MEME II Supports the Cooperative Management of Terminology

Octavio N. Suarez-Munist, Mark S. Tuttle, Nels E. Olson, Mark S. Erlbaum, MD,
David D. Sherertz, Stephanie S. Lipow, William G. Cole, Ph.D., Kevin D. Keck, Alex N. Davis
Lexical Technology, Inc., Alameda, California
William T. Hole, MD, Robert J. Irons, Anil Ganju, Stuart J. Nelson, MD,
National Library of Medicine, Bethesda, Maryland

Health care enterprises need enterprise-wide terminologies to compare, reuse and repurpose health care descriptions. But once they are created, these terminologies need to be maintained and enhanced to sustain their utility and that of the descriptions encoded with them. MEME II (Metathesaurus Enhancement and Maintenance Environment, Version II) supports the required activities and enables enterprises to leverage their investment in terminology and descriptions by permitting remote — extra-enterprise — enhancements to terminology to be incorporated locally, and local —intra-enterprise — enhancements to be shared remotely. MEME II represents all changes to terminologies as data, or "actions," that can be interpreted by an "action engine." These actions, or messages, represent semantic "units of work" that can be interpreted by other copies of MEME II. The exchange of update messages increases the likelihood that the comparability of terminology-based health care descriptions can be sustained.

THE PROBLEM

Caregivers, administrators, researchers and information specialists all need to communicate, share and reuse biomedical descriptions to provide care, manage resources, assess quality, obtain third-party reimbursements and conduct research. Any terminologies used to help create these descriptions must be maintained and synchronized to make these and other kinds of communications possible. This maintenance problem is complicated by the needs of health care enterprises for both internal and external communications. The former include managed care, outcomes, case management, ambulatory care, electronic data interchange, and cost management. The latter include clinical trials, insurance payments, patient transfer and referrals, government reporting, medical research, and community health information networks. Over time, these enterprises and their subdivisions have developed both local terminologies and non-concurrent versions of remote terminologies to help serve these needs. Any lack of concurrency impedes communication by allowing the creation of non-comparable descriptions.

The National Library of Medicine (NLM) Unified Language System[1] (UMLS) Metathesaurus solves part of this problem by merging more than 30 national and international terminologies into a single semantic representation.[2] Because the Metathesaurus sustains the meaning of its Concept Unique Identifiers (CUIs) over time, descriptions created using them remain comparable, provided that the relevant information in the change files accompanying each Metathesaurus release is acted upon appropriately.[3]

Local needs, however, may require terminologies that are not in the Metathesaurus, and any attempt to attach them to a local copy of the Metathesaurus inevitably creates a local maintenance problem and an update penalty (see Campbell, 1996) the next time the Metathesaurus is updated.[4] Some of these local updates may be the very ones the Metathesaurus needs to remain current, but at present there is no reliable way to harvest them. Of course, if every terminology represented and published change in the same way, both update tasks would require less manual effort,[5] but real economies of scale could be achieved if some of the intellectual work required to create updates could be shared and reused algorithmically.[6] Currently, most reuses of terminology updates require manual "reverse-engineering" of the changes between previous and new versions.[7,8]

The last five versions of the Metathesaurus were produced using MEME I, and experience with that version motivated the design of MEME II.[9] MEME I is a *concept engine*, the first to support a direct manipulation interface. But these manipulations changed the state of the MEME I database without creating externally interpretable "actions."

A second effort to create an environment that supports independent enhancement of terminology[10] is proceeding in parallel and an early evaluation of the core methods appears promising.[11,12] The latter citation includes references to other foundational work in the U.S. and Europe. A third effort is focused on guidelines and the models and terminologies used to create and maintain them.[13]

0195—4210/96/$5.00

In summary, there is currently no discipline nor system by which updates to multiple biomedical terminologies from different sites can be shared and reused, but certain conclusions can be drawn about what would be required to remedy the situation. Also, despite the diversity in the terminologies employed and the purposes for which they are used, the problems faced by those who manage these terminologies are similar across all health care enterprises.

Specificity, *coverage*, *richness*, and *connectedness* are qualities that help make a terminology useful. A terminology has "specificity" if its terms are sufficiently precise to differentiate the phenomena being named. A terminology offers "coverage" if it can name all the phenomena being observed. A terminology has "richness" if it carries sufficient information about the phenomena so that users can unambiguously assign a term to any concept. A terminology may be said to offer "connectedness" if it represents for a particular user all useful relationships among the phenomena being modeled. Not surprisingly, different enterprises will place different relative values on these characteristics. Still, the fact that enterprises with very different (though interrelated) purposes happen to share a small number of fundamental requirements for their terminologies (regardless of the relative importance they assign to each of these requirements) opens the way for a scheme of maintenance with potentially broad applicability.

A SOLUTION

MEME II is being developed to improve Metathesaurus maintenance and enhancement productivity. With the change in the UMLS release date to early in each calendar year, MEME II is to be deployed for use during the preparation of the 1998 release of the Metathesaurus. The principal benefits will accrue from MEME II's ability to support the necessary maintenance and enhancement activities concurrently, and from increasing the productivity of some of those activities, e.g., editing, through the introduction of an "undo" function.

The Architecture

As shown in **Figure 1**, MEME II implements the desired functionality with a four-layered architecture. The top layer consists of applications that support the activities required to maintain and enhance the Metathesaurus. These activities are:

Creating New Terms. New terms need to be identified to extend the coverage of an existing terminology and thereby assure its continued currency. Health care domain experts are needed to develop appropriate terms that are new names for existing concepts or names of new phenomena such as newly discovered (or newly differentiated) diseases or new procedures, findings, drugs, or devices.

Enriching Terms. Attributes, e.g., definitions and semantic types, need to be added to clarify the role of terms in existing vocabularies, particularly when a term needs to be related to those in other vocabularies. This may require significant additional information especially if it needs to be made more amenable to machine processing.

Connecting Terms. To be useful in larger contexts, terms need to be related to one another more systematically and existing relationships need semantic labels, e.g., "IS_A". To accomplish this, the concepts represented by these terms must be placed within a context and the relationships among associated concepts must be specified.

Integrating External Knowledge. In order to extend a terminology's coverage and to increase its connectivity, remote terminologies need to be routinely integrated. This requires the use of sophisticated lexical and semantic techniques as well as human review to identify connections between the enterprise's local terminology and external sources.

Publishing. Enterprises that publish all or part of their terminologies are likely to be in a position to enhance those terminologies systematically. That is, their terms are likely to be enriched or, occasionally, corrected, by remote feedback. But this will only work if enterprises, maintain the currency of their published terminologies, preferably automatically. Publishing also allows the users of other terminologies to incorporate the publishing enterprise's concepts and terms into their own terminologies thereby supporting diffusion.

Managing Activities. The best allocation of human and computing resources for all of the preceding activities will necessarily be determined by an enterprise's needs and the characteristics of its terminology. Increasingly, these activities will need to be performed concurrently to maximize human productivity and to meet an enterprise's internal and external requirements.

The rest of the MEME II architecture is (from top to bottom) the molecular layer, implementing commands such as "merge," the atomic layer,

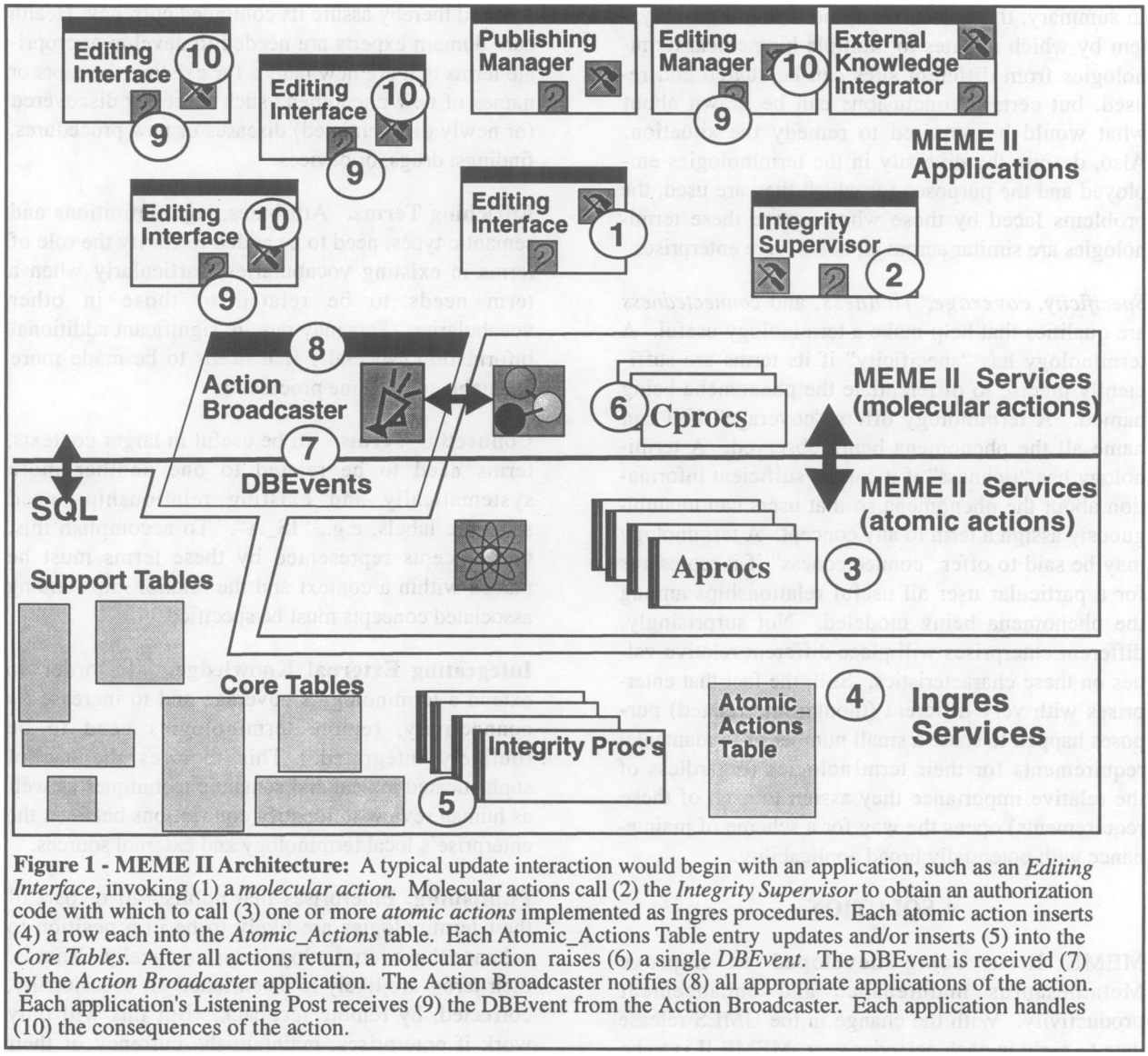

Figure 1 - MEME II Architecture: A typical update interaction would begin with an application, such as an *Editing Interface*, invoking (1) a *molecular action.* Molecular actions call (2) the *Integrity Supervisor* to obtain an authorization code with which to call (3) one or more *atomic actions* implemented as Ingres procedures. Each atomic action inserts (4) a row each into the *Atomic_Actions* table. Each Atomic_Actions Table entry updates and/or inserts (5) into the *Core Tables.* After all actions return, a molecular action raises (6) a single *DBEvent.* The DBEvent is received (7) by the *Action Broadcaster* application. The Action Broadcaster notifies (8) all appropriate applications of the action. Each application's Listening Post receives (9) the DBEvent from the Action Broadcaster. Each application handles (10) the consequences of the action.

implementing commands such as "insert," and the database (Ingres[14]) layer on which all other layers are implemented.

MEME II is designed to be extensible, scalable, and portable. By *extensible,* we mean that the principles of its design are applicable to different terminologies. By *scalable,* we mean that its architecture is able to accommodate significant increases in the size of the terminology. By *portable*, we mean that both clients and servers can be ported without adversely affecting one anothers' activities. The maintenance activities discussed above are performed through an extensible set of actions that are initiated by client applications. These actions are passed from the client to the server, which performs the necessary operations. Only through observance of these procedures can information be changed. This safeguard allows for quick development of applications while maintaining the integrity of the data. In addition, the architecture illustrated in **Figure 1** offers six important features: (1) a multi-user environment, (2) integrity enforcement, (3) support for concurrent activities, (4) automatic external synchronization, (5) an action-based record of changes, and (6) a distributed client/server model.

A Scenario

Figure 2 shows how enterprises maintaining their own terminologies can benefit by publishing their enhancements and then integrating them with the enhancements of others. Changes in published information can be integrated automatically into all participating vocabularies. In this scenario, three

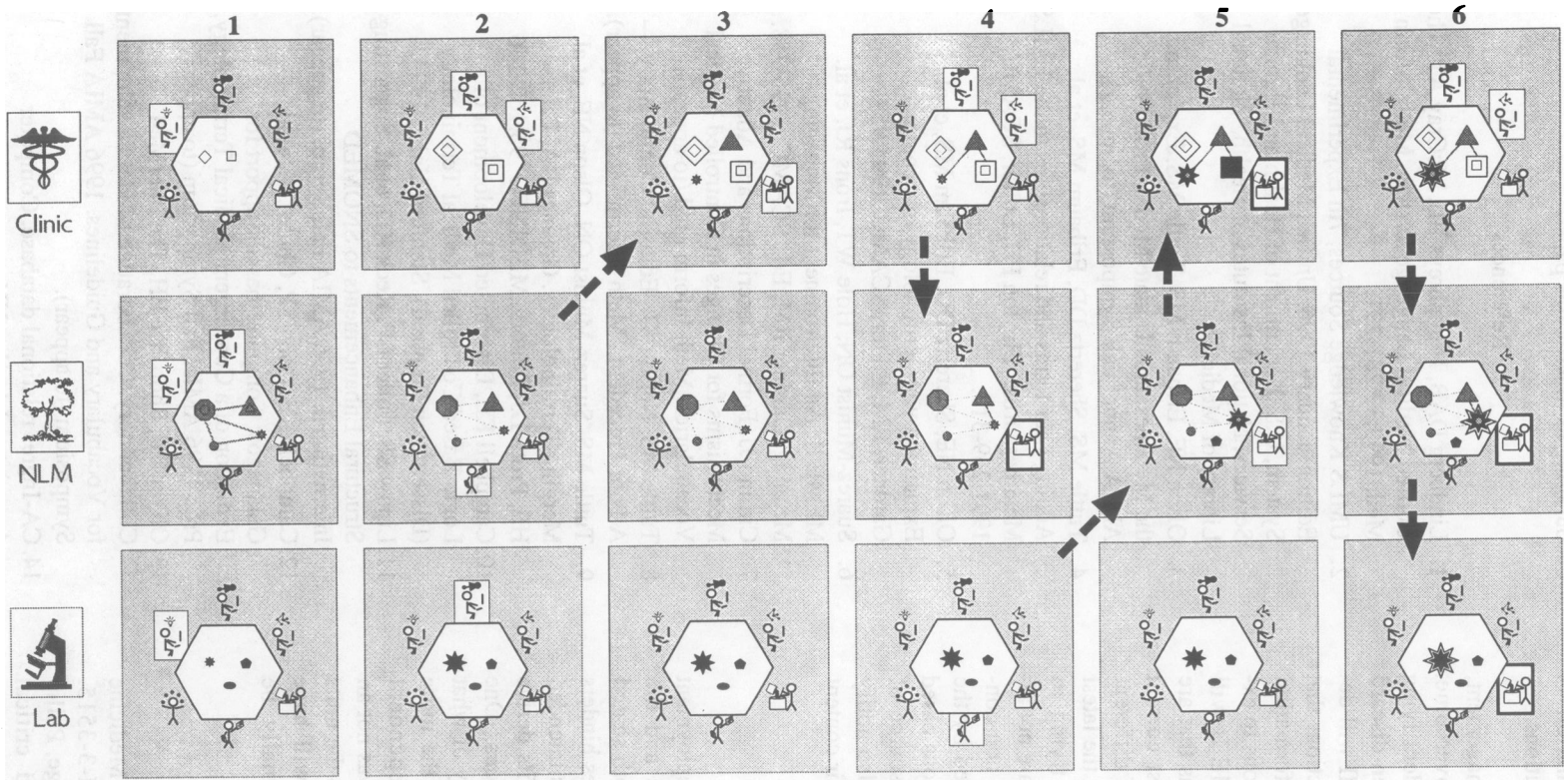

Figure 2 - Scenario for Cooperating MEME II Sites: At a given time at the beginning of a work day, a healthcare **clinic** (top row), the **NLM** (middle row), and a research **laboratory** (bottom row) are working independently at separate MEME II sites.

Column One: the **clinic** and the **lab** are *creating* their own new concept terms while the **NLM** is *enriching* and *connecting* other concept terms.

Column Two: a little later, the **clinic** and the **lab** are each *enriching* the terms they have just created, while the **NLM** is *publishing* the concept terms and relevant connections that it was working on earlier.

Column Three: MEME II notifies the **clinic** of the **NLM**'s advance. The **clinic** then engages in *integration* of this external knowledge.

Column Four: somewhat later in the day, the **clinic** *enriches* and *connects* concept terms; the **NLM** is notified *automatically* by MEME II - non-conflicting information is integrated without any human intervention. Meanwhile, the **lab** is publishing concept terms and connections.

Column Five: as the day draws to a close, the **NLM** is *integrating* the external knowledge it has just received from the **lab** through MEME II, while the **clinic** benefits from an automatic integration of this external knowledge, via the **NLM** through MEME II.

Column Six: at the end of the day, we find the **clinic** *enriching* and *connecting* the concept terms just received from the **NLM** through MEME II; MEME II automatically integrates this new information into the **NLM** and, through the **NLM**, into the **lab**.

different kinds of health care enterprises maintain their own terminologies but benefit from publishing and incorporating one anothers' work. Potential benefits are realized when new information is integrated throughout the cooperating institutions.

To realize the benefits of cooperative management, three operational problems must be overcome. MEME II will detect *redundancy* by identifying identical actions on the same data and then discard the one with the later timestamp. MEME II will detect *disjointedness* by identifying remote actions that do not generate connections to existing information and will deal with it by reporting these actions to experts for a review of connectivity. MEME II will detect *conflict* by identifying remote actions that are "stale," that is, inconsistent with the most current information, and resolve it in a number of different ways. It will implement this by identifying the latest point previous to the inconsistency and trying to identify any actions beyond that point that are not affected by the inconsistency (disjointed or non-conflicting). For those actions that are affected by the inconsistency, it will discard one set of actions based on several user-selectable criteria (time, source of actions, number of actions, etc.). It will also notify affected parties so that they may review their content in light of the conflict detected.

CONCLUSION

The chief benefits of the initial MEME II deployment will be support for concurrent activities and the existence of a log of actions that can be both shared and "undone". The lack of these capabilities hinders MEME I's productivity. While MEME II is intended to serve many purposes, most of its benefits derive from clean implementations of these functions. One emerging benefit is the broader applicability of what started out as a client-server solution to the more general problem of leveraging independent terminology updates. Finally, although it was not an initial requirement, the database actions — the units of work — that will fulfill these needs will have semantic interpretations that can eventually be displayed in human-readable form.

Acknowledgments

Partially supported by National Library of Medicine Contracts N01-LM-03515 and N01-LM-3-3515. Betsy Humphreys, Chris Wenk, Alan Savage, Philip Passarelli, and Tammy Powell provided critical comments on MEME I and insights regarding the requirements for MEME II. We thank Donald Lindberg, MD for interesting discussions regarding technological support for collaborative work and for encouraging us to write this paper.

References

1. Lindberg DAB, Humphreys BL, McCray AT. The Unified Medical Language System. Meth Inform Med, 1993:32:281-291.
2. UMLS Knowledge Sources, 7th Experimental Edition, January 1996, Unified Medical Language System, U.S. Department of Health and Human Services, National Institutes of Health, National Library of Medicine.
3. Olson NE, Erlbaum MS, Tuttle MS. Exploiting the Metathesaurus Update Model. Proc 1996 AMIA Annual Fall Symposium (to appear).
4. Tuttle MS, Sherertz DD, Erlbaum MS, et al. Adding your terms and relationships to the UMLS Metathesaurus. Clayton PD, Editor, SCAMC, 1991:269-317.
5. Olson NE, Sherertz DD, Erlbaum MS, et al. Explaining your terminology to a computer. Gardner, RM, Editor, SCAMC, 1995:957.
6. Suarez-Munist ON, Hole WT, Irons RJ, et al. MEME II: An Environment for Managing Meaning. Gardner RM, Ed., SCAMC, 1995:958.
7. Cimino JJ. Formal Descriptions and Adaptive Mechanisms for Changes in Controlled Medical Vocabularies. Meth Inform Med (to appear).
8. Tuttle MS, Nelson SJ. Editorial Commentary — A Poor Precedent. Meth Inform Med (to appear).
9. Tuttle MS, Suarez-Munist ON, Olson NE, et al. Merging Terminologies. Greenes RA, Peterson HE, Protti DJ, Editors, MEDINFO '95:162-166.
10. Campbell KE, Distributed Development of a Logic-Based Controlled Medical Terminology (Dissertation Proposal), Stanford Univ., 1994.
11. Lipow SS, Fuller LF, Keck KD, et al. Suggesting Structural Enhancements to SNOMED International. 1996 AMIA Fall Symp. (to appear).
12. Campbell KE, Cohn SP, Chute CG, et al. Galapagos: Computer-based Support for Evolution of a Convergent Medical Terminology. Proc 1996 AMIA Fall Symposium (to appear).
13. Oliver DE, Shortliffe EH, The InterMed Collaboratory. Collaborative Model Development for Vocabulary and Guidelines. 1996 AMIA Fall Symposium (to appear).
14. CA-Ingres relational database, Computer Associates, Islandia, NY.