A semantically-aided architecture for a web-based monitoring system for carotid atherosclerosis

Vassileios D. Kolias, Giorgos Stamou, Spyretta Golemati, *Member*, *IEEE*, Giannis Stoitsis, *Member,IEEE*, Christos D. Gkekas, Christos D. Liapis, and Konstantina S. Nikita, *Senior Member, IEEE*

***Abstract*— Carotid atherosclerosis is a multifactorial disease and its clinical diagnosis depends on the evaluation of heterogeneous clinical data, such as imaging exams, biochemical tests and the patient's clinical history. The lack of interoperability between Health Information Systems (HIS) does not allow the physicians to acquire all the necessary data for the diagnostic process. In this paper, a semantically-aided architecture is proposed for a web-based monitoring system for carotid atherosclerosis that is able to gather and unify heterogeneous data with the use of an ontology and to create a common interface for data access enhancing the interoperability of HIS. The architecture is based on an application ontology of carotid atherosclerosis that is used to (a) integrate heterogeneous data sources on the basis of semantic representation and ontological reasoning and (b) access the critical information using SPARQL query rewriting and ontology-based data access services. The architecture was tested over a carotid atherosclerosis dataset consisting of the imaging exams and the clinical profile of 233 patients, using a set of complex queries, constructed by the physicians. The proposed architecture was evaluated with respect to the complexity of the queries that the physicians could make and the retrieval speed. The proposed architecture gave promising results in terms of interoperability, data integration of heterogeneous sources with an ontological way and expanded capabilities of query and retrieval in HIS.**

I. INTRODUCTION

Carotid atherosclerosis is a multifactorial disease and one of the main causes of strokes. To assess a patient's state, a physician needs to consider heterogeneous medical data, such as imaging, biochemical and the clinical history data. Due to the limited capabilities of the current Health Information Systems (HIS), physicians can retrieve similar cases by using only the name or the id of the patient. Search with the semantic content of the patient data, such as anatomy, image characteristics (arterial stenosis, plaque type) and clinical data (glucose, cholesterol levels) is currently not possible. In addition, retrieval of patient data from different sources, such as different hospitals, is limited, due to restricted interoperability among HIS. As a result, the physician has incomplete information about the patient profile and acquisition of the necessary data is a difficult task which slows down the diagnostic process.

A solution to the issues mentioned above is the use of Semantic Web technologies, which can provide a mechanism for creating an annotation infrastructure for data integration by means of ontologies. Such an infrastructure can be used to formalize the data and to create a unified interface, which is accessible from various HIS and thus enables queries and information retrieval from heterogeneous data sources. This will allow physicians to gather all data needed for the diagnostic procedure.

Various studies have underlined the use of heterogeneous data in the diagnosis of carotid atherosclerosis. Several Computer-Aided Diagnosis (CAD) and Content Based Image Retrieval (CBIR) systems have been proposed, such as Analysis [1] and CBMIR [2]. These systems use several image features (texture, motion) and other medical data to retrieve similar cases, but the data repository that is created is not available for querying from other systems (limited interoperability). The use of semantic web technologies in medical imaging has been proposed to provide expanded query capabilities [3], [4], and to create a unified annotation framework with the use of biomedical reference ontologies such as Foundational Model of Anatomy (FMA), Radiological Lexicon (RadLex) and Systemized Nomenclature of Medicine (SNOMED). The importance of semantic technologies for formalizing the data in carotid atherosclerosis has also been highlighted [5], [6]. Although the above works use semantic technologies for formalizing the data, they do not provide a data infrastructure layer to utilize data from heterogeneous sources.

This paper aims to propose a semantically-aided web based-system that is able to create a semantic-based unified interface for data acquisition from heterogeneous data sources and provides expanded query and retrieval capabilities for the case of carotid atherosclerosis. This approach can lead to a semantic data representation layer for this disease and can also enhance the interoperability among HIS in general.

II. SYSTEM ARCHITECTURE

The system's architecture is shown in Fig. 1. It consists of the web application, which is installed at the HIS, the Data Layer, which is generated from the HIS databases, the Ontological Layer, which is the ontology for carotid

V. D. Kolias (e-mail: vaskolias@biosim.ntua.gr), J. Stoitsis and K. S. Nikita are with the Biomedical Simulations and Imaging (BIOSIM) Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, Iroon Polytechniou 9, Athens, 15780 Greece

G. Stamou (e-mail: gstam@cs.ntua.gr) is with the Intelligent Content and Interaction Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, Iroon Polytechniou 9, Athens, 15780 Greece.

S.. Golemati (e-mail: sgolemati@med.uoa.gr), Ch. Gkekas and Ch. D. Liapis are with the Medical School, National Kapodistrian University of Athens, Athens, 10765 Greece

atherosclerosis, and the Ontology-Based Data Access Layer, which provides the connection between the ontology and the data that are stored in the databases. The Data Layer, the Ontological Layer and the Ontology-Based Data Access Layer are described in Section III.

The web application is used by the physician to enter data of a new patient (imaging exams, biochemical tests, clinical profile etc) and to query and retrieve information. The web application using the Ontology-Based Data Access Layer can provide access to the data that are stored in several databases.

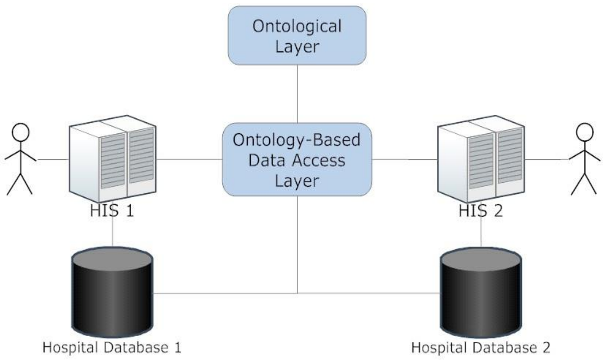

Figure 1. Proposed system architecture

The Ontology-Based Data Access Layer is created by defining a set of mappings between the database schema of the Data Layer and the defined ontologies and then using the Rapid rewriting system [7] to rewrite the user query in terms of the terminological description of the domain. With the use of the Ontology-Based Data Access Layer, the physician can have access to all data through a unified interface, regardless of the database, since the user queries are given in terms of the ontologies.

III. DATA MODELING

The data modeling scheme consists of three layers, as illustrated in Fig. 2: the Data Layer, the Ontology-Based Data Access Layer and the Ontological Layer.

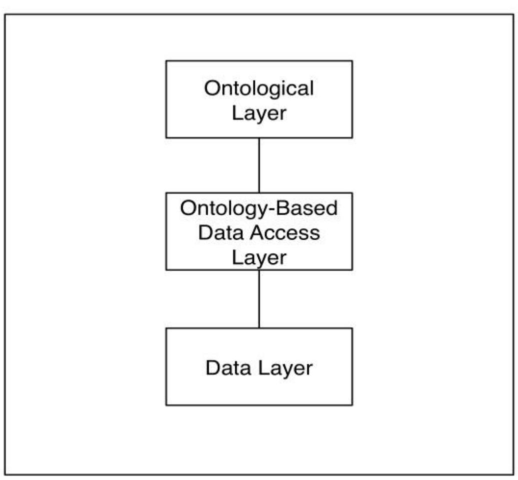

Figure 2. Data modeling scheme

A. Data Layer

The Data Layer is derived from the databases that store the medical data necessary for the clinical monitoring of carotid atherosclerosis. These include imaging exams, reports from imaging exams (stenosis degree, plaque type etc), and the clinical profile of the patient. The data scheme of each HIS database that stores the medical data can be different due to the mapping that is created at the Ontology-Based Data Access Layer.

B. Ontological Layer

The Ontological Layer consists of the carotid ontology. A bottom up methodology was used to develop the carotid ontology. This methodology was selected because the data stored in the databases, combined with the user requirements indicated by the physicians, form the basis of the ontology development. The ontology was developed with the Protégé tool and the OWL 2 QL profile [8], that is known to be tractable, thus the query rewriting can be performed in worst-case polynomial time. The ontology was created by the National Technical University of Athens team in collaboration with physicians from the vascular surgery unit of the Medical School of the University of Athens. The anatomic terms of the carotid ontology were reused from the SNOMED-CT and the FMA ontologies. The main classes of the ontology, as shown in Fig. 3, are the following.

- Patient: Class 'Patient' represents the atherosclerotic patients. The subclasses of the class classify each patient based on their clinical profile, the degree of stenosis, and their symptoms. An individual can have more than one patient types except for the types that are disjoint, such as 'AsymptomaticPatient' (patient without symptoms) and 'SymptomaticPatient' (patient with symptoms).
- Artery: Class 'Artery' represents the artery of a patient. 'Patient' is connected with 'Artery' through the 'has Artery' object relationship. 'Artery' subclasses are classified based on the anatomic position (Left Inner Common Artery, Right Inner Common Artery etc), the type of plaque and the stenosis degree.
- Symptoms: Class 'Symptoms' represents the symptoms that determine if a patient is classified as Asymptomatic or Symptomatic. The subclasses of this class are 'Stroke', 'Trans Ischemic Attack' and 'Amaurosis'. The classification of a patient as Asymptomatic or Symptomatic is crucial for the physician.
- Clinical Profile: Class 'Clinical Profile' represents crucial aspects of the patient clinical profile, such as glucose levels, cholesterol levels, BMI, blood pressure levels.
- Stenosis: Class 'Stenosis' represents the degree of the artery stenosis. Its three subclasses are 'HighStenosis', which represents the arteries with a stenosis above 70%, 'MediumStenosis', which represents the arteries that have 50% to 70% stenosis and 'LowStenosis', which represents arteries with a stenosis lower than 50%.

- Image: Class 'Image' represents the image exams of a patient and its two subclasses are 'HighRiskImage' and 'LowRiskImage'. This class is created to encapsulate image findings from possible CAD integration in the system.
- Plaque: Class 'Plaque' represents the type of a plaque that the patient has (hard, soft, vulnerable).
- PlaqueCharacteristics: This class represents the characteristics of the plaque, including the shape of the outline.

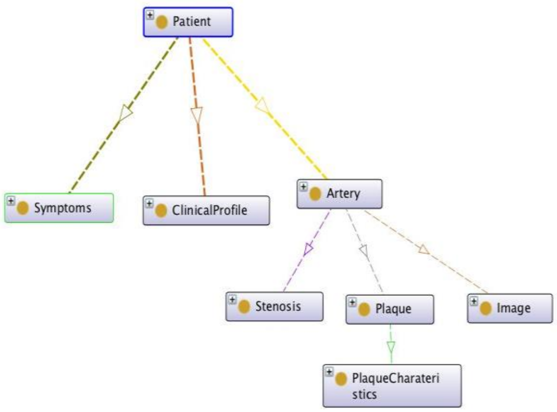

Figure 3. Main ontology classes

C. *Ontology-Based Data Access Layer*

The Ontology-Based Data Access Layer connects the Data Layer with the Ontological Layer, supporting the representation of user queries in the ontology vocabulary. The connection is performed on the basis of a set of semantic mappings connecting the database schema terms to the ontology concepts and roles. With the aid of semantic mappings, classes and roles of the ontology provide a formal, expressive, user-oriented description of the data, thus forming a rich vocabulary for semantically accessing it. Moreover, in this way, all data sources can be accessed with the same queries, following a database schema agnostic framework. As a result, a unified view is created for heterogeneous data sources.

An ontology has two types of statements, TBox statements and ABox statements [9]. TBox is the terminological component of the ontology that contains the conceptualization with a set of facts or individuals. ABox is the set of facts or individuals. With the aid of semantic mappings, the data sources can be viewed as ABoxes of the relevant ontologies (that can be either virtual or materialized, i.e. stored in a triple store). Then, with the use of SPARQL queries, which is the querying language used in the Semantic Web, the needed information from all the data sources is retrieved from this ABox. SPARQL is a flexible querying language, that is able to represent complex conjunctive queries that are posed in the triple form: <"Subject"> <"Predicate"> <"Object">, whereas all the terms are taken from the ontology vocabulary. In order for the system to be efficient, the necessary ontological reasoning, that advances the level of semantic access, is performed using query rewriting services (offered by Rapid [7]). Specifically, the query is transformed to a set of SPARQL queries that encode all the axiomatic ontological descriptions. This procedure is called 'query expansion'. Then, the ontology is no more useful and, thus, it can be safely rejected and existing query answering services cab be used. The above query answering strategy, is proved to be complete for the OWL 2 QL fragment of Web Ontology Language. The advantage of the proposed methodology is that the queries of the web application are not affected by changes at the Data Layer because the queries are made at the ABox. As a result, the interoperability among HIS is enhanced.

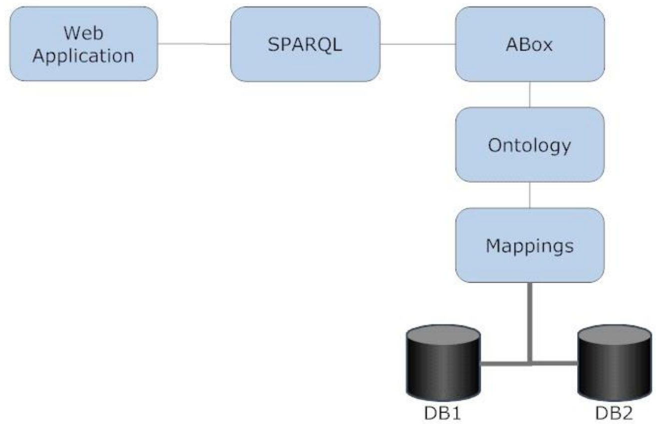

Figure 4. Ontology-Based Data Access Layer

IV. EVALUATION

The proposed architecture was tested in a dataset of 233 carotid atherosclerotic patients from 'Attikon' University Hospital, Athens, Greece. The database was developed with MySQL and included records of imaging exams and clinical profile of each patient.

A. *User Requirements*

The main user requirements identified by the physicians, are

- To make queries, based on the imaging findings, such as stenosis degree, anatomic position, type of plaque
- To make queries based on the clinical profile of a patient (glucose levels, cholesterol levels, hypertension, symptomatic patient, asymptomatic patient)
- To make combined queries both with imaging findings and clinical profile

B. *Example Queries*

In this section, some examples are shown along with a comparison of the proposed architecture with the traditional HIS systems

1) *Query based on image findings*

The test query is "Bring all the patients with high stenosis 'Artery'". The query execution time was 51 ms. In the proposed architecture this query is a single task. The query is represented by a query in the form

<Patient><HasStenosis><HighStenosis>. The system due to the annotation process has the ability to identify the highStenosis Patient. In a traditional HIS which supports only keywords, the system cannot identify the HighStenosis Patients. The system has to be guided by the doctor by entering the stenosis degree in order to create the query. Another limitation is that in many cases, such information cannot be retrieved because the information is kept only in the DICOM header and not in the HIS database.

2) Query based on clinical profile

The test query is “Bring all patients that are asymptomatic and have diabetes”. The query execution time was 91 ms. The query is represented by a query in the form <AsympomaticPatient> <HasClinical> <HighSugarLevel> or <AsymptomaticPatient> AND <DiabeticPatient>. The annotation process and the ontology define which patients have the states that are mentioned in the query. In a traditional HIS system this query is very difficult because this information is usually stored in free text and in the reports and, as a result, the accuracy of the query depends on the user’s input.

3) Query combining image findings and clinical profile

The most complex query was “Bring all asymptomatic patients with diabetes and medium stenosis 'Artery'”. The query execution time was 153 ms. With the proposed architecture this can be represented by the query <AsympomaticPatient> AND <MediumStenosisPatient> AND <DiabeticPatient>. With the semantic search, the proposed architecture just retrieves the patients that are annotated with the three classes. In a traditional HIS system, the retrieval of such information requires the doctor to define the stenosis degree (as in query (1)) and also the system to make keyword search (as in query (2)) with the drawbacks that are mentioned above.

V. CONCLUSION

The proposed semantically-aided web-based system enables (a) the development of an application that allows semantic search for carotid atherosclerosis with the use of an appropriately designed ontology, and (b) the enhancement of interoperability among HIS by creating a unified view of heterogeneous data sources independent of the design of the data source. Current HIS systems cannot support semantic search of the data and the changes at the data sources require several changes in the application level, thus limiting interoperability and making the integration of several data sources a difficult task.

The proposed web-based system can be used to create a unified view of several data sources enhancing interoperability of systems and giving access to a wealth of information for physicians both for diagnostic tasks and research purposes. This can lead to semantic repositories not only for carotid atherosclerosis but also for other diseases and medical cases.

Future research work refers to two main directions. Firstly, the integration of clinical guidelines from American and European Vascular Societies,will allow physicians to better monitor carotid atherosclerotic patients. This can be achieved by increasing the expressiveness of the ontology to an OWL profile that allows the representation of rules such as OWL-RL or OWL-EL [8]. The second direction refers to the integration of medical image analysis tools in the architecture. This will be very useful for the diagnostic procedure and also will create a semantic repository for carotid atherosclerosis that can connect low level image characteristics, such as texture, with high level entities of the image. Such a repository can be used to create a semantic CAD system for the disease and also to provide a solution for the so called semantic gap problem [10].

ACKNOWLEDGMENT

This work was conducted using the Protégé resource, which is supported by grant GM10331601 from the National Institute of General Medical Sciences of the United States National Institutes of Health.

REFERENCES

[1] J. Stoitsis, S. Golemati, and K. S. Nikita, "A modular software system to assist interpretation of medical images&application to vascular ultrasound images," *IEEE Trans. Instrum. Meas,* vol. 55, pp. 1944-1952, 2006.

[2] G. Ioakim, E. Kyriacou, A. A. Sofokleous, C. Chistodoulou, and C. S. Pattichis, "An MPEG-7 image retrieval system of atherosclerotic carotid plaque images," in *Bioinformatics & Bioengineering (BIBE), 2012 IEEE 12th International Conference on*, 2012, pp. 512-516.

[3] G. Kyriazos, I. Gerostathopoulos, V. Kolias, J. Stoitsis, and K. Nikita, "A semantically-aided approach for online annotation and retrieval of medical images," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, 2011, pp. 2372-2375.

[4] M. Möller, S. Regel, and M. Sintek, "Radsem: Semantic annotation and retrieval for medical images," in *The Semantic Web: Research and Applications*, ed: Springer, 2009, pp. 21-35.

[5] A. Gastounioti, V. Kolias, S. Golemati, N. N. Tsiaparas, A. Matsakou, J. S. Stoitsis*, et al.*, "CAROTID – A web-based platform for optimal personalized management of atherosclerotic patients," *Comput Meth Prog Bio* vol. 114, pp. 183-193, 2014.

[6] C. Doulaverakis, M. Papadogiorgaki, V. Mezaris, A. Billis, E. Parissi, I. Kompatsiaris*, et al.*, "IVUS image processing and semantic analysis for Cardiovascular Diseases risk prediction," *Int J Biomed Eng Technol,* vol. 3, pp. 349-374, 2010.

[7] A. Chortaras, D. Trivela, and G. Stamou, "Optimized query rewriting for OWL 2 QL," in *Automated Deduction–CADE-23*, ed: Springer, 2011, pp. 192-206.

[8] B. Motik, B. C. Grau, I. Horrocks, Z. Wu, A. Fokoue, and C. Lutz, "Owl 2 web ontology language: Profiles," *W3C recommendation,* vol. 27, p. 61, 2009.

[9] T. R. Gruber, "A translation approach to portable ontology specifications," *Knowledge acquisition,* vol. 5, pp. 199-220, 1993.

[10] J. S. Hare, P. H. Lewis, P. G. B. Enser, and C. J. Sandom, "Mind the gap: another look at the problem of the semantic gap in image retrieval," in *Electronic Imaging 2006*, 2006, pp. 607309-607309-12.