Chapter 4

Semi–Automatic Ontology Construction by Exploiting Functional Dependencies and Association Rules

Luca Cagliero  
Politecnico di Torino, Italy

Tania Cerquitelli  
Politecnico di Torino, Italy

Paolo Garza  
Politecnico di Milano, Italy

ABSTRACT

This paper presents a novel semi-automatic approach to construct conceptual ontologies over structured data by exploiting both the schema and content of the input dataset. It effectively combines two well-founded database and data mining techniques, i.e., functional dependency discovery and association rule mining, to support domain experts in the construction of meaningful ontologies, tailored to the analyzed data, by using Description Logic (DL). To this aim, functional dependencies are first discovered to highlight valuable conceptual relationships among attributes of the data schema (i.e., among concepts). The set of discovered correlations effectively support analysts in the assertion of the Tbox ontological statements (i.e., the statements involving shared data conceptualizations and their relationships). Then, the analyst-validated dependencies are exploited to drive the association rule mining process. Association rules represent relevant and hidden correlations among data content and they are used to provide valuable knowledge at the instance level. The pushing of functional dependency constraints into the rule mining process allows analysts to look into and exploit only the most significant data item recurrences in the assertion of the Abox ontological statements (i.e., the statements involving concept instances and their relationships).

DOI: 10.4018/978-1-4666-3610-1.ch004
INTRODUCTION

The outstanding growth of both context-aware environments and user-generated content coming from social network communities prompted the investigation of new ways to represent domain knowledge and its relationships. Semantic Web tools provide the instruments to significantly enrich the knowledge representation through a wide range of semantics-based models. These models, often called ontologies, support users in understanding the meaning of a resource and the related domain. Ontologies are fully comprehensive models for describing domain-specific knowledge. Knowledge representation entails (i) shared agreement of meaning, (ii) term disambiguation, and (iii) domain description through concepts and relationships. The contribution of both advanced data mining algorithms and semantics-based knowledge representation may enhance the knowledge discovery process in a broad range of application contexts, such as social behavior analysis, knowledge discovery from user-generated content, and Web service personalization.

Useful ontologies for a given application domain can be either provided by domain experts or (semi-)automatically inferred from the data of interest. Although the Semantic Web already provides a full technological stack to access semantics-based resources, most of the existing approaches, such as the creation of ontologies in a Web Ontology Language like the OWL (World Wide Web Consortium, 2009), still heavily rely on the human intervention. Hence, the machine-driven construction of meaningful ontologies is becoming an increasingly appealing target in several research fields, including information retrieval, data mining, and data summarization. For example, the exponential growth of social media like blogs and social network services has significantly increased the need of useful ontologies to efficiently support the analysis of large data volumes. Thus, novel and more efficient approaches to automatically construct useful ontologies tailored to the analyzed data are desirable.

This paper presents a novel and effective semi-automatic approach to construct ontologies tailored to structured data. Structured datasets are data collections whose content is organized by means of a schema that describes the relevant data features of interest. For instance, a relational dataset schema is characterized by a set of attributes which describes the main data features. Similarly, an XML dataset is characterized by a set of tags (elements). For the sake of simplicity, we will denote the data features belonging to the dataset schema as attributes in the rest of this paper for both relational and XML data. Ontology construction commonly entails a two-level analysis: (i) an intensional data analysis, to represent shared data concepts and their relationships, and (ii) an extensional data analysis, to represent instances of concepts (i.e., the individuals) and their associations. Different approaches can be used to model and represent ontologies. For instance, Description Logic (DL) (Baader, Calvanese, McGuinness, Nardi, & Patel-Schneider, 2003) can be profitably used to represent ontologies. Ontology representation based on Description Logic relies on two main components: (i) the Tbox component, which includes intensional statements about general concepts, and (ii) the Abox component, which includes both extensional statements about the individuals of those concepts and membership statements, which bind instances with their concepts.

To support domain experts in constructing meaningful ontologies, expressed by using the Description Logic and tailored to the input structured data, the DOnG (Dependency-driven Ontology Generator) framework has been designed and developed. It exploits two well-founded database and data mining techniques, i.e., functional dependency discovery and association rule mining. These techniques provide valuable correlations hidden in both the data schema and content. Although these correlations have not the semantic
19 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage:
www.igi-global.com/chapter/semi-automatic-ontology-construction-exploiting/76172?camid=4v1

www.igi-global.com/e-resources/library-recommendation/?id=1

Related Content

Benchmarking in the Semantic Web
www.igi-global.com/chapter/benchmarking-semantic-web/28859?camid=4v1a

Semantic Processing of Monitoring Data in Industrial Applications
www.igi-global.com/chapter/semantic-processing-monitoring-data-industrial/28850?camid=4v1a

Nesting Strategies for Enabling Nimble MapReduce Dataflows for Large RDF Data
www.igi-global.com/article/nesting-strategies-for-enabling-nimble-mapreduce-dataflows-for-large-rdf-data/113712?camid=4v1a

The Development of an Optimised Metadata Application Profile
Paul Walk (2017). Developing Metadata Application Profiles (pp. 16-36).
www.igi-global.com/chapter/the-development-of-an-optimised-metadata-application-profile/175865?camid=4v1a