Expert Systems with Applications 39 (2012) 8522–8536

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

# Improving electronic health records retrieval using contexts

Belen Prados-Suárez [a], Carlos Molina [b,*], Carmen Peña Yañez [c], Miguel Prados de Reyes [c]

[a] Department of Languages and Computing Systems, University of Granada, Granada, Spain
[b] Department of Computer Sciences, University of Jaen, Jaen, Spain
[c] Computer Science Department, San Cecilio Hospital, Granada, Spain

ARTICLE INFO

Keywords:
Electronic Health Records (EHR)
Context
Pertinence
Contextualized access
Fuzzy logic
Access improvement

ABSTRACT

This paper aims to solve a recently arose problem, related to the access to the Electronic Health Records (EHR) in the Hospitals. Due to the digitalization of the information contained in the medical records, and the growing availability of devices that directly generate digital documents to include in it, the EHR are becoming unmanageable. Even more, to find a concrete item of information relevant for a given assistance act is a very hard, difficult and time-consuming task. To solve it we propose here the definition of *contexts of access* to the EHR, to exploit the logical division of the information inside each document in the EHR into *data groups*, and the computation of the *pertinence* of each data group to each context. It allows us to prioritize, the information in the EHR, even at a concrete data item level, according to the situation from which it is acceded. This way when the medical personnel is involved in an assistance act, the most relevant information for it is the one first showed, being able to widen the search but always according to the relevance. With it we not only improve the accessibility to the EHR and make easier the work of the doctors, but also enable other applications like the ubiquitous computation or the mobility, using devices like tablet PC's and PDA's.



## 1. Introduction

The use of EHR has become a reality in the everyday practice of most of the hospitals, so it is possible to find in the literature a great variety of proposals to implement the EHR in different specialities like pediatrics, nursery, family care, emergencies, radiology, elder care or outpatient consultation (as can be seen in Ginsburg (2007), Cho, Staggers, & Park (2010), Gagnon et al. (2010), Karahoca, Bayraktar, Tatoglu, & Karahoca (2010), Erdal et al. (2009), Pung et al. (2009) and Vishwanath, Singh, & Winkelstein (2010), respectively); and also, under different regulations depending on the country, like in the Korean or the Czech medical systems (in Cho, Kim, Kim, Kim, & Kim (2010) and Nagy et al. (2010), respectively).

However in most of these proposals, as well as in other studies regarding the satisfaction of the users about the implantation of the EHR like McAlearney, Robbins, Hirsch, Jorina, and Harrop (2010), Ross (2009), Vishwanath et al. (2010), Svanaes, Das, and Alsos (2008) and Vest and Jasperson (2010); or even about the comparison of different EHR systems as Bisbal and Berry (2009) and Flores Zuniga, Win, and Susilo (2010), there is a remarked problem: the great amount of information that is accumulating in the EHRs is making arise problems of access. Since all the information is always available, it is becoming really difficult to access a concrete information item required, even in relatively simple situations. It becomes really serious for situations like the urgencies, where the decisions must be taken within seconds and the relevant information for the concrete case should be immediately available to support them. Even more, the problem will get worse due to the increasing use of new medical machines and devices like PACs, that automatically generate documents to be included in the EHRs (Prados & Peña, 2003; Prados et al., 2010).

To solve this situation the storage and access cannot be restricted only to information with high clinical value, since depending on the assistance act the information needed may change. This problem is so relatively recent, that up to this moment it is quite difficult to find proposals that really face its whole extent. Most of them just focus on the definition of data structures and documents to organize the information provided by the EHR, offering navigation systems on it, like Jerding and Stasko (1998) and McAlearney et al. (2010). However they do not constitute a solution, since they allow logical and structured access to the information, but they do not avoid the uncomfortable selections steps and the successive screen-shots to reach the desired information (Prados de Reyes, Carmen Peña Yáñez, & Suárez, 2006).

In the medical research community it is clear that "having a good access to the information needed benefits the quality of the

* Corresponding author.
E-mail addresses: belenps@ugr.es (B. Prados-Suárez), carlosmo@ujaen.es (C. Molina), carmenpy@decsai.ugr.es (C. Peña Yañez), prados@decsai.ugr.es (M. Prados de Reyes).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2012.01.016

attention received by the patients" (Adams, Adams, Thorogood, & Buckingham, 2007). In addition, it is being pointed the importance of taking into account the situation or *context* from which the access is being performed (Weinstock, 2010), as a means to improve the access to the EHRs.

As an example, Collins and Speedie (2008) propose to use an *infobutton* engine to manage the clinician and patient context in order to provide concise answers to frequent questions posed by clinicians. "Infobuttons are information retrieval tools that help clinicians to fulfill their information needs by providing links to on-line health information resources from within an electronic medical record (EMR) system" (Del Fiol & Haug, 2009). These models are usually based on classification models to predict clinician's decisions. However, these models are restricted to very concrete topics like the "medication infobutton data, used to predict medication-related content topics (e.g., dose, adverse effects, drug interactions, patient education) that a clinician is most likely to choose while entering medication orders in a particular clinical context" (Del Fiol & Haug, 2009).

Other proposals related to the definition of contexts do not face explicitly nor directly the problem posed. They are mainly focused on the knowledge mobilization (Ray & Wimalasiri, 2006) and the ubiquitous computation (Judd & Steenkiste, 2003; Kang, Lee, Ko, Kang, & Lee, 2006); or on the standardization of Hospital Information Systems and the exchange on information between them (Cayir & Nuri Basoglu, 2008; Lahteenmaki & Kaijanranta, 2009; Nagy, Preckova, Seidl, & Zvarova, 2010b). Proposals in the first cases, can rarely be applied to the Hospital Information Systems, since they are mainly based in the use of sensors to identify the context (Kang et al., 2006) or to provide information according to the device used so other applications can perform pervasive computing (Judd & Steenkiste, 2003). In addition, none of these proposals are designed nor useful for the immense databases of EHR. Proposals in the second case, instead of focusing on identifying the information that is really needed to be exchanged, are centered on adapting the system, its structures, contents and interfaces to different regulations and standards like HL7 (Data exchange standard, 2011), DICOM (Image storage standard, 2011; Open-EHR, 2011), SNOMED-CT (Nomenclature medical standard, 2011), or the most recent proposal of the European Committee for Standardization: the ISO 13606 regulation. It leads them to forget and even obviate the needs of the health professionals, who are the real users of the system, and whose work improvements have more repercussion and impact in the quality of the medical assistance provided to the patients.

Nevertheless, the problem of the access to concrete information items of interest in huge databases, do is addressed explicitly in other environments, like business, legacy and e-government (in Chaker, Chevalier, Soule-Dupuy, & Tricot (2010), Mao & Benbasat (2001) and Bohm, Wolf, & Krcmar (2010), respectively). A German office of digital services to the citizens has detected the problem through deep studies (Bohm et al., 2010), but still has not proposed a solution to it. In the business framework this situation has also arose, as indicated by Chaker et al. (2010) and Buchholz, Hochstatter, and Linnhoff-Popien (2007), and the proposals to solve it are based on the improvement of the information retrieval by the definition of different contexts and business models (as in Jung, 2009), changing the actual access mode and adapting it to the real information needs of the acceding user. However, these proposals are too young and are still in their first development phases, so it is soon to extend and adapt them to other type of systems.

This is why we propose to analyze the daily practice of the medical staff, and follow the same philosophy as these solutions, to improve their access to the information in the EHR based on the information they usually request in each assistance act. Following the work of a doctor we can find a great variety of situations with different purposes: from a deep study of a complex diagnosis process in his office, to a simple revision of the last consultation inform in a control of evolution process, passing through the requirement of very concrete data in the response to an emergency. As can be seen, we face a wide variety of activity contexts, with quite different requirements of information. In other words, we have different sets of relevant documents or information items of the EHR, depending on the *context* we are involved in.

Our proposal is based on the study of the access patterns so the information showed to the user can be *context-sensitive*. This way the system would only show to the doctor the information that is relevant to his/her present context. However it must be taken into account that the information needs are not static, and they may change along the time, so it is possible that a piece of information that today is important will be useless in the future. Moreover, the age of the data has influence too: there are cases like some analysis that must be repeated if the last result of the same type of analysis is older than a few months, since the results may change. Hence, all of these aspects must be considered when defining the *pertinence* of the information items to the *contexts*.

According to all of it two problems must be faced. On one hand it is necessary to identify the contexts of access. On the other hand the information relevant for each of them must be identified. To solve the first problem can be found some proposals focused on the context modeling like Bobillo, Delgado, and Gómez-Romero (2008), Chaker et al. (2010), Ehsan, Amini, and Jalili (2009), Garcia-Morchon and Wehrle (2010) and Chu, Johnson, and Kangarloo (2000); but most of them are just theoretical models too complicated to be integrated in an existing system, and also require such complex algorithms that make them not suitable for an hospital information system. Even more if the system has to be updated continually to adapt to new needs. Proposals to face the second problem are mainly oriented to identify the relevant information inside documents as Järvelin and Kekäläinen (2000), Jones (2004), Cao et al. (2006) and Kanoulas, Pavlu, Dai, and Aslam (2010) propose; but due to the great amount of data involved in the Hospital Information Systems (hundreds of millions of records) it is not possible to use them. We need a very efficient way to contextualize the access and decide which information is relevant on each situation.

In this paper we present our proposal to do it, exposing first the support we have used as base, as well as the general structure of a EHR, in Section 2. Next, in Section 3, we show how to define the contexts, as much automatically as on demand, and how to identify them when a doctor accesses the system. Then, in Section 4, we propose a method to compute the pertinence of the information groups or items to the contexts, considering its several aspects. Based on it we establish a priority ordering in the use of pertinence values, and we show how it is used to improve the access to EHR, in Section 5. After it, in Section 6, we exemplify and summarize our proposal, and we make some comments about the implementation in Section 7. Finally we present some results and also our conclusions in Sections 8 and 9, respectively.

## 2. Background

First of all we must indicate that the proposal presented here has been developed in collaboration with the University Hospital San Cecilio from Granada, and that we have based their Electronic Health Record System, and used it as reference.

This system stores around 800.000 EHR, containing more than 50 millions documents. In the future it is expected to have a fast increase in the size, due to the inclusion of new types of documents from two sources: old documents that still have not been digitalized (scanned images, MRI, etc.) and new documents generated from the recently and future acquired devices and equipments like PAC's.

In this section we briefly show the characteristics of this EHR system, as well as the structure of the Electronic Health Records stored on it.

## 2.1. Electronic Health Records structure

The information stored in the EHR is structured according to the *Reference Model* given by the ISO-13606 (2008). According to this standard, the elements of the hospital information systems are organized according to an Ontology with a class *structure* that gives rise to the following classes:

- *Folder*: This class represents the divisions at the highest level inside the clinical history. In our case these divisions are the *assistance acts* and the *pathologies*, so all the documents in the EHR are grouped into assistance acts or pathologies, and both classifications coexist.
- *Section*: This class of the standard represents logical groupings of information, each one representing a set of data with an uniform informative clinical guidance (Fig. 1), and corresponds to each *document* stored in the EHR. Examples of documents are from a blood analysis to a preanaesthetic study, or from an admission document to a X-ray test.
- *Entry*: According to the standard each entry represents a clinical observation or a set of them. It corresponds to what we call *data groups* (i.e. the hematology information in a blood analysis).
- *Cluster* and *Element*: These classes correspond to what we call *data items*. The difference between these classes is that the first one is used to represent a unique observation or action (a data item) that requires a complex structure like a list, a table or a temporal series (i.e. an electrocardiogram); whereas the second class represents a unique and simple value, instance of some of the types defined by it (i.e. the percentage of hematocrit in a blood analysis).

As indicated in Prados and Peña (2003) and Prados-Suárez, Revuelta, Peña Yáñez, and Fernández (2008), each document is characterized by a set of properties like:

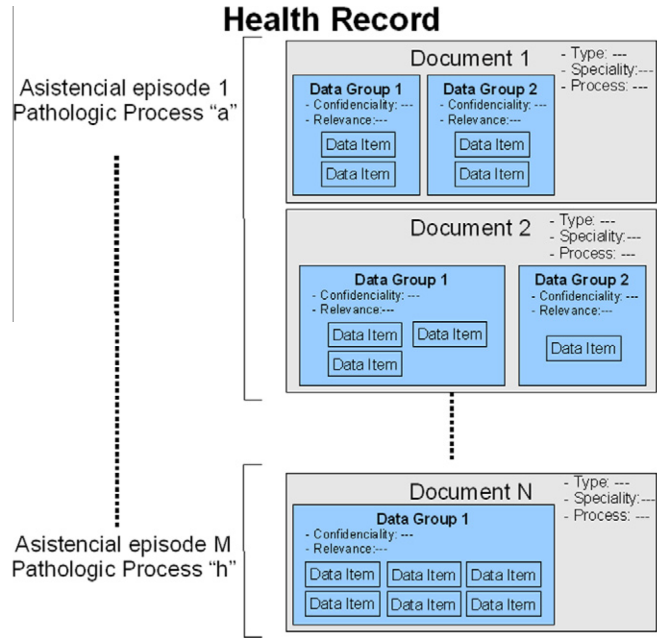


**Fig. 1.** Logical organization of the EHRs.

- the type (exploration, anamnesis, epicrisis, checkup, nursing control, intervention, external,… ).
- the speciality (medical speciality as surgery, cardiology and so on, nursing, administrative, etc.).
- the pathological or clinical process (documents about pregnancy, cataract, diabetes, etc.).
- …

In our system, documents are organized according to *assistance episodes* (admissions, outpatient consultation, emergency assistance, day hospital,…) in a chronological or medical ordering, depending on the assistance processes. The documents are classified according to the types, considering 1500 different documents classes in the system: intervention sheet, progress sheet, nursing

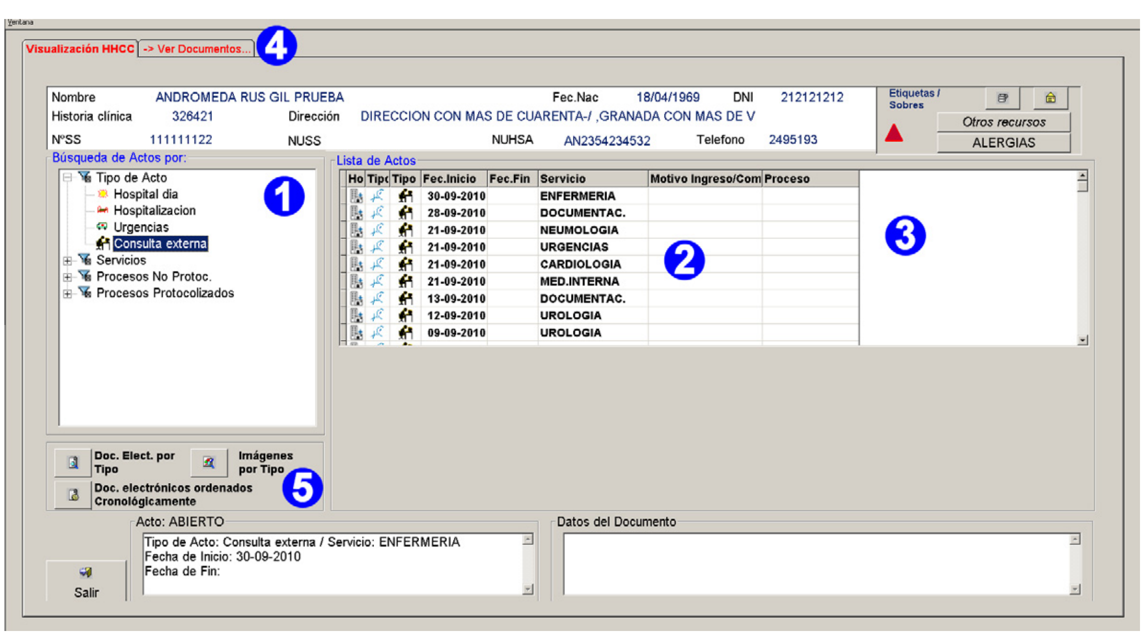


**Fig. 2.** Screenshot example of the applications.

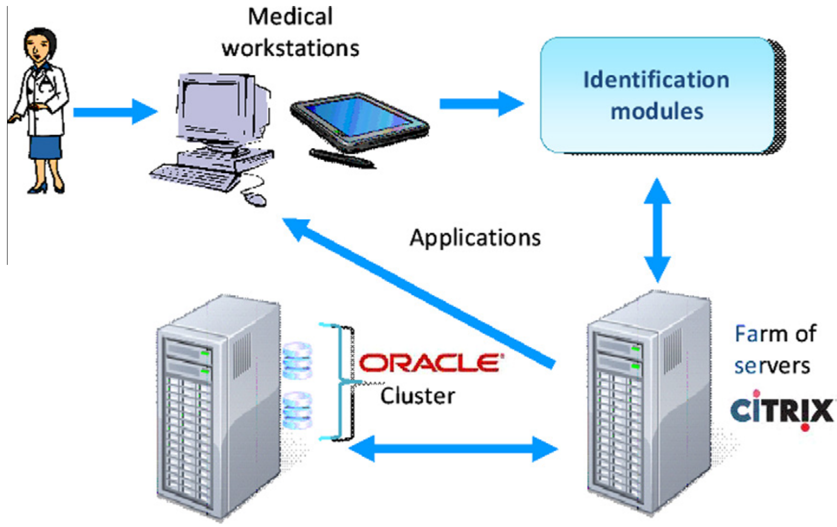


**Fig. 3.** Structure of the system.

sheet, pregnancy process, diabetes protocol, radiological report, and so on. An example of it can be seen in Fig. 2, where the interface that the medical personnel use to access the EHR is shown.

When a doctor is looking for a concrete data item, from a specific test made to the patient, he/she must select (on the square marked with a number 1 inside a circle in Fig. 2) the type of assistance act that is performing. Then a search in the square marked with 2 in Fig. 2 must be done, according to the date of the test or the medical speciality in which it was done. Once the assistance act in which the test was made is found, the doctor has to find among the documents generated in that act, the one with the results of the test, in the interface part marked with 3 in Fig. 2. With it the document is recovered and finally it must be scanned to find the item of interest, by clicking on the tab marked with a number 4 in Fig. 2. Though it is possible to use different ordering criteria (marked with number 5 in Fig. 2), it can be seen that it is a long tedious task that consumes a lot of the doctor's time, and that must be repeated each time the doctor needs to know something about each patient attended.

### 2.2. Data groups

Inside the documents there are data items that can also be grouped into small logical units, that we call *data groups*, when they are related under a clinical point of view (Fig. 1).

Each data group inherits the general properties of the document where it is contained. In addition, as shown in Fig. 1, each data group has its own specific properties, like the relevance level (for the concrete patient and episode), the confidentiality level, etc.

Examples of data items for a blood analysis or a preanaesthetic study are: erythrocyte, hemoglobin, corpuscular volume, amylase, GGT, HDL-cholesterol, LDL-cholesterol or VLDL-cholesterol, in the first case; and Hypertension, cardiopathy, electrocardiogram, radiologic study or echography, in the second case. These data items can be grouped into the data groups general biochemistry and lipid information, for the first type of document; and risk factors or additional tests, for the second type. Here we would like to remark that the information of EHR and patient's identification is a "special" data group, common to all the documents. Due to it, it is discarded from the processes explained later.

This logical organization of documents and their content, allows the processing and analysis of the information, as much at *document* level as at individual *data items* level or *data group* level. Here we consider the data groups as the minimum unit of information, since a single data item can be managed as a data group with just one element.

### 2.3. EHR information system

The structure of the system is shown in Fig. 3. The users access the system using medical workstations. These are normal PCs, light PCs (or net PCs), medical devices like the X-ray systems or the ultrasound scans, or the most recently incorporated terminals as the Tablet PCs and PDAs. The user then log on the system and access to a Citrix[1] farm of servers where the applications are executed. All the data are stored in a data base cluster using Oracle DBMS.[2] The screen-shot of the Doctor's interface once logged is shown in Fig. 2.

This system, as legally demanded, stores each access to the EHR, indicating the data acceded and, in case of modification, the modified data; the staff member acceding; and the assistance situation (called "*controlled assistance situation*") in which the access occurs. From now on, we will call this access data base as *Retrospective Access Data Base (RADB)*. The number of records stored in the RADB is in the order of hundreds of millions.

We support the work proposed here on the registers of this data base since, as we will see next, their analysis allows to know which information has been acceded and the related context.

## 3. Contexts

We call *Context* to a situation in the Doctor–Patient relationship inside an assistance act, requiring an access to the information previously stored in the EHR.

To contextualize the access to the EHR we first need to establish the set of possible situations or *contexts* where that access may occurs. Then, to exploit the contextualized access system, it is necessary to count on a mechanism to identify the context in which the medical staff is involved.

### 3.1. Context definition

The contexts can be defined under three criteria or a combination of them:

[1] http://www.citrix.com.
[2] http://www.oracle.com.

• P-
athological process: In this case the contexts are defined based on a diagnosed pathology that requires monitoring and it is included in the EHR as such process. Some of these processes are defined by the Regional Health Administration and others by the hospital services themselves. It must be taken into account that several medical specialities can be involved in the same process. Some examples are the pregnancy process, the cataract process, the diabetes process, etc.
- Medical speciality: Here the contexts are defined according to the specificity of each medical speciality (pediatrics, gynecology, nursery, cardiology, etc.).
- Kind of assistance: The context definition here is based on the environment where the assistance process takes places. The following cases can be distinguished:
  – Diagnostic study.
  – Surgical intervention.
  – Post-surgical revision.
  – Evolutive revision.
  – Room visit.
  – Treatment revision of outpatient consultation.
  – Analytical control.
  – Urgent assistance situation.

According to these three criteria, we ask different medical doctors to identify the contexts on each speciality. The set of contexts obtained has been reviewed by different groups of medical doctors to validate the results for their corresponding speciality.

### 3.2. Context identification

Once we have the contexts, it is necessary to have an automatic method to identify when a medical doctor accesses an EHR in a given context. By means of different interviews with the medical staff we have identified some characteristics of the accesses they perform, that are important in this process:

- Speciality of the medical staff like cardiology, ophthalmology, internal medicine, emergency, administration, nursing, and so on.
- Position of the medical staff. There are different positions for each type of medical personnel like. Some examples are: resident (from first to fifth year), facultative, section manager or head of service for the medical doctors; the categories of management technician, administrative technician, section manager or head of service for the administrative personnel; or the nursing position or nursing supervisor in that department.
- Type of the medical workstation. The data about the type of terminal used to perform the access, gives a lot of information about the type of assistance act in which the medical staff is involved. This attribute has several parts:
  – The type of the terminal. This value gives information about the hardware used (PC, PDA, patient's room terminal, computer associated to a concrete equipment like the X-ray machines or ultrasound scan, etc.).
  – The medical unit associated. Each terminal is associated to an unit (gynecology, pediatrics, etc.) for management reasons; but this information helps to identify the context. As an example if a cardiologist is acceding an EHR from a computer associated to the emergency unit, the context could be a cardiology emergency.
  – Physical location. It helps to concrete even more the type of context in which the medical staff is involved. In the previous example, if the terminal acceded is located in the observation room in emergencies, the context is different from the case when the access is performed from the surgery room.
- The kind of the present patient's appointment. For each appointment with a doctor, the information about its type is stored. There are around 50 usual types of appointments like first visit, checkup, scheduled visit, urgent visit, extern emergency, admission, several types for the different complementary tests and explorations, inter-consultation, movement between services, and so on. There are also some other types less usual or even rare but also considered in the system, like radiologic surgery.
- Last visit of the patient. This information in some cases helps to predict the cause of the next appointment. As an example, always after a surgical intervention there is a post-surgical checkup.

To identify the context with these attributes we use a rule-based system built on the past accesses data base (RADB). To get the data necessary to build it, a question about the context was added to the normal application (Fig. 2), so the medical staff could answer it each time an EHR was acceded. The system showed a list of contexts, filtered by a very simple criteria as the medical speciality, and the doctor chose the one that best fitted his/her access. This data collecting process, was active for 6 months and we have gotten an answer rate of 24.5% (about half million records).

With all the information collected, the RIPPER algorithm (Cohen, 1995) was used to build the rule-based system. This is a well known algorithm that produces good results, even compared to more recent proposals. This method basically consists on building a list of rules in an iterative way, based on the information gain. After it, a pruning process on the rule list is performed, which improves the classification rates of the unknown cases.

To verify that the results of the algorithm can be used, we have first tested it using a 10 cross-folder validation. In this process we got an average classification rate of 89.32%, which can be considered a good percentage, taking into account the high number of contexts (classes). Once the quality of the rule set has been proved, we have built a new model, but in this case using all the records. With it we have obtained an ordered list of rules, with near two hundred elements and we search for the present context by sequentially verifying this list and stopping at the first satisfied rule.

Though it may seems to be a considerable number of rules, we optimize the search on it by saving the rules in a data base where all the antecedent's variables are stored, as well as their corresponding consequents, in addition to their order. Hence, to obtain the context only a simple query must be done, and so the context identification supposes no significative overhead in the access process.

## 4. Pertinence

Once the set of considered contexts is defined it is necessary to identify the relevant information for each one. The relevance of a concrete data group for a given context is what we call *pertinence*: the more needed or interesting the data group is for the context, the higher is its pertinence to the context.

There is a great variety of factors to consider to compute this pertinence like:

- The regulations about each clinical process. Usually the information relevant for each act of some pathologies (not of all), is fixed by protocols set by governmental institutions, by the hospitals or by the medical services.
- The opinion of the concrete doctor. In addition to the regulations, each doctor can consider that, from his/her point of view, there are other items that must also be taken into account.

- The own history of a concrete patient. Some data groups without significance for the majority of the patients, may have a great and especial importance for a given patient.
- The aging of the information. With time there are tests that loose their validity, because they are too old or because there are new tests of the same type.
- The access patterns. It is possible that the medical staff starts to access frequently a concrete data group for a given situation, and that they are not informed or "conscious" of it, so they do not include it through any of the previous ways. Taking into account this aspect of the pertinence, new patterns of access can be discovered and also the system can also automatically adapt to them.

The system must be capable of representing and bringing together all these aspects of the pertinence. The way we propose to do it is shown in next sections, where we present a method to capture and calculate each of these aspects of the pertinence.

### 4.1. Static pertinence: regulations, doctors and patients

As static pertinence we understand those set by medical criteria or given by the medical staff. Therefore, we consider three types of static pertinence, corresponding to three of the its aspects mentioned above.

On one hand, the medical criteria and regulations that determine which information must be always taken into account for a given process or pathology.

On the other hand, there personal opinions or even research studies of the doctors, that lead them to find a specific information item especially relevant for all the patients they see in a concrete *context*.

In addition, must be considered the concrete data group is particularly important for a given patient but not for the rest.

Hence, we include in the system three degrees of pertinence defined by doctors or medical criteria: one associated to the regulations $\left(P^{R}_{Dc} \in [0,1]\right)$, another one related to the personal opinion of the doctor $\left(P^{C}_{Dc} \in [0,1]\right)$ and the other one associated to the specific patient $\left(P^{P}_{Dc} \in [0,1]\right)$.

### 4.2. Time pertinence

The pertinence of a group of data (and the implicit document) will depend too on the date of creation. It is logic that the results of an analysis will be more important if it was completed a few days before than if it was performed a year ago.

However the influence of the age will not be the same for all document types: some type of analysis may be valid for several months meanwhile others are valid for years.

Hence we propose to modify the pertinence of the document depending on a established age threshold in months. If the document is younger than the threshold we want the time pertinence to be high (value grater than 0.7). If it is older, we want the time pertinence to decrease and give a low value.

To fulfill this restriction we propose the next definition.

**Definition 1.** Been $D$ a document and $A$ the age of the it (express in months), we calculate the *Time pertinence of document D* as

$$P_T(D) = e^{-\frac{\log_B(A)}{e}} \quad (1)$$

where $B \in ]1,+\infty]$ is a parameter defining the decreasing strength.

Fig. 4 shows the behavior of the function according to the value $B$. Let note that the value of $B$ is the point where the function has the first value under 0.7 (high pertinence).

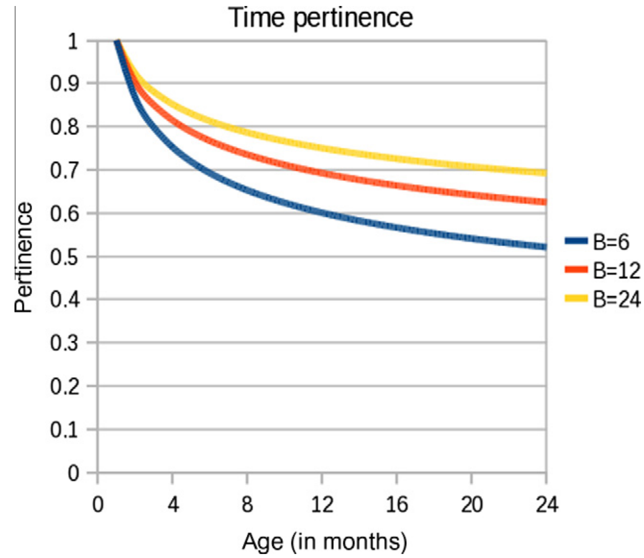


**Fig. 4.** Behavior of the time pertinence $P_T(X)$, according to different values of $B$.

### 4.3. Dynamic pertinence

In most of the situations the doctors will not give a static pertinence for a document neither a patient. So we need to learn the pertinences. Hence we propose to do it according to the accesses stored in the RADB database.

Due to the great number of records in the RADB database we need a very efficient process. If we consider that we want the system to be dynamic and to update on-line the pertinences according to the new accesses, the efficiency requirement is even more important.

As we have mentioned, different methods to calculated the relevance of preferences can be found in literature but the great complexity that they have, makes them not valid for our system. It has lead us to propose the new method explained next.

To calculate to pertinence we propose to use an adaptation of the Vector Space Model Salton, Wong, and Yang (1975). This technique comes from the Documentary Computing, concretely from the automatic indexation methods and retrieval systems Gil-Leyva and Rodríguez-Muñoz (1966). It is used to determine which descriptors are more specific or discriminate better between documents.

The discrimination value classifies terms in the text according to their capability to distinguish some documents from others in a given collection; i.e., the discrimination value of a term depends on how the average distance between the documents changes when a content identification is set for the term. Therefore, the best words are those resulting in a higher distance.

The basic idea of this model lays in the construction of a matrix or table of information items and documents, where the rows are the terms and the columns correspond to the documents acceded.

The rows would correspond to the terms that would be expressed according to the occurrences (access frequency) of each information item.

Applying it to our case, we consider as documents (columns) the possible *Contexts* and as terms (rows) the *data groups* inside the documents. Hence, the table with the access frequencies will be like the one show in Fig. 5, where $tf_{ij}$ represents the number of accesses to the *data group i* in the *context j*, and

$$tf_j = \sum_{i=1}^{N} tf_{ij} \quad (2)$$

gives information about the total accesses for *context j*.

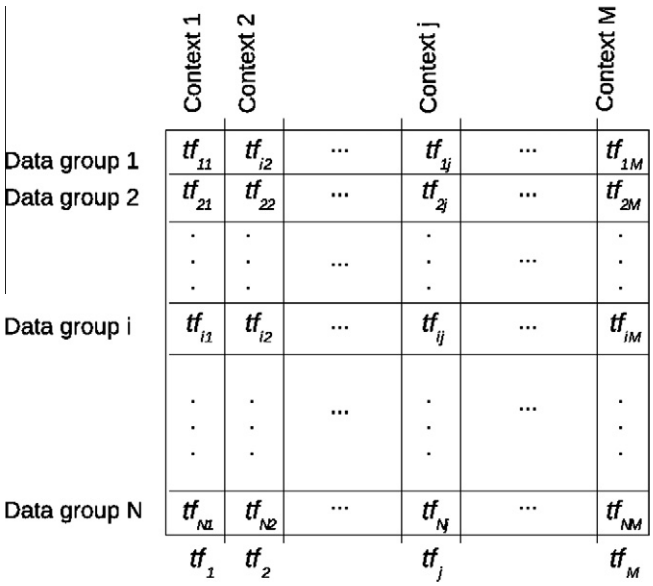


**Fig. 5.** Frequency table for *data groups* and *Contexts*.

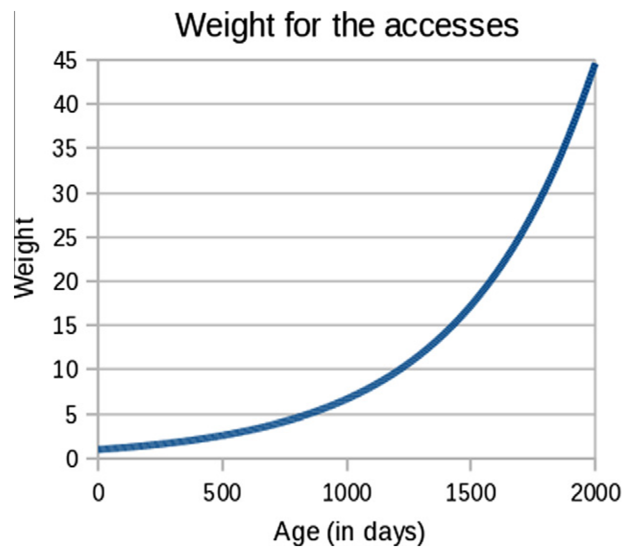


**Fig. 6.** Evolution of the influence according to the distance in days to the reference date.

However in our situation it is not enough, since we need the recent accesses have to a higher influence than older ones when calculating the pertinence. This is why we propose to measure the relevance according to the time as the weight function shown in Fig. 6. Let $D_R$ be a reference date to consider relevant or not the information for the system and $D_A$ represent the access date. In that case, we propose the following function to calculate the weight for a given date ($D_A$):

$$W(D_A) = 2^{\frac{D_A - D_R}{365}} \tag{3}$$

where the date difference is calculated in days. This way an access made today will have more influence that the accesses of the last year but, as the time goes by, less influence than future accesses. The equation establishes that given two accesses with a year of difference, the newer one will have double influence than the older one.

This definition introduces in the system two important and useful capabilities:

- The pertinences will be updated according to the aging of the access and the decreasing relevance.
- The system will be adapted automatically to future accesses patterns and needs, that can even allow us to define new contexts.

Then the system will update the pertinence of the data groups having more influence the newer accesses and enable the system to adapt to future needs.

As shown in Fig. 6, the influence is an increasing function. This may introduce some problems of representation or loss of precision when the values are very high. The definition of the influence that we have proposed allows us to avoid this problem in an easy way. We only have to move the reference date ($D_R$) and adapt the stored values to get an easy and quick adaptation: if we add one year to the reference date and divide all the values by 2, we get the same frequency and we reduce the magnitude of the stored values.

We can repeat this process as many time as needed as long as the final reference date is previous to the next access to be stored. The proposed system will do this each new year keeping the reference day with one year length to actual date. The update of the values only need to change the reference date (one update sentence) and the stored values (a set of very simple update sentences) which would need just a short time.

Now we have the frequency, we propose an adaptation of the *inverse document frequency* of the Vector Space Model (Salton et al., 1975) to measure the pertinence of a data group to a context, based on the information stored in the RADB database.

**Definition 2.** Let $C$ be a context and $X$ a data groups, the *restrospective pertinence* is

$$P_R^C(X) = \left[\frac{tf_{XC}}{tf_C}\right]^{1/4} \tag{4}$$

The idea behind this pertinence is to consider relevant a *data group* if the number of accesses to it is high in comparison to the total number of accesses.

## *4.4. Global pertinence*

Once we have the different considered aspects about the pertinency and defined a way to compute them, we need to aggregate the information given by them into a single value. To do it, we obtain a global pertinence of a data group to a given context as in next definition.

**Definition 3.** Let $X$ be a *group of data* in a document $D$, and $C$ a *context*, we define the *global pertinence of X to C* as

$$P_G^C(X) = \left(P_{Dc}^R(X) \oplus P_{Dc}^C(X) \oplus P_{Dc}^P(X) \oplus P_R^C(X)\right) \otimes P_T(D) \tag{5}$$

where

- $P_{Dc}^R$ is the pertinence set by medical doctors according to the regulations,
- $P_{Dc}^C(X)$ is the pertinence set by medical doctors for the data group to the context under their personal point of view,
- $P_{Dc}^P(X)$ is the pertinence set by medical doctors for the data group to a given patient,
- $P_R^C(X)$ the retrospective pertinence according to prior accesses,
- $P_T(X)$ the pertinence considering the age of the document,
- $\oplus$ and $\otimes$ a t-conorm and a t-norm, respectively.

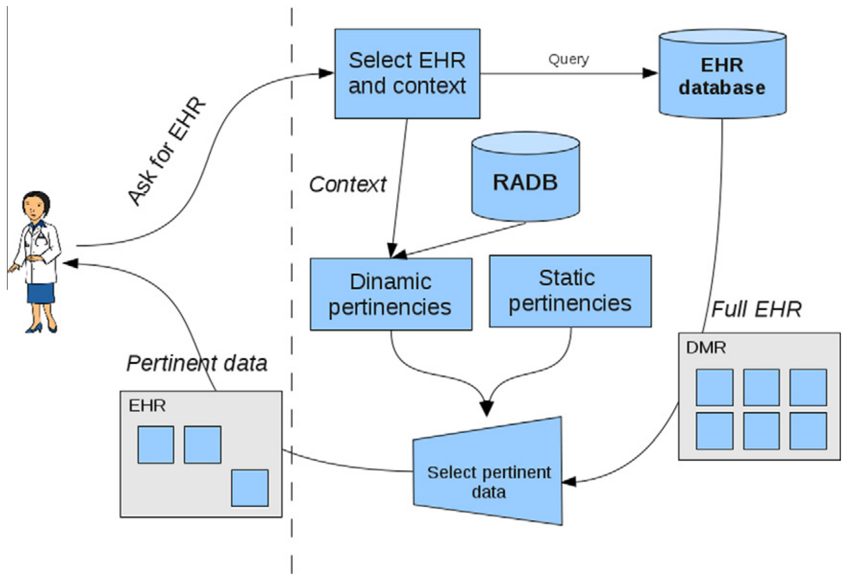


**Fig. 7.** Scheme of the contextualized query process.

For the system we have chosen the *maximum* and the *minimum* as t-conorm and t-norm because of their simplicity, and therefore, efficient and fast calculation as well as they are quite extended.

Hence, we include in the system three degrees of pertinence defined by doctors or medical criteria: one associated to the regulations $\left(P^R_{Dc} \in [0,1]\right)$, another one related to the personal opinion of the doctor $\left(P^C_{Dc} \in [0,1]\right)$ and the other one associated to the specific patient $\left(P^P_{Dc} \in [0,1]\right)$.

## 5. Contextualized access system

With all the elements to implement the contextualized access to the EHR, we show next how we propose to provide this access by presenting the use of the proposed method, as well as the update process that allows the system to automatically adapt to new needs.

### 5.1. Access to the system

An scheme of the access process in shown in Fig. 7.

- The doctor starts the process by logging in the system.
- Using the information about the terminal and the schedule of the doctor, the system gets the *context* for this access using the simple rule system.
- The doctor identifies the patient in the system to access his/her EHR.
- The system gets the EHR and queries the static and dynamic pertinences for all the data groups that appear in his/her EHR. The result of aggregating these pertinences and the time pertinence as shown in Eq. (5) is used to order the data.
- Finally, the system selects the first data groups and returns them to the doctor ordered y priority, as well as a way to access the other data groups if the doctor needs them.

### 5.2. Update process

The system is updated on each access so the pertinences are adapted continually to reflect doctors' needs. In this process no manual intervention is needed and a few records are changed so it needs a short time to be executed. The update process is as follow:

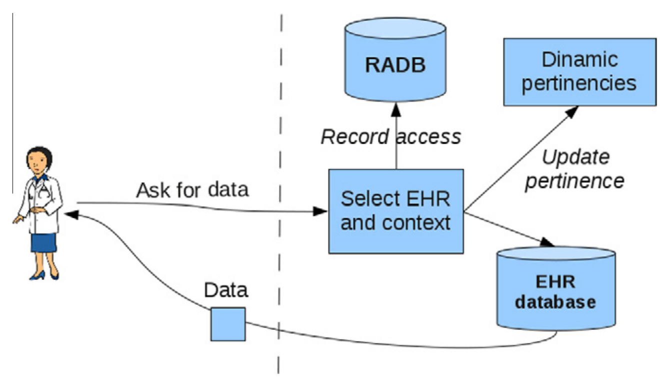


**Fig. 8.** Scheme for pertinences update process.

- When the doctor logs in the system, the context of the access is calculated as mentioned above.
- The doctor asks for a data group of a specific patient.
- The system then gets the required data and returns them to the doctor. At the same time the system logs the access in the RADB table and updates the frequency table used to obtain the *retrospective pertinence* with this new access. Only two records are changed: the accesses to the data group in this particular context ($tf_{ij}$) and the total number of accesses to the context ($tf_j$).

A scheme summarizing of the process is shown in Fig. 8.

After every access to the system, the dynamic pertinence is automatically updated so the next time a context is acceded the pertinence of the data groups to the context is computed considering the most updated information. As can be seen it is done with a very low computation cost. In addition this updating process allows, not only to give the most updated information, but also to discover new patterns of accesses, which give us the chance to define new contexts.

## 6. Example

In this section we show some examples to clarify the proposed method. Here we present two contexts and a set of five different documents with several data groups, and indicate how we compute their pertinence to each context. In a real case the number of contexts as well as the number of documents and data groups will be considerably much higher, as shown in Sections 3.1 and 2. With these very simplified examples our aim is jut to clarify and show the correspondence between the formal notions presented here and the medical terminology. The set of documents and data groups, are shown in Table 1, and the context selected are only two:

- *C*1: Emergency after catheterization surgery
- *C*2: Traumatology pre-surgery appointment.

In Table 2 the previous accesses are shown (a simplification of the RADB), as well as the associated frequency table is shown in Table 3.

**Table 1**
Documents and data groups contained on each of them. *B* is the parameter to define the decreasing strength.

| Code | Document types | Data groups | *B* |
|---|---|---|---|
| DT1 | Electrocardiogram | g1 | 3 |
| DT2 | Blood Analysis | Coagulation (g2) | 3 |
| DT2 | " | Immunology (g3) | 3 |
| DT2 | " | Biochemistry (g4) | 3 |
| DT3 | Discharge report | g5 | 24 |
| DT4 | Thorax radiography | g6 | 6 |
| DT5 | Surgery report | g7 | 24 |

In the next section we present in detail two examples for two different patients.

### 6.1. Examples of retrieval

#### 6.1.1. Patient 1

In the first example we consider a patient that recently had a catheter surgery. The last time he came to the hospital was for

**Table 2**
Recorded accesses to data groups in EHRs ($W(D)$ is the weight of the access according to the date as shown in Fig. 6 considering $D_R$ = 01/01/2008).

| Date | Context | Data group | $W(D)$ | Date | Context | Data group | $W(D)$ |
|---|---|---|---|---|---|---|---|
| 22/01/08 | c1 | g7 | 1.04 | 24/05/09 | c1 | g4 | 2.63 |
| 16/01/08 | c1 | g6 | 1.03 | 21/05/09 | c1 | g7 | 2.61 |
| 16/01/08 | c2 | g6 | 1.03 | 26/05/09 | c2 | g3 | 2.64 |
| 24/02/08 | c1 | g7 | 1.11 | 26/06/09 | c1 | g7 | 2.80 |
| 03/02/08 | c1 | g7 | 1.06 | 18/06/09 | c2 | g4 | 2.76 |
| 22/02/08 | c1 | g6 | 1.10 | 01/06/09 | c2 | g1 | 2.67 |
| 20/03/08 | c2 | g1 | 1.16 | 17/07/09 | c1 | g3 | 2.91 |
| 03/03/08 | c1 | g7 | 1.12 | 23/07/09 | c2 | g6 | 2.95 |
| 26/03/08 | c2 | g1 | 1.18 | 06/07/09 | c1 | g5 | 2.85 |
| 19/04/08 | c1 | g7 | 1.23 | 08/08/09 | c2 | g1 | 3.04 |
| 04/04/08 | c1 | g7 | 1.20 | 28/08/09 | c1 | g5 | 3.15 |
| 21/04/08 | c1 | g6 | 1.23 | 22/08/09 | c1 | g4 | 3.12 |
| 14/05/08 | c1 | g3 | 1.29 | 18/09/09 | c1 | g3 | 3.28 |
| 01/05/08 | c2 | g6 | 1.26 | 13/09/09 | c1 | g7 | 3.25 |
| 22/05/08 | c2 | g6 | 1.31 | 07/09/09 | c1 | g3 | 3.22 |
| 01/06/08 | c1 | g7 | 1.33 | 02/10/09 | c2 | g6 | 3.37 |
| 27/06/08 | c2 | g2 | 1.40 | 10/10/09 | c1 | g2 | 3.42 |
| 03/06/08 | c2 | g3 | 1.34 | 07/10/09 | c1 | g1 | 3.40 |
| 25/07/08 | c1 | g6 | 1.48 | 06/11/09 | c1 | g7 | 3.60 |
| 25/07/08 | c1 | g7 | 1.48 | 08/11/09 | c1 | g5 | 3.62 |
| 11/07/08 | c1 | g7 | 1.44 | 05/11/09 | c1 | g7 | 3.60 |
| 16/08/08 | c1 | g7 | 1.54 | 06/12/09 | c1 | g5 | 3.81 |
| 17/08/08 | c2 | g2 | 1.54 | 17/12/09 | c1 | g7 | 3.90 |
| 26/08/08 | c2 | g6 | 1.57 | 19/12/09 | c1 | g5 | 3.91 |
| 05/09/08 | c2 | g1 | 1.60 | 18/01/10 | c1 | g5 | 4.14 |
| 12/09/08 | c1 | g3 | 1.62 | 10/01/10 | c1 | g5 | 4.08 |
| 08/09/08 | c2 | g1 | 1.61 | 07/01/10 | c1 | g5 | 4.05 |
| 12/10/08 | c1 | g6 | 1.72 | 25/02/10 | c1 | g3 | 4.45 |
| 11/10/08 | c2 | g6 | 1.71 | 14/02/10 | c2 | g6 | 4.36 |
| 23/10/08 | c2 | g2 | 1.75 | 04/02/10 | c2 | g2 | 4.27 |
| 06/11/08 | c1 | g3 | 1.80 | 15/03/10 | c2 | g4 | 4.60 |
| 04/11/08 | c2 | g7 | 1.79 | 16/03/10 | c1 | g7 | 4.61 |
| 03/11/08 | c1 | g3 | 1.79 | 28/03/10 | c1 | g5 | 4.72 |
| 18/12/08 | c1 | g7 | 1.95 | 20/04/10 | c2 | g3 | 4.93 |
| 16/12/08 | c1 | g3 | 1.94 | 01/04/10 | c1 | g3 | 4.75 |
| 28/12/08 | c1 | g7 | 1.99 | 28/04/10 | c2 | g2 | 5.00 |
| 24/01/09 | c1 | g7 | 2.09 | 09/05/10 | c1 | g3 | 5.11 |
| 08/01/09 | c1 | g7 | 2.03 | 09/05/10 | c2 | g4 | 5.11 |
| 24/01/09 | c1 | g7 | 2.09 | 26/05/10 | c2 | g2 | 5.28 |
| 19/02/09 | c1 | g3 | 2.20 | 07/06/10 | c2 | g2 | 5.40 |
| 19/02/09 | c1 | g1 | 2.20 | 24/06/10 | c1 | g5 | 5.58 |
| 25/02/09 | c1 | g3 | 2.22 | 17/06/10 | c2 | g2 | 5.50 |
| 01/03/09 | c1 | g5 | 2.24 | 11/07/10 | c2 | g5 | 5.76 |
| 21/03/09 | c1 | g3 | 2.33 | 06/07/10 | c2 | g2 | 5.71 |
| 08/03/09 | c2 | g1 | 2.27 | 28/07/10 | c1 | g7 | 5.95 |
| 01/04/09 | c2 | g7 | 2.38 | 16/08/10 | c1 | g5 | 6.17 |
| 07/04/09 | c2 | g3 | 2.40 | 03/08/10 | c2 | g2 | 6.02 |
| 25/04/09 | c1 | g7 | 2.49 | 02/08/10 | c1 | g5 | 6.01 |

**Table 3**
Frequency table.

| Document types | Data group | Acum | | Frequency | | $P_C^C$ | |
|---|---|---|---|---|---|---|---|
| | | C1 | C2 | C1 | C2 | C1 | C2 |
| DT1 | g1 | 5.60 | 13.53 | 0.03 | 0.13 | 0.42 | 0.60 |
| DT2 | g2 | 3.42 | 41.88 | 0.02 | 0.39 | 0.37 | 0.79 |
| ” | g3 | 38.93 | 11.31 | 0.22 | 0.11 | 0.69 | 0.57 |
| ” | g4 | 5.75 | 12.47 | 0.03 | 0.12 | 0.42 | 0.58 |
| DT3 | g5 | 54.33 | 5.76 | 0.31 | 0.05 | 0.74 | 0.48 |
| DT4 | g6 | 13.24 | 17.56 | 0.07 | 0.16 | 0.52 | 0.64 |
| DT5 | g7 | 55.53 | 4.17 | 0.31 | 0.04 | 0.75 | 0.44 |
| | Total | 176.80 | 106.68 | | | | |

**Table 4**
Electronic health record of patient 1.

| Document | Creation date | Document type |
|---|---|---|
| 1 | 01/08/2010 | DT1 |
| 2 | 07/03/2008 | DT1 |
| 3 | 31/07/2010 | DT5 |
| 4 | 02/08/2010 | DT3 |
| 5 | 01/08/2010 | DT2 |
| 6 | 15/06/2010 | DT2 |
| 7 | 15/06/2010 | DT4 |

the post-surgery review but know he is feeling really bad (much pain in the breast). He comes to the Emergency service and a resident doctor in the fifth year (R5) is going to check up him. Now they are in the emergency room and the doctor wants to access the patient's EHR.

The first stage of the approach is to identify the access context. According to the information about the attributes used to identify the context, the first rule that is satisfied is:

```
IF Speciality = Cardiology AND
Last_visit=‘‘cateterismo post-surgery review’’ AND
Unit=‘‘Emergency’’
THEN Context=‘‘Emergency after cateterismo surgery’’ (C1)
```

The electronic health record for this patient is in Table 4. Let us note that the EHR has several documents and some of them are of the same type but with different age. We now consider an access from each of the two contexts mentioned above, and show the pertinent data groups in each case.

The first step is always to keep just the newer document of each type. Hence, the documents selected to work with are 1, 3, 4, 5, and 7, whereas documents 2 and 6 are discarded.

With this subset of documents we calculate the different pertinences for each of the data groups to the context $C1$. Table 5 shows the values for each type and the global pertinence in the column $P_G^{C1}$, applying the Definition 3.

The next step is to order the data groups according to the global pertinence values. The last column of the table collects the ranking. If we make blocks of 4 data groups to create levels of preference, the result for the query will include the data groups *Doc*3.*g*7, *Doc*4.*g*5, *Doc*5.*g*3, and *Doc*7.*g*6. Whereas, *Doc*5.*g*4, *Doc*1.*g*1 would be in the next block of data groups. The number of considered data groups (size of the blocks) may change according, as an example, to the capacities of the medical workstations (in a PDA four data groups may be enough because of limitation of the screen but in a PC the number could be greater).

Before returning this set to the user, we refine the answer to determine if in some concrete case it is preferable to show the entire document instead of its pertinent data groups. We do it if more than a given percentage of the data groups inside a document are selected in the answer. In such case we consider the rank of the most pertinent data group. As an example, if the percentage is set to the 60%, the documents *Doc*3, *Doc*4, *Doc*7, that only contain one data group and it has been found pertinent, will replace in the solution the pertinent data group. However *Doc*5 has three data groups from which only one has been found pertinent. Therefore, in the final solution this data group will be kept. This way, the initial solution {*Doc*3.*g*7,*Doc*4.*g*5, *Doc*5.*g*3,*Doc*7.*g*6} will be replaced by {*Doc*3,*Doc*4, *Doc*5.*g*3,*Doc*7}.

Finally we would like to remark that in the real system we establish several preference levels, by grouping the pertinent data groups into sets of a fixed size, 10. If the information needed does not appear within the first block of 10 data groups, the next set is shown; and this process is repeated until the information is found. In addition, in any moment the user can access the complete EHR by the traditional navigation system.

To show the differences lets suppose that the appointment is in a different situation and the inferred context is $C2$. In that case the pertinences are different, as shown in Table 6. The data groups

**Table 5**
Pertinences for patient 1's EHR for context 1.

| Doc. | Date | Document type | Data group | $P_{D_C}^{C1}$ | $P_{D_C}^{P}$ | $P_T$ | $P_C^{C1}$ | $P_G^{C1}$ | Rank |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 01/08/10 | DT1 | g1 | 0.00 | 0.00 | 0.79 | 0.42 | 0.42 | 6 |
| 3 | 31/07/10 | DT5 | g7 | 0.00 | 0.00 | 0.92 | 0.75 | 0.75 | 1 |
| 4 | 02/08/10 | DT3 | g5 | 0.00 | 0.00 | 0.92 | 0.74 | 0.74 | 2 |
| 5 | 01/08/10 | DT2 | g2 | 0.00 | 0.00 | 0.79 | 0.37 | 0.37 | 7 |
| ” | ” | ” | g3 | 0.00 | 0.00 | 0.79 | 0.69 | 0.69 | 3 |
| ” | ” | ” | g4 | 0.00 | 0.00 | 0.79 | 0.42 | 0.42 | 5 |
| 7 | 15/06/10 | DT4 | g6 | 0.00 | 0.00 | 0.75 | 0.52 | 0.52 | 4 |

**Table 6**
Pertinences for patient 1's EHR for context 2.

| Doc. | Date | Document types | Data group | $P_{D_C}^{C2}$ | $P_{D_C}^{P}$ | $P_T$ | $P_C^{C2}$ | $P_G^{C2}$ | Rank |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 01/08/10 | DT1 | g1 | 0.00 | 0.00 | 0.42 | 0.60 | 0.60 | 3 |
| 3 | 31/07/10 | DT5 | g7 | 0.00 | 0.00 | 0.75 | 0.44 | 0.44 | 7 |
| 4 | 02/08/10 | DT3 | g5 | 0.00 | 0.00 | 0.74 | 0.48 | 0.48 | 6 |
| 5 | 01/08/10 | DT2 | g2 | 0.00 | 0.00 | 0.37 | 0.79 | 0.79 | 1 |
| ” | ” | ” | g3 | 0.00 | 0.00 | 0.69 | 0.57 | 0.57 | 5 |
| ” | ” | ” | g4 | 0.00 | 0.00 | 0.42 | 0.58 | 0.58 | 4 |
| 7 | 15/06/10 | DT4 | g6 | 0.00 | 0.00 | 0.52 | 0.64 | 0.64 | 2 |

**Table 7**
Electronic health record for patient 2.

| Document | Creation date | Document type |
|---|---|---|
| 1 | 01/09/2010 | DT4 |
| 2 | 01/09/2010 | DT1 |
| 3 | 25/08/2010 | DT2 |
| 4 | 01/07/2010 | DT3 |
| 5 | 15/06/2009 | DT2 |
| 6 | 20/06/2009 | DT5 |

pertinent for this answer would be the set $\{Doc5.g2, Doc7.g6, Doc1.g1, Doc5.g4\}$. In that case 2 of the three data groups in document 5 are pertinent. Therefore, the refined set would be $\{Doc5, Doc7, Doc1.g1\}$.

### 6.1.2. Patient 2

This patient has a fracture in a rib that needs surgery. Most of the preparatives have been done (pre-surgery analysis as blood test and cardiogram) and now he is at the doctor office to check that everything is right for the surgery. The doctor is a trauma facultative in his/her office accessing the EHR for the analisis results. The EHR for this patient is shown in Table 7. The patient has a chronic disease that makes his defenses be very low all the time. In this case we consider that by medical recommendation, one type of document groups (the defenses analysis inside the blood test) is specially important for that patient: the data group $g3$ in document type $DT6$. Hence, for this type we consider the *static pertinence given by doctor to this document type for this patient* $\left(P^P_{D_C}\right)$ with the value $P^P_{D_C} = 0.7$ that means this type of documents is especially important for this patient independently from the access context.

As in the previous case, we identify the context using the rule base stored in the system. The first rule that is satisfied in this case is the next:

```
IF Speciality = Trauma AND
Present_visit=''pre-surgery''
THEN Context=''trauma pre-surgery review'' (C2)
```

Under this assumption, an access from context $C2$ will result in the set of pertinences shown in Table 8. As in the previous example, we have first selected only one document of each type, considering the newer one if there are several for a type.

**Table 10**
Updated frequency table for access 1.

| Document type | Data group | Acum | $P_C^{C1}$ |
|---|---|---|---|
| DT1 | g1 | 5.6 | 0.42 |
| DT2 | g2 | 3.42 | 0.37 |
| ” | g3 | 38.93 | ***0.68*** |
| ” | g4 | 5.75 | 0.42 |
| DT3 | g5 | 54.33 | 0.74 |
| DT4 | g6 | 13.24 | 0.52 |
| DT5 | g7 | ***62.41*** | ***0.76*** |
| | Total | ***183.68*** | |

**Table 11**
Updated frequency table for access 2.

| Document types | Data group | Acum | $P_C^{C2}$ |
|---|---|---|---|
| DT1 | g1 | ***20.41*** | ***0.65*** |
| DT2 | g2 | 41.88 | ***0.78*** |
| ” | g3 | 11.31 | ***0.56*** |
| ” | g4 | 12.47 | 0.58 |
| DT3 | g5 | 5.76 | ***0.47*** |
| DT4 | g6 | 17.56 | ***0.63*** |
| DT5 | g7 | 4.17 | 0.44 |
| | Total | ***113.57*** | |

In this access, as Table 8 shows, the pertinent data groups found are $\{Doc6.g10, Doc3.g2, Doc3.g3, Doc1.g6, Doc2.g1\}$. If we refine the answer considering the complete documents, the returned set would be $\{Doc3, Doc1, Doc2\}$. In this answer the static pertinence has been very important. Without this information the final pertinence for $Doc3.g3$ would be 0.57 and only the data group $g2$ would have been returned.

If the access to patient 2's EHR would be done from context C1, the result would be different. Table 9 collects the pertinence values. Following the same process as in previous example, the answer of the system would be $\{Doc4, Doc6, Doc3.g3, Doc1\}$.

### 6.2. Example of update

In the previous sections we have shown the answer of the system when a doctor accesses an EHR. Once the doctor selects a

**Table 8**
Pertinences of patient 2's EHR for context 2.

| Doc. | Date | Document type | Data group | $P_{D_C}^{C2}$ | $P_{D_C}^{P}$ | $P_T$ | $P_C^{C2}$ | $P_G^{C2}$ | Rank |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 01/09/10 | DT4 | g6 | 0.00 | 0.00 | 1.00 | 0.64 | 0.64 | 3 |
| 2 | 01/09/10 | DT1 | g1 | 0.00 | 0.00 | 1.00 | 0.60 | 0.60 | 4 |
| 3 | 25/08/10 | DT2 | g2 | 0.00 | 0.00 | 0.79 | 0.79 | 0.79 | 1 |
| ” | ” | ” | g3 | 0.00 | 0.70 | 0.79 | 0.57 | 0.70 | 2 |
| ” | ” | ” | g4 | 0.00 | 0.00 | 0.79 | 0.58 | 0.58 | 5 |
| 4 | 01/07/10 | DT3 | g5 | 0.00 | 0.00 | 0.88 | 0.48 | 0.48 | 6 |
| 6 | 20/06/09 | DT5 | g7 | 0.00 | 0.00 | 0.73 | 0.44 | 0.44 | 7 |

**Table 9**
Pertinences for patient 2's EHR for context 1.

| Doc. | Date | Document type | Data group | $P_{D_C}^{C1}$ | $P_{D_C}^{P}$ | $P_T$ | $P_C^{C1}$ | $P_G^{C1}$ | Rank |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 01/09/10 | DT4 | g6 | 0.00 | 0.00 | 1.00 | 0.52 | 0.52 | 4 |
| 2 | 01/09/10 | DT1 | g1 | 0.00 | 0.00 | 1.00 | 0.42 | 0.42 | 5 |
| 3 | 25/08/10 | DT2 | g2 | 0.00 | 0.00 | 0.79 | 0.37 | 0.37 | 7 |
| ” | ” | ” | g3 | 0.00 | 0.70 | 0.79 | 0.69 | 0.70 | 3 |
| ” | ” | ” | g4 | 0.00 | 0.00 | 0.79 | 0.42 | 0.42 | 6 |
| 4 | 01/07/10 | DT3 | g5 | 0.00 | 0.00 | 0.88 | 0.74 | 0.74 | 1 |
| 6 | 20/06/09 | DT5 | g7 | 0.00 | 0.00 | 0.73 | 0.75 | 0.73 | 2 |

document the system updates automatically the dynamic pertinences. In this section we show two examples of these updates.

First, suppose that the system returns a set of documents and the user selects one of these documents. Considering the patient 1 and context *C*1, let's assume the *Doc*1.*g*2 is chosen, on 27th of September of 2010. The frequency table is updated and all the pertinences for this context are updated too. According to the date, the weight *W*(*D*) for *Doc*1.*g*2 is 6.68 and the new frequency table (with the corresponding dynamic pertinences) is shown in Table 10. The values that have changed are in italic. In this case two pertinences have changed and they will be taken into account for the future accesses.

If the user selects a document that is not in the data set, for example *Doc*2.*g*1 in *C*2, the process is similar and the result is shown in Table 11. In this case the pertinences for five of the data groups change too. The pertinence of this type of document (*Doc*2.*g*1) has increased to reflect the access to this document and the other four has been decrease.

## 7. Implementation

As we have mentioned before, in this kind of system the efficiency is very important. In this section we briefly comment the implementation for our proposal and we analyze the efficiency of the final system.

Next we explain the implementation of the two most relevant elements in the system, the rule base used to identify the context and the frequency table to select its pertinent documents, as well as the processes needed to access them.

### 7.1. Rule base

Each time a hospital's staff member accesses the EHR system, the first step is to identify the context of the access. As mentioned in Section 3.2, the rule base and the process to select the context have been implemented inside the data base so with just a single, simple and fast query the answer to this question is obtained.

The rule base is implemented using one table inside the database. Each record represents a rule, and we store the order learnt by RIPPER, the values for each attribute presented in Section 3.2, and the context. If one rule has no value for an attribute then a NULL value is stored. Therefore the table has 9 columns (one for the rule order, seven for the attributes and one for the context) and near two hundreds records.

When an access occurs, the context is identified by scanning the table looking for the first rule satisfied. The SQL sentence that gets the context in this way uses the 7 attributes of the access to build a select clause for Oracle DBMS as follows:

```
SELECT context FROM RuleBase WHERE
(Speciality=‘‘value1’’ or Speciality IS NULL) and
(position=‘‘value2’’ or position IS NULL) and...
and (last_visit=‘‘value7’’ or last_visit IS NULL)
ORDER BY ord ASC HAVING ROWNUM<=1
```

The efficiency of the query is improved defining indexes on the table over the seven attributes. Therefore the time needed to answer the query is very small.

### 7.2. Frequency table

The other table needed is the *frequency table*, used to know the pertinent documents for each context. This table is stored in the database too, with the following attributes:

- *DG*: internal code for each data group type.
- *Context*: Context
- *Weight*: Total weight for all the accesses to this data group type (*DC*) in this context (element $tf_{XC}$ in Eq. (4)).
- *RP*, *DcP*: the static pertinences (established by regulations and doctors) for the data group to the particular context respectively.

The primary key is (*DG*, *Context*). In the DG domain we include a new code ($-1$) that do not represent any data group type. This value will be used to store the aggregation of all the weights for all the data group types in a particular context. This value is the element $tf_C$ in Eq. (4). The number of records stored in the table is $|contexts| \times (|DG| + 1)$.

To improve the queries efficiency we define indexes over both attributes individually and together (3 indexes). To improve the query over the table we define clusters on the table (physical blocks to store related records) considering the context value (all the records storing the frequency of each data group type in a particular context are stored physically together).

#### 7.2.1. Query for the pertinent data groups

To know the frequency of the five most pertinent data group types to a concrete context we only need to execute the following select sentence:

```
SELECT DG, Weight FROM FrequencyTable WHERE
context = C
ORDER BY Weight DESC HAVING ROWNUM<=6
```

The first row will give the value of $tf_C$ and the five next rows will give the values $tf_{XC}$ for the five most pertinent data group types. Let us note that the query is very simple and needs a very low execution time.

The five most relevant data groups for a concrete patient in a particular context are obtained using the value of the function $P_G^C(X)$ and ordering the data groups according to its value. To speed up the process we have implemented the function inside the database using PL/SQL (the Oracle language for stored functions and procedures) as PGC. With these values the query has to join the frequency table and the EHR table, execute the function and order the data groups. Let be *pid* the ID for the patient, *C* the related context, *TOTAL* the value of $tf_C$, and *RP*, *DcP* and *PP* the static pertinence (regulations, doctor and patients respectively) of this context for this particular patient. Then the select query has the following structure:

```
SELECT *,
PGC (FrequencyTable.DcP,FrequencyTable.RP,DPP,
FrequencyTable.Weight,TOTAL,EHR.Date)
as Pertinence FROM FrequencyTable,EHR WHERE
EHR.PID = pid and
FrequencyTable.Context = C and
/* JOIN */
EHR.DG = FrequencyTable.DG
/* Order */
ORDER BY Pertinence DESC HAVING ROWNUM<=5
```

With one query to know the values of $tf_C$ (very simple) and this second query we get the pertinent data groups. All the computation is made inside the database server so this process does not introduce any computation overload on the terminal.

#### 7.2.2. Update of the frequency table

There are two processes that change the frequency table:

- Each time a data group is acceded (very often).
- Each time the reference date changes (once a year) the frequencies stored are updated.

For the first update is performed by executing just one sentence. Let be *id* the code for the data group, *C* the context and *W*(*A*) the weight for this new access. Then the update sentence is:

```
UPDATE FrequencyTable SET Weight = Weight + W(A)
WHERE DG in (id,−1) AND Context = C;
```

As can be seen this sentence is simple a very fast to execute since only two records are modified: the one related to the data group and the value of $tf_C$.

The second update will be done once a year (at most) and it will imply to modify all the records in *Frequency* table. As we have mentioned in Section 4.3, changing the reference date by one year the update sentence would be:

```
UPDATE FrequencyTable SET Weight = Weight/2;
```

Though this update is more time consuming than the previous one, it is assumable since it will only be executed once a year and just blocks one table for a very simple processing.

## 8. Results

To test our proposal we have designed the test explained next. For each context we have selected a set with the most frequent assistance acts. Then we have monitored these access for a month to get the data groups accessed on each case. Then we have performed the accesses using the new system and we have counted the number of times that all the information needed was present within the first selection of pertinent data groups, the number of occasions in which it has been needed to access the second set of pertinent data groups, the same with the third set, and so on. In the 81.56% of the accesses all the information requested was in the first set, in the 4.12% of the cases, the needs of information where satisfied with the second set of data groups, the 1.91% of the accesses needed to reach a set of data groups in the third, and finally in the remaining 12.41% of the cases it was necessary to reach at or over the fourth block.

To compare the performance of the proposed system regarding the previous one, we have compared the number of "clicks" required to obtain all the information needed in both systems (we compare the number of clicks and not the time, because in the navigation based systems the time depends on the skills of the user). To do it we have considered that each access to a set of data groups requires one click in the new system. In the old system, we have considered that each document acceded needed one click (to open the document) and that the search of each document at least needed three clicks (to select the type of assistance act, the concrete assistance act and the set of documents generated on it, as shown in Fig. 2). According to it, and taking as an example an assistance act that needs data groups of information contained in five different documents, with the traditional navigation systems 20 clicks would be needed; whereas with the new one, according to the previous percentages, in the 81.56% of the cases only one click is needed. Just with this data it can be appreciated the great gain of time that the proposed system offers.

Some possible improvements that can be made in the system are the next:

- Adapt the size of the sets of data groups showed in each priority level, according to the context. This way the contexts like a pre-surgery study, that usually need to access more than 10 data groups, would show in each level sets of 20 or 30 data groups; whereas other contexts could keep smaller sets, like a simple outpatient consultation. It would only require to add a column to the table of contexts.
- Considering that in Section 3.2 we mentioned that the context identification had a percentage of success of the 89.32% it is possible that some of the accesses over the third set of pertinent documents were due to a bad identification of the context. Considering that the training was made with only a participation of the 24%, improving the training with a higher participation the identification of the context could be better.

Finally, we must remark that the system is still on its development and improvement stage, so it is not still completely implanted. Anyway the doctors that have tested it have commented that it is "very useful", "quite comfortable" and, on top of it, "really very time saving".

## 9. Conclusions

In this paper we have proposed a new paradigm to access to the EHR systems, based on a contextualized access to the information, in such a way that the information is ordered according to their preference or relevance for the assistance act or pathologic process in which the medical staff in involved. It is therefore specifically thought to improve the availability of information for medical practice, satisfying their specific needs and offering with it a faster and more efficient access to the really relevant information for a concrete assistance act, avoiding the handling of a huge quantity of information superfluous or unnecessary for it.

We have also presented a method defined the contexts and also to identify them each time a doctor accesses the system, based on a rule system stored in a data base. In addition we have proposed a technique to efficiently define the pertinence of data groups to the context and update it on each new access, so the system can automatically adapt to changes on the access patterns and to the new doctor's needs of information.

In addition, we have shown some details of the implementation as well as options to improve the accessibility. We have also showed some examples of how the system works and the results obtained are quite encouraging as the statistics obtained in the test of the system remak, as well as the satisfied opinion of the medical staff that have used it.

With this proposal several new research lines have been opened, since we enable and make easier to provide the system with new capabilities. This is the case of the knowledge mobilization, where the contextualized access solves the limitation in the use of mobile devices to perform complex accesses to great volumes of information, since with this access the few data really needed for a concrete assistance act are easily available and visualized in this type of devices, being unnecessary the navigation through the EHR.

Another example of this advantage is the possibility of providing the system with new functionalities for research purposes (El Fadly et al., 2010), just defining the corresponding contexts. In addition it has open the possibility to provide the citizens with a personalized access to their own medical data which, as Charters (2009) and Ruland, Brynhi, Andersen, and Bryhni (2008) indicate, it is a growing demand. Moreover, this new access paradigm can be applied to enable the interoperability between health record systems, by using the contexts to define the archetypes that the ISO 13606 requires.

Finally, we would like to remark the viability of this proposal, that is proved with the actual programming line in the San Cecilio University Hospital from Granada.

## Acknowledgment

The research reported in this paper was partially supported by the Andalusian Government (Junta de Andalucía) under project P07-TIC03175 "Representacin y Manipulación de Objetos Imperfectos en Problemas de Integracin de Datos: Una Aplicación a los Almacenes de Objetos de Aprendizaje" and also by the Spanish Government (Science and Innovation Department) under project TIN2009-08296. We would also like to thank their collaboration to the medical personnel that is participating in the development of the system.

## References

Adams, A., Adams, R., Thorogood, M., & Buckingham, C. (2007). Barriers to the use of e-health technology in nurse practitioner-patient consultations. *Informatics in Primary Care, 15*(2), 103–109.

Bisbal, J., & Berry, D. (2009). An analysis framework for electronic health record systems. *Methods of Information in Medicine, 48*(1).

Bobillo, F., Delgado, M., & Gómez-Romero, J. (2008). Representation of context-dependant knowledge in ontologies: A model and an application. *Expert Systems with Applications, 35*, 1899–1908<http://portal.acm.org/citation.cfm?id= 1401266.1401463>.

Bohm, K., Wolf, P., & Krcmar, H. (2010). Context oriented structuring of egovernment services – An empirical analysis of the information demand of expatriates in germany. In *43rd Hawaii International Conference on System Sciences (HICSS), 2010* (pp. 1–10).

Buchholz, T., Hochstatter, I., & Linnhoff-Popien, C. (2007). Distribution strategies for the contextualized mobile internet. *Electronic Commerce Research and Applications, 6*(1), 40–52<http://www.sciencedirect.com/science/article/B6X4K-4K5HV8M-1/2/90c68f322aa155b9f62aaafa2917c25b>.

Cao, Y., Xu, J., Liu, T.-Y., Li, H., Huang, Y., & Hon, H.-W. (2006). Adapting ranking svm to document retrieval. In *SIGIR'06: Proceedings of the 29th annual international ACM SIGIR conference on research and development in information retrieval* (pp. 186–193). New York, NY, USA: ACM.

Cayir, S., & Nuri Basoglu, A. (2008). Information technology interoperability awareness: A taxonomy model based on information requirements and business needs. In *Portland International Conference on Management of Engineering Technology, 2008, PICMET 2008*. (pp. 846 –855).

Chaker, H., Chevalier, M., Soule-Dupuy, C., & Tricot, A. (2010). Improving information retrieval by modelling business context. In *Third international conference on advances in human-oriented and personalized mechanisms, technologies and services (CENTRIC), 2010* (pp. 117–122).

Charters, K. (2009). Challenges of electronic medical record extracts for a personal health record. *Studies in Health Technology and Informatics, 146*, 197–201.

Cho, I., Kim, J., Kim, J., Kim, H., & Kim, Y. (2010). Design and implementation of a standards-based interoperable clinical decision support architecture in the context of the korean ehr. *International Journal of Medical Informatics, 79*(9), 611–622.

Cho, I., Staggers, N., & Park, I. (2010). Nurses' responses to differing amounts and information content in a diagnostic computer-based decision support application. *Computers, Informatics, Nursing: CIN, 28*(2), 95–102.

Chu, W., Johnson, D., & Kangarloo, H. (2000). A medical digital library to support scenario and user-tailored information retrieval. *IEEE Transactions on Information Technology in Biomedicine, 4*(2), 97–107.

Cohen, W. W. (1995). Fast effective rule induction. In *Proceedings of the twelfth international conference on machine learning* (pp. 115–123). Morgan Kaufmann.

Collins, B., & Speedie, S. (2008). Attaching context sensitive infobuttons to an ehr options and issues. AMIA. In *Annual symposium proceedings/AMIA symposium AMIA Symposium 914*.

Data exchange standard (2011). <http://www.hl7.org>. http://<www.HL7.org>.

Del Fiol, G., & Haug, P. (2009). Classification models for the prediction of clinicians' information needs. *Journal of Biomedical Informatics, 42*(1), 82–89.

Ehsan, M., Amini, M., & Jalili, R. (2009). Handling context in a semantic-based access control framework. In *International conference on advanced information networking and applications workshops, 2009, WAINA'09* (pp. 103 –108).

El Fadly, A., Lucas, N., Rance, B., Verplancke, P., Lastic, P.-Y., & Daniel, C. (2010). The reuse project: Ehr as single datasource for biomedical research. *Studies in Health Technology and Informatics, 160*(Pt. 2), 1324–1328<http://www.biomedsearch.com/nih/REUSE-project-EHR-as-single/20841899.html>.

Erdal, S., Catalyurek, U., Payne, P., Saltz, J., Kamal, J., & Gurcan, M. (2009). A knowledge-anchored integrative image search and retrieval system. *Journal of Digital Imaging: the Official Journal of the Society for Computer Applications in Radiology, 22*(2), 166–182.

Flores Zuniga, A., Win, K., & Susilo, W. (2010). Functionalities of free and open electronic health record systems. *International Journal of Technology Assessment in Health Care, 26*(4), 382–389.

Gagnon, M., Desmartis, M., Labrecque, M., Lgar, F., Lamothe, L., Fortin, J., et al. (2010). Implementation of an electronic medical record in family practice: A case study. *Informatics in Primary Care, 18*(1), 31–40.

Garcia-Morchon, O., & Wehrle, K. (2010). Efficient and context-aware access control for pervasive medical sensor networks. In *8th IEEE international conference on pervasive computing and communications workshops 2010 (PERCOM Workshops)* (pp. 322–327).

Gil-Leyva, I., & Rodríguez-Mu ñoz, J. (1966). Tendencias en los sistemas de indizacin automtica. estudio evolutivo. *Revista Espaola de Documentacion Cientffica, 19*(3), 273–291.

Ginsburg, M. (2007). Pediatric electronic health record interface design: The pedone system. In *40th annual hawaii international conference on system sciences, 2007, HICSS 2007* (pp. 139).

Image storage standard. (2011). <http://www.medical.nema.org>.

ISO-13606 (2008). Iso 13606: Electronic health record communication.

Järvelin, K., & Kekäläinen, J. (2000). Ir evaluation methods for retrieving highly relevant documents. In *SIGIR'00: Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 41–48). New York, NY, USA: ACM.

Jerding, D., & Stasko, J. (1998). The information mural: a technique for displaying and navigating large information spaces. *IEEE Transactions on Visualization and Computer Graphics, 4*(3), 257–271.

Jones, K. S. (2004). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation, 60*, 493–502.

Judd, G., & Steenkiste, P. (2003). Providing contextual information to pervasive computing applications. In *Proceedings of the First IEEE international conference on pervasive computing and communications, 2003. (PerCom 2003)* (pp. 133–142). doi:10.1109/PERCOM.2003.1192735.

Jung, J. J. (2009). Contextualized query sampling to discover semantic resource descriptions on the web. *Information Processing & Management, 45*(2), 280–287<http://www.sciencedirect.com/science/article/B6VC8-4V8FFJY-/2/6f9094e73f868971acee4f7cfb78f225>.

Kang, D., Lee, H., Ko, E., Kang, K., & Lee, J. (2006). A wearable context aware system for ubiquitous healthcare. In *Conference proceedings: Annual international conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in medicine and biology society conference* (Vol. 1, pp. 5192–5195).

Kanoulas, E., Pavlu, V., Dai, K., & Aslam, J. (2010). Modeling the score distributions of relevant and non-relevant documents. *Advances in Information Retrieval Theory*, 152–163.

Karahoca, A., Bayraktar, E., Tatoglu, E., & Karahoca, D. (2010). Information system design for a hospital emergency department: A usability analysis of software prototypes. *Journal of Biomedical Informatics, 43*(2), 224–232.

Lahteenmaki, H., Leppanen, J., & Kaijanranta, J. (2009). Interoperability of personal health records. In *Conference proceedings: Annual international conference of the IEEE Engineering in Medicine and Biology Society. IEEE engineering in medicine and biology society conference* (p. 1726-9).

Mao, J.-Y., & Benbasat, I. (2001). The effects of contextualized access to knowledge on judgement. *International Journal of Human–Computer Studies, 55*(5), 787–814<http://www.sciencedirect.com/science/article/B6WGR-4582BWS-7/2/cb7b598f793b5017ca950d7e785f369c>.

McAlearney, A., Robbins, J., Hirsch, A., Jorina, M., & Harrop, J. (2010). Perceived efficiency impacts following electronic health record implementation: An exploratory study of an urban community health center network. *International Journal of Medical Informatics, 79*(12), 807–816.

Nagy, M., Hanzlfcek, P., Preckov, P., Rfha, A., Dioszegi, M., Seidl, L., & Zvrov, J. (2010). Semantic interoperability in czech healthcare environment supported by hl7 version 3. *Methods of Information in Medicine, 49*(2), 186–195<http://www.ncbi.nlm.nih.gov/pubmed/19936441>.

Nagy, M., Preckova, P., Seidl, L., & Zvarova, J. (2010b). Challenges of interoperability using hl7 v3 in czech healthcare. *Studies in Health Technology and Informatics, 155*, 122–128.

Nomenclature medical standard, S., 2011. <http://www.nlm.nih.gov/research/snomad/snomad_main.html>.

Open-EHR open electronical health records, 2011. <http://www.openehr.org>. <www.openehr.org>.

Prados, M., & Peña, M. (2003). Sistemas de Informacion hospitalarios. Organizacion y gestion de Proyectos. EASP.

Prados de Reyes, M., Carmen Peña Yáñez, M. A. V. M., & Suárez, M. B. P. (2006). Generation and use of one ontology for intelligent information retrieval from electronic record histories.

Prados, M., Peña, M., Prados, B., Martinez, B., Ortigosa, J., & Delgado, A. (2010). Electronical health record (ehr) representation through ontology, mobility, accesibility and interoperability usefulness.

Prados-Suárez, B., Revuelta, E., Peña Yañez, C., Molina Fernàndez, C. (2008). Ontology based semantic representation of the reports and results in a hospital information system. In *Proceedings of the ICEIS 2008* (pp. 300–306).

Pung, H. K., Gu, T., Xue, W., Palmes, P. P., Zhu, J., Ng, W. L., Tang, C. W., & Chung, N. H. (2009). Context-aware middleware for pervasive elderly homecare. *IEEE Journal on Selected Areas in communications*, Institute of Electrical and Electronics Engineers Inc., The 27, pp. 510–524.

Ray, J. & Wimalasiri, P. (2006). The need for technical solutions for maintaining the privacy of ehr. In *Conference proceedings: Annual international conference of the IEEE Engineering in Medicine and Biology Society. IEEE engineering in medicine and biology society conference* (Vol. 1, pp. 4686–4689).

Ross, S. (2009). Results of a survey of an online physician community regarding use of electronic medical records in office practices. *The Journal of Medical Practice Management: MPM, 24*(4), 254–256.

Ruland, C., Brynhi, H., Andersen, R., & Bryhni, T. (2008). Developing a shared electronic health record for patients and clinicians. *Studies in Health Technology and Informatics, 136*, 57–62.

Salton, G., Wong, A., & Yang, C. S. (1975). A vector space model for automatic indexing. *Communications of the ACM, 18*(11), 613–620.

Svanaes, D., Das, A., & Alsos, O. (2008). The contextual nature of usability and its relevance to medical informatics. *Studies in Health Technology and Informatics, 136*, 541–546.

Vest, J., & Jasperson, J. (2010). What should we measure? Conceptualizing usage in health information exchange. *Journal of the American Medical Informatics Association: JAMIA, 17*(3), 302–307.

Vishwanath, A., Singh, S., & Winkelstein, P. (2010). The impact of electronic medical record systems on outpatient workflows: A longitudinal evaluation of its workflow effects. *International Journal of Medical Informatics, 79*(11), 778–791.

Weinstock, M. (2010). For hospitals and meaningful use, context is everything. *Hospitals & Health Networks AHA, 84*(8), 20–21.