

Recoin

Relative Completeness in Wikidata

Vevake Balaraman

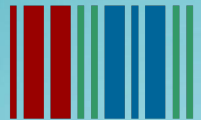
balaraman@fbk.eu

Simon Razniewski

Max Planck Institute for
Informatics

Werner Nutt

Free University of Bozen-Bolzano



WIKIDATA

**DATA QUALITY
DAYS 2022**

Introduction

- Knowledge bases (KB) such as Wikidata allow to store data in (almost) endless levels of detail for an entity
 - (Obama, pet, Bo)
 - (Trump, pseudonym, John Miller)
- Assessing the completeness of such data is important to perform data curation.
- How to assess if the KB covers all the information of an entity? (or) How to assess the completeness of various classes in Wikidata?
- E.g., **“How complete is data about Antoine Griezmann in Wikidata?”**

Whose Wikidata is more complete?



ReCoin

- Capturing recall of properties in comparison with other similar subjects.
- ***sim***(S_1, S_2) - computes a (weighted) set of similar subjects ***S'***.
- ***score***(S, S') - computes a score / rank for the completeness of S with regard to a set of comparison subjects ***S'***.
- **Result** : *score* indicating the percent of relative completeness of entity.
- *level* indicating one of 5 levels of information captured in KB for an entity (Level 1 to 5 indicating least to most).
- Features Property suggestion tool based on the similarity function.

ReCoin - Antoine Griezmann



- Main page
- Community portal
- Project chat
- Create a new item
- Recent changes
- Random item
- Query Service
- Nearby
- Help
- Donate

Tools

- What links here
- Related changes
- Special pages
- Permanent link
- Page information

Item

Discussion

Read

More ▾

Search Wikidata



Antoine Griezmann (Q455462)

French association football player

▾ Most relevant properties which are absent

Property ID	Label	Relative	Add Claim
P734	family name	15.11%	+
P1412	languages spoken, written or signed	12.5%	+
P118	league	5.36%	+
P1814	name in kana	2.34%	
P69	educated at	1.58%	+
P1532	country for sport	1.42%	+
P166	award received	1.23%	+
P1889	different from	0.94%	+
P1477	birth name	0.74%	
P172	ethnic group	0.6%	+





instance of

human

▶ 1 reference

image

Tim Berners-Lee (cropped).jpg

▼ 0 references

sex or gender

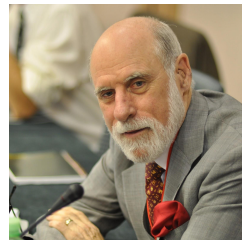
male

▶ 1 reference

country of citizenship

United Kingdom

▶ 1 reference



occupation

computer scientist

▶ 1 reference

engineer

university teacher

▶ 1 reference

professor

▼ 0 references

researcher

▼ 0 references

work

computer science

▼ 0 references

95% complete

80% complete

Recoin

Accessing ReCoin

- Available as a Gadget for all WikiData users
 - Login → Preferences → Gadgets → “Recoin” under section "Gadgets/Wikidata-centric"
- Special version: IDs only
 - `importScript('User:Vvekbv/recoin_id.js');`
- API's
 - <https://tools.wmflabs.org/recoin/getmissingattributes.php?lang=en&subject=Q15074414&n=10>
 - https://tools.wmflabs.org/recoin/getmissingattributes_id.php?lang=en&subject=Q15074414&n=10

Further readings:

- Wiki Workshop at The Web Conference 2018
 - ReCoin: Relative Completeness in Wikidata
- ReCoin Plugin: <https://www.wikidata.org/wiki/Wikidata:ReCoin>
- ESWC P&D 2017
 - Assessing the Completeness of Entities in Knowledge Bases

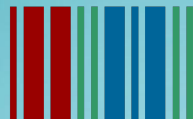
Thanks for your attention!

Vevake Balaraman

balaraman@fbk.eu

Credits

Simon Razniewski.



WIKIDATA

**DATA QUALITY
DAYS 2022**