Efficient management of distributed and dynamic ontologies

Femke Ongenae

Supervisor(s): Filip De Turck, Tom Dhaene

I. Introduction

Ontologies capture knowledge about a certain domain in a formal and structured manner. They allow easy representation, reuse and sharing of knowledge. An example ontology, that models diseases and their symptoms, is shown in Fig. 1. Techniques, e. g. tableau algorithms, exist to reason on these ontologies. This way, the correctness of the model can be checked and new knowledge can be deducted, such as relations between concepts that were not known before. In recent years, ontologies have become increasingly popular in different domains such as knowledge representation, biomedical informatics, natural language processing and the Semantic Web.

This increase in attention has lead to the creation of many ontologies for real-world domains. These are often very large and complex. For example, the Gene Ontology contains around 22.000 concepts.

The size and complexity of these ontologies cause a lot of problems with respect to maintenance, reuse and validation. The scalability also becomes an issue, as reasoning is a resource-intensive process, even for small ontologies. Additionally, most visualization and modeling tools cannot handle such large ontologies in an appropriate manner. Furthermore, ontologies are often subject to changes, as knowledge about a certain domain evolves over time. However, the size and complexity

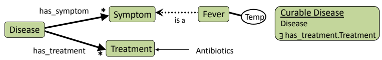

Figure 1. An example ontology

of these ontologies makes it difficult for domain experts to keep an overview of the ontology and discover new or missing knowledge.

II. Main Research Objectives

To address the previously mentioned issues, a distributed platform to reason on large-scale and dynamic ontologies is under development. The following scalable, resilient and efficient algorithms support this platform:

• *Partitioning and distribution algorithms:* To handle the problems caused by the large scale of most real-world ontologies, efficient algorithms are being developed to partition an ontology in a set of coherent modules. These modules are distributed across different servers in a resilient manner and used independently from each other [1], [2]. However, they still contain information about their relationships with other modules. By measuring parameters about the usage of the ontology, the partitioning is iteratively optimized.

• *Adaptive algorithms to support self-learning ontologies:* To support the continuous growth and change of an ontology, algorithms are being developed that indicate where knowledge is missing in the ontology. This missing knowledge can be delivered by an application, such as a neural network, or by a domain expert. A filtering algorithm determines if this new knowledge is relevant enough to be added to

F. Ongenae is with the Department of Information Technology (INTEC), Ghent University - IBBT. E-mail: Femke.Ongenae@intec.ugent.be .

the ontology and where exactly in the model it should be added, without making the whole ontology inconsistent.

• *Algorithms for the management of dynamic ontologies:* A constantly changing ontology must be able to dynamically evolve into a new version. An efficient version control system, which can handle a distributed ontology, is being developed. To ensure consistency, new knowledge is added to the ontology with a certain probability. When more data becomes available that contradicts or confirms this new knowledge, the probability is decreased or increased. Techniques were investigated to efficiently represent and reason on probabilistic knowledge in an ontology.

III. Use Case: Detection Of Sepsis

The developed platform and algorithms will be used to automatically detect sepsis in patients in the Intensive Care Unit (ICU) of a hospital, as visualized in Fig. 2. Sepsis [3] is a severe inflammatory response of the body to an infection. In the US, approximately 750.000 people are diagnosed with sepsis every year, with a mortality rate of 30%. Early detection is crucial to the survival of the patient, as it reduces the mortality rate to 15%. The exact symptoms of sepsis are unknown. However, there is a guideline that lists 25 possible indicators of sepsis. Continuously monitoring these, often time-dependent, parameters is difficult.

In collaboration with Ghent University Hospital, an existing hospital ontology [4] is expanded to represent the symptoms of patients and their relationships with diseases. A lot of knowledge in this domain is hidden in parameter trends. Consequently, techniques are being developed to represent and reason on time-dependent knowledge in an ontology. Sequences of symptoms, which are known to cause sepsis, are modeled in the ontology. The resulting large-scale ontology is partitioned and distributed across the network to resolve scalibility, maintenance and validation issues.

The previously mentioned indicator, filter and knowledge discovery algorithms are used

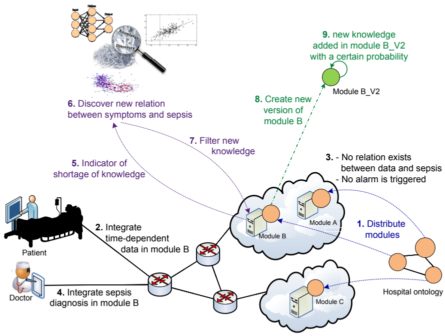

Figure 2. Applying the platform to detect sepsis

to discover new guidelines or symptom trends that indicate the presence of sepsis. The new knowledge is added to the ontology with a certain probability. This is supported by the distributed version control system.

IV. Summary

In summary, the goal of this project is on one hand to develop a distributed platform to reason on large-scale and dynamic ontologies. On the other hand, this platform will be used to automatically detect sepsis more early.

Acknowledgments

F. Ongenae is funded by the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT).

References

[1] Femke Ongenae and Stijn Verstichel et al., "OTAGen: A tunable ontology generator for benchmarking ontology-based agent collaboration," in *Proc. of IEEE COMPSAC*, 2008, pp. 529–530.

[2] Stijn Verstichel and Femke Ongenae et al., "An autonomous service-platform to support distributed ontology-based context-aware agents," *Accepted for Expert Systems*, 2008.

[3] Derek C. Angus and Walter T. Linde-Zwirble et al., "Epidemiology of severe sepsis in the united states: Analysis of incidence, outcome, and associated costs of care," *Crit. Care Med.*, vol. 29, no. 7, pp. 1303–1310, 2001.

[4] Femke Ongenae and Matthias Strobbe et al., "Design of a semantic person-oriented nurse call management system," *Int. J. of Web and Grid Services*, vol. 4, no. 3, pp. 267–283, 2008.