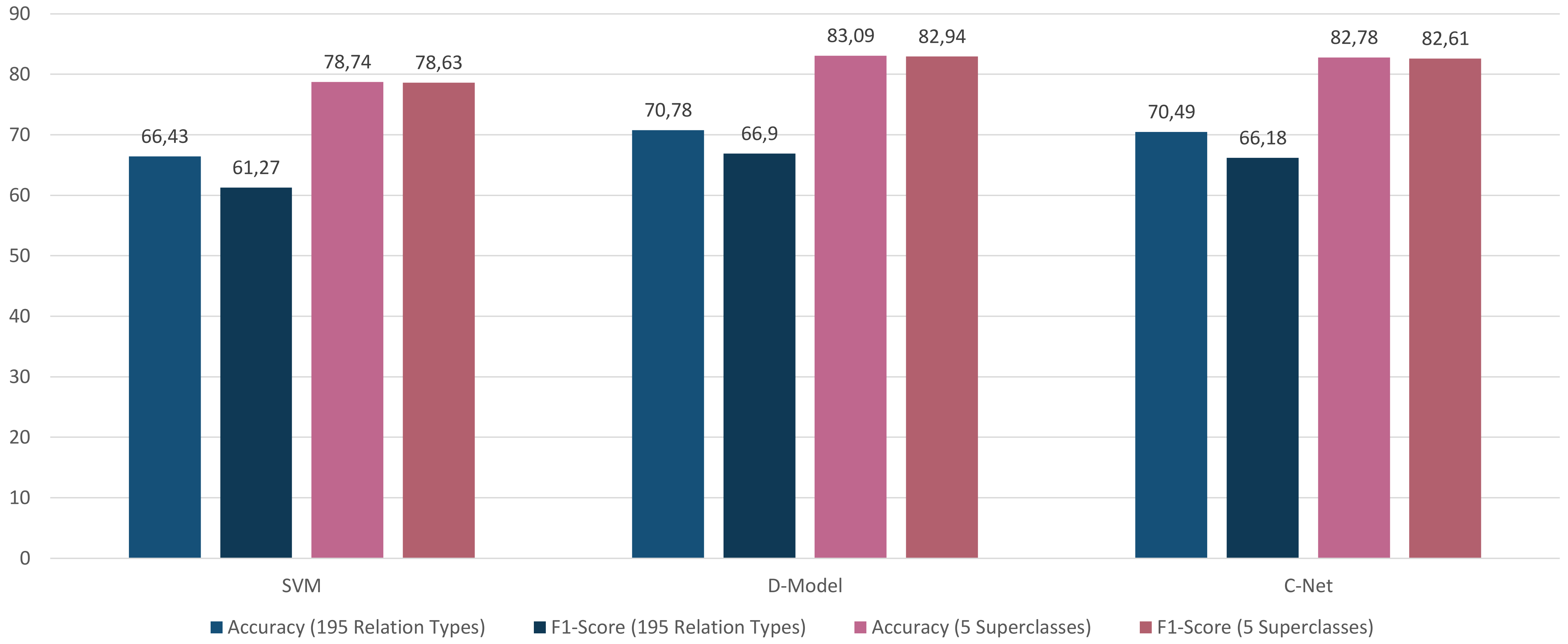
MeSH2Matrix Generation

Relation Class	Relation Types	Relations	Matrices	Rate
Non-Biomedical Non-Symmetric	156	17,758	**9,423**	**0.531**
Biomedical Non-Symmetric	53	27,429	**17,931**	**0.654**
Non-Biomedical Symmetric	12	9,000	**6,353**	**0.706**
Biomedical Symmetric	3	1,441	**801**	**0.556**
Taxonomic	3	25,372	**11,961**	**0.471**



Variables in function of the number of PubMed publications about a given association: Number of semantic relations (A, Log-Scale), Rate of semantic relations returning matrices (B)

Biomedical Relation Classification



SISONKE-BIOTIK

Data Availability

For reproducibility purposes, our source code and dataset are currently available at https://github.com/SisonkeBiotik-Africa/MeSH2Matrix

Data Engineering and Semantics
هندسة البيانات و دلالاتها

RELATION EXTRACTION AND VALIDATION

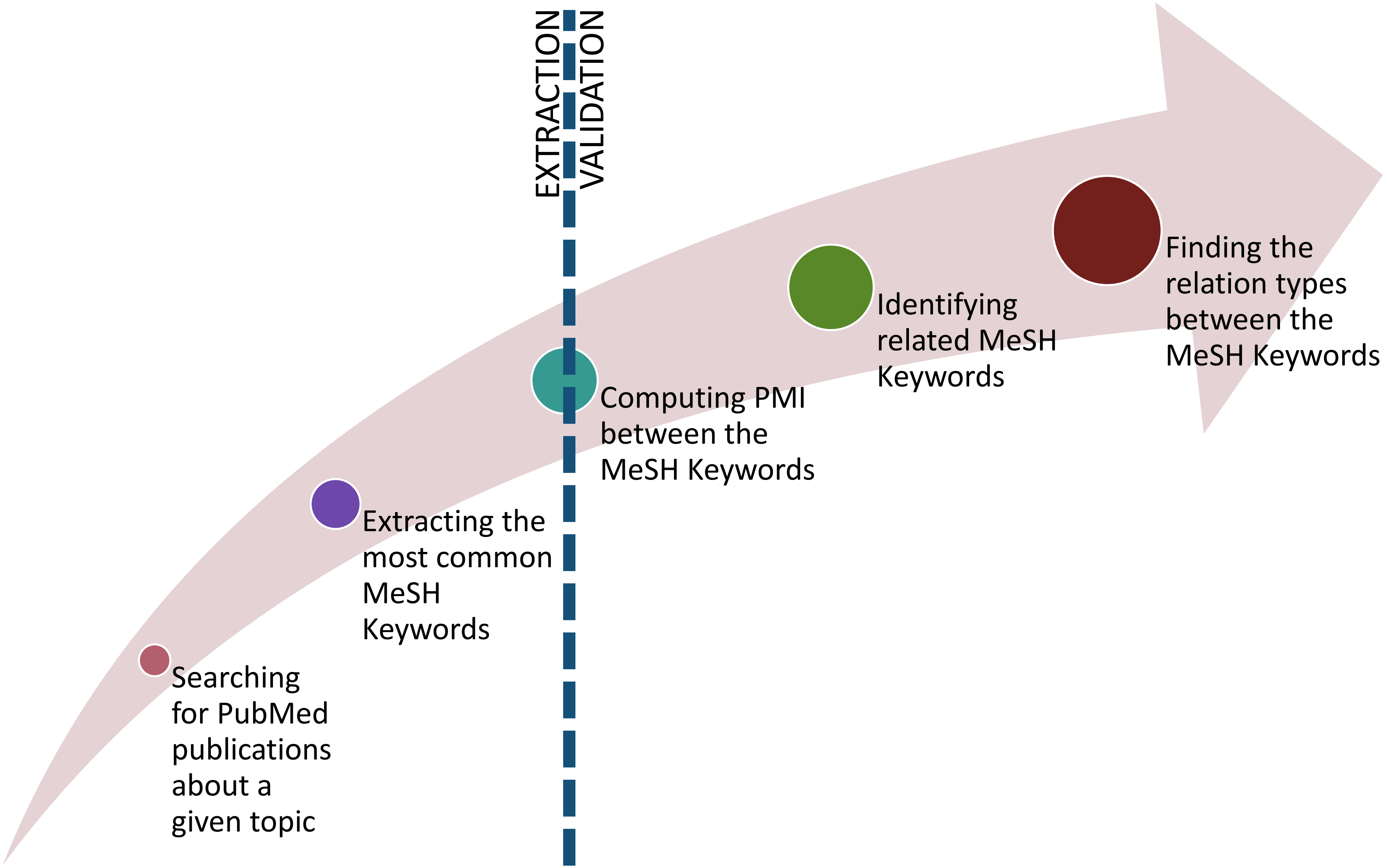
MESH2ONTOLOGY

POINTWISE MUTUAL INFORMATION

- A simple measure of association between entities.
- In computational linguistics, PMI has been used for finding collocations and associations between words.
- MeSH Keywords are predefined and formatted. There is no need for advanced methods for identifying associations.

[illegible]

PROCESS FOR RELATION EXTRACTION AND VALIDATION



EXTRACTION

VALIDATION

Searching for PubMed publications about a given topic

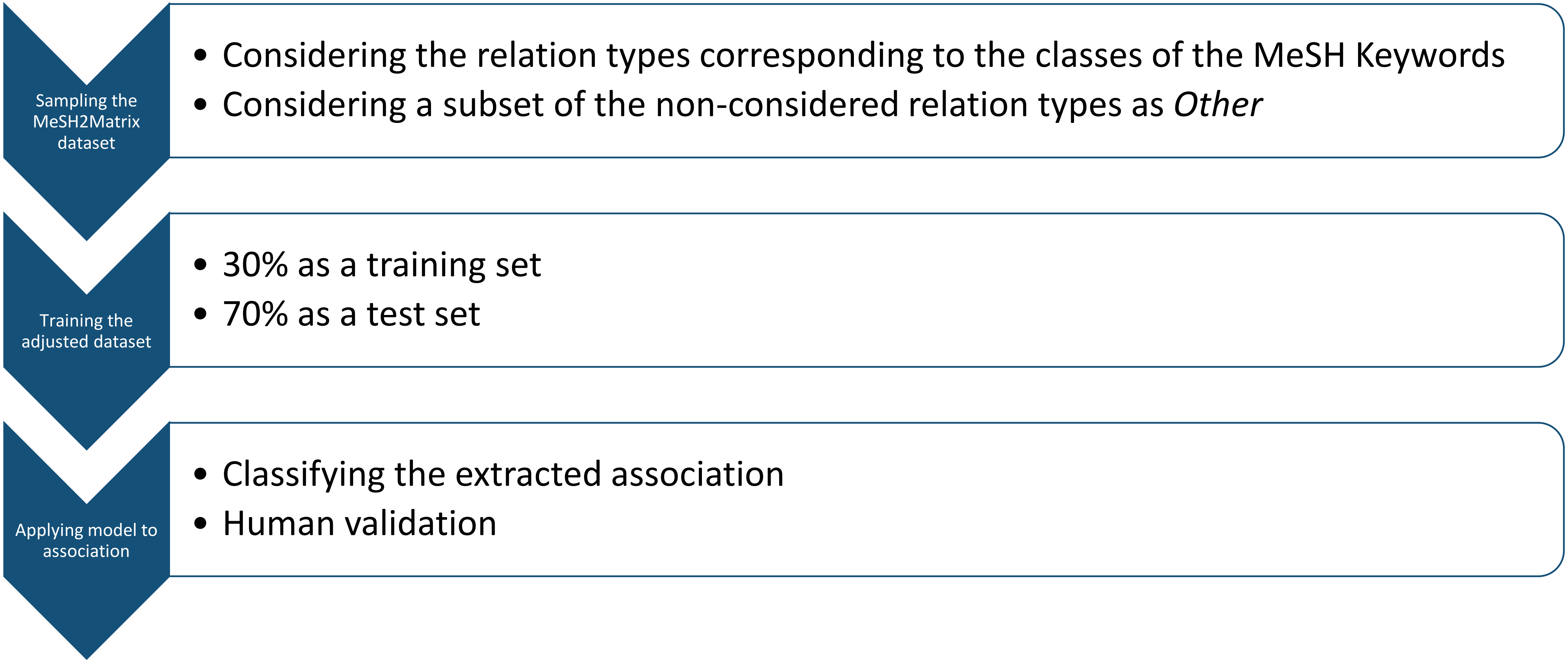
Extracting the most common MeSH Keywords

Computing PMI between the MeSH Keywords

Identifying related MeSH Keywords

Finding the relation types between the MeSH Keywords

FINDING RELATION TYPES BETWEEN MESH KEYWORDS



Sampling the MeSH2Matrix dataset

- Considering the relation types corresponding to the classes of the MeSH Keywords
- Considering a subset of the non-considered relation types as *Other*

Training the adjusted dataset

- 30% as a training set
- 70% as a test set

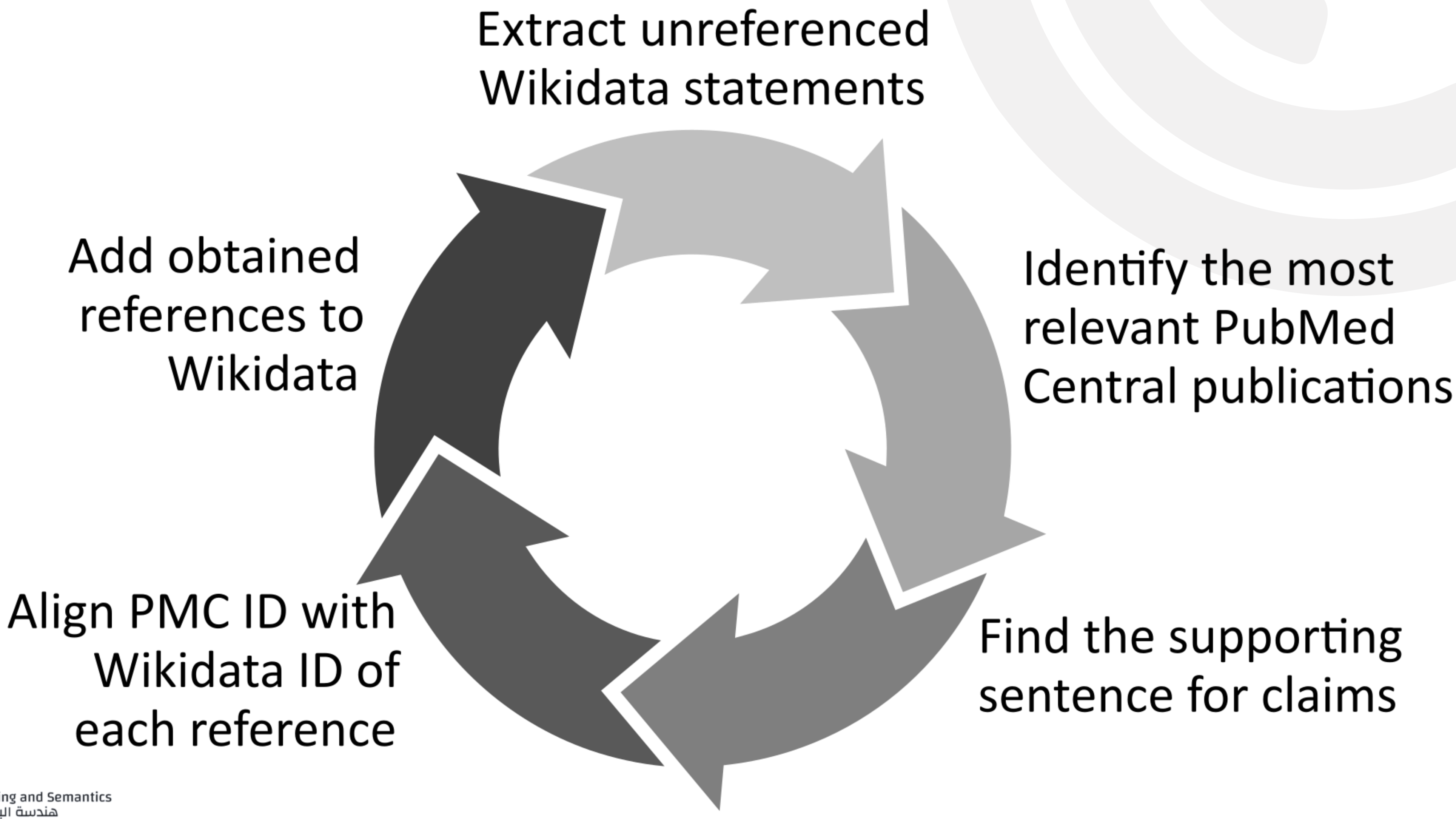
Applying model to association

- Classifying the extracted association
- Human validation

REFERENCE IDENTIFICATION

REFB

PROCESS FOR REFERENCE IDENTIFICATION



Extract unreferenced Wikidata statements

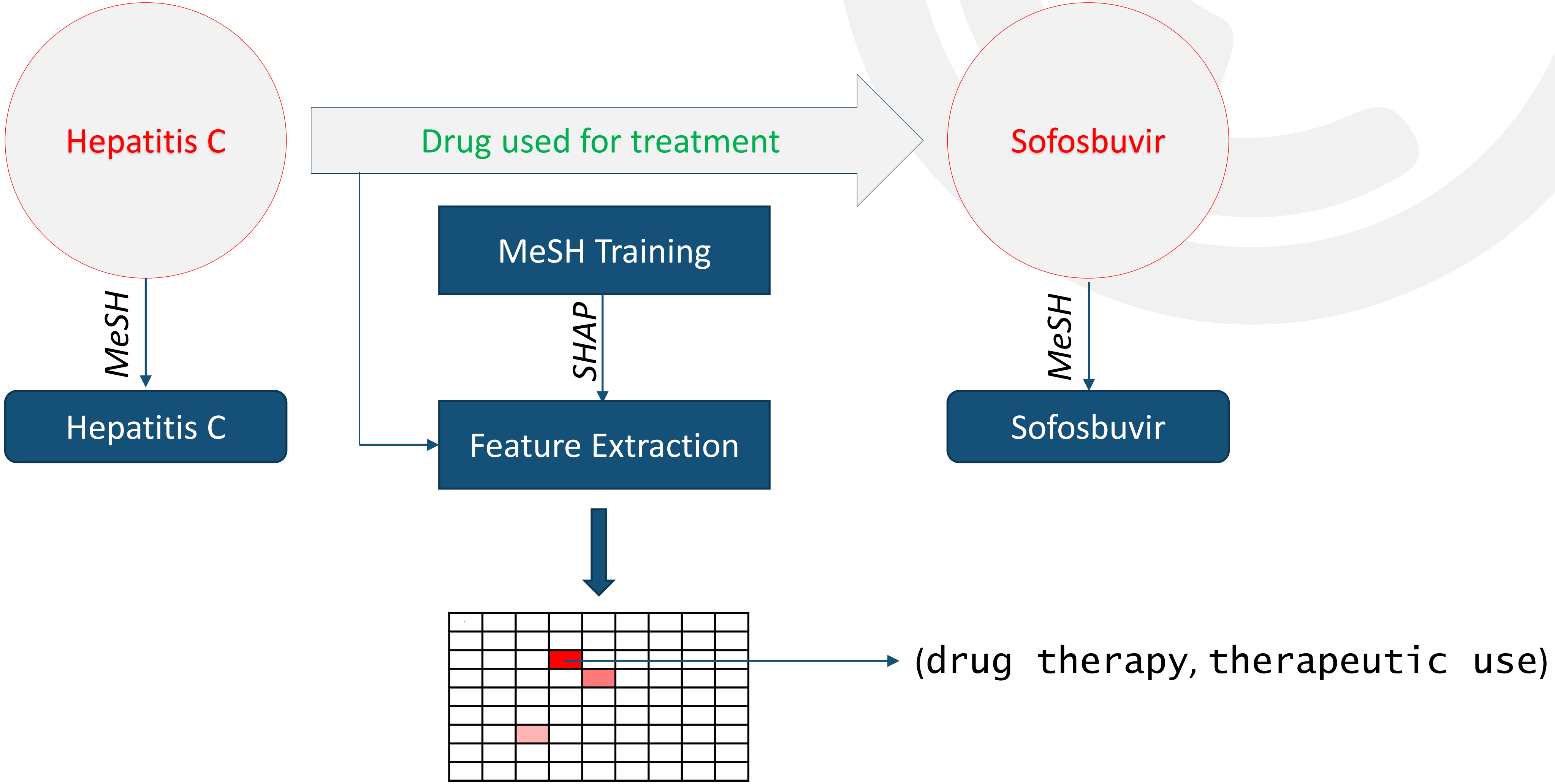
Identify the most relevant PubMed Central publications

Find the supporting sentence for claims

Align PMC ID with Wikidata ID of each reference

Add obtained references to Wikidata

PRINCIPLES



Hepatitis C

Drug used for treatment

Sofosbuvir

MeSH

MeSH Training

SHAP

MeSH

Hepatitis C

Feature Extraction

Sofosbuvir

(drug therapy, therapeutic use)

WIKI CRED

Data Availability

For reproducibility purposes, our source code and dataset are currently available at https://github.com/Data-Engineering-and-Semantics/refb

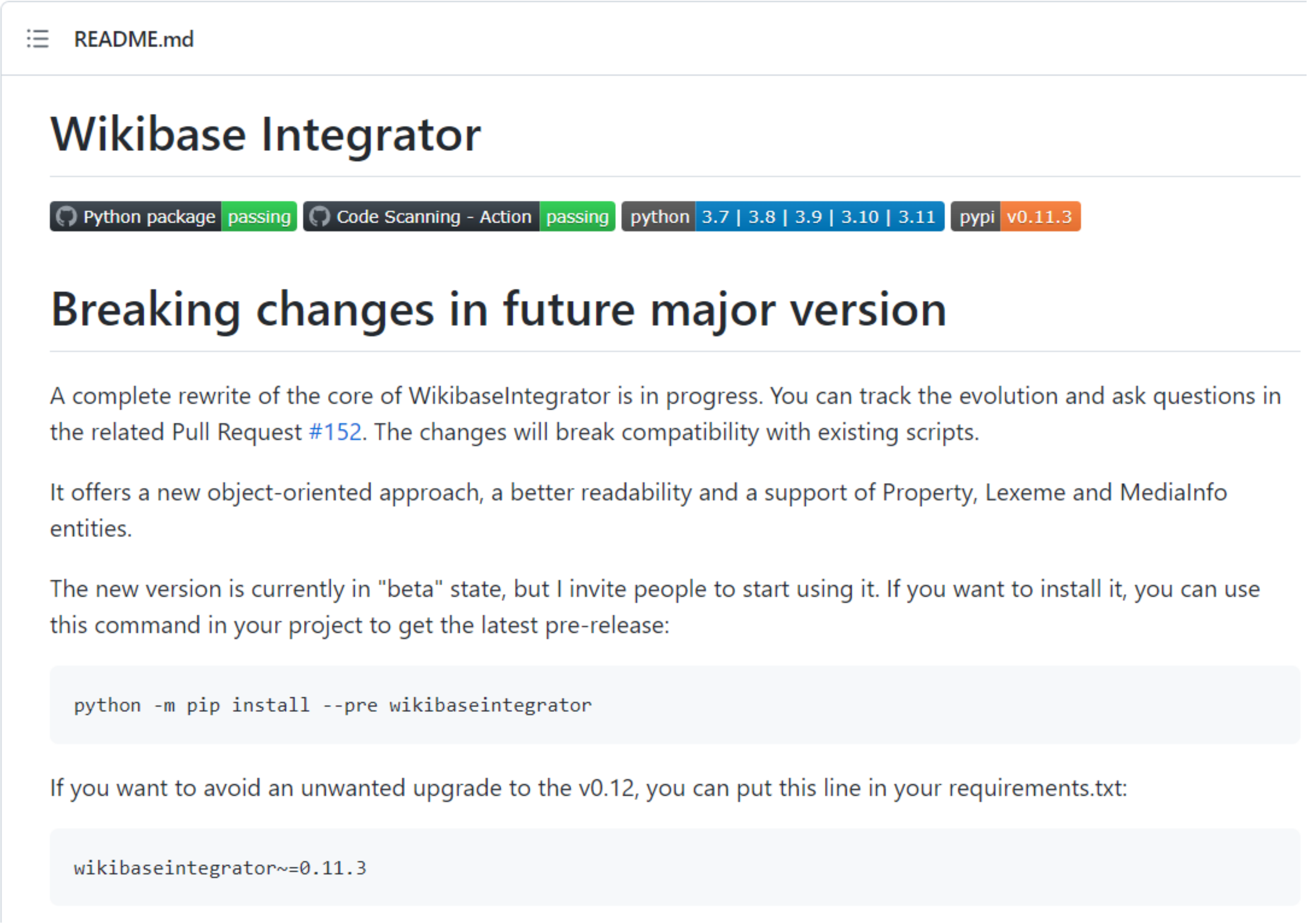
Data Engineering and Semantics
هندسة البيانات و دلالاتها

TOOLS

TOOLS FOR BOT CREATION

WIKIBASE INTEGRATOR

HTTPS://GITHUB.COM/LEMYST/WIKIBASEINTEGRATOR



Wikibase Integrator

Python package passing | Code Scanning - Action passing | python 3.7 | 3.8 | 3.9 | 3.10 | 3.11 | pypi v0.11.3

Breaking changes in future major version

A complete rewrite of the core of WikibaseIntegrator is in progress. You can track the evolution and ask questions in the related Pull Request #152. The changes will break compatibility with existing scripts.

It offers a new object-oriented approach, a better readability and a support of Property, Lexeme and MediaInfo entities.

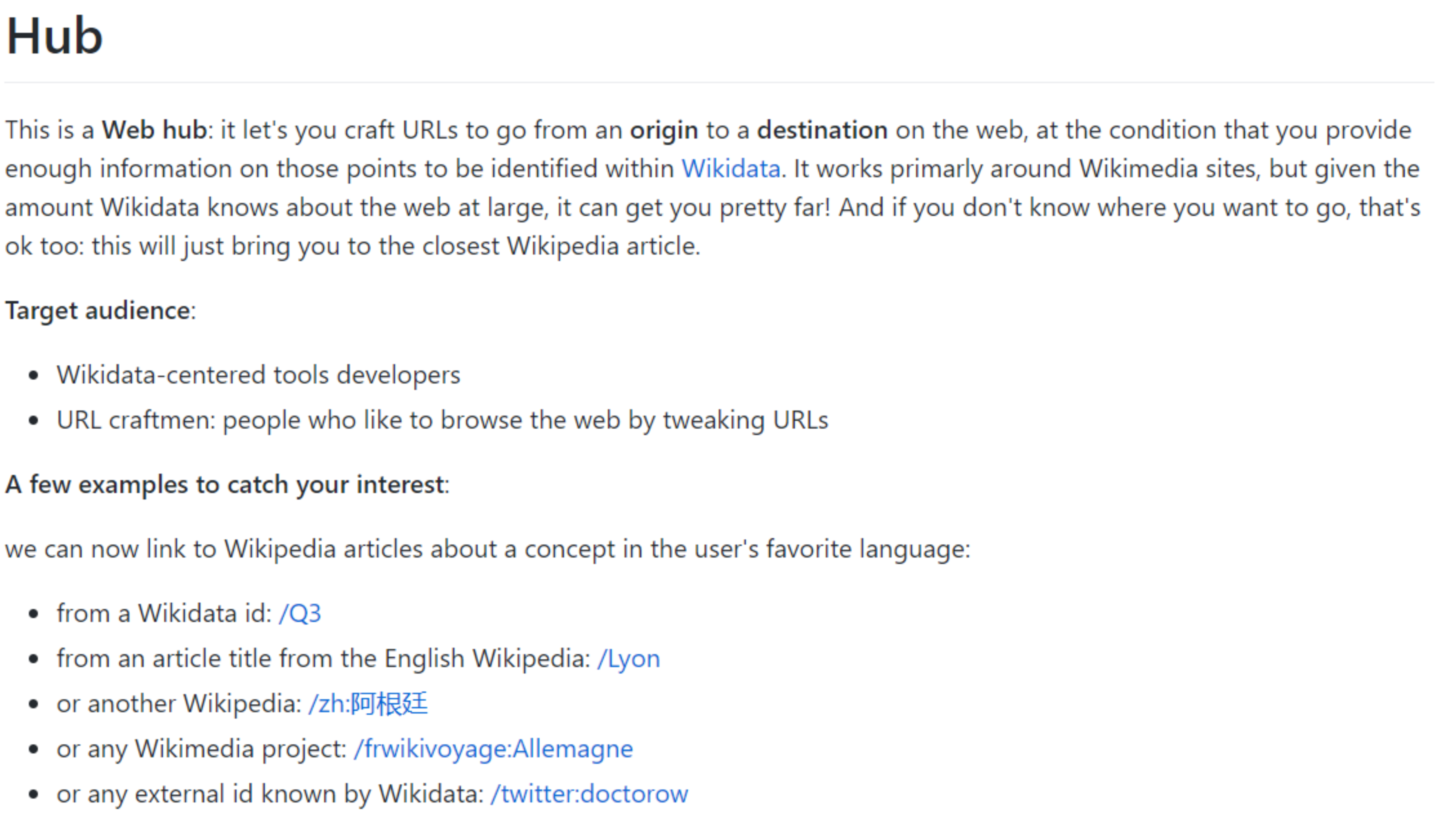
The new version is currently in "beta" state, but I invite people to start using it. If you want to install it, you can use this command in your project to get the latest pre-release:

```
python -m pip install --pre wikibaseintegrator
```

If you want to avoid an unwanted upgrade to the v0.12, you can put this line in your requirements.txt:

```
wikibaseintegrator~=0.11.3
```

WIKIDATA HUB
HTTPS://HUB.TOOLFORGE.ORG/



Hub

This is a **Web hub**: it let's you craft URLs to go from an **origin** to a **destination** on the web, at the condition that you provide enough information on those points to be identified within Wikidata. It works primarly around Wikimedia sites, but given the amount Wikidata knows about the web at large, it can get you pretty far! And if you don't know where you want to go, that's ok too: this will just bring you to the closest Wikipedia article.

Target audience:

- Wikidata-centered tools developers
- URL craftmen: people who like to browse the web by tweaking URLs

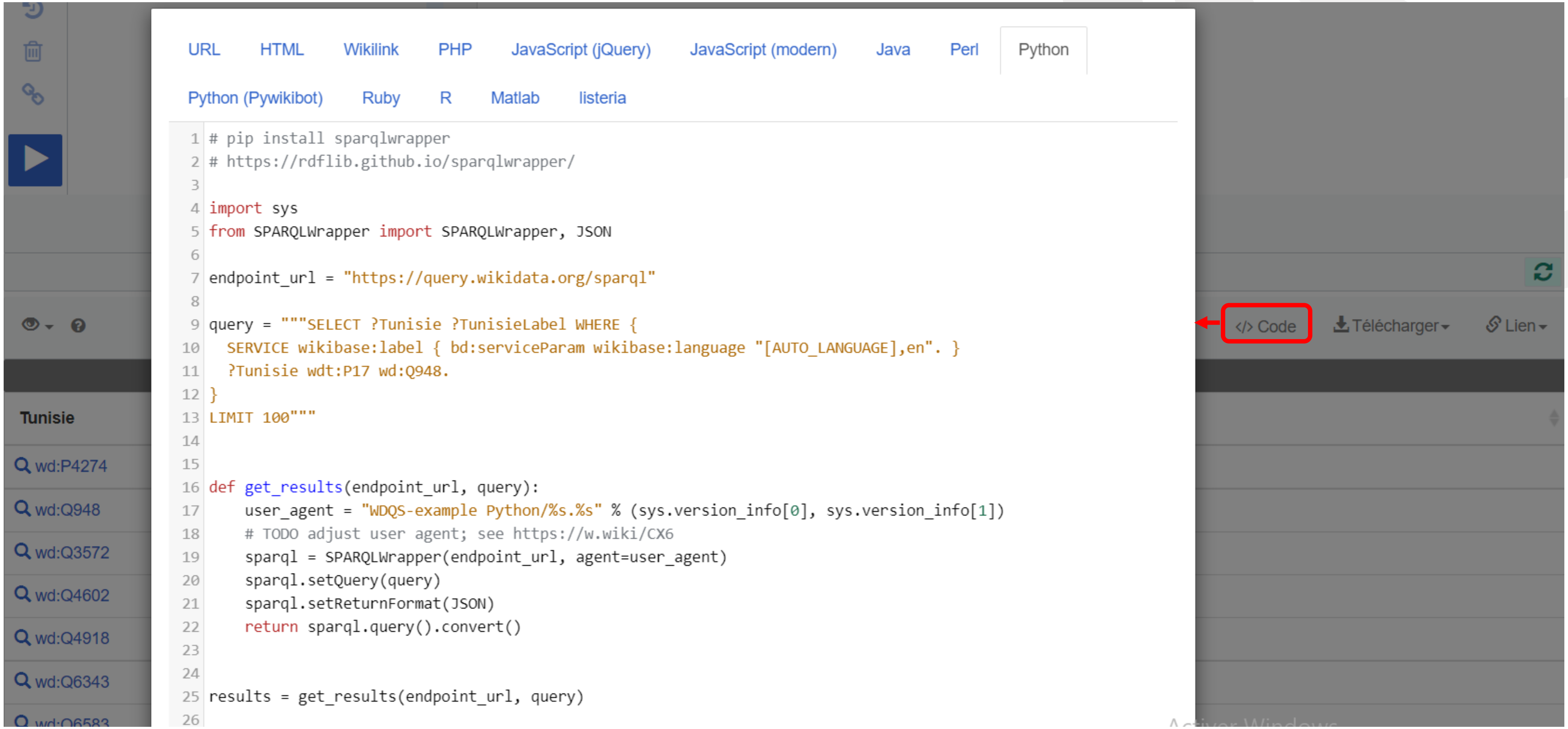
A few examples to catch your interest:

we can now link to Wikipedia articles about a concept in the user's favorite language:

- from a Wikidata id: /Q3
- from an article title from the English Wikipedia: /Lyon
- or another Wikipedia: /zh:阿根廷
- or any Wikimedia project: /frwikivoyage:Allemagne
- or any external id known by Wikidata: /twitter:doctorow

WIKIDATA QUERY SERVICE

HTTPS://QUERY.WIKIDATA.ORG/

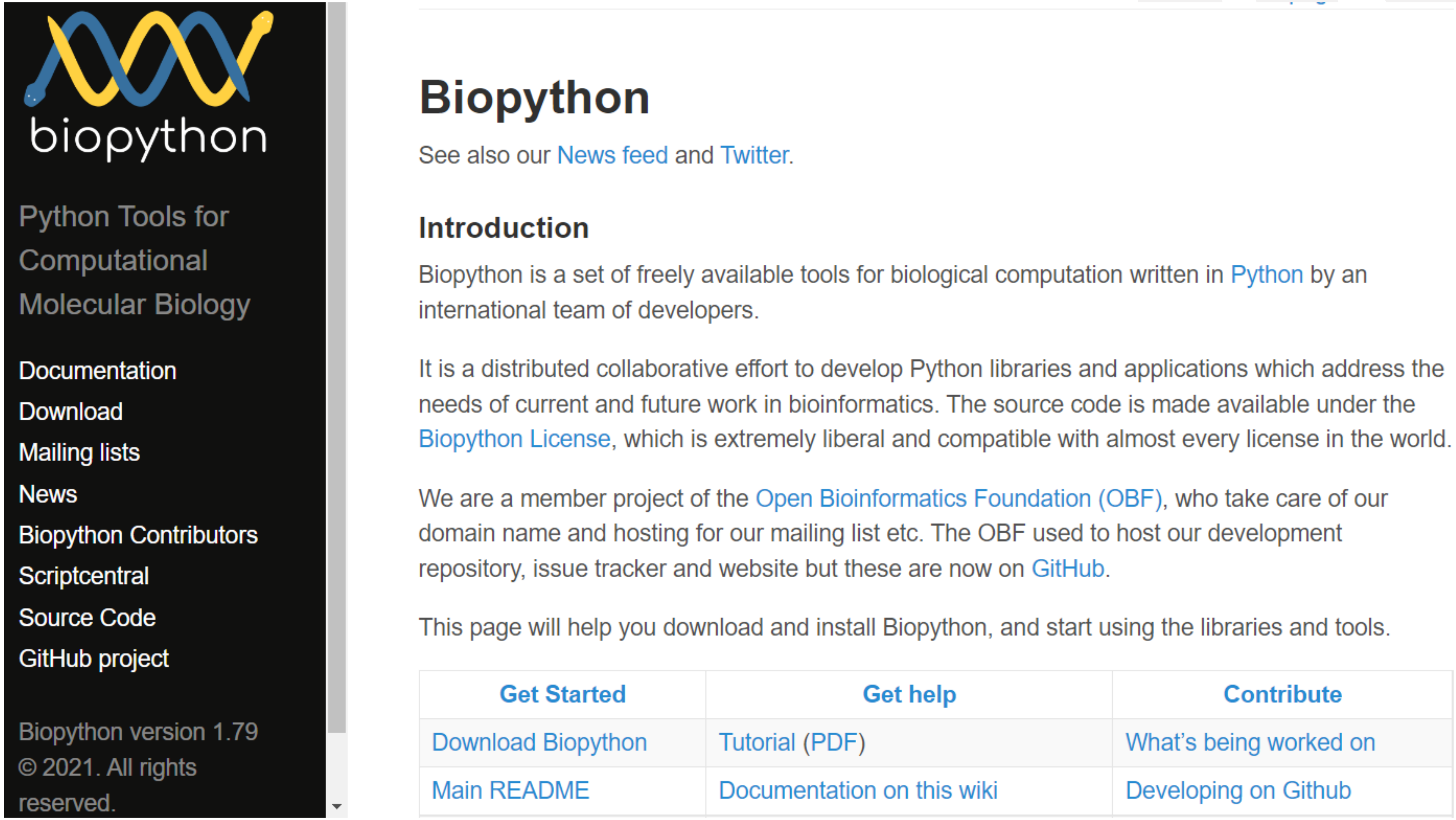


```python
# pip install sparqlwrapper
# https://rdflib.github.io/sparqlwrapper/

import sys
from SPARQLWrapper import SPARQLWrapper, JSON

endpoint_url = "https://query.wikidata.org/sparql"

query = """SELECT ?Tunisie ?TunisieLabel WHERE {
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
  ?Tunisie wdt:P17 wd:Q948.
}
LIMIT 100"""


def get_results(endpoint_url, query):
    user_agent = "WDQS-example Python/%s.%s" % (sys.version_info[0], sys.version_info[1])
    # TODO adjust user agent; see https://w.wiki/CX6
    sparql = SPARQLWrapper(endpoint_url, agent=user_agent)
    sparql.setQuery(query)
    sparql.setReturnFormat(JSON)
    return sparql.query().convert()


results = get_results(endpoint_url, query)
```

Tunisie
wd:P4274
wd:Q948
wd:Q3572
wd:Q4602
wd:Q4918
wd:Q6343
wd:Q6583

BIOPYTHON

HTTPS://BIOPYTHON.ORG/



biopython

Python Tools for Computational Molecular Biology

- Documentation
- Download
- Mailing lists
- News
- Biopython Contributors
- Scriptcentral
- Source Code
- GitHub project

Biopython version 1.79 © 2021. All rights reserved.

Biopython

See also our News feed and Twitter.

Introduction

Biopython is a set of freely available tools for biological computation written in Python by an international team of developers.

It is a distributed collaborative effort to develop Python libraries and applications which address the needs of current and future work in bioinformatics. The source code is made available under the Biopython License, which is extremely liberal and compatible with almost every license in the world.

We are a member project of the Open Bioinformatics Foundation (OBF), who take care of our domain name and hosting for our mailing list etc. The OBF used to host our development repository, issue tracker and website but these are now on GitHub.

This page will help you download and install Biopython, and start using the libraries and tools.

Get Started	Get help	Contribute
Download Biopython	Tutorial (PDF)	What's being worked on
Main README	Documentation on this wiki	Developing on Github

REFERENCES

- **Turki, H.**, **Shafee, T.**, **Hadj Taieb, M. A.**, **Ben Aouicha, M.**, Vrandečić, D., Das, D., & Hamdi, H. (2019). Wikidata: A large-scale collaborative ontological medical database. *Journal of Biomedical Informatics*, *99*, 103292. doi:10.1016/j.jbi.2019.103292.
- **Turki, H.**, **Hadj Taieb, M. A.**, **Shafee, T.**, Lubiana, T., Jemielniak, D., **Ben Aouicha, M.**, Labra Gayo, J. E., Youngstrom, E. A., Banat, M., Das, D., & **Mietchen, D.** (2022). Representing COVID-19 information in collaborative knowledge graphs: the case of Wikidata. *Semantic Web*, *13*(2), 233-264. doi:10.3233/SW-210444.
- **Turki, H.**, **Dossou, B. F. P.**, **Emezue, C. C.**, **Hadj Taieb, M. A.**, **Ben Aouicha, M.**, Ben Hassen, H., & Masmoudi, A. (2022). MeSH2Matrix: Machine learning-driven biomedical relation classification based on the MeSH keywords of PubMed scholarly publications. In *Proceedings of the 12th International Workshop on Bibliometric-enhanced Information Retrieval co-located with 44th European Conference on Information Retrieval (ECIR 2022)* (Forthcoming).
- **Turki, H.**, Jemielniak, D., **Hadj Taieb, M. A.**, Labra Gayo, J. E., **Ben Aouicha, M.**, Banat, M., **Shafee, T.**, Prud'Hommeaux, E., Lubiana, T., Das, D., & **Mietchen, D.** (2022). Using logical constraints to validate statistical information about COVID-19 in collaborative knowledge graphs: the case of Wikidata. *PeerJ Computer Science* (Forthcoming).
- National Institutes of Health (2019). *MEDLINE®/PubMed® Data Element (Field) Descriptions*. National Library of Medicine. https://www.nlm.nih.gov/bsd/mms/medlineelements.html.
- Church, K., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational linguistics*, *16*(1), 22-29.

CREDIT

- https://commons.wikimedia.org/wiki/File:SPARQL,_Be_Connected_to_Wikidata_-_Day_01_-_Wikidata_Presentation_02.jpg
- https://commons.wikimedia.org/wiki/File:Wikimedia_Foundation_logo_-_vertical_(2012-2016).svg
- https://commons.wikimedia.org/wiki/File:Grande_Mosqu%C3%A9e_de_Sfax_09.jpg
- https://commons.wikimedia.org/wiki/File:Mus%C3%A9e_de_Sfax.jpg
- https://www.sciencedirect.com/science/article/pii/S1532046419302114
- https://commons.wikimedia.org/wiki/Category:COVID-19_Study_of_Wikidata
- https://commons.wikimedia.org/wiki/File:13-11-02-olb-by-RalfR-03.jpg

Data Engineering and Semantics
هندسة البيانات و دلالاتها

THANK YOU

✉ TURKIABDELWAHEB@HOTMAIL.FR

HTTPS://DESLAB.ORG