Modeling Domain Ontology for Occupational Therapy Resources Using Natural Language Programming (NLP) Technology to Model Domain Ontology of Occupational Therapy Resources

Ahlam F. Sawsaa, School of Computing and Engineering, University of Huddersfield, Huddersfield, UK
Joan Lu, School of Computing and Engineering, University of Huddersfield, Huddersfield, UK

ABSTRACT

Creation and development of a formal domain ontology of Occupational Therapy (OT) requires the prescription and formal evaluation of the results through specific criteria. Methontology of development ontologies was followed to create OTO ontology, and was implemented by using Protégé-OWL. Accuracy of OTO ontology was assessed using a set of ontology design criteria. This paper describes a software engineering approach to model domain ontology for occupational therapy resources (OT) using Natural Language Programming (NLP) technology. The rules were written to annotate the domain concepts using Java Annotation Patterns Engine (JAPE) grammar. It is used to support regular expression matching and thus annotate OT concepts by using the GATE developer tool. This speeds up the time-consuming development of the ontology, which is important for experts in the domain facing time constraints and high workloads. The rules provide significant results: the pattern matching of OT concepts based on the lookup list produced 403 correct concepts and the accuracy was generally higher. Using NLP technique is a good approach to reducing the domain expert’s work, and the results can be evaluated. This study contributes to the understanding of ontology development and evaluation methods to address the knowledge gap of using ontology in the decision support system component of occupational therapy.

Keywords: Information Extracting, Natural Language Programming (NLP), Occupational Therapy (OT), Ontology, Regular Expression

DOI: 10.4018/ijirr.2013100106
1. INTRODUCTION

Ontologies have become the main components of the semantic web. (Berners–Lee, 2000); the creator of World Wide Web considers the ontology to be a critical part of the semantic web.

The semantic web offers semantic annotations that describe web resources explicitly. These annotations are based on ontologies that represent domain knowledge through defining concepts and the semantic relations between those concepts. It requires standards of machine-processable representations of ontology. The standards for this purpose, such as Resource Description Framework (RDF) (Handschuh & Staab, 2002), Web Ontology Language (OWL) (Gasevic, et al., 2006), have been defined by the World Wide Web Consortium (W3C) (Sawsaa & Lu, 2010, Semantic Web, 2005). Consequently, ontology is a foundation that is central to the growth of the semantic web, that provides a common knowledge for correspondence and communication among heterogeneous systems. Furthermore, it is useful for different applications to share information among heterogeneous data resources (Alberto, et al., 2002).

Recently, Information Extracting (IE) has received significant interest due to the number of web pages emerging on the internet containing unstructured data. Because of the amount of information available on the internet, it has become necessary to have a tool for extracting it. Many specialists in the field of IE have worked to find suitable tools, such as Wrappers, that classify interesting data and map them onto appropriate formats such as XML or relational databases. Furthermore, some HTML-aware tools are based on inheriting the structural features of documents so as to extract the data.

Natural Language Programming (NLP) is a technique used by many tools to extract the data in natural language documents. Tools like GATE use techniques such as part-of-speech tagging, filtering, or lexical semantic tagging to link relevant information, and identify relationships among phrases and sentence elements within text (GATE, 2013). Each of these tools has advantages and disadvantages. A comparative analysis of the existing tools for data extraction is needed to assess their capabilities.

Occupational therapy needs a formal language to identify certain concepts in the field, which are still problematic, to make communication easier. This paper provides a brief background of IE tools to justify the use of NLP technique. It is used to extract concepts in the field and to speed up the process of building the Ontology of Occupational Therapy (OTO); during this study the CREOLE plug-ins from GATE in the IE system is used. It also shows how the JAPE grammar has been implemented by detailing the rules we use to annotate IS concepts. The paper is structured as follows: In Section 2, the background of ontology and IE is discussed. In Section 3, the methods used to develop domain ontology of OT and extracting OT concepts are discussed, to show how they were constructed. Section 4 is the implementation that shows how the domain knowledge is acquired for creating the corpus, Gazetteer, and how the JAPE rule is implemented. Our discussion and evaluation is in Section 5. Finally, in Section 6, we draw conclusions and make suggestions for future work.

2. BACKGROUND

2.1. Information Extracting IE

A number of studies have shown that applications of IE can be used to annotate documents that are written in natural language. Certainly, the growing number of IE tools that can be used to annotate concepts, such as SHOE, Annota, Annozilla, MnM, Ontomat, COHSE, Melita, and GATE, makes it easy to process machine-readable text (Srihari, 2002). A comparison of these tools shows that they provide distinct methods of IE (Alberto et al., 2002; Sawsaa, 2010; Calzolari, 2013), as illustrated in Table 1.

Table 1 shows that there are many tools such as SHOE, Annota Annozilla and KIM ontotext providing automatic annotation of
State-of-the-Art Information Retrieval Tools for Biological Resources
www.igi-global.com/chapter/state-of-the-art-information-retrieval-tools-for-biological-resources/198546?camid=4v1a

MapReduce Based Information Retrieval Algorithms for Efficient Ranking of Webpages
www.igi-global.com/chapter/mapreduce-based-information-retrieval-algorithms/75911?camid=4v1a