TOWARDS A SEMANTIC SYSTEM FOR MANAGING CLINICAL PROCESSES

Ermelinda Oro
DEIS, University of Calabria, Via P. Bucci 41/C, Rende (CS), Italy

Massimo Ruffolo
ICAR-CNR, University of Calabria, Via P. Bucci 41/C, Rende (CS), Italy

Keywords: Knowledge Representation and Reasoning, Knowledge Management, Ontology, Workflow, Data Mining, Workflow Mining, Decision Support System, Health Care Information System, Clinical Process, Medical Knowledge.

Abstract: Managing costs and risks is an high priority theme for health care professionals and providers. A promising approach for reducing costs and risks, and enhancing patient safety, is the definition of process-oriented clinical information systems. In the area of health care information systems, a number of systems and approaches to medical knowledge and clinical processes representation and management are available. But no systems that provide integrated approaches to both declarative and procedural medical knowledge are currently available. In this work a clinical process management system aimed at supporting a semantic process-centered vision of health care practices is described. The system is founded on an ontology-based clinical knowledge representation framework that allows representing and managing, in a unified way, both medical knowledge and clinical processes. The system provides functionalities for: (i) designing clinical processes by exploiting already existing and ad-hoc medical ontologies and guideline base; (ii) executing clinical processes and monitoring their evolution by adopting alerting techniques that aid to prevent risks and errors; (iii) analyzing clinical processes by semantic querying and data mining techniques for making available decision support features able to contain risks and to enhance cost control and patient safety.

1 INTRODUCTION

Across the world the issue of patient safety, medical errors prevention and healthcare risk management is a very challenging and widely studied research and development topic. It stimulates a growing interest in the computer science researchers community.

A promising approach for reducing errors and risks, and enhancing patient safety, is the definition of process-oriented clinical information systems. In fact, healthcare services and practices are characterized by complex clinical processes in which high risk activities take place. A clinical process can be seen as a workflow where clinical (e.g. treatments, drugs administration, guidelines execution, medical examinations) and general (e.g. patient enrolment, medical record instantiation, risk evaluation) activities and events occur. Clinical processes and their activities are, also, characterized by specific and sophisticated medical knowledge. Systems that provide integrated functionalities for representing and managing medical knowledge and for designing, executing (taking into account risks rules and conditions) and analyzing clinical processes, can change clinical practices and help diffusion of a process and quality awareness in healthcare organizations.

Currently, as described in Section 2, in the field of healthcare information systems, a number of approaches to medical knowledge and guidelines representation and management have been proposed. Existing systems and approaches suffer of the following shortcomings: (i) they have a lack of mechanisms for errors and risks handling and prevention; (ii) they do not use the same formalism for representing and managing both medical knowledge and clinical processes, hence, they are not able to exploit in a fully unified way declarative and procedural knowledge during execution and monitoring of clinical process; (iii) they do not allow to organize clinical processes and their element as an ontology; (iv) they do not allow to modify and customize represented knowledge and to execute clinical process in a flexible and agile way.
This work describes the prototypical implementation of a Clinical Process Management System (CPMS) that aims at supporting a semantic process-centered vision of health care practices. The system is founded on an Ontology-based Clinical Knowledge Representation Framework (OCKRF) that allows to express in a combined way medical ontologies, clinical processes and errors and risks rules. More in detail, the OCKRF allows the CPMS to provide methods for: (i) creating ontologies of clinical processes that can be queried and explored in a semantic fashion; (ii) expressing errors and risks rules (by means of reasoning tasks) that can be used (during processes execution) for monitoring processes; (iii) executing clinical processes and acquiring clinical process instances by means of either workflow enactment (pre-defined process schemas are automatically executed) or workflow composition (activity to execute are chosen step-by-step by humans); (iv) monitoring clinical processes during the execution by running reasoning tasks; (v) analyzing acquired clinical process instances, by means of querying and inference capabilities, in order to recognize errors and risks for patients.

The CPMS adopts a semantics approach for representing and managing both static and dynamic aspects of medical knowledge. So it enables better design, execution, control and management of clinical processes and related errors and risks rules. CPMS delivers health care professionals semantic decision support functionalities able to contain risks (due to medical errors and adverse events) in order to enhance patient safety.

The remainder of this paper is organized as follows. Section 2 describes related work in the field of healthcare information systems. Section 3 sketches the OCKRF which the CPMS is founded on. Section 4 depicts system features and architecture by using an application example. Finally, Section 4 concludes the paper and sketches future work directions.

\section{Related Work}

In the recent past, a strong research effort has been taken to provide standard representations of declarative and procedural medical knowledge. In the following, available approaches and systems for medical ontologies and clinical process representation and management are described.

Medical knowledge representation area provides one of the most rich collection of domain ontologies available worldwide. A very famous and widely adopted thesaurus is Mesh the Medical Subject Headings classification (MESH). It provides a controlled vocabulary in the fields of medicine, nursing, dentistry, veterinary medicine, etc. MeSH is used to index, catalogue and retrieve the world’s medical literature contained in PubMed. Another classification, that has become the international standard diagnostic classification for all medical activities and health management purposes, is ICD10-CM (ICD; WHO) the International Classification of Diseases Clinical Modification, arrived to its 10th Revision. The most comprehensive medical terminology developed to date is SNOMED-CT (SNOMED), the Systematized Nomenclature of Medicine Clinical Terms, based on a semantic network containing a controlled vocabulary. Electronic transmission and storing of medical knowledge is facilitated by LOINC, the Logical Observation Identifiers Names and Codes (LOINC), that consists in a set of codes and names describing terms related to clinical laboratory results, test results and other clinical observations. Machine-readable nomenclature for medical procedures and services performed by physicians are described in CPT, the Current Procedural Terminology (CPT), a registered trademark of the American Medical Association. A comprehensive meta-thesaurus of biomedical terminology is the NCI-EVS (NCI-EVS) cancer ontology. Some medical ontologies are, also, due to European medical organizations. For example, CCAM the Classification Commune des Actes Médicaux (CCAM), is a French coding system of clinical procedures that consists in a multi-hierarchical classification of medical terms related to physician and dental surgeon procedures. A classification of the terminology related to surgical operations and procedures that may be carried out on a patient is OPCS4, the Office of Population Censuses and Surveys Classification of Surgical Operations and Procedures 4th Revision (OPCS-4), developed in UK by NHS. The most famous and used ontology in the field of healthcare information systems is UMLS, the Unified Medical Language System (UMLS), that consists in a meta-thesaurus and a semantic network with lexical applications. UMLS includes a large number of national and international vocabularies and classifications (like SNOMED, ICD-10-CM, and MeSH) and provides a mapping structure between them. This amount of ontologies constitutes machine-processable medical knowledge that can be used for creating semantically-aware health care information systems.

The evidence-based medicine movement, that aims at providing standardized clinical guidelines for treating diseases (Sackett et al., 1996), has stimulated the definition of a wide set of approaches and languages for representing clinical processes. A well known formalisms is GLIF, the Guideline Inter-
change Format (GLIF). It is a specification consisting of an object-oriented model that allows to represent sharable computer-interpretable and executable guidelines. In GLIF3 specification is possible to refer to patient data items defined by a standard medical vocabularies (such as UMLS), but no inference mechanisms are provided. Proforma (Sutton and Fox, 2003) is essentially a first-order logic formalism extended to support decision making and plan execution. Arden Syntax (Pryor and Hripcsak, 1993; Peleg et al., 2001; HL7) allows to encode procedural medical knowledge in a knowledge base that contains so-called Medical Logic Modules (MLMs). An MLM is a hybrid between a production rule (i.e. an “if-then” rule) and a procedural formalism. It is less declarative than GLIF and Proforma, its intrinsic procedural nature hinders knowledge sharing. EON (Musen et al., 1996) is a formalism in which a guideline model is represented as a set of scenarios, action steps, decisions, branches, synchronization nodes connected by a “followed-by” relation. EON allows to associate conditional goals (e.g. if patient is diabetic, the target blood pressures are 135/80) with guidelines and subguidelines. Encoding of EON guidelines is done by Protégé (Protégé) knowledge-engineering environment.

The group of formalisms, presented above, aim at representing either static (ontologies) or procedural (guidelines) medical knowledge. They pay less or no attention to the combined specification and management of both static and procedural aspects of medical knowledge. Furthermore, they do not hold ad-hoc mechanisms that allow errors and adverse events prevention. In fact, no reasoning facilities able to exploit patient and disease status, prescribed cares and drugs, current activities to execute, etc. are provided. This limitations hinder clinical processes monitoring aimed at discovering situations that could create risks for patients during processes execution.

Two interesting healthcare information systems that provide comprehensive framework for managing clinical guidelines are DeGeL (Shahar et al., 2003) and SEBASTIAN (Kawamoto and Lobach, 2005). DeGeL is focused on providing automated support for the specification and implementation of clinical guidelines in the treatment of patients, particularly those with chronic conditions such as diabetes, hypertension and depression. SEBASTIAN (System for Evidence-Based Advice through Simultaneous Transaction with an Intelligent Agent across a Network) captures medical knowledge in XML documents known as Executable Knowledge Modules (EKMs). An EKM encapsulates medical knowledge in a machine-executable format that can be used to generate patient-specific inferences useful for clinical decision support (CDS). EKMs use a patient information model based on the HL7 Reference Information Model (HL7), and medical concepts are preferentially defined using standard vocabularies included in UMLS. These systems are mainly designed to support decisions during diagnosis or guidelines application but do not support complex reasoning over available knowledge for risk management scopes. A further remarkable system is ASTI (Séroussi et al., 2001) because it is the only existing system that tackles the problem of errors prevention in prescriptions. The ASTI project, in fact, has been focused on the design of a guideline-based decision support system to help general practitioners avoid prescription errors and comply with best therapeutic practice, specifically in the treatment of chronic diseases including hypertension.

From the above discussion of related works emerges that already existing approaches and systems: (i) have a lack of mechanisms for errors and risks handling and prevention; (ii) do not use the same formalism for representing and managing both medical knowledge and clinical processes, hence, they are not able to exploit in a fully unified way declarative and procedural knowledge during execution and monitoring of clinical process; (iii) do not allow to customize represented knowledge and to execute clinical process in a flexible and agile way. The system described in the following section aims at overwhelming these limitations.

3 THE FRAMEWORK

The Ontology-based Clinical Knowledge Representation Framework (OCKRF) allows to represent machine-executable and flexible model of declarative (static) and procedural (dynamic) medical knowledge. The framework is organized in tree layers as shown in Figure 1.

The first layer, called OCKRF meta-model is founded on an ontology-based approach to medical knowledge representation that allows to express in a combined way medical ontologies and clinical processes. The adopted approach is grounded on a meta-model that merges expressive power of ontology and workflow representation formalisms. More in detail, the meta-model is expressed by means of a formalism based on an object-oriented version of datalog that holds typical ontology representation constructs.
like: classes, relations, attributes, objects (instances). Furthermore, querying and reasoning capabilities of Datalog allow to query represented ontologies and execute reasoning tasks over them in a semantic fashion. The meta-model exploits a flow-graph oriented workflow modeling approach (Figures 2 and 3) inspired to the JPDL (JPDL). In a flow-graph based approach a workflow is represented by a labeled directed graph whose nodes correspond to the activities to be performed, and whose arcs describe the precedences among them. The key idea, which the presented framework is based on, is that elements of the workflow meta-model (i.e. processes, nodes, tasks, events, transitions, actions, decisions) are expressed as ontology classes. So, by using the adopted approach ontologies of clinical processes can be obtained. So, by using the adopted approach: (i) clinical processes and their elements can be organized as an ontology (for instance, possible events can be defined as a taxonomy in which each class of event has a specific set of attributes); (ii) workflow elements like nodes, tasks, activities, events etc. are represented in terms of ontology classes so types of their parameters can be other classes, hence clinical process can be managed in a semantic fashion; (iii) decisions can be defined as reasoning tasks that involve not only the execution state but also the knowledge represented in the medical ontologies (for example, at each prescription a reasoning task can check if the administered drug is compatible with allergies and/or with the state of the patient); (iv) special reasoning tasks aimed at controlling possible risks and errors conditions can be defined for a process (reasoning tasks generate events that are properly handled in the process).

The second layer is constituted by medical ontologies and clinical processes expressed by the formalism provided in the first layer. In particular: (i) medical ontologies represent concepts related to different medical domains (e.g. diseases, drugs and their interactions, prescription criteria, medical examinations, medical treatments, laboratory terms, anatomy, patients administration concepts, possible errors). Medical ontologies can be obtained by importing already existing ontologies and thesaurus or designed by hand (for example, concepts regarding patients, wards, hospital, could depend from the specific hospital or ward). For instance, concepts related to breast neoplasm, needed in the process presented in the Section 4.1, has been imported from the International Classification of Diseases (ICD10-CM), the Anatomical Therapeutic Chemical (ATC) classification system, and the Medical Subject Headings (Mesh) Tree Structures. (ii) clinical processes are both medical guidelines and clinical practices depending from the specific ward. Clinical processes can be imported from already available guideline bases or designed by hand when they represent clinical practices followed in a specific ward for caring a given disease. Medical ontologies and clinical processes are stored in the system knowledge base.

The third layer consists of a set of medical ontologies and clinical process instances stored in a knowledge base. Instances are generated during medical ontologies definition and/or clinical process execution.
Instances are stored in the knowledge base and made available for querying and reasoning. More in general, the OCKRF allows methods for: (i) creating ontologies of clinical processes that can be queried and explored in a semantic fashion; (ii) expressing errors and risks rules (by means of reasoning tasks) that can be used (during processes execution) for monitoring processes; (iii) monitoring clinical processes during the execution by running reasoning tasks that enable to prevent errors and adverse events that can cause risks for patients; (iv) to define semantic Electronic Medical Records schemas (i.e. Meta EMR); (v) analyzing acquired clinical process instances, by means of querying and inference capabilities, in order to recognize errors and risks for patients. For example, advanced browsing capabilities of the medical knowledge base by using concept-based queries that enables to retrieve clinical processes and guidelines by using available medical ontology concepts.

The OCKRF is implemented by the DLP+ ontology language (Ricca and Leone, 2007) that beside complete and expressive ontology representation features, holds also powerful ASP reasoning capabilities (Eiter et al., 1997; Leone et al., 2006; Ricca and Leone, 2007) over represented knowledge.

4 SYSTEM DESCRIPTION

In this section are described architecture (Figure 4) and functionalities of the the Clinical Process Management System (CPMS) prototypical implementation. The prototype has been obtained by combining the JBPM engine (JPDL) and the DLV+ system (Ricca and Leone, 2007). The prototype is designed to follow a clinical processes life-cycle model based on 3 phases: (i) representing (importing) medical

clinical processes and ontologies; (ii) executing and monitoring clinical processes; (iii) acquiring, querying and analyzing clinical process instances. Each phase is implemented by an ad-hoc software module as described in the following.

4.1 Clinical Processes Design

The Design module mainly exploits the DLV+ system. It provides functionalities for defining medical ontologies and clinical processes by: (i) import facilities that allow to acquire already existing ontologies and guidelines; (ii) direct "on-screen" drawing and manual specification functionalities based on an ontology and process editor. The editor enables agile guidelines representation and browsing. It contains a rules editor that allows to define ontology constraints and/or reasoning tasks used to control risks and errors during processes execution. Process elements (e.g. nodes, tasks, decisions) that do not belong to a specific process can be, also, represented. Acquired and/or represented schemas and instances are stored in a knowledge base and can be queried by using querying and meta-querying capabilities of DLV+ system.

In order to briefly describe how the design module works, in the following is presented, an application of the system to a real case concerning a clinical process for caring the breast neoplasm. The clinical process (depicted in Figure 5) is referred to the practices carried out in the oncological ward of an Italian hospital, for this reason it is not a general guideline but a specific clinical process adopted in the domain of the considered ward.

The clinical process is organized in the following 10 activities and sub-processes:

1. The task node acceptance models the patient enrollment. The patient arrives to the ward with an already existing clinical diagnosis of a breast neoplasm. This node is manually executed by an oncologist that collect patient personal data that are
stored as new instances of the classes that describe process and patient information.

2. The group node *anamnesis* represents a set of anamnesis activities that can be executed without a specific order. Activities in the group are: (i) *general anamnesis*, in which physiological general data (e.g. allergies, intolerances) are being collected; (ii) *remote pathological anamnesis*, concerning past pathologies; (iii) *recent pathological anamnesis*, in which each data or result derived from examinations concerning the current pathology (or pathologies) are acquired.

3. The task node *initial clinical evaluation* allows to acquire the result of an examination of the patient by a local oncologist.

4. The decision node *more clinical test requested* represents the decision of the physician about the necessity to perform or not additional examination on the patient.

5. The group node *other exams* models possible additional clinical tests. Each node of the group correspond an exam. If requested these tests are conducted to find out general (or particular) conditions of patient and disease not fully deductible from the test results already available.

6. The task node *therapeutic strategy definition* models guideline to follow for caring the given neoplasm. At design time the physician picks a guideline (among those available in the knowledge base) that depends upon actual pathology state as well as other collected patient data. The semantic indexing strategy allows the retrieval of guidelines by using concept-based queries. At each guideline corresponds the prescription of related drugs. So, the selection of a guideline implies computation of doses, which may depend on patient’s biomedical parameters, such as body’s weight or skin’s surface. Cross-checking doses at execution time (by ad hoc reasoning tasks) is fundamental here, because if a wrong dose is given to the patient the outcome could be lethal.

7. The task node *informed agreement sign* models the agreement of the patient concerning understanding and acceptance of consequences (either side effects or benefits) which may derive from the chosen chemotherapy, and privacy agreements.

8. The sub-process node *chemotherapy administration*, models the guideline to execute for caring the patient.

9. The decision node *therapy ended* models a control about effect of the therapy and the possibility to stop, continue or change cares.

10. The task node *discharging* models the discharging of the patient from the ward and allow to acquire final clinical parameter values.

By using the rule editor, for each clinical process (or its elements) a set of risk and error conditions can be described in terms of ontology constraints and/or reasoning tasks. So during both manual and automatic execution of clinical processes, these conditions can be executed in order to check possible risks and errors that are going to happen. If a condition is verified, the system generates an event that alerts the actor that is executing the activity, so risks and errors can be prevented.

4.2 Clinical Processes Execution and Monitoring

The *Execution & Monitoring* module provides functionalities for the assisted execution of clinical processes and the acquisition of process instances. The module is mainly constituted by the JBPM engine that interact with DLV+ system.

Clinical process execution is performed in two ways: (i) by using a workflow enactment strategy. In this case, a process schema, designed and stored in DLV+, is imported in JBPM and automatically executed involving actors that can be humans or machines (e.g. legacy systems supplying results of medical examinations); (ii) by using a dynamic workflow composition strategy. In this case, nodes to execute are selected step by step by choosing the most appropriate one in a given moment. Nodes are chosen by using semantic querying capabilities of DLV+ system and executed by JBPM. Queries allow to specify patient clinical data and each significant information available in medical ontologies (e.g. drug interaction, allergies, etc). So, queries exploit patient clin-
Clinical data coming from anamnesis and medical examination, and each information available in the particular moment within ontologies and already executed process activities. The execution of each activity is performed either: (i) automatically by executing custom nodes that run java code designed to involve other existing systems (e.g. batch acquisition of clinical results from a computer posted in a laboratory); or (ii) manually by a web based graphical user interface that shows forms to human process actors. During manual execution physician and nurses can store values of activity parameters by filling forms by hand. They can insert the values by writing them in the related textual fields. Special queries and reasoning tasks can be triggered to user input in order to check the correctness of entered values. The execution generates both process and ontology instances that are stored in the DLV+ knowledge base. Since, process elements (e.g. nodes, decisions, transitions, tasks) are represented by means of ontology classes, process instances are in turn constituted by ontology class instances. This way process schemas and instances intrinsically constitute an ontology. Ad hoc extensions permit the interaction between JBPM and DLV+. So reasoning, querying and meta-querying over schemas and available instances are possible.

Clinical process monitoring is based on the automatic execution and verification, on live clinical processes, of ontology constraints and reasoning tasks modeled at design time. This way, the system can generate events that inform the user about exceptions and unusual or undesired behaviors. The availability of domain medical ontologies concerning drugs' interactions, side effects and contraindications could reduce dramatically the probability of fatal mistakes. Furthermore, the set of activities executed until a given moment, the current state of a patient, etc. can be retrieved and visualized by physician and nurse in order to check the evolution of cares.

By considering the example introduced in the previous paragraph, at execution time, for each node of the clinical process and for each ontology concept involved in the process, instances are created. For example results of the anamnesis and the check-up are stored as instances of the noticed pathologies where attributes are filled with the related observed values. During guideline execution, due to drug inherent toxicity, before each dose administration a check about drug administration is performed by analyzing process evolution, patient conditions, and information about the particular drug contained in the drug ontology (e.g. maximum absolute ratings for certain drug in a "whole life" or other ratings relative to biomedical parameters of the patient). Problems related to drugs administration can be immediately flagged as dangerous, or even lethal, and then associated risk is notified to the oncologist.

4.3 Clinical Processes Analytics

The Analytics module aims at allowing analysis of the clinical processes instances after their acquisition. The execution of clinical processes, in fact, makes available process instances organized as an ontology. This way a large amount of semantically enriched data becomes available for querying and retrieval. Analysis are possible by reports composed by using semantic querying capabilities of DLV+ applied to process instances contained in the system knowledge base.

5 CONCLUSIONS AND FUTURE WORK

This work describes the prototypical implementation of a semantic clinical process management system founded on an ontology-based clinical knowledge representation framework. The system allows to jointly represent both medical knowledge (by means of medical ontologies) and clinical processes (by means of a ontology-based workflow representation approach). The system allows: (i) creating ontologies of clinical processes that can be queried and explored in a semantic fashion; (ii) expressing errors and risks rules (by means of reasoning tasks) that can be used (during processes execution) for monitoring processes; (iii) manual process execution in which each clinical activity to perform in a given moment is chosen by physician on the base of the current configuration of patient and disease parameters; (iv) automatic execution by means of the enactment of a designed process schema; (vi) automatic monitoring of clinical processes execution by running ad-hoc reasoning tasks that exploit knowledge represented in medical ontologies and clinical processes to check possible error and risk causes. The execution of clinical processes allows to acquire process and ontology instances that are stored in a knowledge base. Acquired instances can be analyzed by means of reports obtained by querying the system knowledge base. Currently the system has been applied to the clinical process depicted in Figure 5. The practical application shown that the system enables better health care decision making capabilities that allows health care professionals to improve risks and errors prevention.
The main challenging future research and development problems are the definition of an efficient query engine working on the conjunct representation of workflows and ontologies and the definition of further monitoring and analytical techniques. In particular, existing process instances can be organized in datasets that can be analyzed by means of data and workflow mining techniques aimed at discovering patterns related to risks and adverse events. In particular, workflow mining techniques are able to classify clinical process instances on the base of their behavior and, possibly, to suggest new schemas (precess re-enengineering) able to reduce risks for patients and the impact of errors. Furthermore, new import/export features will be implemented in order to make the system compliant with already existing health care information systems standards.

REFERENCES

CCAM. Classification Commune des Actes Médicaux (CCAM). http://www.ccam.sante.fr/


