Chapter 53

Knowledge Acquisition Through Ontologies from Medical Natural Language Texts

José Medina-Moreira
University of Guayaquil, Ecuador

Oscar Apolinario-Arzube
University of Guayaquil, Ecuador

Katty Lagos-Ortiz
University of Guayaquil, Ecuador

María del Pilar Salas-Zárate
University of Murcia, Spain

Harry Luna-Aveiga
University of Guayaquil, Ecuador

Rafael Valencia-García
University of Murcia, Spain

ABSTRACT

Ontologies are used to represent knowledge and they have become very important in the Semantic Web era. Ontologies evolve continuously during their life cycle to adapt to new requirements and needs, especially in the biomedical field, where the number of ontologies and their complexity have increased during the last years. On the other hand, a vast amount of clinical knowledge resides in natural language texts. For these reasons, building and maintaining biomedical ontologies from natural language texts is a relevant and challenging issue. In order to provide a general solution and to minimize the experts' participation during the ontology enriching process, a methodology for extracting terms and relations from natural language texts is proposed in this work. This framework is based on linguistic and statistical methods and semantic role labeling technologies, having been validated in the domain of diabetes, where they have obtained encouraging results with an F-measure of 82.1% and 79.9% for concepts and relations, respectively.

DOI: 10.4018/978-1-7998-1204-3.ch053

INTRODUCTION

In the biomedical context, where the generation of new knowledge is continuous, ontologies have raised interest in the entire biomedical community (Rubin, Shah, & Noy, 2008). Ontologies are used for organizing and sharing biological knowledge as well as integrating different sources of knowledge in order to provide interoperability among different research communities. For example, MEDLINE system, which is based on the MESH medical ontology, is the largest medical knowledge base available over the Internet. Due to the fact that in medical domain new terms are continually emerging, the update and extension of MESH is necessary to avoid its obsolescence. In the last years, ontologies have been applied to several areas such as natural language processing (Paredes-Valverde, Valencia-García, Rodríguez-García, Colomo-Palacios, & Alor-Hernández, 2015), sentiment analysis (Salas-Zárate, Valencia-García, Ruiz-Martínez, & Colomo-Palacios, 2016), bioinformatics (Mayor & Robinson, 2014), business intelligence (Rodríguez-García et al., 2016), recommender systems (Colombo-Mendoza, Valencia-García, Rodríguez-González, Alor-Hernández, & Samper-Zapater, 2015), and cloud computing (Rodríguez-García, Valencia-García, García-Sánchez, & Samper-Zapater, 2014a) (Rodríguez-García, Valencia-García, García-Sánchez, & Samper-Zapater, 2014b), to mention a few.

Nevertheless, building and maintaining long-living ontologies is still a challenge for the scientific community, which requires a considerable human effort and development time. An ontology lifecycle includes several phases such as creation, maintenance, evaluation, consistency checking, etc. and there are several ongoing attempts to address these issues by providing automatic ontology management methods (Petrucci, Ghidini, & Rospocher, 2016). Among them, several ontology learning and ontology evolution methods have been developed.

Ontology learning has been an active research area that has grown according to the needs of creating ontologies in a fast and reliable way. Ontology Learning (also named ontology generation or ontology extraction) is a knowledge acquisition activity that relies on (semi-) automatic methods to transform unstructured (e.g. corpora), semi-structured (e.g. folksonomies, html pages, etc.) and structured data sources (e.g. databases) into conceptual structures (Petrucci et al., 2016).

On the other hand, ontology evolution (also named ontology enrichment) can be seen as the activity of enriching ontologies by adding or modifying concepts and relations. More specifically, in (Stojanovic, Maedche, Motik, & Stojanovic, 2002) ontology evolution is defined as the timely adaptation of an ontology to modified business requirements, to trends in ontology instances and patterns of usage of the ontology-based application, as well as the consistent management/propagation of these change to dependent elements.

Ontology evolution is considered a process due to the fact that a change in the ontology can cause inconsistencies in other parts of the ontology, as well as in other dependent applications (Abcckcr & Stojanovic, 2005), for that reason, the process of ontology evolution contemplates some tasks such as for example resolving the ontology changes, ensuring the consistency of the underlying ontology and facilitating the changes and management of the ontology to the user (Stojanovic et al., 2002).

This paper focuses on the tasks of extracting terms and semantic relationships from natural language texts in order to simplify the process of building and updating the concepts and relationships in biomedical ontologies. The proposed framework is based on linguistic and statistical term extraction methods and semantic role labeling technologies, having been validated in the diabetes domain. Many biomedical ontologies are structurally quite simple, since they consist of only a few relationship types apart from the taxonomical relationship that constitute the backbone of an ontology. One of the main contributions of this paper is the provision of a method for increasing the typology of semantic relationships.

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/knowledge-acquisition-through-ontologies-from-medical-natural-language-texts/243156

Related Content

Application of Machine Learning and Artificial Intelligence Techniques for IVF Analysis and Prediction

Satya Kiranmai Tadepalliand P.V. Lakshmi (2019). *International Journal of Big Data and Analytics in Healthcare (pp. 21-33).*

www.irma-international.org/article/application-of-machine-learning-and-artificial-intelligence-techniques-for-ivf-analysis-and-prediction/247456

SBASH Stack Based Allocation of Sheer Window Architecture for Real Time Stream Data Processing

Devesh Kumar Laland Ugrasen Suman (2020). *International Journal of Data Analytics (pp. 1-21).*

www.irma-international.org/article/sbash-stack-based-allocation-of-sheer-window-architecture-for-real-time-stream-data-processing/244166

Commercial Banks' Digital Paradigm and Customers Responses in the UAE

Muhammad Jumaa (2020). *International Journal of Data Analytics (pp. 68-79).*

www.irma-international.org/article/commercial-banks-digital-paradigm-and-customers-responses-in-the-uae/244170

The Contribution of Information Science Through Scientific and Technological Knowledge in Intellectual Property

Jorge Lima de Magalhães, Flavia Maria Lins Mendes, Adelaide Maria de Souza Antunesand Zulmira Hartz (2019). *Handbook of Research on Expanding Business Opportunities With Information Systems and Analytics (pp. 265-279).*

www.irma-international.org/chapter/the-contribution-of-information-science-through-scientific-and-technological-knowledge-in-intellectual-property/208568

Hazmat Transport Safety and Alternative Transport Modes: A Study of US Accidents Between 1990 and 2010

Luca Zamparini, Genserik Reniersand Michael Ziolkowski (2018). *Intelligent Transportation and Planning: Breakthroughs in Research and Practice (pp. 544-562).*

www.irma-international.org/chapter/hazmat-transport-safety-and-alternative-transport-modes/197149