Ontology versus Semantic Networks for Medical Knowledge Representation

ABDEL-BADEEH M. SALEM, MARCO ALFONSE
Computer Science Department
Ain Shams University
Faculty of Computers and Information Systems
CAIRO, EGYPT

Abstract: - The efficient Semantic Web, needs knowledge to be presented in a well defined form which enables people and software agents to understand it. There are many knowledge representation techniques such as rules, frames, scripts, semantic network and ontology. The purpose of this paper is to illustrate the difference between semantic network and ontology techniques for medical knowledge representation. The paper presents a development of semantic net and ontology for lung cancer using PROLOG and Protégé-OWL respectively. The results show that ontology supports representing knowledge semantically for building a robust knowledge-based system.

Key-words: **-** Knowledge Representation, Medical Informatics, Semantic Network, Ontology, Lung Cancer, Artificial Intelligence.

1 Introduction

A semantic network or net [1] is a graphic notation for representing knowledge in patterns of interconnected nodes and arcs. Computer implementations of semantic networks were first developed for artificial intelligence and machine translation, but earlier versions have long been used in philosophy, psychology, and linguistics. What is common to all semantic networks is a declarative graphic representation that can be used either to represent knowledge or to support automated systems for reasoning about knowledge. On the other hand, An Ontology [2, 10 and 11] is a representation vocabulary, often specialized to some domain or subject matter. It is a representation of a set of concepts within a domain and the relationships between those concepts. It is used to reason about the properties of that domain, and may be used to define the domain. Ontologies are used in artificial intelligence, the Semantic Web, software engineering, biomedical informatics, library science, and information architecture as a form of knowledge representation about the world or some part of it. This paper studies the semantic networks and ontology as knowledge representation techniques in medical domain. In this respect we developed a semantic network as well as an ontology for lung cancer.

2 Main Aspects of Semantic Networks

There are six common kinds of semantic networks: (1) Definitional networks emphasize the subtype or is-a relation between a concept type and a newly defined subtype. The resulting network, also called a generalization or subsumption hierarchy, supports the rule of inheritance for copying properties defined for a super type to all of its subtypes. Since definitions are true by definition, the information in these networks is often assumed to be necessarily true. (2) Assertional networks are designed to assert propositions. Unlike definitional networks, the information in an assertional network is assumed to be contingently true, unless it is explicitly marked with a modal operator. Some assertional netwoks have been proposed as models of the conceptual structures underlying natural language semantics. (3) Implicational networks use implication as the primary relation for connecting nodes. They may be used to represent patterns of beliefs, causality, or inferences. (4) Executable networks include some mechanism, such as marker passing or attached procedures, which can perform inferences, pass messages, or search for patterns and associations. (5) Learning networks build or extend their representations by acquiring knowledge from examples. The new knowledge may change the old network by adding and deleting nodes and arcs or by modifying numerical values, called weights, associated with the nodes and arcs. (6) Hybrid networks combine two or more of the previous techniques, either in a single network or in separate, but closely interacting networks. Some of the networks have been explicitly designed to implement hypotheses about human cognitive mechanisms, while others have been designed primarily for computer efficiency

3 Main Aspects of Ontology

An ontology is a formal explicit description of concepts in a domain of discourse (classes (sometimes called concepts)), properties of each concept describing various features and attributes of the concept (slots (sometimes called roles or properties)), and restrictions on slots (facets (sometimes called role restrictions)). An ontology together with a set of individual instances of classes constitutes a knowledge base. Ontologies are built to share common understanding of the structure of information among people or software agents, to enable reuse of domain knowledge, to make domain assumptions explicit, to separate domain knowledge from the operational knowledge and to analyze domain knowledge. Sharing common understanding of the structure of information among people or software agents is one of the more common goals in developing ontologies. For example, suppose several different Web sites contain medical information or provide medical e-commerce services. If these Web sites share and publish the same underlying ontology of the terms they all use, then computer agents can extract and aggregate information from these different sites. The agents can use this aggregated information to answer user queries or as input data to other applications. Enabling reuse of domain knowledge was one of the driving forces behind recent surge in ontology research. For example, models for many different domains need to represent the notion of time. This representation includes the notions of time intervals, points in time, relative measures of time, and so on. If one group of researchers develops such an ontology in detail, others can simply reuse it for their domains. Additionally, if we need to build a large ontology, we can integrate several existing ontologies describing portions of the large domain. We can also reuse a general ontology and extend it to describe our domain of interest. Making explicit domain assumptions underlying an implementation makes it possible to change these assumptions easily if our knowledge about the domain changes. Hard-coding assumptions about the world in programming-language code makes these assumptions not only hard to find and understand but also hard to change, in particular for someone without programming expertise. In addition, explicit specifications of domain knowledge are useful for new users who must learn what terms in the domain mean. Separating the domain knowledge from the operational knowledge is another common use of ontologies. We can describe a task of configuring a product from its components according to a required specification and implement a program that does this configuration independent of the products and components themselves. Analyzing domain knowledge is possible once a declarative specification of the terms is available. Formal analysis of terms is extremely valuable when both attempting to reuse existing ontologies and extending them. Often an ontology of the domain is not a goal in itself. Developing an ontology is akin to defining a set of data and their structure for other programs to use. Problem-solving methods, domain-independent applications, and software agents use ontologies and knowledge bases built from ontologies as data. Ontologies are commonly encoded using ontology languages [3] such as RDF, RDFS and OWL which is endorsed by the World Wide Web Consortium. It is based on two semantics: OWL-DL and OWL-Lite semantics are based on Description Logics, which have attractive and well-understood computational properties, while OWL-Full uses a novel semantic model intended to provide compatibility with RDF Schema.

4 The Lung Cancer

Lung cancer is a disease of uncontrolled cell growth in tissues of the lung. This growth may lead to metastasis, invasion of adjacent tissue and infiltration beyond the lungs. The vast majority of primary lung cancers are carcinomas of the lung, derived from epithelial cells. Lung cancer, the most common cause of cancer-related death in men and the second most common in women, is responsible for 1.3 million deaths worldwide annually according to World Health Organization [5]. The most common symptoms are shortness of breath, coughing (including coughing up blood), and weight loss. The main types of lung cancer are small cell lung carcinoma and non-small cell lung carcinoma. This distinction is important because the treatment varies; non-small cell lung carcinoma (NSCLC) is sometimes treated with surgery, while small cell lung carcinoma (SCLC) usually responds better to chemotherapy and radiation. The most common cause of lung cancer is long term exposure to tobacco smoke. The occurrence of lung cancer in non-smokers, who account for fewer than 10% of cases, appears to be due to a combination of genetic factors, radon gas, asbestos, and air pollution, including second-hand smoke.

Lung cancer may be seen on chest x-ray and computed tomography (CT scan). The diagnosis is confirmed with a biopsy. This is usually performed via bronchoscopy or CT-guided biopsy. Treatment and prognosis depend upon the histological type of cancer, the stage (degree of spread), and the patient's performance status. Possible treatments include surgery, chemotherapy, and radiotherapy. With treatment, the five-year survival rate is 14% [6]. For more information about lung cancer see [7].

5 Developing the Lung Cancer Semantic Network

The lung cancer semantic network is implemented using PROLOG which is a logic programming language often associated with artificial intelligence and computational linguistics. Figure (1) shows the semantic network of the lung cancer. In this figure we define the lung cancer in terms of it symptoms, stages, causes and how it can be diagnosed, staged and treated.

The physical attributes of the lung cancer can be represented in logic using PROLOG as follows:

1. The *isa* relationship (The super classes and subclasses) can be described as:

isa("Cancer","Disease").
isa("Lung Cancer","Cancer").

2. The Properties such as "*has_stage*" and "*treated_by*" can be described as:

objectproperty("hasMstage","Lung_Cancers","M_stage").
objectproperty("hasNstage","Lung_Cancers","N_stage").
objectproperty("hasTstage","Lung_Cancers","T_stage").
objectproperty("treatedBy","Disease","Treatment").

3. The Instances (see figure 2) can be described as follows:

instance("Smoking","lung_cancer_causes").
instance("Asbestos","lung_cancer_causes").
instance("Radon_gas","lung_cancer_causes").
instance("Chemicals_and_substances","lung_cancer_causes").

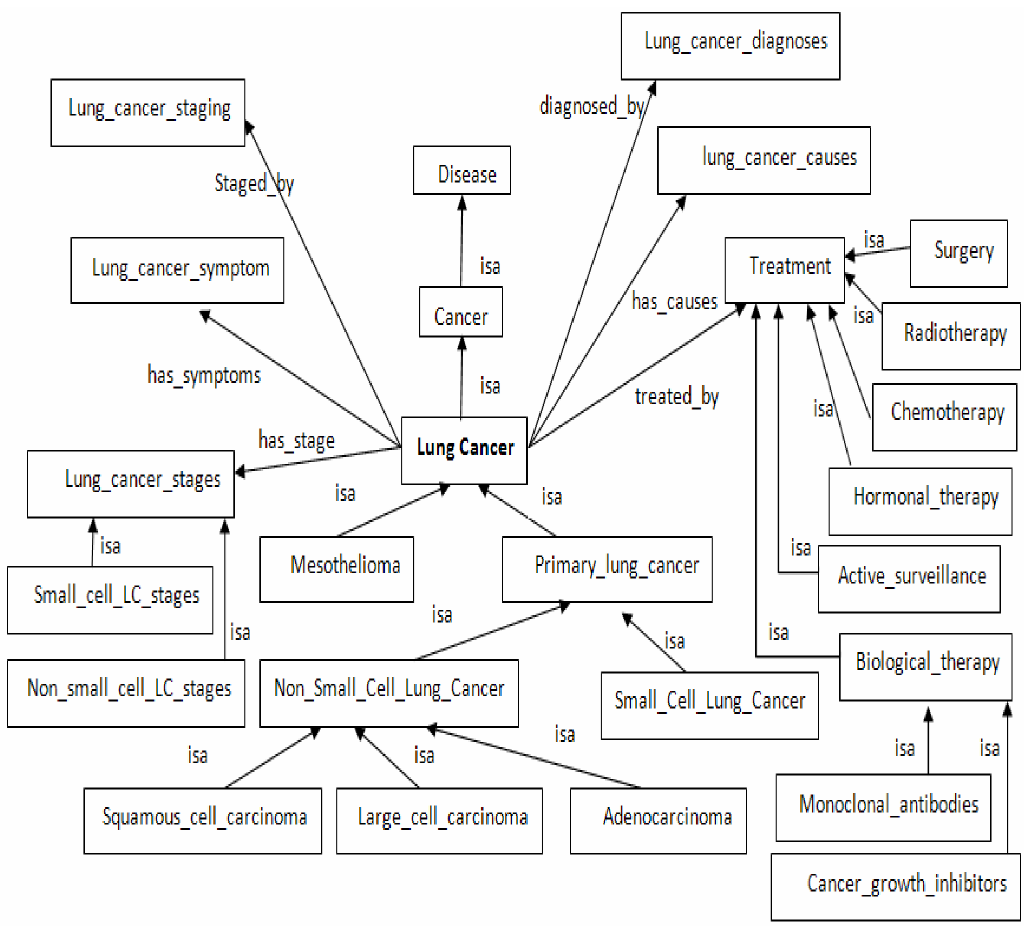

Figure 1: The Semantic Net of the Lung Cancer

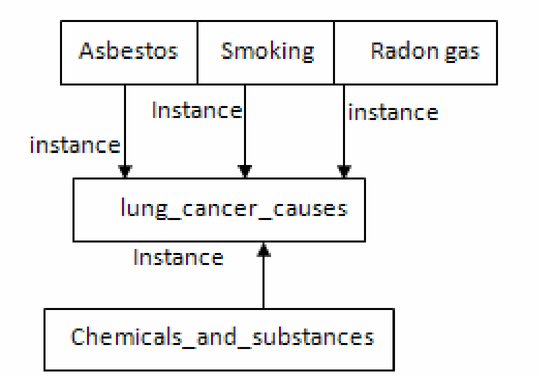

Figure 2: Instances of Lung Cancer Causes

In what follows the advantages of Semantic Networks can be summarized in the following points:

1. Hierarchical organization of knowledge.
2. Semantic Networks realized as massively parallel networks or agents may provide the appropriate framework for modeling reflexive reasoning (isa-link).
3. Easy to visualize.
4. Related knowledge is easily clustered.

Regarding the disadvantages of Semantic Networks as mentioned in Wood's paper 'What's in a link' (1975) shows ambiguities and unclarities in Semantic Networks:

1. No semantics for semantic networks.
2. Different interpretations for the same network.
3. No standards about node and arc values.

Actually what is needed is a common understanding of the semantic in a semantic network which leads to ontology.

6 Developing the Lung Cancer Ontology

The lung cancer ontology is implemented using Protégé-OWL editor version 3.2.1 [4] which is a free, open-source platform that provides a growing user community with a suite of tools to construct domain models and knowledge-based applications with ontologies. It supports the creation, visualization, and manipulation of ontologies in various representation formats. Protégé can be extended by way of a plug-in architecture and a Java-based Application Programming Interface (API) for building knowledge-based tools and applications. This ontology is encoded in OWL-DL as a representation language because of its reasoning and expressive power. See [8] for details about the lung cancer ontology.

Figure 3 shows the lung cancer class hierarchy and its properties as shown in Protégé. As the figure shows, the lung cancer class is described in terms of its causes, symptoms, diagnosis, staging and treatment.

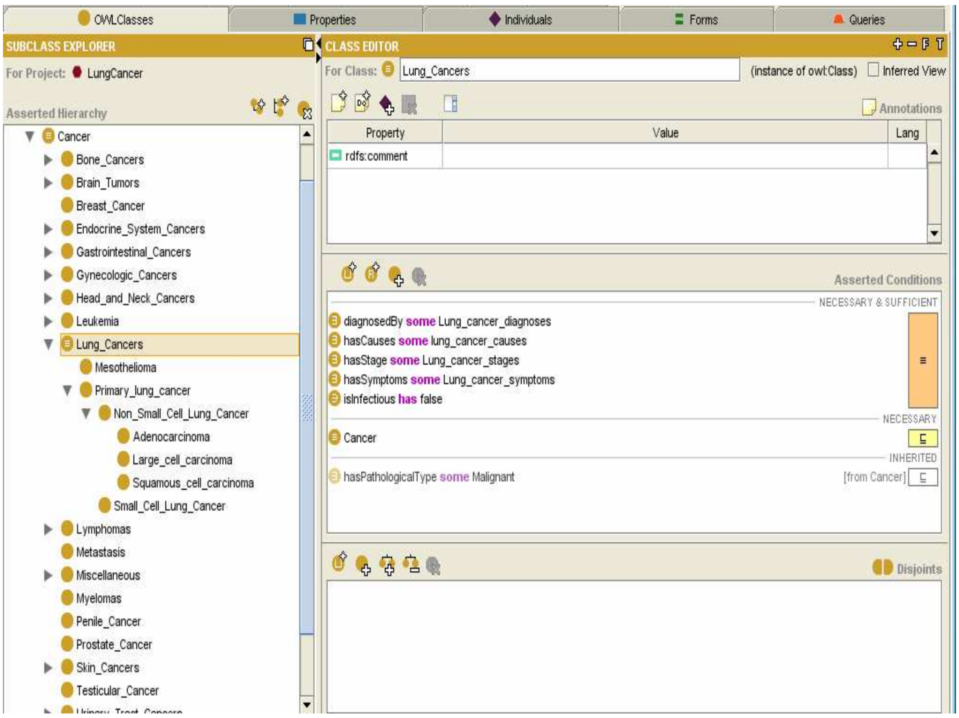

Figure 3: Class Hierarchy of Lung Cancer

7 Semantic Network versus Ontology

As shown in the previous two sections, the semantic net is good for representing the superclass-subclass relationships and it is also useful for representing instances for each class but what about restrictions on relationships and Property Characteristics?

Using OWL, It is possible to specify property characteristics, which provides a powerful mechanism for enhanced reasoning about a property as follows:

1. TransitiveProperty:
   If a property, P, is specified as transitive then for any x, y, and z:
   P(x,y) and P(y,z) implies P(x,z).
2. SymmetricProperty:
   If a property, P, is tagged as symmetric then for any x and y:
   P(x,y) iff P(y,x)
3. FunctionalProperty
   If a property, P, is tagged as functional then for all x, y, and z:
   P(x,y) and P(x,z) implies y = z
4. InverseFunctionalProperty
   If a property, P, is tagged as InverseFunctional then for all x, y and z:
   P(y,x) and P(z,x) implies y = z

In addition to designating property characteristics, it is possible to further constrain the range of a property in specific contexts in a variety of ways:

1. The value constraint *owl:allValuesFrom* restriction requires that for every instance of the class that has instances of the specified property, the values of the property are all members of the class indicated by the owl:allValuesFrom clause.
2. The value constraint *owl:someValuesFrom* defines a class of individuals x for which there is at least one y (either an instance of the class description or value of the data range) such that the pair (x,y) is an instance of P. This does not exclude that there are other instances (x,y') of P for which y' does not belong to the class description or data range.
3. The value constraint *owl:hasValue*, allows us to specify classes based on the existence of particular property values.
4. The cardinality constraint *owl:cardinality*, permits the specification of exactly the number of elements in a relation.

The OWL can also provide Boolean combinations (union, intersection, complement) of classes and the means to specify a class via a direct enumeration of its members. See [9] for details about the OWL syntax and semantics it can provide. In the lung cancer domain, we need to say that the lung cancer can be diagnosed by one or more of the following methods: Bronchoscopy, Computerized Tomography (CT) Scan, or Spiral CT Scan. To express this knowledge we need to use the value constraint *owl:someValuesFrom* as shown in the OWL definition of lung cancer (see previous section). These entire features are missing when using the semantic net to represent the knowledge in hand, so ontologies are best suited to represent knowledge for the semantic web due to its powerful representation.

Table (1) describes the main differences between semantic networks and ontologies for representing knowledge for medical domain.

Table 1: Semantic Network vs. Ontology

Features	Semantic Network	Ontology
Unique Name Assumption	If two objects have different names, they are assumed to be different.	There is no assumption on whether objects are the same or different unless there is an explicit statement about specifying the relationship.
Open vs. Closed World Assumption	Nothing can be entered into it until there is a place for it in the corresponding template.	Anything can be entered into ontology unless it violates one of the constraints.
Assertion vs. Classification	Defining facets on a slot at a class, or defining a constraint on a slot at the top level, makes a statement about all instances of those describing necessary conditions for instances of that class.	There are effectively two kinds of statements about classes: a) those that are true of all individuals in a class, and b) those that are collectively necessary and sufficient to define the class.
Ability to define Rules	Rules can be applied when implementing semantic net using PROLOG.	No rules can be applied when implementing ontology using OWL.
Expressive Power	There are no restrictions on relationships and property characteristics.	It allows some restrictions; anonymous classes; necessary and sufficient conditions; expressions

8 Conclusions

This paper discusses the different aspects of semantic networks and ontology for a medical domain; namely, lung cancer. In semantic network representation, there is no formal semantics, no agreed-upon notion of what a given representational structure means, and there is no ability to express property characteristics. The ontology representation defines the vocabulary of a domain with which queries and assertions are exchanged among agents. The semantic network of lung cancer was developed using PROLOG and the lung cancer ontology was developed using Protégé-OWL. The results show that ontology provides a robust knowledge representation technique for building medical knowledge-based systems.

References:

[1] Sowa, John F., ed. "Principles of Semantic Networks: Explorations in the Representation of Knowledge", Morgan Kaufmann Publishers, San Mateo, CA, 1991.
[2] Y. Ding. "Ontology: the Enabler for the Semantic Web". Journal of Information Science. 27 (6), 2001.
[3] X. Su and L. Ilebrekke, "A Comparative Study of Ontology Languages and Tools", In Proceeding of the 14th Conference on Advanced Information Systems Engineering (CAiSE'02), Toronto, Canada, May 2002.
[4] http://protege.stanford.edu/
[5] World Health Organization: http://www.who.int/en/
[6] Minna JD, "Harrison's Principles of Internal Medicine", PP 506–516, McGraw-Hill, 2004.
[7] http://www.cancerbackup.org.uk/Cancertype/Lung
[8] Abdel-Badeeh M. Salem, Marco Alfonse, "Building Web-Based Lung Cancer Ontology", Proceedings of 1st National Symposium on e-Health and Bioengineering - EHB 2007, PP 177-182, IAŞI, ROMANIA.
[9] OWL Web Ontology Language Reference: http://www.w3.org/TR/owl-ref/
[10] Noy N.F., McGuinnes D. L. "Ontology Development 101: A Guide to Creating Your First Ontology". Stanford Knowledge Systems Laboratory Technical Report KSL−01−05 and Stanford Medical Informatics Technical Report SMI−2001−0880, March 2001.
[11] María Auxilio, "An Overview of Ontologies", Center for Research in Information and Automation Technologies Technical Report, Interactive and Cooperative Technologies Lab, Universidad De Las Américas Puebla, México, March 2003.