

Re-constructing Hidden Semantic Data Models by Querying SPARQL Endpoints

María Jesús García-Godoy^(✉), Esteban López-Camacho,
Ismael Navas-Delgado, and José F. Aldana-Montes

Departamento de Lenguaje y Ciencias de la Computación, Universidad de Málaga,
Andalucía Tech, Ada Byron Research Building, 29071 Málaga, Spain
{mjgarcia, jfam}@lcc.uma.es
<http://khaos.uma.es/>

Abstract. Linked Open Data community is constantly producing new repositories that store information from different domains. The data included in these repositories follow the rules proposed by the W3C community, based on standards such as Resource Description Framework (RDF) and the SPARQL query language. The main advantage of this approach is the possibility of external developers accessing the data from their applications. This advantage is also one of the main challenges of this new technology due to the cost of exploring how the data is structured in a given repository in order to construct SPARQL queries to retrieve useful information. According to the reviewed literature, there are no applications to reconstruct the underlying semantic data models from an SPARQL endpoint. In this paper, we present an application for the reconstruction of the data model as an OWL (Ontology Web Language) ontology. This application, available as Open Source at <http://github.com/estebanpua/ontology-endpoint-extraction> uses a set of SPARQL queries to discover the classes and the (object and data) properties for a given RDF database. A web application interface has also been implemented for users to browse through classes, properties of the ontology generated from the data structure (<http://khaos.uma.es/oe>). The ontologies generated by this application can help users to understand how the information is semantically organized, making easier the design of SPARQL queries.

Keywords: Semantic Web · Ontology · OWL · Linked Open Data · Endpoint