Developing the GIMI Mammography Ontology in OWL 2 using Protege 4

Igor Toujilov and Paul Taylor

Centre for Health Informatics and Multiprofessional Education,
University College London,
i.toujilov@ucl.ac.uk, p.taylor@chime.ucl.ac.uk

Introduction

Radiology is moving rapidly from analogue to digital images. The roll out of Picture Archiving and Communications Systems means that more and more hospitals have access to archives of digital images. One obvious application of such image database is in training Since a large archive can contain a much greater range of appearances than a radiologist might encounter in the course of a typical training rotation, a computer tool based on the archive can greatly enhance a radiology trainee’s learning. Another application is in decision support, since the database contains images with different appearances and known diagnoses. If we can match the appearance of a new image to that of a set of database images, we have clues that can be used in diagnosis. Image databases can also be used in research, e.g. to train machine learning algorithms or in epidemiological surveys. Each of these applications requires an accurate description of the image contents. We argue that if the databases are to be used for many different purposes, then we need to put a great deal of thought into the description, and that ontology lanaguages provide a suitable approach to representing the required concepts.

We are primarily interested in the annotation of mammograms, or breast X-rays, initially to support a computer-based learning tool in which trainees’ assessments of mammograms are compared with experts’ annotations. For this application we want to be able to:

- detect a variety of errors, including failure to detect an abnormality, the misclassification of an abnormality or an inappropriate management plan following the assessment of an abnormality.
- give useful feedback about trainee’s decisions
- select appropriate cases for use in training a given individual
- provide instruction in the core principles of breast radiology

This work was carried out in the context of a larger project (GIMI: Generalised Infrastructure for Medical Informatics), which developed and trailed novel middleware allowing fine-grained control of access to data [Slaymaker et al.], a project of the UK Technology Strategy Board's Collaborative Research and Development programme.. We therefore wish to take a web services approach to our application, in which the image database and the training tool are separate but inter-operable applications.

A number of projects have already developed ontologies for radiology. The largest initiative of this kind is probably the RSNA RadLex terminology [Langlotz]. This was developed specifically to support the sharing of images for training applications, however the most current accessible version does not cover mammography. A ontology for mammography is proposed in [Hu et al.], however, it is derived from a terminology BIRADS which was developed as a standard for reporting mammograms, and which therefore omits much of the detail that is required to support training. In [Qi et al.], an ontology for mapping breast pathology and anatomy concepts to their radiographic appearance is proposed, but training-related concepts are not covered.

A number of other ontologies have been proposed to cover biomedicine more generally, including the NCI thesaurus, SNOMED CT and the Foundational Model of Anatomy (FMA). But these have only minimal conceptualisation of the breast, and does not cover subgross breast anatomy that is important to represent the parenchymal patterns. The NCI thesaurus and SNOMED CT deal poorly with image while the FMA has no radiological concepts at all.

Methods

Amongst ontology representation formalisms, OWL 2 is the most preferable for us because it provides advanced logical features (property cardinality restrictions, data ranges, etc.) that are absent in many other formalisms. However, due to the relative novelty of OWL 2, it is only beginning to be adopted in practical applications.

This presentation discusses the experience of using Protege 4 as a tool for ontology development in OWL 2. The aim was to implement as much as possible in the description logic, allowing a reasoner to handle almost all high level features, so that only some routine operations that cannot be implemented in OWL are handled at the application-level. This should allow a very flexible architecture, where most changes and extensions can be made at the level of the ontology, without requiring reprogramming of the application. This approach means that the ontology is much more than just a terminology, coding scheme, or taxonomy.

The ontology has the following distinctive features:

- Separation of visual appearance of radiographic features from their conceptual descriptions
- An extension OWL module of radiology training knowledge (learning points, and expert's and trainee's graphical annotation comparison
- Models of parenchymal patterns (Mammograms can be classified as having one of four parenchymal patterns, determined by the proportions of various 'building blocks' that they contain. The patterns affect interpretation strategy, and may also relate to risk.)

The ontology was developed in a number of iterations. Early work, described in Sun et al. involved a review of 400 cases annotated by radiologists at St Georges NHS Trust [Sun et al.]. Later work drew mainly on standards such as the Breast Imaging Reporting And Data System (BI-RADS), and a well-known textbook on the early detection of cancer [Tabar et al.].

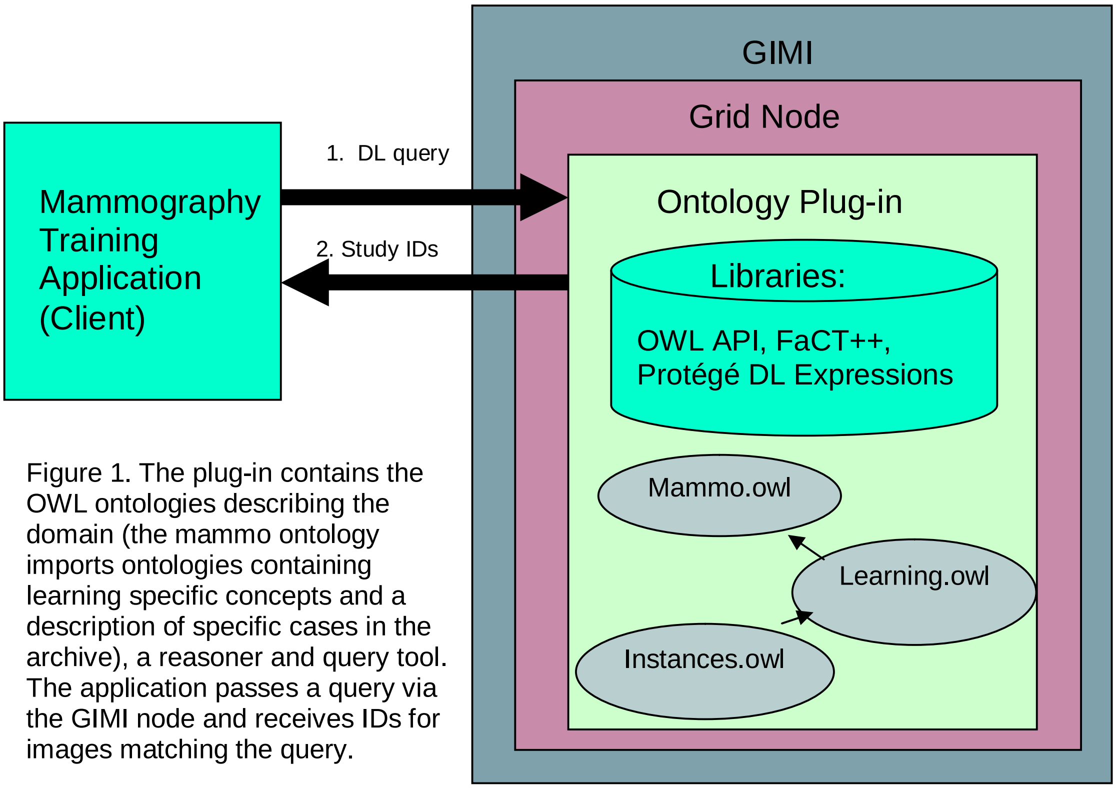

Figure 1. The plug-in contains the OWL ontologies describing the domain (the mammo ontology imports ontologies containing learning specific concepts and a description of specific cases in the archive), a reasoner and query tool. The application passes a query via the GIMI node and receives IDs for images matching the query.

Our training application uses a plug-in architecture. The ontology is uploaded to a secure GIMI node within a plug-in implemented using OWL API, FaCT++, and Protégé DL Expressions (Manchester Syntax) libraries. A client application, that provides training materials, can query the ontology on the server side using DL queries. This is illustrated in Figure 1.

Results and Discussion

The ontology has been developed to support a number of use cases. One example involves generating feedback for a trainee who is annotating a mammographic region. The e-learning application uses just-in time generation of an OWL module for this. This process involves the following sequence. The application calculates a measure of the match between trainee's and expert's geometrical annotations, instantiates the trainee's annotation appearance and the measured correspondence in the training session ontology module, and runs a reasoner. The reasoner classifies the measured correspondence as a mismatch, or undetermined match, or good match. In the case of match, the application collects the result of matching with the help of a DL query. Then with the help of other DL queries, the application can determine how trainee's annotations are consistent with the expert's ones.

Another important use case is the modelling of parenchymal patterns. The five patterns are defined according to relative proportions of four mammographic building blocks in the normal breast [Tabar et al.]: nodular densities, linear densities, homogeneous structureless densities, and radiolucent areas. A pattern is an individual characteristic of a person during a particular mammographic study. The ontology contains conceptualisation of subgross anatomic structures and building blocks that can be mapped to each other with the help of DL queries. For example, it is possible to resolve the following questions in DL: "How does adipose tissue appear on a mammogram?", and "Which anatomic structures do linear densities correspond to?" Study instances are automatically classified according to parenchymal patterns. Abnormal temporal changes of parenchymal patterns are also detected by a reasoner.

The GIMI mammography ontology can be accessed at http://gimimammography.svn.sourceforge.net/.

References

M. Slaymaker, D. Power, D. Russell, and A. Simpson. On the Facilitation of Fine-Grained Access to Distributed Healthcare Data. Proc. of the 5th VLDB Workshop. Secure Data Management, 2008, 169–184

C. P. Langlotz, RadLex: A new method for indexing online educational materials, *RadioGraphics 26* (6), 2006, 1595-1597.

B. Hu, S. Dasmahapatra, P. Lewis, & N. Shadbolt, Ontology-based medical image annotation with description logics, *Proceedings of the 15th IEEE International Conference on Tools with AI,* 2003, 77-82.

D. Qi, E. Denton, & R. Zwiggelaar, Linking image structures with medical ontology information, S. Astley, M. Brady, C. Rose, and R. Zwiggelaar, (eds), *Digital Mammography,* Springer Verlag, 2006, 399-406.

Sun,S., Taylor,P., Wilkinson,L., Khoo,L. (2008). An Ontology to Support Adaptive Training for Breast Radiologists. Proc. *International Workshop on Digital Mammography*, Krupinski,E. (ed.) Springer Verlag, 257-264

L. Tabar, T. Tot, and P. Dean, Breast Cancer - The Art and Science of Early Detection with Mammography. Stuttgart: Georg Thieme Verlag, 2005.