Shankar RD, Tu SW, Martins SB, Fagan LM, Goldstein MK, Musen MA.
Integration of textual guideline documents with formal guideline knowledge bases.
Proc AMIA Symp. 2001;:617-21.

Integration of Textual Guideline Documents with Formal Guideline Knowledge Bases

Ravi D. Shankar, MS[1], Samson W. Tu, MS[1], Susana B. Martins, MD[2]
Lawrence M. Fagan, MD, PhD[1], Mary K. Goldstein, MD [2,1] Mark A. Musen, MD, PhD[1]

[1]Stanford Medical Informatics, Stanford University School of Medicine, Stanford, California, USA
[2]VA Palo Alto Health Care System, Palo Alto, California, USA

Abstract

Numerous approaches have been proposed to integrate the text of guideline documents with guideline-based care systems. Current approaches range from serving marked up guideline text documents to generating advisories using complex guideline knowledge bases. These approaches have integration problems mainly because they tend to rigidly link the knowledge base with text. We are developing a bridge approach that uses an information retrieval technology. The new approach facilitates a versatile decision-support system by using flexible links between the formal structures of the knowledge base and the natural language style of the guideline text.

Guideline Documents in Guideline-based Decision-Support Systems

Clinical guidelines encapsulate the best medical practices, and are becoming ever more important in managing the explosion of medical knowledge and in promoting evidence-based medicine. A clinical guideline contains the intentions of the guideline, medical background, patient eligibility criteria, procedural statements such as clinical algorithms and drug recommendations, evidence for the advisories, treatment cost-benefit analysis, and references. The guidelines are generally published in textual format via print and electronic media. Recent reports on medical errors are fueling the adoption of clinical guidelines as tools to promote patient safety.

In recent years, the healthcare community has shown an increased interest in automated support for guideline-based care in clinical settings. As a result, many types of guideline decision-support systems have been proposed. One type of systems uses markup methodologies to serve guideline text excerpts relevant to patient context. These document-centric systems use languages like HTML and XML, and allow clinicians to browse through documents as easily as they browse the World Wide Web. Markup methods are a significant improvement over paper-based guidelines or linear on-line text documents. Marked-up documents allow easy browsing and more focused linkage between concepts and text. New ways of linking text to information systems allow alternative ways of presenting information at points of care.

On the other end of the decision-support spectrum are systems that use complex knowledge models to facilitate patient-specific therapeutic advisories. We called these systems knowledge-base-centric. In order to build a knowledge-base-centric system, a domain expert extracts information from the guideline text, interprets it, and then encodes it in some computable formalism in a knowledge base. A computer system uses patient data to interpret the knowledge base, generates a therapeutic advisory, and if required, justifies the advisory. Well-structured advisories can be justified partly by presenting relevant free-text extracted from the guideline document. In this approach, guideline documents play two major roles: 1) they are the source of the knowledge base and 2) they help in generating explanations.

Both of these approaches encounter difficulties in integrating guideline text with decision-support systems. Systems in one approach attempt to overcome their deficiencies by adopting methodologies from the other approach. We believe that the natural language style of the marked up documents and the rigid computer formalizations of the knowledge base are orthogonal to one another, and that simply forging a middle ground at the implementation level will not solve integration problems. Thus, a new approach is crucial to integrating the systems optimally. This paper presents an analysis of the document-centric and knowledge-base-centric approaches to highlight the integration issues. The paper then describes an alternative *bridge approach*, which builds on the philosophies of the document-centric approach and the knowledge-base-centric approach. The new approach uses an information markup, search and retrieval technology to bridge the orthogonal divide between the knowledge base and the guideline text.

Shankar RD, Tu SW, Martins SB, Fagan LM, Goldstein MK, Musen MA.
Integration of textual guideline documents with formal guideline knowledge bases.
Proc AMIA Symp. 2001;:617-21.

The Document-Centric Approach

The literature contains many descriptions of document-centric systems. They differ from one another by the amount of formal structure that they impose on a document.

An example of this approach is *ActiveGuidelines, which was* developed at Epic Research Institute [1]. ActiveGuidelines imposes minimal structure on marked-up guideline documents. The system is integrated with a computerized patient record. When a patient context triggers eligibility criteria in a guideline, the guideline is made available to a clinician. Its recommended orders are marked up with tags that contain information needed to process the orders. The clinician can browse the guideline text and click on orders that she decides to accept, and, at the end of a session, she can have the orders automatically routed to the order-entry facility of the computer-based patient record system.

The *GEM* system [2], developed at Yale University, takes a very different approach. GEM's goal is to create a comprehensive mark-up system for annotating a guideline document. GEM's intended use is for health service research and clinical decision support. One problem that GEM had to overcome was that the organization of guideline text documents is generally geared towards a human audience, and may not be suitable for building decision-support systems. GEM overcomes this problem by defining a guideline ontology that allows the document to be marked up for different views. With this system, heterogeneous information at different levels of granularity is organized efficiently and can be retrieved efficiently. The guideline ontology includes a rich collection of tags describing the purpose, intended audience, development methods, and evidence base of guidelines. The decision-support elements in GEM consist of constructs for decision tables and for modeling GLIF2 [3] clinical algorithms.

There are several advantages to the document-centric approach of modeling guidelines. For example, marking up the document, and hence building a knowledge base, can be performed to any level of granularity. As a result, implementing the decision-support system can happen in stages, allowing for rapid system prototyping.

Moreover, guideline documents contain information that caters to different users, including nurses, physicians, medical technicians, hospital administrators and patients. The document-centric approach can preserve these relationships by using appropriate ontology, and it can offer multiple views of each guideline document.

A major concern when using the document-centric approach to build decision-support systems is that the mark-up methodology does not lend itself to representing complex decision logic. For example, ActiveGuidelines only formalizes the applicability criteria of guidelines it supports: it requires clinicians to read text and apply a guideline manually. GEM has decision tables that can formalize rule-like recommendations, but decision tables are insufficient for modeling complex procedures. The GLIF2 guideline model is not sufficiently detailed to be computable. The GEM tries to overcome this deficiency by exploring the possibility of integrating GEM markup with computable GLIF3 currently under development [4].

One approach would be to impose an ontology on the original document for the purposes of decision-support. However, this approach can be cumbersome. For example, there are times when the knowledge elements that are needed for decision support do not cleanly map to the guideline text. This difficulty is especially true for rules that are created through expert interpretations of the text. For example, the JNC 6 hypertension guideline document has several tables enumerating considerations for prescribing antihypertensive medications. The knowledge in these tables must undergo major restructuring if a computer system is to apply the relevant content to specific patient situations. Also, the information needed to make a patient-specific decision may be in disparate parts of the document. Moreover, the guideline knowledge required to build a decision-support system can come from more than one document. Combining information from more than one document can be difficult in the document-centric approach. A formally encoded guideline knowledge base can be used to encode information from multiple marked up documents.

The Knowledge-Base-Centric Approach

In contrast to the document-centric approach, the knowledge-base-centric approach follows a formal representation methodology to encode the guideline text in a knowledge base. There are several proposed formalizations of this approach, including GLIF3 and EON [5]. ATHENA DSS is a hypertension care system that we had developed based on the EON architecture as part of the ATHENA project at the VA Palo Alto Health Care System [6]. ATHENA DSS is a representative decision-support system using this approach. We used the EON guideline model to build ATHENA's guideline knowledge base. The model has knowledge constructs to define medical concepts and relationships, and to encapsulate decision logic in guideline clinical algorithms. With the help of domain experts, we manually extracted decision-support information from the JNC VI hypertension guideline document and the VA hypertension guideline document, interpreted the text when needed, and built the

Shankar RD, Tu SW, Martins SB, Fagan LM, Goldstein MK, Musen MA.
Integration of textual guideline documents with formal guideline knowledge bases.
Proc AMIA Symp. 2001;:617-21.

hypertension guideline knowledge base using the constructs in the EON guideline model. ATHENA DSS uses patient data to interpret the hypertension guideline, and provides patient-specific recommendations. It also generates coherent explanations justifying its recommendations. In a survey we conducted among primary care health professionals participating in a guideline skills workshop [7], we determined the necessity of showing guideline text as part of the explanation. Therefore, we included relevant excerpts of the guideline text as part of all explanations.

The EON system uses Protégé-2000 [8], a frame-based knowledge-engineering environment, to model all declarative domain knowledge. The system links textual supporting materials to knowledge elements, which are used to generate patient specific recommendations. For example, the system models indications and contraindications of antihypertensive agents as relations between medical conditions and medications. It links references to supporting material such as guideline text and clinical evidence from the literature to frames representing such relationships (Figure 1). It also links a knowledge element to the guideline text by using html tags. To produce this system, we identified parts of the text that describe the knowledge element, and labeled them. Next, we added the html references to the knowledge element's supporting-material structure. EON's WOZ explanation system [7] generates explanations for the decision-support system's recommendations system by using a knowledge base of explanation strategies and links to supporting materials.

The strategy of using the guideline knowledge model to drive the integration of text with the decision-support system has a number of advantages. For example, it serves the need to provide explanations for its recommendations, and it links elements in the knowledge base to the document text. Given the nature of the guideline model, ATHENA DSS can easily integrate information from different documents. We needed only to annotate the decision-support portions of the guideline document that we modeled — not necessarily the whole guideline document.

Adding textual references to the knowledge base was a cumbersome process. Most of the difficulty stemmed from the fact that the organization of the guideline documents did not always correspond to the structure of the knowledge base. For example, the JNC VI guideline report is organized by conditions, and lists drugs for treating each one. The knowledge base models this information in the opposite way: it is organized by drug, and lists the conditions for which each one is useful. Mapping such elements in the knowledge base to the guideline text is not a straightforward process.

Some rules in the knowledge base are created

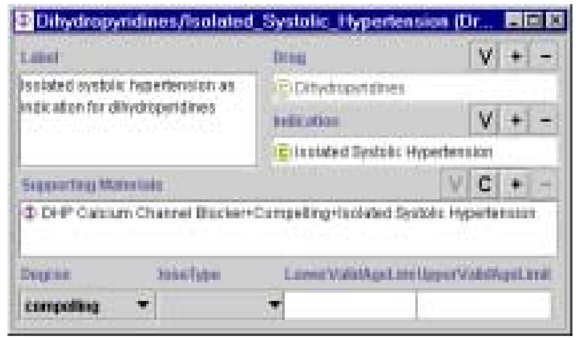

Figure 1. A knowledge element in the EON guideline model. Isolated systolic hypertension is modeled as a compelling indication for dihydropyridines. Supporting materials for the indication are linked to the knowledge element.

through expert interpretations of guideline text. These rules do not map cleanly to the original guideline text. One way to overcome this difficulty is to document the interpretations and map the rules to them. This approach would have introduced problems in ensuring validity (as the document did not originate from a standards body) and maintenance of the overall knowledge base.

The guideline document can change when new medical evidence is introduced. The knowledge base can change because of guideline modeling changes. In both cases, the links from the knowledge base and the markup tags in the guideline document need to be updated and verified. In this approach, there is no easy way to accomplish this task.

The Bridge Approach – Integrating Structured Knowledge Bases with Marked-up Text

When integrating guideline text with decision-support systems, the following characteristics are desirable:

- Incorporating the flexibility of document markups to organize documents to aid in knowledge modeling and in generating explanations.
- Using the preciseness of knowledge base elements to define guideline decision logic.
- Bridging the divide between the document layout and knowledge base structure to facilitate easy mapping between the two.
- Reducing modeling requirement and increasing the robustness of knowledge bases by presenting marked-up text when either patient data or structured guideline knowledge element is missing.
- Maintaining the links between the knowledge base and the guideline document when the

Shankar RD, Tu SW, Martins SB, Fagan LM, Goldstein MK, Musen MA.
Integration of textual guideline documents with formal guideline knowledge bases.
Proc AMIA Symp. 2001;:617-21.

guideline model changes or when new medical evidence is introduced.

To facilitate decision-support using guideline text, the document-centric approach such as GEM tries to forge the marked up documents with a guideline model. To integrate guideline text with the knowledge base, the knowledge-base-centric approach such as ATHENA DSS tries to explicitly link the knowledge base elements with the guideline text. The mapping and maintenance problems with these approaches mostly stem from the fact that the approaches attempt to create hard and concrete links between the knowledge base and guideline text. We are building a bridge approach that improves on the integration methodology of ATHENA DSS. The approach replaces, where appropriate, absolute text references in the supporting material of the knowledge elements with abstract descriptions of the knowledge elements. The descriptions are resolved against a marked up guideline text using an information retrieval system to fetch excerpts of the text that are relevant to the knowledge element. Thus, in the bridge approach, the link between knowledge base elements and the relevant guideline text are soft and abstract.

To implement the soft link methodology, we are currently investigating ACQUIRE (now called QueryEdit), an information retrieval technology that was built in our laboratory [9]. ACQUIRE aids in the mark-up of medical text and retrieval of complex concepts from online medical text. ACQUIRE uses a *concept model* and a *query model*. The concept model defines a set of concepts that is used to markup the text. The query model encapsulates a set of specific questions that can be used to search the marked text. The questions are formulated using the terms in the concept model. Domain experts can use ACQUIRE to identify sections of text that contain elements from the concept model, and formulate relevant questions that the end users may ask. End users search the text using these questions. The search engine matches the set of concepts in each question to relevant points in the document.

ACQUIRE has been used successfully to mark up and search sections of a clinical therapeutics textbook, Antimicrobial Therapy and Vaccines, that deals with infectious disease treatment. Because the textbook refers to therapeutic recommendations for specific patient conditions, its content-type is similar to a clinical practice guideline document. ACQUIRE has handled descriptions of therapeutic elements [10], but may need to be extended to markup and search of clinical practice guideline documents.

The bridge approach adapts the ACQUIRE system to mark up guideline documents with relevant medical domain and guideline concepts defined in the guideline model. The approach introduces the notion of *clinical queries*. These queries are questions that a clinician might pose in order to gain insight into guideline recommendations at each point in the clinical algorithm. The bridge approach associates appropriate clinical queries to elements in the knowledge base. ACQUIRE can search the guideline documents to satisfy the queries, and output relevant excerpts of the document. Thus the clinical queries form the bridge between the knowledge base and the guideline text.

Discussion

The novelties of the bridge approach are 1) it associates clinical queries with the elements in a knowledge base, and 2) it uses information retrieval technology to satisfy the queries by retrieving relevant excerpts of the guideline document. By using clinical queries, the system avoids absolute references to the guideline text. Thus, the link between the knowledge base and the guideline text is more dynamic, which makes it easier to maintain. For example, if the marked up guideline document changes, the bridge approach can serve excerpts from the new document using the same queries in the knowledge base.

The bridge approach also facilitates 1) modeling applicable portions of a guideline to the necessary level of detail, and then 2) supplementing the formal representation with relevant excerpts from the textual guideline document. In some situations, a domain expert might wish to use this system to create a guideline model that is computer-interpretable at only a high level, relying on the full text of the guideline to provide the details that are difficult to specify formally. Furthermore, encoding the entirety of a guideline or protocol at its most detailed level of granularity can be a daunting task — one that is not always necessary for a specific application. For example, when modeling hypertension guidelines for implementing a system like ATHENA DSS, there may be no perceived need to model pregnancy issues explicitly, as the patient population at the Palo Alto VA is predominantly comprised of older males. Such a system would not be useful to a clinician managing hypertension in pregnant women. However, by using the queries at that knowledge element, the bridge approach can supplement the partial recommendations generated by the guideline execution engine with portions of the textual guideline concerning hypertension and pregnancy.

Note that ATHENA DSS generates partial recommendations when patient data are missing from medical records. Ideally, these recommendations should be augmented with text from the guideline document, which can indicate the actions the system

Shankar RD, Tu SW, Martins SB, Fagan LM, Goldstein MK, Musen MA.
Integration of textual guideline documents with formal guideline knowledge bases.
Proc AMIA Symp. 2001;:617-21.

might have recommended if more data were available. For example, if serum creatinine values are absent, the existing decision-support system assumes that the value is normal, generates advisories, and also will urge the clinician to check the value. The bridge approach, instead of making assumptions about missing data, can present excerpts of the guideline text describing this situation.

This information retrieval technology facilitates aggregating information from multiple documents or information from multiple points in the same document. At the same time the marked-up documents are available for browsing and other uses.

An important question that we face with this approach is how a domain expert can ever anticipate all the relevant queries. We believe that the structure of the knowledge base will facilitate methodical creation of the queries, and make the process more manageable. A larger question we need to answer is how the bridge approach fits in with the existing explanation system WOZ.

The document-centric approach is strong in organizing the documents for multiple views. The knowledge-base-centric approach is strong in aggregating and modeling information in guideline documents for decision-support. Using an information retrieval technology, the bridge approach builds on the formalization strengths of knowledge base and the expressiveness of guideline text, and brings the two closer.

Acknowledgments

This work has been supported, in part, by grant LM05708 from the National Library of Medicine, and VA grants HSR&D CPG-97-006, CPI 99-275, and RCD-96-301. The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs. We thank Valerie Natale for her valuable editorial comments.

References

[1] Tang, PC and Young, CY. ActiveGuidelines: Integrating web-based guidelines with computer-based patient records. *Proceedings of AMIA Annual Symposium*, 2000, 843-848.

[2] Schiffmann RN, Liaw Y, Brandt CA, Corb GJ. Computer-based guideline implementation systems: A systematic review of functionality and effectiveness. *Journal of the American Medical Informatics Association*, 1999: 6(2), 104–114.

[3] Ohno-Machado L, Gennari JH, Murphy S, Jain NL, Tu SW, Oliver DE, et al. The GuideLine Interchange Format: A model for representing guidelines. *Journal of the American Medical Informatics Association*, 1999: 6(2), 104–114.

[4] Peleg M, Boxwala AA, Ogunyemi O, Zeng Q, Tu S, Lacson R, Bernstam E, Ash N, Mork P, Ohno-Machado L, Shortliffe EH and Greenes RA. GLIF3: The evolution of a guideline representation format. *Proc AMIA Symposium*, 2000; supplement, 645-649.

[5] Musen MA, Tu SW, Das AK, Shahar Y. EON: A component-based approach to automation of protocol-directed therapy. *Journal of the American Medical Informatics Association*, 1996: 3(6), 367–388.

[6] Goldstein M, Coleman R, Advani A, Munroe P, Gholami P, Lavori P, and Hoffman BB. Implementing practice guidelines for hypertension: effect of computer- enerated patient-specific recommendations for drug therapy. *Medical Decision Making*, 1999. 19: 529.

[7] Shankar, R.D., Martins, S.B., Tu, S.W., Goldstein, M.K., Musen, M.A. (2001). *Building an Explanation Function for a Hypertension Decision-Support System*. Medinfo 2001, London, UK, submitted.

[8] Noy, N. F., Grosso, W. and Musen, M. A. (2000). *Knowledge-Acquisition Interfaces for Domain Experts: an Empirical Evaluation of Protege-2000.* Proceedings of the Twelfth International Conference on Software Engineering and Knowledge Engineering (SEKE2000), Chicago, IL, 177–186.

[9] Dugan, J., Berrios, D. C., Liu, X., Kim, D. K., Kaizer, H. and Fagan, L. M. (1999). *Automation and Integration of Components for Generalized Semantic Markup of Electronic Medical Texts.* Proceedings of the AMIA Annual Symposium, Washington, DC, 736–740.

[10] Cucina RJ, Shah MK, Berrios DC, Fagan LM. Formulation Of A Template Query Set For Querying Clinical Information Retrieval Systems. *J Investig Med*, 2001; 49(1):45A (abstract)