USING ONTOLOGIES WITH HIPPOCRATIC DATABASES
A Model for Protecting Personal Information Privacy

Esraa Omran, Albert Bokma
Department of Computing, Engineering & Technology, Sunderland University, Sunderland, U.K.

Shereef Abu Al-Maati
Computer Science and Engineering, American University of Kuwait, Kuwait

Keywords: Hippocratic database, Ontology, privacy, Health personal information.

Abstract: In the age of identity theft and the increased misuse of personal information held in databases, a crucial topic is the incorporation of privacy protection into database systems. Several initiatives have been created to address privacy protection in various forms, from legislation such as PIPEDA to policies such as P3P. Unfortunately, none of these effectively enforce protection of data. Recent solutions have emerged to enforcing data privacy & protection such as the Hippocratic database. But this technique has proved complex in practice. To overcome this deficiency we propose to use personal information ontologies in combination with Hippocratic databases. This method introduces a new way in reducing the complexity and in clearly identifying terms of privacy in the database architecture.

1 INTRODUCTION

Data privacy is a growing concern among researchers for numerous sectors such as healthcare, finance, e-commerce, and government. Many countries have recognized the importance of privacy protection and have set laws and acts such as the *Personal Information Protection and Electronic Documents Act* (*PIPEDA*) (University of Alberta, 2005). We will present an overview of this and its principles, and we discuss the Platform for Privacy Preferences Project (P3P) (W3 website). The P3P system enables Websites to express their privacy practices in a standard format that can be retrieved automatically and interpreted easily by user agents. P3P user agents allow users to be informed of site practices (in both machine- and human-readable formats) and to automate decision-making based on these practices when appropriate. Thus users need not read the privacy policies at every site they visit (W3 website). But setting the rules does not guarantee their application. Therefore, researchers have thought of applying privacy protection at the technical level in what is now known as Privacy-enforcing technology that aims at making privacy protection guidelines and laws an integrated part of the technology. In addition, this technology ensures that privacy laws and guidelines are practically applied to the data. Thus, an information system is designed to embed components that allow monitoring compliance of the system to privacy rules, guidelines and conditions. One of the leading technologies in this field was the Hippocratic database (Agrwal 2002) by Agrawal where a new idea is presented based on the classical Hippocratic Oath for medical doctors: “And about whatever I may see or hear in treatment, or even without treatment, in the life of human beings – things that should not ever be blurted out outside – I will remain silent, holding such things to be unutterable” and applying this to managing database access. The Hippocratic Database requires authorization for access to the datab it manages. This authorization makes data users access the attributes only for the usage purposes specified in the privacy policies. A more sophisticated mechanism to control access is achieved by integrating it to Personal Information ontology. As the personal information ontology will give specified meaning to each purpose, this will help in reducing the numerous number of purposes because it will forbid redundant purposes. In addition, the ontology will help in clarifying the

Omran E., Bokma A. and Abu Al-Maati S. (2009).
USING ONTOLOGIES WITH HIPPOCRATIC DATABASES - A Model for Protecting Personal Information Privacy.
In *Proceedings of the 11th International Conference on Enterprise Information Systems - Information Systems Analysis and Specification*, pages 376-382
DOI: 10.5220/0002001803760382

relations between the users and the purposes that they have the permission to access. This idea gives a new dimension to database security and privacy.

In this paper, we introduce a prototype implementation addressing several key issues in privacy management, and we demonstrate this prototype in the context of healthcare data management, a sector in which maintaining the privacy of individual information is of essential importance.

The remainder of this paper is organized as follows. Section2 describes the Hippocratic database techniques. Section 3 presents the Personal Information Ontology. Section 4 discusses different scenarios for Information Collecting and storing, information sharing, Data Retrieval and compliance auditing with the Hippocratic technique. Finally, a brief discussion of future work is presented.

2 HIPPOCRATIC DATABASE

The Platform for Privacy Preferences (P3P) provides a privacy policies specification and exchange but unfortunately it does not provide any mechanism to ensure that these promises are applied consistently in the internal data processing. On the other hand, K-anonymization has been investigated to meet the needs for statistical research. Its main idea is to suppress or generalize the personal information data in order not to disclose the identity of the record holders. But the k-anonymity can not be used in general database access managing applications as it was developed for statical researchers to perform reliable analysis on real data.

On the other hand, classical Hippocratic databases have been introduced as systems that enforce privacy policies by technical means. The Hippocratic database idea has emanated from the traditional Hippocratic Oath which is concerned with safeguarding the privacy of a patient's health information. A Hippocratic database includes privacy policies and authorizations that associate with each attribute and each user the usage purpose(s).

Privacy protecting access control deals with privacy policy specification and private data management systems (Sabah Al-Fedaghi, 2007). The purpose is the main factor in the Hippocratic database. A request to access to data depends on access purpose, and accessing permission is determined after comparing such a purpose with the intended purposes of that data in the privacy policies stored tables. Each user has authorizations for a set of access purposes. For example, nurses can access the patient health record for temperature and weight recording while doctors can access it for treatment purposes and so on. Though of its advantages in data management access, it has significant problems such as the complexity of users-purposes management and the purposes huge variety.

In this paper we will build on the classical idea of Hippocratic databases in investigating a new system. This will be achieved by integrating a suitable personal information ontology with the classical Hippocratic database. Management of attributes and users purposes is a complicated issue in the classical Hippocratic database.

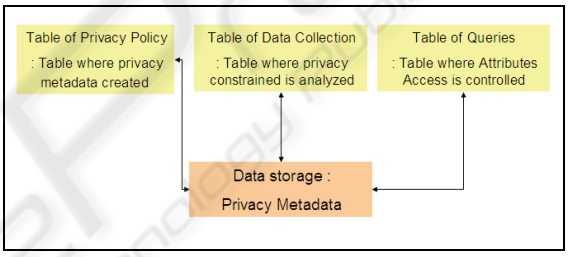


Figure 1: The infrastructure of the Hippocratic database.

For privacy protection, the Hippocratic database stores privacy metadata. *Privacy metadata* contains privacy policies and privacy authorizations. *Privacy policies* refer to administrator's policies for data collection and usage, including usage purposes. *Privacy authorizations* are like passport data administrators gives to data users to ensure that data are accessed only according to privacy policies. These authorizations specify usage purposes and authorized users for each attribute of a data and allow data users to access the attributes only for the usage purposes specified in the privacy policies.
The Hippocratic database applies privacy checking during data collection and query processing. Data providers specify privacy preferences during the data collection. *Privacy preferences* represent the data provider's intention (usage purposes). Hippocratic database stores data records along with the usage purposes included in the privacy preferences if the privacy preferences match privacy policies only. During query processing, the Hippocratic database checks that data users access is in accordance with the data provider's intention. Hippocratic database ensures this by first checking

usage purposes specified in privacy authorizations at the schema level, and then, by checking the usage purposes stored with the data at the record level.

3 PERSONAL INFORMATION ONTOLOGY

In addition to our first definition, privacy is also the right of individuals to determine for themselves when, how and to what extent information about them is communicated to others.

Ontologies as a method for creating a machine processable model of a domain has been studied since the early 1990s. This allows to create a model of a domain by specifying the concepts in it and their relationships and typically involves the modelling of classes and subclasses to create a class hierarchy to which properties can be attached and where observations or data from the domain can be associated to in terms of instances. Thus a shared conceptualisation can be created that can be used to manage information about a domain and share this between users or applications.

Using ontologies has been proposed (Gurber, 2001) as a solution for managing the intrinsic heterogeneity present in knowledge from different sources. For our purpose we will adopt the logical theory view of ontology, and the constraining axioms will play a crucial role in defining the semantic role in our database design. While we allow the communicating agents (doctors, nurses etc...) to have their own data access based on Hippocratic database and ontology, we will require the existence of a common ontology expressive enough to interpret the concepts in all agents' ontologies. The principal benefit of this approach is that it provides a formal base for reasoning about the properties of systems that perform automated knowledge and data translation based on a shared ontology.

In this paper we construct an ontology for personal information, therefore we will next have an overview on a Personal Information Act, PIPEDA.

The *Personal Information Protection and Electronic Documents Act* was enacted to establish national rules for personal information protection in the private sector and establishes, as law, the Canadian Standards Association's Model Code for the Protection of Personal Information, which encompasses the following principles: accountability; identifying purposes; consent; limiting collection; limiting use, disclosure, and retention; accuracy; safeguards; openness; individual access; and challenging compliance (University of Alba, 2005).

PIPEDA has been phased in over a three year period: 2001, 2002 and 2004. *PIPEDA* defines personal information to mean identifiable information about an individual and personal health information is defined from (University of Alba, 2005) as follows:

(a) information concerning the physical or mental health of the individual;
(b) information concerning any health service provided to the individual;
(c) information concerning the donation by the individual of any body part or any bodily substance of the individual or information derived from the testing or examination of a body part or bodily substance of the individual;
(d) information that is collected in the course of providing health services to the individual; or
(e) information that is collected incidentally to the provision of health services to the Individual.

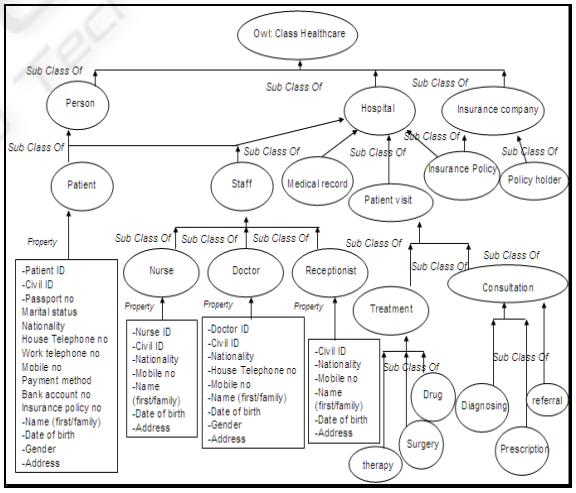
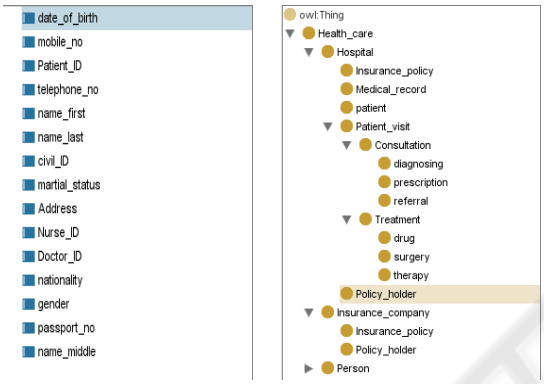


Figure 2: Investigated Personal Information Ontology.

Taking care of the above principles and consulting experiences from the International Health Clinic in Kuwait, we have investigated our Personal Health Information Ontology in Fig. 2 and 3 and its application as a layer on the top of the Hippocratic database as in Fig 5. The ontology has been constructed using the Protégé Owl tool (Fig 3) in

order to be connected in the future to a real project from the health sector. The Web Ontology Language (OWL) is part of the growing stack of W3C recommendations related to the Semantic Web. The Semantic Web is a vision for the future of the Web, in which information is given explicit meaning, making it easier for machines to automatically process and integrate information available on the Web. So when we construct our ontology using OWL environment, we make it more general to be used in other health care projects. This usage will enrich our ontology as each project will add to the ontology new concepts. For example: For the medical record subclass it could add new properties that we haven't stated in more detail yet.



(a) (b)

Figure 3: The Investigated Personal Information Ontology in protégé OWL (a-classes/b-properties).

The ontology has been created in a way that goes through all the concepts that have been proved to be the most common and reliable in health care, depending on the knowledge of an expert in the field (who designed the database for many hospitals in Kuwait). In addition, the ontology has been built to the information abstracted from real records at the International Clinic in Kuwait and forms from the internet and countless meetings with physicians, nurses and reception officials have been held in order to build a reliable ontology.

In this paper we have constructed a health system ontology based (using visual basic) connected to an Access database. This system will help physicians and researches in classifying diseases ontologically. The benefit of this classification appears clear during search process. The ontology helps in saving time and effort during the search process, as the ontology defines the disease asymptotes clearly for each case. For example, in Figure 4 we have shown the window that appears when we press the "search disease information" button. This window offers many helpful options to find disease information from the patient's database. This will help the physicians in diagnosing; help the researchers in finding useful statistics without affecting the patient's privacy.

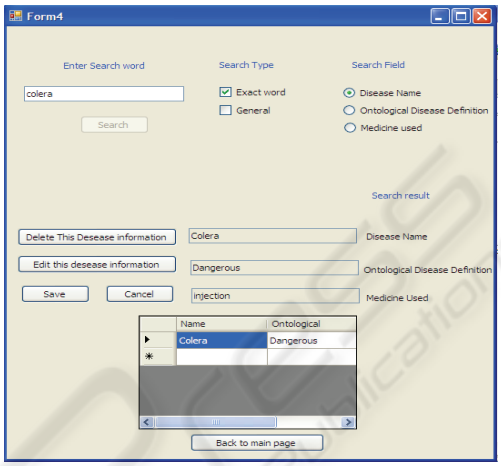


Figure 4: Search process using the Investigated ontological Health system.

4 DISCUSSION OF ONTOLOGY HIPPOCRATIC INTEGRATION

This section will discuss scenarios (comparable to (Agrawal, 2002) scenarios) to show the expected advantages from integrating the Personal Information ontology into the Hippocratic database. The ontology will add a new dimension to the Hippocratic database which is *classification*. The classification that the ontology provides, will define first the meaning of each data access purpose, and then which user is permitted to access for which purpose(s).

The effectiveness of the Ontology-Hippocratic integration will clearly appear through the scenarios. As this integration should improve the query performance as it will help in saving time and effort. In addition, it should improve the privacy saving as it improves the data access management. The scenarios will mainly concentrate on the health sector.

Electronic health records have attracted considerable interest by researchers, because of their numerous benefits, including improving health care delivery by allowing timely and accurate access to information by those involved in patient care,

reducing medical errors and adverse health events, augmenting security of patient information; and enhancing availability of information to support health system planning and reform as well as research (University of Alba, 2005).

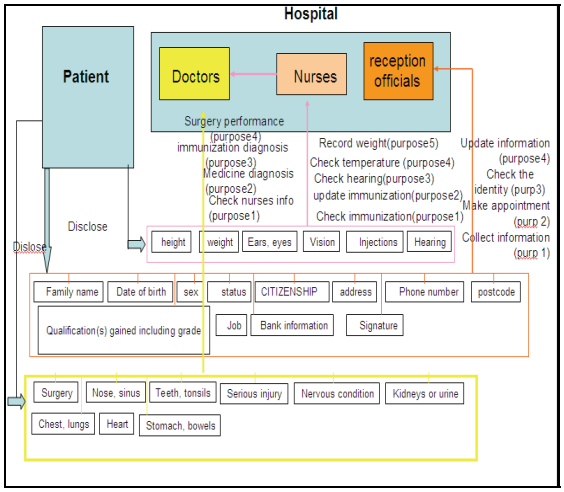


Figure 5: The prototype (Integrating ontology into the Hippocratic database).

4.1 Scenario for Information Collecting and Storing with the Proposed Approach

John wants to register for health care. First a receptionist collects his personal information (such as Name, birth date, etc). The personal information ontology will help in this step by specifying the information to be collected and the exact meaning of each piece of information in order to collect the necessary information. Then the receptionist stores his personal information. Then, the system processes these information in order to disclose/share them by Doctors and nurses as appropriate. The Hippocratic ontology based functionality will be in making a decision of whom to share what information. Here is where the role of ontology appears; the ontology will give each piece of personal information a clear and specific definition, then according to this definition, the personal information based ontology tables will be connected to their appropriated purposes.

4.2 Scenario for Information Sharing with the Proposed Approach

Here we will present a scenario demonstrates how the proposed appraoch can be used to facilitate policy-compliance information sharing among multiple organizations. Sara is a professor at a University Medical School with access to International's patient database under a joint research agreement. She is currently working on a project to evaluate the effect of pollution on the blood pressure. To begin her research, Sara logs into web portal and submits the following SQL query to the International Hospital database:

*Select * from patients where blood pressure>190*

Without Hippocratic ontology base controls, Sara would be given total access to the records of all patients with blood pressure > 190. This is a violation of International privacy policy and EU data protection laws, because not all patients have consented to reveal their health information to third parties for research purposes. With the proposed approach in place, the system rewrites Sara's query to comply with International's data disclosure policy and **patient** opt-in and opt-out choices. Thus, the system filters out the personal data that patients did not opt to share with third parties for drug research purposes and returns the remaining data that is responsive to the query. Again, the ontology would play a role in the construction of the SQL views, as SQL views perform a better classified access to the database. Ontology will save time and effort consumed in deciding what view to appear to each user.

4.3 Scenario for Data Retrieval with the Proposed Approach

A year after his annual physical exam, John has a car accident and his back has been affected. His doctor sends him to RadioTech Labs, an International affiliate, to have a series of X-rays performed. The lab technician types John's's name into a computer terminal and requests access to his medical records. In the absence of the proposed approach controls, the technician could potentially see all of John's personal health records stored in the International Clinic's database. However, with the proposed system in place, the application returns only John's contact information and the records of his latest hospital visit, but no other health records. This complies with International's data disclosure policy and John's privacy preferences.

4.4 Scenario of Compliance Auditing

George is an actor and a patient at the International Hospital. A local magazine discloses portions of George's personal health records, indicating that he

has been treated for depression. He believes that the International Hospital is responsible for this unlawful disclosure and threatens to sue the hospital under national data protection laws. The Hospital manager is very concerned about this case and requests that Mary, the database administrator, immediately provide him with an accounting of all who have accessed George's personal health data and to make all the efforts to know if any one in the hospital has disclosed George's health information.

Mary logs into the audit interface of our proposed system to begin the investigation. Mary would first like to know the identities of all persons who have accessed George's medical information in the past year. She starts with the allowed (user/purpose) combinations. In order to focus her search, she chose to know those who can access his health information only not other personal records (e.g., address, telephone number, payment information). This task was to be very difficult and takes more time without the aid of the proposed method. She has tracked and analyzed the queries that have accessed George health information. Especially those who have accessed his nervous records. Mary notices that the results show a large number of queries accessing George's medical records, but not all of those queries revealed the diagnosis of depression or his prescription for anti-depression medication. She has noticed in particular, that a nurse called Sally has accessed George's nervous record 50 times in one month. Though, she has no purpose for that as she has finished from giving him injections. It appears to Mary, that Sally is accessing George's nervous records without reasons for 50 times in one month using the same query. That's why she suspects in her and submits a report of all her database analysis to the manager. He did his investigations and he discovered that she has disclosed the depression information to the magazine. Without the aid of the proposed system, Mary would spend months looking in files and may have gotten no result at the end. The proposed system has shortened the way to Mary and made her job much easier than analyzing an ordinary database.

5 CONCLUSIONS

In this paper, we have opened the door for research in using ontology in data access management. We have integrated the Hippocratic database method with investigated personal information ontology in order to provide better privacy security. We achieved that by first giving a presentation to the Hippocratic ontology based technology and how it could play significant role in protecting the privacy of personal health records without sacrificing the value of information for diagnosis, treatment, or research purposes. Our presentation demonstrates how this technology enables efficient management, sharing, and processing of sensitive data in compliance with the principles of the *PIPEDA* and other data protection acts and laws. We have also discussed number of scenarios to demonstrate the importance of the new method. Finally we have presented some technical challenges that have been addressed. We have demonstrated this method as a possibility for privacy protection technologies and overcoming its difficulties and problems (Bertino, 2005). We hope that the technology outlined herein serve as a base for modern health records infrastructures and encourage the researches in applying ontology in information management security.

6 FUTURE WORK

The investigated system has been built in a prototype implementation and will shortly be applied to a real health project in order to prove its reliability. In addition the new method will be compared with traditional methods from literature such as classical Hippocratic database and k-anonymity (Sweeny, 2002).

REFERENCES

Sabah Al-Fedaghi, January 29th - 2nd February, 2007, *"Beyond Purpose-Based Privacy Access Control", The 18th Australasian Database Conference,* Ballarat, Australia.

R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu. , VLDB 2002, *"Hippocratic databases".*

Rakesh Agrawal and Christopher Johnson.*" Securing Electronic Health Records without Impeding the Flow of Information" IBM Almaden Research Center*

University of Alberta, Health Law Institute, University of Victoria, School of Health Information Science, April 2005, *"Electronic Health Records and the Personal Information Protection and Electronic Documents Act", Report prepared with generous funding support from the Office of the Privacy Commissioner of Canada.*

http://www.w3.org/P3P/

M. Richardson, R. Agrawal, P. Domingos, October 2003, *"Trust Management for the Semantic Web", 2nd Int'l Semantic Web Conf.,* Sanibel Island, Florida.

Thomas R. Gruber, 1991, *"The role of common ontology in achieving sharable, reusable knowledge bases", In Richard Fikes, James A. Allen, and Erik Sandewall, editors, Proceedings of the Second International Conference, Principles of Knowledge Representation and Reasoning.*

K. LeFevre, R. Agrawal, V. Ercegovac, R. Ramakrishnan and Y. Xu. , VLDB 2004 *"Limiting disclosure in Hippocratic databases".*

Elisa Bertino and Ravi Sandhu, JANUARY-MARCH 2005 *"Database Security—Concepts, Approaches, and Challenges", IEEE Transactions on Dependable and Secure Computing, VOL. 2, NO. 1.*

L. Sweeney,2002, *"K-anonymity: A model for protecting privacy", International Journal on Uncertainty, Fuzziness, and Knowledge Based Systems,* pp. 557-570.