Expert Systems with Applications 39 (2012) 13085–13092

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Using agents to parallelize a medical reasoning system based on ontologies and description logics as an application case

Alejandro Rodríguez-González [a,*], Javier Torres-Niño [a,1], Gandhi Hernández-Chan [a,1], Enrique Jiménez-Domingo [a,2], Jose Maria Alvarez-Rodríguez [b,3]

[a] Computer Science Department, Universidad Carlos III de Madrid, Av. Universidad 30, Leganés, 28911 Madrid, Spain
[b] Computer Science Department, Universidad de Oviedo, C/Calvo Sotelo, S/N, Oviedo, 33007 Asturias, Spain

ARTICLE INFO

Keywords:
Parallel
Distributed
Reasoning
Expert system
Ontologies
Semantic web
Medical diagnosis

ABSTRACT

Reasoning over ontologies is a hard task where the efficiency in terms of time usually depends on the ontology expressivity and the reasoner used. Depending on the domain and the techniques used to design the ontology, this fact could be a decisive factor to decide if the reasoning platform is useful or not. In this paper we describe an application case of a multi-agent architecture to parallelize a medical diagnosis system based on ontologies and description logics. The aim of this paper is to show the improvement of the time efficiency of the system in which is based this work using the proposed architecture. Results show that the use of the new architecture (which is based in the use of a master-slave architecture and the use of multi-threading) shows that global efficiency can be highly improved.

1. Introduction

The creation of systems which perform medical diagnosis has been increased in the last years if we check the growing number of research papers which have been published recently. The researches presented in these papers point out several types of techniques which have been studied and used trying to know which technique is more suitable for this domain. Some examples of these types of technologies applied to the design and development of medical diagnosis systems are rule-based systems (Barnett, Cimino, Hupp, & Hoffer, 1987; Graber & Mathew, 2008) database-based systems (Aronson, 1997), probabilistic systems (Charniak, 1983; Ganeshan, Johnson, Shaw, & Wood, 2000), probabilistic and rule-based systems (Buchanan, 1984), logic representation (Adlassing, Kolarz, Scheithauer, Effenberger, & Grabner, 1985), fuzzy logic (Chen, 1994), genetic algorithms (Anastasio, Yoshida, Nagel, Nishikawa, & Doi, 1998; Garrell i Guiu, Golobardes i Ribé, Bernadó i Mansilla, & Llorà i Fàbrega, 1999; Vinterbo & Ohno-Machado, 2000), neural networks (Alves, Neves, Maia, & Nelas, 2001; Brause, 2001; Gil, Johnsson, Garcia Chamizo, Paya, & Fernandez, 2009; Temurtas, Yumusak, & Temurtas, 2009; Yan, Jiang, Zheng, Peng, & Li, 2006) or other artificial intelligence techniques (Çomak, Polat, Güneş, & Arslan, 2007; Huang, Chen, & Lee, 2007; Marinakis, Marinaki, & Dounias, 2008).

In spite of this, the use of Semantic web and related technologies is not very common. There are some works (Djedidi & Aufaure, 2007; Podgorelec, Grašič, & Pavlič, 2009) but most efforts in Semantic web are focused in the knowledge representation field. Nevertheless, even given some well-argued opinions about if Semantic web is ready for healthcare domains (Wroe, 2006), it is necessary to study the Semantic web underlying technologies like description logics in the diagnosis domain to know if its appliance could be a good idea, even when there are some authors which clearly states that description logics are not enough (Werner Ceusters, Barry Smith, & Jim Flanagan, 2003) to some applications.

In the study of the appliance of description logics to medical diagnosis domain is not only necessary to know if the modeling which description logics allow is enough to design an accurate system. We also need to know if the performance that this technology offers does not represent a problem. For this reason, in this paper we present a complete study over a current system (Rodríguez-González et al., 2011) based in description logics using parallelization and multithreading with the aim of increasing its performance and answer the question about the suitability of multithreading and parallelization on this kind of systems.

The remainder of the paper is as follows. Next section (2) outlines relevant literature in the area. After background section, in Previous Work (3) section explains the work in which is based the current research. Problem Definition and Solution (4) presents the problem to address and how will be the problem tackled.

* Corresponding author. Tel.: +34 91 624 6257.
E-mail addresses: alejandro.rodriguez@uc3m.es (A. Rodríguez-González), javier.torres@uc3m.es (J. Torres-Niño), gandhi.hernandez@uc3m.es (G. Hernández-Chan), enrique.jimenez@uc3m.es (E. Jiménez-Domingo), josem.alvarez@weso.es (J.M. Alvarez-Rodríguez).
[1] Tel.: +34 91 624 5936.
[2] Tel.: +34 91 624 6257.
[3] Tel.: +34 98 510 3397.

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.05.093

In results section (5) we will show the outcomes which we have obtained with the proposed system and the conclusions. Finally, future work (6) is discussed.

2. Background

There are several works which have dealt with the problematic of parallelizing and distributing the computing in reasoning systems. Some of them have focused their efforts in the improvement of time efficiency; some others have tried to process big amounts of data, but all of them with the specific aim of obtaining better results.

Guo and J. (2006) present a description of an approach based on a large OWL ontology in order to specify kinds of reasoning can be performed separately in each partition and then combine results in order to achieve complete answers. They deal with the problem of increasing expressivity and thus increased reasoning complexity. Their most important goal is to develop an approach that helps existing systems to overcome memory limitations by adopting a divide-and-conquer approach. It means partition an OWL ABox in smaller pieces that could be processed separately but guaranteeing the completeness of the reasoning with the combination of the results.

Maedche, Motik, and Stojanovic (2003) presents an integrated framework for managing multiple and distributed ontologies on the Semantic web which is based on representation model for ontologies. They provide features for reusing ontologies and for evolving them while retaining the consistency. In their work they present a conceptual modeling approach called OI models and now to deal with multiple and distributed OI models.

In the work of Li, Zeng, Kotoulas, Urbani, and Zhong, (2009) they present three parallel applications that cope with performance problems on Semantic web: LarKC, MaRVIN, and Reasoning-Hadoop. They argue that given the enormous amount of data in the web and its distribution it is necessary to parallelize the reasoning process besides developing new forms of reasoning. LarKC is an open architecture and a generic platform for massive distributed reasoning that emphasizes on scalability through the execution of an open set of software components (plug-ins). It works as a scalable workflow engine for reasoning tasks where in each workflow it is possible to find several components which are responsible for diverse processing tasks, and ten they can be distributed among several nodes and work in parallel. MaRVIN is a parallel and distributed platform for processing large amounts of RDF data. It brings forward a method called divide-conquer-swap to do inference tasks, and deals with the problem that the Semantic web data is hard to partition using the traditional divide-and-conquer strategies. The third application is a parallel rule-based RDFS/OWL reasoning system called Reasoning-Hadoop. Hadoop is an open source framework mainly used for massive parallel data processing that implements the MapReduce programming model. In the Reasoning-Hadoop project reasoning has been implemented with a sequence of MapReduce algorithms that includes data and rule partitioning.

Bock (2008) makes the observation that the available reasoners do not exploit the benefits of parallel computation techniques and proposes two possibilities for applying parallel computation techniques to ontology reasoning. The first one is based in working with independent ontological modules and parallelizes its processing, but the challenge for reasoning on different parts of ontologies in parallel is to identify those modules that do not influence each other in terms of conclusions that can be derived by not considering other modules. The second one is working with parallel architectures and adjusts the reasoning algorithms to them. It consist to take advantage of multi-core and shared-memory machines to overcome the communications problem of parallel reasoning by providing computation results of one chuck to other processors via common repository.

Stoffel, Taylor, and Hendler (1997) makes a description of an environment for supporting very large ontologies that can be used on single pcs, workstations, clusters of workstations and high-end parallel supercomputers. The proposed architecture allows answer complex queries in very large ontologies in fraction of a second in parallel super computers. The system structure is divided in three layers. The lowest is based on a RDBMS that manages all IO operations and is also used to maintain the relationships in the ontology. The second layer consists in a set of efficient inference algorithms. This level sends basic request to the database level. The relational table returned is used by the inference modules and the result produced is sent back to the first level for storage in the RDBMS. The third level is a general-purpose user interface that allows the user to insert, delete and update information in the ontology and also to pose queries.

Bonacina (1999) presents a taxonomy of parallel theorem-providing methods based on the control of search, the granularity of parallelism, the nature of the method and analyze how the different approaches to parallelization affect the control of search.

Serafini, Borgida, and Tamilin (2005) presents three proposals. The first is in the semantic front, they propose a relatively small change in semantics which localizes inconsistency and preserver directionality of knowledge import. The second is an approach which views the bridge rules connecting two local ontologies as describing operator that propagates knowledge in the form of DL subsumption axioms. The third is a distributed tableaux algorithm that determines the satisfiability of a SHIQ concept in the context of the local axioms of an ontology and the extra knowledge imparted by the bridge rules.

Serafini and Tamilin (2005) present a system called DRAGO (Distributed Reasoning Architecture for a Galaxy of Ontologies) which is a peer-to-peer architecture in which every peer registers a set of ontologies and provide reasoning services for them. The system supports the assignment of semantic mappings to the registered ontologies and performs reasoning with such ontologies coupled with semantic mapping in a distributed manner. MacIntosh, Conry, and Meyer (1991) present a distributed reasoning system called DARES, an automated reasoning system for distributed environments that gives an agent the ability to reason beyond the limitations of its local knowledge. Its environment can be viewed as a collection of distributed agents that cooperate to perform automated reasoning about the domain. Hendler (2001) affirm that the agent technology in combination with ontologies could increases the efficiency of the task and reduce the human intervention. He sees a great number of small ontological components consisting a largely of pointers to each other rather than a few large, complex, consistent ontologies shared by a great number of users. In his work argues that agent-based computing will become much more practical and that distributed computer programs interacting with nonlocal web-based resources might eventually become the dominant way in which computers interact with humans and each other. He says that web services might be one of the most powerful uses of web ontologies and will be a key enabler for web agents. He argue that a well-known problem with the web is that finding the many available web services is difficult, so it is necessary to create a machine-readable ontologies used by agents that will be able to find these web services and automate their use.

3. Previous work

In this work a multi-agent architecture is proposed to parallelize a reasoning system based on semantic technologies and ontologies