Semantics-Based Information Modeling for the Health-Care Administration Sector: The Citation Platform

Aristidis G. Anagnostakis, Maria Tzima, George C. Sakellaris, Dimitrios I. Fotiadis, *Member, IEEE*, and Aristidis C. Likas, *Senior Member, IEEE*

Abstract—An information brokerage environment for effective information structuring, indexing, and retrieval in the health-care administration sector is presented. The system is based on ontology modeling, natural language processing, extensible markup language, semantics analysis, and behavioral description. Semantics-based information acquisition is achieved through the uniform modeling, representation, and handling of domain-specific knowledge, both content-based and procedural. The system has been validated using information located on several repositories in the web and its performance is reported in terms of precision and recall.

Index Terms—Knowledge management, ontology modeling, semantics analysis.

I. INTRODUCTION

THE OVERWHELMING amount of administrative information in the health-care sector stresses the need for the development of knowledge integrative methodologies to facilitate meaningful acquisition and sharing of information. The requirement for seamless, personalized, and semantically accurate information diffusion is based on effective information processing [1].

To capture and represent the knowledge existing in health-care administrative information repositories, a semantic framework capable of describing in a uniform way both the information content and the potential procedural flow must be built [2], [3]. Such a semantic framework can be represented with the utilization of ontologies, which form an information model, consisting of domain-specific vocabulary bound through relationships to represent the cognitive meaning of the domain information [4]. The role of ontologies as a means for knowledge representation, information handling, indexing, and retrieval is highly recognized and constantly gains ground under the semantic-web [5]–[8].

We have witnessed a surge of interest in fields closely related to ontology engineering that tackle the discovery and automatic creation of complex multirelational knowledge structures [9], [10]. The natural language community tries to acquire word semantics from natural language texts [11], database researchers tackle the problem of schema induction [12], and people building intelligent information agents study learning of complex structures from semistructured input [13]. Efforts in the machine learning community pursue the induction of more concise and more expressive knowledge representation structures [14], [15].

The utilization of the extensible markup language (XML) infrastructure intends to make the information self-containing and provides a standard format for data exchange [16]. The deployment of XML for information wrapping and automated machine processing has become the most advisable means for text and multimedia data, since all types of content may be uniformly represented and exchanged [17]. Those advanced wrapping structures and mechanisms impose total control on the coherency and consistency of the introduced model.

For several years, the use of natural language processing (NLP) for meaningful concept extraction has been studied [4]. Although a reasonable level of performance has been achieved, the utilization of the identified concepts to facilitate effective information delivery is still very limited. Standard NLP techniques such as lexical scanning, parsing, and morphological analysis are used for automated analysis of natural language texts [18]. Most of the developed NLP techniques deal only with the syntactic processing leaving out the underlying semantics [19]. Ontologies and semantics networks can be used to refine the semantic accuracy of NLP. However, their use for semantic-based information indexing and acquisition is very limited, mainly due to the lack of a solid framework [13], [20], [21].

In this paper, a platform which integrates knowledge acquisition methods with ontologies' semantic modeling is presented. The platform is entitled CITATION and it can be used for processing of administrative information in the health-care sector, which can be found in several places on the web. The CITATION platform consists of four components. 1) The ontology server: The artifacts describing the administrative information in health care are represented in the internal ontologies as directional colored graphs. Mathematical notations and semantic categories are used to set the internal coherency constraints and verify the integrity [22]. 2) The information assimilation mechanism: Textual information from external resources is assimilated using NLP and the resulting XML file is equivalent to the initial content enriched with statistical and syntactical information. 3) The information indexing and retrieval mechanism: The potential of the ontology model is used
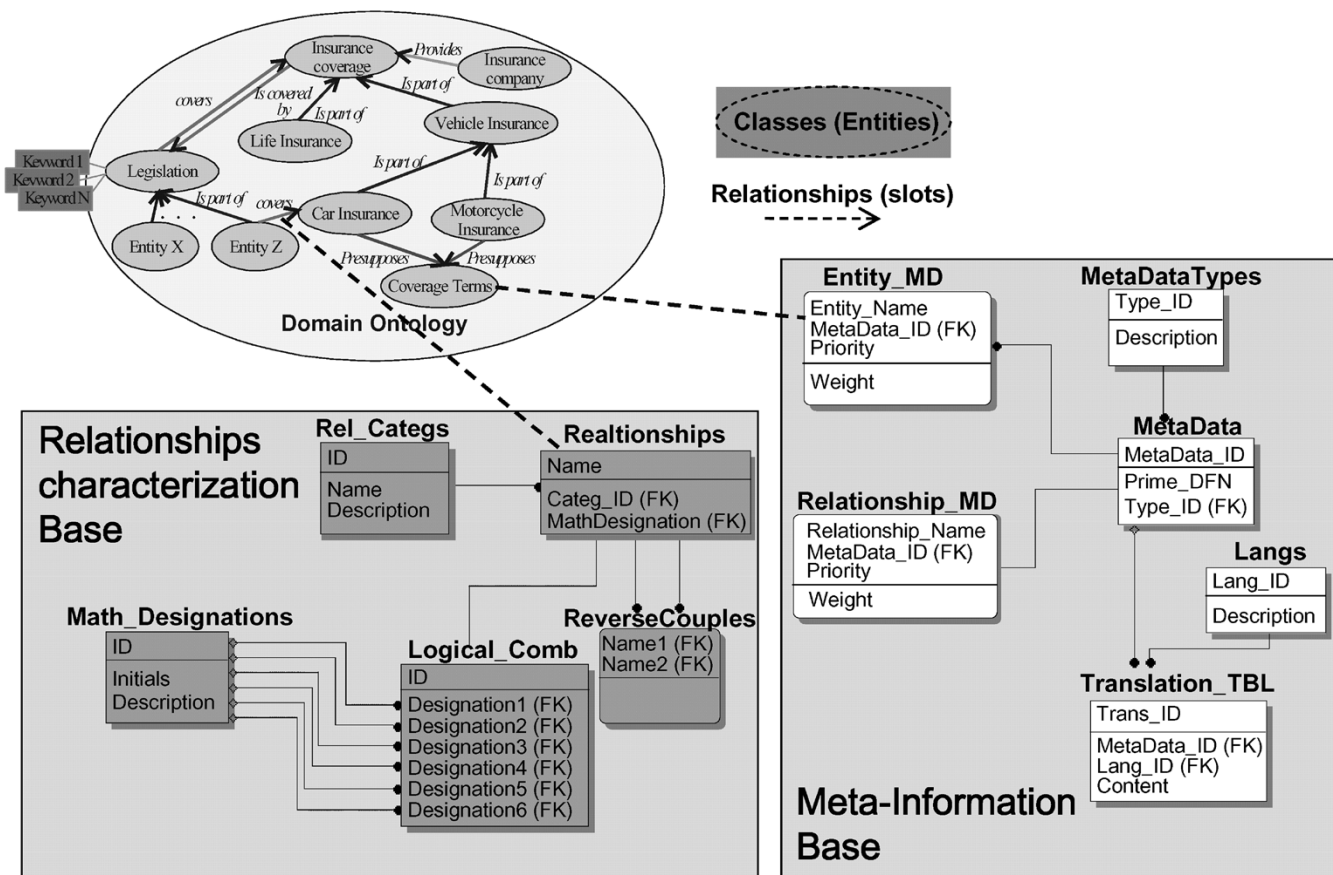
Fig. 1.   Ontology Model Knowledge Base.

for semantic-based indexing and acquisition, through the XML representation, which permits for coherent one-to-one mapping of the information to the corresponding internal entity-relationship (E-R) representation. 4) The querying mechanism: Natural language queries can be augmented using lexical thesauri such as UMLS, WordNet [23]–[25]. The constructed weighted graphs are matched against the internal domain ontologies to achieve semantic correlation.

The platform is capable of handling information on legislation, regulations, and procedures in the health-care administration sector. CITATION methodology has been applied to four pilot installations. The efficiency of the proposed platform is quantitatively evaluated using administrative health-care information from several repositories in the web.

II. SYSTEM DESCRIPTION

A. Basic Concepts Modeling

CITATION is based on Standford's Protégé 2000 framework [26] which is extended to represent special features, peculiarities, and constraints underlying the health-care administrative information. It promotes a uniform ontology representation by commencing a customized model for the efficient handling of the introduced semantics.

The introduced CITATION ontology model (COM) treats ontologies as directed multicolored graphs, whose nodes correspond to ontology entities, while the edges represent some sort

of relationship among them [28]. Using English as the core definition language, the ontologies provide indexing for multilingual content (English, Italian, and Greek). The core schema of the ontology base is depicted in Fig. 1.

Three major characteristics are identified in the definition of the COM: the abstract entities (classes/concepts), the relations holding among those, and the meta-information describing both the abstract entities and the relations [22], [29].

Representing abstract entities was straightforward, since there is a well-established understanding of the term: each entity corresponds to one piece of concept, and it is adequately realized through a unique class [20].

Relations are considered to be binary in the COM. In combination to the commonly used mathematical designations [30], relations are assigned additional semantics characterizations. The mathematical designation (Symmetric, Antisymmetric, Reflexive, Irrefrexive, Transitive, Weak-transitive) sets the formal constraints for automated integrity validation, while the semantic characterization defines the retrieval behavior. Table I depicts the meaningful combinations of the above for the CITATION model.

For the relations, three large semantic categories emerged, namely Hierarchical (the "is-part-of" relationships), Semantic Generalization (of the type "acacia-tree"), and Sequential (relationships defining sequence in time). Each of the imported relations is denoted and taxonomized accordingly.

	Symmetric	Antisymmetric	Reflexible	Irreflexible	Transitive	Weak Transitive
Symmetric		X	✓	✓	✓	✓
Antisymmetric	X		X	X	✓	✓
Reflexible	✓	X		X	✓	✓
Irreflexible	✓	X	X		✓	✓
Transitive	✓	✓	✓	✓		
Weak Transitive	✓	✓	✓	✓		

Based on the semantic characterizations, semantically augmented information retrieval is facilitated. For example:

- The identified entity of interest being part of a "*time-sequence*" chain will cause the excitation of the preceding entities as well.

  Thus, for the time sequence chain

  $$\ldots \text{application} \rightarrow \text{certification} \rightarrow \textbf{intake} \ldots$$

  where "$\rightarrow$" indicates the relationship "*presupposes*" and the identification of the entity "**intake**" as a high-interest entity will cause the excitation of the entities "certification" and "application" as well.

- The entity of interest being part of an "*is-part-of*" hierarchy will cause references to lower level entities to be created.

  Considering for instance the hierarchical chain

  $$\ldots \text{clinic} \rightarrow \text{department} \rightarrow \textbf{hospital} \ldots$$

  where the "$\rightarrow$" denotes an "*is-part-of*" relationship and the identification of the entity hospital causes references to the entities "department" and "clinic" to be created as well.

To enhance the information processing, semantic indexing, and acquisition, a flexible metadata structure capable of representing a wide range of initial information was developed based on the XML framework [12]. The CITATION document representation schema (CDRS) provides the means for the semantics-based acquisition and interchange of the health-care administrative information [32]. The mapping between the initial information and the bearer XML object is achieved through the unique tagging of every semantic unit and subunit identified.

Metadata describing the initial sources in terms of location and content definition are kept; version-keeping mechanisms are also implemented. The CDRS structure, based on the Dublin Core initiative, introduces elements that provide resourceful representation of information objects with respect to the specific domain of discourse [33]. CDRS instantiates the representative object classes to hold essential parts of the initial content along with semantically related meta-information, while maintaining the one-to-one mapping to the internal content base. The CDRS schema is shown in Fig. 2.
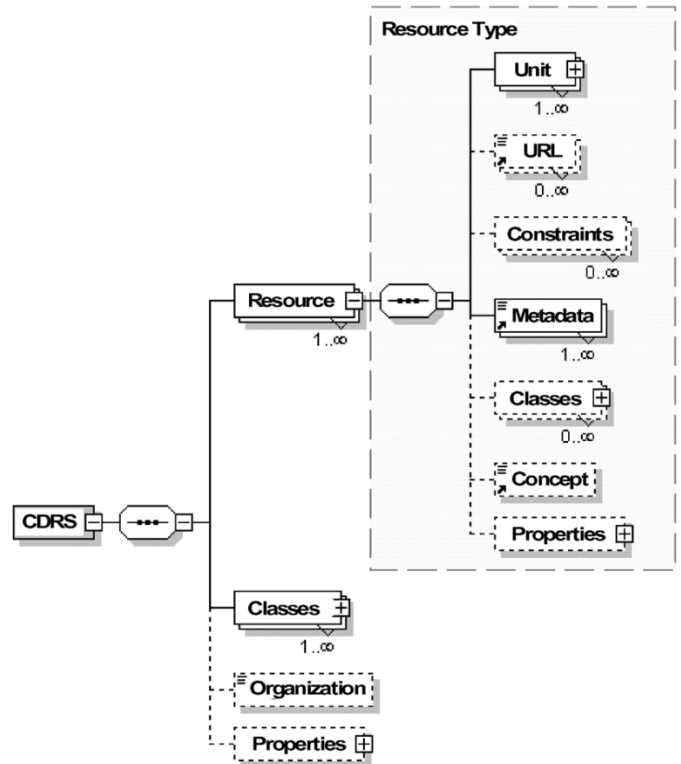

Fig. 2.   CDRS schema.

B. CITATION Global Architecture

The CITATION platform, in order to deliver semantics-based information indexing, incorporates several features. A standardized document type definition for health-care administration is introduced to define explicitly the developed XML schema (CDRS) and provide uniform information representation and sharing. Through the corresponding XML wrapper, the initial information is mapped against the internal ontology [17], [31]. Tokens from the users' queries are being extracted, augmented by external lexicons such as the UMLS and WordNet, filtered by the user's personal characteristics, and matched against the internal ontology as well. The architecture of the platform and the information flow is depicted in Fig. 3.

The system consists of the following modules.

Ontology Server: Generic rules are recognized and stored uniformly. The module supervises the coherency of the domain ontologies based on specific integrity constraints set by the COM. The core of the mechanism is based on the Protégé library [26], which provides an advanced ontology-development and knowledge-acquisition framework. A front-end based on the simple object access protocol has been developed in the form of WebServices. The implementation makes the core services of the CITATION platform accessible by any standalone or web-based application. Mechanisms for the upkeep and augmentation of the model itself have been developed, based on representation graphs. Semiautomated creation of relations among the newly inserted and pre-existing entities is supported.

Conformation Module: Queries are posed in natural language. The tokens of interest are extracted, augmented using external lexicons (UMLS and WordNet), and filtered to fit each individual user profile. The weighted correlation graph is
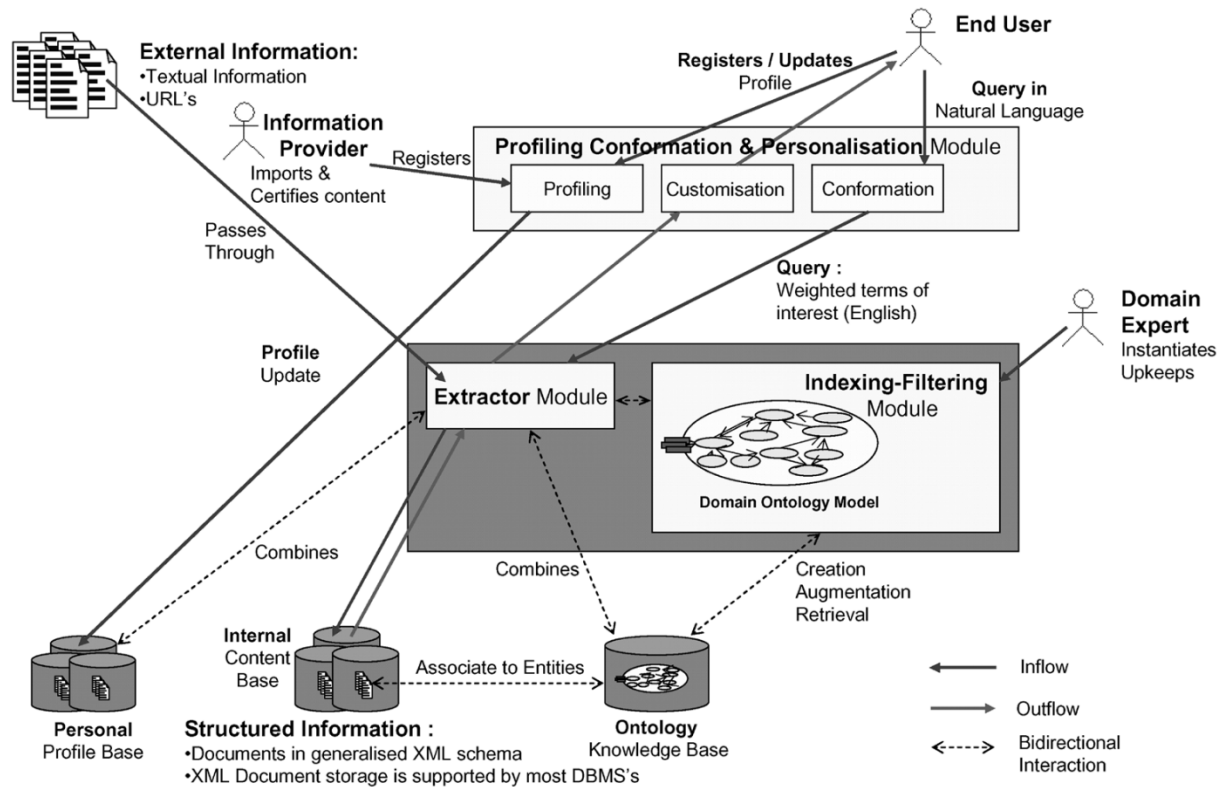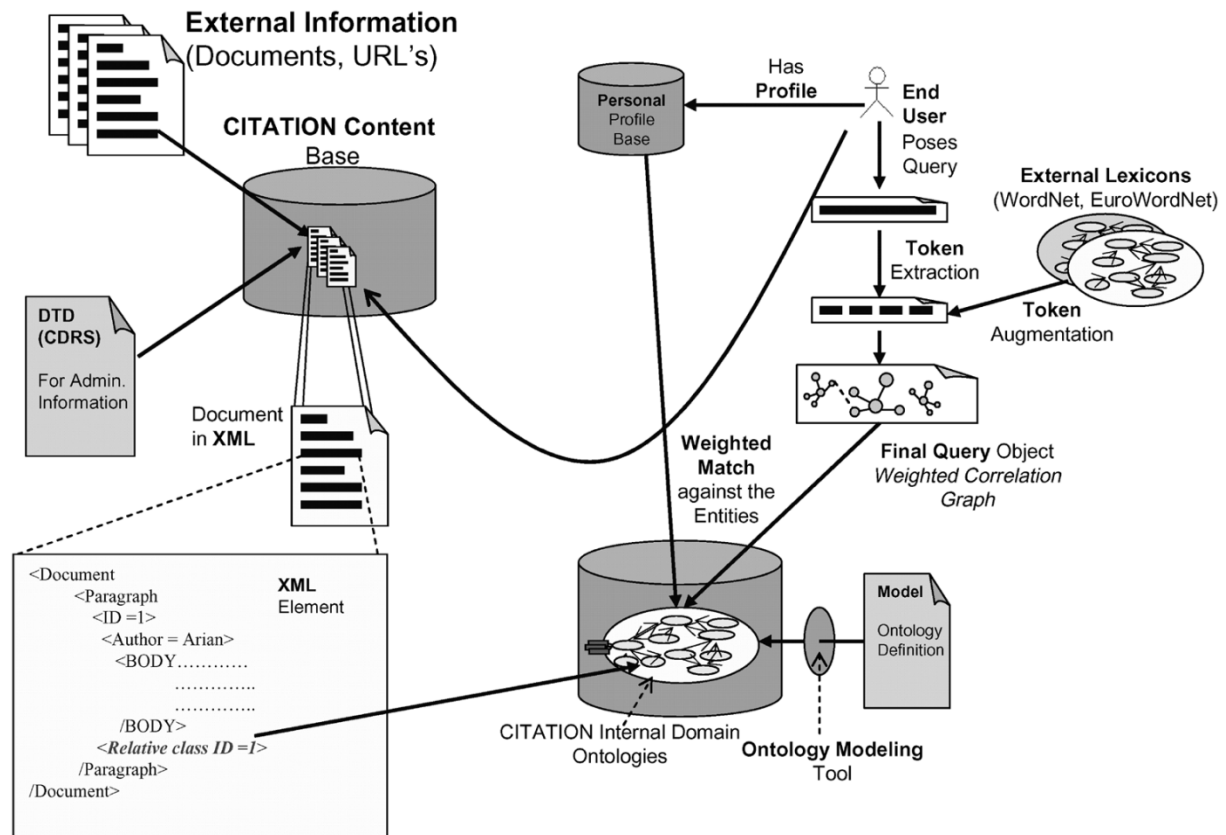
Fig. 3.   CITATION architecture.

Fig. 4.   Semantics-based information retrieval process.

formed and matched against the entities of the internal domain ontologies.

Customization—Presentation Module: It delivers comprehensive information in conformance to the user-facing
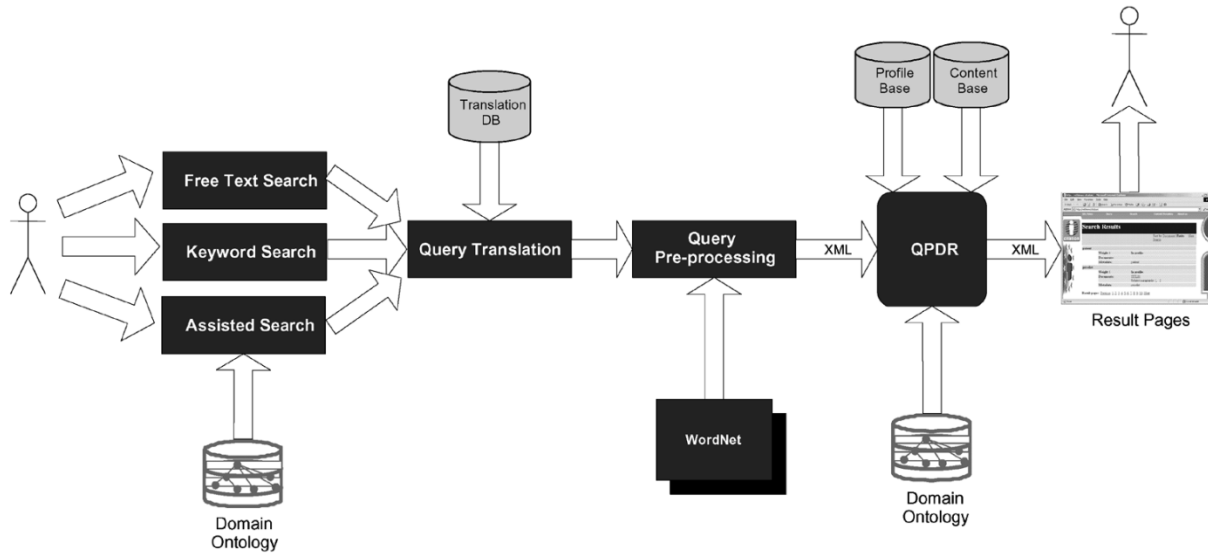
Fig. 5. QPDR functionality.

equipment capabilities and the user's personal characteristics and preferences.

Extractor Module: Morphological and syntactic analysis of the external information sources is performed, utilizing external tools [18] to form the XML equivalent. Additional statistical information is used to enrich the initial content.

Information Representation, Indexing, and Retrieval Module: The ontologies are used for semantic indexing and acquisition. The intitial information filtered by the extractor module is related against the ontologies in a weighted manner. The XML equivalent of the initial information is stored in a solid E-R representation [13]. One-to-one mapping is achieved through naming convensions based on the XML elements hierarchy.

C. Processes

Two major processes are included: Information extraction and query processing and document retrieval (QPDR), which are described in detail below.

Information Extraction: Administrative information is currently vastly textual. Therefore, advanced text manipulation methodologies have been employed [11], [13]. Information extraction is conducted in three stages: 1) analysis of the imported documents (preprocessing based on statistical and syntactical analysis); 2) structuring of the information according to the CDRS (semantics-based segmentation and feature extraction); and 3) storage of the XML equivalents in the E-R repository. For preprocessing, the TextAnalyst component library has been integrated in the system. The output forms a multilevel frequency-based structure of elements (letters, syllables, stems, morphemes, words, and phrases). Small deviations on specific parameters such as the link threshold (expressing the frequency of the concepts co-occurrence) proved to have significant effect on the efficiency of the extractor module.

QPDR: Natural language and assisted querying modes are provided. The QPDR mechanism preprocesses the query to extract the tokens of interest, structures the outcome in XML,

TABLE II
VALIDATION PROCESS CONTINGENCY TABLE

	Relevant	Not Relevant
Extracted	$G \cap R$	$G \cap R'$
Not Extracted	$G' \cap R$	$G' \cap R'$

TABLE III
STANDARD DEVIATION FOR PRECISION AND RECALL OVER THE QS

	HotBot®		CITATION Total	
	Precision	Recall	Precision	Recall
STD	0,0013	0,0037	0,0114	0,0519

augments the tokenized information utilizing external thesauri (WordNet and UMLS-Lex), commits a token-by-token translation based on external dictionaries (EuroWordNet), filters the tokens according to the user's profile, and performs a weighted match against the entities of the domain ontology [19]. The semantics-based information retrieval process is shown in Fig. 4, while the QPDR functionality is depicted in Fig. 5.

The CITATION platform supports multilingual information retrieval; the core language being English, domain applications in Italian and Greek have been also demonstrated.

III. VALIDATION

For the validation of the system, the standard metrics precision and recall have been accumulated to provide with quantitative effectiveness estimation. Tests have been conducted separately for the Extractor module and for the system Query mechanism, to estimate both the validity of the external information assimilation process and the performance of the platform in information retrieval, respectively. Recall indicates the proportion of relevant material extracted, while precision the proportion of extracted material relevant to the query.

TABLE IV
INFORMATION RETRIEVAL RESULTS SUMMARY

| | HotBot® Results | | CITATION IR Results Summary | | | | | |
| | | | Direct | | Implied (max 2-hops) | | Total | |
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
$\overline{QS_1}$	0,4918	0,8571	1,0000	0,6857	0,1489	0,2000	0,3478	0,9143
$\overline{QS_2}$	0,4414	0,8596	0,8571	0,8421	0,1231	0,1404	0,4628	0,9825
$\overline{QS_3}$	0,4603	0,7838	0,8667	0,3514	0,3333	0,3243	0,4902	0,6757
Average	0,4645	0,8335	0,9079	0,6264	0,2018	0,2216	0,4336	0,8575

Precision is defined as

$$\text{Precision} = \frac{|G \cap R|}{|G|}$$

and recall as

$$\text{Recall} = \frac{|G \cap R|}{|R|}$$

where G is the set of the extracted data d and R is the set of the data relevant to the posed query.

The contingency table used for the validation process is given in Table II.

Precision represents the fraction of relevant documents within the retrieved documents while recall represents the fraction of relevant documents that were retrieved. Increase of recall introduces more irrelevant hits, thereby reducing precision. Increase of precision reduces recall by potentially removing relevant hits. An ideal goal for a retrieval engine is to increase both precision and recall. Consequently, the best values for precision are obtained if all retrieved documents are relevant and for recall if all relevant documents are actually retrieved.

The results have been cross-validated comparing those with the Intkomi Corporation general purpose search engine HotBot.

IV. RESULTS

Several tests have been conducted to evaluate the efficiency of the developed mechanisms; both the extractor module and the overall infrared (IR) process have been evaluated using the aforementioned precision and recall attributes [28]. Quantitative tests have been conducted to assess the effectiveness of the CITATION platform using a reference corpus of 1000 selected documents located on several repositories in the web. Each of those was *a priori* judged by six health-care professionals for its relevance to 21 queries categorized to three query sets (QSs), relevant to 1) insurance coverage, 2) health facilities location, 3) hospital admission procedures.

Each of the queries was posed independently through the CITATION query mechanism operating in natural language mode and through the HotBot general purpose search engine for comparison.

The documents and the results were annotated as "*a priori* relevant" and "*a posteriori* relevant" to the queries, respectively, via voting. Out of the initial set, 496 documents were identified as relevant to at least one of the queries by the majority of the users.

A sparse reference ontology of 95 terms was invoked, 25% of which (24 terms) were related to more than one of the rest

TABLE V
CITATION VERSUS HOTBOT PRECISION AND RECALL

| | CITATION vs. HotBot® | | | |
| | CITATION Direct | | CITATION Total | |
	Precision	Recall	Precision	Recall
Average Increase	95,45%	-24,84%	-6,67%	2,87%

through one or more of seven distinct relationships. For each query $q_i, (1 \leq i \leq N)$, the set of the *a priori* relevant documents R_i was defined and the result-set G_i was produced by both engines. The results have been accumulated for each QS, by averaging the $\text{Precision}_i = (|G_i \cap R_i|)/(|G_i|)$ and the $\text{Recall}_i = (|G_i \cap R_i|)/(|R_i|)$ for $N = 7$ (the number of queries per QS).

The standard deviation over the QS for the precision and recall was low (order of 10^{-2}) both for the CITATION query mechanism, and for the HotBot engine (Table III).

For the CITATION mechanism evaluation, the retrieved documents presenting an explicit keyword relationship to the posed query were considered to be *Directly* related, while the rest were considered *Implicitly* related. During the test, a maximum of two-hop distance from the initially identified term was considered. The accumulated precision and recall figures are presented in Table IV.

Compared to the results of the generic search machine utilized, the CITATION engine presents considerably higher precision (95%) with a moderate loss in the recall accuracy (24%) to the *Directly* matched documents.

Considering the overall results of the CITATION engine (both *Implicit* and *Direct*), a total average of 2.8% increase was observed to the recall accuracy which corresponds to new, implicitly related information, no otherwise identified. The respective corresponding precision reduction was 6.6% (Table V). Fig. 6 depicts the overall precision and recall as a function of the distance (number of hops) from the initially identified terms, and is indicative of exponential reduction.

The extractor module was evaluated independently as well. A concrete corpus of 21 (7 per QS) documents was imported several times; the link relevance parameter which expresses the joint frequency of the concepts co-occurrence was gradually increased, resulting in more extracted keywords per iteration. The accuracy of the extraction process was evaluated based on the actually relative keywords previously identified by the users in each of the test documents. The results indicate that a recall accuracy of "1" can be achieved with not much precision loss.
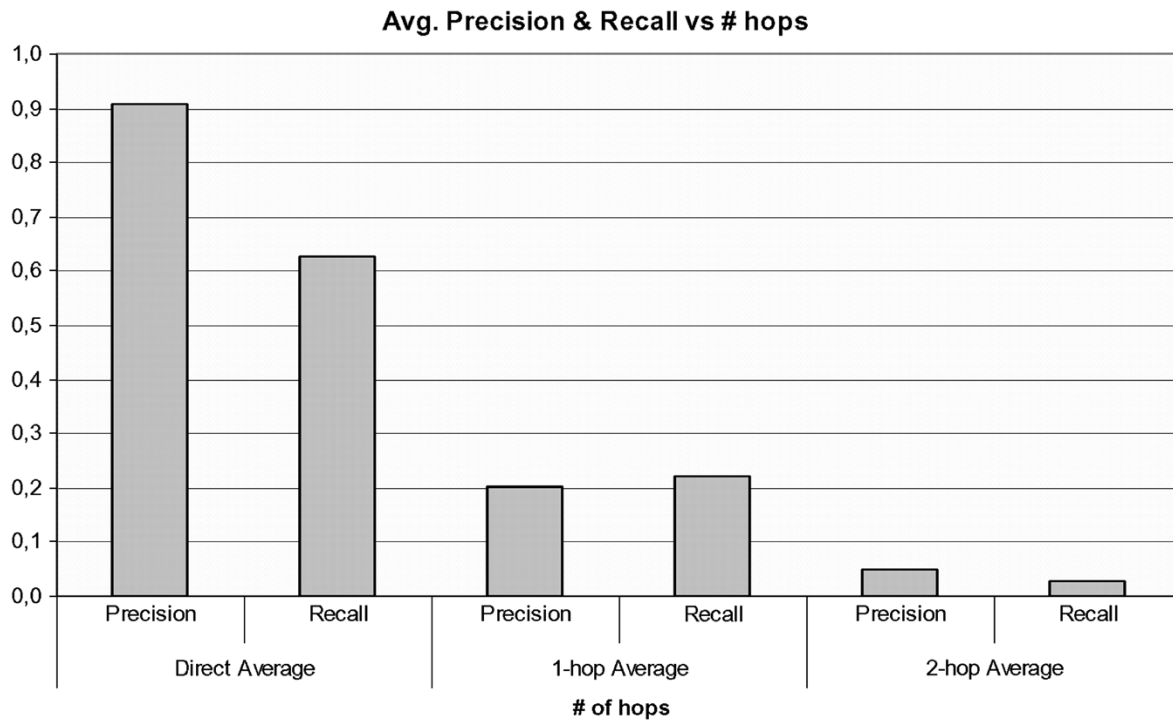
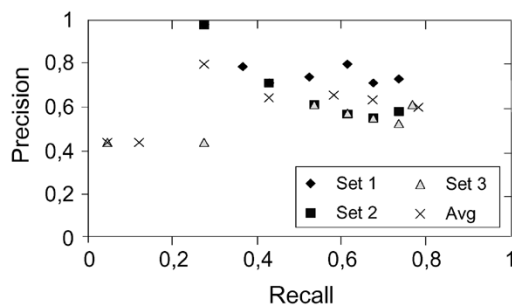Fig. 6. Average precision and recall versus number of hops.

Fig. 7. Extractor module precision versus recall.

The accumulated precision versus recall is depicted in Fig. 7. In most cases, the precision value of the performed queries as well as their average is higher than the respective recall values, which is indicative of the semantics mining potential of the extractor module.

V. DISCUSSION

Ontology-based computing is emerging as a natural evolution of existing technologies to cope with the information onslaught. A considerable amount of effort has been placed in standardizing the representation (UMLS, Open GALEN) and the interchange of medical information (HL7, CEN 13 606) [24], [34]. However, the utilization of the well-defined medical standards for managing the intensive heterogeneous health-care administrative information is still very limited [14], [30].

The CITATION platform takes advantage of the potential of the ontology-driven semantics representation and the expressiveness of XML and implements an applicable framework for the uniform representation, semantics-based retrieval, and structured interchange of health-care administrative information. The platform can be easily expanded to other fields such as medical information.

A solid model and a graphical user interface to support uniform ontologies management was developed. The ontology model is designed to allow for constant restructuring and facilitate the depiction of effective ontologies, over wider spectra of "domains of discourse." The association rules and constraints identified manage to capture and represent a significant portion of the health-care administration knowledge, creating a solid ambient for semantic-based information retrieval. Yet, the introduced model is capable of representing domain ontologies over a wide spectrum of administrative activity.

Enhanced capabilities in semantics-based health-related information sharing and retrieval are the outcome. The system focuses on the automated structuring and provision of semantic brokerage to administrative health-care information. However, advanced IR performance is achieved as well (Table IV). The system was proved advantageous since: 1) In the *Direct* search significantly higher precision (95%) is achieved, while at the same time the recall is less than 25% reduced compared to free text search. This is indicative of the effective filtering of the extractor module. 2) In *Total*, an increase of 2.8% is achieved to the overall recall. This is very important, since it corresponds to the retrieval of *Implicitly*-only related documents, for which no obvious textual association exists. Both the precision and the recall were exponentially reduced with the number of hops from the initially identified matching terms (Fig. 6).

Our results indicate that ontologies and semantics networks can be applied on the health-care administration sector both to

improve quality of the information management, and enhance the efficiency of the information retrieval. The indications so far are stimulating.

VI. CONCLUSION

This paper describes a real-world information brokerage environment for multilingual health-care administration information. The proposed platform automatically structures documents in XML format, filters, and indexes them with ontologies that provide contextual and procedural information over specific domains. It has been proven that using this information, query mechanisms provide more accurate semantics-based information acquisition.

The CITATION platform introduces a uniform way to structure and exchange health-related administrative information, based on 1) the development and establishment of a solid framework for health-related administrative information interchange and 2) the elaboration of an ontology model setting the rules and constraints for the adequate capturing and representation of the semantics over a wide variety of domains of discourse, focused on the meaningful information retrieval.

The benefits of applying structural information and ontology integrative methodologies to health-care administrative information have been quantitatively evaluated.

REFERENCES

[1] M. Joubert, S. Aymard, D. Fieschi, F. Volot, P. Staccini, J. Robert, and M. Fieschi, "ARIANE: Integration of information databases within a hospital intranet," *Int. J. Med. Informatics*, vol. 49, pp. 297–309, 1998.

[2] J. Swa, *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Pacific Brove, CA: Brook/Cole, 2000.

[3] A. Goni, E. Mena, and A. Illarramendi, "Quering heteregeneous and distributed data repositories using ontologies," in *Information Modeling and Knowledge BaseIX*, P.-J. Charrel and H. Jaakkola, Eds: IOS Press, 1998, pp. 19–34.

[4] U. Hahn, M. Romacker, and S. Schultz, "How knowledge drives understanding—Matching medical ontologies with the needs for medical language processing," *Artif. Intell. Med.*, vol. 15, no. 1, pp. 25–51, 1999.

[5] Y. Ding, D. Fensel, M. Klein, and B. Omelayenko, "The semantic web: Yet another hip?," *Data Knowl. Eng. J. Elsevier*, vol. 41, no. 2–3, pp. 205–227, Jun. 2002.

[6] A. Maedche and S. Staab, "Discovering conceptual relations from text," in *Proc. ECAI-2000*. Amsterdam, The Netherlands, 2000.

[7] D. McGuinness, "Ontologies come of age," in *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*, D. Fensel, J. Hendler, H. Lieberman, and W. Wahlster, Eds. Cambridge, MA: MIT Press, 2002.

[8] S. El-Beltagy, W. Hall, D. De Roure, and L. Carr, "Linking in context," in *Proc. 12th ACM Conf. Hypertext and Hypermedia*, 2001, pp. 151–160.

[9] J. Liebowitz, *Knowledge Management: Learning from Knowledge Engineering*. Boca Raton, FL: CRC Press, 2001.

[10] B. Gaines and M. Shaw, "Integrated knowledge acquisition architectures," *J. Intell. Inf. Syst.*, vol. 1, pp. 32–35, 1992.

[11] R. A. Falbo, G. Guizzardi, A. C. Natali, G. Bertollo, F. F. Ruy, and P. G. Mian, "Toward semantic software engineering environments," in *Proc. 14th Int. Conf. Software Engineering and Knowledge Engineering*, Ischia, Italy, Jul. 15–19, 2002.

[12] J. Cook, "XML sets stage for efficient knowledge modeling," in *IEEE IT Professional 2*, May/Jun. 2000, pp. 55–57.

[13] P. Johannesson, "A method for transforming relational schemas into conceptual schemas," in *10th Int. Conf. Data Engineering*, M. Rusinkiewicz, Ed.. Huston, TX, 1994, pp. 115–122.

[14] V. Kashyap and A. Sheth, "Semantic heteregeoneity in global information systems: The role of metadata, context, and ontologies," in *Cooperative Information Systems: Tends and Directions*, M. Papazogloyu and G. Schlageter, Eds., 1998, pp. 139–178.

[15] P. Glasser. (1999, Jan.) The knowledge factor. *CIO Mag.* [Online]. Available: http://www.cio.com/archive/010199/know.html

[16] S. Decker *et al.*, "The semantic web: The roles of XML and RDF," *IEEE Internet Comput.*, vol. 4, no. 5, pp. 63–74, Sep./Oct. 1999.

[17] M. Erdmann and R. Studer, "How to structure and access XML documents with ontologies," *Data Knowl. Eng.*, vol. 36, pp. 317–335, 2001.

[18] Megaputer Intelligent TextAnalyst® (2004). [Online]. Available: http://www.megaputer.com/

[19] *Handbook of Computational Linguistics*, R. Mitkov, Ed., Oxford Univ. Press, Oxford, U.K., 2001. Term extraction and automatic indexing, C. Jacquemin, D. Bourigault.

[20] V. Sugumanana and V. Storey, "Ontologies for conceptual modeling: Their creation use and management," *Data Knowl. Eng. J. Elsevier*, vol. 42, pp. 41–64, 2002.

[21] B. Chandrasekaran, J. Josephson, and V. Benjamins, "What are ontologies and why do we need them?," *IEEE Intell. Syst.*, vol. 14, no. 1, pp. 20–26, Jan./Feb. 1999.

[22] Theory FRAME-ONTOLOGY Stanford University (1999). [Online]. Available: http://www.ksl.stanford.edu/knowledgesharing/ontologies/html/ frame-ontology/

[23] T. Andreasen, "Query evaluation based on domain-specific ontologies," in *20th IFSA/NAFIPS Int. Conf. Fuzziness and Soft Computing (NAFIPS 2001)*, Vancouver, BC, Canada, 2001, pp. 1844–1849.

[24] The Unified Medical Language System (2004). [Online]. Available: http://www.nlm.nih.gov/ research/umls/

[25] G. Miller *et al.*, "Introduction to wordnet: An on-line lexical database," *Int. J. Lexicography*, vol. 3, pp. 235–244, 1990.

[26] E. Grosso, H. Eriksson, R. Fergerson, S. Tu, and M. A. Musen, "Knowledge modeling at the millennium—The design and evolution of protégé-2000," in *Proc. KAW-1999*, Banff, Canada, 1999.

[27] N. F. Noy and M. A. Musen, "PROMPT: Algorithm and tool for automated ontology merging and alighment," in *Proc. 17th Nat. Conf. Artificial Intelligence (AAAI 2000)*. Austin, TX, 2000.

[28] R. Jasper and M. Uschold, "A framework for understanding and classifying ontology applications," in *Proc. 12th Workshop Knowledge Acquisition, Modeling, and Management (KAW'99)*, vol. 1, Banff, Canada, 1999, pp. 4 9 1–4 9 20.

[29] O. Corcho, M. Fernández-López, and A. Gómez-Pérez, "Methodologies, tools, and languages for building ontologies. Where is their meeting point?," *Data Knowl. Eng. J. Elsevier*, vol. 46, no. 1, pp. 41–64, Jul. 2003.

[30] M. Gruninger and M. S. Fox, "Methodology for the design and evaluation of ontologies," in *Proc. Workshop Basic Ontological Issues in Knowledge Sharing*, IJCAI-95, Montreal, QC, Canada, 1995.

[31] H. Assadi, "Construction of a regional ontology from text and its use within a documentary system," in *Proc. Int. Conf. Formal Ontoloyu and Information Systems (FOIS'98)*, Trento, Italy, 1998.

[32] OMG XML Metadata Interchange (XMI) Specification (2001). [Online]. Available: http://www.omg.otg

[33] The Dublin Core Metadata Initiative (2004). [Online]. Available: http://dublincore.org/

[34] The OpenGALEN (2004). [Online]. Available: http://www.opengalen.org/

Aristidis G. Anagnostakis was born in Arta, Greece, in 1974. He received the Diploma degree in computer and informatics engineering, and the M.Sc. degree, both from the University of Patras, Greece.

He worked for the R&D Department of Intracom S.A. and for the IT Department of the Altec Communication Networks as an Informatics Engineer. Since 2001, he has been with the Medical Technology and Intelligent Information Systems Unit, Department of Computer Science, University of Ioannina, Greece, in charge of the technical coordination on a number of research projects. His current interests include knowledge management, ontology modeling, natural language processing, and biosignal analysis.

Maria Tzima was born in Ioannina, Greece, in 1974. She received the B.S. degree in computer studies from the University of Brighton, U.K.

Since October 2000 she has worked with the Medical Technology and Software Development Unit, Department of Computer Science, University of Ioannina, Greece, as a Technical Manager of several national and European projects of the IST program.

Dimitrios I. Fotiadis (M'01) was born in Ioannina, Greece, in 1961. He received the Diploma degree in chemical engineering from National Technical University of Athens, Greece, and the Ph.D. degree in chemical engineering from the University of Minnesota.

Since 1995, he has been with the Department of Computer Science, University of Ioannina, Greece, where he currently is an Associate Professor. He is the director of the Unit of Medical Technology and Intelligent Information Systems. His research interests include biomedical technology, biomechanics, scientific computing, and intelligent information systems.

George C. Sakellaris was born in Piraeus, Greece, in 1977. He received the B.Sc. and M.Sc. degrees in computer science from the University of Ioannina, Greece, in 2000 and 2004, respectively.

His current research interests include knowledge management, ontology-based applications, and graphical models in bioinformatics.

Aristidis C. Likas (S'91–M'96–SM'03) was born in Athens, Greece, in 1968. He received the Diploma degree in electrical engineering and the Ph.D. degree in electrical and computer engineering from the National Technical University of Athens, Greece.

Since 1996, he has been with the Department of Computer Science, University of Ioannina, Greece, where he is currently an Assistant Professor. His research interests include neural networks, machine learning, pattern recognition, and intelligent systems for biomedical engineering.