USING COOPERATIVES ONTOLOGIES FOR THE CUSTOMIZATION OF HYBRID MEDIATOR INTERROGATION PROCESS

Cherrat LOUBNA
Abdelmalek Essaadi Universit
Faculty of Science, LaSIT Laboratory Morocco
e-mail: cherrat@uae.ma
Ezziyyani MOSTAFA
Abdelmalek Essaadi Universit
Faculty of Science, LaSIT Laboratory Morocco
e-mail: ezziyyani@uae.ma
Mohammed ESSAAIDI
Abdelmalek Essaadi Universit
Faculty of Science, LaSIT Laboratory Morocco
e-mail: essaaidi@uae.ma
Omar KASSARA
Abdelmalek Essaadi Universit
Faculty of Science, LaSIT Laboratory Morocco
e-mail: kassara@uae.ma
Bennouna MUSTAPHA
Abdelmalek Essaadi Universit
Faculty of Science, LaSIT Laboratory Morocco
e-mail: bennouna@uae.ma

Abstract

The explosion of information sources accessible via the Web created the need for mediation tools between users and heterogeneous information sources on the Web. However, the interface design of these mediators, with taking into account the wide variety of skills and knowledge of users, and the need for them to share their preferences, requires systems able to guide the user through the interrogation process. In this paper, we try to raise this challenge by proposing, on one hand, the use of a cooperative ontology's models, which adapt with the evolution of user's profiles and the dynamic change of integrated sources, and secondly ensure sharing the knowledge between users which will facilitate analysis of information and improve the data quality research process in space and time.

Key-words: *ontology, mediator, data warehouse, hybrid, profile, focus, PMD, health, medical domain, patient, AXMed, interoperability, security*

JEL Classification: C_{63}

Introduction

The rapid development experienced by the autonomous distributed information systems, and the emergence of the Web networks enable users to access an increasingly growing and highly heterogeneous data that does not necessarily meet their expectations for their needs and preferences or for the level

of quality information. The design and the development of a system that is both flexible and efficient based on mediator/adapter's architecture is necessary (Cali and al., 2003). The purpose of such a system is to intercept users requests and locate data and services which are the most appropriate to meet the demand of the client, to pass parameters, to invoke the service and to return the result in a transparent manner to users's via the adapters involved. The mediator thus provides a centralized sources view and adapters provide uniform access to sources (Kostadinov and al., 2005).

In another aspect, the concept of customization (Bouzeghoub and al., 2005) of the information access process improves the quality of information obtained through a mediator and will give the user adequate information to their preferences, interests, or more generally to their profile. In this paper, we discuss a technique for building the user's profile as part of a statistical approach using the user's behaviour as a source for predicting the implicit model. This technique is based specifically on the interaction between the profile dimensions represented by historical research sources integrated by the mediator and the user's interests. In this context, profiles and the interest can be much richer, because the sources are heterogeneous, with application areas that are complementary.

A possible improvement consists in incorporating the concept of ontology in the integration and interrogation process via the mediator (Dung Nguyen, 2004). Indeed, the creation and use of ontologies would evaluate the data meaning relatively to the exploitation by users. Thus, it will be possible to link the operations of extracting and presenting data based on users' profiles. Another contribution of the use of ontology would be a better evaluation and optimization of results. In the following sections, we develop in detail the concept of ontologies development and exploitation in the process of interaction and interrogation based on a classification of ontologies.

Application field and proposed integration scenario

To demonstrate the practical value of our methodology and bring up all its facets, we will apply it in the medical and health field where the treated data are of a great sensitivity, especially with the existence of a multitude sources specific to a medical fields and the difficulty of evaluating the reliability of medical information. But we point out that although the case study concerns the medical field, the concerns about the customization of integration and querying process of mediators using cooperatives ontologies are neither dedicated nor specific to this sector; methodology and the proposed solutions are applicable to a wide range of applications. We use throughout this article, an application example, relied on a medical application that tracks the different treatments of patients medical document (PMD): The PMD is a computerized file that contains information about the care was administered to a patient. It aims to facilitate patient access to their health data and to encourage information sharing and cooperation among health professionals.

In the practice of medicine, the exchange of data and knowledge in relation to a patient's medical document, among the different categories of users, creates a problem of dysfunction and disagreement in the treatment of medical document and analysis results. Indeed, the data treated by the medical systems are often used with more than one meaning in the same domain, as different conceptualizations must be specified and different names must be given to concepts in order to distinguish their meanings.

To address these problems, medical ontologies (Charlet, 2007) define a way to achieve efficient to store and communicate knowledge and general medical information about the patient. These ontologies have been appropriate to the available support for archiving, processing and transmission of knowledge, such as demand for the reuse and sharing of patient data.

We identify the first critical question to ask: How can we determine in an automatic, selective and optimized manner the certified medical information resources and have them interrogate through a specific system in order to satisfy all medical aspects for any questioning of potential users?

Hybrid mediator

In response to this question, we have proposed the realization of a hybrid integration system, combining both physical and virtual integration of data, capable of integrating multiple data sources specific to patients. These data can be used by patients, administrators, specialists and doctors. The architecture we developed is distributed across multiple data warehouses. Each warehouse (Hamdoun and al., 2007) (Hammer and al., 1995) manages the data in a specific region (analysis, consultation Radiology, etc.). The data for each warehouse are extracted from existing health systems in the region to build the patient's medical document. The user can also access the medical information resources on the web to complete its request. That part of interrogation is accomplished through a set of virtual mediators. Each mediator manages a set of sources whose theme is close. The hybrid mediator handles the task of scheduling the execution of sub-queries of the initial request between the data warehouse and the mediator. To automatically exploit the distributed resources among different data warehouses and mediators, the definition of a meta-ontology at the level of the hybrid mediator, an ontology of domain at the level of each mediator and an ontology for the data warehouse is necessary. Indeed, the definition and the use of these data by multiple types of users (Doctors, Patients, etc.) gives different semantic interpretations, and to harmonize the exploit of these data, the use of ontology concept is essential. The main purpose of this structuring is to optimize the interrogation phase in space and time.

Figure 1 shows the architecture of hybrid integration system. The core of the hybrid mediator is based on the infrastructure of AXMed mediator that has been developed in the LaSIT Laboratory (Ezziyyani and al., 2005).

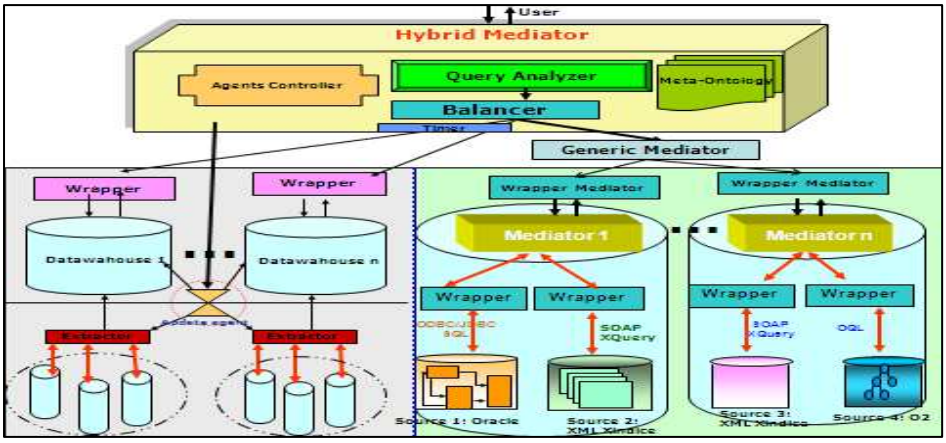

Fig. 1. *Architecture of the hybrid system of mediation*

Ontologies classification

In the context of the mediation system that we propose, ontologies have a threefold objective to represent in a generic and reusable manner a semantic of a domain.

On one hand, through a shared ontology, they provide a common understanding used to share knowledge or data from multiple resources to ensure the interoperability of information systems. They also allow a precise description of data sources using the standard concepts of domain, which helps prevent errors in interpretation because all sources have a common semantics described by the shared ontology (domainontology).

On the other hand, using the concepts of an ontology of profiles to describe the interest, as meta-information to ensure understanding and dynamic adaptation of requests. They provide a dynamic understanding of the data relating to user profiles (Kindo and al., 1997). The integration methodologies unify the views of different users by an ontology that is consistent with changes in the semantics of data based on user profiles and exploitation mode.

Besides the two goals mentioned above, ontologies can also customize and optimize the query process based on the semantics of the query. The semantics of the query is interpreted according to the contents of the queries ontology, which can limit the exploration of data sources to those who have information whose semantic corresponds to the query.

Finally, to ensure the knowledge sharing between the three ontologies mentioned before, we use a meta-ontology for describing the semantic link between different ontologies and the mapping between local ontologies and the global ontology.

Subsequently, we focus on the description of the different ontologies with the deployment mode.

Shared Ontology

The global ontology allows describing the application domain by providing a shared vocabulary between different sources integrated by the mediator. The semantic of each source to be integrated is described by its own ontology and the different ontologies are connected together by a shared common vocabulary (Visser and al., 1999). The space usage of ontologies is divided into two parts: the communication between people with varying points of view and different needs, and interoperability between users who need to exchange data and tools. The ability of these ontologies to share and reuse knowledge is exploited in the construction and use of knowledge-based systems. The usefulness of shared ontologies in the domain of data integration is that they establish semantic links between different elements of the sources (Jellouli and al., 2008). They can also serve as a model for querying the integrated system when used to describe the global Schema.

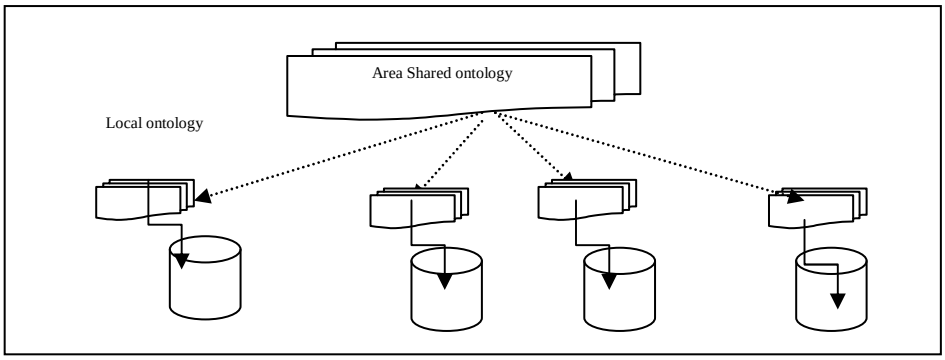

Fig. 2. *Shared ontology architecture*

Profile ontology

The information personalization is a major challenge for the computer industry. The relevance of the information provided, its intelligibility and its adaptation to user preferences and usage are a key factors for success or rejection of such information systems. For this, we focused on the modelling of user preferences in information research area in several heterogeneous resources integrated via mediators. So for the definition of an architecture for the user profile (Amato and al.) taking into account the dimension of focus based on historical interrogation, then the integration of this profile into an information search process to customize the results returned by the mediation system. The use of profile ontologies allows us to:

– facilitate access to multiple information sources;

– exchange information to improve, and participate in a collaborative process of distributed resolution;

– distribute and balance the workload.

Hierarchical organization of interests

The public or shared interest is common to all users, but can be customized dynamically as and when users query the mediation system forming appropriate interest centres to different groups of the system. Indeed, the public interest concerns a specific area (research or not), provided by experts, and will be used by the system for the construction of users interests classified by categories. These will include knowledge and data shared among a users group of the same mode of operation of the system concerned.

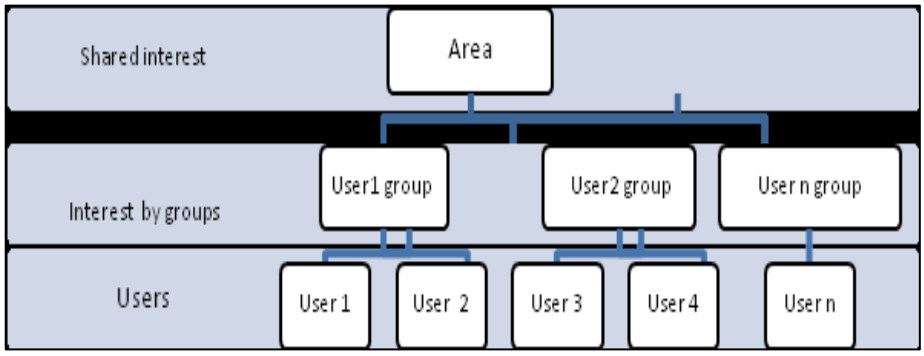

Fig. 3. *Hierarchy of interests*

In our application context, we are particularly interested in site users to manage online PMD. We identified three groups among those users for which our approach can improve the effectiveness of their work online: administrative, doctors and patients. The main common point of these users is ignorance of the accurate purposes of navigation when you start it or the composition of the original document of the PMD.

It is therefore necessary to compose the patient's medical document not only in own way to each user, but also as and when the navigation of that user. To do this, the use of cooperative medical ontologies by reusing the techniques from personalization, customization and adaptive query allow optimizing resources and improving the query process of mediation systems.

Interest dimensions

The interest is to consolidate in a knowledge base the personalized research results in relation with the use profile. This knowledge is used by the monitoring tool for communicating the news of sources on a subject or in a specific area. The

interest is therefore a continuous updating of knowledge of a user or group of users vis-à-vis the data sources integrated by the mediator.

In order to evaluate the degree of change contexts related to the user interests, we describe a context with an adaptive ontology where the focus is described by four dimensions. The first represents the history of user interactions with the mediator. The second characterized its recurrent needs as information and is inferred and evolves from the first dimension. The third defines the security exploitation of interfaces and access to data sources. Finally, the fourth includes knowledge in the semantics, data and functionalities offered by the mediation system.

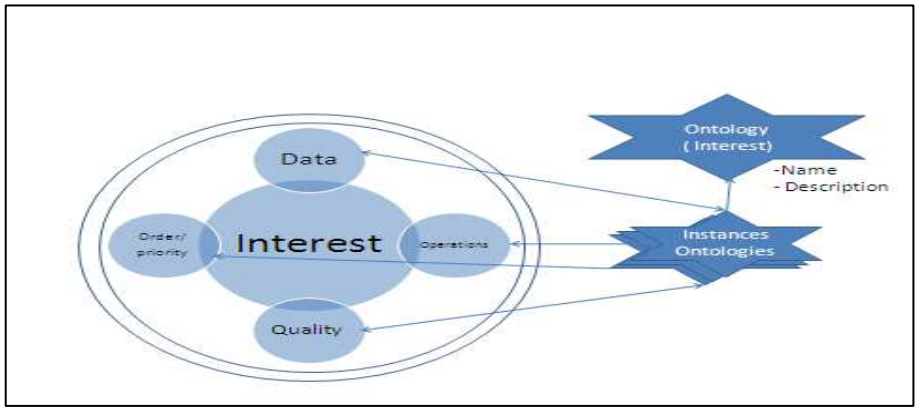

Fig. 4. *The interest*

Below is a class diagram of the focus that implements relationships between domains and users.

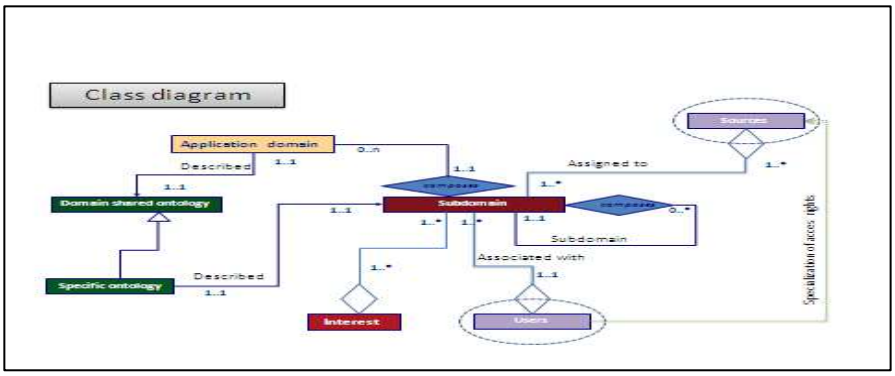

Fig. 5. *Class diagram of the interest*

Queries ontology

For this classification of ontologies, we were interested also in the consideration of users profiles in the interrogation process whose goal is to improve the relevance in terms of research results in various data sources integrated via the mediator. The solution we provide is based on the use of

ontologies to represent knowledge of the research area covered by queries and user profiles on the selected area. In this context, the query ontology can retrieve the desired information effectively and efficiently from a large collection of data. Especially for mediation systems, where the query is more complex and the answer is not correct. Mediation systems have the function to allow users to access heterogeneous and distributed data that contribute to a unifying response to user queries. Certainly, ontologies will improve the quality of access to this information. However, the overabundance of information and its wide availability has led to a deterioration in their performance both efficacy and efficiency. Indeed, as and when the global response is generating and the interviewed sources are diversifying, the mediation system delivers massive results in response to user queries. In that case, the user is confronted with an information overload in which it is difficult to distinguish pertinent information from secondary information, and even noise. In addition, the evaluation of a query is usually without considering the context and / or specific needs of the user who issued it. The same request made by two different users, produces the same results even if these users have different expectations.

To address this problem, firstly, the use of an evolutionary model of ontology and an adaptation process will allow ontologies to adapt to changing knowledge of an area by a user group. Then, the exploitation of this type of ontology to improve the process of interrogation by an incremental optimization algorithm (Figure 7). The parameters of this algorithm are the different structures and separate data sources, and the degree of importance of these entities to extract pertinent data.

Indeed, our ontology model is based on a oriented graph of dependence between entities of data sources integrated by the mediator. The graph defines an execution order of sub-queries after the phase of rewriting the query. Each node in the graph represents an entity of the global ontology of the area. The ontological approach is adopted in the modelling process of collecting data in a specific area of interest from the point of view of a user who queries the mediator in order to retrieve results from multiple sources. The process can be programmed to automatically update the facts of the ontology and optimize the path of execution graph by a user request.

The question that arises at this stage is how do we define an order of execution of sub-queries that is consistent with the query posed and provides pertinent and satisfactory response to the user?

To answer the question we take this example:

Let us suppose a doctor poses the following query: R: “I want to know the medications that the patient X may take”.

The research process involves the execution order follows:

– Search all diseases of the patient X in the data warehouse which contains medical data of the patient X (according to its identifier).

– Search in the Health Integrated Web site of appropriate list of drugs prescribed for each disease.

– Selective search in the integrated pharmaceutical sites of compatibility of prescription drugs for each disease.

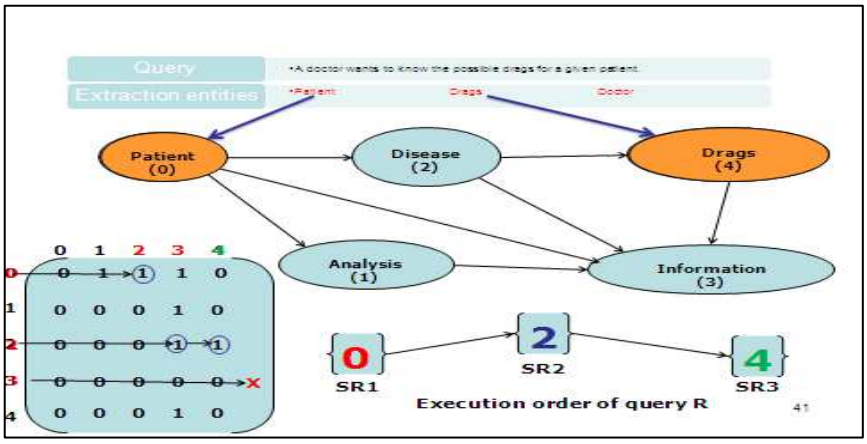

Fig. 6. *Oriented graph of ordering entities and the corresponding matrix*

The data search in the oriented graph of execution order passes in the first through the extraction of area model entities (medical area in our case). Then we proceed with the representation of these entities as a oriented graph which gives the logical order to be followed by the query. We construct the matrix corresponding to the graph which the intersection between the rows and the columns represents a path and we search the shortest path among possible paths that give a pertinent answer and satisfy the query posed by the user. A point to note is that the research data in our situation is incremental (Np Full), the algorithms that must operate in this type of research is very complicated, but since we have limited the number of sources that we will search by defining an execution order and all predefined ontologies by taking into account the user profile, the search becomes much less complicated, which shows the practical value of our methodology.

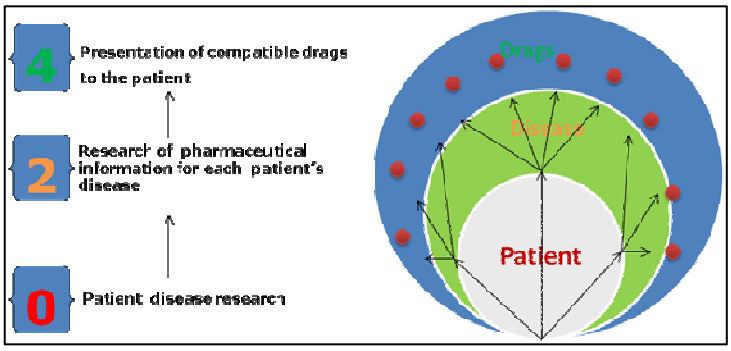

Fig. 7. *Stages of research data*

Meta-ontology

The Meta-ontology allows us to combine the basic generic knowledge in a shared ontology. The types of knowledge adapted to our context can be summarized into five types:

– Profile knowledge: They provide information on civil identification, geographical location and some characteristics.

– Mediator knowledge: They allow easy manipulation of the mediator by users.

– Area knowledge: This type concerns knowledge associated to services related to the area and sub-areas of application.

– Queries knowledge: They involve the elements necessary to establish the user's query, which is defined by a semantic basis of research and the research context.

– Restriction knowledge: These are global access restricting constraints to data of sources specific to the user.

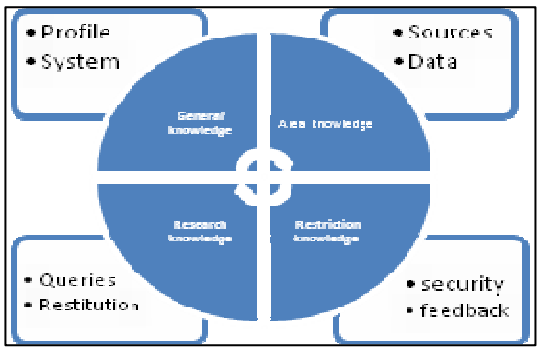

Fig. 8. *Meta-ontology*

Deployment Process

The process consists of intercepting user requests and modifying their behaviour using a set of functional adaptation operators described by the queries ontology. The strategy we have developed has two purposes: (i) integrating knowledge described by the domain ontology in the context of the users in an incremental way and (ii) ensuring the adaptation from a simple description of interest offered by the user dynamically. Indeed, the knowledge relatively to the dimensions of interest and personal data described by the profiles ontology are used to orient the user requests after the optimization and adaptation of integrated data sources. The adaptation data consist in transforming or replacing the data returned by the mediation system which is not usable in the considered context situation by other data adapted to the need of user profiles. This phase selects the sources expected to partially satisfy user requests. The interrogation of selected sources is preceded by an audit of access rights and permission systems defined by the security dimension of data described by the ontology of profiles.

Conclusion

The aim of our approach was to make the information available and optimal anywhere at any time via the mediator. The latter must be used in different contexts depending on the environment, the category of the user and his profile. However, one of the major problems of this type of system concerns the adaptation to the context of use. In an effort to be able to overcome these difficulties, we have proposed in this work a generic and evolutionary strategy for the adaptation of the mediators interfaces to the context of users by defining three types of ontologies.

In a perspective for a personalized access to the information via the mediators, we defined a model based on the classification and cooperation between ontologies for customization and optimization of integration process and interrogation. More precisely, the model is described through the interaction of two aspects. The first reflects the history of interactions with the mediator, represented by an ontology from the application and a users queries on the sources implicitly integrated in the successive search sessions. The second reflects the interests of the user automatically derived from the history of interactions. These interests evolve according to a strategy based on a measure of dependence between all types of users and sources included, to examine the sharing of profiles between users and implicit exchange of data between sources in the process of interrogation and search sequentially.

REFERENCES

Amato G., Straccias U., "User profile modeling and applications to digital libraries", *Abiteboul, A.-M. Vercoustre (Eds.): ECDL'99, LNCS 1696*, p. 184-197, 1999. c Springer-Verlag Berlin Heidelberg 1999.

Bouzeghoub M., Kostadinov D., "Personalisation de l'information: apercu de l'état de l'art et définition d'un modèle flexible de profils ", *CORIA*, 2005, p. 201-218.

Cali A., Calvanese D., Giacomo G., Lenzerini M., "On the Expressive Power of Data Integration Systems", *Proceedings of the 21st International Conference on Conceptual Modeling*, 2003, p. 338-350.

Charlet J., "Ontologies et thésaurus en medicine", *séminaire documentation numérique de l'INTD*, 22 janvier 2007, CNAM, Paris.

Dung Nguyen X., *Intégration de base de données hétérogènes par articulation a priori d'ontologies application aux catalogues de composants industriels*, Thèse de doctorat, Université de Poitiers, 2004.

Ezziyyani M., Bennouna M., Essaaidi M., "Mediator of the heterogeneous information systems based on application domains specification: AXMed Advanced XML Mediator", *IEEE Journal, International Computer Science an Applications IJCSA,* p. 25-45, 2006.

Fitzpatrick L., et Dent M., "Automatic feedback using past queries: Social searching". *Proceedings of the 20th International Conference on Research and Development in Information Retrieval (SIGIR-97),* p. 306-313, Philadelphia, PA, Juillet 1997.

Hamdoun S., Boufarès F., Badri M., "Construction et maintenance des entrepôts de données hétérogènes", *La revue électronique des technologies de l'information*, 23 juin 2007.

Hammer J., Garcia-Molina H., Widom J., Labio W., et Zhuge Y., “The Stanford data warehousing project”. *IEEE Quarterly Bulletin on Data Engineering, Special Issue on Materialized Views and Data Warehousing*, 1995, p. 41-48.

Jellouli I., El mohajir M., Zimanyi E., “Classification conceptuelle et ontologie de domaine pour l’intégration sémantique des données”, *La revue électronique des technologies de l’information e-TI*, Numéro 5 , 5 novembre 2008, p. 2-12.

Kindo T., Yoshida H., Morimoto T., et Watanabe T., “Adaptive personal information filtering system that organizes personal profiles automatically”. *In Proc. of the 15th Int. Joint Conf. on Artificial Intelligence (IJCAI-97)*, p. 716-721, 1997, Nagoya, Japan.

Kostadinov D., Peralta V., Soukane A ., Xue X., “Intégration de données hétérogènes basée sur la qualité”. *INFORSID*, 2005, p. 471-486.

Visser P .R.S., Beer M.D., Bench-Capon J.T.M., Diaz B.M. et Shave M.J.R., 1999, “Resolving ontological heterogeneity in the KRAFT project”. *In Database and Expert Systems Applications*, p. 668-677.