Species Association Knowledge Graph Construction - A Demo Paper

Dina Sharafeldeen 0000-0001-5801-4948, Alsayed Algergawy 0000-0002-8550-4720, and Birgitta König-Ries 0000-0002-2382-9722

Heinz Nixdorf Chair for Distributed Information Systems
Friedrich Schiller University of Jena, Germany
{dina.sharafeldeen, alsayed.algergawy, birgitta.koenig-ries}@uni-jena.de

Abstract. Constructing knowledge graphs for new domains and linking them to existing ones has recently gained significant attention, especially in domains that have experienced a tremendous increase in available data such as biodiversity research. In this demo, we show a semantic data mining framework combining several knowledge bases to help in this task and show the feasibility of our framework using real-world datasets from a large-scale biodiversity project.

Keywords: Knowledge Graph, Association rules, Data mining

Biodiversity is a multidisciplinary, challenging research area that has experienced a tremendous increase in the number of datasets [3]. Therefore, it is quite challenging, on the one hand, to extract valuable hidden knowledge from these complex datasets and on the other hand, to make these datasets linkable to other sources of knowledge to gain new insights. Knowledge graphs can be processed in various ways, leading to applications such as semantic search, question answering, and entity resolution [1]. Using robust techniques for knowledge graph construction automation is crucial and useful in different domains. Motivated by [5], Page proposed a biodiversity knowledge graph [2]. It is combining and interlinking information about biodiversity entities, such as taxa, taxonomic names, publications, people, species, sequences, images, and collections. Many questions in biodiversity can be framed as paths in this graph. We proposed a semantic data mining framework [4] for knowledge graph construction to extract hidden knowledge from species datasets combined with other knowledge sources like Encyclopedia of Life (EOL)[1] and Global Biotic Interactions (GloBI)[2]. Combining these sources results in a knowledge graph that will be accumulatively constructed and refined.

In our demo, visitors will be able to construct an association knowledge graph from species abundance datasets. After they load the species abundance dataset, they can select the attributes that specify the date of observation. Then, they can select the radius of the species plot, which is the area where the species is observed. Afterwards, we perform entities(species) disambiguation by linking

[1] https://eol.org/

[2] www.globalbioticinteractions.org

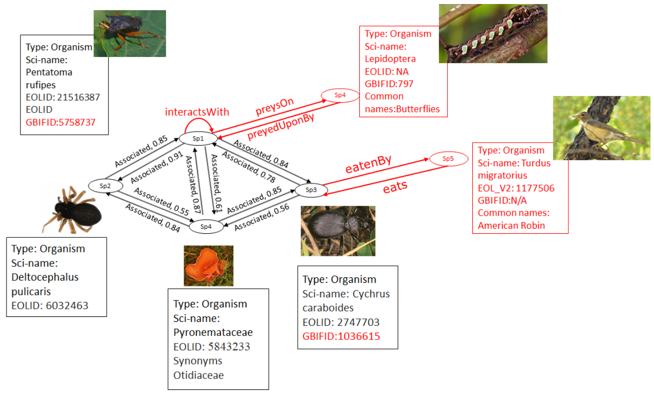

Fig. 1: Part of the knowledge graph after enrichment with GLOBI

to EOL. Then, the data is transformed into transactions to fit the association rules extraction algorithm. Each transaction contains all the species that exist in the same plot and on the same date. Then, the association rules mining algorithm is applied to extract the species association rules. In addition, these rules are represented in RDF format. Then, the initial association knowledge graph is constructed as shown in Fig.1 (the entities, properties, and the relations in the black color). Furthermore, we enrich the constructed graph by linking to EOL getting more information like images and other accepted scientific names. Moreover, linking to GloBI enriches the constructed knowledge graph. This enrichment is achieved by promoting some of the extracted species co-occurrence to concrete interactions. In another way, more enrichment is done by adding new entities with their relationships from the GLOBI knowledge base, as shown in Fig.1. As future work, we are working to demonstrate the general applicability of our tool with more datasets in the biodiversity domain and other domains.

References

1. Kejriwal, M.: What is a knowledge graph? In: Domain-Specific Knowledge Graph Construction (2019)
2. Page, R.: Towards a biodiversity knowledge graph. Research Ideas and Outcomes **2**, e8767 (2016)
3. Shah, A.: Why is biodiversity important? who cares?—global issues. Global Issues: Social, Political, Economic and Environmental Issues That Affect Us All—Global Issues **18** (2009)
4. Sharafeldeen, D., Algergawy, A., König-Ries, B.: Towards knowledge graph construction using semantic data mining. In: The 21st International Conference on Information Integration and Web-based Applications & Services (In press) (2019)
5. Szekely, P., Knoblock, C.A., Slepicka, J., Philpot, A., Singh, A., Yin, C., Kapoor, D., Natarajan, P., Marcu, D., Knight, K., et al.: Building and using a knowledge graph to combat human trafficking. In: International Semantic Web Conference. pp. 205–221. Springer (2015)