Applied Soft Computing 11 (2011) 1908–1923

Contents lists available at ScienceDirect

Applied Soft Computing

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

Merging domain ontologies based on the WordNet system and Fuzzy Formal Concept Analysis techniques

Rung-Ching Chen*, Cho-Tscan Bau, Chun-Ju Yeh

Department of Information Management, Chaoyang University of Technology, 168, Jifong E. Rd., Wufong Township, Taichung County 41349, Taiwan, ROC

ARTICLE INFO

Article history:
Received 19 September 2008
Received in revised form 10 February 2010
Accepted 19 June 2010
Available online 6 July 2010

Keywords:
Ontology merge
Ontology mapping
WordNet
Formal Concept Analysis
Fuzzy ontology

ABSTRACT

Many different contents and structures exist in constructed ontologies, including those that exist in the same domain. If extant domain ontologies can be used, time and money can be saved. However, domain knowledge changes fast. In addition, the extant domain ontologies may require updates to solve domain problems. The reuse of extant ontologies is an important topic for their application. Thus, the integration of extant domain ontologies is of considerable importance. In this paper, we propose a new method for combining the WordNet and Fuzzy Formal Concept Analysis (FFCA) techniques for merging ontologies with the same domain, called FFCA-Merge. Through the method, two extant ontologies can be converted into a fuzzy ontology. The new fuzzy ontology is more flexible than a general ontology. The experimental results indicate that our method can merge domain ontologies effectively.

1. Introduction

While the development of World Wide Web has allowed people to easily access information, but still have many drawbacks remain. Due to large quantities of information available, locating the desired information can be time consuming. The next generation of the web requires a convenient and efficient way to improve this situation. In 1998, Berners-Lee proposed the notion of Semantic Web [1]. The core technology of a Semantic Web is an artifact called ontology, and the Semantic Web relies on formal ontologies to structure data for machine understanding. In other words, the Semantic Web is a mode of communication between machine and users. The construction of domain ontologies relies on different experts, different tools, different techniques, and different languages. Domain ontologies include many differences and conflicts even though ontologies exist in the same domain. Moreover, the extant domain ontologies may need to be updated to solve domain problems. Thus, the manner of integrating ontologies is a very important issue.

Gruber indicated that an ontology is an explicit specification of a conceptualization [2]. Hendler defined an ontology as a set of knowledge terms, including vocabulary, semantic interconnections, and some simple rules of inference and logic for a particular topic [3]. Ontology is a formal conceptualization of the real world. In general, an ontology consists of concepts, attributes, relations, operations, axioms, and instances. However, an ontology lacks information to deal with uncertainty in problems. Fuzzy ontology is better suited to describe domain knowledge for solving uncertainty reasoning problems [4–10]. An example is shown in Fig. 1. The concept of Ball has the attributes of color and size; the ball has the operation of throw and kick to be controlled; the ball has an axiom of entity; the ball has the instances of baseball, basketball and volleyball; and the ball has relations to each element. If the gray boxes have fuzzy information in the part-of relationship between the *Equipment* and *Game Equipment*, the *Ball* might be similar to the *Game Equipment*.

Fenza et al. [10] presented a hybrid framework for achieving a fuzzy matchmaking of Semantic Web services. They indicated the matchmaking activity exploits a mathematical model, the fuzzy multiset to suitably represent the multi-granular information. Lee et al. [6,7] presented two methods: an ontology-based computational intelligent multi-agent system and an ontology-based intelligent decision support agent (OIDSA) for Capability Maturity Model Integration (CMMI) assessment. They used ontology model to represent the CMMI domain knowledge that will be adopted by the computational intelligent multi-agent. So, the CMMI ontology is predefined by domain experts, and created by the ontology generating system. Their experimental results indicate that the ontology-based computational intelligent multi-agent can effectively summarize the evaluation reports for the CMMI assessment [6]. They also utilized the fuzzy inference agent computes the similarity of the planned progress report and actual progress report, based on the CMMI ontology, the project personal ontology, and natural language processing results. Their experimental results show that the OIDSA can work effectively for project mon-

* Corresponding author. Tel.: +886 4 23323000x7701; fax: +886 4 23304902.
E-mail addresses: crching@cyut.edu.tw, crching@mail.cyut.edu.tw (R.-C. Chen).

1568-4946/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.asoc.2010.06.007

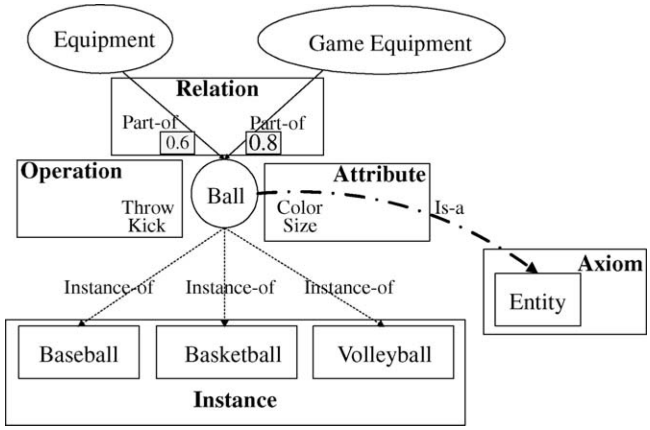

Fig. 1. An example of fuzzy ontology.

itoring and control of CMMI [7]. Reformat and Ly proposed an application of ontology, in the sense of the Semantic Web, for development of computing with words based systems capable of performing operations on propositions including their semantics. The ontology-based approach is very flexible and provides a rich environment for expressing different types of information including perceptions. It also provides a simple way of personalization of propositions [9]. Calegari and Farina presented a concept network based on the evolution of a dynamical fuzzy ontology. A dynamical fuzzy ontology can manage vague and imprecise information. Fuzzy ontologies were defined by integrating Fuzz Set Theory into ontology domain, so that a truth value is assigned to each concept and relation. They examined the case where the truth values change in time according to the queries executed on the represented knowledge domain [8].

The integration of ontologies has become an important research subject in recent years. Many tools [11–13] and techniques [14,15] have been used to construct domain ontologies, however various ontologies are constructed using different tools or techniques despite existing at the same domains. Nevertheless, experts use various tools and languages to create an ontology that will have different architectures even though having the same or similar concepts in the domains. Therefore, ontology integration becomes an important task. Choi et al. [16] have divided the ontology into global ontology, local ontology, and domain ontology. In their definition, a global ontology has a domain topic schema lacks detailed content. It also names upper ontology such as Cyc (Cycorp) [17], SUMO (Suggested Upper Merged Ontology) [18] and WordNet [19]. The scale of local ontology is smaller than that of global ontology. Local ontology has topic schema in a certain specific field, and the local ontology may belong to different domains. Domain ontology has plentiful or particular information to apply to specific task. In this paper, we focus on the merging of domain ontologies having the same domain topics.

The rest of the paper is organized as follows. Section 2 outlines related research. Section 3 describes the FFCA-Merge method. Experiments and discussions are presented in Section 4. Finally, Section 5 presents conclusions and future work.

2. Related research

2.1. Domain ontology merge and alignment

The field of domain ontology merging and alignment generally includes the following six strategies [20]. (1) *Strategies based on linguistic matching*: They complete an integration task according to the linguistic meaning of words. (2) *Structure-based strategies*: They adopt the structural information relating super-concepts and sub-concepts. (3) *Constraint-based strategy*: They complete the integrative task according to the constraints in each concept. (4) *Instance-based strategy*: They complete the integrative task according to the external instance from a database or the Internet. (5) *Auxiliary-based strategy*: They use auxiliary information such as WordNet to complete the integrative task. (6) *Hybrid-based strategy*: They combine different strategies to complete the integrative task.

Noy and Musen [21] proposed a SMART system to match similar class names based on linguistic characteristics. The system is semi-automatic and makes suggestions to users to resolve conflicts. In addition, Noy and Musen [22] proposed a PROMPT system based on linguistics and structural knowledge to determine similarities between terms. PROMPT is a semi-automatic system that generates a list of suggestions for users. Next, Noy and Musen developed an Anchor-PROMPT system [23] to identify pairs of related terms from source ontologies. The system then compares the terms to identify similar terms and to generate a set of new pairs of similar terms. Anchor-PROMPT is a semi-automatic method. After analysis of terms, it can display the results, allowing users to merge source ontologies. Chalupsky's OntoMorph system [24] provides a powerful rules language for ontology merging and generating a knowledge-base translator. OntoMorph is also a semi-automatic system. However, users may require more background knowledge to use the powerful rules language for ontology mapping as it does not facilitate use of the initial list. McGuinness et al. proposed a Chimaera system [25]. The system is an ontology merging tool based on an ontolingual ontology editor. If the linguistic match can be found, ontology merging can be achieved automatically. The system is difficult to implement automatically because it is not easy to find all linguistic matches from uncertain data.

Ichise et al. have developed a HICAL (Hierarchical Concept Alignment system) system [26] that provides a concept hierarchy management for ontology merging and alignment. The method is semi-automatic and uses machine-learning techniques to align multiple concept hierarchies; it also exploits data instances in overlap mapping. The drawback of the system is that categorizing different words under the same concept may result in ambiguity. FCA-Merge (Formal Concept Analysis Merge) [27] consists of the following three steps: (1) Extract instances from documents and generate formal contexts. (2) Output the pruned concept lattice by the Titanic algorithm. (3) Establish the merged ontology based on the pruned concept lattices.

FCA-Merge is also a semi-automatic method. The system translates concept lattices into the merged ontology, but a domain expert must make revisions manually. The CMS (CROSI Mapping System) system [28] consults external linguistic resources, feature selections, multi-strategy similarity aggregators, and similarity evaluators. The system is automatic but suffers from the drawback that while it provides many techniques for the user, and user can choose one or more techniques for mapping ontologies, the user may not know how to choose suitable techniques for specific situations.

The ontologies merging and aliment, mentioned above, have two drawbacks: (1) Such techniques are hard to achieve automatically. (2) The operation of merging ontologies lacks fuzzy information. In this paper, we use FFCA [29,30] and WordNet techniques to merge and align domain ontology automatically. After the merging operations, the system translates the merged ontology to a fuzzy ontology. The membership values of the fuzzy concepts can be used in flexible applications and can process uncertain information. FCA-Merge is an ontology merging method depending on the FCA, and our system improves upon FCA-Merge to generate a fuzzy ontology. The experiments indicate our method is useful.