Chapter 12
Applications of Ontologies and Text Mining in the Biomedical Domain

A. Jimeno-Yepes
European Bioinformatic Institute, UK

R. Berlanga-Llavori
Universitat Jaume I, Spain

D. Rebholz-Schuchmann
European Bioinformatic Institute, UK

ABSTRACT

Ontologies represent domain knowledge that improves user interaction and interoperability between applications. In addition, ontologies deliver precious input to text mining techniques in the biomedical domain, which might improve the performance in different text mining tasks. This chapter will explore on the mutual benefits for ontologies and text mining techniques. Ontology development is a time consuming task. Most efforts are spent in the acquisition of terms that represent concepts in real life. This process can use the existing scientific literature and the World Wide Web. The identification of concept labels, i.e. terms, from these sources using text mining solutions improves ontology development since the literature resources make reference to existing terms and concepts. Furthermore, automatic text processing techniques profit from ontological resources in different tasks, for example in the disambiguation of terms and the enrichment of terminological resources for the text mining solution. One of the most important text mining tasks that exploits ontological resources consists of the mapping of concepts to terms in textual sources (e.g. named entity recognition, semantic indexing) and the expansion of queries in information retrieval.

INTRODUCTION

The development of ontologies is a time-consuming task. The formal representation of domain knowledge is one important step and the acquisition and confirmation of terms representing relevant concepts is another one. If textual and semantic resources such as the scientific literature can be exploited for the design and development of ontologies, work efficiency can be improved.

Large document collections like the World Wide Web or the biomedical scientific literature (Medline) are readily available, however neither one is currently available in a semantically structured representation. The extraction of information from both sources requires text mining solutions and could be beneficial for the ontology development since such literature resources contain a significant portion of the domain knowledge.

Text mining does not only contribute to the generation of ontological resources, but also text mining profits from the use of ontological resources. For example, the use of an ontology can support the disambiguation of terms that represent different concepts and furthermore, the use of synonyms linked to individual concepts can enlarge the coverage of a text mining solution. The most basic text mining tasks that integrate ontologies into text mining are the mapping of concept labels to terms in textual sources (e.g. named entity recognition) and the expansion of query terms in information retrieval. As shown later, the combination of ontologies with text mining solutions leads to benefits in different IT approaches and their combined exploitation is developing into a dedicated research topic.

In this chapter, we analyze the symbiosis between text mining and the ontological field, namely: which text mining techniques are profitable for ontological development and, how ontologies can enhance the coverage and precision of existing text mining solutions. In this context, we introduce and discuss the different techniques from both fields as well as their interaction. Accord-

ingly, the chapter is divided into two main parts: one part is dedicated to explore the relevance of ontologies for text mining, and the other one to describe the contributions of text mining solutions to the ontology development lifecycle.

The contents of this chapter are specially aimed to bioinformatics and computer science researchers, showing them the state-of-the-art and the new opportunities that are arising from the combination of text mining and ontology-based technology.

USE OF ONTOLOGIES IN TEXT MINING

Text mining is the processing and analysis of data stored in textual representation. Text mining extracts facts from text to fill databases or to improve exploitation of document content through better retrieval or navigation in the document. Text mining consists of two main sub-tasks: information retrieval (IR) and information extraction (IE). IR techniques aim at recovering relevant documents from a large textual repository in order to satisfy the user's information need expressed by his retrieval query. Information extraction solutions glean facts from a set of documents.

In text mining systems, IR and IE are usually interlinked (e.g. Figure 1). IR is used to retrieve relevant documents or parts of the document (e.g., paragraphs or sentences) to be possibly further processed by IE methods. The other way around, IE may feed identified results into an IR system to produce better results. For example, the IR system can generate an enriched index based on the results from the IE system to allow better performance. In the following sections, we present in more detail all involved text mining components and demonstrate different usages and exploitations of the ontological resources to this end.

Related Content

An Efficient Complex Event Processing Algorithm Based on NFA-HTBTS for Massive RFID Event Stream
Jianhua Wang, Shilei Lu, Yubin Lanand Lianglun Cheng (2018). *International Journal of Information Technologies and Systems Approach (pp. 18-30).*
www.irma-international.org/article/an-efficient-complex-event-processing-algorithm-based-on-nfa-htbts-for-massive-rfid-event-stream/204601

ICT Standardization
Kai Jakobs (2018). *Encyclopedia of Information Science and Technology, Fourth Edition (pp. 4679-4691).*
www.irma-international.org/chapter/ict-standardization/184174

Exploration on the Operation Status and Optimization Strategy of Networked Teaching of Physical Education Curriculum Based on AI Algorithm
Yujia Wang (2023). *International Journal of Information Technologies and Systems Approach (pp. 1-15).*
www.irma-international.org/article/exploration-on-the-operation-status-and-optimization-strategy-of-networked-teaching-of-physical-education-curriculum-based-on-ai-algorithm/316892

Challenges to the Development of Information Infrastructures: A Synthesis of the Findings
(2012). *Perspectives and Implications for the Development of Information Infrastructures (pp. 115-135).*
www.irma-international.org/chapter/challenges-development-information-infrastructures/66259

Towards Knowledge Evolution in Software Engineering: An Epistemological Approach
Yves Wautelet, Christophe Schinckusand Manuel Kolp (2010). *International Journal of Information Technologies and Systems Approach (pp. 21-40).*
www.irma-international.org/article/towards-knowledge-evolution-software-engineering/38998