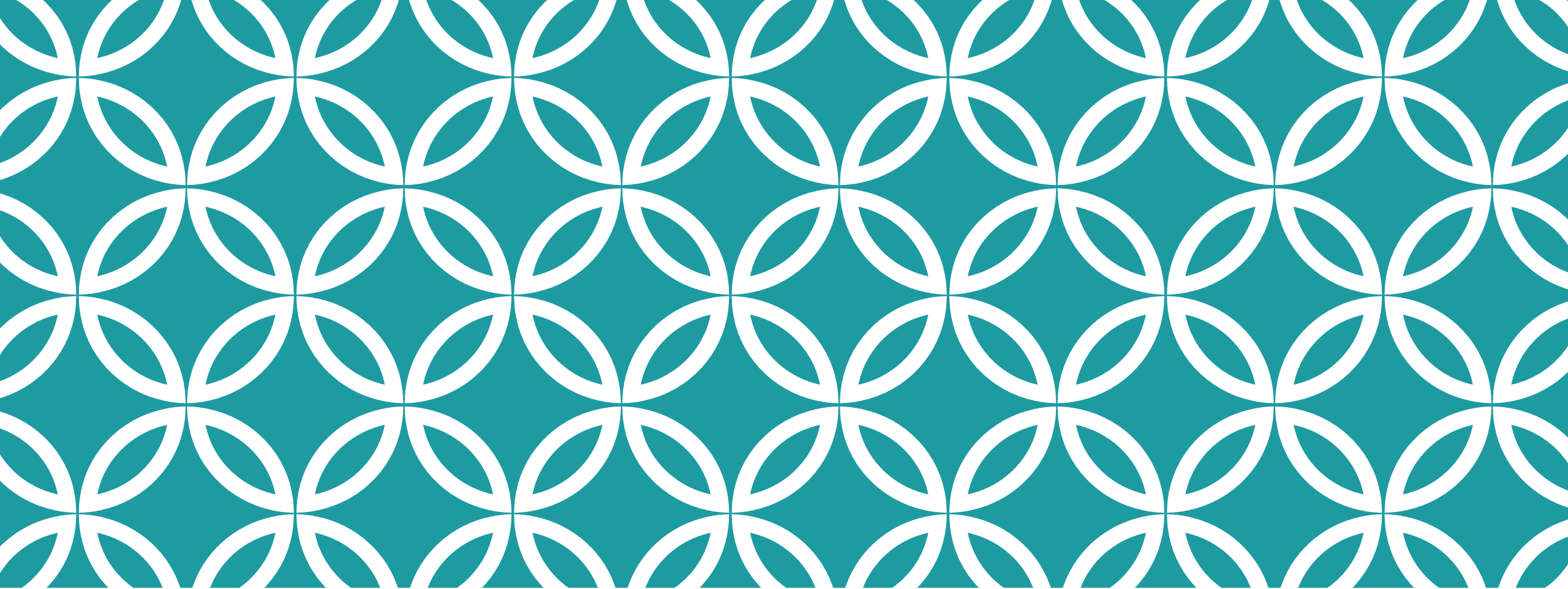
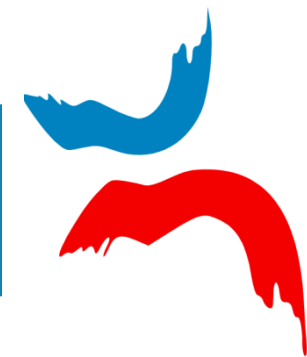


# WIKIDATA AND HEALTH: CURRENT SITUATION AND PERSPECTIVES

Houcemeddine Turki  
Faculty of Medicine of Sfax, University  
of Sfax, Sfax, Tunisia  
Mina Theofilatou  
University of Patras, Patras, Greece



**ABOUT US**



# ABOUT US



**Houcemeddine Turki**

Born: May 24, 1994 in  
Sfax, Tunisia



**Mina Theofilatou**

Born: March 7, 1967 in  
Montreal, Canada

## Houcemeddine Turki (User:Csisc)

Medical student, University of Sfax (Pre-Med 2018)

Wikimedian since 2009 (IP editing since 2007)

Active in Wikipedia, Wikidata, Wikimedia Incubator, Meta-wiki,  
and Wikimedia Commons

Member, Wiki Project Med, Wikimedia R&D Tunisia, Wikimedia  
and Libraries User Group Steering Committee, Wikilndaba  
Steering Committee

## Mina Theofilatou (User:Saintfevrier)

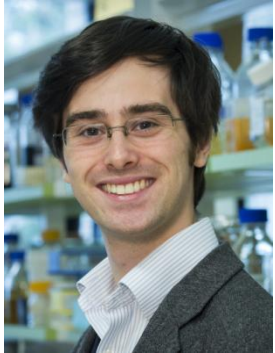
Computer Engineer, University of Patras (M.Sc. Eng. 1991);  
Computer Science Teacher, Argostoli Evening High School

Member, Wiki Project Med, Wikipedia Education Program

Wikimedian since 2007

Active in Wikipedia, Wikidata, Meta-wiki and Wikimedia  
Commons

# TEAM MEMBERS



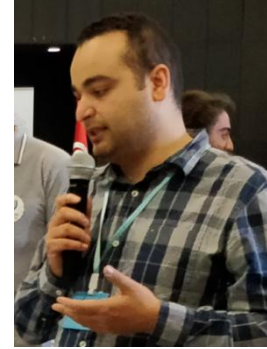
Thomas Shafee  
La Trobe University



Daniel Mietchen  
University of Virginia



Diptanshu Das  
Institute of Neurosciences, Kolkata



Helmi Hamdi  
Université de Sherbrooke



Mohamed Ali Hadj Taieb  
University of Sfax



Mohamed Ben Aouicha  
University of Sfax



## Other teams

Benjamin M. Good, Andrew I. Su, Elvira Mitraga, Sebastian Burgstaller-Muehlbacher, Andra Waagmeester, Sebastien Lelong et al. Genomic information on Wikidata ; Microbial information on Wikidata

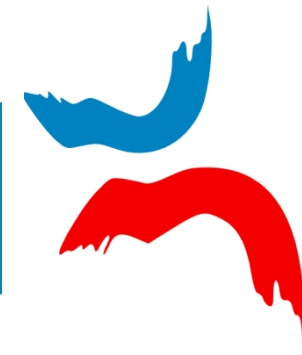
Alexander Pfunder, Tobias Schönberg, John Horn, Richard D. Boyce, Matthias Samwald, Şerkan Ayzaz et al. Drug information of Wikidata ; Wikidata for Medical Wikipedia

## TO CITE THIS PART

Vrandečić, D., & Krötzsch, M. (2014). Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57 (10), 78-85.

Lih, A. (2016). *Introduction to Wikidata: Structuring Wikipedia and Beyond for (G)LAM*. Washington DC: Smithsonian Institution.

# WIKIDATA IN BRIEF



# PRINCIPLES

A structured representation of the sum of all human knowledge

Facts are represented as statements in Wikidata. These statements are in form of triples:

- Subject – Predicate – Object
- Item – Property – Value
- Thing – Relationship – Thing



# WIKIDATA TRIPLES: ITEM — PROPERTY - VALUE

## Item

Identified by an automatically assigned Q-number

Can be created by anyone

Often corresponds to a Wikipedia article or concept

Examples: Human (Q5), Disease (Q12136)

## Property

Identified by an automatically assigned P-number

Controlled vocabulary for consistency

Proposal, discussion and approval process

For example, see: [Wikidata:List\\_of\\_properties](#)

## Value

A value can be an item, a string, a URL, a time, a period, a location or a quantity

# A PLATFORM FOR BOTH HUMANS AND MACHINES

## Human-friendly

User interface

([https://www.wikidata.org/wiki/{entity\\_id}](https://www.wikidata.org/wiki/{entity_id}))

Queryable using Wikidata query service

(<https://query.wikidata.org>)

## Machine-friendly

Machine-readable interface available (e.g. RDF and JSON)

([https://www.wikidata.org/wiki/Special:EntityData/{entity\\_id}.{format}](https://www.wikidata.org/wiki/Special:EntityData/{entity_id}.{format}))

API service

(<https://www.wikidata.org/w/api.php>)

Downloadable as XML, JSON or RDF dumps

([https://www.wikidata.org/wiki/Wikidata:Database\\_download](https://www.wikidata.org/wiki/Wikidata:Database_download))

CC-0 Licence (Can be used, processed and integrated in other projects without any legal issue)

Automated and semi-automated methods of mass editing for Wikidata



## TO CITE THIS PART

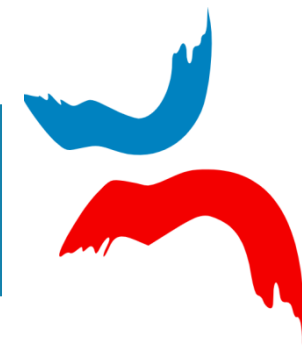
Turki, H., Shafee, T., Hadj Taieb, M. A., Ben Aouicha, M., Vrandečić, D., Das, D., & Hamdi, H. (2019). Wikidata: A large-scale collaborative ontological medical database. *Journal of Biomedical Informatics* (in press).

Brasileiro, F., Almeida, J. P. A., Carvalho, V. A., & Guizzardi, G. (2016, April). Applying a multi-level modeling theory to assess taxonomic hierarchies in Wikidata. In *Proceedings of the 25th International Conference Companion on World Wide Web* (pp. 975-980). International World Wide Web Conferences Steering Committee.

Diefenbach, D., Migliatti, P. H., Qawasmeh, O., Lully, V., Singh, K., & Maret, P. (2019, May). QAnswer: A Question Answering prototype bridging the gap between a considerable part of the LOD cloud and end-users. In *The World Wide Web Conference* (pp. 3507-3510). ACM.

Nielsen, F. Å., Mietchen, D., & Willighagen, E. (2017, May). Scholia, Scientometrics and Wikidata. In *European Semantic Web Conference* (pp. 237-259). Springer.

# MEDICAL INFORMATION ON WIKIDATA

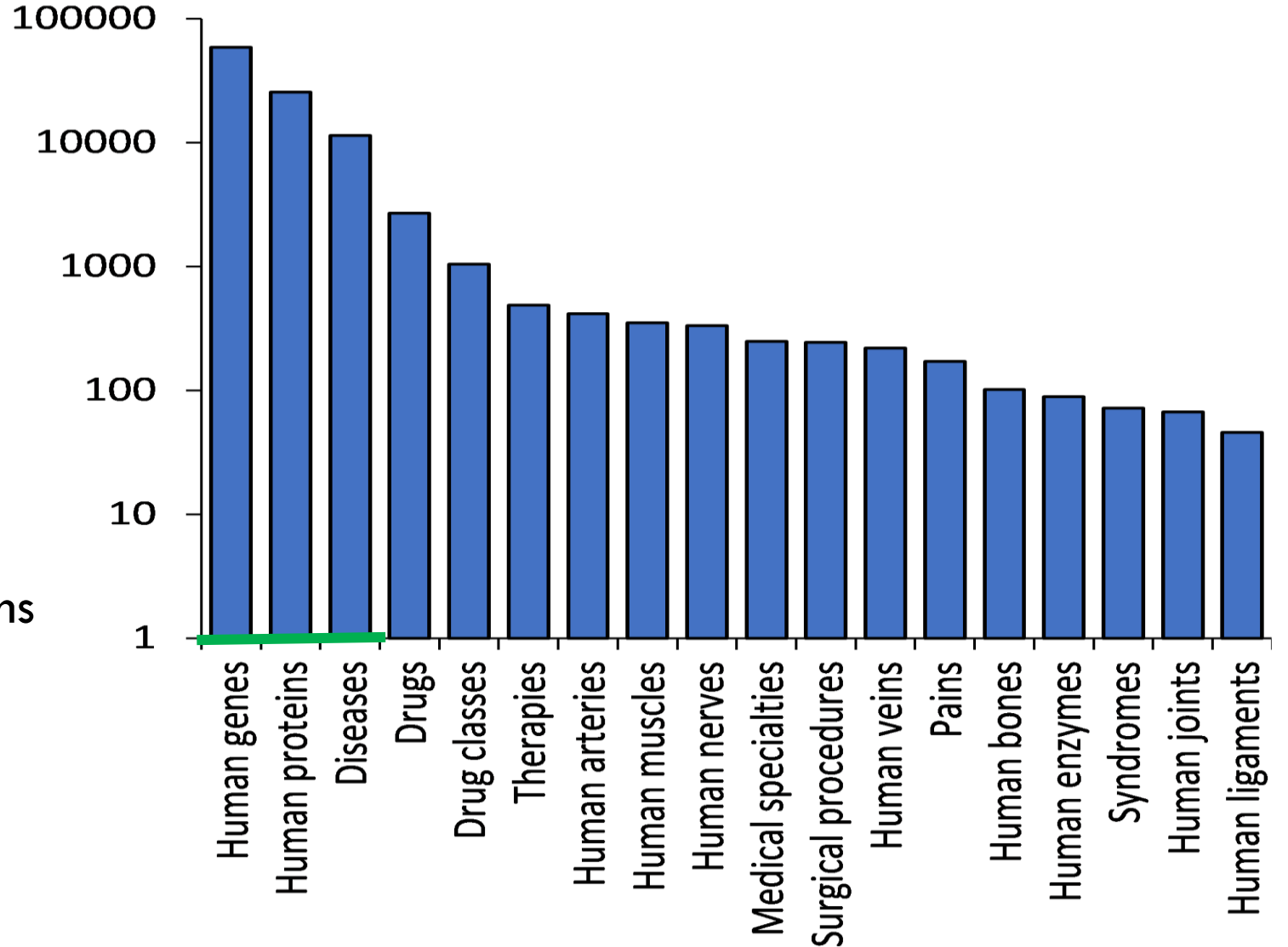


# BIOMEDICAL ITEMS PER TYPE

Significant representation of human genes, human proteins and drugs

Lack of coverage of other types of biomedical items

Several types of clinical items are missing: classifications (E.g. Gharbi classification), formulas, medical signs (E.g. back pain), valves, anatomical parts (E.g. Iliac crest), tests and diagnosis methods (E.g. Valsalva maneuver, Köplik's spots)

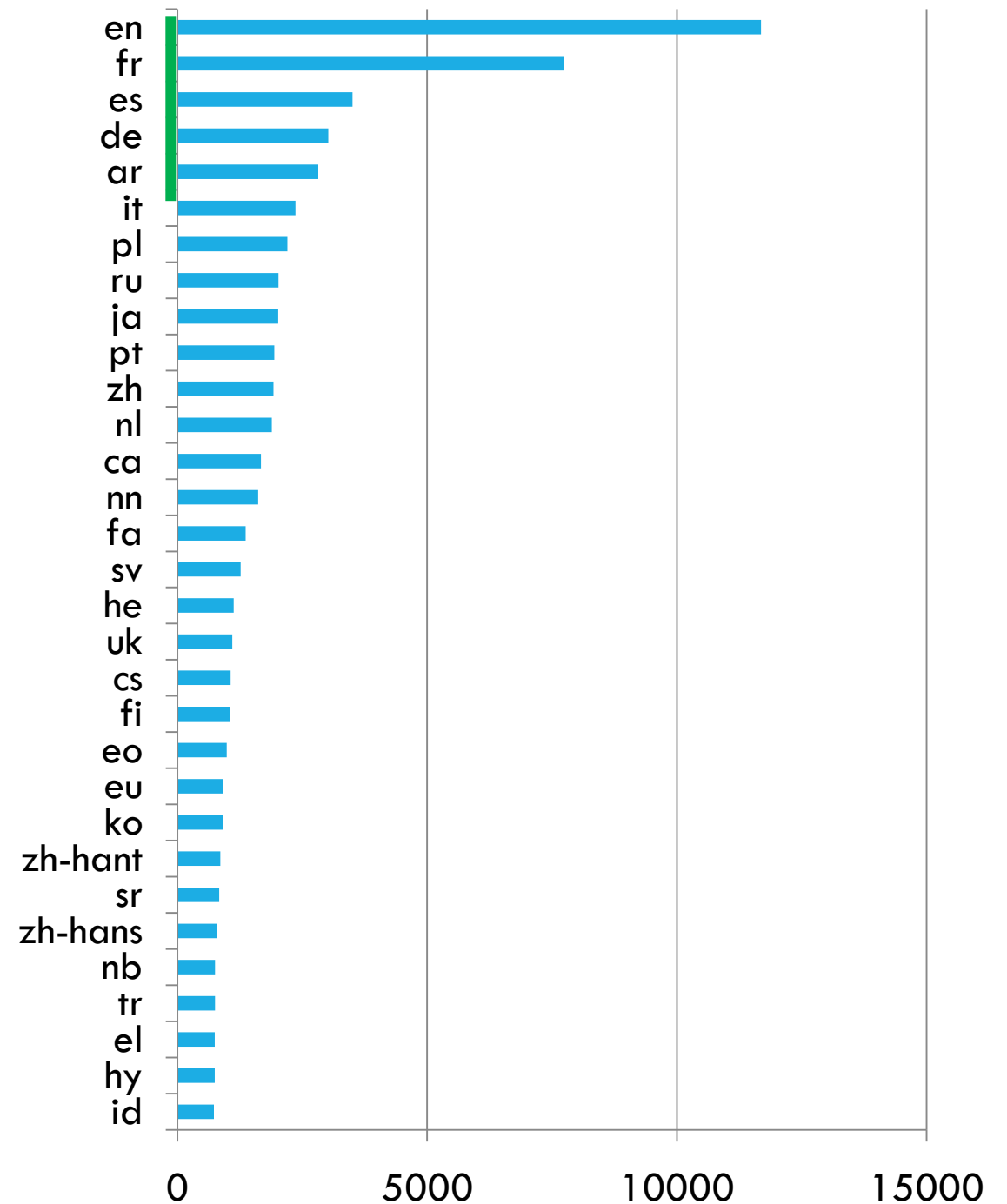


# LANGUAGE SUPPORT

## NUMBER OF LABELS FOR DISEASES IN WIKIDATA

Excellent representation of labels only in English and French (to lesser extent, also Spanish, German, and Arabic)

Efforts needed to ameliorate language support for biomedical data in Wikidata



# LANGUAGE SUPPORT

## SOURCES FOR LABELS OF BIOMEDICAL ITEMS

Wikipedia or Wiktionary

Online and Offline Lexicons in Local Languages

Massive Open Online Courses and Medical Textbooks in Local Languages

Medical Experts

## LIMITATIONS

Surgical nomenclature vs. Anatomical nomenclature

Regular changes in the names of medical items

# MEDICAL WIKIDATA PROPERTIES

Wikidata currently includes hundreds of Medical properties

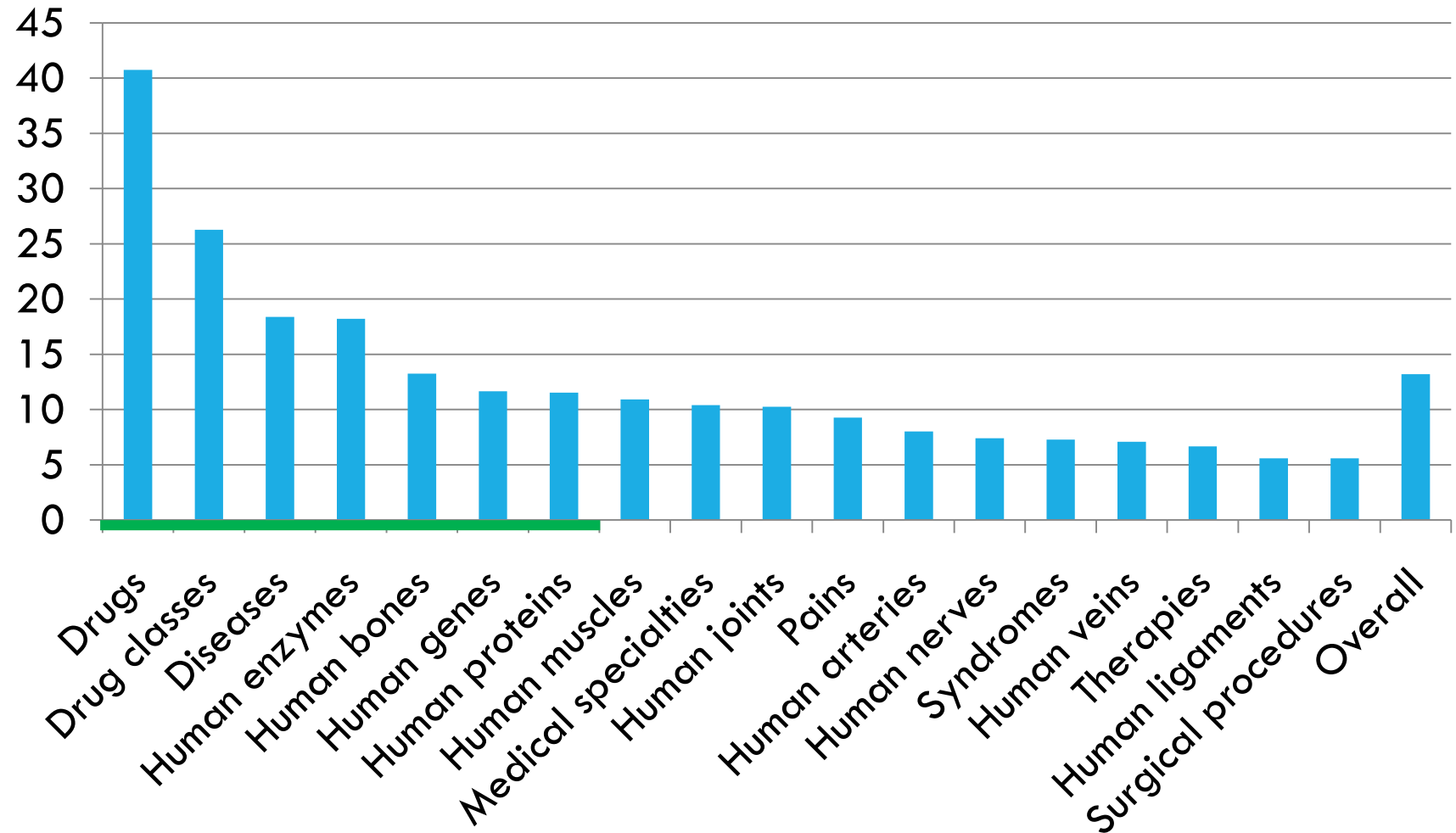
These properties have various purposes including:

- Definition of anatomical relations between Wikidata items
- Specification of pharmacological data of a drug
- Specification of epidemiological and clinical data of a disease
- Specification of human genomic and proteomic information
- Matching between Wikidata items and biomedical entries in external databases

# STATEMENTS PER BIOMEDICAL ITEM

Significant representation of Wikidata statements related to drugs, human diseases, human enzymes, human bones and human proteins

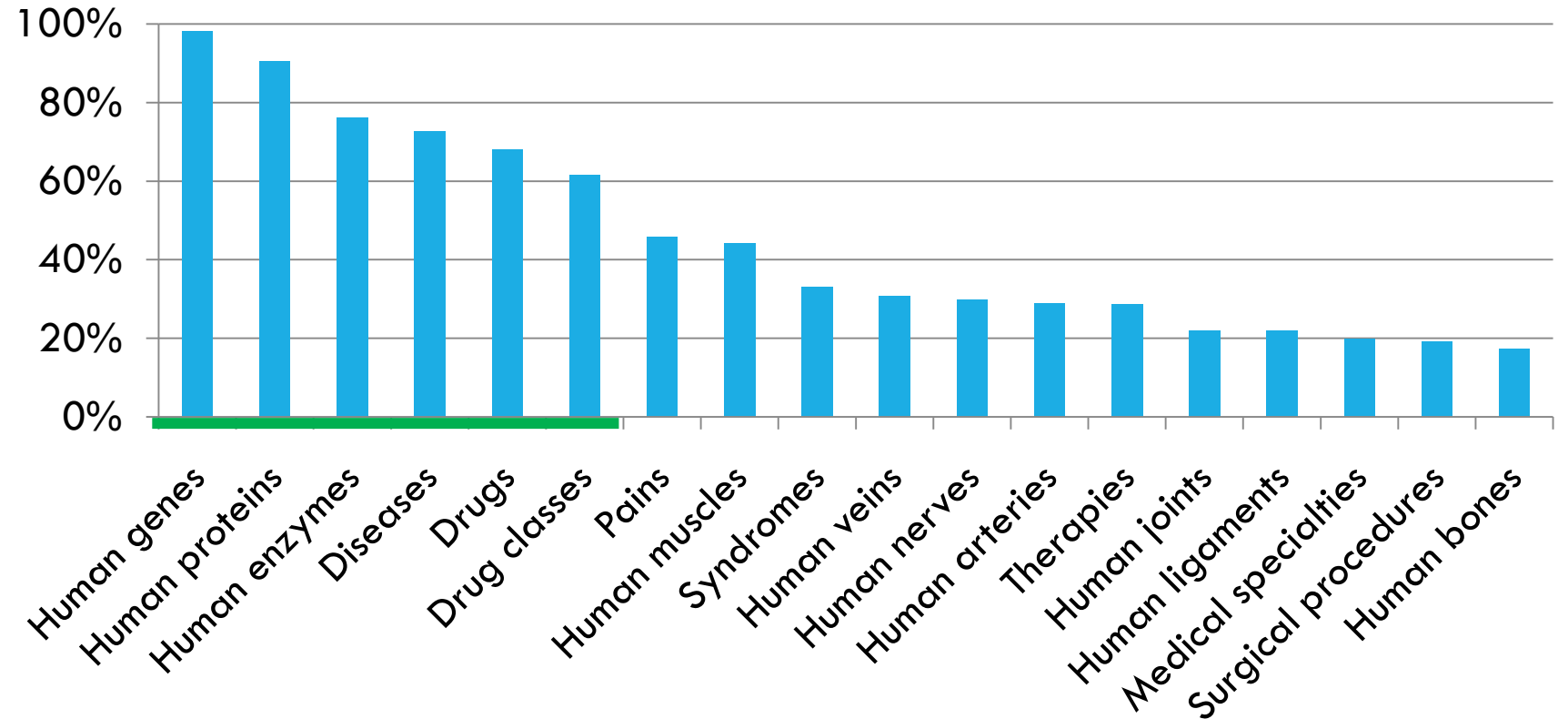
Lack of coverage of Wikidata statements related to other types of biomedical items



# REFERENCE SUPPORT FOR BIOMEDICAL ITEMS

Excellent reference support for Wikidata statements related to human genes, human proteins, human enzymes, diseases and drugs

Limited reference support for Wikidata statements related to other types of biomedical items





# WIKIDATA BIOMEDICAL USAGE EXAMPLES

## 1) CLINICAL DECISION SUPPORT

Biomedical information seeking (Direct search, Wikidata query service)

Identification of drug-drug interactions within prescriptions

Diagnosis of diseases based on Electronic Health Records

Determination of the prognosis of diseases

Development of Wikidata and Social Media-based Methods for the surveillance of epidemic diseases

Automation of Systematic Reviews and Meta-analyses

# WIKIDATA BIOMEDICAL USAGE EXAMPLES

## 2) NATURAL LANGUAGE QUESTIONS

[HTTPS://QANSWER-FRONTEND.UNIV-ST-ETIENNE.FR](https://qanswer-frontend.univ-st-etienne.fr)

QAnswer   [About](#) [FAQ](#)

Confidence : 70 %

[SPARQL LIST](#)

[DID YOU MEAN](#)

[DIRECT ANSWER](#)

Is this the right answer ?  Yes  No

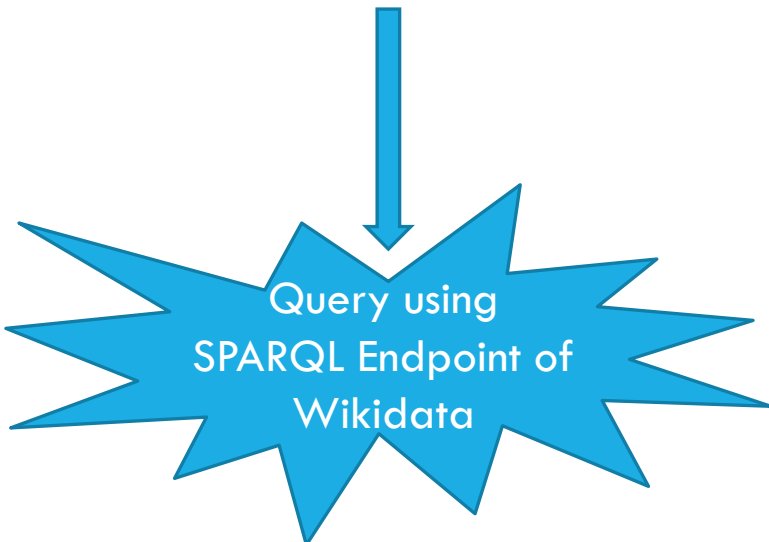
*instance of / medication (substance used to diagnose,...)  
/ medical condition treated / leishmaniasis (disease caused by parasites...)*

**pentamidine** +  
W

**Itraconazole** +  
W

**Ketoconazole** +  
W

Natural language questions are converted to machine-readable triples based on a dataset of sample questions



# WIKIDATA BIOMEDICAL USAGE EXAMPLES

## 3) STRUCTURED ABSTRACTS OF BIOMED RESEARCH PUBLICATIONS

[HTTPS://TOOLS.WMFLABS.ORG/SCHOLIA/WORK/Q48672086](https://tools.wmflabs.org/scholia/work/Q48672086)

### Supports the following statement(s)

Statements in Wikidata supported by references to this work. Only a maximum of around 2000 statements are shown.

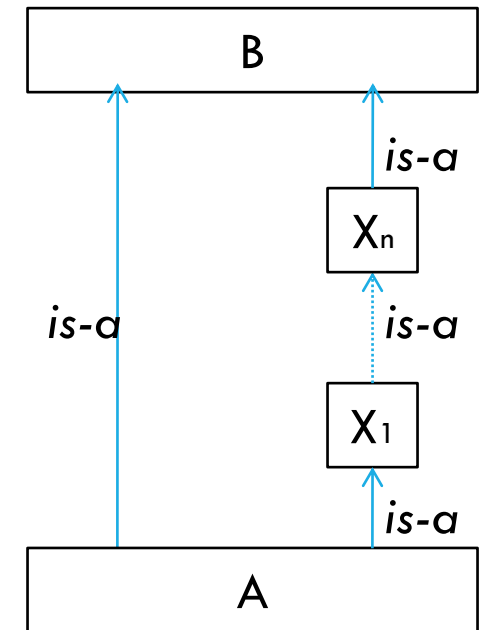
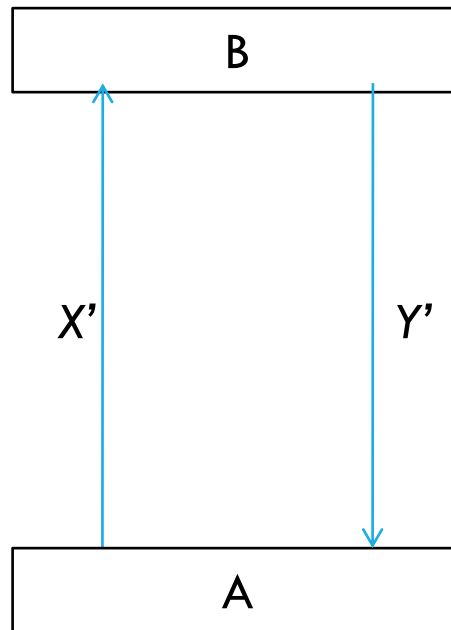
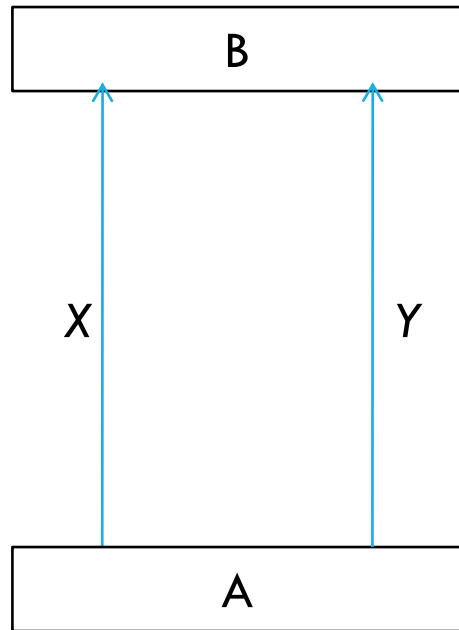
Show  entries

Search:

Item	Property	Value
<a href="#">smoking</a>	<a href="#">risk factor</a>	<a href="#">substance abuse</a>

[Edit on query.Wikidata.org](#)

# COMMON DEFICIENCIES IN WIKIDATA PROPERTIES



A and B are biomedical items  
X, Y, X' and Y' are unrelated Wikidata properties  
*is-a* is instance of (P31) or subclass of (P279)

# OTHER DEFICIENCIES IN BIOMEDICAL USAGE CASES

Wikidata does not take into consideration the chronological evolution of the symptoms of diseases.

- Acute appendicitis is characterized by an epigastric pain in early stages and by a right lower quadrant pain in late stages.

Wikidata does not give weights to biomedical associations according to importance scale.

- A pathognomonic sign of a disease should not have the same weight as a non-specific sign of that disease. As well, gold standards to cure or identify diseases should not be considered the same as other treatments or diagnosis methods.

Wikidata does not include radiological signs and ECG and EEG biomarkers for diseases

- S1Q3 is an ECG biomarker for pulmonary embolism

Wikidata does not assess the quality to the references used to support biomedical statements

- Recent clinical trials, systematic reviews and meta-analyses should be considered as better than other types of research publications in supporting biomedical Wikidata statements

## TO CITE THIS PART

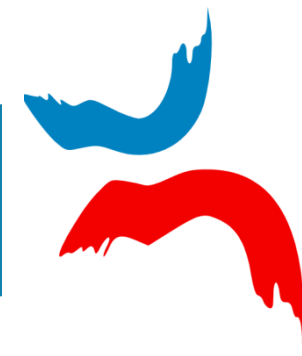
Krötzsch, M., Simancik, F., & Horrocks, I. (2014). Description logics. *IEEE Intelligent Systems*, 29(1), 12-19.

Zhang, Y., Lin, H., Yang, Z., Wang, J., Zhang, S., Sun, Y., & Yang, L. (2018). A hybrid model based on neural networks for biomedical relation extraction. *Journal of Biomedical Informatics*, 81, 83-92.

Turki, H., Hadj Taieb, M. A., & Ben Aouicha, M. MeSH qualifiers, publication types and relation occurrence frequency are also useful for a better sentence-level extraction of biomedical relations. *Journal of Biomedical Informatics*, 83, 217-218.

Koopman, B., Russell, J., & Zuccon, G. (2018). Task-oriented search for evidence-based medicine. *International Journal on Digital Libraries*, 19(2-3), 217-229.

**ENHANCING (BIOMEDICAL)  
INFORMATION PROVIDED BY WIKIDATA**



# REQUIRED TASKS

Adding Description Logics to Wikidata properties to eliminate the organizational deficiencies of the information provided by Wikidata

Developing Bibliometric-Enhanced Information Retrieval-based tools to verify, enrich and add reference support to Wikidata statements



# ADDING DESCRIPTION LOGICS TO WIKIDATA PROPERTIES

The definition of description logics for properties contributes to the elimination of inconsistencies within Wikidata information

E.g. if  $A \xrightarrow{x} B$  then  $B \xrightarrow{\cancel{x}} A$ .

Adding OWL2 support to Wikidata properties can solve this matter. These description logics can be used later to create tools that send notifications to admins when critical inconsistencies within Wikidata information exist

# BIBLIOMETRIC-ENHANCED INFORMATION RETRIEVAL TO VERIFY, ENRICH AND ADD REFERENCE SUPPORT TO WIKIDATA STATEMENTS

As Wikidata is meant to be a structured representation of the sum of all human knowledge, the use of citation indexes (WoS, Scopus, Medline...) to enrich, verify and add reference support to it makes sense.

In fact, citation indexes involves the metadata of millions of scientific publications and are consequently a trusted representation of the sum of all human knowledge.

# METADATA OF SCIENTIFIC PUBLICATIONS THAT CAN BE USED TO ENRICH WIKIDATA

Title

Abstract

Keywords (such as MeSH headings)

Identifiers (such as PubMed ID)

Publication types

Publication Year

Keyword co-occurrence frequency

# EXAMPLE

## ADDING REFERENCE SUPPORT TO BIOMEDICAL STATEMENTS

Search Pubmed with query  
(using API, results returned as XML)



Check number of results



Get PubMed ID of most relevant result



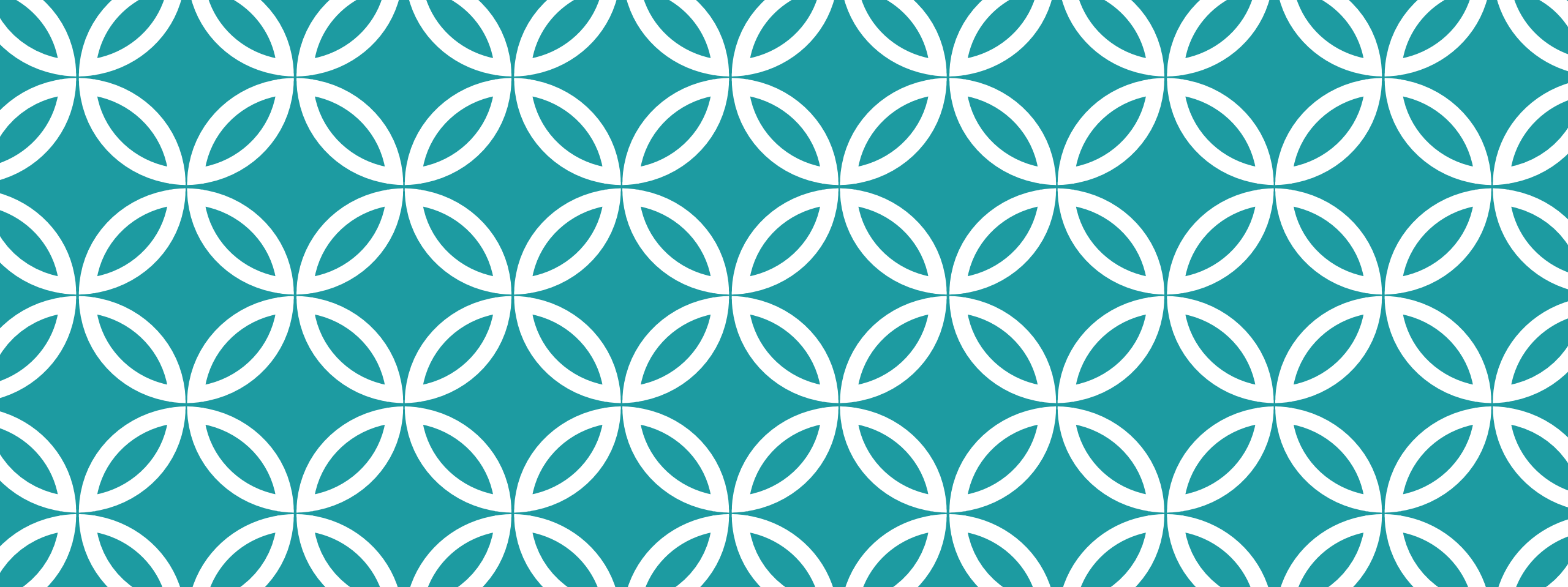
Suggest as reference for Wikidata statement

# CHALLENGES

Finding the best PubMed query that can return required results with the best precision (Choosing filters...)

Define an excellent method for the automatic verification of relations (IC-based measures...)

Decide the guidelines for the verification of biomedical relations by scientists



Houcemeddine Turki  
User:Csisc  
turkiabdelwaheb@hotmail.fr

Mina Theofilatou  
User:Saintfevrier  
saintfevrier@gmail.com

**THANK YOU**

Questions

