Towards a Case-Based Reasoning Method for openEHR-Based Clinical Decision Support

Hajar Kashfi and Jairo Robledo Jr.

1 Department of Applied Information Technology
Chalmers University of Technology
SE–412 96, Gothenburg, Sweden
hajar.kashfi@chalmers.se

2 Department of Oral Medicine and Pathology
Institute of Odontology
Sahlgrenska Academy
University of Gothenburg
SE–405 30, Gothenburg, Sweden
jairo.robledo@gu.se

Abstract. In 2007, a team of informaticians and specialists in dentistry in Sweden started a project to develop a CDSS based on openEHR for an oral disease named dry mouth. Since openEHR is an emerging standard, designing a clinical decision support system (CDSS) based on it is an unexplored research area. According to our findings, so far, very few (almost none) openEHR-based CDSSs have been released. The methodological approach applied in developing an openEHR-based CDSS is presented in this paper. This includes typical activities in developing CDSSs in addition to the activities one needs to carry out in order to develop an openEHR-based system. In the first phase of this project, the focus has been on openEHR archetype design, knowledge acquisition, and choosing a suitable KRR method based on the available legacy patient records, i.e., a knowledge intensive case-based reasoning method, and the extracted general domain knowledge. We also propose an architecture for such a system with the aim of benefiting from the structured openEHR-based patient data in reasoning.

Key words: Clinical decision support, openEHR, archetype, Case-based reasoning

1 Introduction

Presently, the use of computerized approaches to improve quality of health care is widespread in the clinical domain. Electronic health records (EHR) and CDSS are two complementary approaches to improve quality of health care. One of the success factors of CDSSs is observed to be their integration into EHRs [1–5] and since there are various international EHR standards (such as openEHR) being developed, it is important to take these standards into consideration while developing CDSSs [6].
Developing CDSSs involves challenges such as representing clinical knowledge, keeping it updated and performing the reasoning [6, 4, 3, 5]. At the time of introducing various EHR standards and calls for moving from standalone CDSSs towards CDSSs that are integrated into EHR system [6, 4, 3, 5], developers should adopt new approaches in developing such systems. This is of course an interesting research problem to see how the presence of these standards may influence developing CDSSs.

In this paper, in addition to presenting a methodological approach in developing openEHR archetypes and an openEHR-based CDSS, the authors propose utilizing the openEHR-based stored data in the reasoning process in a CDSS. The proposed approach is one way to realize the integration of EHR systems and CDSSs as recommended by researchers in the domain. The introduced architecture for such a CDSS is not still fully implemented and the paper includes only the initial findings of this aspect of the research.

The paper is structured as follows. The background information about openEHR and the oral disease for which the CDSS is going to be developed is given in Section 2. Methods and materials applied in this study for designing an openEHR-based CDSS are presented in Section 3. This includes the activities carried out in this methodological approach. Results of the activities are given in Section 4. A discussion is provided in Section 5. Finally, we end with a conclusion and the future directions of the study in Section 6.

2 Background

2.1 The openEHR Approach

openEHR is an open specification standard for managing EHRs so that the problem of shareability and computability of information in the clinical domain is overcome [7]. The openEHR approach emphasizes the role of clinicians in organizing domain knowledge in the form of different clinical concepts such as observation, evaluation, instruction and action [7]. This approach suggests a two-level architecture for clinical applications to separate knowledge and information layers in order to overcome the ever-changing nature of clinical knowledge. Patient data is stored in a generic form, which is retrievable in heterogeneous clinical applications based on some constraints named archetype. An archetype, which is designed by domain experts, defines constraints on data in terms of types, values, relation of different items and so on [7]. Archetypes are used for data validation and sharing [7].

Very few methodological approaches are documented to guide developing archetypes or openEHR-based applications. A well-known methodological approach for developing archetypes is the one presented by Leslie et al. [8], which includes these steps: (i) Identifying clinical concepts (ii) Identifying existing archetypes (iii) Creating new archetypes if necessary.
2.2 Dry Mouth

The research work presented here is actually the outcome of a real world project in a need for developing a CDSS for an oral disease named dry mouth. Dry mouth or xerostomia is “the abnormal reduction of saliva and can be a symptom of certain diseases or be an adverse effect of certain medications” [9]. There are various causes for dry mouth, of which certain previous treatments on the patient, drugs and diseases can be mentioned. Dry mouth is typically managed with saliva substitutes, yet these days, clinicians can find more systemic treatments for this condition [9]. A number of specialists in dentistry from The Sahlgrenska Academy at Gothenburg University\(^3\) proposed a need for getting an automatic support for diagnosis and treatment of this disease. In their opinion, many general dentists are not aware of systemic therapies available for dry mouth, and it is not so easy for them to find potential causes for the disease in order to administer the optimal treatment.

3 Materials and Methods

So far in the project, the focus has been on openEHR archetype design, knowledge acquisition, and choosing a suitable knowledge representation and reasoning (KRR) method, based on the available legacy patient records and the available external domain knowledge. Details of the activities and their outputs are discussed below.

3.1 Preparing the Assessment Questionnaire

Our domain experts were not able to independently develop the archetypes as suggested by Leslie et al. [8], especially since they were new to openEHR. Hence, we decided to use a different approach that suited our domain experts. As part of their everyday work, specialists in dentistry at Sahlgrenska have access to an online application named mForm that provides them with facilities to develop their own examination forms (assessment questionnaires). These forms act as data entry interfaces for the clinical system available for patient data gathering in Sahlgrenska. Hence, our domain experts were already familiar with the concept of independently developing their own examination forms. This fact, initiated our approach for developing openEHR archetypes based on clinical questionnaires.

3.2 Domain Concept Modeling and Designing openEHR archetypes

The questionnaire created in the previous step, was used as a basis to produce domain concept models and finally archetypes. Domain concept diagrams were used mainly for communicating the concepts and the relation between them.

\(^3\) http://www.sahlgrenska.gu.se/english
These models were created using a mind-mapping tool\(^4\) in collaboration between domain experts and informaticians. In our approach, brainstorming sessions were held with domain experts to iteratively prepare the domain concept diagrams. Finally, the domain concept models were used by informaticians to create archetypes. This was done using the available openEHR tools.

![Diagram of domain concept model](image)

**Fig. 1.** A part of the domain concept model diagram

### 3.3 Domain Knowledge Gathering and Related Patient Data

At first, we held interviews with domain experts and also studied the related material to find out more about dry mouth, and related concepts. However, we soon found out that, as informaticians, we would not be able to efficiently extract knowledge from existing evidences. Therefore, in parallel to domain concept modeling and archetype creation, a domain expert was asked to gather such information.

For this purpose, PubMed\(^5\) was searched for the following terms: ("xerostomia" OR "dry mouth" OR "Sjögren’s Syndrome") AND ("therapy" OR "treatment"). Also individual terms were combined, e.g. "xerostomia" AND "treatment"; "xerostomia" AND "therapy", etc.

Moreover, the main dentistry database at Sahlgrenska was also searched by domain experts to find dry mouth related patient cases. Details about this are given in the following:

\(^4\) http://xmind.org  
\(^5\) http://www.pubmed.gov
3.4 Selecting the KRR Method

In this project, based on the amount of legacy patient records and domain knowledge available, and some other motivations (see Section 4.4), we chose a knowledge intensive case-based reasoning (CBR) as the knowledge representation and reasoning method for the CDSS.

Case-based Reasoning One of the reasoning methods that has been used in the clinical domain is case-based reasoning (CBR) [10, 11]. Begum et al. [11] explain CBR and its application in the clinical domain as follows: “The CBR is inspired by human reasoning, i.e. solving a new problem by applying previous experiences adapted to the current situation. A case (an episodic experience) normally contains a problem, a solution, and its result. The CBR is an appropriate method to explore in a medical context where symptoms represent the problem, and diagnosis and treatment represent the solution”.

In this method, the solution to previous problems is adapted to the new problem [11, 12]. Cases and indexing information are stored in the knowledge-base, and reasoning is carried out by doing indexing, matching and adapting and storing new cases. Indexing efficiency is a key issue in this reasoning method [3, 13].

Case-based reasoning may be considered to be a data intensive method, since it starts with a set of cases for training. However, it is very different from other data intensive approaches. The knowledge that is extracted from experts in knowledge-intensive methods can be subjective. On the other hand, if one only relied on objective knowledge, e.g. the knowledge extracted from evidence, the valuable experience of experts in domain is not used. CBR uses both objective and subjective knowledge for reasoning [11]. In other words, reasoning from previous cases is done in this method [11, 12]. In the case-based knowledge-base, cases (Knowledge is recorded in form of cases) and indexing information are recorded [3].

One class of CBR methods is a hybrid approach called knowledge intensive-CBR where the reasoning process is enhanced by benefiting from reasoning on the existing general knowledge [12]. The CBR cycle that includes retrieve, reuse, revise and retain [12] is depicted in Figure 2 (note the “general knowledge” close to the “previous cases” in Figure 2).

3.5 Creating Artificial Cases

In a need for generating some dry mouth patient cases, we used the assessment questionnaires generated before, to develop data entry forms using the mForm application. Manual patient entry was performed by one domain expert with the aim of creating some cases to be used in the reasoning process. However, this was done in order to have a starting point for the CBR process. After using the system for real cases, the system will be trained based on actual cases and there will be no need for artificial cases.
3.6 Proposing an Architecture Based on Two Existing Frameworks

A framework is a set of classes and design solutions and one can see it as a partial design and implementation of an application [14]. Using a framework, a huge amount of development time can be saved in a project. For developing an openEHR-base CDSS it is most efficient to use the existing frameworks to benefit from the services provided for managing the openEHR-based patient records. This is the same for CBR. Therefore, this CDSS is going to be built on top of two frameworks namely opereffa (an openEHR application development framework) and JColibri (a CBR application development framework).

opereffa [15] is one of the existing frameworks for developing openEHR-based applications. The framework manages tasks needed for loading archetypes, validating them and storing data, based on the openEHR reference model. JColibri [14] is a framework for developing CBR applications. It includes basic functions and algorithms needed in a CBR application. Both of the frameworks are Java-based (our preferred language) and they have been developed using a layered architecture that makes them suitable for our purpose.

4 Results

The workflow of the activities in the first phase of the project is depicted in Figure 3. As shown in this Figure, the outputs of the activities are: the questionnaire, the domain concept models, the archetypes, general domain knowledge, and artificial dry mouth patient cases. Also as mentioned in the previous section, based on the gathered information (general domain knowledge, and the available legacy patient records) a KRR method was selected for the CDSS namely
knowledge-intensive CBR. More details about the results are given in the following.

Fig. 3. The methodological approach in developing an openEHR-based clinical decision support system.

4.1 The Questionnaire, Concept Models and Archetypes

The questions covered various aspects of patient data including history of other diseases and drugs, related lab results, diet habits, age and sex. The questionnaire consisted of 6 main sections and a total of 41 questions and 15 related laboratory tests. The questionnaire was created by one domain expert and shared with 4 more domain experts to be revised and approved.

A domain concept model was created based on the generated questionnaire. The model was created in close collaboration with clinicians. Various concepts/sections in the questionnaire were mapped to the related openEHR general classes such as observation and evaluation. A sample of the domain concept model is depicted in Figure 1.
The model was later translated to 23 openEHR archetypes by informaticians. Before creating the archetypes, the existing shared openEHR archetypes were searched to find reusable archetypes. Around 10 archetypes were reused and almost all were modified for this purpose.

In addition, having a close collaboration with clinicians while developing these models, the interviews we held with some external domain experts provided an opportunity for us to gather some basic knowledge that a clinician uses in diagnosis and treatment of dry mouth. Furthermore, some external resources were also studies to gather more general information regarding dry mouth. Based on this information a group of rules were created and will be used to enhance the CBR process.

4.2 General Domain Knowledge

More than 6000 references were obtained from the searches; only review articles were used due to their scientific evidence level. From a list of around 1000 articles, 71 were selected. Papers describing treatment strategies in xerostomia, dry mouth or Sjögren’s Syndrome were the main objective of the search.

No guidelines for the treatment of any of these diseases were found after this search. This suggests that there is no global agreement about treatment strategies in xerostomia, dry mouth, or oral manifestations of Sjögren’s Syndrome. However, there are several papers with a high level of scientific evidence about different therapy methods for patients with these diseases, such as systematic reviews or meta-analysis, where some statements can be drawn from.

In contrast to dry mouth and Sjögren’s Syndrome, several treatment guidelines and global treatment consensuses can be found for many other medical conditions. Despite all the literature that has been published so far related to treatment strategies in dry mouth or xerostomia, the available information is far less compared to other common diseases. There is a global agreement in the literature that xerostomia is a common and significant (or maybe the most common) side effect of many commonly prescribed drugs. However, according to the literature, it is difficult to establish relative incidence rates for xerostomia for a particular medication. This is due to the fact that because reported rates depend on how the professional collects the information from the patient, how patients react to a specific kind of drug, the type of drugs being taken, the cause for which the drug is being taken, the possible presence of contributing factors (such as Sjögren’s Syndrome, radiation therapy, or other immunological conditions) to the disease, the dose of the medication, etc. So, as an example, it is not possible to affirm that xerostomia will be present if a patient is taking 3 or more systemic drugs; neither about the percentage of patients who will present with xerostomia if taking 3 or more drugs. Nevertheless, the risk for xerostomia increases with the number of drugs being taken. Some studies have described up to 500 medicaments that may have caused xerostomia, and those drugs listed have been reported to cause xerostomia in 10% or more of patients.
4.3 Legacy Patient Records

The database of oral medicine at Sahlgrenska contains more than 20,000 patient records and images. Nevertheless, there are only around 100 dry mouth related cases in the database. According to our domain experts, the information related to dry mouth is missing for most of the patient cases, especially information related to diagnosis and treatment of the disease. One might be able to extract most of the historical information of patients from the available repositories but still other vital information is still missing. In other words, there are not a reasonable number of high quality (complete) dry mouth cases in the repository. As a solution to this problem, a number of artificial cases were created by the domain experts to improve the process of CBR.

4.4 Knowledge Representation and Reasoning

There are two main classes of choices for selecting a KRR method, i.e. data intensive methods and knowledge intensive methods. Answers to the following questions would help in choosing the suitable method in a specific project.

– Do we have enough data to be used in data intensive methods?
– Do we have enough domain experts to extract the knowledge or sufficient structured domain knowledge in order to adopt a knowledge intensive method?

Unfortunately in this project, the answer to the both questions was “no”. Yet, as a result of previous activities (see Section 4.1) we had some basic knowledge that, though not sufficient for adopting knowledge intensive method, could be used for simple rule-based reasoning. Additionally, the domain experts involved in the project indicated that, while they are not able to provide us with probability numbers or exact rules in how to treat dry mouth, they can create some dry mouth patient cases. This yielded in a decision for adopting a knowledge intensive-CBR for the dry mouth CDSS.

As illustrated in Figure 2, general knowledge is used in CBR to improve the reasoning process [12]. This general knowledge can be represented in many ways, one of which is a set of rules. With a rather small set of rules (general domain knowledge) generated as a result of previous activities, we can support the CBR method we selected and benefit from a hybrid method. The reasons for selecting the knowledge-intensive CBR method are explained in Section 5.1.

Two domain experts were responsible for creating the cases. For this purpose the aforementioned questionnaire was used to create an online data entry form using the tools available at the clinic. Total of 14 cases were created at this stage.

4.5 Archetypes versus Domain Knowledge

In the dry mouth domain, a sample of general knowledge would be: *people who use 3 or more drugs at the same time usually get dry mouth.* This kind of information cannot be extracted from archetypes, but can be extracted from literature, or gained from the clinicians’ mind, that is why as result of domain knowledge
gathering and domain concept modeling activities we generated some general domain knowledge.

Based on the openEHR approach, clinicians would be responsible for creating archetypes even though our experience revealed that sometimes it is too optimistic to think that this task is done only by a group of clinicians. Especially in cases where we plan to use archetype data for clinical decision support. This means that one should be aware of the fact that beside attributes (items in archetypes) one needs some knowledge to be used in CBR. As shown in Figure 3, the general knowledge was extracted from literature and from the meetings with the domain experts (the meetings were basically held for designing the archetypes). This information was added to the models that were created for domain knowledge as descriptions for each item.

4.6 The software architecture proposal

A layered architecture is proposed for this application. As depicted in Figure 4, the top layer is the view layer or in other words the user interface. Below that, there is a mapper layer that is responsible for mapping the GUI components to the opereffa framework classes, also to connect them to the JColibri framework. Automatic generation of cases out of archetypes will be done in this layer. JColibri manages the CBR process and will have access to patient data repository. This repository will be shared between JColibri and opereffa. Patient data is stored in the database based on openEHR reference model. Each case/patient data will be stored using the related archetype that acts as a constraint on data. All the processes related to openEHR are managed by opereffa. JColibri and opereffa layers are not aware of each other, and everything between them is managed via the mapper layer. So far, part of the mapper layer that is responsible for mapping the view layer to the openEHR underlying framework is implemented and tested. The implementation of the whole application is still in progress.

Fig. 4. The proposed architecture.
5 Discussion

As mentioned before, there are very few documented methodological approaches to guide developing archetypes or openEHR-based applications. This includes those published by Marcos et al. [16] and Leslie et al. [8]. The most well known methodological approach for developing archetypes is the one presented by Leslie et al. [8]. This approach however emphasizes the role of clinicians and how they may apply existing openEHR tools to browse existing archetypes or developing new ones. Our approach is different in its view on adopting more traditional clinical approaches for gathering information such as clinical questionnaires.

Additionally, when it comes to CDSSs, there are a few studies that deal with how openEHR offers opportunities for CDSSs. Most of these efforts however seem to be more focused on integrating clinical guidelines into openEHR archetypes or utilizing archetypes for representing clinical guidelines [16–18] or to enhance archetypes by including knowledge representation capabilities to them [19]. To our knowledge, there is no study that has been focused on benefiting from the well-structured openEHR-based patient data for adopting data intensive reasoning methods in CDSSs or methods such as CBR that rely on previous cases to carry out the reasoning process, as our approach does.

As this study shows, because of the openEHR novelty, it is likely that in many projects, clinicians are more familiar with the traditional clinical approaches such as creating clinical questionnaires compared to the more complex process of creating archetypes or templates. Therefore in such cases, other approaches for developing archetypes can be applied that are more compatible with the capabilities of the clinicians involved in the project.

The well-defined concepts in openEHR help provide opportunities for informaticians to use these concepts for knowledge extraction in this manner. In case of developing a CDSS, the activities in developing archetypes can be in form of more informed interviews and/or brainstorming sessions with domain experts where not only domain concept models and eventually archetypes are generated but also the available general domain can be extracted from clinicians’ minds.

5.1 Why Case-based Reasoning?

CBR is considered to be a suitable method to be used in CDSSs especially in the clinical domain [10, 11]. The concept of case is a concept that is used in medicine as well as for training and discussing treatment of an individual patient; moreover clinical guidelines include practice cases [10].

In addition, CBR is the most suitable KRR method to be used for this CDSS not only because of the advantages and practicability of this method for this project considering the amount of data and knowledge available, but also for the similarity found between openEHR archetypes and cases in CBR.

Case description is analogous to archetypes in openEHR as discussed in Section 5.2. This approach did not require explicit knowledge to be represented and also the reasoning process is not a black box and can be understood by users. Therefore, the developed system can be used for training dentistry students as
mentioned before. Table 1 shows a brief comparison between the selected method and other existing approaches.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Method</th>
<th>No need for deep domain knowledge</th>
<th>No need for huge data volume</th>
<th>Easy Knowledge Acquisition</th>
<th>Objective Knowledge</th>
<th>Subjective Knowledge</th>
<th>Easy to maintain</th>
<th>Use of past experiences</th>
<th>Suitable for Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Intensive</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Knowledge Intensive</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Knowledge intensive-CBR</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table 1. Motivations for selecting the knowledge intensive-CBR method.

5.2 openEHR archetypes and cases in CBR

As depicted in Figure 2, two types of knowledge are applied in a CBR cycle, domain-dependent knowledge or general knowledge, and specific knowledge that is encapsulated in cases. Archetypes provide us with all the specific knowledge we need for CBR. It is natural to see that no reasoning knowledge is included in the concept models or archetypes, but they could be used for extracting general knowledge of the domain for instance for extracting basic rules that are applied for diagnosis.

openEHR defines different classes of patient data. These classes are observation, evaluation, instruction and action. In openEHR observation is a structure to record any information that is extracted from the world outside the clinician’s mind [7]. This includes patient history of diseases and other treatments and symptoms and signs of the disease. In contrast, evaluation type is used to store the decision made by the clinician that is done in her/his mind. Instruction is a set of tasks that should be done on a patient; for instance prescription or orders. Action is used to record information about the action taken on the patient based on the instructions. Figure 5 illustrates how an openEHR-based patient record can be mapped to a case in CBR.

On the other hand, representation of the cases in CBR includes Description of the problem and the Solution. Description of the problem is analogous to the observation part of the patient data, including information about clinical history, symptoms and signs, and also lab values. Each case in CBR should include information about the solution to the specified problem (query) in that case. This solution in CBR is analogous to the openEHR evaluation, instruction and action in openEHR-based patient data.

6 Conclusion and Future Work

To develop an openEHR-based CDSS, one should carry out not only the typical CDSS development activities but also activities suggested by openEHR commu-
Fig. 5. The sample archetyped data

...nity for providing a solid underlying layer for storing and retrieving sharable clinical data. The openEHR activities start with designing archetypes by involving domain experts. In this study, we found out that the approach suggested by the openEHR community [8] is not applicable because of the capabilities of the clinicians involved and we needed to apply our own approach, which was designing archetypes based on clinical questionnaires.

Moreover, as in all CDSSs, a knowledge representation and reasoning method should be selected. There are some criteria for selecting a KRR method for a CDSS that we applied for selecting CBR for this project. CBR is a suitable reasoning method for clinical domain since it is analogous to the concept of individual patients, known as cases, which are also used for training medical students. Clinicians see each patient as a case and even use this term for sharing patient data among colleagues. Additionally, CBR applications can be used for education in clinical domain.

Applying a CBR method in an openEHR-based CDSS is an interesting open research direction, but needs connecting two different frameworks. Cases have similarities to archetypes; therefore they can be generated automatically from them and be used for reasoning purposes. openEHR archetypes help in the knowledge extraction process, but the classical bottleneck of knowledge acquisition in clinical domain still exists. The next phase of the project is to implement the rest of the mapper layer and investigate the reasoning functionality.
References