Mined-knowledge and Decision Support Services in Electronic Health

Kamran Sartipi, Mohammad H. Yarmand, and Douglas G. Down
Dept. Computing and Software, McMaster University
Hamilton, ON, L8S 4K1, Canada
{*sartipi, yarmanmh, downd*}*@mcmaster.ca*

Abstract

Large organizations in various information domains are constantly facing the challenges of growing size, new business requirements, and customer demands for service agility. As an example, in the healthcare domain provision of unique electronic health record systems (EHR) for patient identification and health history, integration of regional systems into a nation-wide system, information and service sharing, and security and privacy of patient data have generated a set of new challenges. Canada Health Infoway has proposed an information infrastructure for networked healthcare systems that is based on service oriented architecture (SOA) and provides standards for sharing data and services. In this paper, we investigate the provision of mined-knowledge (results of data mining on patient data), clinical decision support systems, and network visualization and monitoring through SOA. We also address the advantages of SOA implementation using an enterprise service bus in order to accommodate these services. Such services can benefit similar domains such as banking, communications, air traffic control, and transportation.

KEYWORDS: SOA; Decision Support System; Healthcare; Data Mining; Mined-knowledge; eHealth; Infoway; Monitoring; Visualization.

1. Introduction

The advancement of software development through component technology and system integration techniques has resulted in a new generation of very large software systems [10]. This new paradigm has intensified challenges including interoperability of heterogeneous systems, sharing and reusing services, management of complexity, and security and privacy aspects. Examples of such globalized systems include: communication systems, banking systems, air traffic systems, transportation systems, and healthcare systems. These challenges require new development and management technologies and processes that fulfill the emerging demands in the networked systems. In this context, Service Oriented Architecture (SOA) plays a key role in the integration of heterogeneous systems by the means of services that represent different systems' functionality independent from the underlying platforms or programming languages. SOA will contribute in relaxing the complexity, leveraging the usability, and improving the agility of the business services. On the other hand, new services may need to be adopted by the SOA community as standard SOA services have proven essential in specific application domains. Examples of such services include: decision support facilities and mined-knowledge provisions that are crucial in administrative or technical decision making processes.

In this position paper, we propose the enhancement of SOA services to incorporate mined-knowledge interoperability as services, along with data interoperability. The application domains include: financial analysis, tourism, insurance, healthcare, and transportation. The application of the provided mined-knowledge at the point of use would boost the accuracy and convenience of decision making by the administrative personnel. We suggest that, in order to provide such mined-knowledge interoperability for large systems, enterprise service bus technology [11] provides the required facilities. In this context, mined-knowledge refers to the results of applying data mining techniques on data repositories to extract patterns and trends in the data. This mined-knowledge would leverage the quality and effectiveness of decision making on data at the point of use. In this approach, data, mined-knowledge, and decision making operations would be represented as independent SOA services that are accessed through standard service descriptions to be used for different purposes. Moreover, we propose services that allow administrative personnel or government agencies to secure effective control and supervision over the quality of service of the networked systems through activity visualization and identification of distribution patterns of services and bottlenecks. We take healthcare as our example application domain and present the significance of our proposal in the context of the electronic health (eHealth) network of systems.

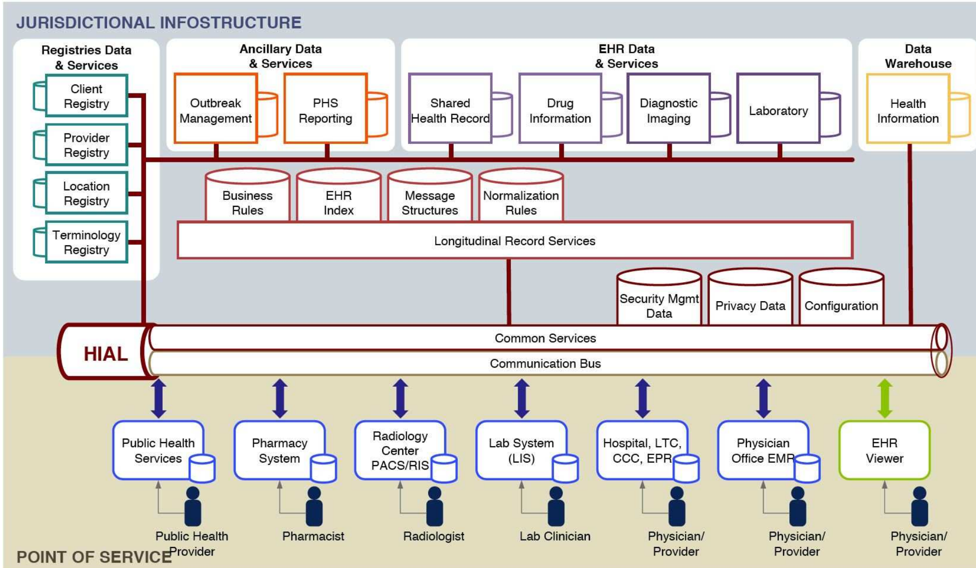

Figure 1. Canada Health Infoway's Infostructure, where the components are defined to provide the SOA services and electronic health record (EHR) technology.

The remainder of this paper is organized as follows. Section 2 presents electronic health systems and service oriented architecture. Section 3 discusses the proposed mined-knowledge services. Section 4 is allocated to the description of supervisory and visualization services as the means for service quality assurance. Section 5 addresses some implementation considerations including PMML and CDA knowledge and data interoperability formats. Section 6 presents related work in the area of SOA and healthcare. Finally, section 7 provides concluding remarks.

2. Electronic health systems and SOA

Modern healthcare is experiencing major changes and as a result traditional conceptions are evolving: from health provider-centric to patient and family-centric; from solitary decision making to collaborative and evidence-based decision making; from decentralized and generalized care to centralized and specialized care [1]. The need for better quality of service, unique identification of health records, and efficient monitoring and administration requires a uniform and nation-wide organization for service and data access. Among other requirements, these changes require extensive assistance from modern software and information technology domains. However, until recently there has been little attention to the IT infrastructure of healthcare systems.

Canada Health Infoway [1] is an organization that provides specifications for a standard and nationwide healthcare infrastructure. Together with public sector partners across Canada, they work to implement and reuse compatible health information systems that support a safer, more efficient healthcare system. The goal is to integrate information systems from different health providers and administrations (e.g., hospitals, laboratories, pharmacies, physicians, and government agencies) within each province, and then connect them to form a nationwide healthcare network with standard data formats, communication protocols, and a unique health history file for each patient; where the health information is accessible ubiquitously, using common services according to different access privileges for patients and providers. Infoway's mission is to foster and accelerate the development and adoption of an interoperable Electronic Health Record (EHR) system [1] which is compatible with standards and communications technologies and includes information relating to the current and historical health, medical conditions and medical tests of its subjects and is typically accessed on a computer or over a network.

Service Oriented Architecture (SOA) consists of the following concepts: *application frontend*, *service*, *service repository*, and *service bus*; each summarized as follows. Application frontends use the business processes and services within the system. A service consists of implementation, service contract, functionality and constraint specification, and service interface. A service repository stores service contracts. A service bus connects frontends to the services. Figure 1 illustrates the realization of an SOA specification into the components of the Infoway's architecture.

- *Application frontend*: components for physician, pharmacy, patient, nurse and EHR viewer are considered as frontend applications.
- *Service*: services are distributed among different parts

of this architecture; for an extended list of these services and their types, refer to [5].

- *Service repository*: there are three main service repositories, namely "registries data and services", "ancillary data and services", and "EHR data and services".
- *Service bus*: HIAL (Health Information Access Layer) is responsible for the service bus functionality.

Considering the heterogeneity and age of the connected healthcare systems in a network, the integration process should be performed through a multi-technology facility, provided by "intermediary services" of an SOA architecture. Different implementations of the intermediary services including "technology gateway" and "Facade" can be used for integration of legacy healthcare systems in the Infoway architecture. A technology gateway mediates between two components developed with different technologies to allow seamless inter-communication by using each other's services. Therefore, it provides a means for convenient integration of legacy systems and performs the required data encoding/decoding for communication. Facade includes a common service access layer which provides uniform access to the services of the underlying components. The above techniques can be applied for the design of a service oriented architecture within the Infoway's HIAL communication mechanism.

Infoway's infostructure is mainly intended for transporting clinical documents through a communication framework. However, this architecture can also be used for other purposes such as tele-medicine, where performance guarantees are required. In such cases the performance of the service bus can be configured using technologies such as Message Oriented Middleware.

As mentioned earlier, a typical SOA architecture consists of four main parts: frontend applications, service repository, services, and service bus. We propose adding new general services to SOA systems to make knowledge available for users throughout the whole system as well as to provide monitoring and supervisory facilities for the administrative personnel. These additions are illustrated in Figure 2 and will be discussed in the following sections.

3. Mined-knowledge services

Current SOA services are either "data-centric", i.e., they transport data between two systems, or "logic-centric", i.e., they encapsulate business rules. However, there has been less attention given to knowledge-based services where the embodied knowledge in a specific application domain would be available through standard services. The main reason is the difficulty of precisely encoding different aspects

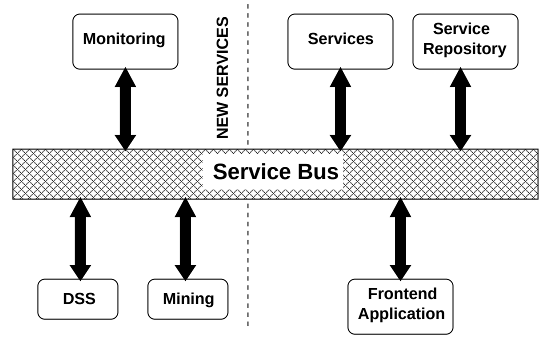

Figure 2. Adding new services to the SOA architecture.

of knowledge so that they can be correctly interpreted at the point of use. Researchers in the healthcare domain are actively working on encoding terminology semantics through terminology systems and clinical guidelines. This has resulted in encoding important business rules and best practice work flows into guidelines to be used by the community.

Another source of valuable knowledge is the knowledge that is extracted from a large data set by applying data mining algorithms. This knowledge includes non-trivial patterns and trends among data that are not easily visible without computing assistance. In this paper we present mined-knowledge interoperability through SOA services.

Decision support system (DSS) component

The proposed DSS component provides two types of services. I) Standard work flow services that reflect the best practices in a domain. In the health care domain, best clinical practices are developed by researchers and practitioners as clinical guidelines to help health care professionals and patients make decisions about screening, prevention, or treatment of a specific health condition. The DSS component offers these work flows to users all over the system. COMPETE [2] is a pioneering Canadian healthcare project in electronic health research that is specialized in developing clinical guidelines based on different available resources. II) Customizable work flow generation services that allow users to define work flows for their enterprises. Such components may provide services for generating new guidelines by professionals as a part of standardization of work flows. This service for generating clinical guidelines would complement the service for accessing a predefined clinical guideline.

Mined knowledge services (Mining)
In flow based decision support systems the input data in various stages of decision making are fed into a decision tree along a decision path from the root node to a leaf node where a decision is made. The choice of branching at each internal node is made by simple or complex logic. This process continues until a final step is reached where a decision is made and a message is generated. In clinical decision support systems (CDSS) for diagnosis of a disease, the required knowledge can be encoded as a flow-based guideline, according to the input data. However, to enhance the task of decision making, additional knowledge would significantly assist the user to make a more accurate decision. This additional knowledge can be in the form of recommendations, alerts, or hidden patterns in data. Such a component can be synthesized using a database of guidelines represented by the service repository and the actual mining action corresponding to the service. In the case of healthcare and the Infoway infostructure, the service repository is placed in the "Registries Data and Services" part and the actual data mining service is located in the "EHR Data and Services" part of the Infoway infostructure.

4. Supervisory and visualization services

The visualization of activities in a large network of systems is crucial for administrative personnel to obtain updated and comprehensive insight into the active or passive relationships among different systems within the network. Examples of such networks include: bus and train transportation systems, air traffic control systems, transactions between financial institutions, and healthcare systems. As an application, consider the integrated network of healthcare systems as a large graph with different types of nodes representing clinical and administrative institutions, and types of edges representing categories of interactions among these institutions. Specific techniques from software reverse engineering can be applied to visualize the static or dynamic architecture of a large region of healthcare infrastructure (e.g., a province) from different view points. In this approach, data mining techniques are used to discover non-trivial patterns of interactions among the nodes of the network and provide means for self-management of the network. These services are illustrated as the "monitoring" component in Figure 2. The steps of the proposed approach are as follows:

- *Healthcare network visualization*. A domain model is required to specify the graph node-types (i.e., different kinds of healthcare institutions such as: hospitals, physicians, pharmacies, laboratories, and government agencies) and graph edge-types as abstractions of health related communications between any two nodes (e.g., lab referral, drug prescription, billing statements, and patient electronic record acquisition). In this context, each edge-type will have several sub-types; for example, an edge-type "drug prescription" can have sub-types that represent a specific category of drugs that can be prescribed. This service will be implemented as follows. A common service in each node of the network is needed so that it enables the "supervisory component" to inquire about the network transaction activities of every node in the network within a certain period of time. On a daily basis, the interactions among the healthcare institutions will be logged and at the end of the day these logs will be sent (through service invocation) to a supervisory component to be analyzed. Consequently, the supervisory component can provide different views of the graph of the healthcare network, where the nodes and edges are color-coded according to the type of institutions and the types of interactions.

- *Discovery of patterns of interactions.* The application of association rules for mining algorithms on the generated graph would identify groups of maximally associated graph nodes according to specific graph edge-types. In this context, maximal association refers to a group of nodes that all share the same services from a maximal group of service providers (e.g., specific categories of medications in pharmacies; and blood tests, X-rays, and ultrasounds in laboratories). A variety of data mining applications can be used to explore non-trivial properties in this network such as: spread of epidemics; distribution patterns of patients in particular regions; or distribution patterns of specific health services. The discovery of such patterns would enable the healthcare administrations and government agencies to restructure the service locations in order to reduce cost of service and to increase the accessibility of services to a larger population. The results would be available through a set of "monitoring and supervisory services" to the healthcare administrative personnel for analysis and policy decisions.

5. Implementation considerations

In this section, the realization aspects of the proposed services are discussed.

Mined-knowledge invocation
For incorporation of mined-knowledge in the steps of the decision flow tree discussed in Section 3, we require a mechanism to identify which parts of the mined-knowledge (stored in a knowledge repository) are required for that

particular situation. This can be handled by using context aware software agents that capture the context of the input data or the kind of work flow process or the privilege of the user, and accordingly trigger events that invoke proper mined-knowledge services. We suggest using the enterprise service bus (ESB) mechanism for implementing such context aware DSS, as the required event-driven features already exist in the ESB-based SOA standards. When a context aware software agent captures a specific context, a service request including the context is sent to the mined-knowledge component. In response, the component returns the requested mined-knowledge from the knowledge repository to be used at the particular decision making step of the DSS.

Mined knowledge and data interoperability
We adopt Predictive Model Markup Language (PMML) [12] to encode mined-knowledge in an off-line operation. A PMML specification is in the form of an XML document and encodes various types of *mining models*, including clustering, regression, and association rules. A mining model refers to the data structures that hold the results of applying data mining operations on data sets. A comprehensive discussion on application of PMML for mined-knowledge interoperability has been presented in [12].

Data mining models also require the input values to meet specific constraints, where these constraints are critical for correct application of the models. As an example the values of a numerical mining attribute should be within prespecified valid ranges, or categorical attributes should not have values outside of the acceptable values set.

We use *schematron validation documents* to validate patient data that are encoded as XML instances (also called Clinical Document Architecture CDA). The validation documents encode the data mining model's input data constraints and hence are tightly bound to the data mining model's PMML document. These two documents (i.e., PMML and schematron) should be ported to the usage site so that the data mining model can be applied on validated data items.

The extracted mined-knowledge from data is stored in knowledge repositories and transported using PMML documents via the service bus.

In order to overcome the problems caused by heterogeneous data formats in participating systems, an interoperable and standard data format is required. In the health care domain the Health Level Seven institute (HL-7) has defined an interoperable data format for clinical documents, namely Clinical Document Architecture (CDA) [7]. Figure 3 illustrates the steps for the proposed data and mined-knowledge interoperability within the context of the healthcare domain.

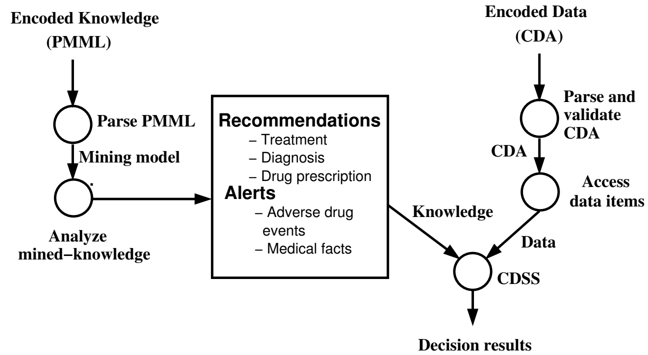

Figure 3. The steps for the proposed data and mined-knowledge interoperability.

Work flow services
Clinical work flow services in the healthcare domain, i.e., clinical guidelines, can be implemented using Guideline Interchange Format 3 (GLIF3) [6] that models clinical best practices as flow charts. GLIF3 guidelines are defined at three levels of abstraction, as follows:

- *Conceptual level*: the first level is a flow chart that represents different states and actions in a structured graph. This level provides an easy to comprehend conceptualization of the guideline.
- *Computable level*: to make the guideline flows computable, the author has to specify the control flow, decision criteria, medical concepts, and relevant patient data.
- *Implementation level*: for GLIF3 guidelines to be actually deployed at an institution site, the patient data and actions should be mapped to institution specific information systems.

6. Related work

In this section, the organizations that either provide or specify SOA services in the healthcare domain will be discussed, along with several projects for which it is suitable to apply SOA services in their architectures.

IBM has strongly supported SOA through investing more than US $1 Billion per year in SOA related projects in various domains, and has secured more than 300 SOA-related patents. IBM HCN (Healthcare Collaborative Network) [8] is a solution that provides an integrated, web-based information network to enable secure transmission of healthcare data. This network helps participants to access patient information to improve decision making across healthcare institutions.

Health Level Seven (HL7) is an ANSI organization accredited to produce clinical and administrative data in the healthcare domain to address interoperability issues. Object Management Group (OMG) [3] is a computer industry consortium which develops enterprise integration standards for a range of technologies. The Healthcare Services Specification Project (HSSP) is a joint project by HL7 and OMG which focuses on service identification for implementation purposes, using SOA principles.

Canada Health Infoway [1] is a government supported not-for-profit organization which has suggested consistent healthcare standards throughout Canada. Infoway's *infostructure* is based on SOA technology and its goal is to connect provincial networks of health systems to form a nationwide network where electronic health records are accessible from different locations.

Evidence-based Guidelines And Decision Support System (EGADSS) [4] is a stand alone application that assists practitioners at the point of care by automatically generating alerts and reminders. EGADSS adopts a language for encoding medical knowledge to define decision making logic as a set of rules encoded in separate modules.

OSCAR, Open Source Clinical Application Resource Service [9] provides tools for clinic operations, prescriptions, and laboratory use. Currently OSCAR is being used by several healthcare institutions in Canada. OSCAR is a web based environment for patient and provider interactions.

Computerization Of Medical Practices for the Enhancement of Therapeutic Effectiveness (COMPETE) [2] is a project that intends to provide computer-based decision support facilities for managing patients with diabetes, hypertension, cholesterol, history of heart disease and/or strokes, and chronic disease. The knowledge-base of COMPETE III consists of a set of guidelines, each represented using a function table. The designers have paid particular attention so that the content of the messages is simple yet accurate, and to prevent the user from an overwhelming number of possible recommendations.

7. Conclusion

The current state of software development technology is moving towards more flexible integration of heterogeneous systems in order to adapt to the challenges of growing networks of very large systems consisting of a variety of platforms and programming languages. In this context, service oriented architectures (SOA) will contribute in relaxing the complexity, leveraging the usability, and improving the agility of business services. In this paper, we proposed a set of new services for the healthcare domain in three categories: flow-based decision support services (DSS), mined-knowledge services, and pattern-driven visualization and supervisory services. We argue that these services are essential in networked healthcare systems in order to enable administrators to control the quality of services. This is achieved through visualization of data interactions among healthcare institutions and extracting patterns to identify certain features such as breakout of disease or distribution of patients in particular regions. Moreover, the amalgamation of mined-knowledge with best practice work flows (as guidelines) will allow professionals to make more accurate decisions, which will in turn have drastic impact on government policy.

References

[1] Canada Health Infoway . http://www.infoway-inforoute.ca/.

[2] COMPETE. Computerization of Medical Practice for the Enhancement of Therapeutic Effectiveness. http://www.compete-study.com/index.htm.

[3] CWM. The Common Warehouse Metamodel . http://www.omg.org/technology/cwm/.

[4] EGADSS. Evidence-based Guidelines And Decision Support System. http://egads.org.

[5] EHRS Blueprint - Infoway Architecture Update. http://www.infoway-inforoute.ca/.

[6] GLIF. Guideline Interchange Format. http://www.glif.org/glif_main.html.

[7] HL-7. Health Level 7. http://hl7.org/.

[8] IBM. http://www.IBM.com/.

[9] Open source clinical application resource service. http://oscarservice.com/index1.html.

[10] ULS. Ultra-Large-Scale Systems; a Workshop at OOPSLA 2006. http://www.dreamsongs.com/Feyerabend/Breathturn2006.html.

[11] D. Chappell. *Enterprise Service Bus: Theory in Practice*. O'Reilly, 2004.

[12] R. S. Kazemzadeh and K. Sartipi. Incorporating data mining applications into clinical guidelines. In *Proceedings of the IEEE International Symposium on Computer-Based Medical Systems (CBMS'06)*, pages 321–326, 2006.