Supporting Population Centered Medical Decision Making: Design Recommendations and Preliminary Assessment

George Potamias [1,2]
Lefteris Koumakis [1,2]
George Charissis [1,3]
Vassilis Moustakis [1,4]
Manolis Tsiknakis [1]
Stelios Orphanoudakis [1,2]

[1]Institute of Computer Science, FORTH
711 10 Heraklion, Crete, Greece
{potamias,koumakis,charissi,moustaki,tsiknaki,orphanou}@ics.forth.gr

Abstract

Architectural design recommendations of an integrated environment for the seamless integration and intelligent processing of distributed and heterogeneous clinical information is presented. This is achieved via the appropriate customization of XML and data-mining technologies. The basic objective of the work is to ease health-care professionals to orient themselves in the disperse information space and enhance their decision-making capabilities. Preliminary assessment of the development shows the technical feasibility of the proposed approach.

1 Prelude: The Scene

Information and communication technologies are contributing to a revolution oriented towards the generation and diffusion of human knowledge embracing expression of it both at individual and social levels. Explosive growth of the World Wide Web (Web) brings together millions of users cutting across geographical, cultural and scientific boundaries.

Web impacts all organizational and social activities. Impact encompasses perceptual changes as well as changes in the ways the involved activities are performed. An area which has experienced radical change, and continues to do so, is health-care delivery. Health-care professionals may now work in a *distributed information environment*, enabling them to instantly access thousands of applications and a wealth of information.

Seamless data integration coupled with *data-mining* operations offer the ability to automatically extract indicative and representative profiles (i.e., interesting patterns) from vast amounts of data. They enable a 'global view' on the disperse and evolving information space, making information inquiries more natural to the user. Utilization and customization of the related technology will decrease the risk for health-care professionals to "get lost in the information space". Moreover, they will enrich and enhance their *decision-making* capability, in an attempt to meet the increasing demand for evidence-*based* medicine.

---

[2] Department of Computer Science, University of Crete, Crete, Greece
[3] Paediatrics Surgery Clinic, University Hospital of Heraklion, Heraklion, Crete, Greece
[4] Department of Production and Management Engineering, Technical University of Crete, Chania, Greece

2 A Medical Inquiry: Actors & Plots

Physician *A* is interested in the determination of associations between laboratory test findings. A *population centered* exploration should give him/her the chance to explore data stored in a variety of clinical sources without the need to comprehend the specifics of how data at each source are organized. Exploration should focus on *semantics* rather than on syntax. Expected results should take the form:

laboratory_test_findings *-associated with-* [different]laboratory_test_findings
and/or
laboratory_test_findings *-are associated with-* [various]diseases

Exploration results such as the above contribute to the formulation of concrete *knowledge*, which despite the inductive flavor it carries, comes in useful in evidence based medicine or in *epidemiology* studies. With the natural assumption that the discovered associations stand for population *health dynamics*, i.e., the formed queries are assigned to potential *health indicators*, the discovered constructs could be dynamically formed and maintained. In this respect *internet-based epidemiology* is enabled and related studies are facilitated (Potamias & Moustakis, 2001a).

However, what does it take for Mr. A to be able to enjoy such associative results? Current information- search and navigation technology and the related query models realize just 'verification' and not *discovery* processes. But, when confronted with disperse information sources and data-repositories of high volume such query models are difficult to form and even more, they become inadequate to capture the potential variations and the rapid changes in the information space. In other words, the users need to recover their '*orientation*' and regain their *supervision* over the information-space.

3 Design Recommendations

Facilitating satisfaction of information and data exploration activities such as the ones listed in the previous section means that relevant infrastructure should be made available. Infrastructure should be able to empower users by enabling them to post complex queries and to receive results on an ad-hoc basis. Enabling ad-hoc exploration over *distributed* and *heterogeneous* (or D & H, for short) clinical sources, calls for paradigm shift in medical information system design and realization. Current medical information systems focus on patient workflow by examining one patient at a time. We need to decouple ourselves from this constraint and move on to system realizations that focus on "many patients at a time." In other words the need for *population medical decision making* is raised. This is not to replace the normal patient long-life follow-up, as achieved via integrated patient electronic healthcare record services (Katehakis, Orphanoudakis, Sfakianakis & Tsiknakis, 2001), but to *add-value* on the decision-making process itself.

Experience with *HYGEIAnet*- the integrated healthcare network at the region of the Crete island (Katehakis, Orphanoudakis & Tsiknakis, 2000), and especially the respective integrated electronic health care record development (Katehakis et al., 2001), has motivated our work on an enabling architecture, named ***HealthObs***, built on a set of services aimed to facilitate and empower medical actors exploration activities over D & H clinical sources. Salient services incorporated in *HealthObs* include: (a) *Access and interfacing* – inquiring may not be satisfied unless effective access to data is achieved; (b) *Information fusion and homogenization* – individual data are "dressed" appropriately so they can be used despite their heterogeneous origin; (c) *Intelligent information processing* – necessary to support discovery over high volumes of data; and (d) *Personalized information delivery* – encompasses presentation and vehicle of delivery. Furthermore, access, interfacing and inquiring over D & H clinical sources should be done with due respect to *privacy* and other *legal* and *ethical* constraints (Grimson, 2001).

4 *HealthObs* – The Realization

Figure 1, below, depicts *HealthObs* realization. Realization encompasses two layers: *middleware* and *application*. Prominent services identified in previous section form the heart of middleware. Implementation is based on standard software standards and components, which include CORBA/CORBAmed ("OMG") - as the communication infrastructure, RDF/XML ("RDF/XML", 2003) - for the *semantic homogenization* and uniform modeling of clinical information, etc. Middleware stands in between the user and the D & H clinical sources. User – *HealthObs* interfacing is achieved via the application layer, which specializes according to the medical domain of interest, e.g., diabetes, cardiology, pediatrics, etc.

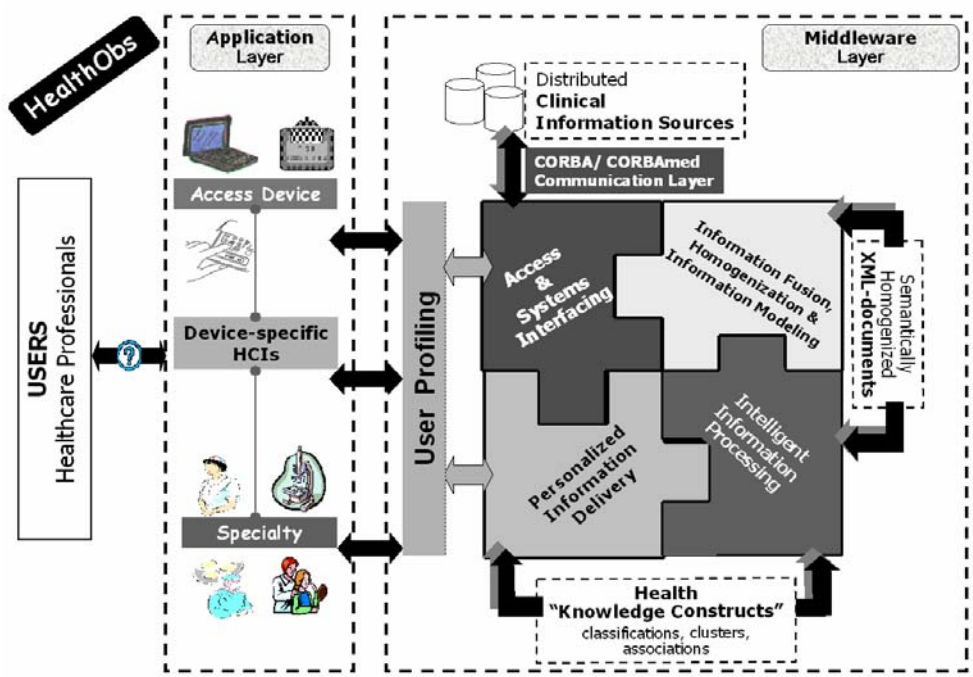

Figure 1: The *HealthObs* Reference Architecture

D & H clinical sources form a federation with middleware in the midst of action. Middleware hides heterogeneity giving the user the impression of uniform structure and behavior across individual sources. This is achieved by equipping systems residing at clinical sources with universal medical clients and by meta-data layering of individual data models. Meta-data layering is facilitated via the use of *Clinical Object Access Service* interface- a standard from the OMG group ("OMG"; www.omg.org).

5 *HealthObs* – Preliminary Assessment

Drawing on *HYGEIAnet* and linked with it electronic health record structure, we conducted a small scale experiment to assess the technical feasibility of *HealthObs*. We targeted experimentation on the discovery of *statistically valid associations* between laboratory test findings from patients in the region of the island of Crete – Greece. We used *Association Rule*

Mining- ARM (Agrawal, Imielinski & Swami, 1993) to enhance modeling of associations between values. ARM functions are appropriately customized to run exclusively on-top of XML-documents (Hristofis et al., 2000). Table 1, below, summarizes (a sample of the) results obtained by using pre-specified support and confidence levels – these levels relate to the statistical validity of induced associations, of course the higher they are set the better.

Table 1: A sample of the discovered clinical association rules
[Posted over the information systems of two remote healthcare centers in the Crete region – concerns an inquiry for associations between biochemical findings – for support >= 60% and confidence >= 80%]

Association Rule	Support (%)	Confidence (%)
OURIA [*Low*] => GLUCOSE [*Low*]	60	85.7
KREATINI [*Normal*] => GLUCOSE [*Low*]	60	85.7
KREATINI [*Normal*] => OURIA [*Low*]	60	85.7
TRIGLIKERIDIA [*Low*] => OURIA [*Low*]	60	100.0

Figure 2, below, shows a snapshot of the (current) *HealthObs* (Java-based) implementation for the discovery of association rules form distributed clinical information sources. Clustering operations are also offered.

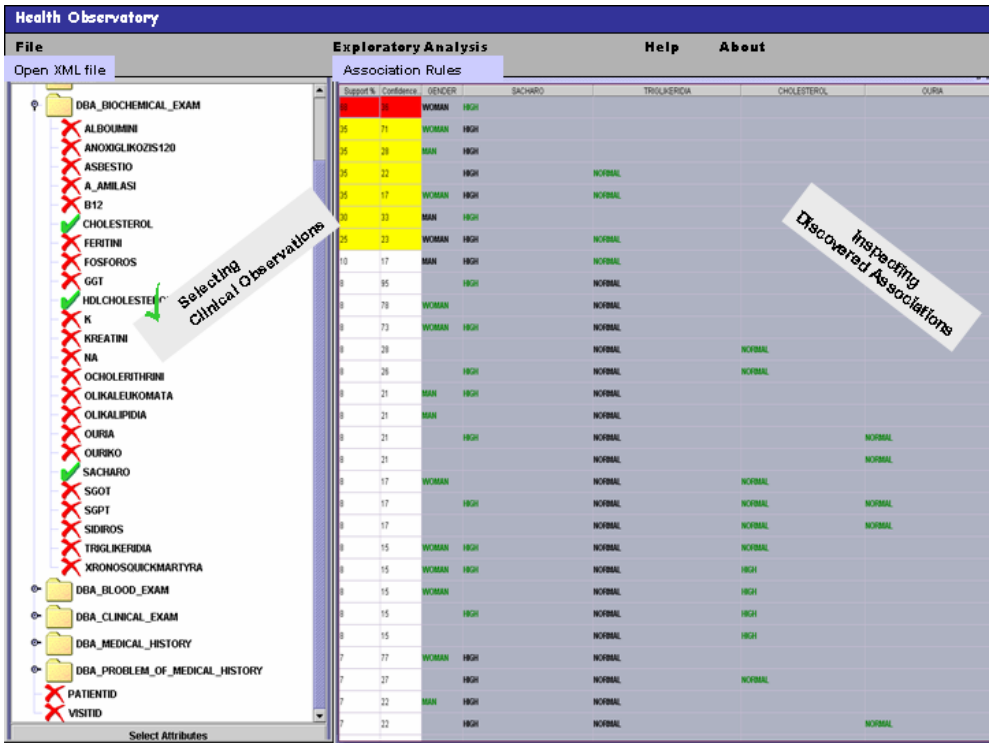

Figure 2: Association Rule Discovery in *HealthObs*

6 *HealthObs* – The Future

The recommended *HealthObs* approach provides for a seamless data integration and intelligent processing environment targeting the healthcare domain. Initial design and preliminary assessment has demonstrated technical feasibility of the proposed architecture.

Similar results were also obtained by applying discovery oriented services on large collections of time series sequences and documents of economic interest (Potamias & Moustakis, 2001b).

Taken altogether we are convinced that the recommended approach is technically feasible and we plan to expand implementation both at middleware and application layer levels to encompass additional association or other data mining approaches and medical domains.

Acknowledgement. The work presented in this paper was partially supported by the IRAIA (IST) project, as well as from other Health Telematics projects within the Institute of Computer Science, FORTH, Heraklion, Crete.

7 References

Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. In *SIGMOD*, (pp. 207-216). Washington D.C., USA.

COAS. (1999). "*Clinical Observations Access Service" (COAS)*, Final Submission, *OMG Document*: http://cgi.omg.org/cgi-bin/doc?corbamed/99-03-25 [accessed 2003 Jan 30].

Grimson, J. (2001). Delivering the electronic healthcare record: a software engineering perspective. In *10th World Congress on Health and Medical Informatics*; September 2-5, London, UK.

Hristofis, K., Potamias, G., Tsiknakis, M., Moustakis, V., & Orphanoudakis, S. (2000). Mining Distributed and Heterogeneous Data Sources: A Project in the Medical Domain. In proceedings *Machine Learning in the New Information Age*, MLnet workshop*, European Conference of Machine Learning (ECML2000)*, (pp. 27-38). Barcelona. Spain.

Katehakis, D.G., Orphanoudakis, S.C., & Tsiknakis, M. (2001). An Architecture for Integrated Regional Health Telematics Networks. In Proceedings *23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Istanbul, Turkey.

Katehakis, D.G., Sfakianakis, S., Tsiknakis, M., & Orphanoudakis, S.C. (2001). An Infrastructure for Integrated Electronic Health Record Services: The Role of XML (Extensible Markup Language). *J. of Medical Internet Research*, 3(1):e7.

OMG/ CORBA/ CORBAmed. The Object Management Group. http://www.omg.org/ [accessed 2003 Jan 30].

Potamias, G., & Moustakis, V. (2001a). Knowledge Discovery from Distributed Clinical Data Sources: The Era for Internet-Based Epidemiology. In proceedings *23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 25-28 October 2001, Istanbul, Turkey.

Potamias, G. & Moustakis, V. (2001b). Mining coherence in time series data. Smith, M. J., Salvendy G., Harris D., and Koubek R. J., editors. In proceedings *9th International Conference on Human Computer Interaction: Symbiosis of Human and Artifact* (pp. 928-932). Lawrence Erlbaum.

RDF/XML. (2003). *RDF/XML Syntax Specification (Revised)*, W3C Working Draft 23. http://www.w3.org/TR/rdf-syntax-grammar/ [accessed 2003 Jan 30].