



WikiCred Demo Hour

February 12, 2020

Fighting misinformation with Wikidata

Houcemeddine Turki

Data Engineering and Semantics Research Unit (DES-Unit), University of Sfax, Tunisia

Wikimedia TN User Group, Tunisia

Wikimedia and Libraries User Group, United States of America

Wikimedia Medicine, United States of America



<https://w.wiki/z47>



Introduction

Misinformation today

MISINFORMATION

Current Situation

- Misinformation is increasing and spreading more quickly over years because of the recent advances in Information Technologies, particularly Social networking sites.
- Misinformation affects the perception of events and facts by individuals and can bias social behaviors and convictions in nations
- Governmental and non-governmental institutions are currently trying to fight misinformation and spread ground truth about facts.

PERSPECTIVE | SOCIAL SCIENCE

The misinformation machine

Derek Ruths

+ See all authors and affiliations

Science 25 Jan 2019:
Vol. 363, Issue 6425, pp. 348
DOI: 10.1126/science.aaw1315

POLICY FORUM | SOCIAL SCIENCE

The science of fake news

David M. J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metz...

+ See all authors and affiliations

Science 09 Mar 2018:
Vol. 359, Issue 6380, pp. 1094-1096
DOI: 10.1126/science.aao2998

RESEARCH ARTICLE

Fake news on Twitter during the 2016 U.S. presidential election

 Nir Grinberg^{1,2,*},  Kenneth Joseph^{3,*},  Lisa Friedland^{1,*},  Briony Swire-Thompson^{1,2},  David Lazer^{1,2,†}

+ See all authors and affiliations

Science 25 Jan 2019:
Vol. 363, Issue 6425, pp. 374-378
DOI: 10.1126/science.aau2706

IN DEPTH | COVID-19

Officials gird for a war on vaccine misinformation

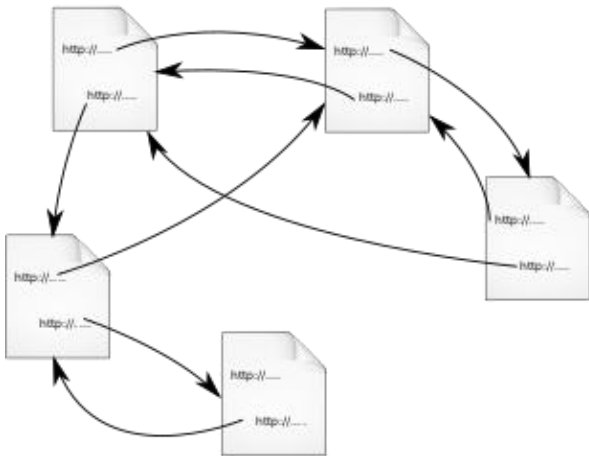
Warren Cornwall

+ See all authors and affiliations

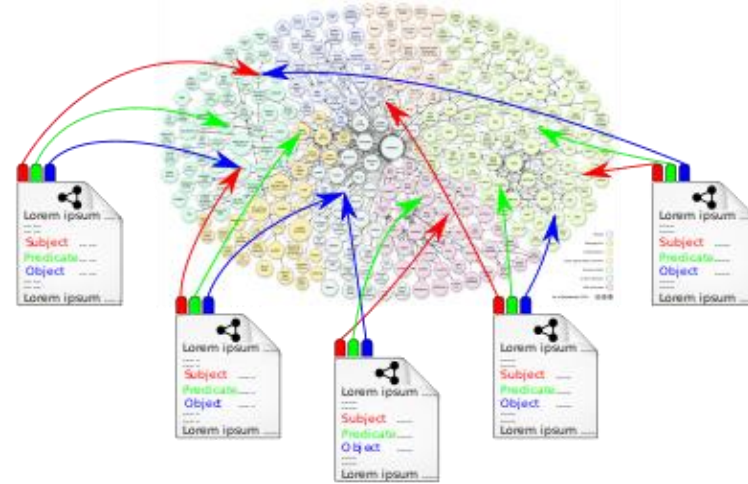
Science 03 Jul 2020:
Vol. 369, Issue 6499, pp. 14-15
DOI: 10.1126/science.369.6499.14

DATA SCIENCE

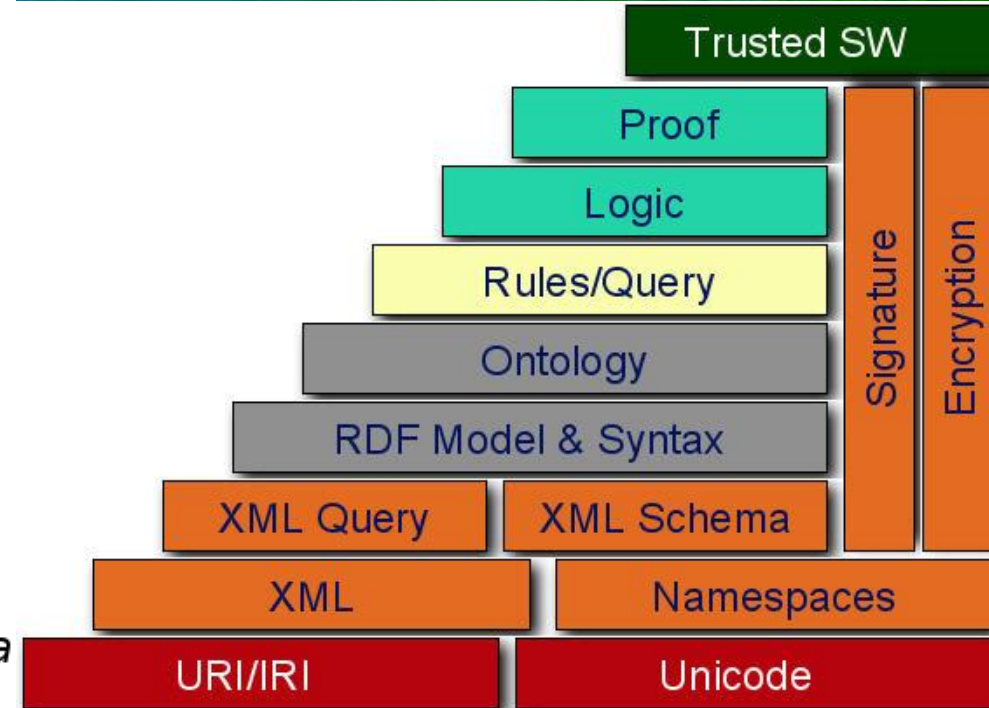
A scholarly field currently fighting misinformation



The traditional web -
A web of documents



The semantic web -
A web of human and machine readable content employing linked data





Wikidata

A large-scale semantic resource to fight misinformation behaviors

STRUCTURED DATA

Statements in the form of triples (RDF Format), Multilingual, Linked to external resources

label — **Douglas Adams** (Q42) — item identifier

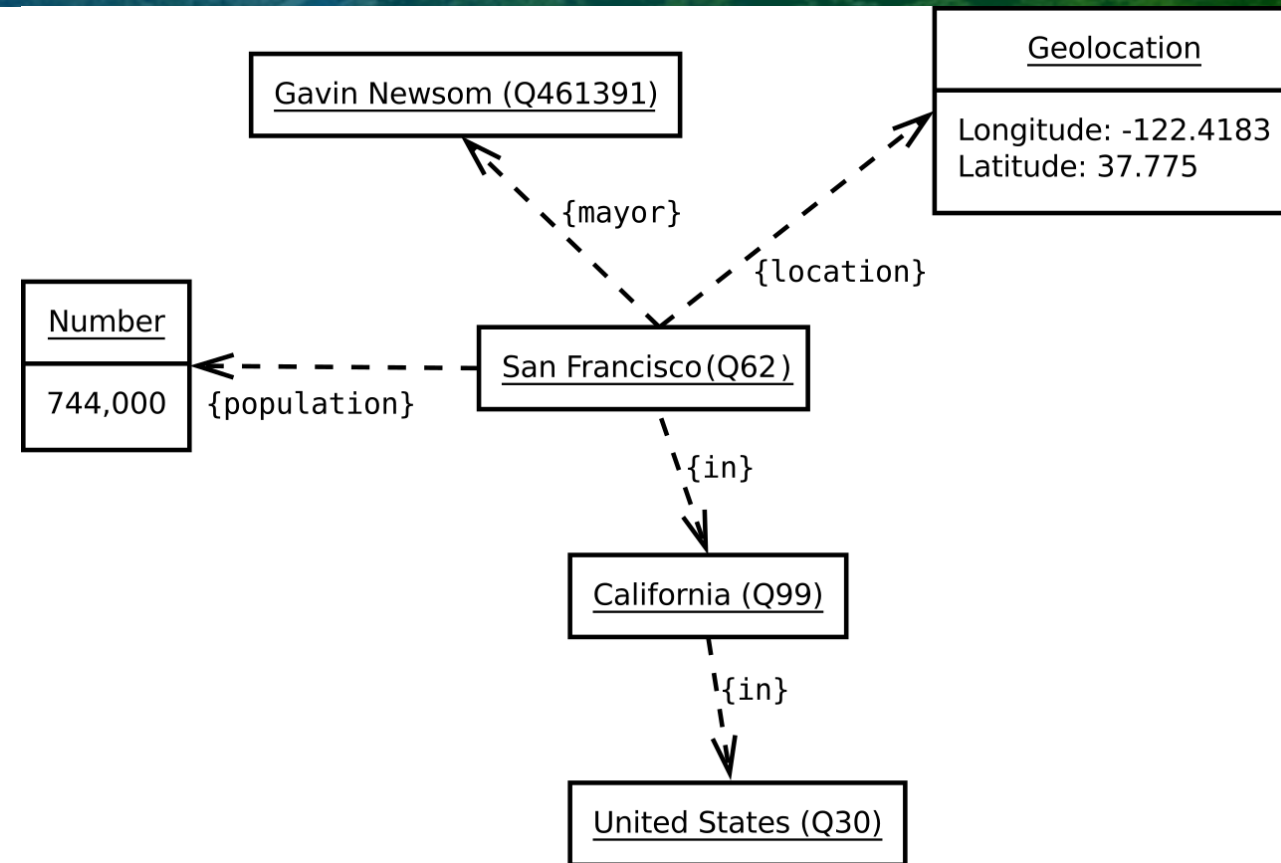
description — English writer and humorist
Douglas Noël Adams | Douglas Noel Adams — aliases
▶ In more languages

property — **educated at** — value

rank —

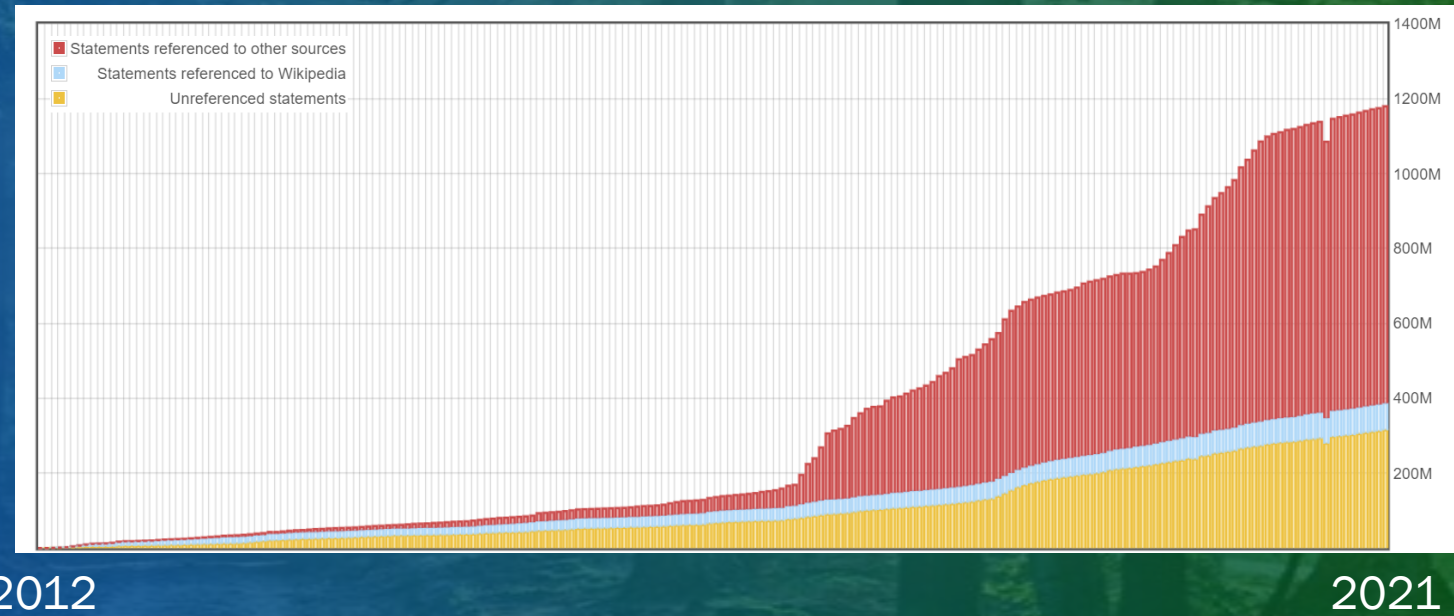
statement group —

St John's College	value
end time	1974
academic major	English literature
academic degree	Bachelor of Arts
start time	1971
qualifiers	
▼ 2 references	
opened references	
collapsed reference	



EVOLVING

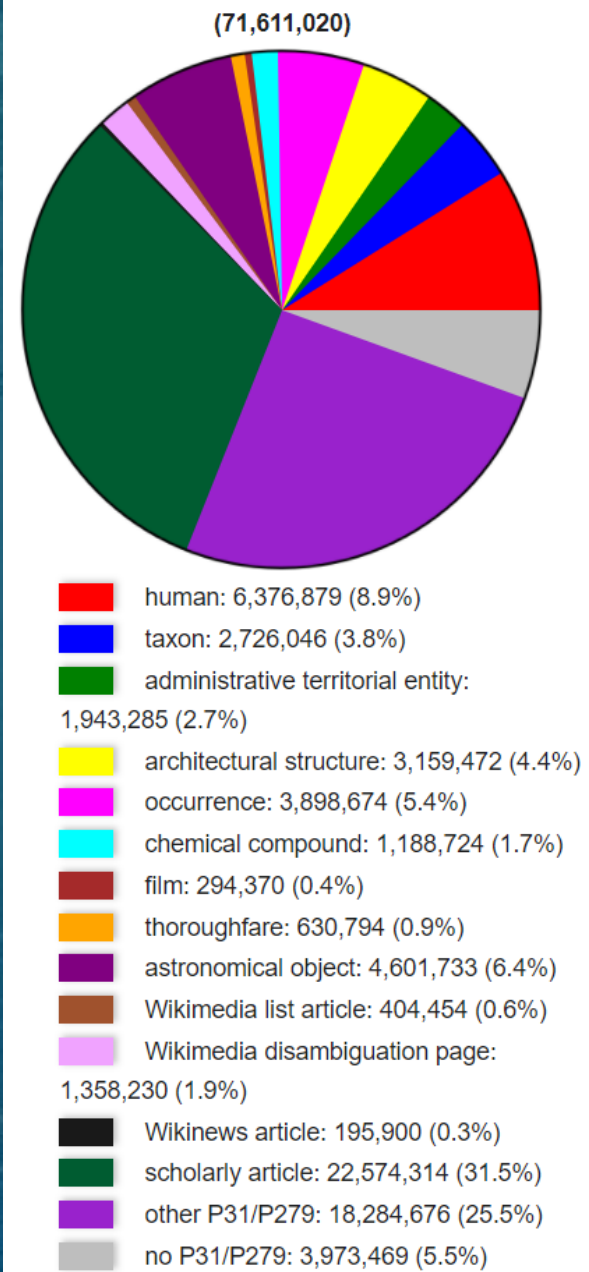
Number of statements



MULTIDISCIPLINARY

Broad coverage

ADD A FOOTER



Wikidata

How to fight misinformation

- Adding references to unsupported statements in Wikidata to verify its output and then using the output to verify facts online.
- Reference data for a dashboard to track misinformation fighting efforts



WIKIDATA

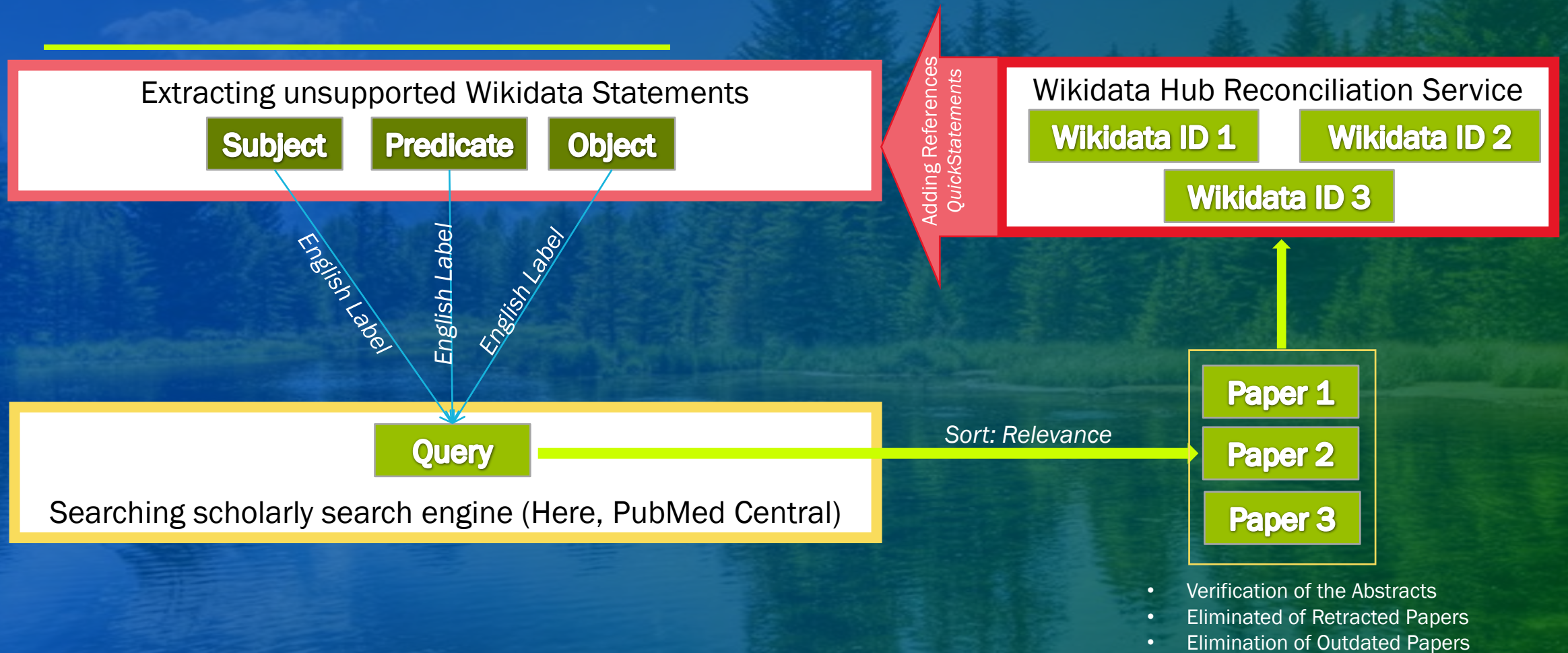




Adding references to Wikidata

Ensuring the verifiability of Wikidata statements

RefB: Citation Index-Based Method



Results

- 5,384 assessed Wikidata statements
- 451 statements assigned 842 references
- 47.4% of the 4,933 relations not assigned a reference are accurate
- 97.3% of the references are accurate
- Reasons for inaccurate references
 - Two items linked by more than one type of semantic relations (e.g. SARS-CoV-2 is different from COVID-19, SARS-CoV-2 causes COVID-19)
 - Reference dealing with hyponym (e.g. Fever is a symptom of infection, The reference deals with fever as a symptom of viral infection)
 - Negation (e.g. Wikidata says that X is a risk factor of Y where the reference says that X is not a risk factor of Y)

Similar Project: Reference Hunt

Based on external identifiers for Linked Open Data resources

Is this source a reliable reference material **and** does the extracted data support the Wikidata Statement?

Wikidata Statement
Item: Carlos Gattiker [Q5042117]
Property: country of citizenship [p27]
Value: Argentina [Q414]

Source Data
Source URL: <https://en.tennistemple.com/-ps5038> (Retrieved: 2020-06-26 20:51:13)
Extracted Data:

```
{  
  "http://schema.org/name": [  
    "Argentina"  
  ]  
}
```



Monitoring Misinformation Fighting Efforts

Building a database and dashboard to track misinformation fighting

What we need is to feed Wikidata with information

Five steps with QuickStatements (<https://quickstatements.toolforge.org>)

- Having a Excel spreadsheet of the misinformation fighting-related items (e.g. newspaper items, scholarly papers)
- Substituting categories with their Wikidata IDs (Search for the Category – Find its Wikidata ID – Replace the Category by its Wikidata ID in the Excel spreadsheet)
- Search for the entries in Wikidata (Use external identifiers to align existing items with their Wikidata IDs)
- Build a model for creating QuickStatements batch lines (Instructions at <https://www.wikidata.org/wiki/Help:QuickStatements>)
- Run the batch on QuickStatements

Demo (4 min.)

Sample Preparation of Batches from an Excel Spreadsheet

Example: COVID-19 information in Tunisia

Wikidata Hub Reconciliation Service

<https://hub.toolforge.org/P356:10.1016/J.JBI.2019.103292?format=json>

Wikidata Query Service

List of works about misinformation related to the COVID-19 pandemic:

<https://w.wiki/yuL>

Co-authorship Network: <https://w.wiki/yuM>

Bubble charts about the source titles of the work: <https://w.wiki/yuN>

Co-occurring topics: <https://w.wiki/yuP>

Topics of interest for the authors up to 2019 : <https://w.wiki/z3J>



Conclusion

Outcomes for this work

Conclusion

- It is easy to use Wikidata to track and encourage misinformation fighting.
- Further efforts are needed to turn Wikidata into a centralized knowledge base for fighting misinformation.
- The advances in word embeddings, semantic similarity and topic modeling will increase the functions of Wikidata in the context of misinformation fighting.

Image Credits

- [File:Data science.png - Wikimedia Commons](#)
- [File:W3c semantic web stack.jpg - Wikimedia Commons](#)
- [File:The Semantic Web Compared To The Traditional Web.svg - Wikimedia Commons](#)
- [File:Datamodel in Wikidata.svg - Wikimedia Commons](#)
- [File:Linked Data - San Francisco.svg - Wikimedia Commons](#)
- [Wikidata:Statistics - Wikidata](#)
- <https://wikidata-toolforge.org/stats.php>
- [File:Wikidata-logo-en.svg - Wikimedia Commons](#)
- [https://commons.wikimedia.org/w/index.php?title=File:What%27s new in Wikidata - Italian WikiCon 2020.pdf&page=13](https://commons.wikimedia.org/w/index.php?title=File:What%27s_new_in_Wikidata_-_Italian_WikiCon_2020.pdf&page=13)

References

- <https://misinfocon.com/refdata-adding-trustworthiness-to-wikidata-d3cc68c21a6f>
- <https://misinfocon.com/refb-on-the-quest-of-references-for-collaborative-knowledge-resources-7efad8c88772>
- <https://github.com/wmde/reference-island>
- <https://www.sciencedirect.com/science/article/pii/S1532046419302114>
- <https://zenodo.org/record/4033382>
- <https://zenodo.org/record/4008359>

THANK YOU!

Houcemeddine Turki

Phone

+21629499418

Email

turkiabdelwaheb@hotmail.fr