Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 127 (2018) 426–435

Procedia Computer Science

www.elsevier.com/locate/procedia

The First International Conference On Intelligent Computing in Data Sciences

Towards a Novel and Generic Approach for OWL Ontology Weighting

Hasna Abioui*, Ali Idarrou, Ali Bouzit, Driss Mammass

Laboratory IRF-SIC, University IbnZohr, Agadir,Morocco

Abstract

Semantic search is qualified – by web-related enterprises as well as, academic research – as a key technology, ensuring important improvements in terms of shared data understanding, while it leads to refined and targeted interpretations. Accordingly, ontologies are the focal asset for a well-functioning semantic search approach, since their ability to share, represent and reuse explicit and semantic domain specification. Nowadays, a multitude of ontologies containing up to hundreds of thousands of concepts are proposed. Thus, our challenge as researchers exceeds conceptualizing or creating ontologies to being able to choose the fitting and suitable one, taking into account specific criteria. This paper comes within the same context as it presents a novel approach for weighting OWL ontologies, in order to choose the most appropriate one from a set of proposed ontologies. Our approach takes into account not only the taxonomic structure, but also the semantic aspect of the ontology. Furthermore, semantic relationships and specific concepts are the favored since they reflect the semantic richness of the ontology.

Keywords: Ontologies; Ontology Weighting; Taxonomic Structure; Semantic Relationships

1. Introduction

Being inspired from ontology – *the science* referring to a fundamental metaphysic branch in philosophy, studying the general properties of what exists –, ontology *the object*, appeared in the early 1990s, and adopted by the computer science community. Since then and, over the years, the notion of ontology becomes the focal point and interest of many computer science research works in several fields, due to its unavoidable role as a means to share, represent and reuse explicit and semantic domain specification. It's qualified as a key technology for the success of not only historical communities that saw it born, namely, knowledge engineering and artificial intelligence, but also various recent applications, essentially, natural language processing, information retrieval and semantic web, where the use of ontologies becomes a very common practice. During the past few years, the conception and development of ontologies

* Corresponding author. Tel.: +212-670-581-533.
E-mail address: hasna.abioui@gmail.com

10.1016/j.procs.2018.01.140

are no longer limited to artificial intelligence laboratories, as long as, domain experts may simply do such tasks, since a variety of editors dedicated to create and edit ontologies exist, like Protégé [1], Swoop [1], OntoEdit [2], etc. Hence, a multitude of either domain ontologies or generic ontologies containing up to hundreds of thousands of concepts are proposed. For instance, biomedical community produced large standard and structured vocabularies, offering a common biomedical terminology which can be shared and reused across various fields. We name some of the most powerful ontologies in clinical and biomedical domains, such as, the SNOMED-CT project [3], which is the result of merging two other ontologies: SNOMED-RT and Clinical Terms, also, GALEN[4], the European project offering the terminology reusability in clinical systems, then, Gene Ontology GO [5], representing a biological terminology which gathers all biological processes, concerning cells, molecules and their components, etc. Thus, either creating or having an ontology regardless of the purpose of its use or the application in which it will take part, is no longer a challenge as well as having the fitting and suitable one for that purpose. Especially, when it comes to choosing between more than one ontology for a specific domain. To this point, finding the most consistent, coherent and rich one becomes the challenge. Against this background, some initiatives stressed the importance of carefully choosing the most suitable ontology among others for a specific application, by setting indicators and criteria that must be taken into account when selecting an ontology. In [6], authors believe that choosing efficiently the appropriate ontology is relative to the compliance with semantic web standards, the relevance of metadata and the quality of the vocabulary describing contents. While [7], [8] and [9] propose ontology evaluation and ranking tools, according to either a set of terms representing the semantic content or a set of metrics describing various aspects and characteristics of ontologies. As can be noted, there are no established criteria to adopt when choosing an ontology, it may depend on the objectives behind the use of the ontology or also the richness of this latter translated by the number of concepts, relations between them, axioms, structures and levels of granularity. In front of all these points we stated, facilitating the ontology selection phase and make it more targeted, in addition to the huge number of ontologies existing for different fields, the need to find the best method for selecting an ontology remains a priority. In this paper, we propose a new strategy aimed to evaluate the importance of an ontology compared to the others from the same application fields. The approach is based on a novel OWL ontology weighting method that we will present too and which exploit all provided information by the ontology, namely, structural and semantic data. The remainder of the paper is organized as follow. The second section sheds light on the general background of the research work and relative works; the third section defines the approach objectives, based on the limits of previous work, then describes the ontology weighting approach with pre-processing and implementation details. Finally, the paper ends with the description of the performed experiments.

2. Background and Related Work

This section provides background information on both main focal points of our research, namely, the concept of ontology, the ontology weighting and an overview of related works.

2.1. Ontology

During the last years, the use of ontologies becomes more prevalent and commonplace – notably, in heterogeneous environment – where it helps to manage and solve heterogeneity problems in case of concept misperception [10] on account, generally, of different data structures and semantic misinterpretation. Thus, ontologies are described as being the practical uniform conceptual model, providing sets of knowledge presented as structured and representative data models of a specific domain. They allow communication, interchanging and interoperability among diversified data systems, offer characteristics and features that promote the reuse and the sharing of data and also, facilitate information retrieval by integrating semantics. All this in order to disambiguate meanings of concepts into resources through using classes, instances of classes, relations between classes, axioms and finally functions.

By digging into the literature, various definitions of the term *ontology* were proposed by scientific communities. [11] links this variety to the multiple inheritances of which the ontology disposes of, that are summarized in four inheritance types: (1) The philosophic inheritance, treating the study of *being* and its properties. (2) The logical inheritance,

[1] Protégé is available at https://protege.stanford.edu/