Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 83 (2016) 179 – 186

Procedia Computer Science

The 7th International Conference on Ambient Systems, Networks and Technologies (ANT 2016)

Conto: a protégé plugin for configuring ontologies

Andrew LeClair[a], Ridha Khedri[a]

[a]*McMaster University, Hamilton, Ontario, Canada*

Abstract

With the extensive proliferation of data, the efficient organization and integration of data has become exceedingly important. To achieve this, ontologies have been employed to govern the data by providing structure and a layer of abstraction over the data sources. However, the ontologies themselves present shortcomings in how they capture the definition of concepts of a domain. The definitions are static and unable to sufficiently conceptualize the dynamic data sources in a way that makes use of the knowledge provided by the data. This limitation carries over to the reasoning processes performed using ontologies.

We introduce Conto, a Protégé plugin that overcomes this limitation by interpreting a concept as an abstract data type. These interpretations allow for new ways of understanding the data, which when coupled with knowledge generation processes leads to the generation of new knowledge that would traditionally be unobtainable.

Crown Copyright © 2016 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Peer-review under responsibility of the Conference Program Chairs

Keywords: Data Governance; Ontology; Concept Interpretation; Protégé; Data Integration

1. Introduction

Traditionally, during the discovery process, data is analyzed to identify parameters such as relevance, required security, and permissions before uploading it to a central repository. However, within the last decade we have seen an increase to the amount of data being collected[1]. The domains that are impacted by this deluge of data range from businesses and the collection of business records[2] to astronomy and the cataloging of interstellar bodies[3]. With this deluge, the traditional methods of data governance are pushed to their limits; classical methods can no longer scale to handle the large quantities of data, and thus require new governing methods. To tackle this issue, the data must be structured so that it can be understood with an agile method before it can be used.

Ontologies have been employed to provide structure to the existing data so that it can be assessed and understood. This results in an efficient and safe way that users can interact with the data, as exemplified through technology such as the Ontology-Based Data Access (OBDA)[4]. The OBDA approach is used to create mappings between a relational database and an ontology. It constantly updates its understanding of the data through queries and rules. For example, considering a university information system, an ontology may define how the concepts of bachelor and

E-mail address: khedri@mcmaster.ca

1877-0509 Crown Copyright © 2016 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Peer-review under responsibility of the Conference Program Chairs
doi:10.1016/j.procs.2016.04.114

masters students are both students. The actual data about the students are maintained in a relational database. The ontology behaves as a layer above the data, and conceptualizes the domain so that the data can be assessed. Therefore, these conceptualizations are crucial for the understanding of the data. The conceptualized objects need to be able to sufficiently define and embody the data. Often times, a form of Description Logic (DL)[5] is used – such as DL-Lite[6] used by OBDA – which defines concepts within the terminological box (TBox). However, the concepts are statically defined. They are initialized with definitions that relate them to other earlier defined concepts and never updated or refined with respect to the data. This static definition is not sufficient for the domain of Big Data where data is constantly being updated and changing.

Dynamically defining concepts within an ontology have recently begun to be explored with the ideas of interpreting concepts[7]. Jaskolka et al.[7] suggest a framework for ontology design which includes an archetype and interpretations component (among other components). Within the archetype component, concepts are defined at an abstract and domain-independent level. The attributes that belong to the concepts are declared with a data type. For example, the concept *StudentGradeType* may be defined as a String in the archetype component. However, within a domain, a data type may be understood in several different ways. The interpretations component defines the concept dependent on how the archetype is to be understood. Returning to our example, *StudentGradeType* as a String may be understood in several ways: a *letter-grade* (e.g., A), as a *percentage* (e.g., 85%), or as a *grade-point average* (e.g., 3.0/4.0). The interpretations component clearly instantiates the concept by giving the archetype an interpretation. From these interpretations component, it leads us to understand that ontologies have a dynamic nature which changes depending on how the concepts are interpreted.

We propose a tool, Conto, which allows for the interpretation of concepts in an ontology with abstract data types. With this notion of interpretation, the understanding of the data changes depending on the interpretation used – a way of understanding a concept in more than one way is introduced. Through the interpretation of a concept, we not only structure the data with some data type and the associated properties, but we also gain the privileges of the functionality of this data type that allows us to learn about the data associated with the concept. With this ability for the user to give an interpretation to a concept, we are enhancing the current ways that users interact with, and understand the data.

The remainder of this paper is organized as follows. Section 2 provides insight to how ontologies are used in the field of data governance. Section 3 further discusses the idea of concept interpretation, surveys current literature on similar research, and briefly touches upon the reasoning using ontologies. Section 4 introduces Conto, discusses the architecture of the tool, as well as illustrates the expected benefits of the tool with examples. Finally, Section 5 gives concluding remarks and provides directions for future work.

2. The Role of Ontologies in Data Governance

There exists three major problems that the proliferation of data has caused: a data source is reshaped due to corrective maintenance, data sources lose their application independence over time, and the data is stored in different sources[8]. Ontologies have been employed to remedy these problems by creating an ontological layer over the data sources[4,8]. Within OBDA and Ontology-Based Data Management (OBDM) we see a three layered architecture composed of an ontology layer, the data sources, and a mapping layer between the two.

The purposes of the ontology layer is to provide a declarative approach for information integration and data governance. With the proper conceptualization of the domain, it promotes the re-usability and sharing of the data[9]. The mapping layer in OBDA and OBDM map the data sources to the ontological concepts, which provides proper definitions to the concepts[8,9].

By using ontologies in data governance, the data sources are allowed to evolve with the addition of more data. The ontology layer provides a skeleton of the domain model, which can accommodate the addition of new data sources or elements therein. This results in a system which continues to operate while evolving as more data is integrated and managed. Thus, it can be seen that the role ontologies play in data governance is that of providing a layer to unify the scattered data sources, as well as providing a means to conceptualize the domain knowledge so that it may be interacted with. The purposes of said interactions, as exemplified with OBDA and OBDM, is to access the data and learn new knowledge[4]. However, as mentioned, there exists challenges in properly creating concepts that reflect the domain which ultimately hinders the ability to acquire new knowledge[8].

Download English Version:

https://daneshyari.com/en/article/485354

Download Persian Version:

https://daneshyari.com/article/485354

Daneshyari.com