SEMANTIC SUPPORT FOR MEDICAL IMAGE SEARCH AND RETRIEVAL

Wang Wei, Payam M. Barnaghi
School of Computer Science and Information Technology
The University of Nottingham Malaysia Campus
{Kcy3ww, payam.barnaghi}@nottingham.edu.my

ABSTRACT
The need for annotating digital image data is recognised in a variety of different medical information systems, covering both professional and educational usage of medical imaging. Due to the high recall and low precision attribute of keyword-based search, multimedia information search and retrieval based on textual descriptions is not always an efficient and sufficient solution, particularly for specific applications such as the medical diagnosis information systems. On the other hand, using image processing techniques to provide search on the content specific data for multimedia information is not a trivial task. In this paper we use the semantic web technologies in medical image search and retrieval process for a medical imaging information system. We employ an ontology-based knowledge representation and semantic annotation for medical image data. The proposed system defines data representation structures which are given well-defined meanings. The meanings are machine-accessible contents which could be interpreted by the software agents to find and retrieve the information based on the standard vocabularies and meaningful relationships between the data items.

KEY WORDS
Medical Data Representation, Medical Imaging, Ontology-based Search, Semantic Annotation

1. Introduction

The image annotation and meta-data explanation can provide high level description for the image data. It is, however, time consuming and thus expensive to "find" different information in the huge annotated image databases. Creating annotation based on automatic feature extraction and image processing techniques provides too low level information for many applications. The difference between the low level feature descriptions provided by image analysis tools and the high level content descriptions required by the applications is often referred to, in the literature, as the “Semantic Gap” [12]. The manually or (semi-)automatic text-based explanation of image data enables the information publishers to fulfil this gap. The description of medical data is normally presented in terms of controlled, but diversely expressed vocabulary and natural language specifications. These descriptions usually require additional human intervention for data processing determination of relevancy between different data items. In order to facilitate machine-based reasoning for better information retrieval additional interpretive semantics must be attached to the data. This requires a move from data-centric approach to knowledge and semantics description models [6].

This paper concentrates on studying of components for building a medical knowledge-base system. In particular, we focus on controlled terminologies and medical ontologies to develop a medical imaging information search and retrieval system supported by the semantic web technologies. The work describes reusable information blocks and information retrieval agent’s design for medical imaging. The semantic web technology improves comprehension of the content and decreases search time and cost to retrieve the required information. It reduces manual discovery and usage of resources and allows software agents to automatically determine the resources, interpret and integrate them to provide an improved information search and retrieval. The information and knowledge representation structures such as the Resource Description Framework (RDF) [1] and Web Ontology Language (OWL) [2] are used to build flexible and interoperable knowledge representation which is able to facilitate progressive information search and retrieval mechanisms. Using the semantic web technologies for data representation provides outstanding advantages over the traditional data models such as XML and relational databases. The semantic web applications employ ontologies to provide a common understanding of a discourse domain. The ontology description using public standards such as RDF(S) and OWL which use XML serialisation provides platform independent, interoperable, scalable and reusable content representation structures. In such ontology meaningful between concepts are defined and enable the software agents to extract explicit knowledge from the implicit information represented through the contextual relationships between the data items. The ontology describes intensive and explicit data and enables the users

to look up concepts from the provided structure and form improved queries, even if they do not have apparent idea about the discourse domain. Consequently, comprehensive queries can be submitted and answered by an information retrieval system. In this paper we use human lung X-ray images, and related disorders specifications to describe the architecture of the medical imaging search and retrieval system and evaluate the results.

2. Ontology-based information retrieval for multimedia data

There are a number of works that have focused on using the semantic web technology and in particular ontology-based image search and retrieval for different domains, such as, animal species [3], artworks [4], [5], breast cancer [6] etc. These systems use ontology-based description of discourse domain which provides a rich, explicit specification of the concepts and their relationships in a particular domain. The main goal is to answer the user queries based on semantic relations that can be inferred from meaningful between the data items

Most of the existing works are concerned on ontology-based knowledge representation and information extraction. However, they do not fully exploit the advantages of the ontology-based information search and retrieval in the context of an integrated semantic annotation and medical information search and retrieval system. In medicine, information system means handling and integrating data in ways that enhance diagnosis and treatment processes. The semantic web architecture enables to implement an emergent database of medical imaging information that has many of the characteristics of flexible systems: it is extensible to support new data explanation attributes, the explanations could be distributable across institutions, the expressions could be established based on common specifications and vocabularies. It also enables the expert user to add new and possibly conflicting knowledge to its existing knowledge-base (using ontology-based mediations) [7], [13], [14]. In the current system, the image data is annotated based on the concepts defined in the domain ontology and the user selects a concept from the ontology and the system performs a progressive search to find related images based on the different attributes.

3. The ontology development

We use the National Cancer Institution (NCI)'s standardised and authoritative meta-thesaurus to describe the domain ontology concepts and relationships. The NCI's meta-thesaurus provides profound domain knowledge compared to the other clinical vocabularies and taxonomies [8]. The meta-thesaurus defines comprehensive concepts related to human anatomy, diseases, findings, disorders, medical techniques and so on. The specified features reasonably satisfy the requirements of the image annotations. We implemented a domain ontology represented in OWL form based on the concepts and relationships taken from NCI meta-thesaurus.

The system also utilises an image annotation and description ontology which is used to specify the data explanation structure and properties for the medical image data. The content description attributes of the annotation model are then associated to the domain ontology concepts for each particular image data. This enables the system to provide the meta-data description according to the standard vocabularies and concepts defined in the domain ontology. The specification of an images data also needs to include information about the media dependent attributes (e.g. size, width, height, resolution, etc) to satisfy both the content selection and presentation requirements. The interdependencies amongst the image data should be encapsulated through set of classes, relations, functions and object constraints for the domain of discourse. Figure 1 illustrate a fragment of the domain ontology and Figure 2 shows the structure of the image annotation and description ontology.

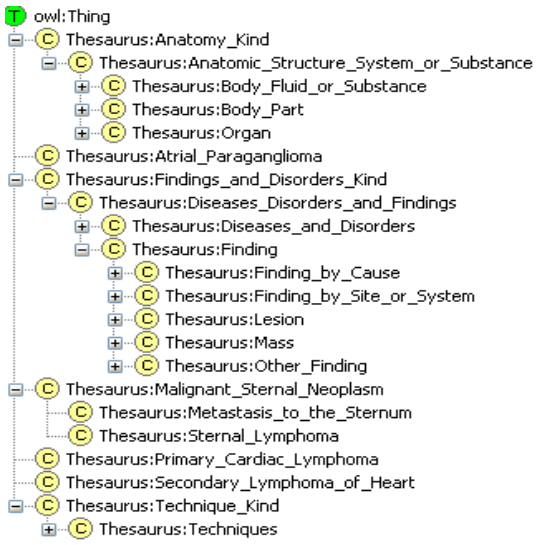

Figure 1. The domain ontology descriptions

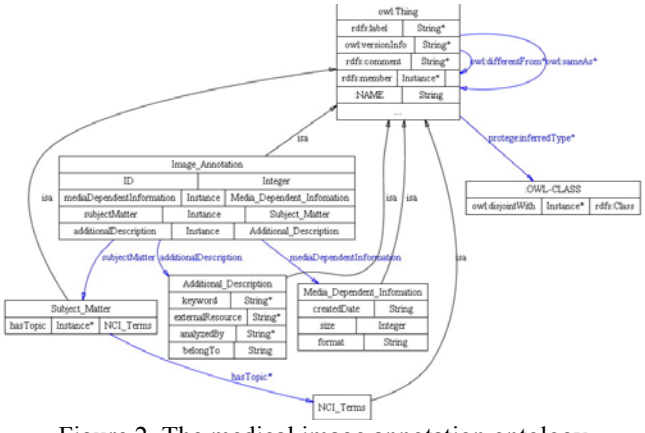

Figure 2. The medical image annotation ontology

The annotation ontology consists of different types of information specified in three groups. The media dependent information component which describes media dependent and technical attributes of image data. The subject matter component is associated to the domain ontology concepts, and describes the intrinsic content-specific features of a medical image. The additional description section provides additional information such as special keywords which are used when there are medical terminology discrepancies between different groups or knowledge representation environments.

4. The medical imaging information system

The system aims to providing a complete record of related data to the medical images, including the patient monitoring data, similar cases from the other patients' records, related documents to enhance diagnosis and treatment. The relevant information is obtained based on the meaningful relationships to the specific record.

The system includes four main layers as shown in Figure 3. The ontology layer contains annotated data and domain ontologies represented in RDF/XML format. The data storage layer provides data manipulation and retrieval facilities. An open source RDF repository [9] is used to store and query RDF/XML representations. The system also employs a relational database to store the binary image data.

The API layer uses RDF repository's application API to query the RDF representations. The application layer contains an annotation tool which is used to load the image objects and add meta-date explanations to the image data according the annotation ontology specifications. The tool enables the users to add new data items to the system and edit/update the existing explanations. The annotation interface provides the specifications based on the structure of the annotation ontology and requires specifying the descriptions according to the domain ontology concepts. As a result, users would be able to annotate images without requiring prior understanding of the OWL and RDF syntax.

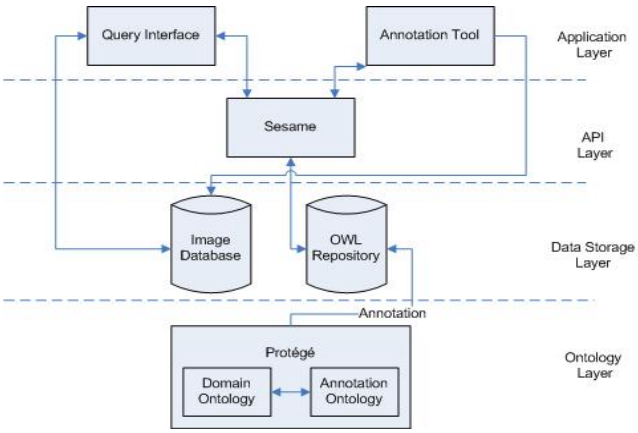

Figure 3. The system architecture

The query interface enables the user to browse the domain ontology concepts and select the terms and relations. This facilitates selection of the properties and terminology for information discovery purpose. The user is able to submit a keyword-based query or to select the terms from the user interface menus according to the domain ontology concepts. The system searches the annotations to find the matching objects. The query menu shows the query topics which correspond to the domain ontology entities. One can browse the topics and select a subject (and/or related sub-topics) from the domain ontology concepts. For example a user can select to see all the images related to a specific disorder or a particular anatomy class. Figure 4 shows the system's user interface. We have implemented a reasoning engine to find and interpret meaningful relationships between the concepts to address the relevant data. The reasoning engine is responsible for searching for semantic associations between the objects based on the different attributes. This enables the medical experts to find similar patterns in the appearance of specific symptoms or other multiple cases based a specific attribute. It helps the medical staff to search and retrieve electronic patient records containing interlinked medical images, physician's annotations, and references to previous medical examination results.

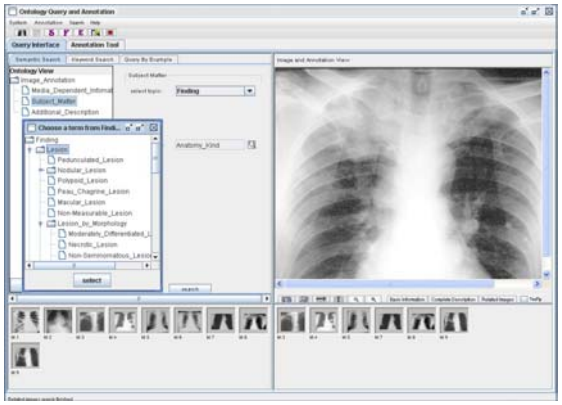
Figure 4. The user interface

The user selects an image, and then the image's properties are shown to the user (such as findings, diseases or abnormalities). The properties are then used to address the related data. The user can focus on a specific aspect of the data to find the relevant information. For example, if a selected image has a "Mass" value for its "finding" attribute (which depicts the main finding of a chest X-Ray image), then the images containing sub-classes of the "Mass" class (according to the domain ontology specifications), such as "Diffusion Mass", "Chest-Wall Mass, etc will be also considered as related data. This provides an enhanced process to discover the relevant information compared to manual and multiple queries to extract the required information. The automatic discovery of the semantic associations and addressing relevant information based on the meaningful relationships between them is also referred to as the

semantic recommendation [10]. In this context, the semantic distance between entities of the domain (i.e. the number of properties that relate one entity to another) could be used to determine the semantic recommendation [10]. In this case, the distance factor would be a relation robustness factor to emphasise more relevant information in a set of related data.

5. Evaluation of the results

To evaluate the results and show the effectiveness of ontology-based search and retrieval for medical data, we have compared the system with MedPix [11]. MedPix is an online medical image database which is provided by the Departments of Radiology and Biomedical Informatics, Uniformed Services University, Bethesda, MD and includes different samples of annotated MRI and CT Scan X-ray images. The contents in MedPix database are organised based on disease category, anatomy, and patient profiles. The objective of evaluation is to examine the relevancy and efficiency of these two different systems in searching and retrieving medical images.

We are also aware that there are other factors which affect the comparison, such as the quality of the annotations, the association of the keywords to the concepts in the domain ontology; the terminology being used in the domain ontology etc. For the comparison purpose several lung diseases and symptoms are deliberately chosen for the testing purpose, such as “Hilar Mass”, “Primary Lesion”, “Inflammation”, “Cough”, “Vascular Neoplasm”, “Neoplasm”.

The search in MedPix returns all the objects which have text-based annotations similar to the queried terms. Search for for “Hilar Mass” in MedPix returned 64 images, and we examined 24 of the images which seemed to be more related to the queried term (highly relevant results). Amongst the 24 images, 4 images were duplicated, 2 images had annotation “without evidence of a mass”, and 2 images were marked having “mass effect”. The rest of the images had keywords such as “pararenal mass”, “spiculated mass”, “ring-enhancing mass”, “fluid filled mass” which were not precisely related to the proposed query term and were only found based on text similarity.

Using the implemented ontology-based search tool, two images containing annotation of “Hilar Mass” were retrieved. After performing a related images search based on the first image (i.e. semantic search), another 7 images which had some kind of relationships to the selected image were also retrieved (the system also specifies the type of relationship). For the second image another 9 images were addressed as relevant images. Although some of the related images retrieved did not have the annotation text similar to the initial query term, they had some meaningful relationships to the proposed query term. The results can be very significant for the users who are searching for such semantically related images at a conceptual level. In keyword-based information search and retrieval systems like MedPix, the images are addressed individually and based on keyword-based query. However different medical experts could have different perspectives to search for the medical data (i.e. searching based on a specific attribute and the relationships) or different terminologies could be used to explain the same image. We believe that using the ontology-based information search and retrieval allows them to select the terms and concepts to be searched from a well-defined domain ontology which defines a standard controlled vocabulary to describe the entities.

6. Conclusion

This paper describes an ongoing research project aiming to implement a flexible, extensible and interoperable framework for medical image annotation, search and retrieval systems. We discuss an ontology-based approach to search and retrieve medical images. The proposed system provides an efficient solution to discover the information about the queried term by using high level knowledge representations and reasoning. We have used the semantic web technologies to represent the meaningful relationships between the entities in a discourse domain and in particular medical information for human lung. We employ an image annotation and description structure which provides enriched medical image data explanations. The annotation attributes are associated to the domain ontology specifications to provide common and standard terminology to describe the medical data. The system employs a reasoning mechanism to search for semantic associations and meaningful relationships between the data objects. The users are able to retrieve related image data based on those relationships and associations rather than only using keyword matching techniques. The system aims to enhance the information discovery process and allows similar data and patterns to be found based on specific symptoms or other attributes. The system helps to improve the medical diagnosis and treatment process. The future work focuses on ranking the semantic search results, and providing information search and retrieval methods that have special emphasis on reasoning biomedical data and in particular medical images. The project also aims to implement semantic web services to communicate with other information resources to acquire relevant biomedical information from distributed heterogeneous resources.

References

[1] The Resource Description Framework (RDF), W3C Recommendation, available at: <http://www.w3.org/RDF>
[2] L.McGuinness, F. Harmelen (editors), Web Ontology Language (OWL), W3C Recommendation, 2004, available at: <http://www.w3.org/TR/owl-features/>

[3] G. Schreiber, B. Dubbeldam, J. Wielemaker, B. Wielinga, "Ontology-Based Photo Annotation", *IEEE Intelligent Systems*, *16*(3) PP 66 -74, 2001.
[4] E. Hyvoenen, A. Styrman, S. Saarela, "Ontology-Based Image Retrieval", *HIIT Publications,* pp. 15-27, 2002.
[5] L. Hollink, G. Schreiber, J. Wielemaker, "Semantic Annotation of Image Collections", *In the Proceedings of workshop on Knowledge Mark-up and Semantic Annotation,* KCAP'03, 2003.
[6] B. Hu, S. Dasmahapatra, P. Lewis, N. Shadbolt, "Ontology-based Medical Image Annotation with Description Logics", *In Proceedings of The 15th IEEE International Conference on Tools with Artificial Intelligence,* pp. 77-82, 2003.
[7] M. Halle, R. Kikinis, Flexible frameworks for medical multimedia, *Proceedings of the 12th ACM International Conference on Multimedia,* pp. 768-775, 2004.
[8] J. Golbeck, G. Fragoso, F. Hartel, J. Hendler, J. Oberthaler, B. Parsia, "The National Cancer Institute's Thésaurus and Ontology", *Journal of Web Semantics, vol. 1,* no.1, 2003.
[9] J. Broekstra, A. Kampman, F. Harmelen, "Sesame: An Architecture for Storing and Querying RDF and RDF Schema", *In Proceedings of the First International Semantic Web Conference (ISWC 2002),* pp. 54-68, 2002.
[10] E. Hyvonen, S. Saarela, K. Viljanen, "Ontogator: Combining View- and Ontology-Based Search with Semantic Browsing", *In proceeding of XML Finland 2003, Open Standards, XML and the Public Sector,* 2003.
[11] MedPix, Free Online Medical Image Database, available at:
<http://rad.usuhs.mil/medpix/medpix.html>
[12] J. van Ossenbruggen, R. Troncy, G. Stamou, Image annotation on the Semantic Web, W3C Working Draft, available at:
<http://www.w3.org/2001/sw/BestPractices/MM/image_annotation.html>
[13] Y. Lee, C. Patel, S. A. Chun, J. Geller, "Compositional knowledge management for medical services on semantic web", *International World Wide Web Conference 2004,* pp. 498 - 499, 2004.
[14] J. Kim , D. D. Feng, T. W. Cai, "A web based medical image data processing and management system", *ACM International Conference Proceeding Series; vol. 9,* pp. 89 - 91, 2000.