Knowledge-Mediated Retrieval of Laboratory Observations

Charles Hsu and Howard S. Goldberg MD
Center for Clinical Computing, Beth Israel Deaconess Medical Center, Harvard Medical School
Boston, MA

Intelligent medical applications including agents, clinical decision support systems, and expert systems can benefit from components that expose the meanings of medical concepts. We have endeavored to create an ontology for laboratory observations and to make the ontology accessible in a distributed environment through a knowledge mediator offering several services. To date we have created two such services, one service to mediate the retrieval of laboratory observations and an auxiliary service to facilitate the mapping of units of measure to LOINC property-types. We report progress and insights on the development of our ontology and related knowledge mediator.

INTRODUCTION

One of the challenging issues facing designers of intelligent medical systems is imparting the ability to draw on the meaning of medical concepts. This is facilitated by organizing medical concepts into an explicit knowledge representation, or ontology, which defines concepts in terms of their relations with other concepts. Ontologies provide a rich representation of real-world objects and can function as a base of knowledge from which applications can draw. A number of papers in the literature have reported on the use of ontologies for clinical applications.[1,2]

Health care organizations are developing applications for use in distributed environments, where objects which comprise an application may reside in different locations on the Internet. Distributed environments have several advantages, including the ability of client applications to use objects requiring a great deal of processing power as well as centralized application updates. As component objects of applications, and as applications which could take significant processing power for classification and querying of concepts, ontologies are prime candidates for deployment as distributed objects.

In the Health Object Library On-line (HOLON) environment, we found a need to provide distributed services to mediate access between agents that understand clinical objects and data repositories that understand HL-7 queries. For example, services are needed to mediate requests such as "retrieve all glucoses," which would be important for an insulin-regulation agent which regularly queried heterogeneous repositories (home monitoring repository, hospital repository, local diabetologist's system) for blood glucose measurements. We recognized this as an opportunity to explore the development of a knowledge mediator that would contain a model of laboratory observations and their mapping to concrete lab tests. Therefore, we decided to create an ontology and knowledge mediator encapsulating the Laboratory Observation Identifier Names and Codes (LOINC) database, a set of over 15,000 laboratory observations, and its related medical concepts. We planned to make the mediator available as a HOLON-compliant component, and in doing so, hoped to assess the feasibility and utility of developing knowledge-mediated services to improve clinical care.

Our project goals include:

- Development of a concept model for LOINC observations and LOINC-observation-related medical concepts (e.g., "glucose", "serum"), while adhering to currently existing concept models where they were applicable;
- Inclusion of LOINC observations and related medical concepts in an ontology;
- Development of a knowledge mediator implementing useful services for utilizing the ontology, such as functions for searching for observations of given qualities (e.g., all tests measuring creatinine in the blood);
- Enabling the availability of services in a distributed environment.

We report here our progress in developing our knowledge mediator.

METHODS

Programming environment (Loom)
Loom 4.0 (ISI) is a Lisp-based knowledge-representation environment used for creating ontologies.[3] Loom provides a KRSS-compliant

1091—8280/99/$5.00 © 1999 AMIA, Inc.

description logic; classifier, recognizer, and production rules; and a query language.[4] Loom runs within Allegro Common Lisp 5.0 (Franz) on a Windows NT platform. The ontology was updated directly through ACL's Lisp interface and also through Ontosaurus 1.5 (ISI), a Web browser for Loom ontologies.[5]

Ontology sources

We created our ontology for the Logical Observation Identifier Names and Codes (LOINC) database release 1.0k (Regenstrief Institute & LOINC Consortium), a set of over 15,000 lab tests and clinical observations.[6] The LOINC database contains fields for identifying each of the different lab tests or clinical observations that can be reported by health care providers and also assigns each of these tests a unique LOINC code. LOINC observations are thus identified either by their unique LOINC code or by a LOINC name that is comprised of usually five different components – component, property measured, time aspect, sample-type, and scale-type – separated by colons.

Knowledge mediator services

In order to encapsulate the Lisp-based ontology and services in a CORBA-compliant environment, we used ILU (Xerox), a CORBA-compatible object request broker providing several language bindings including Lisp. Client applications were developed using the Visigenic (Inprise) object request broker, and laboratory observation services were made available to the CORBA environment through the Visigenic Naming Service.

Large-scale load of LOINC observations

LOINC observations and related medical concepts were imported and semi-automatically classified in the ontology through a Java applet we created with JDK 1.2 (Sun).

RESULTS

Concept models

For laboratory observations, we developed a simple concept model which modeled the observation's relations with the five parts of the LOINC name (see Figure 1). For related medical concepts, we developed a concept model which included the concept's unique Unified Medical Language System (UMLS) concept identifier as well as synonyms through SYNONYM relations and parent concepts through IS and IS-PRIMITIVE relations. We also specified the concept's UMLS preferred name through the PREFERRED-NAME relation and the concept's LOINC name through the LOINC-NAME relation, both defined as child relations of the SYNONYM relation. All LOINC observations were modeled as concepts rather than instances for the purpose of being able to reuse the ontology. Our intention is that the ontology will be used by different applications which can instantiate real observations from these concepts.

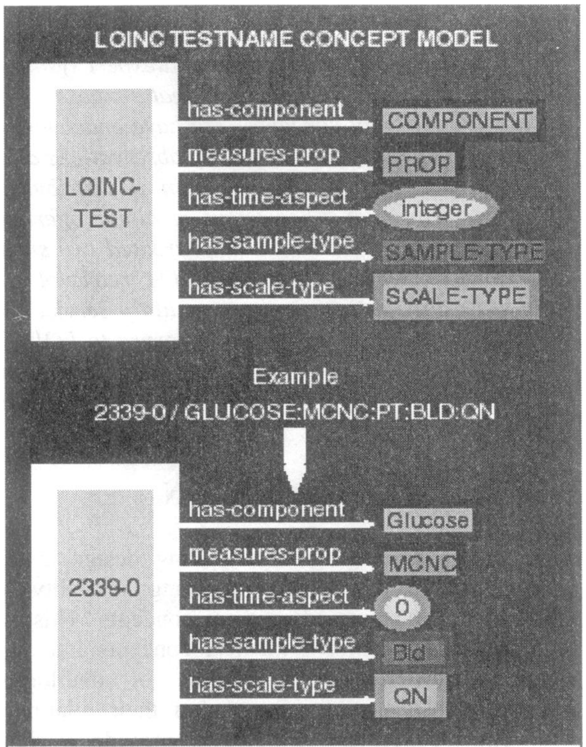


Figure 1 LOINC concept model

Creating the ontology

Discrete ontologic trees were created for LOINC tests and related concepts for each of the different parts of the LOINC name. Concepts were automatically classified by the Loom classifier to produce poly-hierarchical trees. Wherever possible, we used existing terminologic representations that we thought best represented the clinical domain and were a ready source of concepts. For example, in our COMPONENT tree, we reproduced the relevant parts of the Medical Subject Headings (MeSH) classification scheme, and for our ANATOMY tree, we reproduced the relevant parts of the University of Washington's Digital Anatomy (UWDA) project.[7] For portions of our ontology in which existing concept models were unavailable or found to be less applicable, we created our own trees (SAMPLE-TYPE and PROPERTY trees).

Mediator services

We initially focused on building two services. The first service supports our intended application of the knowledge mediator: a general query mechanism for semantic retrieval of LOINC observations. It is clear that the poly-hierarchical tree structure of an ontology would lend to its use for very rich and specific searching in ways that can benefit the delivery of health care. For example, physicians could search for all lab values measuring a particular component in a particular sample type, so that a query for measurements of "glucose" levels in "blood" would not only return LOINC observations matching the concepts "glucose" and "blood" but would also return, for example, LOINC observations for measurements of "glucose" in "venous blood" and of "glucose" in "serum." Agents could also take advantage of this ability, as previously explained, by using the knowledge mediator to find all relevant LOINC observations to query for. This service was successfully implemented by allowing for the retrieval of any terms related through IS, IS-PRIMITIVE, or IS-PART-OF relations (Figure 2).

Service #1: Semantic retrieval

"Retrieve all blood bicarbonate tests"

↓

1959-6 / BICARBONATE:SCNC:PT:BLD:QN
1960-4 / BICARBONATE:SCNC:PT:BLDA:QN
1961-2 / BICARBONATE:SCNC:PT:BLDC:QN
1962-0 / BICARBONATE:SCNC:PT:PLAS:QN
1963-8 / BICARBONATE:SCNC:PT:SER:QN

Service #2: Unit-to-property-type mapping

"What is mg/min a measure of?"

↓

MRAT
(mass rate)

Figure 2 Knowledge mediator services

The second service is intended to facilitate the semi-automated mapping of tests to LOINC observations. Of the five parts of the LOINC name (component, property type, time aspect, sample type, and scale type), the property type is generally not coded into most legacy systems. We therefore developed a service which would provide the LOINC observation property type based on a unit of measure, e.g., given the unit "mg/ml" the service would return "mass concentration." This was done by defining another ontologic tree containing units of measure and their types (e.g., "MG IS-A MASS-UNIT"), and then defining the property-types as having units with numerator of one type and denominator of another (e.g., "MASS-CONCENTRATION HAS-NUMERATOR MASS-UNIT AND HAS-DENOMINATOR VOLUME-UNIT"). The service utilized the Loom classifier to take advantage of these definitions and reason the property type from the unit (Figure 2).

Deploying services in distributed environment

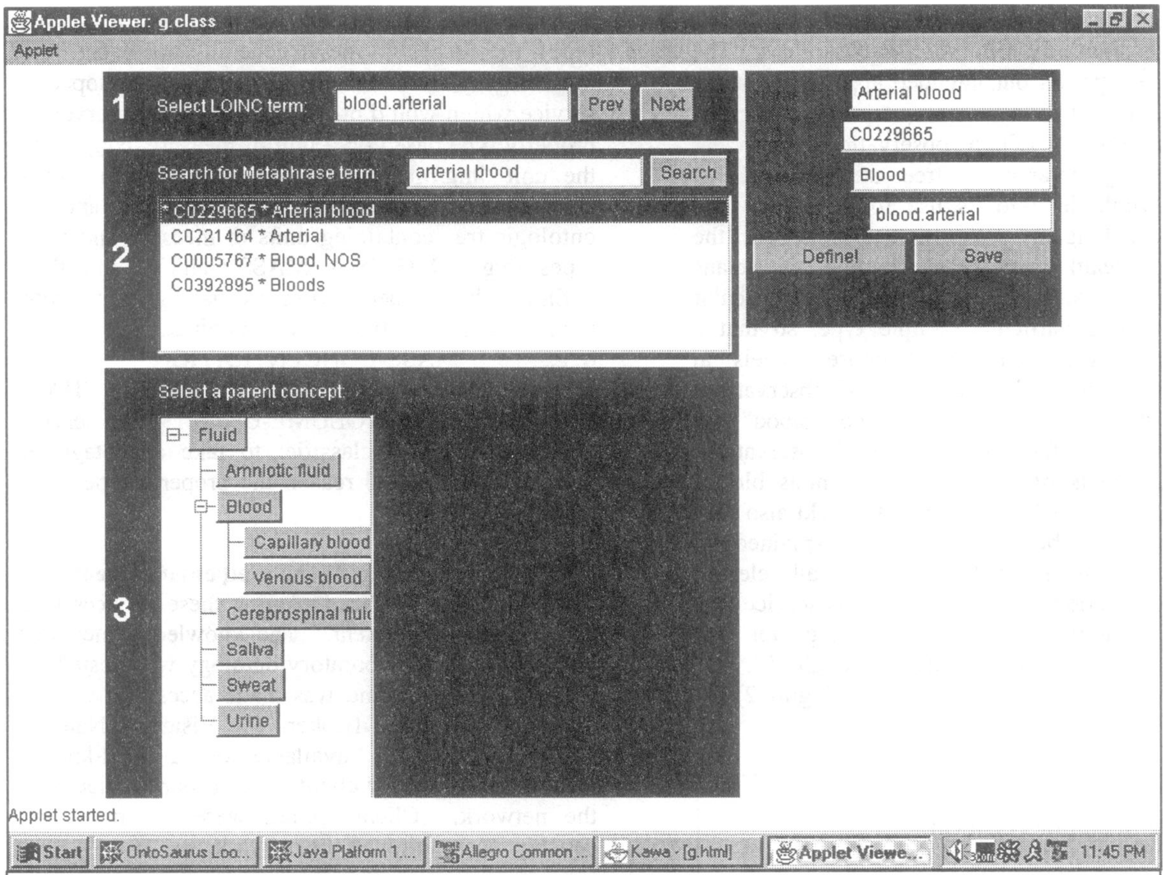
We were successful in deploying these services in a distributed environment. The knowledge mediator encapsulating the laboratory ontology was hosted on its own computer, and was made accessible via the ILU Object Request Broker. The Visigenic Naming Service was made available on a well-known machine in order for clients to find our mediator on the network. Client applets were run from any capable workstation on the network.

Large-scale load of LOINC observations

For large-scale load of the LOINC observations, we are developing an editing interface along with a Java terminology-updater applet to define new LOINC-related medical concepts in a semi-automated manner (Figure 3). In its initial form, the applet performs the following four steps: 1) retrieve a medical term (as a lexical string) from the LOINC database via JDBC/ODBC, 2) query Metaphrase for the term's UMLS concept ID, 3) wait for user input to specify the term's parent concept from the ontology, and 4) define the concept in the ontology. Due to the extensiveness of the Metaphrase search, the applet was constructed to allow users to specify which UMLS concept corresponds to the LOINC term. The applet was successful in defining these concepts and updating the ontology in our distributed configuration.

DISCUSSION

This paper reports our progress toward the development of a knowledge mediator which encapsulates an ontology of laboratory observations and related medical concepts. We have defined viable concept models for laboratory observations and their related concepts by exploiting the LOINC



1 Select LOINC term: blood.arterial

2 Search for Metaphrase term: arterial blood

* C0229665 * Arterial blood
C0221464 * Arterial
C0005767 * Blood, NOS
C0392895 * Bloods

3 Select a parent concept:

- Fluid
  - Amniotic fluid
  - Blood
    - Capillary blood
    - Venous blood
  - Cerebrospinal flui
  - Saliva
  - Sweat
  - Urine

Arterial blood
C0229665
Blood
blood.arterial

Applet started.

Figure 3 Terminology-updater applet

model, implemented them in a Loom-based ontology, developed several initial knowledge mediator services utilizing the ontology, and tested these services in a distributed environment.

One of the strengths of our concept model for medical concepts was its independence from lexical terms, which are subject to change. In particular, in defining our concepts in Loom, we annotated the concept whenever possible with the concept ID for the appropriate UMLS concept, so concepts could be retrieved specifically by concept ID and not only by concept name. This eliminated any dependence of our ontology on the use of any particular concept name. Also, by allowing for synonyms, UMLS preferred names, and LOINC names, concepts can be retrieved by any one of several names, the appropriate UMLS preferred name can be retrieved, and LOINC observation names can be dynamically constructed from its related medical concepts to conform to the LOINC database despite each concept having several names.

Implementing our concept model presented greater challenges than initially appreciated. A significant issue involves the tradeoffs in designing an ontology to be concept-based rather than instance-based. The advantage of this approach is the potential for reuse of the model. This advantage notwithstanding, we found that it was substantially more difficult to manipulate the concept model using Loom's query language, which was designed to manipulate instances. Additionally, in order to support part-whole relations, we had to make assertions (annotations) between concepts. Finally, it remains to be seen how well the current LOINC concept model will be able to accommodate other types of LOINC observations. We began the modeling of the LOINC database by focusing on a subset of LOINC observations which we judged would be simple to model. As we continue to model other subsets of the LOINC database in iterative fashion, however, it is clear that certain observations may be substantially more difficult. For example, ratio tests, such as observations recording a "sodium/potassium" ratio in "blood," may not immediately fit into the current model. What is being measured – a substance or a ratio? Should these observations be retrieved when querying for either or both of the two components in the ratio?

The knowledge mediator was implemented and tested for use in a distributed environment in order to assess the feasibility of access to the ontology from anywhere on the Internet. We reused existing terminologic schemes (MeSH and UWDA) for portions of our ontology in the belief that the sponsoring organizations had already spent extensive time and effort classifying these concepts. It would be interesting to see if independent ontology servers could be implemented for MeSH and UWDA so that the ontology itself could exist in distributed form. Within this paradigm, one could imagine several large ontologies residing in different places, each with its own knowledge mediator, but all accessible via the Internet. Clients could thus access different ontologies through their knowledge mediators and use the information for different clinical applications. With the emergence of a standard set of primitives for defining ontologies, such as those defined in the Open Knowledge Base Connectivity (OKBC) API, it may even be possible to develop generic knowledge mediators which can access multiple ontologies without needing to tailor the knowledge mediator for each individual ontology.[8]

Our strategy for the large-scale load of LOINC observations was to define all LOINC-related medical concepts in the ontology first, and then to automatically define all LOINC observations that referenced these concepts. We therefore created the terminology-updater applet to semi-automatically define component and sample-type concepts such as BLOOD and GLUCOSE, and then used the applet to automatically define all LOINC observations which referenced concepts BLOOD and GLUCOSE. In creating the terminology-updater applet, we defined several additional interfaces for inspecting previously defined concepts, defining new concepts, and saving the ontology to file. Thus, in addition to creating a "dictionary-like" set of knowledge mediator services related to access to the ontology, we also created a set of editing interfaces for updating the ontology. Ultimately, we hope to extend functionality of the terminology-updater applet to be able to define other relations in addition to the currently enabled parent-child IS-PRIMITIVE relations.

CONCLUSION

Our initial progress has shown that a knowledge mediator encapsulating a formal ontology can be developed and deployed in a distributed environment. Knowledge mediators can make the power of a rich knowledge representation available to users or client applications that want to retrieve information without concern for the complexity of the underlying ontology.

ACKNOWLEDGMENTS

Special thanks to Tom Russ (University of Southern California, Information Sciences Institute) for his support in Loom.

This work was supported in part by a cooperative agreement award from the Agency for Health Care Policy and Research and the National Library of Medicine (HS08749), and a contract from the National Institute of Standards and Technology (1995-10-0067N).

REFERENCES

[1] Buchanan BG, Moore J, Forsythe D, Banks G, Ohlsson S. Involving patients in health care: explanation in the clinical setting. SCAMC. 1993: 510-514.

[2] Heathfield HA, Kirby J, Nowlan WA, Rector AL. Pen&pad: a collaborative patient record system for the shared care of the elderly. SCAMC '92, 147-150.

[3] MacGregor RM. Using a description classifier to enhance deductive inference. Proc 7th IEEE Conference on AI Applications, 1991: 141-147.

[4] KRSS Working Group of the DARPA Knowledge Sharing Effort. Draft of the specification for description logics. http://www-ksl.stanford.edu/knowledge-sharing/papers/dl-spec.ps

[5] Swartout B, Patil R, Knight K, Russ T. Toward distributed use of large-scale ontologies. KAW Calgary, 1996.

[6] Forrey AW, McDonarld CJ, DeMoor G, et al. Logical observation identifier names and codes (LOINC) database: a public use set of codes and names for electronic reporting of clinical laboratory test results. Clinical Chemistry 1996;42(1):81-90.

[7] Brinkley JF and Rosse C. The digital anatomist distributed framework and its applications to knowledge based medical imaging. JAMIA 4:165-183, 1997.

[8] Chaudhri VK,Farquhar A,Fikes R, P. D. Karp, Rice JP. Open Knowledge Base Connectivity 2.0. Knowledge Systems Laboratory, KSL-98-06, January 1998.