Medical Infobahn for Europe
A. Hasman et al. (Eds.)
IOS Press, 2000

XML structured clinical information: A practical example

Schweiger R, Bürkle T, Hölzer S, Dudeck J
Department of Medical Informatics, University of Gießen, Heinrich-Buff-Ring 44, D-35392 Gießen, Germany

Abstract. *At Giessen university, a drug formulary comprising drug data and treatment guidelines is supplied to clinical users who can access the drug information by an index of drug substances and drug substance groups. The guideline itself is a textual description with related information such as drug substances and drug brand names. Since clinical users also want to access the information by drug names, we had to extract this information from the textual descriptions. The extraction however caused some effort. In order to not repeat this effort in the future, we used the eXtensible Markup Language (XML) to restructure the information sources. This paper describes our experiences with this kind of legacy to XML conversion and outlines a possible migration path towards the XML technology.*

1. Introduction

At Giessen university, a clinical information system [1,2] has been developed for several years. Besides clinical patient data it comprises a variety of medical information sources. Drug information, hygiene plans and guidelines for blood transfusion, microbiology and nursing care can be used at nearly 2000 clinical workstations [3,4,5,6].

A special emphasis has always been on drug information sources to support doctors when prescribing medication for patients. For this purpose, we established a formulary of drug recommendations or guidelines that is regularly updated by our hospital drug commission. This drug formulary, the so called "Hausliste", is organized into groups of drug substances. To each drug substance group there is a textual description or guideline that contains the related drug substances and drugs. The clinical user can access the information as an "electronic book" [6,7]. The electronic book provides an index of drug substance groups and drug substances and has been implemented on a DOS, Windows and HTML browser platform.

Clinical users are also interested to access the drug guidelines by criteria other than the drug substance. Especially the access by drug names is highly appreciated. Since the drug names are "hidden" within the textual descriptions, we had to invest some effort in order to extract the required information. This paper describes, how the eXtensible Markup Language XML [8] can be used to solve the extraction problem and, what is even more, to avoid the problem for future plans.

2. The problem

The following figure presents an exemplary information source as it is provided and updated by our pharmacist. These drug formulary units will be referred to as drug resources. The drug resource is structured into drug substance groups, drug substances, drug names, drug articles, drug units, drug prices (annulled for publication) and comments.

Stand 17.12.1997

10800 Bronchialmittel - Asthmamittel Rote Liste: Hauptgr. 27

(siehe auch Husten - und Erkältungspräparate, Seite 63)

10810 Beta 2 - Adrenergika

FREINAME Artikelbezeichnung	Handelsname(n) Stat.einheit	Preis/Stk
FENOTEROL (-HBr)	** Berotec(R) **	
Dosieraerosol 0.20mg/Hub 15ml	1 AER	00.00
Lösung LS 0.5% 20ml	1 FLA	0.00
SALBUTAMOL	BronchoSpray Klinge, ** Sultanol(R) **	
Dosieraerosol 0.1mg/Hub	1 AER	0.00
SALBUTAMOL(-SULFAT)	Loftan(R),** Volmac(R) **	
Retardtabl 4mg	TBL	0.00
Retardtabl 8mg	TBL	0.00
TERBUTALIN (-SULFAT)	Aerodur(R), ** Bricanyl(R)(/-Duriles) **	
Amp 0.5mg/1ml	10 AMP	0.00
Dosieraerosol 0.5mg/Hub (Aerodur(R))	1 AER	00.00
Retardtabl 7.5mg	50 TBL	0.00
TOLUBUTEROL	** Brelomax(R) **	
Saft 1mg/5ml 150ml	1 FLA	00.00

Figure 1: One of approximately 200 drug resources as provided by the drug commission. The resource belongs to the drug substance group "Bronchialmittel - Asthmamittel".

The provision of a drug index required the extraction of the drug brand names from the textual drug resources. We developed a parser that reads the line structure of the text and then decides for each line whether it corresponds to a drug substance group, to a drug substance or to a drug article. A drug article line could be identified most clearly by a drug unit (AER, FLA, TBL and AMP). The decision for the other lines was more difficult and required the analysis of preceding and subsequent lines (context). Once the meaning of a line was identified, the line could be further decomposed into its elements. However, since the resource structure was guessed from a few sample files, there were many structural exceptions such as multi-line drug articles (no consequent line structure), comments at an unexpected location or missing drug units. The parser result consequently needed to be checked by a person, and this was still a tedious task. The same procedure repeated as we started to construct a drug dictionary. We started to ask the question: How can we reduce the human effort in future, especially the human checks of the parser results?

3. Method

We decided to restructure the legacy format into an XML format since the XML technology promised a lot of support in terms of XML parsers, XML browsers and XSLT processors. We used our existing drug parser in order to mark up the different drug elements. Most of the contents was represented as XML attributes (NAME, UNIT, PRICE). Non-empty elements (GROUP, SUBSTANCE) are only used for the grouping of related information. The SUBSTANCE element for example contains all related DRUGs and ARTICLEs. The current XML structure reflects the original legacy structure and might be easily changed later with XSLT methods. The XML markup solves the identificaton problem, i.e. we can now reliably identify and extract the information inside the drug formulary without human checks of the extraction results.

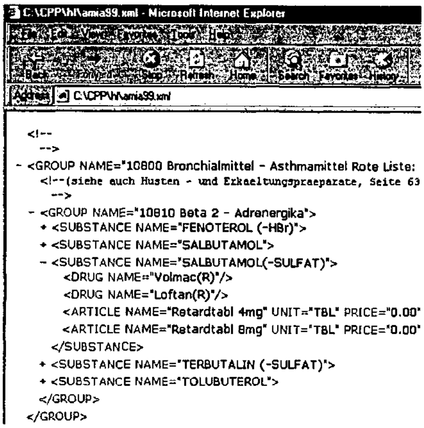

Figure 2: XML structured drug resource. An XML browser allows to navigate through the XML document in a comfortable way. Only the interesting drug substance has been opened (-), the other drug substances remain closed (+). No user defined stylesheet is needed for simple browsing, i.e. we can browse the XML file as it is.

During the unavoidable human checks of the XML drug resources we studied the structures of all 200 resources and summarized them into a compact structure. In future we only have to study one structure instead of 200 structures if we want to know the contents of the formulary. For the electronic description of the drug resource structure the XML technology provides several solutions such as DTDs (Document Type Definition) and XML Schemas.

The last step in the migration from the legacy format towards the XML format is the development of a user interface. The user interface comprises two aspects, the presentation of the drug information at the physician's office and the maintenance of the drug formulary by the drug commission, namely our pharmacist. Both user groups, the contents requester (physicians) and the contents provider (pharmacist), are barely interested in XML markup. The guideline presentation can be easily implemented using a browser platform and corresponding XML stylesheets (XSL, CSS). While the pharmacist had to deal with presentation issues of the legacy format, he can now leave the presentation issue completely to the software engineer, since XML separates contents (XML document) from presentation (stylesheet).

The hard nut in the development of the user interface is the data entry. We currently implement a mechanism that translates between XML documents and HTML forms. This document-form-translation is performed by a so called form server. The form server provides functions for data manipulation, data transparency and data retrieval. Data granules in this context might be documents or document fragments such as single elements and element levels. The form server can generate a new HTML form from a document type, i.e. a DTD or an XML Schema. Since the standard approaches DTD and XML Schema describe a document type rather than a form type, we designed a so called template approach. Beside of structural information, a template may also contain user prompts, data options and URIs for remote server access (fill existing data into the form). The user may now start with an empty form, enter some data and let the form server retrieve a list with the related documents in a context sensitive fashion. The user may then select a document from the list. The form server converts the selected document into an HTML form and provides the user with buttons for data manipulation (create, retrieve and delete XML elements, save the document etc.). Transparency functions play an important role for large documents. The user may e.g. close and open single elements or even complete levels of the XML document.

4. First results

As a first result we generated a list of approximately 700 drug substances and 1.000 drugs that are used in our university hospital. Moreover we succeeded to map the "Hausliste" to the German standard drug formulary, the so called "Rote Liste". The "Rote Liste" provides some additional information such as drug interactions.

5. Discussion

Clinical information today can be split into two large partitions. The first is patient related data which may be contained e.g. in an electronic patient record [9]. The second partition comprises clinical information sources such as electronic textbooks, electronic dictionaries or electronic guidelines. Both types of information are relevant for clinical staff and must be available at the clinical workstation. Funded research programs such as IAIMS [10,11] have especially concentrated on the integration between both kinds of information.

Many authors have emphasized the necessity to capture and to store clinical patient data in a structured fashion and to avoid free text entries. Cimino for example depicts a transition from unstructured patient data to more structured and coded data and emphasizes the significance of universal codes [12]. (see Figure 3)

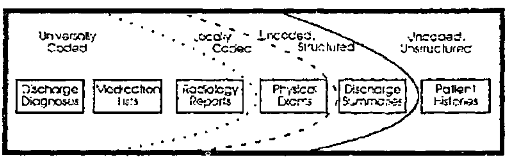

Figure 3: Pushing the border to more structured data. Taken from [12].

However, far less has been published about the structuring of clinical information sources. Health care asks for flexible information structures that can evolve over time. We think that XML provides an excellent answer to this question, because XML makes structures more explicit and easy to manipulate. We have started to adopt the XML technology for the solution of a concrete problem, namely the structuring of our drug formulary. We are convinced that many other information sources might be structured and handled in a similar way.

It turned out, however, that XML does not yet solve the problem of the end user access to XML documents. So called XML editors still confront the user with XML markup and cannot be used in a clinical environment. For this reason we developed a generic approach that might be reused for other clinical domains. Figure 4 shows a very simple form for the acquisition of patient data. This data entry form is generated automatically from a corresponding template that describes the form in terms of structure, user prompts and remote data connections. Such a template is easy to adapt to the latest documentation needs. This way we are able to combine the advantages of structured data with the advantages of flexible documentation. By providing templates for several domains (drugs, surgery, orthopaedics etc.) we might even manage a complete clinical document repository. In that sense, XML turned out a proper means to introduce structure into domains where only little structure has been captured so far.

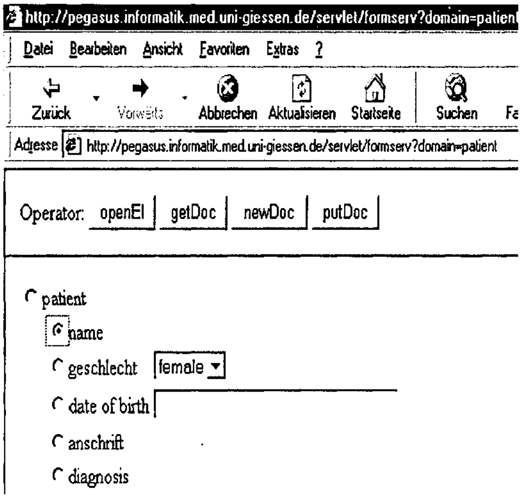

Figure 4: HTML form for the manipulation of XML documents. The name element is selected and the corresponding operators show up in the upper frame.

References

[1] Prokosch HU et al: WING - Entering a New Phase of Electronic Data Processing at the Giessen University Hospital. Meth Inform Med 30 (1991) 289-298.

[2] Michel A: Migration steps from a mainframe based HIS approach to an open HIS environment. In Hospital Information Systems: Current Status and Future Trends. HIS: Design and Development Characteristics, Elsevier, Amsterdam 1995, 267-286.

[3] Prokosch HU et al: From HIS to IAIMS: Expanding the Scope of Information Processing Applications in a German University Hospital. Proc. 18th SCAMC, Hanley & Belfus Inc 1994, 115-119.

[4] Prokosch HU et al: One-Stop-Information-Shopping: Do we meet the Information needs of the hospital staff? Proc. MIE, Lisbon 1994, 362-365.

[5] Bürkle T, Prokosch HU, Dudeck J: Knowledge and Information Based Functions in the Giessen Hospital Information System - Present and Future. In Toward An Electronic Health Record, omnipress Madison/Wisconsin 1996, 100-107.

[6] Bürkle T, Ruan W, Dudeck J: Darstellung klinischen Wissens am Arztarbeitsplatz: Der Arbeitskontext sollte berücksichtigt werden. Proc. 43. Tagung GMDS 1998, 222-225.

[7] Brumhard M, Prokosch HU: Die elektronische Arzneimittelliste - Neue Wege zur Informationsbereitstellung im Gießener Universitätsklinikum. Krankenhauspharmazie 1996, 129-133.

[8] www.w3.org

[9] Borst F: Computer-based patient records, Synopsis. Yearbook of Medical Informatics 1995, 191-194.

[10] Clayton PD, Cimino JJ, Anderson RK et al: Integrated Advanced Information Management System (IAIMS): Accomplishments from a global and local perspective. Hospital Information Systems - Design and Developments Characteristics; Impact and Future Architecture, 1995, 287-311.

[11] Stead WW, Sittig DF: Integrated Advanced Information Management Systems: Strategies and Architectures for Fast Track Development. Hospital Information Systems: Design and Development Characteristics, 1995, 313-330.

[12] Cimino JJ: Data storage and knowledge representation for clinical workstations. Yearbook of Medical Informatics 1995, 289-298.