An Integer Linear Programming-Based Method for the Extraction of Ontology Alignment

Naima El Ghandour, LITIO Laboratory, Department of Computer Science, Université Oran1, Algeria
Moussa Benaissa, LITIO Laboratory, Department of Computer Science, Université Oran1, Algeria
Yahia Lebbah, LITIO Laboratory, Department of Computer Science, Université Oran1, Algeria

ABSTRACT

The Semantic Web uses ontologies to cope with the data heterogeneity problem. However, ontologies become themselves heterogeneous; this heterogeneity may occur at the syntactic, terminological, conceptual, and semantic levels. To solve this problem, alignments between entities of ontologies must be identified. This process is called ontology matching. In this paper, the authors propose a new method to extract alignment with multiple cardinalities using integer linear programming techniques. The authors conducted a series of experiments and compared them with currently used methods. The obtained results show the efficiency of the proposed method.

KEYWORDS

Alignment, Constraint Programming, Heterogeneity, Instance-Based Approach, Ontology, Ontology Matching, Probabilistic Approach, Semantic Correspondences, Semantic Web

INTRODUCTION

With the advent of the web and Internet, the number of users has increased exponentially, generating both large and heterogeneous sets of data (Ouali, et al., 2019). In order to deal with data heterogeneity, the need for the creation and implementation of ontologies becomes more critical, as they provide conceptual models to represent and share knowledge.

Over the years, many definitions have been given on the Ontology term. The closest definition of ontology in the context of this article is proposed in (Studer et al., 1998), who defined ontology as follows: “An ontology is an explicit, formal specification of a shared conceptualization of a domain of interest.” From this definition and considering the purposes of this article, we adopted the definition given in (Rinaldi et al., 2020), where an ontology is considered as a formal definition of concepts and relationships between them, belonging to a given domain.

In the past two decades, ontologies have been the silver bullet technique to represent the semantics of information (Rinaldi et al., 2020), as they are a practical means to conceptualize what is expressed in a computer format.

DOI: 10.4018/IJITWE.2021040102

Copyright © 2021, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
Gradually, traditional search is being replaced by semantic search using ontology based techniques in reading and processing semantic information for effective presentations of query results to the user (ARIS, 2019).

Ontologies are seen as the solution to data heterogeneity on the semantic web. However, these ontologies could themselves introduce heterogeneity. This heterogeneity may occur at the syntactic, terminological, conceptual, and semantic levels (Euzenat & Shvaiko, 2013). Resolving this semantic heterogeneity problem requires identifying semantic correspondences between entities (concepts, instances and properties.) of various ontologies. This process is commonly known as ontology matching and its output as alignment (Euzenat & Shvaiko, 2013).

The goal of ontology matching is to cope with heterogeneity between them. It has a number of application fields as ontology engineering, information integration, peer-to-peer information sharing, web service composition, autonomous communication systems, navigation and query answering on the web (Rinaldi et al., 2020).

Ontology matching generates a set of correspondences, where each single correspondence links two entities from two given ontologies (source ontology, target ontology), and stands for equivalences as well as other relations, such as subsumption, or disjointness. It may also include a degree of confidence (Euzenat & Shvaiko, 2013).

The matching process can be perceived as a process of two successive steps. The first step consists of generating a set of hypothetical matches (an initial alignment) between the entities from both ontologies which is realized generally by applying diverse techniques. The second step consists of extracting the final alignment where the authors contribution is situated (Meilicke & Stuckenschmidt, 2015).

Constantly increasing size and number of ontologies made matching process impossible to be done manually. For that reason, automatic techniques have to be developed to reduce the burden of manual creation and maintenance of alignments (Asim, et al., 2018).

In recent years, a lot of research work on approaches have proposed such alignments, see (Ochieng & Swaib, 2018; Shvaiko & Euzenat, 2011) for state of the art of some recent contributions. These approaches take advantage of various properties of ontologies (names, relationships, instances) and use techniques from different fields, e.g., statistics and data analysis, machine learning, automated reasoning, and linguistics, and most systems often combine these approaches (Euzenat & Shvaiko, 2013).

However, despite the many tools that have been developed so far, and due to the growing number and size of ontologies, finding a robust ontology matching process remains a challenge. This is due to a limitation of the majority existing similarity-based systems, which consists of using similarity measures based on string, linguistic, structure and instance techniques (e.g. Edit distance, WordNet similarity (Pedersen et al., 2004), SimRank (Jeh & Widom, 2002), and Jaccard similarity (Niwattanakul et al., 2013)) in order to identify the matches between ontologies.

On the other hand, the approximate nature of the similarity makes the majority of the identified matches partially correct. Consequently, the alignment extracted from these matches would be incomplete.

Additionally, the uncertainty inherent in the problem of matching leads us to hypothesize that a probabilistic approach is better suited than an approach based on similarity.

Therefore, the objective of this paper is to present a new method to extract the final alignment. This method considers a probabilistic interpretation of the alignment. More specifically, the hypothetical matches generated during the first phase are assumed to be probabilistic. Therefore, this method is suitable for all systems which generate a probabilistic alignment.

The contributions of this paper consist of:

- Presenting a method based on integer linear programming, to extract an optimal final alignment (best alignment) from hypothetical probabilistic matches, generated by a probabilistic approach;