Neurocomputing 146 (2014) 276–290

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A multi-attribute based framework for ontology aligning

Marcin Pietranik [a], Ngoc Thanh Nguyen [a,b,*]

[a] Division of Knowledge and System Engineering for ICT, Ton Duc Thang University, Ho Chi Minh city, Vietnam
[b] Institute of Informatics, Wroclaw University of Technology, Wroclaw, Poland

ARTICLE INFO

Article history:
Received 13 November 2013
Received in revised form
14 February 2014
Accepted 22 March 2014
Available online 28 June 2014

Keywords:
Ontology alignment
Ontology mapping
Knowledge management

ABSTRACT

This paper is a comprehensive description of our proposal of ontology alignment framework and novel evaluation procedure. The main contribution is providing precise definitions of ontologies on every level of their granularity. On top of such foundations we have built a set of consistent algorithms that allow designating clear matches between two ontologies.

The difference between our solution and former approaches to this task is in accepting concepts' attributes and their varying semantics as the grounding level of every workflow that has been defined. Based on the conducted experiment, we can claim that such an approach is useful and reliable. Due to the fact that created framework is the complete redefinition of the problem, we could not incorporate widely used experimental procedure developed by Ontology Alignment Evaluation Initiative. This is the reason why we have created a novel approach to alignment evaluation based on statistical SignTest that can be conveniently used to emphasise differences in quality of designated mappings.

1. Introduction

The following paper is a comprehensive and detailed presentation of our novel ontology alignment framework. This topic is widely discussed in the literature and has its origins in database schema matching. Basically, it can be understood as formulating a method of matching certain parts of some ontology into selected part of some other ontology that somehow are related with each other. This relatedness is tightly connected with the informal definition of ontologies (taken from Thomas Gruber's work [1]) that describes them as an explicit specification of conceptualisation. Such definition assumes that the conceptualisation is the description of how certain universe of discourse is built and how it can be decomposed into elementary objects that interact with each other. The actual specification is the way that this conceptualisation can be expressed in terms of used names, defined relations, level of granularity, etc.

Such an approach implicates that there are many different specifications of the same conceptualisation. Therefore, ontology alignment is the task of designating how different ontologies specify the same parts of underlying conceptualisation. It is a widely discussed topic, very popular in the recent literature. It deals with the problem of selecting these parts of different and independent ontologies that tightly refer to the same part of reality expressed within them. Due to the fact that ontologies are commonly stored using OWL standard [2], the task of designating alignments between them has been frequently simplified to selecting these objects, which are the most similar in terms of used OWL tags. In other words, concerned problem has been reduced to calculating similarities between some primitive elements of OWL syntax – if for given pair of objects this value is higher than accepted threshold, then such pair is identified as "matchable".

Let us assume the existence of two computer systems illustrated in Fig. 1 named respectively *KB* (which utilises some source ontology *O*) and *KB'* (that uses a target ontology *O'*). Further, imagine that some user sends a query to the system *KB'*, but the answer for his request is present not in *KB'*, but in *KB*. In order to fulfill user's requirement there must be a method of designating which fragments of the source ontology correspond to fragments of the target ontology concerned by the query. Therefore, it is necessary to provide a tool that could reliably transform a content of the source ontology into a content of the target ontology. In other words – to provide a way of communication between *KB* and *KB'*. By ontology alignment we understand a method of designating elements of the source ontology that match elements of the target ontology without violating their separation and heterogeneity. Moreover, a user may ask about any element of ontology he wants and pragmatically he is not interested in generalisation, equivalency or similarity between elements of different ontologies. His goal is to obtain the answer for his query, so providing as

* Corresponding author at: Institute of Informatics, Wroclaw University of Technology, Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland.
E-mail addresses: marcin.pietranik@pwr.wroc.pl (M. Pietranik), ngoc-thanh.nguyen@pwr.wroc.pl (N.T. Nguyen).

http://dx.doi.org/10.1016/j.neucom.2014.03.067
0925-2312/

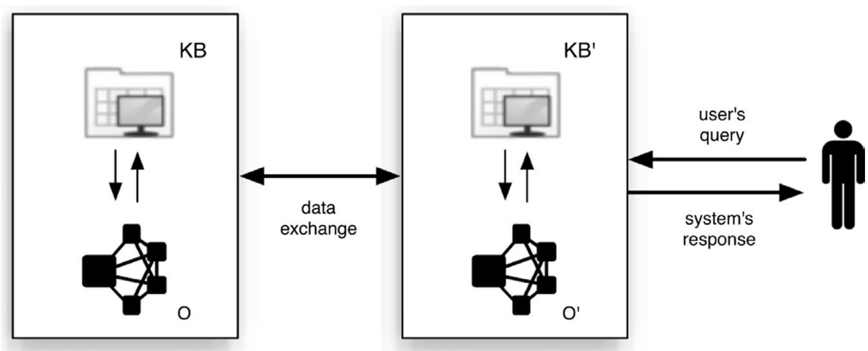

Fig. 1. Case study of ontology integration.

much matching content form the source ontology should be the main purpose of any alignment framework. Our main motivation was not designating literally the most similar concepts, but fulfilling the end user's requirements.

In our work we propose a broader view on ontologies itselves, with careful analysis of their content and how different methods of expressing assumed part of reality relate with each other. It is the direct extension of our previous work [3] enhancing it with aligning every level of ontology granulation, consistent theoretical foundations and detailed experiment's design.

During the initial work we have started with the foundations of the topic by analysing what information can be defined with ontologies and how this information should be expressed. The structuring of our remarks has formed "ontology stack" presented in Fig. 2.

On the highest level of presented diagram we have put "the ontology structure". This element contains definitions of relations between concepts, so description of how these concepts interact with each other. Below there are blocks concerning aforementioned concepts (which are building blocks of considered domain) and their attributes. Concept by itself poses their semantics that are strict implication of relations defined on the higher level of the stack. Attributes are characterised by values they can take. The association of the two is placed in the "attributes's semantics" block of the stack. On this level it is possible to clearly define varying meaning that particular attributes may obtain while being assigned to different concepts. For example, an attribute "*date*" has different semantics if included within a concept "*Person*" and different in a concept "*Book*". This diversity will be formally defined as a function that assigns logic statements to every inclusion of attributes in concepts.

According to presented ontology stack and accepted basic notions (that will be described in Section 3) the ontology alignment framework can be considered complete only if it allows for mapping ontologies on a level of attributes, concepts, relations and instances. We will describe our approach to designating ontology mappings on these four levels, basing solely on created theoretical foundations and a notion of a degree to which it is possible to align elements taken from considered levels of granularity. We will present four different functions (namely $\lambda_A, \lambda_C, \lambda_R$ and λ_I) that are used to calculate these values.

The reminder of the paper is as follows. Section 2 contains an overview of former works that have been done in the field and gives solid background for our work. In Section 3 we will present basic definitions and formal foundation that we have accepted throughout the whole methodology. In Section 4 the set of algorithms for aligning ontologies on attribute, concept, relation and instance levels will be presented. Its first two parts serve as a reminder and brief overview of aligning ontologies on attribute and concept levels. These elements have been described in our

<table>
<tr><td colspan="4">ONTOLOGY STRUCTURE</td></tr>
<tr><td colspan="2">CONCEPTS</td><td colspan="2">ATTRIBUTES</td></tr>
<tr><td>CONCEPTS' SEMANTICS</td><td colspan="2">ATTRIBUTES' SEMANTICS</td><td>ATTRIBUTES' VALUATIONS</td></tr>
</table>

Fig. 2. Ontology stack.

previous publication [3] and are included due to clarity reasons. Also matching on relation level has been introduced earlier in [4], but the following paper is a broader description of created procedure with emphasis put on embedding it within other elements of the framework. Section 5 includes the overview of aforementioned evaluation procedure along with the presentation of gathered experimental results. The paper ends with brief conclusions and a description of planned future works.

2. Related works

As aforementioned in the previous section, ontology matching is a broad and complex task, widely discussed in recent literature. First publication concerning this topic has appeared more than 10 years ago [5], having its origins coming from database schema matching [6]. Up until now the most extensive overview of basic methods, that allow to designate mappings of ontologies, can be found in [7], while recent advances and future challenges have been described in [8]. Authors cover variety of fundamental approaches to analysing commonalities between two or more ontologies, which include among others:

- *String-based approaches* built on top of selection of functions that calculate similarities between strings. This group of functions can be further used to select elements present in ontologies that share closely related descriptions (e.g. names). Despite obvious flaws, such methods are widely incorporated into complex aligning systems, for example: AgreementMaker described in [9].
- *Language-based approaches* analyse the grammar of natural languages expressions. This group of methods spread across tokenisation, stemming or removing stop words from descriptions of objects taken from compared ontologies. Eventually this method serves as a pre-step for string-based methods.
- *Structure-based approaches* are the group of methods that can be understood twofold. The first understanding includes analysing relations (both taxonomical and non-taxonomical) between concepts [10]. Such an approach includes calculating similarities between defined hierarchies and concepts' connections. The second understanding of this group is not a method of designating similarity between elements of ontologies, but propagating partial similarities (e.g. obtained from string-based approach)