Transforming XML-Based Electronic Patient Records for Use in Medical Case Based Reasoning Systems

Syed Sibte Raza ABIDI & Selvakumar MANICKAM
Health Informatics Research Group, School of Computer Sciences
Universiti Sains Malaysia, Penang, MALAYSIA.

Abstract. Electronic patient records (EPR) can be regarded as an implicit source of clinical behaviour and problem-solving knowledge, systematically compiled by clinicians. We present an approach, together with its computational implementation, to pro-actively transform XML-based EPR into specialised Clinical Cases (CC) in the realm of Medical Case Base Systems. The ‘correct’ transformation of EPR to CC involves structural, terminological and conceptual standardisation, which is achieved by a confluence of techniques and resources, such as XML, UMLS (meta-thesaurus) and medical knowledge ontologies. We present below the functional architecture of a Medical Case-Base Reasoning Info-Structure (MCRIS) that features two distinct, yet related, functionalities: (1) a generic medical case-based reasoning system for decision-support activities; and (2) an EPR-CC transformation system to transform typical EPR’s to CC.

1. Introduction

The exploitation of Case-Based Reasoning (CBR) techniques [1] in ‘intelligent’ medical decision-support systems is by now a well-established practice within the medical fraternity [2, 3, 4]. Typically, CBR based medical systems provide ‘analogy-based’ solutions/diagnosis to clinical problems by manipulating the specific knowledge of previously experienced situations, called Cases. Each case is usually described by a set of case-defining attributes, and is associated to a solution (or decision) suggested by a medical practitioner. For instance, for cases pertaining to breast cancer, the attributes describe a collection of macroscopic areas, each of them associated to a collection of histologic areas together with cytological description and so on. The knowledge of CBR based medical systems pertains to domain-specific cases—stored in a structured repository called the case base—acquired over time by noting the problem-solving behaviour of medical practitioners [1].

In this paper, we present an approach, together with its computational implementation, for pro-actively transforming Electronic Patient Records (EPR) into specialised Clinical Cases (CC) which can then be seamlessly incorporated within Medical Case Base Systems [2], and used for deriving CBR-based solutions. The motivation for our approach derives from the fact that we regard EPR as an ‘alternate’, yet implicit, source of clinical behaviour and problem-solving knowledge, systematically compiled by clinicians during episodic visits by patients [5, 6]. We note that EPR comprise the kind of information—i.e. longitudinal patient history, illness-related symptoms & signs, pathological finding, diagnosis (or prognosis) by physicians and a treatment plan—that is usually incorporated in the definition of CC. Henceforth, this brings into relief the opportunity to transform generic XML-based EPR—originating from heterogeneous EPR repositories accessible over the Internet/WWW—to specialised CC. In this way, we are able to systematically
supplement the medical knowledge (acquired through traditional knowledge engineering methods) already present in any medical case base system.

Methodologically speaking, we argue that the transformation of an EPR to CC is not a mere mapping of corresponding attributes from the EPR to CC. Rather, in our work, a correspondence between EPR and CC attributes is established by way of a systematic exercise of structural, terminological and conceptual standardisation. We incorporate a confluence of techniques and resources to achieve the necessary standardisation of CC, this is to ensure the medical completeness and correctness of the derived CC. Structural standardisation is achieved by exploiting the XML, terminological standardisations is achieved by employing UMLS (meta-theasaurus) and conceptual standardisation is acquired by the involvement of medical knowledge ontologies.

2. Transforming EPRs to Clinical Cases: Methodological Issues

Typically, EPR’s contain data/information that reflects events pertaining to the diagnostic-therapeutic observations, decisions, interventions and outcomes—in summary the complete and longitudinal medical history of a patient, ranging from demographic data to episodic diagnostic/treatment information to electro-physiological and image data [5]. An EPR is a document that is both structured and coded—usually using HL7 and ICD10 coding for diagnosis. More structure is now being added to EPR’s—vis-à-vis the usage of XML—due to the growing need to distribute them across multiple institutions over the Internet. Note that the use of XML to represent EPR offers a wide range of interoperability, whereby EPR generated from diverse institutions and applications can be transported over the Internet and moreo manipulated by different applications. In our work, we exploit this intrinsic structure of EPR’s to realise their transformation to other structured and logical formats, such as CBR ‘case’ formats.

The transformation of generic EPR’s to specialised CC is not a straight forward mapping of attributes (and their values) from the EPR to a CC representation. Rather, we propose the need for a EPR-CC Transformation Methodology that takes into account (i) structural (ii) conceptual and (iii) terminological differences between the source EPR’s...
and the target CC. In essence, the requirements circumvented by the EPR-CC transformation methodology needs to address two distinct, yet related, activities:

1. **Data Transformation** addresses the issue of how structurally divergent and heterogeneous EPR representations can be translated to apriori defined CC structure (notwithstanding database platform variations). The EPR-CC data transformation depends on the satisfaction of two constraints: (i) The use of HL7 to store healthcare information in an EPR by the source HIS; and (ii) the use of XML by the source HIS to both specify and transport the EPR over the Internet. Data transformation relies on the XML based EPR’s Document Type Definition (DTD)—the logical structure of the EPR document—which defines the EPR’s semantic tags (as shown in figures 1, 2 and 3). For data transformation, we have implemented a HL7 and XML based bridge to parse the EPR, based on its DTD, to generate a mapping from the EPR to the CC structure.

2. **Data Abstraction** handles and fixes conceptual and terminological mismatches during the EPR-CC transformation process. To address the issues of standardisation of concepts and terminology, we propose (a) the usage of medical ontologies [8] for the abstract transformation of ‘novel’ or ‘non-standard’ concepts (with regards to the clinical cases) within the EPR’s to the standard concepts understood by the CC; and (b) the usage of UMLS, used as a medical thesaurus, to standardise the translation of ‘non-standard’ terms (with regards to the clinical cases) to the native terminology of CC.

In essence, the tenets of our EPR-CC transformation methodology are (a) to exploit the representational and operational flexibility of XML-represented EPRs; (b) to map selected data items from the EPR to corresponding CC attributes; (c) to employ additional knowledge resources, such as UMLS (meta-thesaurus) for vocabulary standardisation and medical knowledge ontology to align medical concepts, if a one-to-one mapping does not exist between the desired CC attributes and the EPR’s data items.

3. **Transforming EPRs to Clinical Cases: Functional Issues**

   We present below the functional architecture of a Medical Case-Base Reasoning Info-Structure (MCRIS) that features two distinct, yet related, functionalities: (1) a generic medical case-based reasoning system for decision-support activities; and (2) an EPR-CC transformation system to transform routine EPR to CC. However, in this paper, we will only focus on the EPR-CC transformation system. To support the above functionality MCRIS comprises three functionally distinct layers. Each layer comprises a number of modules, each responsible for a certain task. Figure 4 shows the functional architecture and the process workflow of MCRIS. We briefly explain the functionality of the layers.
**Data Procurement Layer**: Responsible for (a) procuring EPR from diverse sources over the Internet; (b) cleansing & filtering the procured EPR—i.e. checking for missing items & inconsistent values; and (c) uploading the cleansed EPRs as an intermediate database.

**Case Transformation Layer**: Responsible for the transformation of the ‘uploaded’ EPR into specialised CC, as per the specification of the case base. This layer incorporates several modules: (a) an *EPR Parser* that parses the EPR (represented in XML) based on the EPR’s DTD to explicate the EPR’s semantic tags and the data elements; (b) a *Vocabulary Standardiser* using the UMLS meta-thesaurus, to standardise the medical terminology between the EPR and CC, as per the controlled vocabulary—represented as the Clinical Case-MetaMap—of the MCRIS. Put simply, the Clinical Case-MetaMap denotes all the CC Attributes together with the range of allowed CC Values for each attribute, such that it ensures that the non-standard attribute labels of the EPR can be mapped on to the CC definition; and (c) a *Concept Standardiser* employing medical knowledge ontologies to standardise medical concepts between the EPR and CC definition—to account for the fact that a medical concept in the EPR is (or needs to be) represented as the generalisation/specialisation of a concept in the CC definition. Two different types of medical ontologies are available here: (1) standard medical ontologies available within the medical literature and (2) ontologies pro-actively derived by us from medical coding schemes such as ICD10, MSH99 and so on (see figures 5 and 6); and (d) a *Case Compiler* that completes the ‘automatic’ EPR-CC transformation by transferring the EPR’s attribute values to the corresponding CC’s attribute value-slots. To record the EPR-CC transformation, the case compiler makes use of an internal Transformation-Map, that contains the mapping of the Clinical Case-MetaMap to EPR Attributes and Values with respect to the EPR’s DTD. Finally, the user is provided an on-screen explanation to validate the EPR-CC transformation, if desired.

**Case Based Reasoning Layer**: Responsible for the delivery of decision-support services based on the knowledge contained in the case base by employing built-in case-based reasoning strategies. Here, we incorporate standard CBR features [1] within MCRIS.

<table>
<thead>
<tr>
<th>ICD10</th>
<th>MSH99</th>
<th>AOD95</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symptoms, signs and abnormal clinical and laboratory findings</td>
<td>Diseases (MeSH Category) [C]</td>
<td>Health and Disease [G]</td>
</tr>
<tr>
<td>General symptoms and signs</td>
<td>Symptoms and General Pathology [C23]</td>
<td>Symptom [GF]</td>
</tr>
<tr>
<td></td>
<td>Signs and Symptoms [C23.888]</td>
<td>Physical symptom [GF2]</td>
</tr>
</tbody>
</table>
4. The EPR-CC Transformation Process

We mentioned earlier that the EPR-CC transformation requires the systematic and ‘correct’ selection and filtering of relevant information—vis-à-vis the CC’s specification and (controlled) vocabulary—from the source EPR. The data entities involved in the EPR-CC transformation are all the CC Attributes (say CCA), the range of allowed CC Values (say CCV) for each attribute, the EPR Attributes (say EA) and EPR Values (EV) with respect to the EPR’s DTD. The EPR-CC transformation process is carried out according to the following scheme, observing the given sequence of transformation tasks:

1. **Direct Mapping of CCA to EA:** For each CCA a mapping to a matching EA is determined by comparing the Case MetaMap with the EPR’s DTD. IF an exact match is found then it is recorded in the transformation-map as an *Exact Match*.

2. **Indirect Mapping of CCA to EA:** In this case, we determine whether the target ‘unmatched’ CCA is represented in the source EPR using some variant terminology. The UMLS meta-thesaurus is used to find all the CCA’s synonyms, which are then compared with the EPR’s DTD to determine whether a similar EA exists. If a match is found it is recorded in the transformation-map as a *Synonymous Match*.

3. **Ontological Mapping of CCA to EA:** In case, the above-mentioned mapping strategies fail, we use a medical knowledge ontology (figure 6), based on standard medical coding schemes (figure 5), to attempt to inductively match a CCA with any possible EA. A medical knowledge ontology determines the ontological synonyms of the ‘unmatched’ CCA, which can then be compared with the EPR DTD to establish whether there exist any EA that can be regarded as the conceptual specialisation or generalisation of the CCA in question. If a successful ontological match is achieved then it is recorded in the transformation-map as an *Ontological Match*.

5. Conclusions

In our work, we have managed to leverage upon ‘information rich’ EPR, accessible over the Internet, to enhance the (medical) knowledge of traditional medical case-based reasoning systems. In this way, we have presented a novel facet and utility of routinely collected EPRs, whereby they can be transformed from mere information resource to a diagnostic decision-support resource. MCRIS is now complete and is under trial at the Oncology Department, University Hospital, Kota Bharu. The physicians involved in the trial are quite interested and impressed with the quality of the CC automatically generated...
from the EPRs. We intend to publish the trail results as a follow-up paper, soon. On a side note we have managed to demonstrate the efficacy of (a) XML as a representation language for medial information/data and (b) medical knowledge ontologies for conceptual mapping. Indeed, the confluence of these technologies have rendered sophistication and ‘intelligence’ to the EPR-CC transformation methodology; note that our approach is a departure from the typical ‘look-up table’ based translation method.

References