MONITORING EVOLUTION IN HEALTHCARE DOMAIN

Varlamis Iraklis
Department of Informatics,
Athens University of Economics and Business,
76 Patision St., 10434, Athens, GREECE
varlamis@aueb.gr

Apostolakis Ioannis
National School of Local Government,
Technical University of Crete,
211 Peiraios St., Tavros, Athens,
GREECE
apost@ekdd.gr

Abstract

Healthcare is a highly evolving domain, which attracts the interest of both researchers and practitioners. In such an evolving area, new trends appear daily and older standards become out of fashion very fast. The paper presents a methodology for seizing knowledge from rich information sources of the healthcare domain and processing it in order to improve the domain understanding for organizations and individuals. To achieve that, we suggest the development of technologies for systematically & automatically updating, dynamic knowledge maps of healthcare domain. These maps will act as advanced business intelligence support for market, products & technology watch and will represent a novel knowledge management practice (technology evolution assessment). The methodology may lead to a knowledge management platform for the intelligent monitoring, planning, forecasting of technological evolution in highly dynamic domains. Doctors, researchers and healthcare organizations that utilise the platform will achieve state-of-the-art awareness in research and development and will increase competitiveness in their domain.

Keywords: monitoring, knowledge management, ontology evolution.

Introduction

The vast amount of information that is currently available in the Healthcare domain, the multitude and diversity of information providers and the absence of common standards for information exchange, result in a world Babel that does not promote common efforts and decelerates progress. The first step towards the integration of information is the definition of formal terminologies, ontologies and standards. Standards provide the common base for communication whereas ontologies offer a formal description of the domain at a certain time. Terminology works as a shibboleth for individuals and companies wishing to join the domain and ontology as a synopsis of the domain knowledge. However, the attempt to preserve a stable state for a long period of time in a highly evolving domain is unfeasible and unworkable, since new types of devices and products are developed, new markets emerge, and new areas of interest are introduced by the scientific community. The continuous monitoring of the domain is necessary for discovering new trends and profit of them or for detecting declining and uninteresting objectives and stop giving effort on them.

The current work focuses on Healthcare, a technology intensive domain, with rich informational background, plenty of efforts for standards and semantics definition, but can be extended on other domains. Healthcare systems and applications undergo constant change, so the relevant terminology needs to change in parallel, in order to allow understanding of the produced information and the integration of the amassed knowledge. For example, each year the FDA (Federal Drug Administration) approves about 150 new drugs and thousands of changes to the "label" of existing drugs, the VHA (Veterans Health Administration) manages label changes and drug "packaging" changes, and the NLM (National Library of Medicine) maintains an index of references to proposed or approved medications in the world's biomedical literature (SNOMED Clinical Terms – Wang et al 2001).

Individuals or organizations that wants to play an active role in the healthcare domain must first have a good understanding of the domain and then be able to monitor its evolution. In the following we address the general knowledge management process in terms of: what type of knowledge is important, where and by whom this knowledge is produced, the nature of the knowledge (tacit or not) and under what work scenarios this knowledge needs to be utilized.

Background

The two fundamental steps for successful knowledge management are information acquisition and knowledge representation. However we should also take care of the knowledge evolution over time in order for the collected knowledge to be helpful to users. In this paper we focus on the information that is available on the web, so for the acquisition we will refer to tools that extract information from web pages.

The data stored electronically is not only HTML or XML pages but also text files, mail files, product tables, etc. Therefore, if we want to extract data, which is important to the users, we need to have a powerful tool to analyse such documents and obtain the relevant information. The most common way to extract information from the Web is by generating a *wrapper* for every set of pages that share the same format. Wrappers (Chawathe et al. 1994, Ashish & Knoblock, 1997, Hammer et al, 1997, Kushmerick et al., 1997) parse unstructured data and map it into a structured (i.e. database, knowledge base) or semi structured form (i.e. data matrix or tagged language).

Ontologies play a crucial role in enabling the processing and sharing of Web-based knowledge between applications. They specify a domain-specific vocabulary of entities, classes, properties and functions, and the relationship between them. They provide a shared and common understanding of a domain that can be communicated across people (for human collaboration) and application systems (e.g. for the communication between agents). An example of ontology for healthcare is MeSH (Medical Subject Headings http://www.nlm.nih.gov/mesh/meshhome.html) a taxonomy of biomedical terms that is updated annually.

As stated by Tuttle et al (Tuttle et al, 2001), what is missing in healthcare and biomedicine is a way to link its progress and experience with that occurring in the Semantic Web community. Resource Description Framework (RDF, 2005) is an emerging standard used by the Semantic Web community for storing metadata. RDF's broad goal is to define a mechanism for describing resources that makes no a priori assumptions about a particular application domain or the associated semantics. In healthcare knowledge management systems, RDF can automate resource discovery and management. RDF is the basis for several other ontology-definition languages such as **RDFS** (RDFS, 2004) and **DAML+OIL** (DAML-OIL, 2001) that define the primitives for creating ontologies.

Benefits from knowledge monitoring

The benefits from creating, updating and utilizing a knowledge warehouse are: a) the enhancement of creativity and innovation, b) the strengthening of position and competence and c) the increased responsiveness. The ability to grasp the dynamic profile of the discipline may translate to impressive gains of wealth and employment security. Healthcare is a domain with hundreds of disciplines. Companies hardly maintain a national identity as they span activities all over the world. Slow reaction to developing dynamics may cost the viability of business and/or thousands of jobs. Doctors typically assess the evolution of their disciplines by reading journals, attending conferences and sometimes by hearsay. Instead, the web (as well as other resources) offers scattered and distributed information that is impossible to analyse manually. It is preferable for a doctor to browse an ontology for new medicines that cure a disease, than searching the web sites of the pharmaceutical companies for new products. It would be more helpful if the ontology links to a web page that describes the medicine or to the web site of the pharmaceutical company that produces the medicine. Moreover, doctors can easily locate equipments or treatments that are not applied any more and replace them.

The proposed methodology suggests the combination of intelligent agents together with dynamic ontologies in a knowledge management platform that will systematically and automatically extract knowledge from data sources and will synchronize domain knowledge (as expressed in domain ontologies) to the present needs taking into consideration the associations emerging from knowledge processing. The developed knowledge management platform can accommodate for any scientific or engineering field. The users of such platform

can essentially monitor the relative position of a discipline with regards to other fields, businesses and markets. More specifically, a knowledge management platform that adopts the proposed methodology is expected to:

(i) Discover "hidden" dependencies across heterogeneous domains and potential for knowledge transfer across these domains,

(ii) Better prepare the professionals to the new challenges, and increase their efficiency,

(iii) Guide researchers to better synchronise their research with the needs of the market,

(iv) Help young scientists to become better equipped to face the needs of their profession,

(v) Guide universities and institutes to revise curricula and offer up-to-dated education and knowledge, and

(vi) Aid healthcare industry to exploit opportunities and profit out of the knowledge.

An additional outcome of this process is the establishment of a domain knowledge base, rich in semantics, which can be used as the Healthcare standard that interconnects all case specific standards. Standards for products and services (Product Data Utility / PDU / www.CHeStandards.org/pdu and United Nations Standard Products and Services Code – UNSPSC / www.unspsc.org), customer identification (global location number – GLN), market classifications (Class of Trade standard), and more (see Healthcare Level 7 web site http://www.hl7.org) can be mapped to a common knowledge base and updated in a regular basis.

Roadmap to knowledge monitoring

In the following we detail on the steps that comprise the knowledge management process as a whole. Knowledge assessment is performed after each step to guarantee the quality of the produced results.

Knowledge development

Domain description: It would be unrealistic and impractical for a knowledge management system to process all the information that exists in a domain and produce knowledge from scratch. Therefore, an initial domain description from experts in a commonly accepted schema, is a prerequisite of the knowledge management process. These static ontologies form the "core" ontology system. They will be further developed and refined in order to obtain dynamic evolution attributes. Ontology creation tools will assist experts in the definition of concepts, instances and their relations. The criteria that the ontology should validate are: **clarity** (concepts, relations and axioms should map the real world needs), **coherence** (axioms and relations must not contradict each other), **extendibility** and **adaptability** (to new and declining concepts).

Selection of sources: The second step involves the selection of data sources. Web sites with explicit structure, dynamic and frequently updated content will be ideal sources. The reliability and awareness of the source provider are very crucial for the success of the whole process. The knowledge of domain experts can prove very helpful in the selection of the appropriate sources. The HL7 and NLM web sites are useful sources, as well as the web sites of large pharmaceutical and healthcare systems producers, healthcare quality centres (Center for Medicare and Medicaid Services' – CMS, Outcome and Assessment Information Set – OASIS - Shaughnessy et al, 2002 etc). Alternatively search agents may crawl the Web for relevant knowledge sources according to the core ontology (Varlamis et al, 2004).

Data acquisition and information extraction: This task encompasses the continuous extraction of data from the selected sources and the collective processing of extracted data to produce valuable information. Data can be easily fetched from websites that have explicit and constant structure with the use of wrappers. Tools that extract information from the web must be employed, e.g. intelligent search engines, intelligent agents, wrappers etc.

Information refinement to produce knowledge: Data mining tools will be used to produce knowledge from the collected information. Any data mining technique that is applicable in web data can be used to detect patterns in the data and infer new relations (Maedche &

Staab, 2000). The most useful among data mining applications is the induction of axioms and associations between concepts. They can be used to predict new trends, discover relations between seemingly unrelated concepts (drugs and diseases) etc. The results can be presented standalone to the user (i.e. an rule that matches an orthopaedical accessory with an injury), or can be incorporated to the ontology (as a relation between the accessory and the type of injury).

Knowledge preservation

Selection of knowledge schemata: The selection of the appropriate knowledge schemata is very important, since knowledge produced in the previous step will not be so useful if not combined or disseminated. The format selected for the codification of information should be: a) complete in order to maintain the abundance of produced information, b) scalable in order to provide information in different granularity levels, c) flexible so as to allow projection of knowledge to different aspects and d) extendible and adaptive to changes in the domain knowledge. A knowledge schema that satisfies the above criteria will have wide acceptance and will facilitate communication of knowledge within the healthcare domain.

Knowledge representation: An extended analysis on the use of ontologies for representing medical knowledge has been performed by Ceusters (Ceusters et al, 2005, Smith & Ceusters, 2003 and Ceusters et al, 2004). We suggest the use of open knowledge representation models such as RDF and DAML-OIL for the storage of knowledge emanating from the various healthcare information systems. These open models are the only solution for collecting the abundance and diversity of healthcare standards under a common knowledge umbrella. The requirements that should fulfill a Web Ontology Language are summarized in Heflin et al, 2002.

Knowledge accumulation: Data warehouses must be employed to store amassed knowledge. The continuous processing of the stored knowledge will allow the detection of changes and the update of the ontology that is explained in the following.

Knowledge updating

The accumulated knowledge must be evaluated in terms of usability and correctness. The produced knowledge (rules, concepts, etc.) must be compared to the existing knowledge. Comparison will reveal contradictions, modifications and possibly errors. Once the knowledge is validated it should be incorporated to the existing repository (Tzitzikas et al, 2001). Although, the knowledge update can be done automatically in certain cases (and this is a field of current research in ontologies), the intervention of human experts can be very helpful in this direction.

Generic architecture

The methodology presented above can be implemented in a knowledge management platform for the Healthcare domain. The platform comprises a suite of tools for information acquisition and processing, knowledge representation and management. The inputs of the platform are the core ontology and an assortment of information sources (content-rich web sites). Users may build their own simple ontologies using the appropriate tools. Custom ontologies are handled as personalized views of the domain that can be used as queries to restrict the acquired knowledge into the users' field of interest.

The outcome of the knowledge management process (Figure 1) will be dynamic ontologies in well specified engineering domains ("knowledge maps") that users can employ to locate information and improve their understanding on the structure and dynamics of a technology or business domain. Unlike a conventional search engine, the analysis brings knowledge into a context that improves the business efficiency. In addition to the benefits of the content map, users will be able to execute Boolean and advanced queries through an ontology-driven search engine. The knowledge classification scheme must use a meta-data approach based on XML (i.e. the Dublin Core schema) in order to facilitate discovery of electronic

resources. The metadata approach implies that produced knowledge will be available not only to human users, but also to automated tools and other knowledge management platforms.
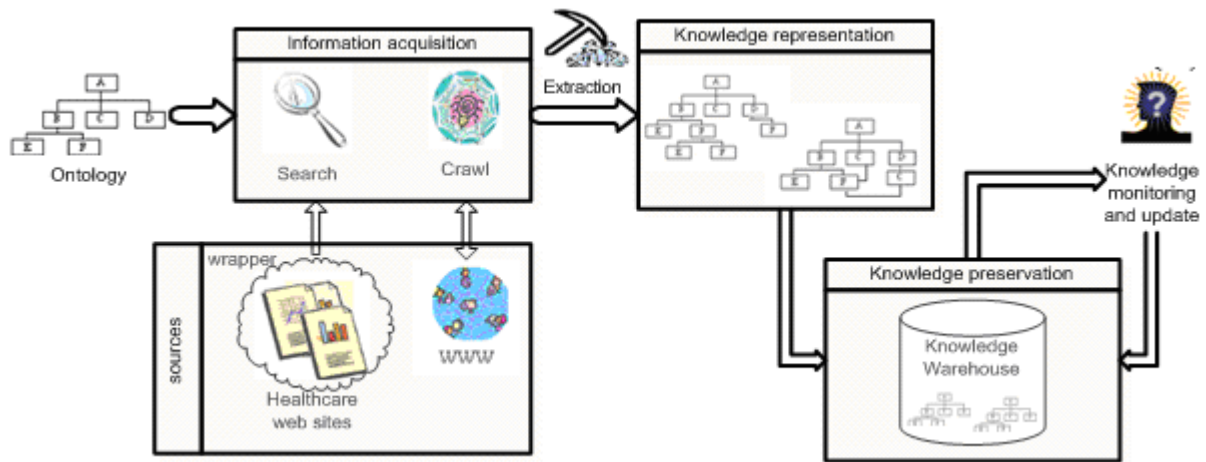

Figure 1. The knowledge management process

The advantage of the dynamic ontology compared to the core ontology is that its content and structure is under continuous change in order to follow the evolution of the domain. The information sources are associated to the domain concepts so users have an up-to-date overview of the domain and a rich informational repository on any specific subject at the same time. However, the management of concepts in an evolving ontology is a tedious task that must be carefully designed.

Usage Scenario

A typical user scenario (Figure 2) may start with the registration of healthcare practitioner, researcher or company to the knowledge management platform. The logged user may then use the ontology editor to create a new ontology. Even a simple taxonomy containing the concepts of interest to the user is adequate. The newly created ontology could be then found among the other ontologies of the logged user, together with the "core" ontologies offered by the platform. Alternatively, the user selects an existing ontology from the KMP and begins editing it. New nodes can be created, and for each node the user declares relations and also provides values for these relations. New instances for a node are also added and for each instance the user can either select one of the relations declared for the node or create new relations. When the ontology editing is finished, the user saves the ontology and presses the search button to perform the generic web search.

The retrieved results are saved in the KMPs database and are also displayed organised under the ontologies nodes and the user views the results that correspond to a specific ontology node by clicking on this node. These results are further analysed in a semantic way, using the ontology terms that characterize them.

Figure 2. Usage scenario of the Knowledge management platform.

Conclusions

Knowledge Management is concerned with acquiring, maintaining, and accessing knowledge of a domain. It aims to exploit the intellectual assets for greater productivity, new value, and increased competitiveness. Due to globalisation and the impact of the Internet, many healthcare organizations are increasingly geographically dispersed and organized around virtual teams. The use of ontologies will allow structural and semantic definitions of documents providing completely new possibilities: Intelligent search instead of keyword matching, query answering instead of information retrieval, document exchange between departments via ontology mappings, and definition of views on documents. The proposed methodology will lead to platforms that discover dynamic relationships between healthcare concepts (healthcare products and services, facilities and systems, technological achievements etc) and reveal emerging trends and new concepts. The latter increase wealth by better decision-making and provide social services by better professional placement.

References

Ashish, N., Knoblock, C.A., *Wrapper generation for semi-structured internet sources*. Workshop on Management of Semistructured Data, 1997.

Ceusters W, Smith B, Kumar A, Dhaen C., *Mistakes in Medical Ontologies: Where Do They Come From and How Can They Be Detected?*, in Pisanelli DM (ed) "Ontologies in Medicine. Proceedings of the Workshop on Medical Ontologies, Rome October 2003" IOS Press, Studies in Health Technology and Informatics, vol 102, 2004

Ceusters W. Smith B, De Moor G, *Ontology-Based Integration of Medical Coding Systems and Electronic Patient Records,* to appear in the 9th international congress of the European Federation for Medical Informatics (MIE), 2005.

Chawathe, S., García-Molina, H., Hammer, J., Ireland, K., Papakonstantinou, Y., Ullman, J., Widom, J,.*The TSIMMIS Project: Integration of Heterogeneous Information Sources*. In Proc. of IPSJ Conference, pp 7-18, 1994.

Connolly, D., Harmelen, F., Horrocks, I., McGuinness, D., Patel-Schneider, P., Stein,L.A., *DAML+OIL Reference Description*, http://www.w3.org/TR/daml+oil-reference, 2001

Hammer, J., Garcia-Molina, H., Cho, J., Aranha, R., Crespo, A., *Extracting semistructured information from the web*, Workshop on Management of Semistructured Data, 1997.

Hefflin, J., Volz, R., Dale, J., *Requirements for a Web Ontology Language*, W3C Working Draft 07 March 2002, http://www.w3.org/TR/2002/WD-webont-req-20020307/

Kushmerick, N., Weld, D., Doorenbos, R., *Wrapper induction for information extraction*, Proceeding of IJCAI, 1997.

Maedche, A., Staab, S., *Mining Ontologies from Text. Knowledge Acquisition, Modeling and Management*, 12th International Conference, EKAW, LNCS 1937, Springer 2000.

Resouce Description Framework, http://www.w3.org/RDF/

Brickley, D., Guha, R.V., "RDF Vocabulary Description Language 1.0: RDF Schema", http://www.w3.org/TR/rdf-schema/, 2004

Shaughnessy, P., Crisler, K., Hittle, D., Schenkler, R., Conway, K., West, L., Powell, M., & Richard, A. *OASIS and Outcome-based Quality Improvement in Home Health Care: Research and Demonstration Findings, Policy Implications, and Considerations for Future Change*. Center for Health Services Research, University of Colorado Health Sciences Center. 2002

Smith B, Ceusters W. *Towards Industrial-Strength Philosophy; How Analytical Ontology Can Help Medical Informatics*. Interdisciplinary Science Reviews, vol 28, 2003.

Tuttle, M., Brown, S., Campbell, K., Carter, J., Keck, K., Lincoln, M., Nelson, S., Stonebraker, M., *Semantic Web As "Perfection Seeking:" A View from Drug Terminology*, 2001.

Tzitzikas, Y., Spyratos, N., Constantopoulos,P., *Mediators over Ontology-based Information Sources*. 2nd Int. Conference on Web Information Systems Engineering, WISE 2001.

Varlamis, I., Vazirgiannis, M., Halkidi, M., Nguyen, B., THESUS, a Closer View on Web Content Management Enhanced with Link Semantics. IEEE TKDE, Vol 16(6), 2004.

Wang, A., Sable, J. H., & Spackman, KA., *The SNOMED Clinical Terms development process: refinement and analysis of content*. Proceedings of the American Medical Informatics Association Symposium, 2001.