Knowledge Representation for Platform-Independent Structured Reporting

Charles E. Kahn, Jr., M.D., Phiem N. Huynh, B.S.
Section of Information and Decision Sciences, Department of Radiology,
Medical College of Wisconsin, Milwaukee, Wisconsin

Structured reporting systems allow health care providers to record observations using predetermined data elements and formats. We present a generalized language, based on the Standard Generalized Markup Language (SGML), for platform-independent structured reporting. DRML (Data-entry and Report Markup Language) specifies hierarchically organized concepts to be included in data-entry forms and reports. DRML documents serve as the knowledge base for SPIDER, a reporting system that uses the World Wide Web as its data-entry medium. SPIDER generates platform-independent documents that incorporate familiar data-entry objects such as text windows, checkboxes, and radio buttons. From the data entered on these forms, SPIDER uses its knowledge base to generate outline-format textual reports, and creates datasets for analysis of aggregate results. DRML allows knowledge engineers to design a wide variety of clinical reports and survey instruments.

INTRODUCTION

The realization of the computer-based patient record (CPR) has become an important goal for providers of health care [1]. The success of CPR systems depends in large part on the mechanisms for capturing clinical information. Physicians typically record their clinical observations as handwritten notes. Although quick and convenient, this form of documentation suffers from illegibility, poor formatting, missing information, and limited availability [2]. Transcription of dictated reports may introduce typographical errors, and adds delay between the time the physician generates the report and the time it becomes available to others. Although natural-language reports may allow more flexible expression, they require sophisticated processing techniques or tedious manual review to summarize the entered information.

Structured data entry is the process of recording observations in a predefined format using a standardized set of concepts. SPIDER (Structured Platform-Independent Data Entry and Reporting) is a developmental system that uses the World Wide Web for platform-independent structured entry of radiology reports [3]. SPIDER incorporates a knowledge base of hierarchically organized reporting concepts, two specialized software modules, and a WWW server. SPIDER generates structured data-entry forms that incorporate familiar, graphical objects such as text windows, checkboxes, and radio buttons. The system stores the data entered on these forms, creates datasets for analysis of aggregate results, and uses its knowledge base to generate outline-format textual reports.

We present a simple, generalized language for describing the concepts to be included in structured data entry forms and reports. This language, the Data-entry and Report Markup Language (DRML), is based on the Standard Generalized Markup Language (SGML). Applications of DRML include radiology results reporting and a health status questionnaire.

LANGUAGE SPECIFICATION

Standard Generalized Markup Language

The "grammar" of DRML is defined using the Standard Generalized Markup Language (SGML). SGML is an international standard for describing the format and content of documents: it serves as a meta-language for describing document markup languages, and forms the basis of several widely accepted languages for interchange of documents [4-6]. HTML — the Hypertext Markup Language that undergirds the World Wide Web — is an application of SGML [7]. Just as HTML allows users to create hypertext documents without worrying about how to display the documents, DRML seeks to simplify the task of constructing data-entry applications.

An SGML document consists of a declaration, a Document Type Definition (DTD), and the document instance. The declaration defines some of the basic properties of SGML for a particular installation. The DTD, which is written in SGML, defines the rules for marking up a class of documents. It thus defines the structure of a document, in this case, the specification of a data-entry system. The DTD specifies the names of elements that are allowed, how often an element

0195—4210/96/$5.00 © 1996 AMIA, Inc.

may appear, and the order in which elements must appear. The DTD describes the contents of elements, that is, the names of other elements that are allowed to appear inside them.

Document elements are delimited by starting and ending "tags": keywords enclosed in angle brackets. The ending tag includes a forward-slash (/) character. For example, a document's title is demarcated by the `<TITLE>` starting tag and `</TITLE>` ending tag. Attributes that further describe the document element can appear within the angle brackets. Thus, for a given application — which can be viewed as a class of documents — the DTD defines the allowable tags and attributes. DRML is defined as an SGML Document Type Definition.

Data-entry and Report Markup Language

The structure of an DRML document is shown schematically in Figure 1. Each document has a title followed by one or more reporting elements, each of which may be a single datum (`ITEM`), a collection of data elements (`GROUP`), or a `REPEAT` block. An `ITEM` is the fundamental data-entry element. There are three types of data items: binary, numeric, and textual. Binary items are indicated by the `<BIN>` tag; they attain a Boolean value of 1 (true) or 0 (false). For numeric (`<NUM>`) items, one can specify the units of measure (*e.g.*, "cm") and number of dimensions. Textual (`<TXT>`) items allow entry of one or more lines of narrative-text data. The name of the reporting concept follows the starting tag; *e.g.*, `<NUM>Size`. Ending tags for individual data items (*e.g.*, `</NUM>`) are optional; if omitted, the concept is assumed to end at the next concept's starting tag.

Reporting concepts are organized hierarchically and can be nested to arbitrary depth. If the `OPTION` attribute is specified for a group, then only one of the group's items can be selected at one time. The text immediately following the `<GROUP>` tag specifies the parent concept; subconcepts are bounded by `<GROUP>` and `</GROUP>` tags. Each concept in the hierarchy is a modifier of its parent concept and can only be used in conjunction with its parent. In DRML, all concepts are treated as one semantic type rather than being classified into "findings," "diagnoses," or other categories. This hierarchical approach has been used successfully in a system for reporting pelvic ultrasound examinations [8,9].

Groups of concepts can be repeated. Repeat blocks reduce redundancy in an DRML document; they consist of the `<REPEAT>` tag, the repeated concepts, the `<FOR>` tag, a list of parent concepts, and the `</REPEAT>` tag (Figure 1). For example, a DRML document for renal ultrasonography can use a Repeat block to define identical features for Right Kidney and Left Kidney (Figure 2). Repeated elements can be displayed in tabular format (by row or column) or

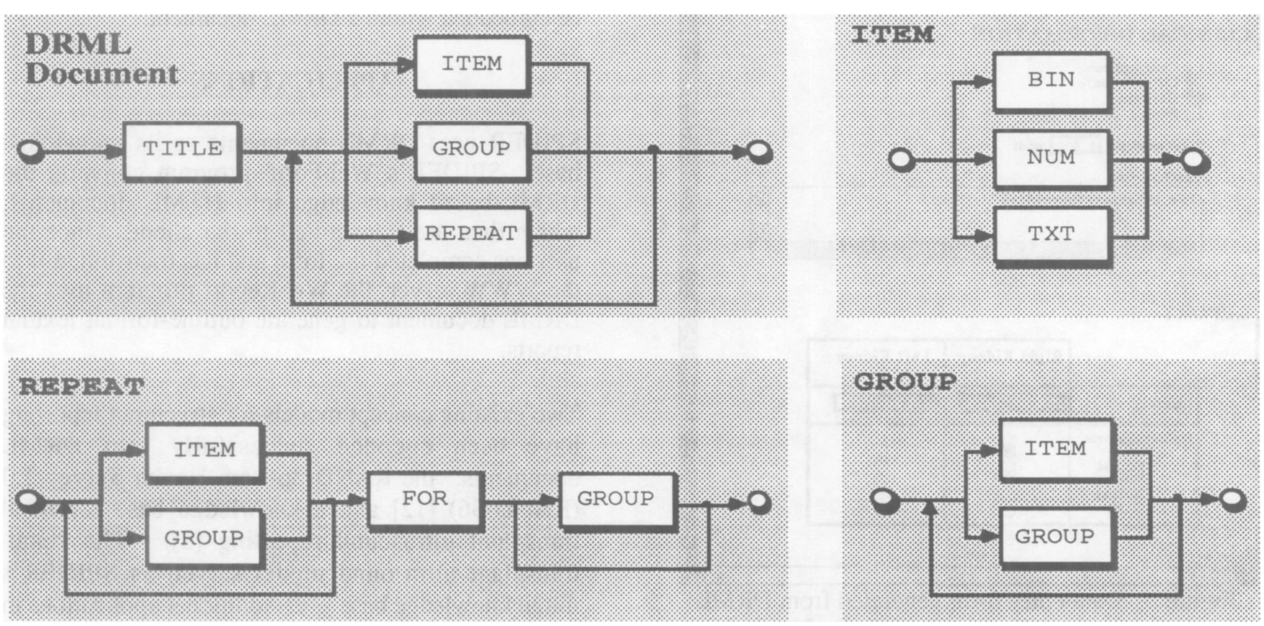

Figure 1. Schematic view of DRML document structure. Paths indicate how DRML elements can appear within the document and as components of other elements. For example, a GROUP contains one or more components, each of which can be either an ITEM or a GROUP.

```
<drml>
<title>Kidney Ultrasound</title>
<group>CLINICAL INFORMATION
 <bin>Right flank pain
 <bin>Left flank pain
 <bin>Hematuria
 <bin>Fever
 <num units="mg/dl">Serum creatinine
 <txt rows=3 cols=50>Other Indications
</group>
<repeat table=row>REPORT
 <num units="cm">Size
 <group option>Hydronephrosis
  <bin>None
  <bin>Minimal
  <bin>Mild
  <bin>Moderate
  <bin>Severe
 </group>
<for>
 <group>Right Kidney</group>
 <group>Left Kidney</group>
</repeat>
</drml>
```

Figure 2. Example DRML document encoding knowledge base for renal ultrasound reporting.

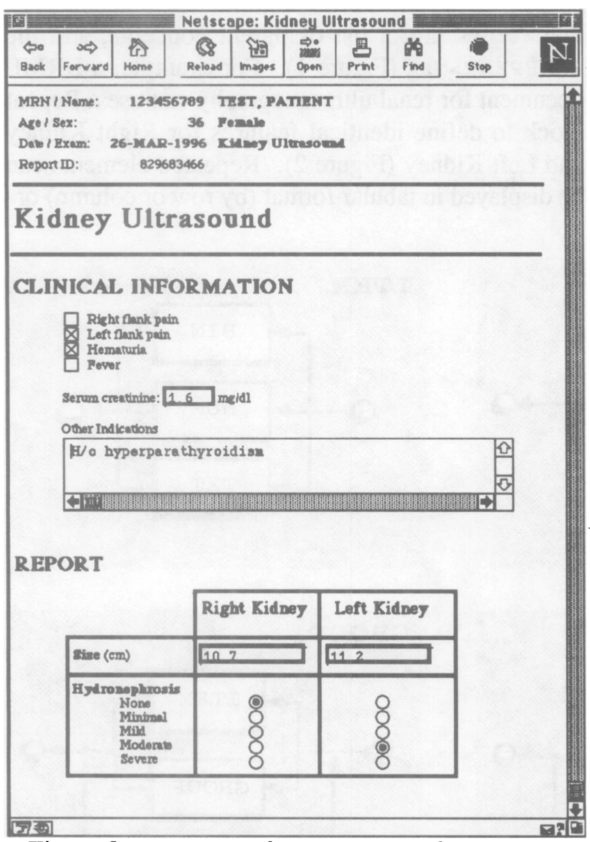

Figure 3. Data-entry form generated from DRML document above. Note use of tabular format for Right and Left Kidneys, and use of radio buttons for Hydronephrosis.

```
CLINICAL INFORMATION
    Left flank pain
    Hematuria
    Serum creatinine: 1.6 mg/dl
    H/o hyperparathyroidism

REPORT
    Right Kidney
        Size: 10.7 cm
        Hydronephrosis: None
    Left Kidney
        Size: 11.2 cm
        Hydronephrosis: Moderate
```

Figure 4. Outline-format report generated by SPIDER from data entered in Figure 3.

as regular, indented text. SPIDER generates data-entry forms (Figure 3) and reports (Figure 4) using the specifications in the DRML document.

Concepts can be linked to external vocabularies such as the UMLS Metathesaurus [10] or ACR index [11]; knowledge engineers can specify whether a reporting concept is narrower than, broader than, or equivalent to the external concept. One advantage of this approach is that patient encounters can be indexed automatically as their findings are recorded. To provide a "socket" into which data from external sources can be acquired without manual entry, DRML tags can specify a Uniform Resource Locator (URL) from which a concept's value should be taken. Textual notes can be inserted into DRML documents, and some HTML text-formatting codes, such as those for boldface type, italics, and fixed-width fonts, can be embedded within a DRML document.

APPLICATIONS

SPIDER uses DRML documents as its knowledge base. SPIDER's WebForm program converts the DRML-based knowledge into HTML documents, which then are transmitted to the client. Once the user has completed the form and has transmitted it to the WWW server, the WebReport program uses the DRML document to generate outline-format textual reports.

Two existing concept models for structured reporting have been encoded successfully into DRML documents: the RAND 36-Item Health Survey 1.0 (RAND-36) [12] and a knowledge base for renal transplant ultrasound reporting [3]. The renal-transplant ultrasound model served as SPIDER's initial knowledge base [3]. Additional applications in ultrasound reporting are now under development. In addition, SPIDER is being developed to serve as a user interface for ISIS (Intelligent Selection of

Imaging Studies), a case-based decision support system that helps primary-care physicians select diagnostic imaging procedures [13].

DISCUSSION

The World Wide Web (WWW, or simply the "Web") provides a medium for rapid, economical dissemination multimedia data — such as text, images, illustrations, audio, and video — from a source computer ("server") to a destination computer ("client") via the Internet [7,14]. At its most basic, the Web forms a global medium for publication. Web documents also can contain data-entry forms that allow users to send information to the Web server; the Common Gateway Interface allows servers to run programs to process this information. There are a growing number of medical applications of the Web, including resources for education, clinical reference, and retrieval of clinical information [15-18].

Structured reporting systems can be acceptable to physicians and can improve the efficiency of the reporting process [9,19]. Systems for structured reporting generally have not been implemented outside the institutions where they were developed (e.g., [8,9,20,21]). One of the barriers to development of systems for structured data entry has been the dependence on particular computer hardware and software platforms. The lack of diffusion of this technology may be due, in part, to difficulty in "programming" such systems or changing the set of reporting concepts. Developers of structured-reporting systems also have been limited by the need to concentrate on the user interface as well as the system's content.

The use of structured reporting may help satisfy the increasing demands for information by which to judge the quality and effectiveness of medical interventions, particularly in managed-care organizations. These demands for information, and the ability to retrieve it consistently, may lead physicians to more readily embrace structured reporting techniques. There is evidence that structured reporting systems can help improve the process of care by reducing the time and expense of dictating and transcribing radiology reports [19].

Preliminary experience with SPIDER has indicated that WWW technology can be used to implement systems for structured entry and retrieval of medical data [3]. In SPIDER's preliminary evaluation, three ultrasound technologists and eight radiologists — with little familiarity with computer systems or Web client software — entered clinical reports rapidly and easily [3]. SPIDER's ratings for overall impression, screen design, ease of learning the system, and system capabilities exceeded those of a clinically deployed, SuperCard-based ultrasound reporting system that uses hierarchically organized reporting concepts [9].

Web-based systems offer a uniform mechanism for interaction between physicians and computer systems. Increasing numbers of physicians are gaining access to the Internet and are becoming acquainted with the Web. With SPIDER, physicians can enter observations using a familiar software package — the user's preferred Web client program — rather than having to learn a new software system. Web client and server software is in the public domain and is readily available via the Internet; this can significantly decrease the time and resources needed to develop workable reporting systems. In a real clinical environment, the Web's encryption and password-protection features could be used to provide appropriate security for communication between a host computer and a physician at a remote site without the need for a dedicated, restricted-access communications link.

DRML allows designers of structured reporting applications to define one file that specifies the format of both the data-entry form and the printed report. Developers can concentrate on report content rather than user-interface programming. The language contains features to integrate reporting applications with existing vocabularies and data sources.

Future enhancements to DRML and SPIDER include automatic construction of a report's table of contents, incorporation of in-line images, and enhanced interaction among data-entry objects using Java-language applets. Also being considered is a way to aggregate common features of various imaging findings. For example, descriptions of perinephric fluid collections and hydronephrosis both include features such as severity and change from previous examination. Such modifiers could be placed in windows that would have a fixed on-screen location; this approach has been adopted successfully in the PEN-Ivory project [22].

Additional information about DRML and SPIDER is available at <http://www.mcw.edu/midas/spider>.

Acknowledgments

Supported in part by The Whitaker Foundation (Biomedical Engineering Research Grant to C.E.K.) and the National Library of Medicine (grant LM05705 to the Medical College of Wisconsin). The authors thank David Grosse and James Lowe for computer-systems support, and Douglas S. Bell, M.D., for critical review of the manuscript.

References

1. Dick RS, Steen EB, ed. *The Computer-Based Patient Record: An Essential Technology for Health Care*. Washington, DC: National Academy Press, 1991.
2. Shortliffe EH, Barnett GO. Medical data: their acquisition, storage, and use. In: Shortliffe EH, Perrault L, Wiederhold G, Fagan L, ed. *Medical Informatics: Computer Applications in Health Care*. Reading, MA: Addison-Wesley, 1990: 37-69.
3. Kahn CE, Jr, Wang K, Bell DS. Structured entry of radiology reports using World Wide Web technology. *RadioGraphics* 1996; 16:683-691.
4. International Standards Organization. *ISO 8879:1986 Information processing — Text and office systems — Standard Generalized Markup Language (SGML)*. Geneva: International Standards Organization, 1986.
5. Goldfarb CF. *The SGML Handbook*. Oxford: Clarendon Press, 1990. (Rubinsky Y, ed.
6. van Herwijnen E. *Practical SGML*. (2nd ed.) Boston: Kluwer Academic Publishers, 1994.
7. Berners-Lee T, Cailliau R, Luotonen A, Nielsen HF, Secret A. The World-Wide Web. *Communications of the ACM* 1994 76-82.
8. Bell DS, Greenes RA, Doubilet PD. Form-based clinical input from a structured vocabulary: Initial application in ultrasound reporting. In: Frisse ME, ed. *Proceedings of the 16th Annual Symposium on Computer Applications in Medical Care*. New York: McGraw-Hill, 1992: 789-791.
9. Bell DS, Greenes RA. Evaluation of UltraSTAR: performance of a collaborative structured data entry system. In: Ozbolt JG, ed. *Proceedings of the 18th Annual Symposium on Computer Applications in Medical Care*. Philadelphia: Hanley-Belfus, 1994: 216-221.
10. Lindberg DAB, Humpreys BL, McCray AT. The Unified Medical Language System. *Methods Inf Med* 1993; 32:281-291.
11. American College of Radiology. *Index for Radiological Diagnoses*. (rev. 3 ed.) Reston, VA: American College of Radiology, 1986.
12. Bell DS, Kahn CE, Jr. Assessing health status via the World Wide Web. In: Cimino J, ed. *Proceedings of the AMIA Annual Fall Symposium*. Philadelphia: Hanley & Belfus, 1996: (in press).
13. Kahn CE, Jr, Anderson GM. Case-based reasoning and imaging procedure selection. *Invest Radiol* 1994; 29:643-647.
14. Schatz BR, Hardin JB. NCSA Mosaic and the World Wide Web: global hypermedia protocols for the Internet. *Science* 1994; 265:895-901.
15. Glowniak JV. Medical resources on the Internet. *Ann Intern Med* 1995; 123:123-131.
16. Richardson ML. A World-Wide Web radiology teaching file server on the Internet. *Am J Roentgenol* 1995; 164:479-483.
17. Kahn CE, Jr. CHORUS: a computer-based radiology handbook for international collaboration via the World Wide Web. *RadioGraphics* 1995; 15:963-970.
18. Cimino JJ, Socratous SA, Clayton PD. Internet as clinical information system: application development using the World Wide Web. *J Am Med Informatics Assoc* 1995; 2:273-284.
19. Doubilet P, Greenes RA, Bell DS, Dichter MS, Benson CB, Fener EF. Computer-based reporting of US examinations: a time- and money-saving alternative to dictating. *Radiology* 1995; 197(P):160 (abstract).
20. Bernauer J, Gumrich K, Kutz S, Lindner P, Pretschner DP. An interactive report generator for bone scan studies. In: Clayton PD, ed. *Proceedings of the 15th Annual Symposium on Computer Applications in Medical Care*. New York: McGraw-Hill, 1991: 858-860.
21. Kuhn K, Zemmler T, Reichert M, Heinlein C, Roesner D. Structured data collection and knowledge-based user guidance for abdominal ultrasound reporting. In: Safran C, ed. *Proceedings of the 17th Annual Symposium of Computer Applications in Medical Care*. New York: McGraw-Hill, 1993: 311-315.
22. Poon AD, Fagan LM, Shortliffe EH. The PEN-Ivory project: exploring user-interface design for the selection of items from large controlled vocabularies of medicine. *J Am Med Informatics Assoc* 1996; 3:168-183.