

IWEBCARE: AN ONTOLOGICAL APPROACH FOR FRAUD DETECTION IN THE HEALTHCARE DOMAIN

**Panos Alexopoulos¹, Xanthi Benetou², Tassos Tagaris², Panos Georgolios¹,
Kostas Kafentzis¹**

¹ *IMC Research*
e-mail(s): {palexopoulos, pgeorgolios, kkafentzis}@imc.com.gr
Greece

² *Institute of Communication and Computer Systems*
e-mail(s): {xbenetou, tassos}@biomed.ntua.gr
Greece

Abstract: The European iWebCare project (FP6-2004-IST-4-028055) aims at designing and developing a flexible fraud detection web services platform, which will be able to serve e-government processes of fraud detection and prevention, in order to ensure quality and accuracy and minimize loss of health care funds. This paper discusses the approach this project adopts and which involves the introduction of a fraud detection methodology combining business process modelling and knowledge engineering as well as the development of an integrated fraud detection platform combining an ontology-based rule engine and a self-learning module based on data mining.

Key words: e-Government, e-Health, Knowledge-based systems, Fraud detection

1. INTRODUCTION

Fraud is an issue with psychological, economic and legal ramifications for both the public and private sector spanning geographic regions. The last EHFCN (European Healthcare Fraud and Corruption Network)¹ conference produced agreement among members on a common definition of fraud: “Civil fraud is the use or presentation of false, incorrect or incomplete statements and/or documents, or the non-disclosure of information in violation of a legally enforceable obligation to

¹ <http://www.efhcn.org>

disclose, having as its effect the misappropriation or wrongful retention of funds or property of others, or their misuse of purposes other than those specified”.

Particularly in the healthcare domain, fraud is committed when someone intentionally submits, or causes someone else to submit false or misleading information for use in determining the amount of health care benefits payable. Healthcare fraud could be committed by dishonest health care providers such as physicians, dentists, labs, and medical equipment suppliers or by plan members themselves.

The main consequence of healthcare fraud is the raise of the cost of health care benefits for everybody. According to the Deputy Health Minister of Scotland Lewis Macdonald² the potential losses to healthcare across Europe from fraud and corruption are estimated to be at least 30 billion euros each year and may be as high as £100 billion. For most employers, fraud increases the cost of providing benefits to their employees and, therefore, their overall cost of doing business. That translates into higher premiums and out-of-pocket expenses as well as reduced benefits or coverage.

In the context of the iWebCare project we are mainly interested in the technological aspect of fraud-fighting. In that area organizations and agencies seek multiple layers of health care fraud detection methods and tools ranging from rule-based systems to predictive modelling approaches. Our approach involves the introduction of a fraud detection methodology which combines the areas of business process modelling and knowledge engineering and which is based on the technologies of knowledge-based and data mining systems.

In this paper we intend to illustrate the combined methodology of our approach and focus on its knowledge engineering aspect by presenting the ontological model on which the iWebCare fraud detection platform is based. The rest of the paper is organized as follows. The next section discusses the iWebCare approach to detection of fraud and attempts a short survey on existing healthcare standards and ontologies that we took in mind while building our fraud detection ontology. Section 3 explains the need for the combined methodology that we introduce and describes how this methodology actually works. Section 4 provides an analytical description of the iWebCare fraud detection ontology while section 5 highlights the applicability of this ontology (and of the whole approach) to a specific case study. Finally, section 6 summarizes our approach, discusses open issues (especially modeling and technology limitations etc), and presents future work within the iWebCare project.

2. FRAUD DETECTION IN THE HEALTHCARE DOMAIN

2.1. iWebCare technological approach

² <http://www.scotland.gov.uk/News/Releases/2006/11/09152619>

In general, the IT fraud detection systems in the healthcare domain fall into two main categories: those that detect anomalies before claims are paid and those that sift through batches of claims after they are paid. The first are usually based on rules and prediction models while the latter utilize data mining techniques. Rules and prediction models are used for detecting irregularities and errors in payment claims and allowing thus the insurance organizations to deny the satisfaction of these claims. Similarly, in predictive modeling (Zukerman and Albrecht 2000), historical data is used to build profiles of fraudulent behavior in order to detect future occurrences of the same behavior based on the similarity to the existing profiles.

However, rule-based systems and predictive modeling can only defend against known (or predicted) fraud types. Data mining systems, on the other hand, utilize large datasets in order to discover unknown patterns of suspicious or fraudulent behavior. Additionally, data mining systems provide the foundation of predictive modeling as they are used for creating new fraud profiles.

The iWebCare approach aims to combine the rule-based and data mining approaches in order to build a system that performs rule-based fraud detection and learns new rules through a self-learning module based on a number of data mining and machine-learning techniques.

Ontologies play a vital role in this combined approach as they influence both the rule-based and data mining aspects. Apart from the rules, a really important component of a rule-based system is its knowledge base. Ontologies³ are knowledge models that represent a domain and are used to reason about the objects in that domain and the relations between them. Thus, a knowledge base may use an ontology to specify its structure (entity types and relationships) and its classification scheme.

Ontologies, however, are also very important to the data mining area as they can be used to select the best data mining method for a new data set (Tadepalli et al 2004). When new data is described in terms of the ontology, one can look for a data set which is most similar to the new one and for which the best data mining method is known, this method is then applied to the new data set. In this way, there is no need for trying out every known method on the new data set, but the one (or few) that is most promising can be directly selected.

2.3. Survey of healthcare ontologies/standards

Creating a knowledge model for a given domain from scratch is most of the times a very difficult and time/resource consuming task especially as far as the knowledge acquisition process is concerned. Therefore, in any such effort, the existence of already established and commonly accepted standards, classification schemes and ontologies regarding this domain should be taken seriously in mind.

³ <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>

In the case of healthcare existing medical classifications, terminologies and taxonomies include the International Classification of Diseases (ICD)⁴, the ATC system⁵ and the SNOMED CT system⁶. The ICD classification is an international standard diagnostic classification for all general epidemiological and many health management purposes. It provides codes to classify diseases and a wide variety of signs, symptoms, abnormal findings, complaints, social circumstances and external causes of injury or disease.

On the other hand, the Anatomical Therapeutic Chemical (ATC) system is a system for classification of medicinal products according to their primary constituent and to the organ or system on which they act and their chemical, pharmacological and therapeutic properties. It provides a global standard for classifying medical substances and serves as a tool for drug utilization research.

Finally, SNOMED (Systematized Nomenclature of Medicine) is a system of standardized medical terminology developed by the College of American Pathologists (CAP). It is a “comprehensive and precise clinical reference terminology that provides unsurpassed clinical content and expressivity for clinical documentation and reporting” and it allows a consistent way to index, store, retrieve, and aggregate clinical data across specialties and sites of care”.

3. IWEBCARE METHODOLOGY

The iWebCare proposed fraud detection methodology suggests that fraud is actually an operational risk for an organization and as such it should be treated through a risk management process. Risk management⁷ (RM) is the process whereby public organizations may methodically address the risk associated to their activities with the goal of achieving a sustained benefit within each activity and across their portfolio of activities. The focus of RM is to identify, measure and treat these risks in order to reduce their probability of happening.

In a similar fashion, the iWebCare methodology defines a process for identifying, measuring and treating fraud in the context of e-government services. This process comprises three steps, namely establishment of the fraud context, identification of fraud within this context and incorporation of this information into the ontological model of the iWebCare fraud detection platform.

Establishment of the fraud context within an organization involves defining the type of fraud the organization wishes to fight and identifying the business processes fraud occurs upon. This is done through a business process modelling⁸ procedure which records the fraud susceptible business processes of the organization and their

⁴ <http://www.who.int/classifications/icd/en/>

⁵ <http://www.whocc.no/atcddd/>

⁶ <http://www.snomed.org>

⁷ Risk Management Guide”, WCO , June 2003

⁸ <http://www.euml.org>

context. On the other hand, fraud identification involves the description of potential fraud cases that could occur within the organization and of corresponding detection methods. This identification is done in two ways, namely by acquiring organizational knowledge regarding fraud from experts and by utilizing data mining methods in order to extract unknown fraud patterns.

The final step of the methodology involves incorporating the knowledge derived from the two previous steps into the ontological model of the iWebCare platform so that it can be utilized by the fraud detection system. This step requires following a knowledge engineering procedure compatible to the structure of the iWebCare ontology.

4. IWEBCARE ONTOLOGY

As it has already been described in previous paragraphs, the purpose of the iWebCare ontology is to be used for fraud detection in the healthcare domain. The detection is to be performed by means of rule-based reasoning and the ontology should provide the necessary knowledge on which these rules would be based.

For building the ontology we followed a bottom-up development approach (Noy and McGuinness, 2001) and we defined and utilized an ontological architecture based on layers. The reason for this was that the healthcare domain is quite vast and the development of a complete healthcare ontology would require not only a significant amount of time, resources and domain experts. Furthermore, especially when it comes to fraud, there is a significant overlapping with other domains such as law, finance, sociology etc, which increases the complexity and effort for the construction of a “global healthcare fraud ontology”.

For all these reasons, instead of building such a global ontology we decided to build an ontology that would cover initially a couple of specific domains (see section 5) but which should easily be extended to other domains as well. The latter was made possible through a multi-layer architectural design of the ontology which makes the latter adaptable, extendible and to some degree reusable.

4.1. Layered architecture of the iWebCare ontology

The overall architecture of the domain specific ontology consists of three independent but interconnected layers each of which defines its own set of ontologies.

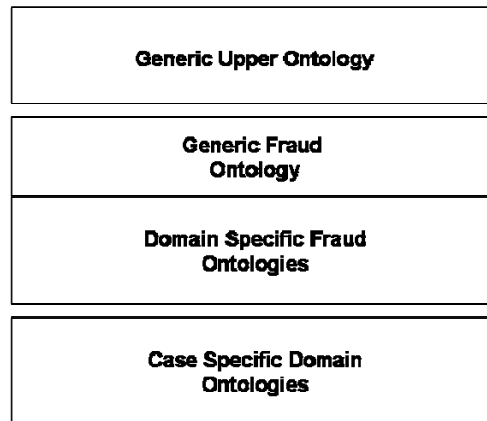


Fig. 1. iWebCare Ontology Layered Architecture

The bottom layer (or domain specific layer) consists of domain ontologies which model the business processes of the specific case studies that are examined for fraud. The main purpose of this layer is to provide the basic knowledge on which the fraud detection rules are going to act.

On the other hand, the middle, or fraud specific, layer comprises ontologies which model fraud related knowledge such as fraud types and fraud detection processes. The content of these ontologies reflects the knowledge of fraud domain experts and it is primarily used as the basic means for expressing the fraud detection rules that these experts provide.

The middle layer can also be considered as having two sublayers, a case-specific one and a generic one. The domain-specific sublayer models the fraud characteristics of the use cases' domains, while the generic sublayer provides more abstract and generic knowledge and provides the basis for applying the iWebCare's approach into virtually any fraud susceptible field.

Finally, the upper layer, namely the Generic Upper Ontology, captures generic and domain-independent knowledge that helps minimize redundancy and duplication of knowledge within the overall ontology.

5. CASE STUDY: The Social Security Fund of TSAY

TSAY⁹ is the insurance body of all healthcare professionals in Greece and its interest concerning healthcare fraud is detected in the prescription reimbursement domain. The reimbursement process requires that a TSAY's member initially purchases the drugs he needs from some pharmacist paying only a percentage of the actual cost and then the pharmacist claims the rest of the money from TSAY. However, it is often the case that the prescriptions TSAY is asked to reimburse contain erroneous or deliberately inaccurate data so that larger sums of money can be claimed or inappropriate drugs can be prescribed.

⁹ <http://www.tsay.gr>

Detection of fraud within TSAY was made possible by applying the iWebCare methodology. Initially, a complete business process model of the organization's prescription related processes was created (establishment of fraud context) and then the fraud identification phase took place. Both the fraud context and the fraud description were used to create the domain specific fraud ontology and the case specific domain ontology for the TSAY case.

More specifically, the business process analysis identified as highly relevant the following processes:

- The issuance of prescription booklets by TSAY
- The issuance of prescriptions by doctors
- The inspection of prescriptions by the ministry of health
- The filling of prescriptions by the pharmacists
- The reimbursement process for filled prescriptions

Based on these processes, a procedure of extracting concepts and relations from the prescription reimbursement process and the processes it interacts with, took place. The result was an initial TSAY domain ontology that represented the underlying knowledge of the way TSAY reimburses its members' prescriptions.

The next step of the iWebCare methodology involved the acquisition of case specific fraud detection rules by corresponding domain experts. The rules identified during this process comprised two main categories, namely auditorial rules and medical rules. Auditorial rules tried to detect incomplete prescriptions and invalid or miscalculated data while medical rules tried to detect prescriptions in which the data are inconsistent from a medical point of view.

An example of an auditorial rule is when a prescription contains no diagnosis at all for the drugs that it prescribes and an example of a medical rule is when the diagnosis written on the prescription is not included in the indications of the prescribed drugs.

From these rules a prescription-specific fraud ontology was created and incorporated into the third layer of the iWebCare ontology architecture.

6. CONCLUSIONS

In the previous sections we tried to give an overview of the iWebCare project and highlight its great potential impact in the field of fraud detection in the healthcare domain. We claimed that the most important aspect of the project was the innovative fraud detection methodology it introduces and which combines uniquely three different scientific fields, namely Business Process Modeling, Knowledge Engineering and Data Mining. We believe that this methodology can be easily adapted to virtually any e-government domain (apart from the healthcare one) and we intend to show this in the future by applying the methodology to additional cases studies.

Furthermore, of great importance is the WebCare ontology and particularly its layered architecture. The significance of the latter is not only that it makes the platform domain-independent but that it also serves as a common framework for modeling fraud-related knowledge and detection methods.

REFERENCES

- Noy N. F., D. L. McGuinness (2001). Ontology development 101: A Guide to Creating Your First Ontology. *Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001.*
- Tadepalli S., A.K. Sinha, N. Ramakrishnnan (2004). Ontology driven data mining for geosciences. *Proceedings of 2004 AAG Annual Meeting, Denver, USA, 2004.*
- Zukerman I., D.W. Albrecht (2000). Predictive statistical user models for user modeling. *User Modeling and User-Adapted Interaction* 11(1-2), 5-18