Toward Reusable Software Components at the Point of Care

Mark S. Tuttle, David D. Sherertz, Nels E. Olson, Stuart J. Nelson, M.D., Mark S. Erlbaum, M.D., Kevin D. Keck, Alex N. Davis, Octavio N. Suarez-Munist, Stephanie S. Lipow, William G. Cole, Ph.D.
Lexical Technology, Inc., Alameda, California
Lawrence M. Fagan[a], Ph.D., M.D., Richard D. Acuff[a], Colleen E. Crangle, PhD, Mark A. Musen[a], M.D., Ph.D., Samson W. Tu[a], Gio C.M. Wiederhold[b], Ph.D., Robert W. Carlson[a], M.D.
[a]Stanford University School of Medicine, Stanford, California
[b]Computer Science Department, Stanford University, Stanford, California

*An architecture built from five software components — a **Router, Parser, Matcher, Mapper,** and **Server** — fulfills key requirements common to several point-of-care information and knowledge processing tasks. The requirements include problem-list creation, exploiting the contents of the Electronic Medical Record for the patient at hand, knowledge access, and support for semantic visualization and software agents. The components use the National Library of Medicine Unified Medical Language System to create and exploit lexical closure — a state in which terms, text and reference models are represented explicitly and consistently. Preliminary versions of the components are in use in an oncology knowledge server.*

THE PROBLEM

Point-of-care software is difficult to deploy successfully. The chief constraint on its use is time. Caregivers seeing four or more patients per hour may have at most a few minutes per patient to exploit software made available to them. One way to increase the chances of success is to make the most of any caregiver input and anything known about the patient at hand.[1] For example, the task of maintaining a patient's problem list can benefit from access to a problem dictionary[2] or from suggestions computed from an encounter note.[3]

The National Library of Medicine (NLM) Unified Medical Language System (UMLS) Knowledge Sources are designed to help solve these and other point-of-care "symbolization" and "symbol linking" problems.[4] But these resources lack a standard or default deployment architecture. If there were such an architecture, evaluations of clinical systems would be more comparable, improvements to the architecture, its components, or the UMLS, could be more easily identified and disseminated, and economies of scale more easily realized. Thus, the problem is how to create and deploy a component-based architecture that fulfills certain point-of-care text processing needs. Recent experience led us to conclude that this architecture should also support semantic visualization,[5] make use of software agents[6] and leverage locally developed domain models.[7]

This problem is a generalization of more specific point-of-care software problems encountered in our work on the following objectives: 1) The development and distribution of a *Problem-list Toolkit (Pl/Tk)* as part of the NLM / AHCPR (Agency for Health Care Policy Research) EMR (Electronic Medical Record) Cooperative Agreement,[8] 2) the use of the contents of the EMR for the patient at hand to help anticipate and fulfill caregiver knowledge needs from a *knowledge server*;[9] 3) the development and deployment of *standards* by which an EMR, a knowledge server, and health care enterprise departmental computers could exchange data, information and knowledge relevant to the patient at hand;[10] 4) integrated use of *interfaces* that support speech, pen, and 3-D input;[11,12] 5) the development, exploitation and maintenance of *domain models* that capture local care practices;[13] and 6) the creation, deployment and maintenance of *software agents* that operate in support of care.[14]

FROM MEANING TO TERM

Figure 1 depicts a view of the "clinical symbolization" problem that emerged from attempts to fulfill a vision described by M.S. Blois, Ph.D., M.D. (1918-88) in the 1980s.[15] Blois proposed that computers act as aids to symbolization, and, more generally, that research and development in medical informatics should focus on the creation, exploitation and properties of patient descriptions that are both human-readable and machine-processible. The view identifies three distinct "worlds" — the patient ("reality"), the caregiver ("mind"), and the computer ("symbols").[16] Implicit in this view is the assertion that concepts in the caregiver's mind are the only things that connect "reality" with "symbols." Our objective is to develop and deploy reusable software components that help caregivers create, leverage, and repurpose clinical symbols relevant to the patient.

0195—4210/96/$5.00 © 1996 AMIA, Inc.

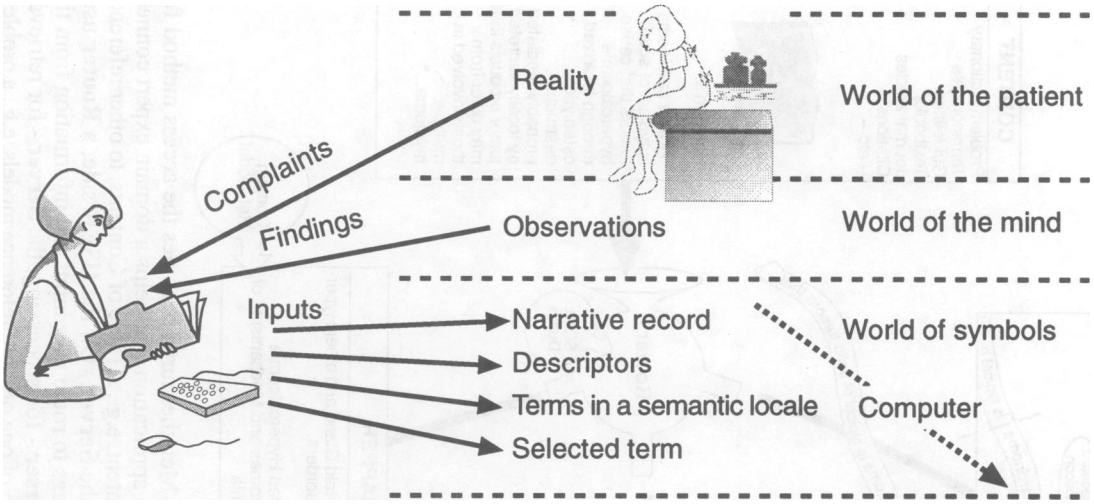

Figure 1 - From Meaning to Term: Given a narrative record of a patient encounter, computers can suggest potentially relevant descriptors by extracting the nominal phrases. Given some relevant descriptors, computers can suggest one or more potentially relevant semantic neighborhoods in terminologies or reference models, e.g., in ICD, SNOMED, or MeSH, or the Table of Contents of a textbook, by matching the descriptors using the UMLS. Given some semantic neighborhoods, computers can suggest a appropriate terms or headings by using information from the EMR to help display the neighborhood. In general, while users must both create symbols and judge their relevance, computers can help translate relevant symbols, e.g., narrative text, into other symbols, e.g., terms from a terminology. (See examples.)

A COMPONENT-BASED ARCHITECTURE

An architecture is an arrangement of components intended to serve a specified purpose. A component-based architecture places special emphasis on the reusability and replaceability of its constituent components by specifying both their behavior and their input-output relationships.

A **Content Machinery** architecture appears in **Figure 2**. Shown there are three Content access methods — (1) concept, (2) schema, and (3) model, arranged vertically — and five components that support these access methods. The left-right symmetry is what produces the *lexical closure* that is exploited by the access methods. Lexical closure permits *the authoritative text to be retrieved — coming in from the right — is processed the same way as the point-of-care text doing the retrieving — coming in from the left.* The top (left-right) paths *link* potential "Concept" names, e.g., those found in a problem list (on the left), with authoritative occurrences of those names (on the right). The middle (left-right) paths support Schema *access*, e.g., the retrieval of the name of an ICD code. The lower (left-right) paths *manipulate* models, e.g., terminology hierarchies or tables of contents.

(1) "Concept"-name Linking

The *top paths* try to link any text fragment to an enhanced version of the current UMLS Metathesaurus residing in the Server.

(2) Schema access

The *middle paths* support Schema access to the Server using either standard or custom query interpreters; these interpreters insert Content (on the right) or retrieve Content (on the left).

(3) Model manipulation

The *lower paths* support access to Reference Models known to the Server through an application called the Mapper. On the right the Mapper helps domain experts integrate Reference Models with one another; on the left the Mapper helps users exploit the Server using Reference Models, e.g., by supporting browsing. On the left any point-of-care Reference Model has special status.

A Scenario

Part of **Figure 2** is a scenario that poses the problem of connecting some of what is known about the patient at hand with an appropriate section of *a* textbook of medicine.[17] On the left are fragments of the patient record including an entry from the problem list; on the right is a potentially relevant fragment of the textbook. Displayed in the Mapper on the right is the local table of contents for the fragment of the textbook; in the Mapper on the left is a portion of a local model of breast cancer.

Five Components

The following five components, shown in **Figure 2**, are as simple as we can make them without

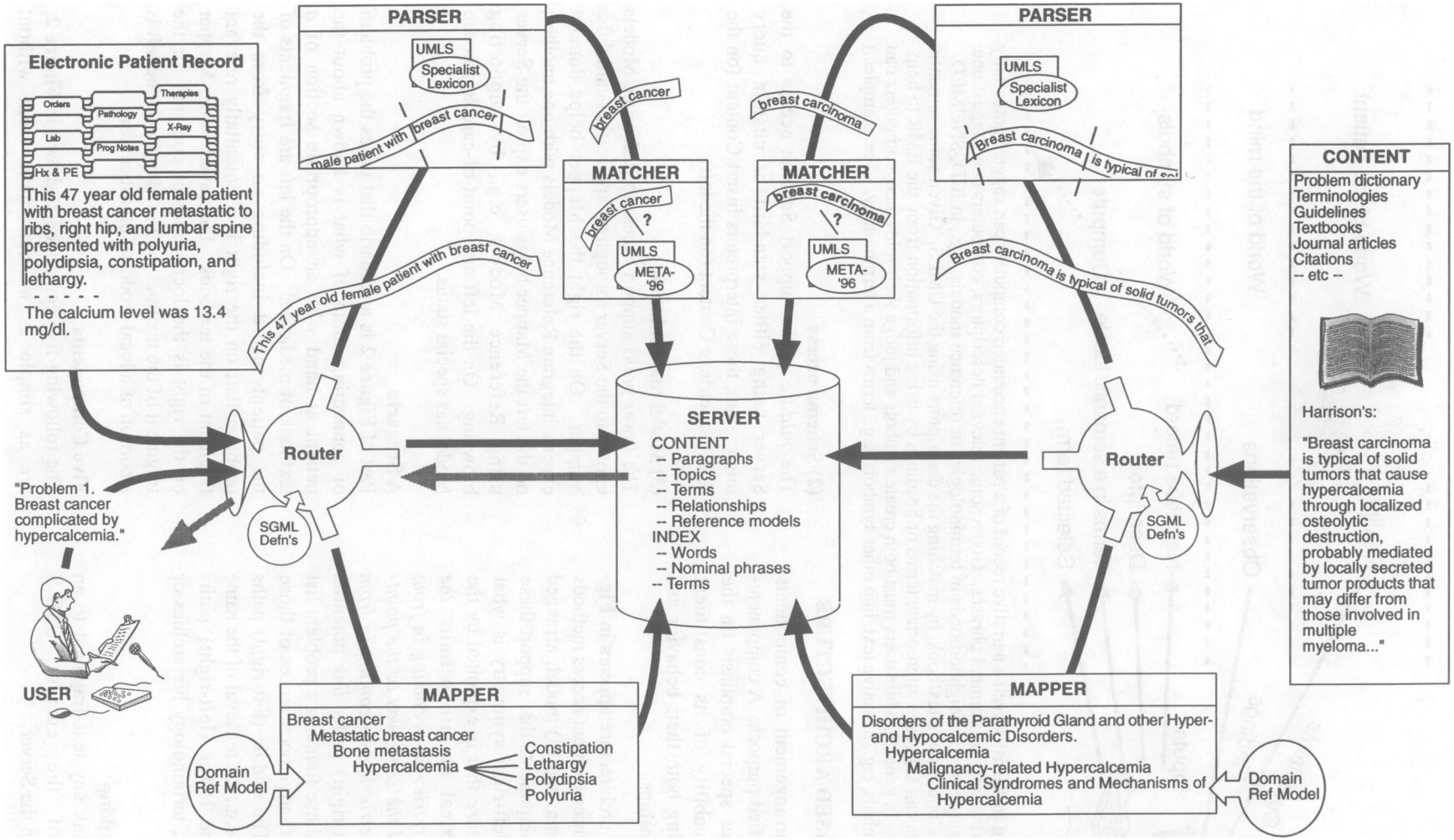

Figure 2 - Content Machinery: A symmetric component-based architecture matches user input and the EMR for the patient at hand to authoritative content in a **Server**. The symmetry guarantees that user input and the EMR are processed in the same way as authoritative content. On the **CONTENT** side, a **Router** uses content-specific SGML tags to route text and images to one or more of the **Parser** - for processing, the **Server** - for storage, or the **Mapper** - for connection to reference models. The **Parser** converts text into nominal phrases and words, and, when signaled by tags, terms and titles. The **Matcher** uses lexical matching to convert the output of the **Parser** into an enhancement of the 1996 UMLS Metathesaurus. The enhanced Metathesaurus becomes the access method for the **Server**. The **Mapper** is an application that helps a domain expert connect the structural components of content, e.g., a Table of Contents, to other reference models, e.g., ICD, already in the **Server**. On the **USER** side, a **Router** uses point-of-care-specific SGML tags to route user input and information from the EMR to one or more of the **Parser** - for processing, the **Server** - for retrieval, or the **Mapper** - for initialization of one or more reference models, e.g., a problem dictionary. Locally designed domain reference models, e.g., for an ambulatory care breast cancer clinic, can have special status in the **Mapper**.

compromising the purpose of the architecture. Each component uses one or more of the UMLS Knowledge Sources and computer programs — the latter are used as de facto standards.

The **Router** uses explicit tags to send text and objects through one or more Server access methods. Specifically, on each side of the diagram, the **Router** directs text and objects through one or more of the upper (Concept), middle (Schema) or lower (Reference Model) paths, depending on explicit tags, e.g., SGML, in a standard representation. On the **USER** (left) side, the **Router** directs user input and information from the **EMR** to one or more of the **Parser, Server,** or **Mapper** so as to generate retrievals from the **Server**. On the **CONTENT** (right) side, the **Router** directs content to one or more of the **Parser, Server** or **Mapper** so that it can be made explicitly available using words, nominal phrases, terms, queries and Reference Models.

The **Parser** extracts retrieval units from text. Specifically, on both sides of the diagram, the **Parser** produces words, normalized words, nominal phrases, and normalized nominal phrases from its input. Nominal phrases are short noun phrases presumed to name concepts. These phrases are extracted using a syntax-driven parser. A simply form of such a parser, employing the "barrier word" method,[18] leverages, the UMLS Specialist Lexicon. Use of other syntax-driven parsing methods, e.g., "minimal commitment parsers,"[19] needs to be explored.

The **Matcher** collapses words and nominal phrases into lexical classes, and, where possible, semantic classes. Using norm[20] — a lexical processing program released with the UMLS Knowledge Sources, the synonym classes in the UMLS Metathesaurus, the **Matcher** accepts potential terms as input, and outputs suggested matches to canonical terms. If known, a Metathesaurus Concept Unique Identifier is output as well; otherwise a server identifier is returned indicating that the server recognizes the input as a potential descriptor, i.e. concept name.

The **Mapper** presents Reference Models for linking by humans and exploitation by humans and agents. On the user side, the **Mapper** is a Reference model Browser where the user can expect to find existing Models initialized for inspection, given the output of the **Parser** and **Matcher** for the patient at hand. One of the Reference Models should be care-centered and encode local notions of best practice. On the content side, the **Mapper** helps link incoming reference models with those already in the **Server**. Because it is being designed to have important formal properties, an enhanced version of SNOMED International[21] may play a key role in the **Mapper**, especially for agents.

The **Server** stores Content created and retrieved using one or more of the three access methods. All content — terminologies, guidelines, textbooks, journal articles, citations, reference models, etc. — is both represented homogeneously and made accessible using the enhancement of the current version of the Metathesaurus created by the **Parser** and **Matcher**. While enterprise applications might take advantage of SQL3[22] access, the principal goal of the **Server** is to try to support semantic access as provided by the other four components operating on structured clinical content.

RESULTS

Fellows in the Pediatric Oncology Clinic at the National Institutes of Health (NIH) Clinical Center (CC) began experimental use of a mobile, wireless pen-based computer23 to access the NIH CC EMR (TDS, AllTel Healthcare Systems, Inc.) and an Oncology Knowledge Server (OKS) in March, 1996. The version deployed at the NIH CC contains PDQ — a database of statements about cancer and cancer therapy, the textbooks DeVita, et al., Principles and Practice of Oncology and Pizzo, et al., Principles and Practice of Pediatric Oncology, about half of CANCERLIT (500,000 citations, including abstracts), and the 1996 UMLS Metathesaurus.

CONCLUSIONS

The proposed Content Machinery represents much less than what most users want, but much more than is available now. As such it represents an important starting point for the deployment of reusable software components in health care. Many health care enterprises have some approximation of an EMR, e.g., results reporting, and deployment of the Content Machinery described here in conjunction with existing systems would represent a way to build on existing investment. This investment could be further leveraged if it included a nationally linked problem dictionary. The biggest current need is for components that exploit the notions of similarity available in the Metathesaurus. Users may browse through similar terms in various Reference Models, but there is no systematic notion of similarity for, say, a software agent. The problem is not with the Parser or Matcher; instead new mechanisms need to be developed to control exploitation of the non-synonymous relationships available in the Server.

Adoption of standard point-of-care text processing practices, especially those that exploit the UMLS Knowledge Sources, would focus attention on the remaining substantive enterprise differences in care and technology. Widespread use of the Router, Parser, Matcher, Mapper, and Server would lead either to their rapid improvement or to their replacement by better components, a better architecture, or better knowledge sources. While the impediments to enterprise-wide deployments of technology in health care are formidable, the potential gains that would result from technologic and content economies of scale should more than justify the required investment.

Acknowledgments

Partially supported by NCI Contracts #N44-CO-33071, #N44-CO-40550, #N44-CO-51031, #N43-CO-51008, NIH Grant 1R41 CA 65426-01, and by NLM Contract N01-LM-3-3515. J.B. Lippincott Company contributed electronic versions of two textbooks processed by IS Grupe, Inc. Earlier versions of the Parser and Matcher were used to help index selected American College of Physicians publications. Portions of this material were presented at the Harvard CME Course "Clinical Computing," C. Safran, MD, Director, and during a visit to the University of Missouri at Columbia, J. Mitchell, PhD, host.. J. Thomas, PhD contributed to the exposition.

References

1. Blois MS, The physician's personal workstation. MD Computing, 1985: 2:22-26.
2. Zelinger J, Rind DM, Caraballo E, et al. Categorization of Free-Text Problem Lists: an Effective Method of Capturing Clinical Data. Gardner RM, Ed., SCAMC, 1995:416-20.
3. Tuttle MS, Nelson SJ. The Role of the UMLS in 'Storing' and 'Sharing' Across Systems. Ball MJ, Silva JS, Douglas JV, et al., Eds, Special Issue: The Health Care Professional Workstation, Int. J. Bio-Med Comp, 34(1-4), Jan, 1994:207-37.
4. Lindberg DAB, Humphreys BL, McCray AT. The Unified Medical Language System. Methods of Information in Medicine, 1993:32:281-291.
5. Cole WG, Sherertz DD, MS Tuttle. Semantic Visualization of Oncology Knowledge Sources. Gardner RM, Ed., SCAMC, 1995:67-71.
6. Wiederhold GCM. Mediators in the Architecture of Future Information Systems. Computer, March, 1992:38-49.
7. Musen MA, Tu SW, Hasan Z. Reuse of the EON Architecture to Automate Protocol-Base Care: From AIDS to Breast Cancer. 1996 AMIA Fall Symposium (to appear).
8. NLM+AHCPR Grant HS/LM08751.
9. Tuttle MS, Sherertz DD, Fagan LM, et al. Bringing Knowledge to the Point of Care. HIMSS '95:101-20.
10. Tuttle MS. Bringing Knowledge to the Point of Care. ANSI Healthcare Informatics Standards Planning Panel, Crystal City, MD, Jan, 1994.
11. Sherertz DD, Tuttle MS, Olson NE, et al. Accessing Oncology Information at the Point of Care: Experience Using Speech, Pen, and 3-D Interfaces with a Knowledge Server. MEDINFO '95: 792-5.
12. Crangle CE, Fagan LM, Sherertz DD, et al. Conversational Access to On-Line Cancer Information: An Adaptable Speech Interface. 1996 AMIA Fall Symposium (to appear).
13. Musen MA, Tu SW, Das AK, Shahar Y. EON: A component-based approach to automation of protocol-directed therapy. JAMIA, in press.
14. Tuttle MS, et al. Intelligent Software Agents that Access PDQ & CANCERLIT. Final Report,1996.
15. Blois MS. Information & Medicine: The Nature of Medical Descriptions. University of California Press, Berkeley, 1984.
16. Ogden CK, Richards IA. The Meaning of Meaning. Harcourt, Brace, Jovanovich, 1923:11.
17. Isselbacher KJ, et al. Harrison's Principles of Internal Medicine. McGraw Hill, NY, 1994.
18. Nelson SJ, Olson NE, Cole WG, et al. Recognizing New Medical Knowledge Computationally. Safran, C, Ed., SCAMC, 1993:409-13.
19. Tom Rindfleisch, Lister Hill Center, National Library of Medicine, personal conversation.
20. McCray AT, Srinivasan S, Browne AC. Lexical Methods for Managing Variation in Biomedical Terminologies. Ozbolt, J, Ed., SCAMC, 1994:325-239.
21. Campbell KE, Cohn SP, Chute CG, et al. Galapagos: Computer-Based Support for Evolution of a Convergent Medical Terminology. 1996 AMIA Fall Symposium (to appear).
22. SQL3 is a proposed extension to SQL2 partially supported by the Illustra Object-Relational database system used to implement the OKS.
23. Acuff RD, Fagan LM, Rindfleisch TC, et al. Integration of Pen-based Computer Technology in Clinical Settings (Demonstration, 1st Prize). Safran C., Ed., SCAMC, 1994:1042.