Creating an ontology to integrate clinical evidence databases

Jan Carl Beeck Pepper, Ivan Luiz Marques Ricarte (Supervisor)

Departamento de Engenharia de Computação e Automação Industrial (DCA)
Faculdade de Engenharia Elétrica e de Computação (FEEC)
Universidade Estadual de Campinas (Unicamp)
Caixa Postal 6101, 13083-970 – Campinas, SP, Brasil

{jcbeeck,ricarte}@dca.fee.unicamp.br

Abstract – The medical scientific information is scattered and growing, this information generally does not reach to the hands of health personnel who make decisions. Our goal is to integrate the approach of evidence-based medicine (EBM) with the electronic health record (EHR) in order to bring the former information into use in daily basis. Our partial results are the identification and creation of a specific computer ontology to solve the problem that we are addressing. The expected results that we want to achieve are: use our crafted computer ontology and identify the best evidence that exist according to the data found in the EHRs.

Keywords – Evidence-Based Practice, Electronic Health Records, Computer Ontology

1. Introduction

Healthcare professionals have daily to take decisions related to their patients, and ideally such decisions should be based on their experience and on the most current information available. This is the essence of Evidence-Based Practice (EBP) or, specifically in the case of physicians, Evidence-Based Medicine (EBM). One process to promote EBP is knowledge translation, defined by the Canadian Institutes of Health Research as a dynamic and iterative process that includes the synthesis, dissemination, exchange and ethically sound application of knowledge to improve health, provide more effective health services and products, and strengthen the health care system [12]. Synthesized information applicable to healthcare, known as evidences, are kept in knowledge repositories known as evidence databases.

Even with availability of evidences, one of the main obstacles to EBP is the huge volume of information. Health professionals are unable to keep up with the new knowledge in their fields, since approximately one million new articles are published yearly only in the health field [10]. A previous study [1] also pointed out that they should have knowledge about the available databases, be trained on how to use them and on how to formulate carefully structured clinical questions about issues related to specific patients, and have enough time to search the evidence databases to support their clinical decisions.

In this work, an approach to transparently integrate and querying evidence databases using ontologies is proposed. A computer-based ontology is an explicit specification of a conceptualization, which is the objects, concepts, and other entities that are presumed to exist in some area of interest and the relationships that hold them [4]. By having the knowledge about evidences expressed as an ontology, one should be able to create applications using Semantic Web technologies to transparently integrate evidence databases and to find, based on the health professional notes about a specific patient, which is the best evidence to support a clinical decision.

2. Literature review

A literature review was performed to evaluate the state of the art related to the relationship of clinical notes taken by health professionals, usually registered in Electronic Health Records (EHR), and evidence-based practice, or more specifically to the access of evidence databases to support clinical decisions.

This review followed the principles of systematic reviews, as stated in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework [9]. Thus, the main concepts related to the literature search, Electronic Health Record and Evidence-Based Practice, were clearly defined. Then, for each concept, the corresponding search terms were established. As an example, the concept Electronic Health Record could also be found in the literature by the terms Electronic Medical Record as well as by the corresponding abbreviations EHR and EMR. The search was performed on bibliographic databases considered relevant for this field: PubMed, ScienceDirect, ACM Digital Library, and IEEExplore.
From 339 papers found in these bibliographic databases, 2 were duplicate. From the 337 remaining papers, 323 were excluded by considering that the title was not related to the topic of interest; further analysis of full text in the remaining set of 14 papers yield to the exclusion of 6 additional papers, resulting in the selection of 8 papers for the detailed review.

Table 1 synthesizes the result of this review.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Problem</th>
<th>Proposed solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mendonça et al., 2001 [8]</td>
<td>Search query formulation</td>
<td>Controlled terminologies</td>
</tr>
<tr>
<td>Hsu et al., 2012 [5]</td>
<td>Clinical text mining</td>
<td>Ontologies and semantic annotations</td>
</tr>
<tr>
<td>van Valkenhoef et al., 2013 [13]</td>
<td>Clinical decision support</td>
<td>Unified data model for evidences</td>
</tr>
</tbody>
</table>

Table 1. Summary of literature review

From what we have seen from the review, we can state that the proposal to integrate information through knowledge capture, whether through mentioning the potential of using ontologies or conceptual graphs. There also exists systems that are already developed to achieve the use of EBM, however none of them mention that they were implemented with a computer ontology crafted specifically for the integration of evidences with EHR, which is the problem that we want to solve.

3. Proposal

Our proposal is to facilitate the use of evidences by health professionals, enabling them to locate clinical evidences having patient information as the starting point for their searches. Thus, the following questions were proposed: How evidences are represented? How evidence databases are organized and searched? How can we integrate searches and results from evidence databases with distinct characteristics?

To achieve this goal and answer these questions, we will develop an ontology to generically describe evidences. This ontology should support the search in distinct evidence databases and act also as an integration point for the patient clinical information.

Methodologically, we will develop a case study with two evidence databases in a selected medical field. Based on this case study, an application will be developed and its validation and extension will be planned.

4. Partial results

The selected medical field for the case study was Physiotherapy, and two clinical evidence databases were selected: PEDro, the Physiotherapy evidence database 1 and the PICO interface for PubMed 2. PICO stands for Population, Intervention, Comparison, and Outcome, a structured way to represent clinical questions [3].

The patient clinical data is represented, in this study, by a code taken from the International Classification of Functioning, Disability and Health (ICF). The two databases support different strategies for searching. PEDro has some locally defined codes, whereas in the PICO interface for PubMed the P and I elements should be used first to construct queries, and only if too many results are obtained the other elements should be considered. Also comparison element has to be considered as the least important part of the query because it often contains a vague word.

For presenting the results, we chose to present a list with results from the databases with the evidence title, date of publication, and link to the full information in the original evidence database.

From this case study, we proposed an ontology structured in five classes: Clinical information, Category, Attributes, Query, and Result. Each

---

1 http://www.pedro.org.au
class has concepts and attributes; we have define the former with a name that unifies its attributes and the latter with the items that we need to capture information. Table 2 summarizes the ontology concepts, with the name chosen for the concept, alternate names (i.e., different names for the same concept as appeared in some evidence databases), proposed classification (one of the five concepts above), and comments (additional attributes or comments).

### Table 2. Table of concepts

<table>
<thead>
<tr>
<th>Name of the concept</th>
<th>Alternate names</th>
<th>Classification</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient information</td>
<td>Patient ID, Sex, Date of birth, Age</td>
<td>Clinical information</td>
<td></td>
</tr>
<tr>
<td>Record information</td>
<td>Patient allergies and critical information, General information, Medication, Diseases, ICD10, ICF, Surgeries, Lifestyle, Socioeconomics</td>
<td>Clinical information</td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>Physiotherapy, General biomedical literature</td>
<td>Category</td>
<td>PEDro, MEDLINE/PubMed</td>
</tr>
<tr>
<td>Codes (PEDro codes)</td>
<td>Subdiscipline, Intervention, Problem, Body part, Topic</td>
<td>Attributes</td>
<td>Orthopaedics (Ex.), Acupuncture, Pain, Thoracic spine, Chronic pain</td>
</tr>
<tr>
<td>PICO</td>
<td>Problem, Intervention, Comparison, Outcome</td>
<td>Attributes</td>
<td></td>
</tr>
<tr>
<td>String query</td>
<td>ICT, String</td>
<td>Query</td>
<td></td>
</tr>
<tr>
<td>List of results</td>
<td>Title, Link</td>
<td>Results</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1 presents a graphical representation of the elements for the proposed ontology.

### 5. Future work

To implement our ontology we are going to use Protégé, which is a free open-source ontology editor and framework for building intelligent systems. The Jena library, a free and open source Java framework for building semantic web and linked Data applications, will be used to create an application to test our hypothetical scenarios to validate our proposal.

### References


