A Layered Approach to Building a Semantic Web Application for Pathological Gross Description*

Hong-Gee Kim[a], Byung-Hyun Ha[b], Jae-Il Lee[c], Myeng-Ki Kim[d]

[a]*Department of Management Information Systems, Dankook University, Korea*
[b]*Department of Industrial Engineering, Seoul National University, Korea*
[c]*Department of Oral Pathology, Seoul National University, Korea*
[d]*Department of Dental Management & Informatics, Seoul National University, Korea*

Abstract

In our research we implemented the PathOnt system for the computerized input of gross description that enables knowledge modelers and users with different roles to collaborate with each other. PathOnt that is built upon the reusable pathological ontology facilitates the interactions not only between end-users of medical information but also between different applications using semantic web markup languages.

Keywords:

Gross Description, Medical Informatics, Ontology, Pathology, Semantic Web

Introduction

Extensive research in recent years has been conducted to build knowledge-based applications that are available on the Web. The Semantic Web is a vision to create a web where the contents of Web pages amount to "readable" or "understandable" knowledge for machines. Web-based ontologies gained through knowledge modeling activities are key-enablers of the vision of the Semantic Web.

Structured knowledge models in medical informatics are applied to such areas as medical terminology, intelligent user interfaces, decision support and semantic indexing, semantics oriented language technology, and system integration of information systems [8]. However, formalizing medical knowledge in the form of ontologies is accompanied with both technical and modeling issues. The technical issues, on the one hand, include connectivity, interoperability, scalability, complexity, and expressivity. On the other hand, the modeling issues are related to the needs for reuse of an external ontology, the requirement for easy understandability by domain experts, and decisions for granularity. This paper introduces a layered approach to building an ontology-based application for the pathological gross description using semantic web technologies. Presenting the layered architecture of our system, we consider different levels in modeling and using medical knowledge, while dealing with at the same time issues of interoperability and connectivity from the technical perspective.

The gross description, usually made by an anatomical pathologist, contains macroscopic findings of pathological tissues such as size, color, texture, and consistency. It may also contain the pathologist's judgment of the disease state of the specimen tissue such as hemorrhage, necrosis, and cystic change. The data in the gross description can be used as either evaluative or supportive information for the final pathological diagnosis. In our research we implemented the PathOnt system for the computerized input of gross description that enables knowledge modelers and users with different roles to collaborate with each other. PathOnt that is built upon the reusable pathological ontology facilitates the interactions not only between end-users of medical information but also between different applications using semantic web markup languages. In the following sections, we will discuss in detail about the layered architecture of PathOnt for the gross description both from the modeling perspective and from the technical perspective.

Knowledge Modeling for the Gross Description

Information Sharing During Pathological Examination

Clinical diagnoses for many medical conditions are made by removing a sample of tissue from the patient and sending it to a pathologist for examination. The pathologist prepares a written report with information, named the gross description, in order to help the primary doctor manage the patient's condition properly. Although the gross description is mainly a job of an anatomical pathologist, the entire process of pathological examination involves quite complicated communicative activities among pathologists, clinicians, and pathological technicians.

The typical gross description contains such information as (1) shape and type of specimen, (2) fixed or unfixed, (3) number of pieces of tissues, and (4) dimensions of specimen. Since the gross description could heavily rely on the pathologist's subjective observation, pathological organizations publicize recommendations for the reporting of specimen in aid of information sharing between different professionals [1][2][3]. The computerized input form of the PathOnt system is to flexibly generate input templates reflecting the recommendations.

Preparing the gross description, the pathologist uses the general background knowledge of anatomy and pathology in order to make judgment of the patient's disease state from

* This study was supported by a grant of the Korea Health 21 R&D Project, Ministry of Health & Welfare, Republic of Korea. (03-PJ1-PG10-51300-0003)

the clinical information provided by a clinician. The gross description may include some evaluative information such as appropriateness of surgical operations and classification of disease stages in progress. It can also spell out orders for additional examinations such as embedding, dissection, decalcification, DNA extraction, and so on, which are, in usual, performed by pathological technicians. In this sense, the gross description can be regarded as a communicative channel among clinicians, pathologists, and lab technicians.

While many pathological reporting systems have been developed to store the patient's diagnosis information in the form of relational database, the gross description is still recorded in the memo form in most hospitals. The PathOnt system implemented in our research not only provides automatic ways of data input and information modeling, but also facilitates communications between different professionals by organizing medical concepts in the form of sharable and reusable ontology.

Reuse of the GALEN Ontology

Today, few raise a question of the need for controlled medical terminology to support clinical information systems in the sense that information using terminology standards facilitate coordinating shared patient care among different healthcare professionals and institutions. However, most existing terminologies are not qualified to represent medical concepts whose semantics are machine understandable since they are too coarse grained to record care of individual patients.

GALEN, one of the best-known medical ontology systems, provides reusable terminology resources in terms of the Common Reference Model, formulated in a specialized description logic, GRAIL. Advocates of description logics maintain that an ontology should contain correct and complete definitions of represented concepts as well as generalization relations among concepts [7][9]. The GALEN ontology in this sense is sufficiently expressive not only to capture rich semantics of the statements used by clinicians but also to provide logical formalism that can further deal with temporal, spatial, epistemic, and inferential aspects of knowledge. Another distinctive feature of GALEN is that it provides loosely coupled development tools to enable users to adapt core terminologies to their specific needs. Building a Semantic Web application for the gross description in this research, we have modeled pathological ontology that is based on the GALEN's Core Reference Model though it is much simpler than the GALEN ontology.

Pathological Ontology for the Gross Description

The authoring tools in GALEN are supposed to be sufficiently easy to use for domain experts in the sense that "intermediate representation" of GRAIL language insulates authors from the details of the description logic [8]. Although representation in GRAIL is easier to understand than full description logics, it is still not only too complex for domain experts to understand but also too fine grained for some specific needs like gross description (Figure 1).

```
[SolidStructure PlanarStructure Substance]
        sensibly hasColour Colour.
SolidStructure
        sensibly hasShape Shape.
[SolidStructure Substance]
        sensibly hasRadioluscence Radioluscence.
```

Figure 1 – GALEN example in GRAIL

In addition, GRAIL language is not interoperable with any other representational formalism that is applied to a web application. Recently developed web ontology languages such as RDFS, DAML+OIL, and OWL are getting more popular to be applied to ontology-based applications. Building a web application for the gross description in our research, therefore, we model a pathological ontology in a limited form based on the GALEN ontology and use representational formalism that populates the Semantic Web. In fact, the description of the gross exam is usually brief so that it is often regarded as the "macroscopic" description. Thus information required for the gross description does not necessarily contain a deep knowledge structure although it can be semantically linked to a deeper level ontology like GALEN or have been simplified from the Core Reference Model (Figure 2).

```
<rdfs:Class rdf:ID="GeneralStructure">
  <rdfs:subClassOf rdf:resource="#Structure"/>
</rdfs:Class>
<rdfs:Class rdf:ID="PlanarStructure">
  <rdfs:subClassOf rdf:resource="# GeneralStructure "/>
</rdfs:Class>
<rdfs:Property rdf:ID="hasColor">
  <rdfs:subPropertyOf rdf:resource="#hasFeature"/>
  <rdfs:range rdf:resource="#Color"/>
</rdfs:Property>
```

Figure 2 – Limited pathological ontology in RDFS

Layered Architecture of the PathOnt System

Connectivity and Interoperability in the Semantic Web

As discussed in the previous section, there are distinct levels of information models, from a fine-grained deep level ontology to only a surface level data description, each of which is handled in a different application. Performing a complex task electronically often requires information exchange among different applications so that data interoperation is essential.

The meaning of information in the semantic web must be explicitly expressed in terms of a semantic metadata language such as RDF for computer programs to "use" it. Data represented in an RDF statement at the meta-level is supposed to specify a semantic value for a property of a resource. At the meta-meta level, RDF Schema is intended to define property names that are used for a domain specific task. DAML+OIL whose syntax is built in the RDF(S) provides richer modeling primitives than RDF(S) since it is based on a model-theoretical semantics beyond that of

RDF(S). DAML+OIL is actually a kind of description logic language furnished with the features of frame-based languages. The OWL web ontology language following DAML+OIL is expected to be most popular.

Conceptualization explicitly represented in the form of an ontology helps both human and machine share and reuse information by reducing the ambiguities of the meanings of the terms used for a given task. An ontology is the result of cooperatively constructing the knowledge by the group of people who are ontology engineers, knowledge modelers, or domain experts. The semantic aspects of the information interoperability problems are possibly overcome by using ontologies for content explication.

Levels of Information Model in PathOnt

The complexity of building a Semantic Web based application can be reduced by regarding information models as a series of layers each of which may have multiple sublayers [6]. These layers include the syntax layer, the object layer, and the semantic layer.

On this view, the semantic layer deals with conceptual modeling and knowledge engineering tasks by providing a way of representing knowledge. The Semantic Web is a web where applications are interoperate in the semantic layer. We can of course think of a higher level of abstraction called ontology layer that provides a deeper level of knowledge representation framework. The conceptual model in the semantic layer can be expressed in RDF Schema. The semantic layer concerns interpretation of the object model in the object layer.

The object layer, then, provides applications with an object-oriented view of their domain. Data objects and binary relationships between them can be manipulated in this layer so that the ordered relationships are graphically represented in a modeling language like UML. The information model in the object layer can be expressed in RDF.

The syntax layer serializes data instances into a sequence of characters or bits. XML provides a generic document model in the syntax layer and instances of this model can be manipulated using APIs such as XML DOM.

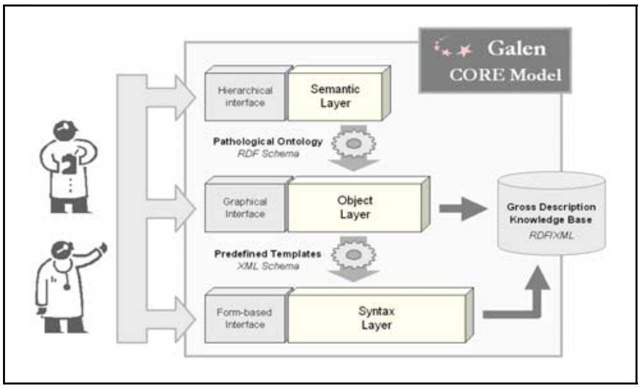

Figure 3 – Layered Architecture of PathOnt

Figure 3 depicts the layered architecture of the PathOnt system. Supporting different levels of information model, our system consists of the three application components: namely, PathOnt-Semantic, PathOnt-Object, and PathOnt-Syntax. At the highest level, the ontology engineer and the pathologist can work together to specify the ontology for the gross description. Using PathOnt-Semantic that automatically generates an RDF Schema file, we modeled the limited form of pathological ontology by referring to the GALEN Core Model. Making the gross description of an individual patient, at the next level, the pathologist can easily visualize the macroscopic findings of the specimen tissue in PathOnt-Object that is stored in an RDF file. At the lowest level, PathOnt-Syntax generates the data input form in XML. The three components of PathOnt do not exist independently but are designed to facilitate communications and collaborations among different professionals.

The Components of PathOnt

PathOnt-Semantic has functionality of defining concepts in the gross description as RDF Schema that is suitable for representing a lightweight ontology. Representing concepts and the relationship between the concepts, PathOnt-Semantic lists Class for resource type, Property for predicate, Domain for slot restriction, and Range for value restriction [5]. Figure 4 illustrates the window to input the pathological ontology in PathOnt-Semantic. The left pane of the PathOnt-Semantic displays the hierarchical structure of Classes and Properties whereas the right pane is to edit the slot information including the values of domains and ranges.

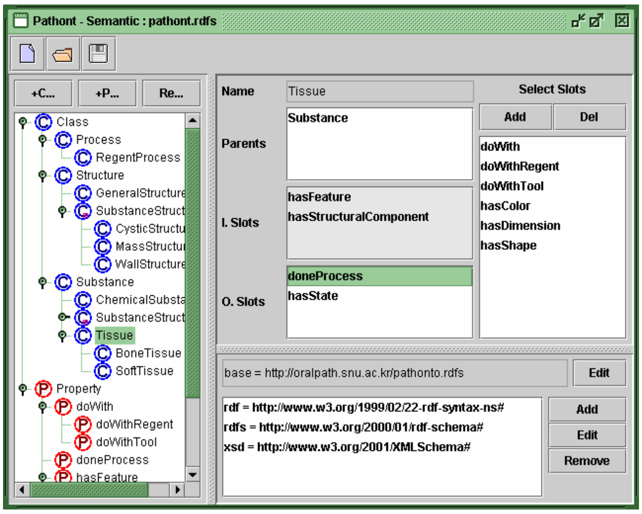

Figure 4 – Pathont-Semantic

The primary user of PathOnt-Semantic is either an ontology engineer working with a pathologist or a pathologist having knowledge of ontology modeling. On the contrary, PathOnt-Object is used to help a pathologist without knowledge of ontology engineering articulate the gross description of the specimen in a visual form. The RDF document created in PathOnt-Object is composed of statements, i.e. triples, which are complying with RDF Schema. The triples can also be represented with nodes and arcs graphically as well as in a sentential format so that the users understand and model the complicated ontological structure more easily. Figure 5 shows how the gross description in a text form is graphically visualized based on the ontological structure that has been made in

PathOnt-Semantic.

Specimen received in formalin is a dark brown soft tissue mass measuring 4.0 x 3.3 x 1.0 cm. It contains cystic area measuring 2.3x1.1cm. The cystic wall is thin and transparent. The representative portions are embedded.

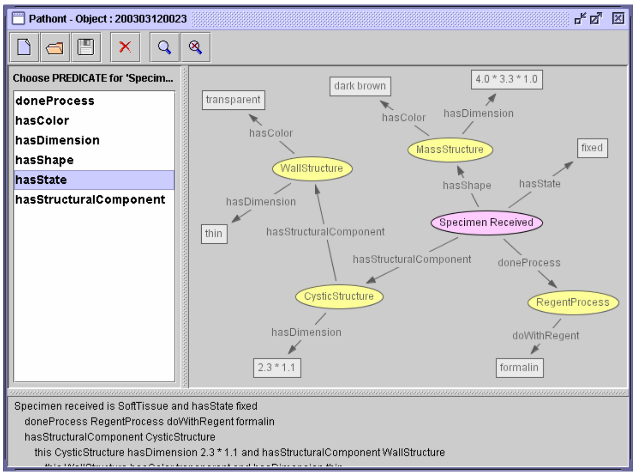

Figure 5 – PathOnt-Object

Another important role of PathOnt-Object is that the information model represented in PathOnt-Object can be stored in the gross description library to be reused as templates for similar cases. Since the template is written internally in a form of XML Schema, it can be transformed into a web-based input form through XSLT without difficulty [4]. As Figure 6 shows, the web-based input form in PathOnt-Syntax can be easily compatible with most common EMR applications. Since PathOnt-Syntax provides the user interface of the ordinary database input form, it can be easily used as the recording tool of gross description by a pathological technician or any other professionals. Its engine mainly creates HTML and CGI code for the web-based input form.

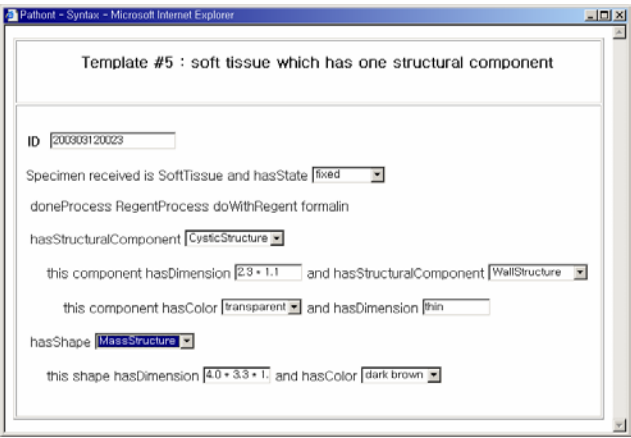

Figure 6 – PathOnt-Syntax

Conclusion

We presented a layered approach to building an ontology-based application for the pathological gross description using semantic web technologies. The layered architecture of our system articulated different levels in modeling and using medical knowledge, while dealing with at the same time issues of interoperability and connectivity from the technical perspective. In our research we implemented the PathOnt system for the computerized input of gross description that enables knowledge modelers and users with different roles to collaborate with each other. PathOnt that is built upon the reusable pathological ontology facilitates the interactions not only between end-users of medical information but also between different applications using semantic web markup languages.

References

[1] Association of Directors of Anatomic and Surgical Pathology. Recommendations for the reporting of breast carcinoma. *Am J Clin Pathol*, 1995; 104(6): 614-619.

[2] Association of Directors of Anatomic and Surgical Pathology. Recommendations for the reporting of resected large intestinal carcinomas. *Am J Clin Pathol*, 1996; 106(1): 12-15.

[3] Association of Directors of Anatomic and Surgical Pathology. Recommendations for the reporting of specimens containing laryngeal neoplasms. *Modern Pathol*, 1997; 10: 384-38.

[4] Decker S, Melnik S, Harmelen F, Fensel D, Klein M, Broekstra J, Erdmann M, Horrocks I. The Semantic Web: the Roles of XML and RDF. *IEEE Internet Computing*, 2000; 4(5): 63-74.

[5] Decker S, Mitra P, Melnik S. Framework for the Semantic Web: An RDF Tutorial. *IEEE Internet Computing*, 2000; 4(6): 68-73.

[6] Melnik S, Decker S. A Layered approach to Information Modeling and Interoperability on the Web. *Proc. of the ECDL'00 Workshop on the Semantic Web*, 2000.

[7] Patel-Schneider P, Building the Ontology Layer of the Web, IEEE Intelligent Systems, 2002.

[8] Rector AL. Medical Informatics. In: Baader F, Calvanese D, McGuinness D, Nardi D, Patel-Schneider P, eds. *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, 2003.

[9] Rector AL, Rogers J. Ontological Issues in Using a Description Logic to Represent Medical Concepts: Experience form GALEN. *Methods of Information in Medicine*, 2000.

Address for Correspondence

Myeng-Ki Kim is an associate professor in the dental college at Seoul National University. His research interests include medical concept representation, medical imaging, and clinical information system. Kim received a PhD and MS from Michigan university, and BS degree from Seoul National University. Interested readers may contact the author, via either meeree@snu.ac.kr or Department of Dental Management & Informatics at College of Dentistry, Seoul National University, 28 Yongon-dong Chongno-gu, Seoul, 110-749 KOREA