RDFsim: Similarity-Based Browsing over DBpedia Using Embeddings

Manos Chatzakis[1,2], **Michalis Mountantonakis** [1,2#,*] and **Yannis Tzitzikas**[1,2]

[1] Institute of Computer Science, FORTH-ICS, Heraklion, Greece

[2] Computer Science Department, University of Crete, Heraklion, Greece

\# Presenting and Corresponding author: Michalis Mountantonakis, email: mountant@ics.forth.gr

ABSTRACT

Browsing has been the core access method for the Web from its beginning. Analogously, one good practice for publishing data on the Web is to support dereferenceable URIs, to also enable plain web browsing by users. The information about one URI is usually presented through HTML tables (such as DBpedia and Wikidata pages) and graph representations (by using tools such as LODLive and LODMilla). In most cases, for an entity, the user gets all triples that have that entity as subject or as object. However, sometimes the number of triples is numerous, thereby it can result in information overload; for example, the webpage about Aristotle in DBpedia contains over 1000 links, making it hard for the user to identify the most important information. Apart from tackling information overload, it is difficult to find relevant entities that are not directly connected through current browsing systems. For instance, the page of Socrates in DBpedia does not include a link to Pythagoras and vice versa; however, we expect that a similarity-based algorithm can possibly detect these two important people of Ancient Greece as similar.

To alleviate these problems, we investigate a similarity-based browsing method for RDF data based on knowledge graph embeddings. Specifically, the creation of knowledge graph embeddings is a machine learning task of learning a low-dimensional representation of a knowledge graph's entities and relations by preserving their semantic meaning. The embedded representations can be used for several applications, including the discovery of the most similar entities of a given entity. By creating knowledge graph embeddings through DBpedia Knowledge graph, we introduce an interactive similarity-based browsing system, called RDFsim, that offers "Parallel Browsing". In particular, it enables the user to see and browse not only the original data of the entity in focus, but also the K most similar entities of the focal entity, offering in this way a method to tackle the information overload. The mentioned parallel browsing is offered through real-time interaction, since all the required data are stored in dedicated indexes that can be accessed through random access file mechanisms and do not depend on the particular method for computing similarity.

Concerning the web application of RDFsim (publicly available in https://demos.isl.ics.forth.gr/RDFsim/), it offers a left frame where the user can see and browse the original data of the entity in focus, and a right frame enabling the user to inspect the K most similar entities of the focal entity in the form of a star-like graph or tag cloud. The user can change the entity in focus by interacting with any of these two frames. The current version of RDFsim offers similarity-based browsing for over 350,000 entities of DBpedia, including entities belonging to Philosophers, Programming Languages, Video Games and Movies. All the details about its implementation, its evaluation and use cases are available in [1].

REFERENCES

[1] Chatzakis, Manos, Michalis Mountantonakis, and Yannis Tzitzikas. "RDFsim: Similarity-Based Browsing over DBpedia Using Embeddings." Information 12, no. 11 (2021): 440.