Chapter 11
Ontology Learning from Thesauri:
An Experience in the Urban Domain

Javier Nogueras-Iso
Universidad de Zaragoza, Spain

Javier Lacasta
Universidad de Zaragoza, Spain

Jacques Teller
Université de Liège, Belgium

Gilles Falquet
Université de Genève, Switzerland

Jacques Guyot
Université de Genève, Switzerland

ABSTRACT

Ontology learning is the term used to encompass methods and techniques employed for the (semi-)automatic processing of knowledge resources that facilitate the acquisition of knowledge during ontology construction. This chapter focuses on ontology learning techniques using thesauri as input sources. Thesauri are one of the most promising sources for the creation of domain ontologies thanks to the richness of term definitions, the existence of a priori relationships between terms, and the consensus provided by their extensive use in the library context. Apart from reviewing the state of the art, this chapter shows how ontology learning techniques can be applied in the urban domain for the development of domain ontologies.

INTRODUCTION

The activity of knowledge acquisition constitutes one of the most important steps at the beginning of the ontology development process. This activity is essential in all the different methodologies for ontology design as a previous step to the conceptualization and formalization phases. As its name indicates, this activity is devoted to gather all avail-

able knowledge resources describing the domain of the ontology and to identify the most important terms in the domain (Gandon, 2002).

To alleviate the work of knowledge acquisition there is an emerging interest in the study of methods and techniques for the (semi-)automatic processing of knowledge resources. The main aim of this automatic processing, known as ontology learning (Gómez-Pérez, Fernández-López & Corcho, 2003; Antoniou & van Harmelen, 2004), is to apply the most appropriate methods to transform unstructured (e.g., text corpora), semi-structured (e.g., folksonomies, HTML pages) and structured data sources (e.g., databases, thesauri) into conceptual structures (Gómez-Pérez and Manzano-Macho, 2003). The methods of ontology learning are usually connected with the activity of ontology population which also relies on (semi-)automatic methods to transform unstructured, semi-structured and structured data sources into instance data (i.e., instances of ontology concepts).

Among all the knowledge resources to be used as an input for ontology learning, thesauri, hierarchical classification standards and such taxonomies are likely the most promising sources for the creation of domain ontologies at reasonable costs (Hepp & de Bruijn, 2007). A thesaurus defines a set of terms describing the vocabulary of a controlled indexing language, formally organized so that the a priori relationships between concepts (e.g., synonymous terms, broader terms, or narrower terms) are made explicit. Additionally, the applicability of thesauri for search and retrieval in digital libraries has promoted the creation and diffusion of well-established thesauri in many different domains. Therefore, thesauri reflect some degree of community consensus and contain, readily available, a wealth of category definitions plus a hierarchy.

During the last years and even within the context of digital libraries and information retrieval, there is a general consensus about promoting the use of more elaborated ontologies. Ontologies

with formal *is-a* hierarchies, frame definitions or even general logical constraints can improve the performance of retrieval systems. As (Fisher, 1998) remarks, the advantage of doing this transformation work between models is that combining formal ontologies with concept-oriented lexical databases can cover a spectrum of functionality which in principle includes all the traditional services of a classical thesaurus, and can offer more. (Soergel et al, 2004) remark that we need to change the use of thesauri into other more formal representations when at least one of the following requirements is needed:

- Improved user interaction with thesauri on both the conceptual and the term level for improved query formulation and subject browsing, and for more user learning about the domain.
- Intelligent behind-the-scenes support for query expansion, both concept expansion and synonym expansion, within one language and across languages.
- Intelligent support for human indexers and automated indexing/categorization systems.
- Support for artificial intelligence and semantic Web applications.

This chapter will be devoted to review the state of the art in ontology learning from thesauri, and to show how these techniques can be applied to practical examples in the urban domain. This domain is a quite interesting one since, although quite technical, it usually involves a number of different disciplines, from architecture to law or transport engineering. Accordingly, it is not obvious to delineate a priori the set of concepts to be captured in an urban ontology which is well reflected by the intricacy and looseness of existing urban thesauri. Furthermore, urban systems have evolved quite rapidly over the last decades due to growing environmental concerns and metropolisation processes. Urban knowledge is hence a very

Related Content

Information Attacks and Defenses on the Social Web

Agostino Poggiand Michele Tomaiuolo (2018). *Global Implications of Emerging Technology Trends (pp. 216-235).*

www.irma-international.org/chapter/information-attacks-and-defenses-on-the-social-web/195831

An Overview of 3GPP Long Term Evolution (LTE)

Elisavet Grigoriouand Periklis Chatzimisios (2015). *Encyclopedia of Information Science and Technology, Third Edition (pp. 6122-6131).*

www.irma-international.org/chapter/an-overview-of-3gpp-long-term-evolution-lte/113069

Parallel and Distributed Pattern Mining

Ishak H.A Meddahand Nour El Houda REMIL (2019). *International Journal of Rough Sets and Data Analysis (pp. 1-17).*

www.irma-international.org/article/parallel-and-distributed-pattern-mining/251898

A Systemic, Participative Design of Decision Support Services for Clinical Research

Alexandra Pomares Quimbaya, Rafael A. González, Wilson Ricardo Bohórquez, Oscar Muñoz, Olga Milena Garcíaand Dario Londoño (2014). *International Journal of Information Technologies and Systems Approach (pp. 20-40).*

www.irma-international.org/article/a-systemic-participative-design-of-decision-support-services-for-clinical-research/117866

Three Parties Engagement of Learning Management System: Students-Lecturer Technology Evidence From Brunei

Fadzliwati Mohiddinand Heru Susanto (2021). *Handbook of Research on Analyzing IT Opportunities for Inclusive Digital Learning (pp. 130-153).*

www.irma-international.org/chapter/three-parties-engagement-of-learning-management-system/278958