BioHackrXiv

Knowledge graphs and wikidata subsetting

Jose Emilio Labra Gayo[1], **Ammar Ammar**[2], **Dan Brickley**[3], **Daniel Fernández Álvarez**[1], **Alejandro González Hevia**[1], **Alasdair Gray**[4], **Eric Prud'hommeaux**[1], **Denise Slenter**[2], **Harold Solbrig**[6], **Seyed Amir Hosseini Beghaeiraveri**[4], **Benno Fünkfstük**[8], **Andra Waagmeester**[7], **Egon Willighagen**[2], **and Liza**[3]

1 WESO research group, University of Oviedo, Spain **2** Maastricht University **3** Google, London, UK **4** Heriot Watt University, UK **5** Janeiro Digital, W3C/MIT **6** Johns Hopkins University **7** Micelio/Gene Wiki **8** TU Dresden

BioHackathon series:
BioHackathon Europe 2020
Virtual conference 2020
Wikidata subsetting

Submitted: 14 Jan 2021

Published by BioHackrXiv.org

Introduction

- Short sentence about Knowledge graphs and wikidata.
- Talk about entity schemas at Wikidata which are based on ShEx (Prud'hommeaux, Labra Gayo, & Solbrig, 2014)

Motivation

TODO

Use cases

Description of different approaches and pipeline

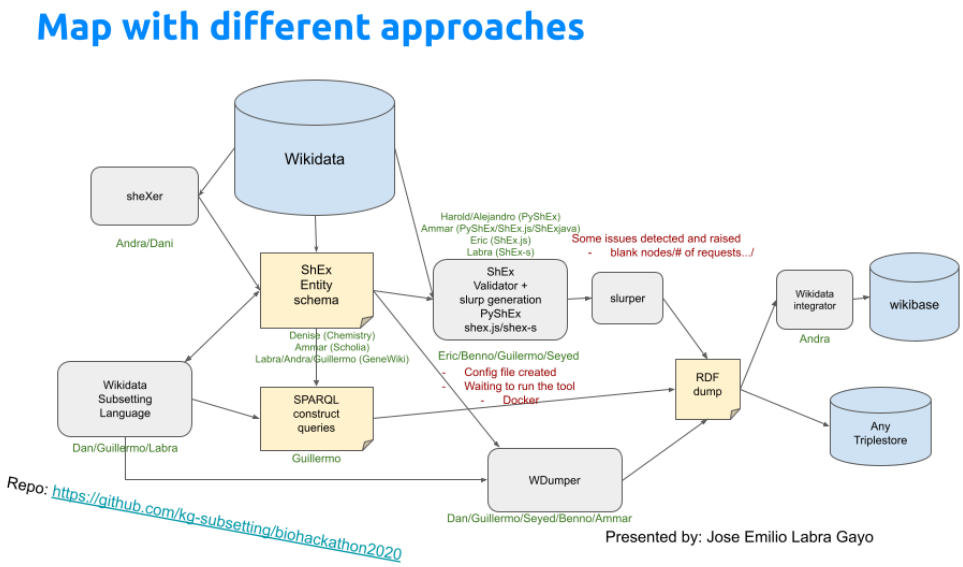

Figure 1: Diagram presenting the different approaches

Describing the subsets

- Using Entity schemas created manually
- Generating entity schemas from instance data

Wikidata subsetting language

- TODO: Describe the idea...
- Generates SPARQL Construct queries

Extraction of subsets from wikidata

Slurping

WDumper

Creating the subset

- Transforming and enriching the RDF dump using schema.org
- WikidataIntegrator

Discussion

TODO: Review...

- Talk about docker image
- Future work: Wikidata subsetting as a service?

Conclusions

References

Prud'hommeaux, E., Labra Gayo, J. E., & Solbrig, H. (2014). Shape expressions: An RDF validation and transformation language. In *Proceedings of the 10th international conference on semantic systems, SEMANTICS 2014* (pp. 32–40). ACM.