COMPUTER SCIENCE DEPARTMENT
FACULTY OF COMPUTERS AND INFORMATION SCIENCE
AIN SHAMS UNIVERSITY

Ontological Engineering Approach for Medical Knowledge Sharing

A Thesis Submitted to Computer Science Department,
Faculty of Computers and Information Science,
Ain Shams University, Cairo, Egypt.

In partial fulfillment of the requirements for
the degree of doctor of philosophy in computer science

By

Marco Alfonse Tawfik
Computer Science Department
Faculty of Computers and Information Science
Ain Shams University, Cairo, Egypt

Under Supervision of

Prof. Dr. Abdel-Badeeh M. Salem
Professor of Computer Science
Computer Science Department
Faculty of Computers and Information Science
Ain Shams University, Cairo, Egypt

Prof. Dr. Mostafa M. Aref
Professor of Computer Science
Computer Science Department
Faculty of Computers and Information Science
Ain Shams University, Cairo, Egypt

August 2014

Acknowledgment

First and foremost, I humbly give my deep thanks to God for giving me the opportunity and the strength to accomplish this work.

I would like to thank all the people who helped me to make this work possible. I would like to give my deep appreciation and thanks to my supervisor, Prof. Dr. Abdel-Badeeh M. Salem, for the confidence that he always had in me, for his wisdom, intelligence and creative vision. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my PhD study. He always gave me the precise advice at the precise moment so that I had both the freedom and support that I needed to do my research.

My sincere thanks also goes to Prof. Dr. Mostafa M. Aref for his valuable and helpful guidance in reviewing and directing the work presented in this research.

Last but not the least, I would like to thank my family and friends who were always with me supporting my decisions, celebrating my successes, and comforting me in my setbacks.

To all of them, thank you.

List of Publications

[1] Abdel-Badeeh M. Salem, Marco Alfonse, "**Web-Based Ontologies for Breast and Lung Cancer**", Proceedings of the 6th International Conference of EURO-MEDITERRANEAN MEDICAL INFORMATICS and TELEMEDICINE (EMMIT 2010), Split, Croatia, PP 52-61, September 26–28, 2010.

[2] Marco Alfonse, Abdel-Badeeh M. Salem. "**Ontological Engineering in Bioinformatics**". Proceedings of 1st International Conference on E-Health and TeleMedicine (ICEHTM 2011), Nicosia, North Cyprus, ISBN: 978-9963-7393-0-1, PP: 95-110, 10-12 October 2011.

[3] Marco Alfonse, Mostafa M. Aref, Abdel-Badeeh M. Salem. "**Ontological Engineering Approach for Creating a Semantic Web Service for Determining the Stage of Breast Cancer**". Proceedings of the 7th International Conference of EURO-MEDITERRANEAN MEDICAL INFORMATICS and TELEMEDICINE (EMMIT 2011), Tirana, Albania, PP: 72 – 78, November 10–12, 2011.

[4] Marco Alfonse, Mostafa M. Aref, Abdel-Badeeh M. Salem. "**An Overview of Ontology Learning From Unstructured Texts**". Proceedings of the Eleventh International Conference on Informatics (INFORMATICS 2011), Rožňava, Slovakia, ISBN 978-80-89284-94-8, PP: 169-174, November 16-18, 2011.

[5] Marco Alfonse, Mostafa M. Aref, Abdel-Badeeh M. Salem. "**Ontology-Based Knowledge Representation for Liver Cancer**". Proceedings of the International eHealth, Telemedicine and Health ICT Forum for Educational, Networking and Business. Luxembourg, G. D. of Luxembourg, ISSN 1818 – 9334, PP: 821-825, April 18-20, 2012.

[6] Marco Alfonse, Kenneth Revett, Abdel-Badeeh M. Salem. "**A Study on Ontological Engineering Approach for Bioinformatics**". EURO-MEDITERRANEAN MEDICAL INFORMATICS and TELEMEDICINE, EMMIT 2012, 8th International Meeting, Bighi Hospital (Old Military Hospital), Malta. December17-19, 2012.

[7] Marco Alfonse, Mostafa M. Aref, Abdel-Badeeh M. Salem. "**Developing an Ontology-Based Multi-Agent Tool for Cancer Diseases Knowledge Management**". International Journal of Applications of Fuzzy Sets and Artificial Intelligence (ISSN 2241-1240), Vol.3, PP: 5-22, 2013.

[8] Marco Alfonse, Mostafa M. Aref, Abdel-Badeeh M. Salem. "**An Ontology-Based Cancer Diseases Diagnostic Methodology**". Proceedings of the 7th European Computing Conference (ECC '13), ISSN: 1790-5109, PP:95-99, Dubrovnik, Croatia, June 25-27, 2013.

[9] Marco Alfonse, Mostafa M. Aref, Abdel-Badeeh M. Salem. "**An Ontology-Based System for Cancer Diseases Knowledge Management**". Accepted for publication in the International Journal of Information Technology and Computer Science (IJITCS), 2014.

Abstract

Medical informatics is concerned with the development of the science and technologies needed for collecting, sharing, reporting, analyzing, and visualizing medical data; and for providing decision-making support for disease prevention, detection, and management. However, the lack of standard vocabulary of medical data hinders effective classification. To meet this challenge, an appropriated way to describe these data is required. Many alternatives have been developed, however, there is not a consensus to use one or a small set of them yet. Artificial Intelligence is a research area, which has proposed important approaches as data structures, relational databases, mathematic logic, procedures, and taxonomies among others. One of these proposals is ontologies.

Ontologies were developed in Artificial Intelligence to facilitate knowledge sharing and reuse. Since the beginning of the nineties ontologies have become a popular research topic investigated by several Artificial Intelligence research communities, including Knowledge Engineering, natural-language processing and knowledge representation. More recently, the notion of ontology is also becoming widespread in fields such as intelligent information integration, cooperative information systems, information retrieval, electronic commerce, and medical knowledge management. The reason ontologies are becoming so popular is in large part due to what they promise: a shared and common understanding of some domain that can be communicated between people and application systems.

This thesis presents the process of building a web-based ontology for the liver cancer. This ontology is built using the Protégé -OWL editing environment and is encoded in OWL-DL format. It was built using Top-Down approach. The main goals behind building this ontology are to allow finding and location information about the liver cancer needed for interested users and domain experts and providing a semantic representation of liver cancer information over the web.

The thesis also includes the process of creating a semantic web service for determining the stage of breast cancer. The service is encoded in OWL-S using the OWL-S editor. This service depends on a breast cancer ontology developed by the authors using Protégé-OWL ontology editor and encoded in OWL-DL format. This service can be used by patients as well as physicians to determine the stage of breast cancer.

Also the thesis introduces the process of developing an ontology-based multi-agent tool for cancer diseases knowledge management. The tool was built using the Java Agent Development framework (JADE) which is the most widespread agent-oriented middleware. The tool is composed of two intelligent agents which implement the client (patient) and server (physician) roles for the cancer diseases. The "Diagnosis Agent" class acts as a physician and the "Client Agent" class acts as a patient. The two agents use a common ontology "Disease Ontology" that defines the vocabulary and semantics which represent the terms that constitute the specific language of the agents. The tool can be used to find the cancer type or determining the stage of a specific cancer. Moreover it can provide treatment solutions according to the cancer discovered or the stage of the cancer determined.

Finally the thesis discusses the process of developing an ontology-based cancer diseases knowledge management system. This system can be used by patients, students and physicians to decide the type of the cancer the patient has, the stage of the cancer and the treatment options. The main contribution of this system is its extensibility to contain all types of cancers and its capability to provide the treatment recommendation for the case at hand. Currently the database of the cancer ontologies of this system has only three types of cancers, which are lung, liver and breast. In future, it is planned to extend the database with all types of cancer ontologies. The system gives 92% correct classification.

Table of Contents

List of Tables

List of Figures

Chapter 1
Introduction

To support the sharing and reuse of formally represented knowledge among Artificial Intelligence (AI) systems, it is useful to define the common vocabulary in which shared knowledge is represented. A body of formally represented knowledge is based on a conceptualization: the objects, concepts, and other entities that are assumed to exist in some area of interest and the relationships that hold among them. A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose. Every knowledge base, knowledge-based system, or knowledge-level agent is committed to some conceptualization, explicitly or implicitly. A specification of a representational vocabulary for a shared domain of discourse - definitions of classes, relations, functions, and other objects - is called an ontology [1]. Ontologies provide shared and precisely defined terms that can be understood and processed by machines. A typical ontology consists of a hierarchical description of important concepts and their relations in a domain, task or service.

Recent work in Artificial Intelligence is exploring the use of formal ontologies as a way of specifying content-specific agreements for the sharing and reuse of knowledge among software entities. Formal ontologies are viewed as designed artifacts, formulated for specific purposes and evaluated against objective design criteria.

1.1. Thesis Motivation

The next generation of the current web is called the sematic web or “Web of data,” the sort of data you find in databases. The ultimate goal of the Web of data is to enable computers to do work that are more useful and to develop systems that can support trusted interactions over the network. The term “Semantic Web” refers to the Web of linked data. Semantic Web technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data [2]. The development of ontologies will be central to this effort. Ontologies are metadata schemas, providing a controlled vocabulary of terms, each with an explicitly defined and

machine processable semantics. By defining shared and common domain theories, ontologies help both people and machines to communicate more effectively. They will therefore have a crucial role in enabling the Semantic Web. The add on value of a worldwide semantic network can now only be demonstrated in small parts due to lack of relevant domain specific ontologies and ontology-based applications.

1.2. Thesis Objectives

The technologies of information systems have been progressing at a rapid pace. Information systems are now being called upon to support knowledge management, and not just to process data or information. Many advances contribute to taking information systems beyond mere data into the realm of knowledge. These include: similarity-based retrieval and browsing, data mining and knowledge discovery, text understanding, and knowledge sharing. However, the key to providing useful support for knowledge management is founded on how meaning is embedded in information models as defined in terms of ontologies. The thesis objective is to study the role of ontologies they can provide to enrich and support most of the AI technologies available today by enabling semantic representation of knowledge in general and medical knowledge in specific. These technologies include web services, multi-agent systems and medical information management systems.

1.3. Thesis Importance

The use of ontologies in medicine is mainly focused on the representation and (re-) organization of medical terminologies. Physicians developed their own specialized languages and lexicons to help them store and communicate general medical knowledge and patient-related information efficiently. Such terminologies, optimized for human processing, are characterized by a significant amount of implicit knowledge. Medical information systems, on the other hand, need to be able to communicate complex and detailed medical concepts (possibly expressed in different languages) unambiguously. This is obviously a difficult task and requires a profound analysis of the structure and the concepts of medical terminologies. It can be achieved by constructing medical domain ontologies for representing medical terminology. The importance of this thesis comes from its contribution in medicine through developing

some medical ontologies and ontology-based medical applications addressing the above issues.

1.4. Thesis Methodology

This thesis depends on a stepwise methodology for building ontologies, a modeling methodology for the formal analysis of the medical domain which has been implemented and an engineering methodology for developing the ontology-based applications [3]. The stepwise methodology divides the ontology development process into a set of phases, which are specification, conceptualization, formalization, implementation, and maintenance. The domain analysis (the modeling process) was performed typically by means of a set of well-defined modeling constructs and primitives, e.g. the notions of concept/class, relations/roles, functions, properties/attributes, constraint/rule types, etc. The engineering methodology is concerned with the design, representation, architecture, and management aspects of the ontology-based applications created. The object-oriented paradigm is the basis for the engineering methodology. This paradigm provides guidance for its adopters (software developers) by encapsulating the complexity of each software module, thus making the products (software programs) more reusable, maintainable, and easy to build.

1.5. Thesis Contribution

Due to the lack of relevant domain specific ontologies and the necessity to formalize medical information and providing ontology-based medical applications, the main contributions of this thesis are:

- Building a web-based domain specific ontology that may help advancing the work done in the process of moving the current web from the first generation to the second generation: the Semantic Web. This ontology represents one of the most common types of cancer, which is the liver cancer.
- Creating a semantic web service for determining the stage of breast cancer.
- Developing an ontology-based multi-agent tool for cancer diseases knowledge management.
- Developing an ontology-based cancer diseases management system.