ONTOLOGY BASED SEMANTIC REPRESENTATION OF THE REPORTS AND RESULTS IN A HOSPITAL INFORMATION SYSTEM

B. Prados-Suárez
Software Engineering Department. University of Granada, Spain

E. González Revuelta
Computer Science Department. Torrecárdenas Hospital, Almería, Spain

C. Peña Yáñez, G. Carmona Martínez
Computer Science Department. San Cecilio Hospital, Granada, Spain

C. Molina Fernández
Software Engineering Department. University of Jaén, Spain

Keywords: Health Ontologies, Hospital knowledge representation, Semantic organization of results, Hospital Information Systems, Semantic retrieval of hospital reports.

Abstract: The main purpose of this paper is to contextualize the access to the huge amount of results and reports that a Hospital Information System (HIS) can reach to have. Our target is to integrate a semantic layer with the HIS, so that the user can employ this layer to access just the precise information needed, under his working context. Here we propose a new navigation system based on the semantic characteristics of the data acceded, their complementary characteristics, properties and relations, providing so the HIS with a new tool to solve an evident problem for the users.

1 INTRODUCTION

Nowadays a characteristic of Hospital Information Systems (HIS) is the huge amount of data they manage, generate and store, as well as their wide variety and typology. Among these data *reports* or *outputs* can be found, which are the direct result of concrete queries, as well as *results*, that are more elaborated data obtained from statistical processes or a basic exploitation of the information in the DB, using the tools offered by the applications of the concrete area or department Prados M., Peña M.C. 2003. Examples of reports are the list of patients waiting for a given surgery or the list of medicines prescribed by each medical speciality; while the average duration of the hospital stay or the efficiency indicators in each medical speciality are examples of results. The set of all these documents and data is what we call ***Universe of Result Reports (URR)***.

A serious inconvenience is that the access to these data must be done by a complicated browse though the menus of different applications, which gives rise to severe problems like:

- There is so much information available that the user can not find what is looking for.
- There are a number of reports and outputs, that are wrongly controlled, and these are sometimes useless.
- There is no proper organization of the information, making quite difficult the retrieval.
- There is also a great amount of redundant results that, in addition, can be acceded from different applications.

Prados-Suárez B., González Revuelta E., Peña Yáñez C., Carmona Martínez G. and Molina Fernández C. (2008).
ONTOLOGY BASED SEMANTIC REPRESENTATION OF THE REPORTS AND RESULTS IN A HOSPITAL INFORMATION SYSTEM.
In *Proceedings of the Tenth International Conference on Enterprise Information Systems - ISAS*, pages 300-305
DOI: 10.5220/0001711003000305

- The user doesn't know the structure of the results, and hence what can he obtain from them and from where (which application) must do it.
- The user obtains something that is not what he/she really needs, or even can not find it.
- The models used to exploit the information changes along the time, as well as the criteria, purposes and requirements.

To solve these problems it would be necessary to define logical structures to organize the URR, designing it under semantic criteria. In this way new navigation systems can be designed to allow the user to focus or widen the information queries Lambrix P., Habbouche M. and Perez M. (2003).

This is the target we face in this paper: *The analysis and design of a system that enables the user to obtain accurate and appropriate information, as well as to create fine and useful results user interface, with as much knowledge as possible.*

Our proposal to face this task is to establish a conceptualization of the URR, by means of Ontologies.

2 BACKGROUND

It is well known that Ontologies are conceptualizations with a variety of utilities, among which we remark two: on one hand, they are quite helpful in decision making tasks like diagnostic decision in Medicine Lambrix P., Habbouche M. and Perez M. (2003); on the other hand, they are useful in Document Retrieval Systems and Semantic Information Retrieval Systems, where information is acceded by its meaning and not just from a given terminological description Berners-Lee, T., Hendler, J.and Lassila, O., (2001).

Today there is a wide variety of applications of Ontologies within specialized fields, like "Health Systems", especially for medical terminology analysis and the extraction of its semantic characteristics Angelova, Galia, (2000). Within this framework, is especially significant their use on "query expansion", to detect the kernel of a user's query and extend it with variants, finding the documents with most relevant information Strzalkowski, Tomek; Lin, Fang; Wang, Jin;and Pérez-Carballo, José (1999). In addition, Ontologies have been used to unify the medical language in multilingual frameworks Scope Web site (2002), and for semantic categorization of medical diagnosis terminology Prados Reyes, M. Peña Yañez C., Vila Miranda A. and Prados-Suarez B., (2006).

However, Ontologies themselves can also be directly useful for the user as a tool to integrate data and applications, offering the possibility to know conceptualization itself, and use it to improve the information access. This is the framework where we present our proposal.

3 SUPPORT MATERIAL

To develop our proposal we have the HIS of the Hospital San Cecilio, in Granada. This system covers the traditional areas of every HIS, and has an extensive catalogue of results and reports, whose contents are structured according to their membership to HIS applications.

This catalogue (our URR), can be assumed as a "catalogue document", where each document is a report that, as every document, can be identified by a set of formal and semantic characteristics. The contents in the URR catalogue have different types, (from statistical reports to reports relating attributes from the data model), but they always have a sample selection.

The representation of these data and their characteristics gives raise to a Knowledge Base, designed according to an Ontology, that will be the support to develop the *Semantic Retrieval System of the universe URR*.

4 HIS STRUCTURE

The actually implemented HIS covers the Economical-administrative and Logistical- assistance fields, managing the Operational Systems of the Hospital Organization, and hence its activity.

Functionally it is structured in several levels, as indicated in Figure 1.

5 METHODOLOGY

To develop the system proposed in this paper we have followed the next steps:

1- **Determine the Type of Universe to Study**: Most of the elements in our URR have small formal differences, like the presentation order of the results, or the grouping criteria. Hence, under a documentary point of view they are different, while from the semantic point of view they have the same value. It lead us to represent the formal differences as

properties in the Ontology, or inside the executed procedure, being more relevant the semantic value.

2- **Definition of the Types of URR Formal Characteristics and the Semantic Valuation Criteria**: The main purpose of this stage is to find which components are interesting and allow the user to access the URR, focusing or widening the queries. The elements found in this stage will have a correspondence in the Ontology design.

3- **Design and Development of the Ontology**: Here we have structured the components according to semantic characteristics, hierarchies of classes and relations between them, according to the universe URR. Hence we focus exclusively on the Operational System, avoiding the inclusion of classes, relations, or properties that don't exist in the operational system, though they would have logical sense.

4- **Ontology Representation**: In this last stage we have implemented the Ontology using the tool *Protégé*, following a representation through conceptual scheme.

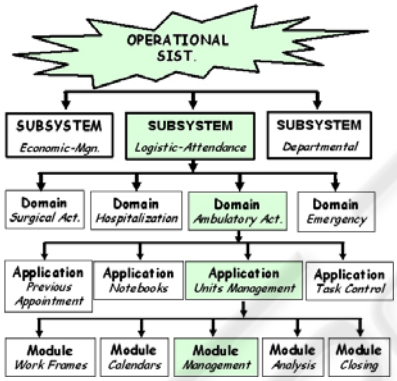

Figure 1: Functional structure levels of HIS.

6 THE UNIVERSE TO STUDY

Analyzing the URR we have found the following possible contents:

- Attributes from the data model tables.
- Attributes generated in the executed procedure, as partial or total results.
- Statistical calculations of different types.

Under a functional point of view, these reports belong to a concrete application of the HIS, and are generated by procedures included on that application as modules or tasks.

6.1 Formal Analysis of the URR

Formally a document from the URR has the following characteristics:

Type: This property can have one of the next values: "query relation report", "operative recount" or "statistical report".

Proprietary Application: This characteristic references the functional module from the HIS logical design that manages the information and procedures related to the concrete target document. This characteristic follows the logical levels structure of the HIS, shown in bold letter in Figure 1 (System – Subsystem – Functional Scope – Application – Module – Procedure). Its values are, as an example, "absenteeism" *Module*, from the "Human Resources Management" *Application*, as well as those in capital letter in Figure 1.

Computing Procedure that Generates (executes) the Document: This attribute is referred to the physical unit from the physical design that generates the target document.

Domain of the Showed Attributes: This is the set of data from the Operational System that are showed in the document.

Generated Data: References the total results or statistical values computed on the execution process.

Query Descriptors: Are those that allow the user to filter the query for a particular document. Usually they are very limited set like date, functional unit, diagnosis, etc., and can be structured according to their meaning.

6.2 Semantic Analysis of the URR

The semantic analysis of the URR offers a perspective about its meaning from a given point of view of interest, like the "business management" perspective from the point of view of the "administrative management".

This task acquires maximum interest if we consider that we have a logical structure of the knowledge, showed in Figure 2, based on several levels (*contexts*, *scenarios*, *images*, *views*), which makes possible the semantic *pertinency* of a given document to a concrete image or view Prados M., Peña M.C., Prados M.B. and Garrido J.M. (2004). As an example, a report about the occupation of the operating rooms is related (*pertinent*) to the "Efficiency of operating rooms use" *image*, which belongs to the "Surgical activity" *scenario* from the "*Medical Care*" *context*.

In addition to the *pertinency*, there may be other characteristics with semantic value like:

Orientation: That determines the type of user to which the document is directed: Medical directors, Management, Central Services…
Confidentiality level: With three possible values: low, medium and high.

7 ONTOLOGY DESIGN

The aim of the Ontology design is to elaborate a proposal according to the existing HIS, but opened to the incorporation of new elements to the URR, that could give raise to new "main classes" (defined later) or instances.

The basic criteria that must be verified are the *hierarchization* (ordering), *modularization* (concept isolation), *abstraction*, *clarity* and *coherency*. Considering it, we have defined the following elements for the design:

- **Domain:** Set of reports and results (URR) that can be obtained from the Hospital Information System. Hence, the concepts or Domain elements will be the Results, corresponding to reports/outputs/lists issued from the variety of information sources, on user demand.
- **Classes:** Each class corresponds to one of the formal or semantic characteristics defined on the URR. Here we define the concept of "*main class*" to mean that the instances of that type of class are elements from the URR, and we establish as *main class* the *semantic categorization*.

This way, as an example we say that an "extern emergency assistance delay report" is an instance of the class "Emergency efficiency", which is a subclass of "Emergency assistance activity", instance at the same time of "Assistance activity" that is a subclass of the "Operational context".

- **Properties: A**ssigned to each element or class of the URR domain, represent the set of characteristics of a given element or set of elements. There are two types of properties: *complex properties*, defined as relations between ontology classes; and *explicit properties*, defined exclusively for a given class or instance, without generating a new class.
- **Instances:** the elements in the URR are assigned to instances of the semantic categorization. They will be represented by the naming of the physical unit of the HIS. There will also be instances corresponding to the individual terminal elements of the class hierarchies, and will be defined in the Ontology as "***direct classes***".

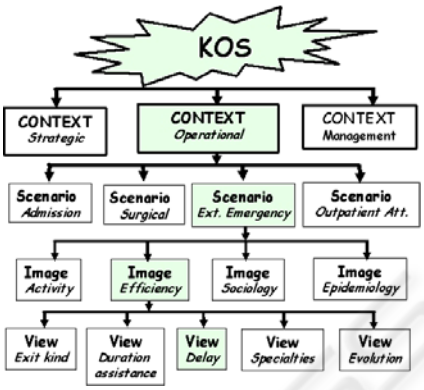

Figure 1: Semantic Categorization scheme. Example: Operational Context Scenarios.

7.1 Identification of Ontology Classes

The main conceptual classes defined in the Ontology are:

Semantic Categorization: Represents the cognitive focus approached by a URR document, and obeys to user criteria formalized following the levels schema in Figure 2. Its instances are the URR documents expressed as a code according to the physical unit or executable computing procedure.
Logical Model Functions: This class represents the organization activities, computerized and structured according to hierarchies of functional requirements. An approach to this class hierarchization is shown in Figure 1. As an example, the control of the "temporary working inability" is a member of "Absenteeism control", member at the same time of "Human resources management", member again of "Administrative management", that belongs to the Economic-Administrative subsystem.
Receiver: Final destination of the document. We have planned three types: "Local", "Extern" and "Publication". The class "Local" generates a hierarchic line following the organic structure of the Hospital government. The "Extern" class represents different institutions that are users of the hospital information, while "Publication" refers different ways to announce the information (web pages, notice board…).
Chronology: Establishes the "periodical" (and period) or "on demand" character.
Confidentiality Level: Unavoidably associated to every hospital document.
Query Strategy: Represents the dimensions through which the user can select or filter the content of a

document It is useful to inform the user about the possibilities to focus the query.
Leaning to decision: Indicates the interest of the document under a strategic point of view, like periodic control, evolution or research.
Semantic Valoration of the Contents: It refers to the set of resulting information in a URR document. Concretely, it refers the physical data model, and the structure and contents of the data tables themselves.
Hierarchies in this class are identified according to three levels: "application", "table type" (historical, movements, master, and base) and "data type" (own attribute, situation attribute, movement attribute).

7.2 Properties Definition

The aim of the properties is to enrich the conceptual structure of the Ontology, to properly define URR documents according to their formal and semantic characteristics. We have two types of properties:

Complex Properties: Are an integral part of the Ontology, and are defined as relations between classes. As an example, a report of the "Attended emergencies", which is a member of "Emergency frequentiation", has the properties of being *confidential*, of *local* use for *Medical management*, can be selected *chronologically, functionally or* by its *characters*, and show the data in the *table* "Emergency movements".
Explicit Properties: Represent qualities of one or several classes, and are useful only when they are assigned to a few classes since, otherwise they would finally become a new class. This type of properties is reserved for very specific characteristics of a class or instance, like a document which totals some concrete values according to a specific calculation. These properties can be easily added and removed to enable or disallow some given characteristic.

7.3 Ontology Generation with PROTÉGÉ

Once made the intellectual tasks to define the Ontology, the use of the tool Protégé Natasha Noy and Samson Tu, (2003) has made possible to easily obtain the conceptual scheme, the properties and characteristics, with the advantage of the possibility of generate a model in XML and RDF. Since this tool is quite useful for maintenance tasks it also offers the possibility of facing the problem in an increasing way.

8 USE OF THE ONTOLOGY IN RETRIEVAL TASKS

As previously mentioned our aim is to enable the access through the user's concepts, to improve the accessibility to the information without having to know the technical or functional structure of the system.

Since every document in the URR is marked by the Ontology according to a set of formal or semantic properties, it can be indexed by a set of descriptors, as in any Documentary System Gil Leyva, I. and Rodríguez Muñoz, J.V. (1996). With it, we can offer the user different access ways:
Simple Documentary Query*:* The user formulates the query equation from which the "query kernel" is obtained, by "query expansion", as a set of descriptors of the concepts in the Ontology. From this query, the reports indexed by this set are obtained Scope Web site (2002).
Focusing by Semantic Categorization*:* It locates the user in the cognitive field and selects the reports pertinent to the target study. It is performed by accessing the class "semantic categorization", whose navigation allows the centre the query and know the properties related to each level or instance.
Ontology Navigation: The user gets into the Ontology structure, knowing its classes, and from them focus or widen the query field or analyze the properties of a given class up to reach the instance level that is the target of the selection.
Query Contextualization: Quite similar to a documentary query, but by descriptors in the knowledge base, returning to the user the context by means of the hierarchical relations and the properties defined in the Ontology, as well as the documents pertinent to that context. Among them, the user will select those interesting for him/her. As can be seen, this is a semantic query procedure.

9 CONCLUSIONS: IMPLEMENTABLE DATA MODEL

The design proposed in this paper is a conceptualization of the URR obtained from the HIS. This design satisfies the needs posed by the habitual users of the HIS, offering them an interface that makes easier the access to the URR contents, allowing the user to know a priori what can be interesting for a concrete study, the characteristics

and limitations of the contents and other documents that can be useful.

Hence, the proposed design has a mainly practical character, and represent a practical application of the Ontologies, restricted but useful to easily solve a problem that can be generalized and extended to other fields, especially in the hospital framework.

The structural richness of the Ontology is conditioned by the characteristics and extent of the HIS, but we think that it is secondary, since the update, maintenance and extent of the Ontology and its data model is easy to bring up to date.

The process to implement the model has been cyclic and iterative: on an initial model and implementation we have iteratively discover relations or classes, those have modified the conceptual model and have unchained the pertinent modifications cascade.

ACKNOWLEDGEMENTS

This work was partially supported by the Junta de Andalucía, under Project P07-TIC 03175.

REFERENCES

Prados M. and Peña M.C. (2003), Sistemas de Información hospitalarios. Organización y gestión de Proyectos. EASP (Escuela Andaluza de Salud Pública), Granada, 1nd edition.

Lambrix P., Habbouche M. and Perez M. (2003). Evaluation of ontology development tools for bioinformatics. Bioinformatic. Vol 19, n 12, pag 1564-1571.

Berners-Lee, T., Hendler, J.and Lassila, O., (2001) The Semantic Web, Scientific American, 284(5):34–43

Angelova, Galia, (2000) Ontologies for Natural Language Processing Applications. In: Simov, Viril; Kiryakov, Atanas (eds.) Ontologies and Lexical Knowledge. Sofia: OntoText Lab, Sirma AI EOOD, p. 1-15.

Strzalkowski, Tomek; Lin, Fang; Wang, Jin;and Pérez-Carballo, José (1999) Evaluating Natural Language Processing Techniques in Information Retrieval. In: Strzalkowski, Tomek(ed.) (1999) Natural Language Information Retrieval. Dordrecht: Kluwer Academic Publishers, p. 113-145. ISBN 0-7923-5685-3.

Scope Web site (2002), http://tecn.upf.es/scope.

Prados Reyes, M. Peña Yañez C., Vila Miranda A. and Prados-Suarez B., (2006) Generation and Use of one Ontology for Intelligent Information Retrieval from Electronic Record Histories. In Proceedings of the 8th International Conference on Enterprise Information Systems. pp 565– 571, Ed. INSTICC, Paphos, Chipre. May 23-27.

Prados M., Peña M.C., Prados M.B. and Garrido J.M. (2004), Gestión de conocimiento en el ámbito hospitalario. EASP (Escuela Andaluza de Salud Pública), Granada, 1nd edition.

Natasha Noy and Samson Tu, (2003) Developing Medical Informatics Ontologies using Protégé, (Tutorial materials from the AMIA 2003 Protégé tutorial). In AMIA Symposium

Gil Leyva, I. and Rodríguez Muñoz, J.V. (1996) Tendencias en los Sistemas de Indización Automática. Estudio evolutivo. Revista Española de Documentación Científica, vol 19, nr 3 p 273-291.