

A Teaching System for Assistance in the Interpretation of Mammograms and the Diagnosis of Breast Cancer

Denise Guliato*, Ricardo S. Bôaventura, Mariangela S. Smedo, Rangaraj M. Rangayyan
and Túlio Augusto A. Macedo

Abstract—This work proposes the design of a teaching system for training students in the interpretation of mammograms and diagnosis of breast cancer. The proposed system integrates an illustrated tutorial on radiology of the breast, that is, mammography, which uses education techniques to guide the user (doctors, students, or researchers) through various concepts related to the diagnosis of breast cancer. The user can obtain informative text about specific subjects, access a library of bibliographic references, and retrieve cases from a mammographic database that are similar to a query case on hand. The information of each case stored in the mammographic database includes the radiological findings, the clinical history, the life style of the patient, and complementary exams. The breast cancer tutorial is linked to a module that simulates the analysis and diagnosis of a mammogram. The tutorial incorporates tools for helping the user to evaluate his or her knowledge about a specific subject by using the education system or by simulating a diagnosis with appropriate feedback in case of error. The system also makes available digital image processing tools that allow the user to draw the contour of a lesion, the contour of the breast, or identify a cluster of calcifications in a given mammogram. The contours provided by the user are submitted to the mammographic database for evaluation. The teaching system is integrated with AMDI — an Indexed Atlas of Digital Mammograms — that includes case studies, e-learning, and research systems. All the resources are accessible via the Web.

I. INTRODUCTION

Medical education is a challenge because it requires a vast range of material; intellectual, visual, and tactile skills; and the incorporation of large amounts of factual information. Traditionally, medical teaching is based on texts, lectures, and bedside teaching, with self-guided individual learning based on books [1]. The traditional medical teaching and individual learning methods can be complemented with electronic learning systems delivered on the Internet, leading to the so-called e-learning paradigm. With the advance of

Web technology, e-learning has been a topic of interest in recent years, and has become an important trend in medical education. Alves et al. [1] describe their experience in designing a multi-agent system for e-learning in the medical arena. The proposed system is based on the client/server network architecture, where the server holds the data warehouse, medical data, and the software agent that delivers medical content; on the client side, the interface is based on internet technology. Shaw et al. [2] and Johnson et al. [3] presented the Adele system, which includes the pedagogical agent Adele, for assisting students as they assess and diagnose medical and dental patients in a clinical setting. In a typical use of Adele, the student is presented with a computer simulation of a clinical problem. The integrated system is downloaded and run on the client's side. Shyu et al. [4] proposed a system to establish a virtual medical school as the platform of an e-learning center, which provides a problem-based e-learning environment, and utilizes the hospital information system to capture and store clinical cases; medical students and residents can access the clinical cases online.

An e-learning platform consists of various complex activities, such as an *authoring system* that helps in creating and exchanging content, a *learning management system* that stores and manages content, and a *run-time system* that interacts with the user. These activities can be perceived and modeled as processes and executed as workflows. Vossen and Westerkamp [5] present a bibliography and a prototypical system using workflow for managing activities of learning.

To add flexibility, interoperability, and reuse, the activities that compose an e-learning environment can be implemented as Web services. A Web service is a stand-alone software component that has a unique URI (Uniform Resource Identifier), that operates on the Web, and offers possibilities to access services in distributed environments [5]. The basic premise is that Web services have a provider and users, and can be combined to build new services with more comprehensive functionality. Various works have been presented in the literature utilizing Web services in the e-learning environment [5], [6], [7], [8], [9], [10].

In this paper, we propose the design of an e-learning system, oriented to a specific problem, for assisting medical students in the interpretation of mammograms and diagnosis of breast cancer, that we call INDIAM (INterpretation and DIagnosis of Mammograms). The activities that compose the proposed e-learning system are modeled as Web services, which can be reused by other medical e-learning

* To whom all correspondence should be addressed

D. Guliato, R. S. Bôaventura and M. S. Smedo are with the Faculty of Computation, Federal University of Uberlândia, Uberlândia, MG, Brazil. guliato@ufu.br, boaventura@lcc.ufu.br, mariangela@comp.ufu.br

T. A. A. Macedo is with the Clinical Hospital of Uberlândia, Federal University of Uberlândia, Uberlândia, MG, Brazil. tamacedo@hotmail.com

R. M. Rangayyan is with the Department of Electrical and Computer Engineering, Schulich School of Engineering, and Department of Radiology, University of Calgary, Calgary, AB, Canada T2N 1N4. ranga@ucalgary.ca

This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico of Brazil and a Catalyst grant from Research Services, University of Calgary. We thank Vinícius de Deus, Federal University of Uberlândia, for assistance with the digital image processing tools.

environments. The information required for modeling the application knowledge (interpretation of mammograms and breast cancer diagnosis) is represented by an ontology that is associated with an illustrated tutorial and a mammographic database. The resources available in INDIAM guide the medical student, step-by-step, in analyzing a mammogram correctly; simulate diagnosis of breast cancer with appropriate feedback; answer questions of the student according to the knowledge modeled by the ontology; and navigate through the illustrated hypermedia tutorial. The proposed e-learning system makes available a graphical interface, via the Web, with digital image processing tools that permit the student to draw the contours of lesions, the contour of the breast, and boundaries of clusters of microcalcifications during activities to simulate interpretation or diagnosis.

This paper is organized as follows: Section II presents the general architecture of the INDIAM e-learning system. Section III provides a description of each component of the application knowledge base. Section IV contains the details of the Web services made available by INDIAM and their functionalities. Section V gives a succinct description of AMDI, the Indexed Atlas of Digital Mammograms, including a mammogram registration module, a research system module, and the e-learning system INDIAM. Section VI presents concluding remarks.

II. AN OVERVIEW OF THE PROPOSED E-LEARNING SYSTEM: INDIAM

INDIAM was designed to assist a medical student or resident in the interpretation of mammograms and diagnosis of breast cancer, and is being developed using Web services. Figure 1 illustrates a general overview of the system. INDIAM makes available a tutorial that uses education techniques to guide the users (doctors, students, or researchers) through concepts related to the diagnosis of breast cancer. It also makes available a module to simulate the analysis and diagnosis of breast cancer using cases retrieved from a mammographic database, and another module for training the student in the interpretation of mammograms. The e-learning system is being developed utilizing an ontology that provides controlled and consistent vocabularies to describe concepts and relationships, thereby enabling knowledge sharing [11]. The system makes available a user-friendly graphical interface that is configured according to the service being provided.

III. THE APPLICATION KNOWLEDGE BASE

The knowledge of the application is composed by a mammographic database, an ontology that incorporates semantics to the Web services, and an illustrated tutorial on breast cancer.

A. The mammographic database

The mammographic database was modeled using the PostgreSQL with Image-handling Extension (PostgreSQL-IE), an eXtended Relational DataBase Management System (XRDBMS), that includes facilities to organize visual and

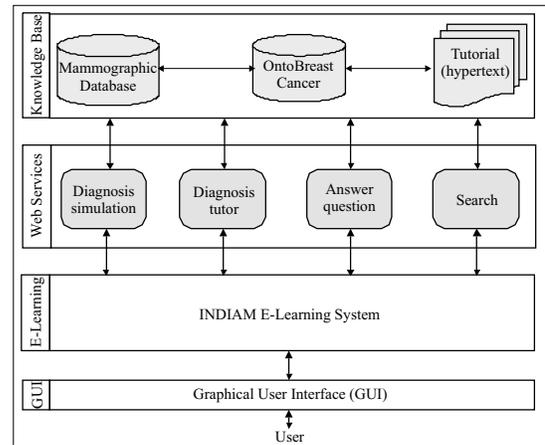


Fig. 1. An overview of the architecture of INDIAM.

conventional data, and answer queries based on image content [12], [13]. It is projected to include cases with all of the available mammographic views, radiological findings, diagnosis proven by biopsy, the patient's clinical history, and information regarding the life style of the patient. Each exam of each case includes four views (two views of each breast: cranio-caudal or CC, and medio-lateral oblique or MLO). To address the teaching aspects, the database links each mammogram with the contour of the breast, the boundary of the pectoral muscle (MLO views only), the contours of masses (if present), the regions of clusters of calcifications and the number of calcifications (if present), and the locations and details of any other features of interest. The contours of masses and regions of clusters of calcifications may be drawn interactively by the expert radiologist when including a new case in the database. The mammographic database also supports the inclusion of several mammographic exams of the same patient performed at different instants of time. This information can be used for temporal analysis of the breast, due to natural modifications that occur during the life of the woman, or to analyze interval cancer.

B. The ontology

An ontology is a formal and explicit specification of an abstraction, i.e., a simplified vision of the world or domain that we wish to represent for a specific purpose. An ontology shares a domain of knowledge, defining a common vocabulary [14]. An ontology can be used by software agents to establish a common agreement regarding the concepts and relationships of a specific domain of knowledge [15]. An ontology defines the primary concepts in an application area, as well as the relationships among those concepts. It is independent of particular algorithms, such as database schema, in that it is application-independent. Unlike a database, an ontology can express complex relationships among the concepts represented [16].

An ontology can be used to formalize the representation of a knowledge domain; to describe a common and defined

vocabulary for data of a given application; to provide unified access to information through ontology-based querying for both human and computational processing; to improve management and integration of application data; and to facilitate data mining.

An ontology can be represented as a graph, where the nodes are connected by directed arcs (arrows). The nodes are concepts (implemented as classes) in the ontology, and the directed arcs are the relationships (implemented as properties) between the concepts. Considering the domain of “Analysis of Mammograms and Diagnosis of Breast Cancer”, the concepts of the proposed ontology include all the relevant information to be analyzed in order to reach a diagnosis regarding breast cancer.

In our system, the specialized knowledge to be learned by the student is represented by an ontology, called OntoBreastCancer, that was modeled by an expert in breast cancer diagnosis. In its current version, OntoBreastCancer possesses 16 classes, 461 instances, and 98 properties, and is implemented in OWL (Ontology Web Language) [17] using the Protégé Framework [18].

The proposed ontology models the knowledge necessary for relating radiological findings (local or global) with the diagnosis of breast cancer according to BI-RADS classification [19]. Figure 2 illustrates a section of Protégé showing the 16 classes of OntoBreastCancer and the properties related to the class *type-of-lesion-margins* and its instances. OntoBreastCancer models assertions such as “lesions with circumscribed margins are associated with BI-RADS categories 2, 3, 4a, and 0”, or “lesions with circumscribed margins are associated with the radiological findings cyst, fibroadenoma and calcified-fibroadenoma”.

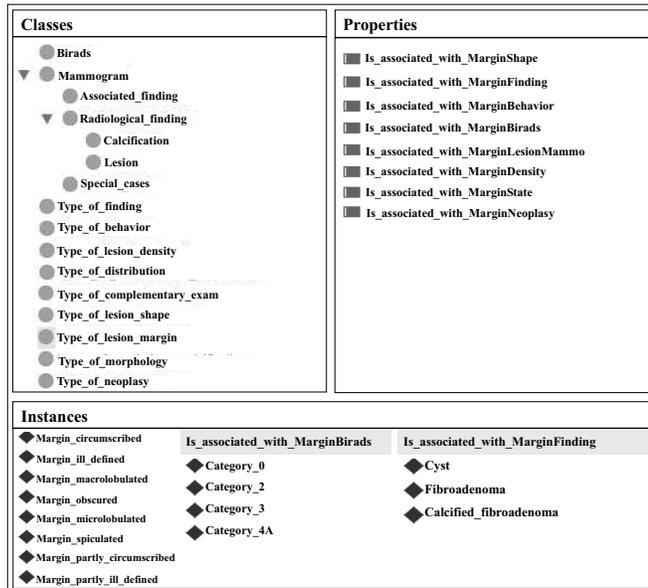


Fig. 2. A section of Protégé with the classes of OntoBreastCancer, the properties of the class *type-of-lesion-margins*, and its instances.

OntoBreastCancer integrates a tutorial on breast radiology

and cases stored in the mammographic database, and incorporates semantics in the searching procedures. Through OntoBreastCancer, the INDIAM e-learning system can answer questions such as:

- What types of radiological findings are associated with malignant lesions?
- Which categories of BI-RADS are associated with radiological findings and special cases with characteristics of malignancy?
- What procedure must be followed in the case of a malignant diagnosis?
- What are the radiological findings associated with BI-RADS category 5?
- How can a lesion associated with an oval form be classified?

Because the information in OntoBreastCancer is exhaustive, a vast range of questions may be asked: the “Answer-question” Web service is designed to provide a flexible and user-friendly interface for this purpose, as described in Section IV-A.

C. The tutorial

The INDIAM e-learning system provides a tutorial on breast radiology using hypertext and hypermedia technologies, which guides the users (doctors, students, or researchers) through the concepts related to the diagnosis of breast cancer. The tutorial is associated with OntoBreastCancer, and includes informative text about specific subjects, provides access to a bibliography, and makes available tools for retrieving similar cases from the mammographic database. The breast cancer tutorial is accessed by the Web services described below. The tutorial uses education techniques for providing the user with an easy strategy for navigating through the content of the tutorial.

IV. THE WEB SERVICES

Web services offer possibilities to access services in a distributed environment via the Web. When associated with an ontology, Web services can incorporate semantics to the related activities. The proposed e-learning system is composed by four principal Web services: the *answer-question Web service* that answers questions from the user based on OntoBreastCancer; the *search Web service* that makes available a tutorial on the diagnosis of breast cancer associated with the ontology; the *diagnosis-tutor Web service* for guiding the beginner student through the steps of analysis of a mammogram; and the *diagnosis-simulation Web service* that makes available a graphical user-interface to simulate the diagnosis of a given mammogram retrieved from the mammographic database.

A. Answer-question Web service

This Web service allows the user to pose questions to INDIAM using an oriented strategy to frame the questions. Although OntoBreastCancer is able to answer any type of question related to breast radiology, the answer-question Web service limits the structure of the questions into two classes:

- Class 1 - In this class, the question relates a class of the ontology with an instance. The question is broken into four grammatical classes: a question type; a noun that represents a class of the ontology; a verb that represents a property; and an adjective that represents an instance of the ontology. An example is “What types of radiological findings are associated with malignant lesions?”. Figure 3 shows the user interface to ask this question.
- Class 2 - In this class, the question relates two classes using an instance. The question is broken into five grammatical classes: a question type; a noun that represents a class; a verb that represents a property; another noun; and an adjective that represents an instance of the ontology. An example is ”Which BI-RADS categories are associated with lesions with spiculated margins?”.

Questions can be connected by the logical operators AND or OR. The question types are limited to What/Which/How, and allow the student to obtain information from the e-learning system by crossing classes of OntoBreastCancer with the associated properties. The inference engine that is being used by the Answer-question Web service is RAP (RDF API for PHP) [20] because of the resources of working with the PHP language.

The student is not required to know the name of the classes, properties, or instances modeled by the ontology. OntoBreastCancer makes available a list of synonyms that makes the task of framing questions easier to the user. A more flexible way to frame a question to the e-learning system is by using a natural language; however, this approach is beyond the scope of the present work.

Fig. 3. Answer-question Web service interface.

B. Search Web service

The search Web service makes available a search engine for helping the student in breast cancer diagnosis. The search engine takes into account the semantic represented in OntoBreastCancer, making the search more precise (reduction of false positives). The search Web service also accesses the tutorial and the mammographic database. Because of this, the search engine can return to the user all the cases stored in the database with the features that the user is learning, along with some theoretical information.

C. Diagnosis-tutor Web service

The diagnosis-tutor guides the user through the mammogram-interpretation fluxogram (Figure 4), designed by an expert radiologist. In a general way, the diagnosis-tutor Web service has the following objectives: i) provide an ergonomic environment for the presentation of the content; ii) guide the students through the content and assist them to reach the correct diagnosis; iii) provide a flexible and friendly environment for answering questions; and iv) present questions to guide the student during the learning process. The diagnosis-tutor Web service communicates with the ontology, mammographic database, and the tutorial to send hints and appropriate feedback to the user. This service uses the answer-question Web service described above to provide the answer-question user interaction when required by the user.

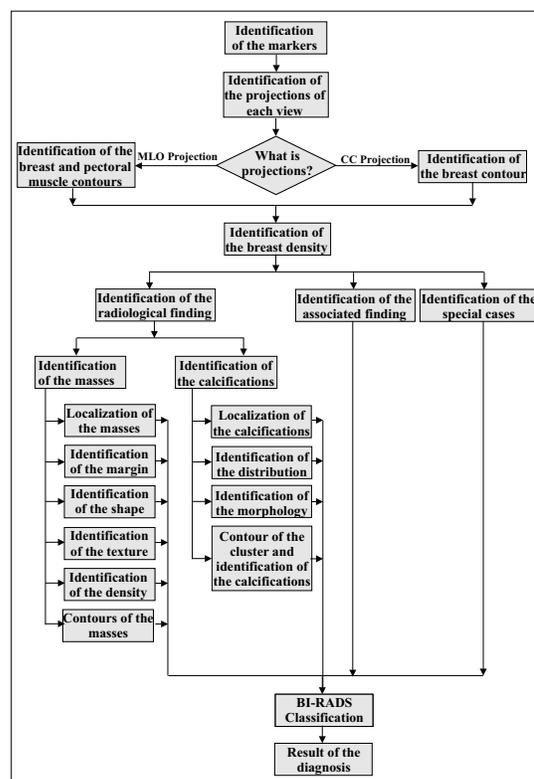


Fig. 4. The steps in analyzing a mammogram represented by a fluxogram designed by an expert radiologist.

The diagnosis-tutor presents to the student a mammographic case (displaying the associated four views) to be analyzed, and limits the actions of the user according to the steps proposed in the mammogram-interpretation fluxogram. First, the student has to identify the markers in the mammographic views. If he or she makes a mistake, the diagnosis-tutor returns a feedback with the correct answer and an illustrated tutorial with the appropriate content to be learnt. The process continues until the student gives the correct diagnosis. Figure 5 shows a stage of the analysis of a mammogram where the student has to identify the breast density of a given view. At this point, the student has already identified the view projection, the pectoral muscle, and the breast contour using the digital image processing tools available in INDIAM. For helping in the analysis, the diagnosis-tutor lists the possible answer options with appropriate links to the tutorial. If wanted, the student can navigate through the tutorial for more information about the subject being learnt. Options for framing questions using answer-question or search Web services are available.

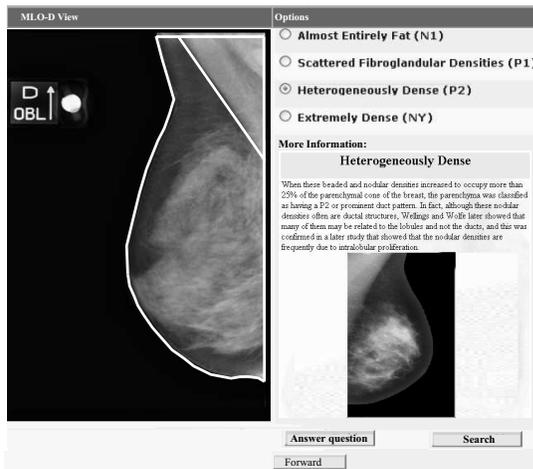


Fig. 5. Step of the diagnosis tutor for analyzing the breast density in a given view.

D. Diagnosis-simulation Web service

The diagnosis-simulation Web service simulates an environment for the interpretation of mammograms and breast cancer diagnosis. The service displays the four views of a mammographic case previously stored in the mammographic database. The student analyzes each pair of views and asks for a specific report form to submit his or her diagnosis. The system makes available four report forms: for identifying global radiological findings, radiological findings associated with lesions, radiological finding associated with microcalcifications, and the final diagnosis report. After filling out the forms, the reports are submitted and checked by the system; the system provides appropriate feedback in the case of errors. In addition to the analysis of the report, this service makes available digital image processing

tools with which the student can draw the contour of the breast, delineate and remove the pectoral muscle (for MLO views), identify clusters of microcalcifications and indicate the position of each one, and draw the contours of lesions. The management system of INDIAM evaluates the degree of similarity between the answer of the user and the data stored in the mammographic database, helping the user if necessary. Figure 6 illustrates the graphical interface for the diagnosis simulation, with the lesion report form.

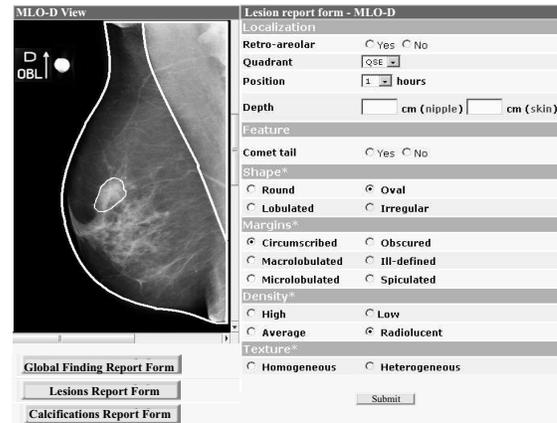


Fig. 6. Graphical interface of the diagnosis simulation service.

V. INTEGRATION OF INDIAM INTO AMDI - THE INDEXED ATLAS OF DIGITAL MAMMOGRAMS

AMDI [21], [22] is a system that integrates modules to permit the addition of new cases into the mammographic database by authorized radiologists, and to assist research and education activities in breast cancer through a flexible and easy-to-use interface via the Web.

The research system called SISPRIM — Sistema de Pesquisa para Recuperação de Imagens Mamográficas, that is, Research System for Retrieval of Mammographic Images — [12], integrated with AMDI, allows physicians and oncologists to study possible statistical correlation between the incidence of breast cancer and the life style of the patient. For this purpose, some of the important items of data supported by AMDI relate to information regarding the patient, including basic data about the clinical history of the patient as well as life style, such as food habits, exercise, diet, and the use of antidepressive medication, tobacco, or alcohol. The SISPRIM research system also allows content-based image retrieval to assist the radiologist in breast cancer diagnosis. The physician can realize queries such as "Return all images that are similar to the given reference image, including the point that the patient uses an antidepressant" or "Return all images that are similar to the given reference image and have a malignant diagnosis".

AMDI provides a tool that enables the user to download cases from the database, so as to make the information available to authorized medical and research communities interested in breast cancer diagnosis.

INDIAM is one of the modules that is integrates into AMDI. Figure 7 shows a schematic representation of the flow of information in AMDI.

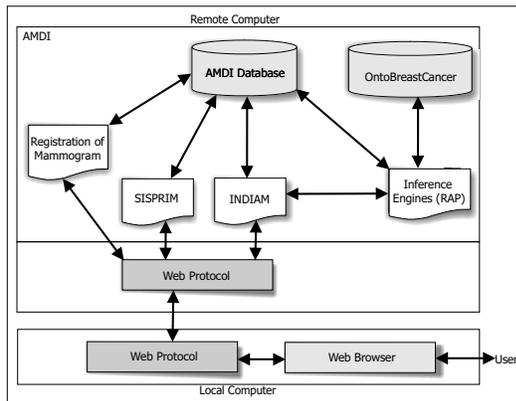


Fig. 7. An overall view of AMDI.

VI. DISCUSSION AND CONCLUSION

This work presented INDIAM, an e-learning system to assist a medical student in the interpretation of mammograms and diagnosis of breast cancer. The application knowledge base of INDIAM includes a mammographic database with complete clinical information and details regarding the life style of the patient; global and local radiological findings; a tutorial based on hypertext; and an ontology called OntoBreastCancer that represents the knowledge of the radiologist for analyzing a mammographic case.

The activities that compose INDIAM are being developed using Web services that take advantage of sharing the Web services with other e-learning environments, facilitating the upgrading of the proposed e-learning system, and reuse of the Web services developed by others providers. INDIAM is composed of four Web services: Diagnosis-tutor is a Web service that guides the student through the stages necessary to reach a diagnosis according to the steps designed by an expert radiologist. Diagnosis-simulation is a Web service that evaluates the diagnosis done by a student and returns appropriate feedback in the case of errors. Both services use digital image processing tools allows the user to draw contours of breast, lesions, and clusters of microcalcifications. The remaining two components are Search and the Answer-question Web services; the latter provides facilities for constructing questions. Further work is being conducted to develop an interface using a natural language. INDIAM is being integrated with AMDI — an Indexed Atlas of Digital Mammograms — an atlas available via the Web that also includes a research system to retrieve images based on content, and a tool for inserting new mammographic cases by registered radiologists.

REFERENCES

[1] Victor Alves, José Neves, Luís Nelas, and Filipe Marreiros. Web-based medical teaching using a multi-agent system. *Proceedings of*

AI2005, the Twenty-fifth SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, Applications and Innovations in Intelligent Systems XIII, Cambridge, UK, 25:14 pages, 2005.

[2] Erin Shaw, W. Lewis Johnson, and Rajaram Ganeshan. Pedagogical agents on the web. *Proceedings of the third Annual Conference on Autonomous Agents, Seattle, WA*, pages 283–290, 1999.

[3] W. Lewis Johnson, Erin Shaw, Andrew Marshall, and Catherine Labore. Evolution of user interaction: The case of Agent Adele. *Proceedings of the 8th International Conference on Intelligent User Interfaces, Miami, Florida*, pages 93–100, 2003.

[4] Fong-Ming Shyu, Ya-Fen Liang, Wei-Ting A. Hsu, Jer-Junn Luh, and Heng-Shuen Chen. A problem-based e-learning prototype system for clinical medical education. *MedInfo, Amsterdam, The Netherlands: IOS Press*, pages 983–987, 2004.

[5] Gottfried Vossen and Peter Westerkamp. E-learning as a web service. *Proceedings of the 7th International Database Engineering and Applications Symposium, Asunción, Paraguay*, pages 242–249, 2003.

[6] Gottfried Vossen and Peter Westerkamp. Maintenance and exchange of learning objects in a web services based e-learning system. *Electronic Journal of e-Learning*, 2(2):293–304, 2004.

[7] Emanuela Moreale and Maria Vargas-Vera. Semantic services in e-learning: an argumentation case study. *International Forum of Educational Technology & Society*, 7:112–128, 2004.

[8] Ian Dickinson and Michael Wooldridge. Towards practical reasoning agents for the semantic web. *International Conference on Autonomous Agents, ACM Press, New York, NY*, pages 827–834, 2003.

[9] Maria Vargas-Vera, Enrico Motta, and John Domingue. AQUA: An ontology-driven question answering system. *Proceedings AAAI Spring Symposium Stanford, CA*, pages 53–57, 2003.

[10] Daniel Moldt and Jan Ortmann. A conceptual and practical framework for web-based processes in multi-agent systems. *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS, New York, NY*, pages 1464–1465, 2004.

[11] Tom R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. *Int. J. Human-Comput. Stud.*, 43:907–928, 1993.

[12] Denise Guliato, Ernani V. Melo, Ricardo S. Bôaventura, and Rangaraj M. Rangayyan. AMDI - indexed atlas of digital mammograms that integrates case studies, e-learning, and research systems via the web. In: *Jasjit S. Suri; Rangaraj M. Rangayyan. (Eds.). Recent Advances in Breast Imaging, Mammography, and Computer-aided Diagnosis of Breast Cancer. Bellingham, WA, SPIE Press*, pages 535–562, 2006.

[13] Michael Stonebraker, Lawrence A. Rowe, and Michael Hirohama. The implementation of Postgres. *Knowledge and Data Engineering*, 2(1):125–142, 1990.

[14] Tom R. Gruber, Greg Olsen, and Jay T. Runkel. The configuration design ontologies and the VT elevator domain theory. *Int. J. Human-Comput. Stud.*, 44:569–598, 1996.

[15] Markus Schumacher. Security engineering with patterns origins, theoretical model, and new applications. *Lecture Notes in Computer Science - LNCS 2754, Springer, New York, NY*, pages 29–44, 2003.

[16] Marck A. Nusen. Medical informatics: Searching for underlying components. *Methods Inf. Med.*, 41:12–19, 2002.

[17] OWL. Web ontology language overview - w3c. [Online]. Available: <http://www.w3.org/TR/owl-features/> Accessed in: Jun. 2006.

[18] Protégé. Documentation. [Online]. Available: <http://protege.stanford.edu/doc/users.html>, Accessed in: Jun. 2006.

[19] American College of Radiology. Breast imaging reporting and data system BI-RADS. 4th Edition, American College of Radiology, Reston, VA, 2004.

[20] RAP RDF API for PHP. [Online]. Available: <http://www.wiwiss.fu-berlin.de/suhl/bizer/rdffapi/tests.html> Accessed in: Jun. 2006.

[21] Denise Guliato, Márcio Caetano, Rangaraj M. Rangayyan, Paulo M. A. Marques, Flávio F. R. Janones, Vinícius de Deus, and Susan C. Lima. AMDI: An indexed atlas of digital mammograms available via the web. In: *7th International Workshop on Digital Mammography, Durham, NC*, pages 69–74, 2004.

[22] Denise Guliato, Ernani V. Melo, Ricardo S. Bôaventura, Flávio R. Janones, Vinícius de Deus, and Rangaraj M. Rangayyan. AMDI: An atlas to integrate case studies, e-learning, and research systems via the web. *Proceeding of the IASTED International Conference on Telehealth. Banff, AB, Canada*, pages 69–74, 2005.