TOWARDS AN OWL MAMMOGRAPHIC ONTOLOGY

T.Podsiadły-Marczykowska*, H.Goszczyńska* and A.Guzik**

* Institute of Biocybernetics and Biomedical Engineering PAS, Warsaw, Poland
** Department of Radiology of Grochowski Hospital, Warsaw, Poland

teresa@ibib.waw.pl

Abstract: Breast cancer is the leading cause of cancer death among women in Poland. At present mammography is the only practical and reliable method for the early diagnosis of breast cancer. Reliable assessment of of diagnosis and patient management methods must be based on large data sets. One of the possibilities for building such tools is the use of ontology-based solutions. The paper presents efforts towards building on OWL mammographic ontology – MammoOnt. Primary use of mammographic ontology is to provide vocabulary and formal definitions of concepts for describing and interpreting breast X-ray films. Its other possible uses are: educational tasks, assistant tool for diagnosis, and semantic content-based search of mammogram database. The paper presents the notion of an ontology, its origin, and its definition within the context of artificial intelligence, as a formal and explicit specification of conceptualization. Ontology development methods, guidelines for good ontology design, cycle of development of mammographic ontology, implementation details and structure of the MammoOnt are described. Although our work is actually a work in progress , the instances of model's concepts are able to represent mammographic findings.

Introduction

Mammography and breast cancer, definition of an ontology, ontologies in biomedicine, intended uses of mammographic ontology, paper organization.

According to recent statistics [5] , breast cancer is the leading cause of cancer death among women in Poland . At present mammography is the only practical and reliable method for the early diagnosis of breast cancer. Assessment of effectiveness of diagnosis and patient management methods must be based on large data sets, so there is a need for tools that enable distributed, web-based collection of breast cancer data. One of the possibilities for building such tools is the use of ontology-based solutions.

The term ontology arises from philosophical tradition, and as a branch of philosophy is related to the study of being. Now the term is gaining a new role in many diverse fields of computing and biomedicine. Within the context of artificial intelligence an ontology is defined as a formal and explicit specification of conceptualization [1]. Conceptualization refers to an abstract model of some domain. Formal means that an ontology is an abstract organization of terms and relationships used as a tool for the analysis and representation of the relevant concepts in a domain of interest. Ontology should adequatly represent domain knowledge. It defines domain concepts and relationships between them, provides a vocabulary that is meaningful to humans and machines

Ontologies has already been defined for many biomedical and clinical domains such as genetics [6,15], anatomy [7] or pathology [8]. In this paper we present the work towards construction of OWL mammographic ontology, called MammoOnt - conceptual model of restricted subfield of radiology.

Primary purpose of our ontology is twofold. The first one is to provide vocabulary and formal definitions of concepts that can be used to describing and interpreting breast X-ray films. The second one is to use ontology as a specification for designing a database for mammography reports.

Other valuable uses of mammographic ontology are: educational tasks, assistant tool for diagnosis and semantic, content based search of mammogram database.

Materials and Methods

Ontology development methodologies, used method, and implementation details.

At present ontology building is more a craft than an engineering task [2]. There is no one , univocaly accepted methodology of ontology creating. TOVE [18,21] Methonology [19], or KBSI IDEF5 [20] are examples of some widely known methodologies. Ontology development methodologies are primarily inspired by enterprise modeling and software engineering. Exhustive survey of ontology developement methods can be found in [2,3,9]. The main common stages that can be derived from existing methodologies consist of following steps:

1. Identification of purpose and scope of the model
2. Knowledge acquisition
3. Informal specifications of concepts
4. Ontology formalization using ontology editor
5. *Evaluation
6. *Maintenance

 ISSN: 1727-1983

MammoOnt has been created basing on above presented cycle of development. Points 5 and 6 are out of reach of our current work *.

Different guidelines have been proposed for good ontology design [1,10 11,12,13]. In our opinion in those proposals there are two essential principles: Clarity [1,11,12,13] and Modularity [10]. Calrity refers to communication, the ontology should be well documented, its definitions and purpose clear to its intended users. Modularity refers to menagement. Correctly constructed ontology should consist of small, internally coherent components. Ontologies developed in this fashion are easier to reuse and suffer less form ontological commitments. Every stage in the development of the mammographic ontology has been guided by those principles.

Identification of purpose and scope. Primary aim of the MammoOnt is to represent medical knowledge about mammographic findings. The domain of the model is clearly defined, it remains to define the reach and granularity of that representation. One of the methods to determine it consists of writing a list of questions which the final knowledge base of the model has to answer. These questions are called competency questions and the method has been described in [21]. The requiremenets for ontology were gathered and formulated as a set of competency questions that the model must answer. Actually about 120 of competency questions have been formulated for mammographic ontology. A few of them are presented in table 1.

Table 1: Examples of competency questions

1. What findings should be described in mammography report
2. What are possible forms of pathologic findings in mammograms
3. What are necessary attributes for all pathologic findings
4. What are attributes of a mass in mammograms
5. What are possible interpretations of a mass finding
6. Is every spiculated mass a breast cancer
7. What are attributes of calcifications in mammograms

Knowledge acquisition. Developing ontology involves identifying the terms used in the domain of interest, because its basic function is to privide a common vocabulary by which users and systems can communicate.

Knowledge for mammographic terminology has been extracted from three main sources: corpus of routine, free-text mammography reports (around 400), interviews with radiologist and analysis of medical literature [4,14,16,17]. During the phase of knowledge acquisition manual methods have been used. Among medical literature concerning breast cancer BI-RADS [4] - standard vocabulary proposed by American College of Radiology deserve special attention. It provides basic lexicon for describing mammographic findings. In our work BI-RADS has been a starting point in construction of mammographic vocabulary. It has to be stated, that there is a discrepancy between terminology found in polish radiology reports and BI-RADS. There are terms in BI-RADS, that were newer used by polish radiologists, like for example calcifications morphology descriptors 'suture' or 'dystrophic', or calcifications distribution descriptors-'regiolnal' and 'segmental'. Calcifications description in BI-RADS does not include such property as density, but calcification density was quite often mentioned in polish reports. In our opinion BI-RADS system has to be extended to a national vocabulary. We are fully aware of the fact, that reaching a consensus within a mammographers community represents a great challenge.

Informal model. After gathering relatively complete vocabulary of mammographic terms, an initial set of concepts, their properties and relationships between them have been identified. Concepts have been structured into subsumtion hierarchy, properties of concepts has been modelled, to allow distinction betwen dirrerent kinds of mammographic lesions.

Ontology formalization. Ontology editors are tools that enable codifying browsing and modifying an ontology. They vary in their architecture and features support, an exhustive survey of ontology editors can be found in [22]. After a review of available ontology development environments we choosed Protege-OWL plugin in version. 3.1 build 205 to formalize and instantiate mammographic ontology. Our selection was based on the tool's expressivness, and flexibility. OWL-DL the most recent development among ontology languages (official W3C recommendation) is based on description logic, has rich set of operators (and, or and negation) and allows for reasoning and inconsistencies checking in an ontology.

Results

Diagnosis of breast cancer involves cooperation of experts from different medical background: radiologists, surgeons, oncologists, histologists and other medical staff. Thats why at the first level of the model we have identified as our basic concepts five general notions from medicine: Anatomy, Clinical Examination, Consultations, Diagnostic Procedures, Medical History, and finally Mammography.

ISSN: 1727-1983 © 2005 IFMBE

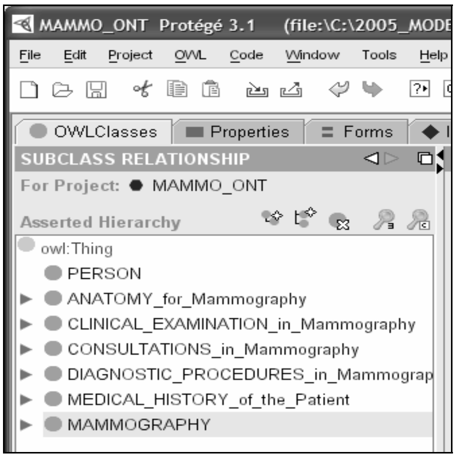

Figure 1: First level of the model - mammographic ontology is divided in seven main modules. The goal of modular design is to achive explicitness in the ontology, and to support reuse and maintainability

The modul Mammography is also divided into three sub-modules: Features of Mammographic Observations, Mammographic Observations and Report (see Fig.2).

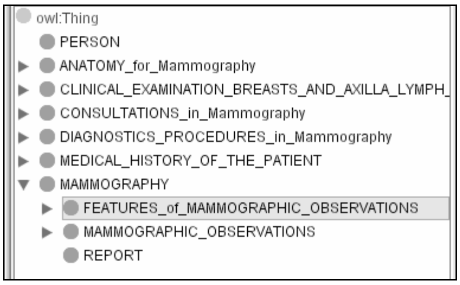

Figure 2: Sub-classes of Mammography class

The modul of Features of Mammographic Observations is crucial for the task of describing and interpreting X-ray breast images.

We identified the necessity of two main conceptual levels for accurate breast findings description. The first one, is the level of Visual Features of an object seen in an x-ray film. This level contains Visual Features of Mammographic Findings that radiologists indentify and describe from the image. The second level captures abstract, non-visible Features of Mammographic Observation: Radiological Diagnosis, Interpretation and Guidelines. Properties of Mammographic Observations are defined as individuals of the classes Visual Features and Non-Visual Features. Every Mammographic Observation is thus described in terms of its visual and non-visual properties (see Figure 3).

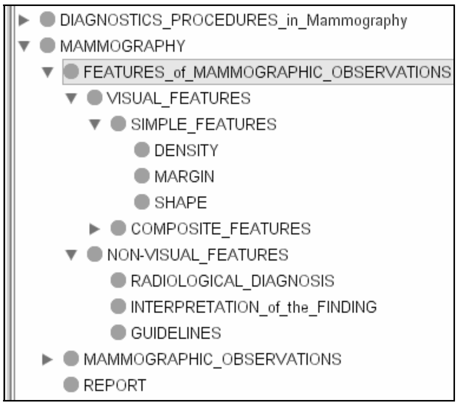

Figure 3: Structure of the modul Feature of MammographicObservations. Two main levels of properties of Mammographic Observations

The sub-modul of Mammographic Observations conteins concepts that should be described in a well constructed, sound mammographic report: Breast Composition, Findings in breasts, and radiological image of Axillary Lymph Nodes (see Figure 4) The contens of that level of mammographic ontology results form the answer to first competency question (see Table1). The class Mammographic Observation is divided into seven sub-classes describing pathologic findings in mammograms. Again the contens of a modul in a level of mammographic ontology results form the answer to competency question, the secod one (see Table1). Individuals of sub-classes of class Finding are with individuals of classes Breast Composition and Axillary Lymph Nodes form properties of class Report.

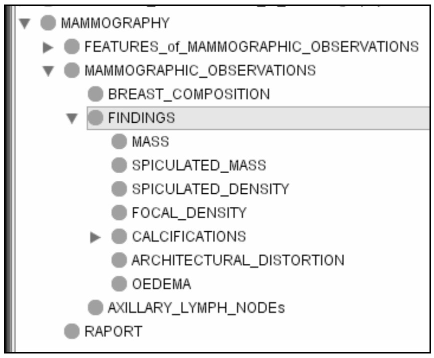

Figure 4: Sub-levels of class Finding, describing possible forms of pathologic finding in mammograms. Individual

Actually MammoOnt contains 119 classes and 105 properties representing mammographic finding. Preliminary evaluation of the model proved its ability to express significant mammographic findings.

Conclusion

Although our work is actually a work in progress , the instances of model's concepts are able to represent mammographic findings.

References

[1] GRUBER T.R., (1995). Towards Principles for the Design of Ontologies used for Knowledge Sharing. Int. Journal of Human and Computer Studies 1995 43(5/6): 907-928

[2] PINTO H.S.,MARTINS J.P., (2004) Ontologies: How Can They be Built? Knowledge and Information Systems, 6: 441–464

[3] BENCH-CAPON J.D., VISSER T.P., (1998) Methodologies for Ontology Development, in Proceedings of IT&KNOWS of the 15th IFIP World Computer Congress, Budapest, Hungary

[4] KOPANS D.B., D'ORSI C.J., ADLER DD.,. (1998) Breast Imaging Reporting and Database System: BI-RADS, American College of Radiology third edition 1998

[5] WOJCIECHOWSKA U., DIDKOWSKA J., TARKOWSKI W., ZATOŃSKI W, (2004) Cancer in Poland in 2002, Warszawa 2004

[6] THE GENE ONTOLOGY CONSORTIUM (2004) The Gene Ontology (GO) Database and Informatics Resources. Nucleic Acids Research, 1, 32, Database Issue pp 258-61

[7] ROSSE C., SHAPIRO I.G., BRINKLEY J.F., (1998): The Digital Anatomist Foundational Model: Principles for Defining and Structuring its Concept Domain, Journal of the American Medical Informatics Association 1998 pp 820-824 pp

[8] PASLARU BONTAS E., SCHLANGEN D., NIEPAGE S., (2005) "Ontology Engineering for the Semantic Annotation of Medical Data," 16'th International Workshop on Databases and Expert Systems Applications DEXA'05 pp 567-571

[9] LOPEZ M., PEREZ A.G. (2002), Overview and Analysis of Methodologies for Building Ontologies, Knowledge Engineering Review, 17(2), pp 129-156

[10] RECTOR A.L., (2003) Modularisation of Domain Ontologies Implemented In Description Logics and Related Fomalisms including OWL , in Proceedings K-CAP'03, October 23-25 pp 121-128

[11] STUDER R., RICHARD V. BENJAMINS R.V., DIETER FENSEL D. (1998) Knowledge engineering: Principles and methods. Data and Knowledge Engineering, 25(1-2):161–197, 1998.

[12] USCHOLD M., , GRUNINGER M., (1996) Ontologies: Principles, Methods and Applications. Knowledge Engineering Review, 11(2) pp 93-155

[13] VAN HEIJST G., SCHREIBER T., WIELINGA. B.J., (1996) Using explicit ontologies in kbs development. International Journal of Human and Computer Studies, 46 (2-3) pp 183-292 ,1996

[14] DEMIRKAZIK F.B., GULSUN F., FIRAT P., (2003) Mammographic Features of Nonpalpable speculated lesions, Clin Imaging Sep-Oct 17(5) pp 293-7

[15] YEH I., KARP P.D., NOY N.F., ALTMAN R.B. (2003) Knowledge Acquisition, Consistency Checking and Concurrency Control for Gene Ontology (GO) Bioinformatics, vol. 19 No2, pp 241-248 2003

[16] D'ORSI C.J., KOPANS D.B (1998) Mammogarphic Feature Analysis, Seminars in Roentgenology, 28, pp 204-229, 1998

[17] DZIUKOWA J. (1998) Mammografia w Diagnostyce Raka Sutka Copyright BEL CORP Scientific Publ. Co. Warszawa 1998

[18] GRUNNINGER M., FOX S, (1994) The Design and Evaluation of Ontologies for Enterprise Engineering. In Workshop on Implementes Ontologies , European Conference on Artificial Intelligence Amsterdam NKL 1994

[19] FERNANDEZ-LOPEZ M., GOMEZ-PEREZ A., (1997) JURISTO N. (1997) METHONOLOGY:From Ontological Art Towards Ontological Engineering. Symposium on Ontological Engineering of AAAI Stanford March 1997

[20] KBSI 94 (1994) The IDEF5 Ontology Description Capture Method InKBSI Raport Texas, 1994

[21] GRUNNINGER M., FOX S., (1997) The Role of Competency Questions in Enterprise Modelling IFIP WG5 Workshop on Benchmarking – Theory and Practice, Trondheim, Norway

[22] DUINEVELD A.J., STOTER R., WEIDEN M.R., KANEPA B., BENJAMINS V.R., (1999) WonderTools ? A Comparative Study of Ontological Engineering Tools, in proceedings of the 12th International Workshop on Knowledge Acquisition, medelling and menagement (KAW'99). Banff, Canada 1999, Kluwer Academic Publishers

ISSN: 1727-1983