

Semi-automatic data migration in a self-medication Knowledge-based system

Olivier Curé

ISIS Laboratory, Cité Descartes – 5, bld Descartes
Champs-sur-Marne – 77454 Marne-la-Vallée Cedex 2 – France
ocure@univ-mlv.fr

Abstract. Self-medication, defined as the act to treat oneself with or without drugs, is a common practice in industrial countries. A study of available computerized solutions in this field highlights that this issue has not been considered with enough attention, although they provide valuable services to both patients and health care organizations. This paper presents XIMSA, a self-medication knowledge-based system, which is supported by a database/ontology collaboration. This collaboration is guaranteed by DBOM, an application-independent system which enables the end-user to design, enrich and maintain an ontology from an existing database. DBOM's functionalities are presented within XIMSA's application domain.

1 Introduction

In order to become profitable and enterprise-like structures, health care organizations (henceforth HCO) need to provide services to all involved actors. Usually HCOs provide a large attention to health care professionals (physicians, pharmacists, etc.) but rarely concentrate their efforts on patients. Most of the time, this leads the patient to a semantic isolation whenever he is confronted with medical information, data and knowledge.

Our collaboration with the clinical pharmacology department at the Cochin hospital in Paris (France) has resulted in the implementation of IMSA (Interactive Multimedia for Auto-medication)[6]. This Knowledge-Based System (henceforth KBS) aims at providing information and services to the general-public on mild clinical signs, related treatments and medications. The latest version of this system, XIMSA (eXtended IMSA) bundles together a drug and symptom database, a self-medication ontology, a simplified patient electronic health record and an inference engine. The results provided by the inference engine depend on the ontology/database collaboration efficiency, which is undertaken by DBOM (DataBase Ontology Mapping), a domain independent application providing data integration and maintenance services in a Semantic Web environment [5].

This paper is organized as follows : section 2 presents the main characteristics of self-medication, section 3 focuses on XIMSA's architecture and functionalities, section 4 proposes an overview of DBOM, section 5 emphasizes a database/ontology

collaboration, section 6 concludes with a discussion on future extensions of the XIMSA and DBOM systems.

2 Self-medication

Self-medication can be defined as the health activities to treat oneself with or without drugs. People self-medicate using information obtained from past health experiences, books, advices, software, web sites, health advertising, radio or TV programs. On the medication side, people usually self prescribe drugs they already have at home and buy Over The Counter (OTC) products. These products are unevenly distributed over therapeutic classes (respiratory and digestive systems drugs are the most self-prescribed).

Self-medication is popular in most industrial countries, e.g. a recent study estimated that 91% of French citizens self-prescribe drugs when confronted to a known symptom [1]. The act to self-medicate is also an interesting financial market which represented 9.7% (2 billion euros) of the global pharmaceutical market for 1999 in France [1]. The French government, in its struggle with the French social security system deficit (14 billion euros for 2003), is indirectly encouraging self-medication through a series of actions : drug switch, lower the reimbursement rate for some drugs, favoring the emergence of the generics market, etc..

HCOs are as much aware of the semantic isolation of most patients as they are aware of the increasing success of self-medication. A logical correlation between these facts partially explain some alarming French figures for 2004: 128,000 hospitalizations due to drug interactions and 10,000 deaths due to drug over and mis-consumption.

The current policy of the French government is to encourage patients to become responsible and (pro)active health actors, but at the same time not much support, in terms of guidelines, books and computer tools, are proposed. Officially, the French healthcare system relies on its physician and pharmacist network to provide information on safe practice of self-medication. We believe that KBS is an alternative that has not been exploited with the proper attention. Such solutions may benefit from rapidly emerging markets, such as high speed Internet access and Internet compliant mobile phones, to reach an important portion of the population.

3 XIMSA

The XIMSA web application proposes self-medication services to the general public and aims to make this health care act a safer one and to free the patient from the semantic isolation related to medical information.

XIMSA' architecture is based on 4 distinctive modules (module interactions are presented in figure 1) :

- the XIMSA database stores symptom and drug data related to self-medication,
- the XIMSA ontology uses the OWL [8] syntax to represent terminological and assertional knowledge in a self-medication context.,
- the Simplified Electronic Health Record (SEHR) stores information (in an XML syntax) concerning data such as clinical antecedents and the history of drug consumption for a particular patient,

- the inference engine makes deduction with respect to the XIMSA ontology, the patient's SEHR and the data acquired during the navigation within XIMSA [6].

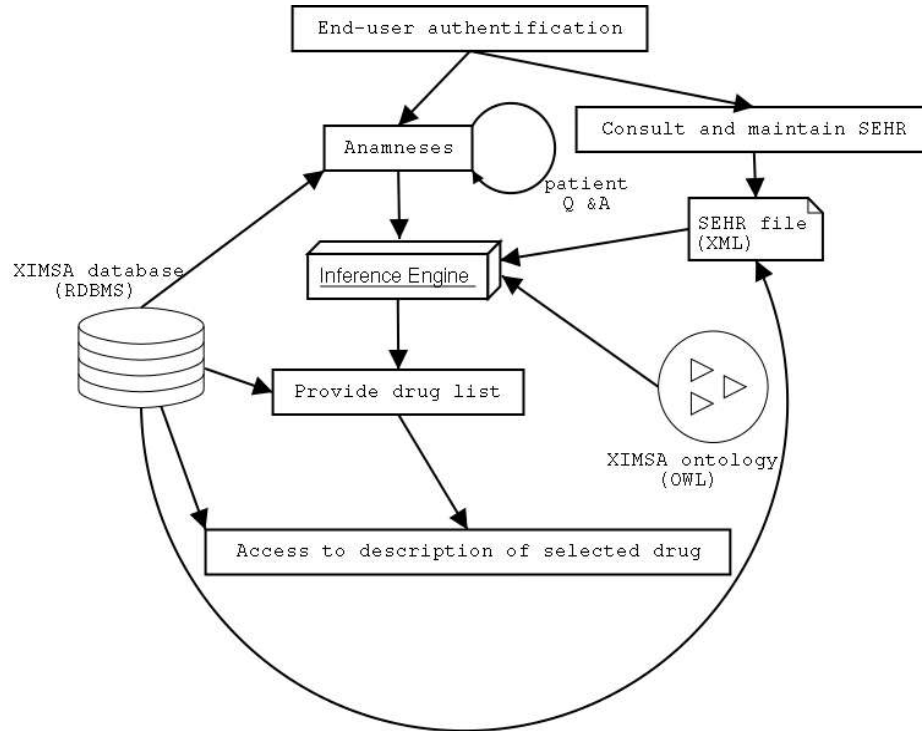


Figure 1 : Interaction-based architecture of XIMSA

The advantages provided by XIMSA's functionalities can be assessed in two categories. The first category is concerned with the quality of treatment which is ensured by :

- the confidence that the system provides advices and drug propositions only for mild clinical signs that can be treated via self-medication,
- the adequacy between the symptom described by the patient and the therapeutic classes proposed,
- the adaptability and accuracy of the drugs provided for a treatment, with respect to the clinical and pharmacological information stored in the SEHR of the patient,
- the value of the drugs presented based on an efficiency/tolerance ratio rating,
- the usage of such systems will increase the overall knowledge of end-users and will improve the communication between patients and healthcare professionals.

The second category is related to the controlling of costs :

- on the patient side, the system provides a direct access to OTC drug prices. Although not reimbursed by the Social Security system, these drugs may cost less to the patients due to their lower prices compared to prescription (partially reimbursed) drugs,
- on the Social Security system side, avoidance of reimbursement are guaranteed on the physician consultation and the drugs proposed.

- Finally, for both patients and the Social Security system, a global visibility of drug prices over all therapeutic classes may encourage the usage of the less expensive generics drugs.

4 DBOM

XIMSA's effectiveness is based on the quality and accuracy of the self-medication ontology and the SEHR. Although the patient/end-user is solely responsible for the value of the SEHR, the ontology's quality is undertaken by DBOM. An important part of this ontology focuses on drug related data, a field with a high update rate which requires storage in a database. Starting from the fact that "databases are similar to knowledge base because they are usually used to maintain models of some domain of discourse"[3], the idea of DBOM is to tackle the problem of database-to-ontology mapping.

This problem has been addressed by several research groups but DBOM's approach is more concerned with the following issues : data storage redundancy and inference efficiency. The consideration of these issues is done at the price of a non-automatic ontology design. For example, in [9] and [4], the goal is to (fully) automate data migration by transforming the relational database model into corresponding ontological structures, while [2]'s architecture is based on an existing ontology structure (non-automatic).

The approach adopted by DBOM is semi-automatic and involves the end-user to select amongst database components (relations, attributes, keys) which are going to map to the ontology structures. This solution ensures that the ontology will not contain concepts, properties and instances unnecessary to the inference engine. The vision of the DBOM system is to develop applications that use the ontology for inference purposes and is able to bind the inference results to the database thus providing valuable information to the end-user. In order to reach this goal, the designer of the mapping must be aware of the database schema and needs a clear vision of the characteristics of the implemented application, including inferences.

A high potential of the DBOM framework is to design domain and application ontologies [7] from existing databases. An important fact about databases are that update operations may also update the domain of discourse. Thus the designed ontology requires to be synchronized with the database. The proposed system offers such features and extends them to permit symmetrical maintenance solutions, meaning that controls are done both ways : from database to the ontology (ontology updating, e.g. adding new instances) and from the ontology to the database (e.g. consistency checking).

The main motivation behind the maintenance features remains in the database / ontology systems separation. This separation requires that the database schema is not modified during mapping processing and enables users of the DBOM framework to benefit from database features which are not available in ontology engineering (concurrency control, transaction, crash recovery, advanced storage techniques and query languages) as well as features of OWL ontologies, and underlying Description Logics properties and functionalities [3].

The description of the DBOM framework is divided into four distinct components: design, enrichment, ontology and database maintenances (see figure 2).

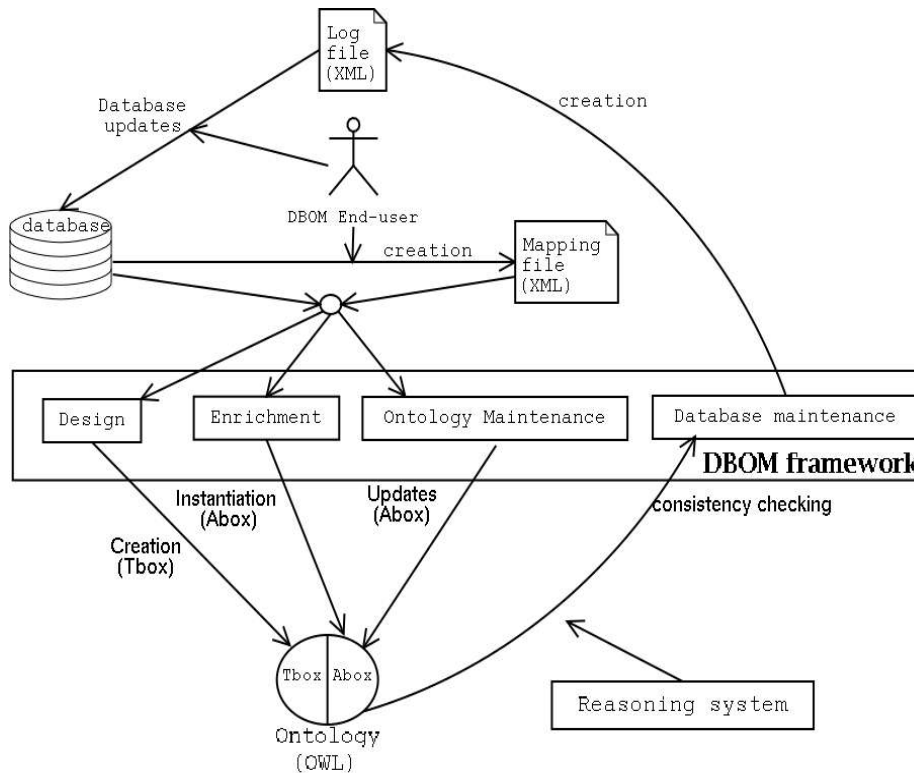


Fig. 2 : Global vision of interactions with the DBOM framework.

The design component supports the creation of the TBox (a set of terminological axioms) in an OWL syntax. This phase is partially done by the end-user who uses his knowledge of the database schema to describe a mapping file (an XML document) to the ontology. This description consists in describing the database relations, attributes and possible conditions implied in the design of the ontology. The approach adopted enables DBOM to generate hierarchies of complex classes and properties in the Tbox.

The enrichment component deals with instantiating the ABox (a set of assertional axioms) with individuals obtained from the database. The performances of this phase are increased due to the storage of SQL queries in the mapping file. A simple parsing of a mapping file enables queries to be executed in the database and thus creates new assertions in the ontology.

The maintenance components are dealing with ABox updates and database consistency checking. The ontology maintenance phase is concerned with (non-schema) updates of the databases, e.g. insertion of a new tuple in the database fires a trigger that may creates a corresponding instance in the ontology. The enrichment phase is usually performed once for an ontology schema while maintenance may be processed several times during the life cycle of a terminology. The last component of DBOM is related to the database maintenance which is ensured by consistency checking of an ABox w.r.t. a TBox. This database maintenance is executed after an effective, at least one trigger has been fired, ontology maintenance. The objective of this maintenance is to ensure that characteristics of new instances are consistent with

the semantics of the ontology, something the DBMS can not process due to its lack of detailed semantics. The current philosophy of the database maintenance is not to act directly on the database. Thus the approach adopted is to propose a log file, in an XML format, to the end-user.

5 Ontology/Database collaboration

The main features (design, enrichment and maintenances) of the DBOM application are used on the XIMSA system. An important part of XIMSA's self-medication ontology has been designed using DBOM based on a drug database which contains all drugs available in France. For each drug, the database regroups all the data of the Summary of Product Characteristics (SPC) plus extra information such as opinions from health care professionals and a drug rating.

The integration of the ontology in XIMSA enables the patient to control drug prescription regarding data contained in his SEHR which is created and maintained within the web-based XIMSA interface and enables the patient to store general (name, gender, date of birth, etc.) and health related (clinical and drug interactions, drugs being prescribed, etc.) information. An accurate and up-to-date SEHR assists the system in providing safe drug prescription to a specific patient. A ruled-based mechanism handles the ontology, the SEHR, and a particular request of the end-user (for example requiring an antitussive drug) to propose hyper links of safe to prescribe drugs. The result of the inference mechanisms provides identifiers of eligible ontology instances. Due to the correspondence between the ontology instances and the database tuples, the end-user can click on a hyper link and obtain all the information stored in the database about this drug, including information not contained in the ontology (e.g. drug price).

Finally, updates of the ontology produce a consistency checking. Lets consider the addition of a new drug in the database where the therapeutic class is not consistent with the Recommended International Non-proprietary Name (RINN, the active molecule) of the drug. Although the statement recorded in the database is valid, its semantic is wrong. The consistency checking of the ontology will report, in a log file, that an inconsistent statement has been added in the database. A study of the log file will enable the database administrator to change the therapeutic class of this drug.

6 Discussion

The efficiency of XIMSA relies on the quality and quantity of information stored in the SEHR. If the end-user/patient inputs sufficient data concerning clinical (e.g. whether he is suffering from certain diseases) and pharmacological (current consumptions, allergies, etc..) aspects then valuable inferences are provided.

In order to ensure an accurate and up-to-date SEHR, health care professionals may become sources of information. Both patients and physicians would benefit from this collaboration : the physician would be aware of all drugs taken by the patient and the patient would be ensured to have a valid, accurate SEHR concerning non self-medication descriptions and treatments. On the HCO side, such a distributed

collaboration may increase the overall quality of self-medication with the possibility to study, understand and control this medical activity.

The DBOM application enables the design and maintenance of high quality ontologies by providing correctness and minimally redundant data. The correctness quality is provided by the capture of the intuitions of domain experts which is facilitated by a conceptual-concerned collaboration with the designer. This collaboration also benefits from a knowledge representation language abstraction and fast access to a realistic, richly instantiated ABox. The minimal redundancy quality is provided by the database/ontology separation considering that the ontology only contains relations and attributes concerned with inference mechanisms.

A study of DBOM also emphasizes economical aspects with the following facts : the database and the ontology can concurrently be accessed and maintained, the knowledge acquisition and updates are done at no extra costs, the guarantee that the system will be adopted by the experts because they were involved in the design of the ontology.

Although the XIMSA auto-medication ontology contains more than 6000 drug products, 1500 RINN and 500 therapeutic classes, we believe that studies with larger ontologies, meaning larger TBoxes and ABoxes, need to be conducted. Performance surveys should also be conducted with such ontologies. The DBOM framework also requires the implementation of a graphical QBE-like solution for the design of database to ontology mapping file.

References

1. AFIPA : "Premier geste". Décembre 1999. On-line at http://www.afipa.org/index/informations/etudes_et_enquetes.html.
2. Astrova, I. : Reverse Engineering of Relational Databases to Ontologies. ESWS 2004: 327-341
3. Baader, F. ; Calvanese, D. ; McGuinness, D. ; Nardi, D. ; Patel-Schneider, P. (editors) : The description logic handbook. Cambridge University Press.
4. Barrasa, J. ; Corcho, O. ; Gómez-Pérez, A. : Fund finder : A case study of database-to-ontology mapping. Proceedings of the Semantic Integration Workshop (ISWC 2003). On-line at <http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-82/>.
5. Berners-Lee, T. ; Hendler, J. ; Lassile, O. : The Semantic Web. Scientific American, Vol. 5/01, 2001. (<http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html>).
6. Curé, O. : Overview of the IMSA project, a patient-oriented medical information system. Codata Journal, Vol 1/ 02, 2002, pp 66-75.
7. Guarino, N. : "Formal Ontology and Information Systems". In the Proceedings of FOIS 1998. Also in *Frontiers in Artificial Intelligence and Applications*, IOS-Press, Washington, DC, 1998.
8. OWL, Web Ontology Language Guide on-line at <http://www.w3.org/TR/owl-guide/>.
9. Stojanovic, L. ; Stojanovic, N. ; Volz, R. : Migrating data-intensive web sites into the Semantic Web. SAC 2002: 1100-1107.