WIKI DATA CON

Systematic Review Automation driven by Wikidata

Data Engineering and Semantics
هندسة البيانات و دلالاتها

جامعة صفاقس
University of Sfax

Houcemeddine Turki
Medical Student, Faculty of Medicine of Sfax
Research Assistant, Data Engineering and Semantics Research Unit
Member, Wiki Indaba Steering Committee
Vice-Chair, Wikimedia Tunisia User Group
Board Member, Wikimedia and Libraries User Group

SISONKE-BIOTIK

Supported by SisonkeBiotik Community
Biomedical Machine Learning Community in Africa

About Us

Data Engineering and Semantics
Research Unit

University of Sfax

Located in Tunisia, North Africa (270 km from Tunis)

Major University in Tunisia

Among the best universities in Africa in Computer Science Research

Data Engineering and Semantics Research Unit

- Recently Created as a part of University of Sfax
- Deals with various aspects of Computer Science ranging from Semantic Technologies, Social Network Analysis and Graph Embeddings to Scientometrics
- Members having publications in highly referred scholarly journals
- Dealing with Wikimedia Projects as resources for driving real-life applications at a low cost

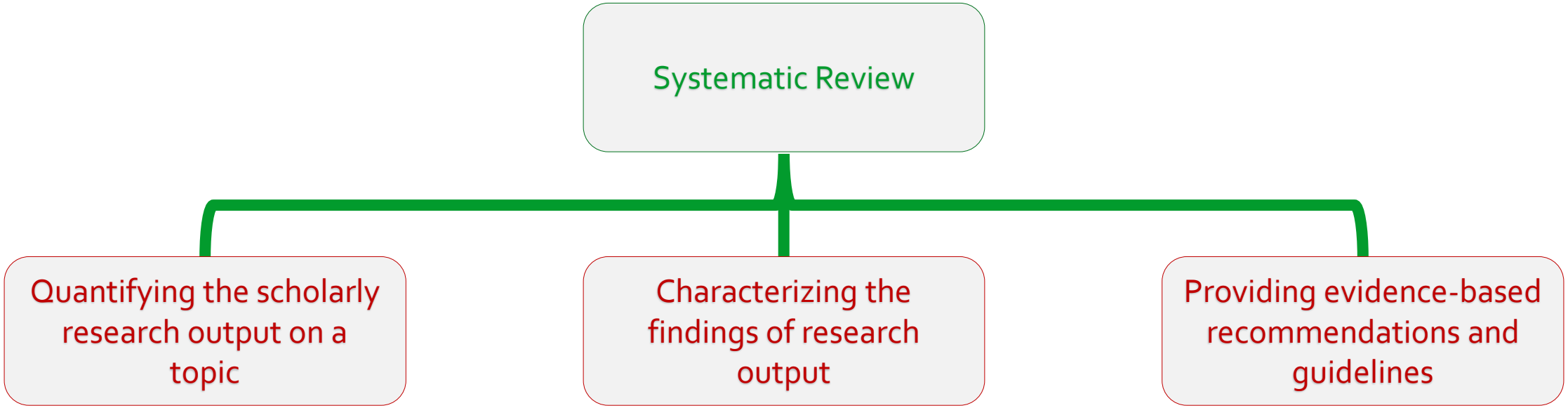
Sisonkebiotik

- Recently created research community aiming to develop healthcare machine learning approaches in Africa
- Hosted in South Africa, Having members across the African Continent
- Follows the Masakhane Community Governance Model
- Currently working on a bibliometric study and a systematic review on machine learning and healthcare in Africa

Introduction

What’s a Systematic Review

Systematic Review



Many guidelines for developing systematic reviews

Series: Living Systematic Reviews

Review article ○ Abstract only
Living systematic review: 1. Introduction—the why, what, when, and how
Julian H. Elliott, Anneliese Synnot, Tari Turner, Mark Simmonds, ... Leslea Pearson
Pages 23-30
Purchase PDF Article preview

Review article ● *Open access*
Living systematic reviews: 2. Combining human and machine effort
James Thomas, Anna Noel-Storr, Iain Marshall, Byron Wallace, ... Leslea Pearson
Pages 31-37
Download PDF Article preview

Review article ○ Abstract only
Living systematic reviews: 3. Statistical methods for updating meta-analyses
Mark Simmonds, Georgia Salanti, Joanne McKenzie, Julian Elliott, ... Leslea Pearson
Pages 38-46
Purchase PDF Article preview

Review article ○ Abstract only
Living systematic reviews: 4. Living guideline recommendations
Elie A. Akl, Joerg J. Meerpohl, Julian Elliott, Lara A. Kahale, ... Leslea Pearson
Pages 47-53
Purchase PDF Article preview

Turk Arch Otorhinolaryngol. 2019 Mar; 57(1): 57–58.
Published online 2019 Mar 14. doi: 10.5152/tao.2019.4058
PMCID: PMC6461330
PMID: 31049257

A Guide for Systematic Reviews: PRISMA

Ayşe Adin Selçuk

Research | Open Access | Published: 01 January 2015

Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement

David Moher, Larissa Shamseer, Mike Clarke, Davina Ghersi, Alessandro Liberati, Mark Petticrew, Paul Shekelle, Lesley A Stewart & PRISMA-P Group

Systematic Reviews **4**, Article number: 1 (2015) | Cite this article

Guidance on Conducting a Systematic Literature Review

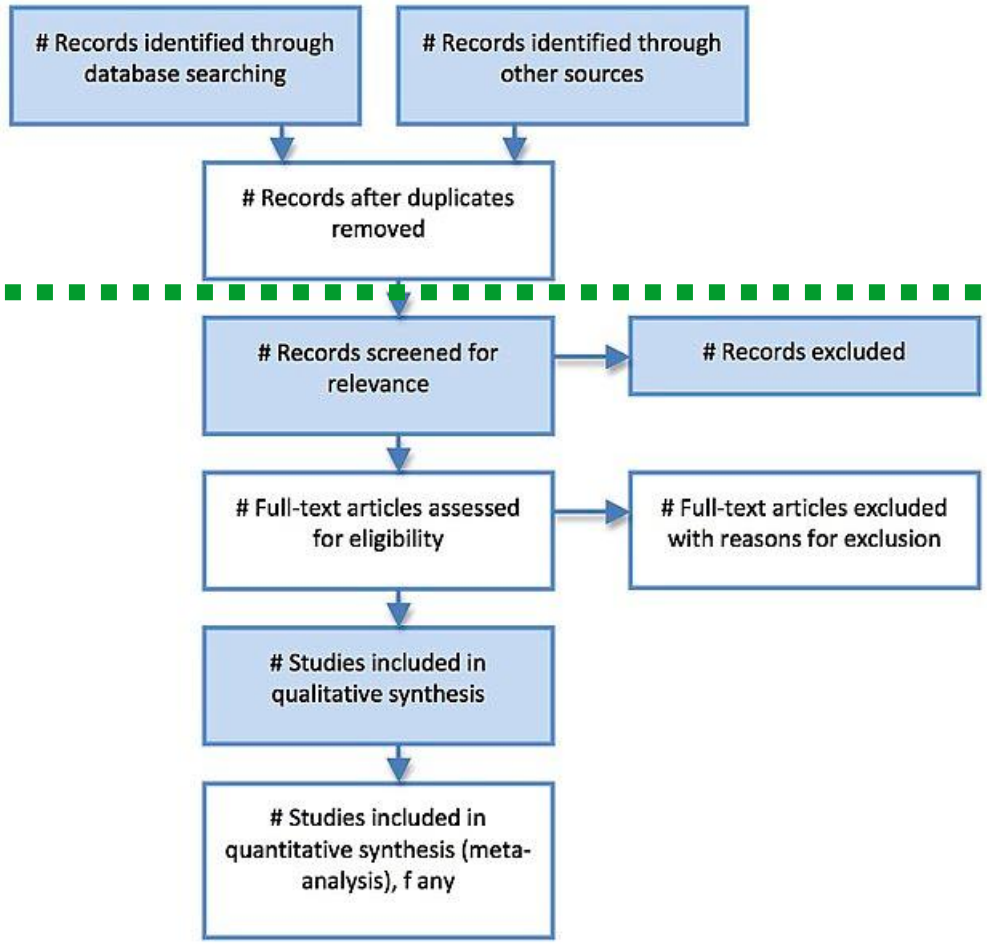
Yu Xiao, Maria Watson

First Published August 28, 2017 | Research Article | Check for updates
https://doi.org/10.1177/0739456X17723971

Reporting guidelines for main study types

Randomised trials	CONSORT	Extensions
Observational studies	STROBE	Extensions
Systematic reviews	PRISMA	Extensions
Study protocols	SPIRIT	PRISMA-P
Diagnostic/prognostic studies	STARD	TRIPOD
Case reports	CARE	Extensions
Clinical practice guidelines	AGREE	RIGHT
Qualitative research	SRQR	COREQ
Animal pre-clinical studies	ARRIVE	
Quality improvement studies	SQUIRE	Extensions
Economic evaluations	CHEERS	

Advanced method for scholarly evidence identification



Data Mining from Databases
Database Alignment

NLP and Information Retrieval
Human Screening

So…

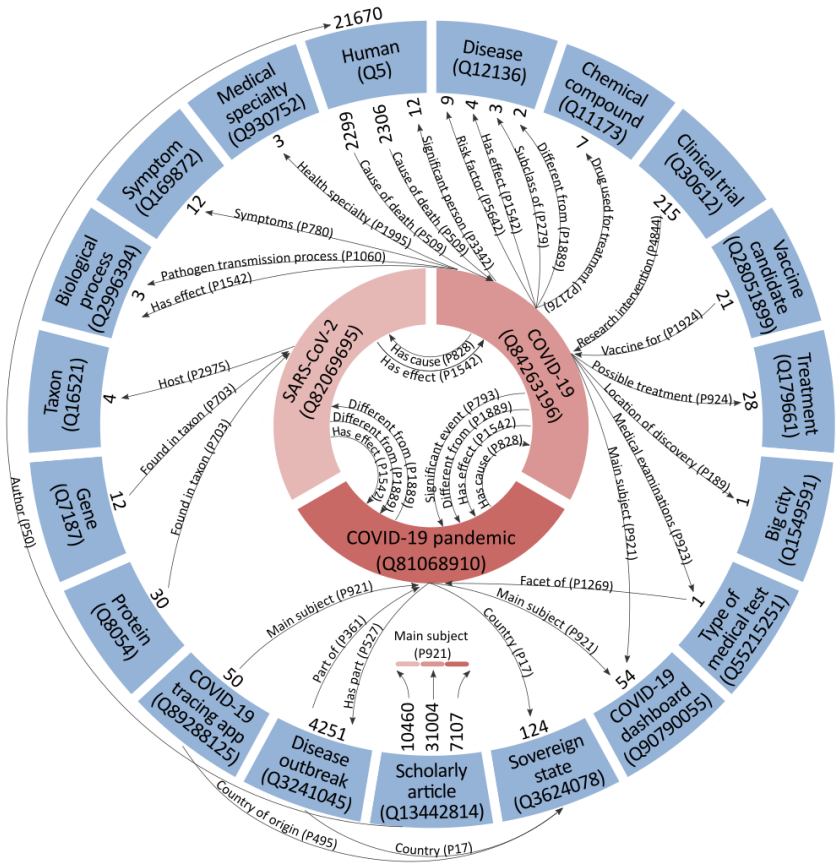
« It is a long and time-consuming work that should be automated to advance scholarly research »

Wikidata

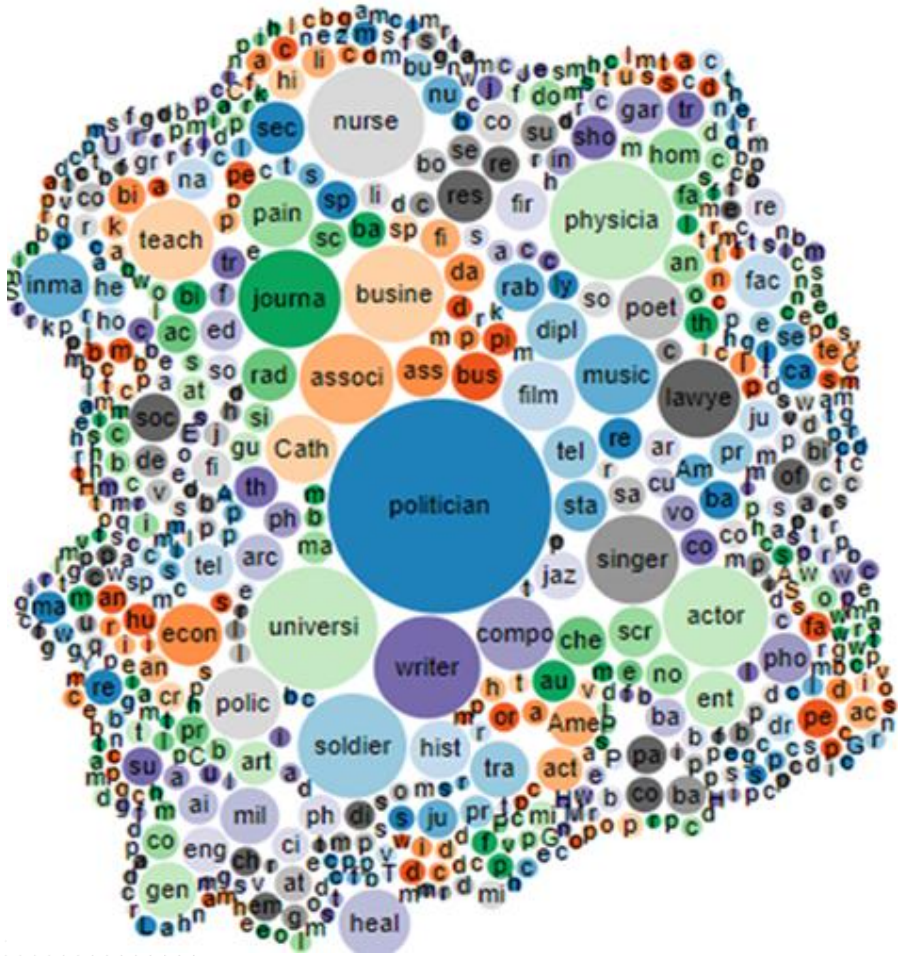
Large-scale resource that can be useful
for systematic review automation

Wikidata as a large-scale semantic database

Multidisciplinary Representation

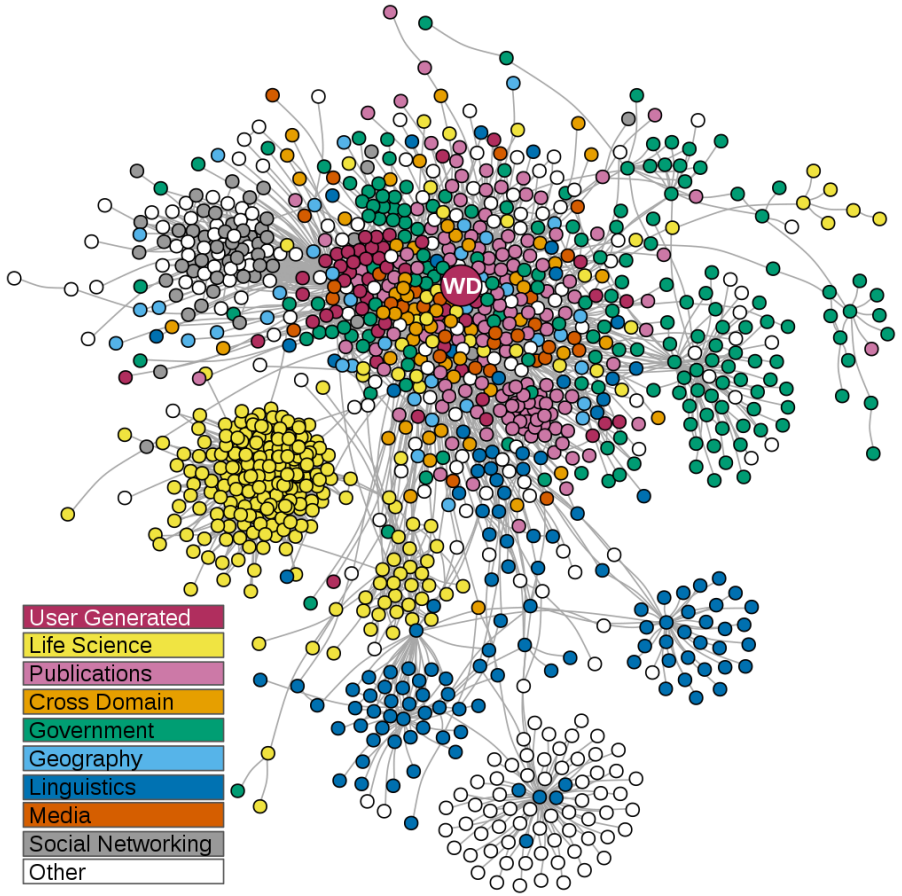


Data Federation and Integration



Wikidata as a large-scale semantic database

Aligned to other databases

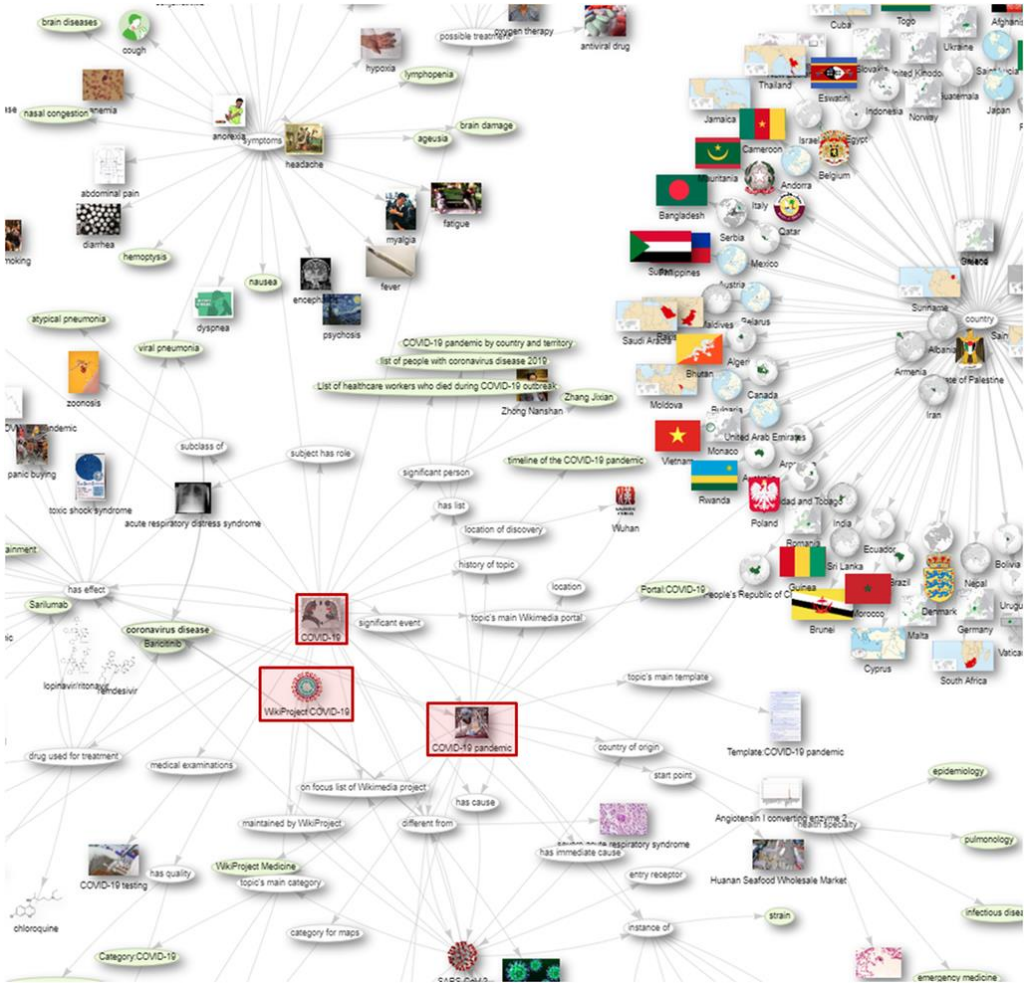


Freely available and reusable

CC0 1.0 Universal (CC0 1.0)
Public Domain Dedication

Wikidata as a knowledge graph

- Fully structured resource where concepts are represented as items and are described using statements in the form of triples.
- RDF Triples: Subject-Predicate-Object
  - Subject: The Wikidata item described by the Statement
  - Predicate: An entity representing the type of the information provided by the statement
  - Object: Can either be another item or a value
- Easy to:
  - Enrich and process (User interface, Wikidata Integrator, QuickStatements…)
  - Validate (Property Statements, Entity Schemas, Comparison with external databases, Logical constraints…)
  - Easy to query (SPARQL) and to download (RDF dumps, WDumper).
- Provide large-scale semantic information for driving computer applications in multiple natural languages.



Our approach

How we used Wikidata to automate several tasks for systematic review creation

Useful Applications of Wikidata for Systematic Review Automation

- Formulating an efficient search query to retrieve research publications from the parsed bibliographic database.
- Eliminating odd records from the initially retrieved publications using citation analysis.
- Adding further evidences to the list of verified publications based on citation analysis.
- Retrieving the features that should be extracted from scholarly publications to study the findings of the research output about the topic.

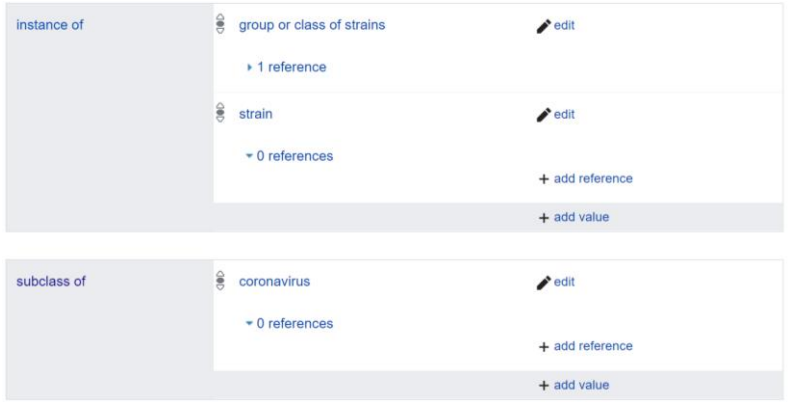
Wikidata involves needed information

» Citation data

wikicite

OpenCitations Bot

» Semantic relations

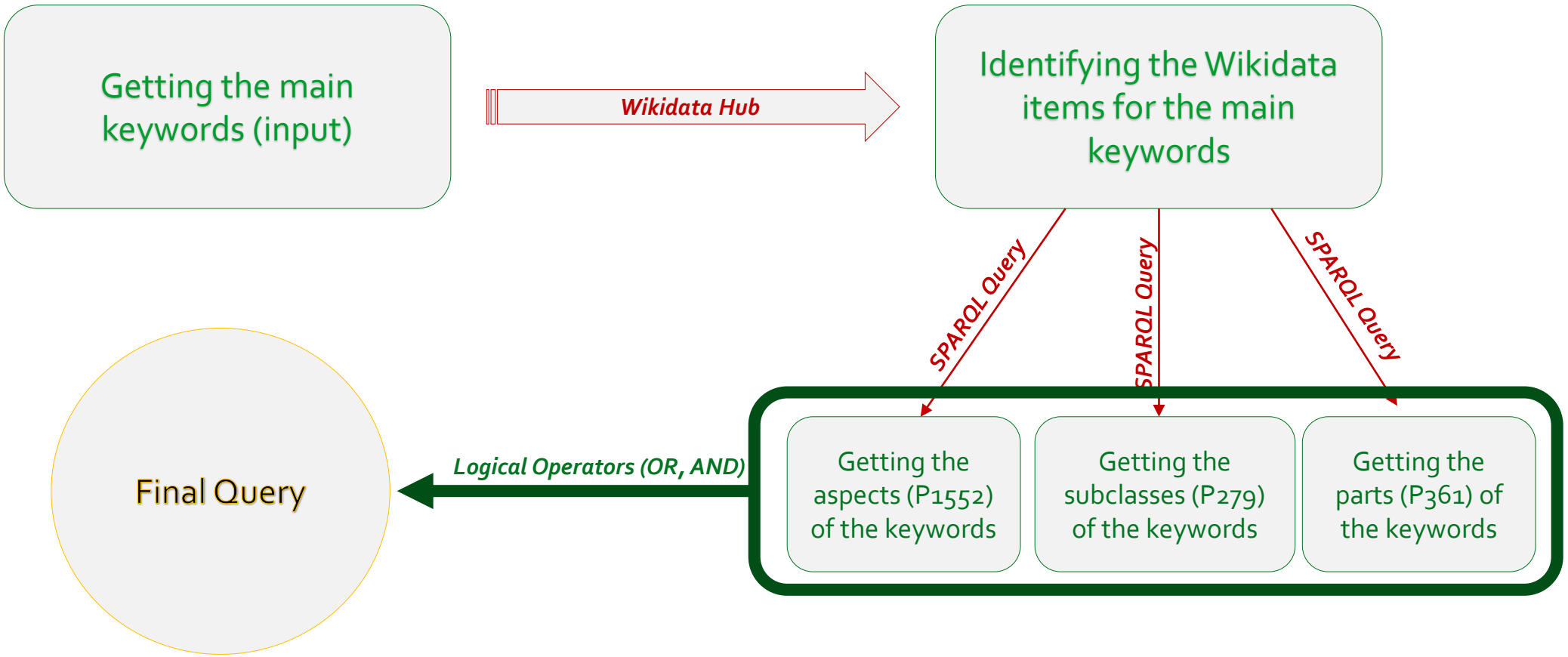


» Labels in Multiple Languages

Language	Label	Description	Also known as
English	SARS-CoV-2	strain of virus causing the ongoing pandemic of coronavirus disease 2019 (COVID-19)	2019-nCoV nCoV2019 nCoV-2019 Wuhan seafood market pn… Wuhan coronavirus Chinese coronavirus 2019nCoV 2019 nCoV NCP Novel coronavirus (2019-n… Severe acute respiratory sy… SARS-CoV2 novel coronavirus COVID-19 virus Wuhan virus Severe acute respiratory sy… Corona Coronavirus Corona virus Acute respiratory syndrom… severe acute respiratory sy… SARS-CoV-2 virus

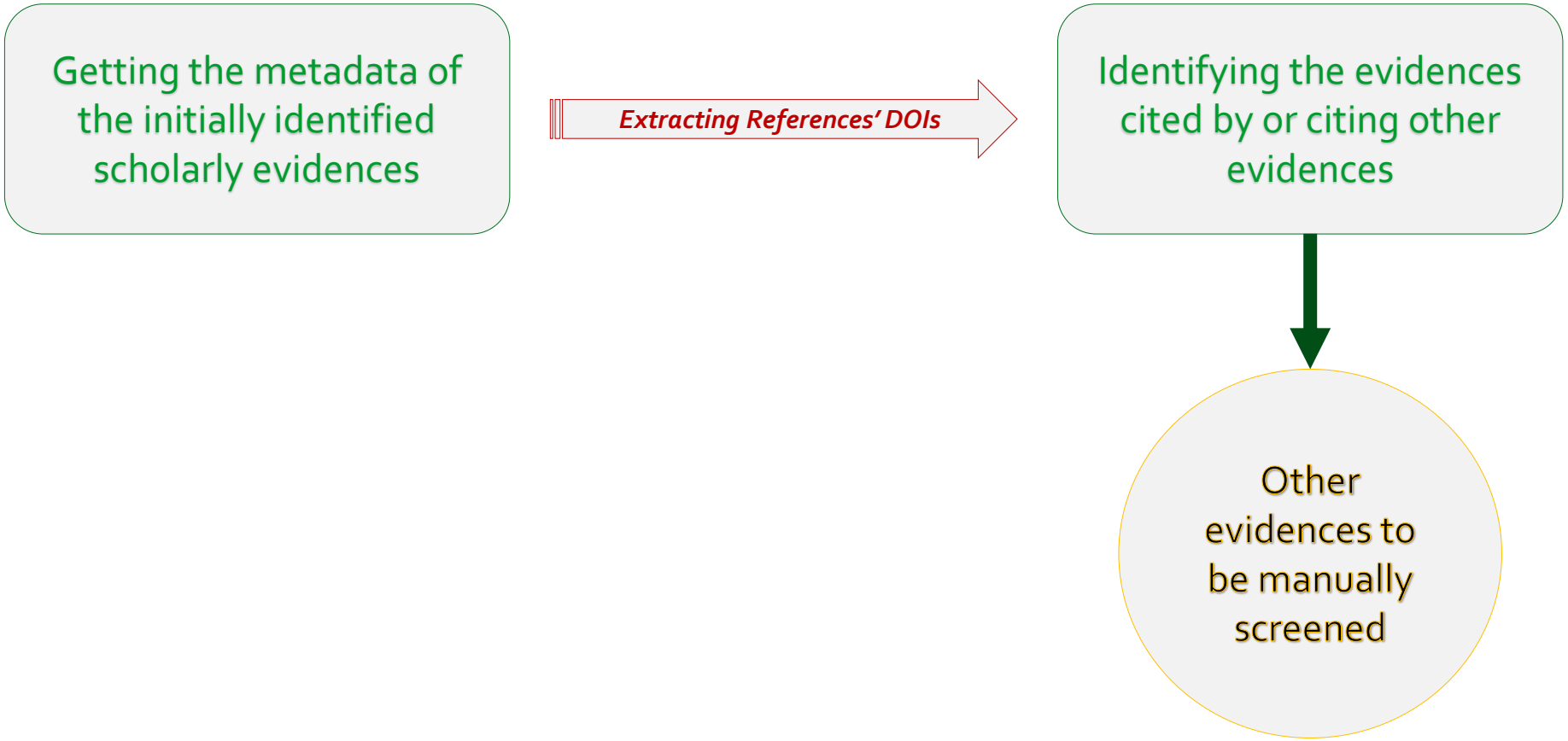
Task 01

Formulating an efficient search query to retrieve research publications from the parsed bibliographic database



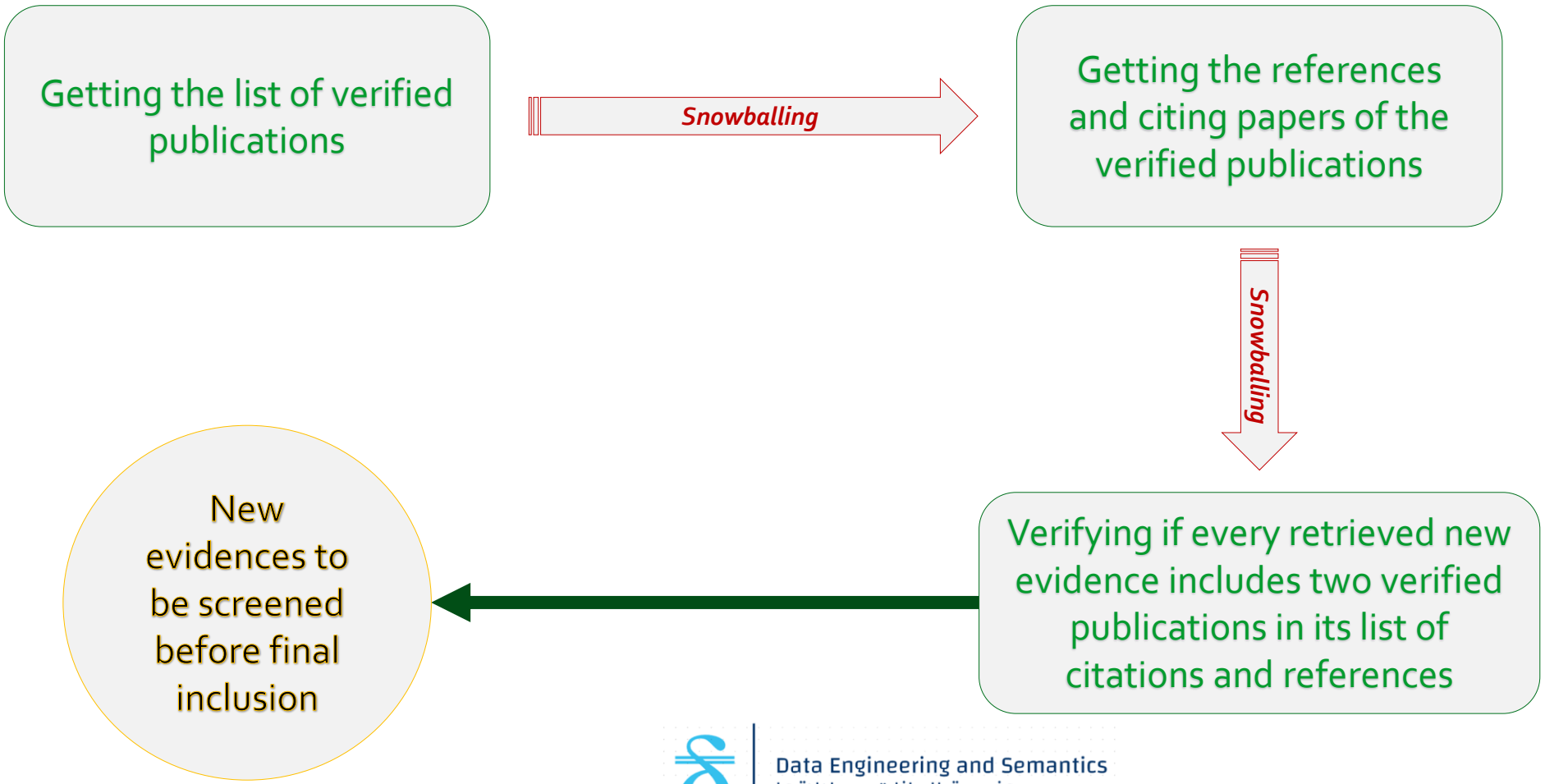
Task 02

Eliminating odd records from the initially retrieved publications using citation analysis



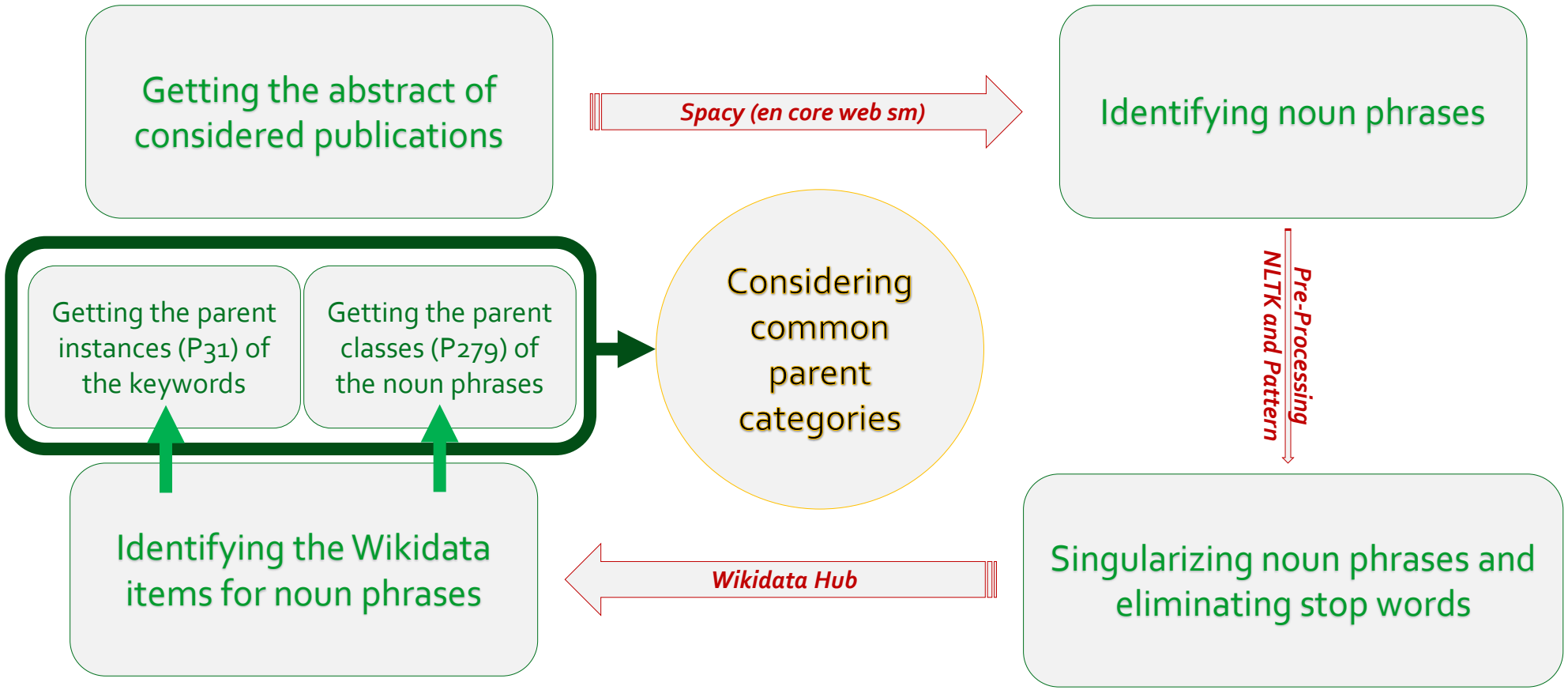
Task 03

Adding further evidences to the list of identified publications based on citation analysis



Task 04

Retrieving the features that should be extracted to study the findings of the research output about the topic



Source Codes

All tasks have been implemented in Python and are available at https://github.com/SisonkeBiotik-Africa/Systematic-Review-Automation. You can find the current development status of the project at https://www.loom.com/share/a415573dc016448dbe2bff220047b707.

Proven efficiency

Task 01: Wikidata-Enhanced Query vs. User Query (Test conducted on October 17, 2021)

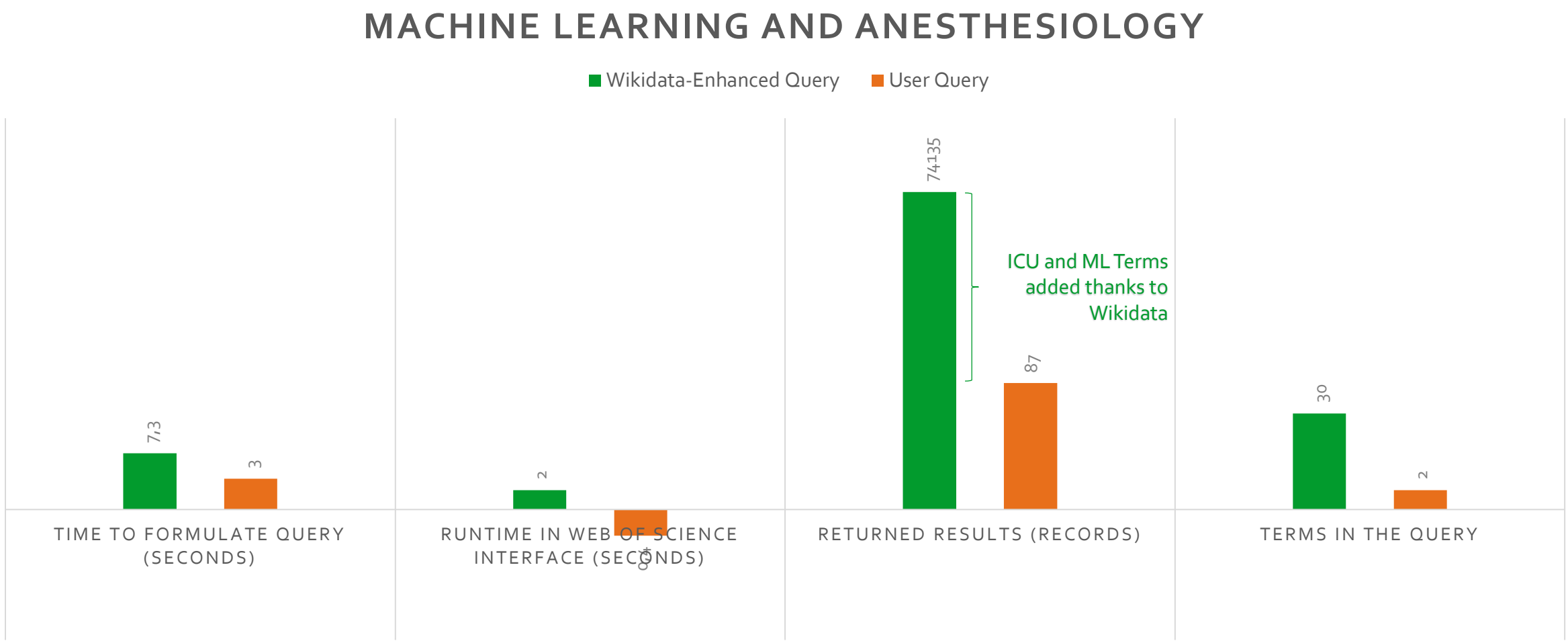


Image Credits

- https://www.equator-network.org/reporting-guidelines/prisma/
- https://commons.wikimedia.org/wiki/File:PRISMA_flow_diagram.jpg
- https://commons.wikimedia.org/wiki/Category:COVID-19_Study_of_Wikidata
- https://commons.wikimedia.org/wiki/File:WikiCite_wordmark.svg
- https://commons.wikimedia.org/wiki/File:OpenCitations_logo.jpg

References

- Elliott, J. H., Synnot, A., Turner, T., Simmonds, M., Akl, E. A., McDonald, S., ... & Pearson, L. (2017). Living systematic review: 1. Introduction—the why, what, when, and how. *Journal of clinical epidemiology*, *91*, 23-30.
- Thomas, J., Noel-Storr, A., Marshall, I., Wallace, B., McDonald, S., Mavergames, C., ... & Pearson, L. (2017). Living systematic reviews: 2. Combining human and machine effort. *Journal of clinical epidemiology*, *91*, 31-37.
- Simmonds, M., Salanti, G., McKenzie, J., Elliott, J., Agoritsas, T., Hilton, J., ... & Pearson, L. (2017). Living systematic reviews: 3. Statistical methods for updating meta-analyses. *Journal of clinical epidemiology*, *91*, 38-46.
- Akl, E. A., Meerpohl, J. J., Elliott, J., Kahale, L. A., Schünemann, H. J., Agoritsas, T., ... & Pearson, L. (2017). Living systematic reviews: 4. Living guideline recommendations. *Journal of clinical epidemiology*, *91*, 47-53.
- Selçuk, A. A. (2019). A guide for systematic reviews: PRISMA. *Turkish Archives of Otorhinolaryngology*, *57*(1), 57.
- Xiao, Y., & Watson, M. (2019). Guidance on conducting a systematic literature review. *Journal of Planning Education and Research*, *39*(1), 93-112.
- Turki, H., Hadj Taieb, M. A., Shafee, T., Lubiana, T., Jemielniak, D., Ben Aouicha, M., ... & Mietchen, D. (2021). Representing COVID-19 information in collaborative knowledge graphs: The case of Wikidata. *Semantic Web*.

جامعة صفاقس
University of Sfax

SISONKE-BIOTIK

Thank You

E-mail: turkiabdelwaheb@hotmail.fr

Phone: +21629499418

Twitter: @Csisc1994

https://dblp.org/pid/176/1531.html

https://www.researchgate.net/profile/Houcemeddine_Turki