Ontology-Mediated Distributed Decision Support for Breast Cancer

S. Dasmahapatra, D. Dupplaw, B. Hu, P. Lewis, and N. Shadbolt

IAM, School of Electronics and Computer Science (ECS),
University of Southampton Southampton, SO17 1BJ, UK

Abstract. We have developed a prototype system to support decision making in Breast Cancer, wherein the varied nature of expertise is modelled by multiple ontologies that provide domain-specific grounding to concepts and relationships used. While the different medical experts need to be co-present at a meeting, our system employs a distributed architecture for handling data and invoking services appropriate for the requirements of this decision-making process. This distributed system is built upon Semantic Web technology, which enables the possibility of Web-based tele-medicine.

1 Introduction

The increasing necessity of incorporating specialist knowledge in critical medical domains has introduced a rearrangement of the sites of patient-doctor engagement. Thus, the patient's condition is described by different specialists, offering complementary views which determine the course of intervention or treatment. In the domain of breast cancer, this multiplicity of views is the norm, so a decision support environment for this domain needs to incorporate the distribution of sources and type of knowledge, and provide methods to collate the case-specific information based on the concepts used in the various specialist domains. In particular, the domains of interest that are relevant involve radiology, X-ray, magnetic resonance images (MRI) and ultrasound images are routinely used; pathology, histopathologists and cytopathologists produce microscopic slides from cells and tissues extracted from the patient; and clinicians, oncologists and surgeons also deliberate on possible interventions they can suggest based on prognostic factors implied by the interpretations of these images.

The integration of heterogeneous types and formats of information sources for the purpose of retrieval is one of the envisaged positive outcomes of Semantic Web technology [1]. In this paper, we describe a prototype system for providing decision support which used Semantic Web technology—domain ontologies for describing the relevant specialist information as indicated, a declarative (ontological) characterisation of the methods and services that are invoked in order to enable relevant information flows within a Web-based distributed environment. This provides a flexible open-ended architecture for the integration of services

S. Miksch et al. (Eds.): AIME 2005, LNAI 3581, pp. 221–225, 2005.

which can be described at the knowledge level, and hence possibly chained together in order the requirements of the users' decision-making deliberations. The application building process is also made more flexible with distributed service implementers having control over their methods which are called through a mediating task registry. This is crucial in a medical application such as ours, as institutional control over data is critical as ethical issues create an independent channel regulating the flow of information within the system.

2 Domain Ontologies for Breast Cancer

The medical domains represented by ontologies in this project are X-ray mammography, MRI, ultrasound imaging and histopathology. These were developed independently of each other using a variety of sources and implemented in the Web Ontology Language (OWL-DL). For X-ray, MRI and ultrasound, we could make use of lexicons developed in standardising efforts of the American College of Radiologists, called BI-RADS[3]. The histopathology ontology was created from UK National Health Service guidelines for reporting [2] in breast cancer cases as well as from papers in medical journals. The National Cancer Informatics (NCI) ontology [4] is a very useful resource, but its sheer size (38MB) and breadth of coverage meant having to throw out most of the terminological concepts that were not relevant to breast cancer. Moreover, the terms that we required in order to provide terminological support for the cases that we were provided by our medical partners were not completely covered in the NCI ontology. In addition to these medical specialist terminological coverage, we also have an ontology to describe basic patient information like age, a description of the procedures that she may have undergone and so on. Within each class in each of the ontologies, sub-concepts are usually introduced not just with expressive names borrowed from the lexicon, but as refinements with respect to values in the range of some discriminatory relation, such as morphological features taking particular descriptive names as values.

The principal role these ontologies play in our system is to provide a framework within which relevant information pertaining to a case can be retrieved and related information reached by navigation. To that end, we have a mechanism of exposing aspects of the web of conceptual terms to the user should it be necessary or desirable (as illustrated in Figure 1).

3 Distributed Decision Support

In this section, we give an overview of the system architecture which allows the inclusion of different knowledge-based services. The system design is a declarative specification based on information flow within the system, which is driven by the demands for the relevant knowledge by the medical users of the system. Access to information is contrained by legal and ethical concerns, and we have assumed local storage under institutional custodianship. Any algorithms or clients that are instantiated through the system retrieve the patient information, such as x-ray images directly from these distributed sources.

As with the domain ontology where the interactions between the different sources of knowledge about a case was organised into the separation of different groups of concepts, here too we develop an "service composition ontology (SCO)" that organises the different services that are invoked and trigger information transfers around the system, so that the client can invoke the available distributed services described. Apart from the client and the server, a general task invocation framework allows remotely developed and appropriately packaged modules to be called upon, should they be relevant to the knowledge handling requirements for use. This too is described in SCO and is run on a remote server.

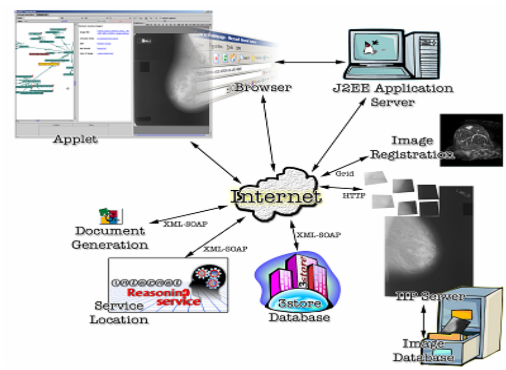

Fig. 1. The System Architecture

Declarative Specifications and Applications

The task invocation sub-system uses a number of different mappings to provide task-level invocation of functionality on disparate systems. The idea in the task invocation sub-system is first and foremost to protect the client application from changes in the remote web-resources. Secondly, it provides a clean interface for the execution of web-resources that are accessed through different interfaces. The flexibility provided by the system also provides a good base for application deployment.

In the task invocation framework, a Task Registry plays the analogous role, but here it is the inputs and outputs of the services that have to remain uncorrupted, not the internal workings of the modules. The Task Registry is a dictionary-like structure that maps tasks on to task implementations. The task's identification (its name and other parameters) is represented by a Task Description. A specific task implementation may be used in more than one task description, although new instances of that implementation are used in each task. Both task descriptions and task implementations can have default arguments and both can have mappings applied to any argument names. This also ensures flexibility between the task descriptions and the task implementations.

Knowledge Services

Our system provides a generic platform to compose various medically relevant services designed particularly for Breast Cancer Domain. A number of such services are provided by some of our partners in the project from other universities and they have been and are being written up elsewhere. These include MRI diagnostic classification tasks, natural language generation from case annotations against our domain ontologies, image registration, and so on. Here, we report on some of the knowledge rich diagnostic support services we have implemented. The architecture accommodates, among other methods, classification services based on features automatically extracted from images or its hand-drawn segments. In addition, in the multi-disciplinary meeting for patient management, the cases already come with expert labels attached—shape features for masses seen in X-rays, for example. We have also built classifiers which take BI-RADS descriptors and other ontological metadata labels for X-ray images as input and classify cases according to their likelihood of being benign or malignant. We have packaged a few classifiers – trained on 1500 cases from the Digital Database for Screening Mammography (DDSM) of the University of South Florida [5] – to run as services that can be accessed remotely via task invocation methods. Also, with the descriptive framework for task invocation in place, the future availability of semantic web service description standards could then be used to advertise such capabilities in appropriate semantic web service repositories.

4 Related Work and Looking Ahead

While there are several aspects of the system we have not described here, we would like to indicate some future directions we would like to go in. As indicated in Figure 1 one of the service modules delivers image feature vectors from X-ray images—these are grayscale histograms, shape, perimeter and area measures on hand-annotated closed curves on the image indicating regions of interest, wavelet features, and so on. Association of features of an image or region of interest with the concepts in the ontology may be manually or semi-automatically generated, or retrieved from legacy data. Association of a feature to concepts defined in the ontology is provided by a simple point and click mechanism. The user highlights the feature which they are going to associate with a concept, and finds the relevant concept in one of the concept browsers to right click and link. The feature vector is stored in a feature database, which provides indexed, feature-dependent retrieval of features, and the unique ID of the feature vector is inserted as an instance of the given concept. This image feature extraction algorithms are provided to the client mainly through the web-service interface. This API provides functionality for storing, retrieving, and comparing feature vectors from images, and automatically provides feature modules with the relevant regions from the source media. There is another use to which we can put these classifiers. As different hospital practices and radiologists use slight variations in ontological terms and relations in describing x-ray images, the collation of information for statistical analysis becomes difficult. We will use these image to concept label classifiers to assist the ontology mapping process, where the system will enable the remote exchange of images.

Recent work in feature extraction for diagnosis is featured in [6], which framework would be amenable to the word-picture matching problem[7]. As for the system presented, a quite similar architecture was described in [8]. The content-based soft annotation (CBSA) tool [10] uses Bayesian classifiers to commute semantic labels between image sets, providing content based semantic labeling. Marques and Barman [9] use feature clustering and ontological labeling to provide semi-automatic image annotation. They map the image into feature space (based on automatically extracted colour, shape and texture features) and retrieve the ontological concepts that been associated with the nearest cluster.

Future work on this system and the requirements of the domain will initially be drawn from the ideas presented above. As a proof of concept of a prototype for this domain, the system succeeds in pulling together web services and Grid based image registration services through an ontologically indexed architecture. While we have had substantial clinical input, we are currently trying to use this framework in a limited clinical setting.

References

1. T Berners-Lee, J Hendler, O Lassila. "The Semantic Web." *Scientific American*, 2001.
2. R Wilson, D Asbury, J Cooke, M Michell and J Patnick (eds.). Clinical Guidelines for Breast Cancer Screening Assessment. NHSBSP Publication No 49, April 2001.
3. American College of Radiology. Breast Imaging Reporting and Data System: BI-RADS.
4. The National Cancer Informatics Thesaurus and Ontology. Available from `http://www.mindswap.org/2003/CancerOntology/nciOncology.owl.gz`
5. M. Heath, K.W. Bowyer, D. Kopans et al. Current Status of the Digital Database for Screening Mammography. *Digital Mammography,* p 457-460, Kluwer Academic Publishers, 1998.
6. P Sajda, C Spence and L Parra. A multi-scale probabilistic network model for detection, synthesis and compression in mammographic image analysis *Medical Image Analysis* 7 (2003) 187204.
7. K Barnard, P Duygulu, N de Freitas, D Forsyth, D Blei and M I Jordan. Matching Words and Pictures. *Journal of Machine Learning Research*, 3, pp 1107-1135, 2003.
8. M Addis, M Boniface, S Goodall, P Grimwood, S Kim, P Lewis, K Martinez, A Stevenson. SCULPTEUR: Towards a new paradigm for multimedia museum information handling. *International Semantic Web Conference (ISWC)* 2003, October 20-23, 2003, FL.
9. O. Marques, and N. Barman. Semi-automatic semantic annotation of images using machine learning techniques. *International Semantic Web Conference (ISWC)* 2003, October 20-23, 2003, FL.
10. E. Chang, K. Goh, G. Sychay, and G. Wu. CBSA: Content-Based Soft Annotation for Multimodal Image Retrieval Using Bayes Point Machines. *IEEE Transactions of Circuits and Systems for Video Technology*, 13(1):26-38, January 2003.