Clinical Data Entry

Erik M. van Mulligen, H. Stam, A.M. van Ginneken
Dept. of Medical Informatics, Erasmus University Rotterdam, The Netherlands

Routine capture of patient data for a computer-based patient record system remains a subject of study. Time constraints that require fast data entry and maximal expression power are in favor of free text data entry. However, using patient data directly for decision support systems, for quality assessment, etc. requires structured data entry, which appears to be more tedious and time consuming. In this paper, a prototype clinical data entry application is described that combines free text and structured data entry in one single application and allows clinicians to smoothly switch between these two different input styles. A knowledge base involving a semantic network of clinical data entry terms and their properties and relationships is used by this application to support structured data entry. From structured data, sentences are generated and shown in a text processor together with the free text. This presentation metaphor allows for easy integrated presentation of structured data and free text.

INTRODUCTION

In the last five years, the data model for a computer-based patient record system called ORCA (Open Record for CAre) has been published in several journals [1,2]. This data model has been implemented in the I4C (Integration and Communication for the Continuity of Cardiac Care) project in the 4th framework programme of the European Commission [3]. This project aims at a multimedia clinical workstation for cardiology integrating a computer-based patient record system with data from "legacy" systems in health care environments. The ORCA data model has been designed with the demands of the evolving health care systems in mind. ORCA uses a knowledge base that represents in a semantic network all clinical terms that can be used and how they can be combined into meaningful expressions. In this paper, a first version of an application for capturing clinical data and storing it in the ORCA computer-based patient record system is described. One of the objectives of this application is to explore how the richness of the ORCA data model can be used by clinicians in an easy way, i.e., so that they can use it during patient encounters [4].

Applications for clinical data entry can be classified either as using free text (possibly with natural language processing) or as using structured data entry. Structured data entry restricts clinicians to a predefined vocabulary of medical terms to capture patient data [5,6,7,8]. However, data entered in a structured fashion are easier to use for clinical research and quality assessment, for exchange with other care providers (indicating what parts of the structured record should be exchanged), for integrated decision support systems, for accessing online biomedical literature, etc. [9]. Free text data entry gives clinicians maximum freedom on what they want to store, but it limits the additional value one might get from having patient data stored in a computer. Furthermore, free text data entry resembles much more the way clinicians interact with the paper-based patient record. Due to the limited use of free text data, our initial application had a strong focus on structured data entry. Free text entry was supported, but with the intention to express only information that is not covered by the predefined terms in the semantic network. Upon request to adapt ORCA for use by a wider range of clinicians, including non-academics, free text entry has also enabled as an alternative for structured data entry. The implication of this approach to our clinical data entry application is discussed.

Experience with clinical data entry applications taught that they are typically not designed for browsing and skimming the record [10]. Crucial for the clinical data entry application has been flexibility of data entry according to the physician's workflow and predictable navigation through the computer-based patient record.

MATERIAL AND METHODS

Architecture

The I4C multimedia clinical workstation follows the client-server paradigm. The server (implemented on Windows NT server) contains

1091—8280/98/$5.00 © 1998 AMIA, Inc.

the ORCA computer-based patient record system (implemented in the MS-SQL database management system) and software to access data (Laboratory data, ECGs, and angiograms) contained in other systems in the hospital environment (legacy systems). The database management system will, when requested for data contained in a legacy system, start communication with a legacy system and retrieve the requested data. The client side (implemented on Windows95 with Delphi) is a set of applications that communicate with ORCA. This set of applications contains administrative applications, applications for viewing laboratory data, ECGs, and angiograms, and applications for entering data, such as visits, medication, and clinical data.

Data model

The clinical data entry application has been built on top of the ORCA data model. This model uses a so-called direct and indirect data model. The data model has been implemented in a relational database management system (MS-SQL server) and uses tables for storing clinical data. Those clinical data sets (e.g., medication, laboratory, and diagnosis) that typically can be described with a fixed set of descriptors are implemented in the direct data model using table attributes for each descriptor. However, those elements in the record that can not be described with a fixed set of descriptors (e.g., patient history or physical examination) are stored in the indirect data model.

Patient data consist of a tree of instances in the data model. Interpretation of the meaning of the instances only requires the thesaurus of the semantic network, not the network itself. The essence of the indirect data model is separation between data representation and content. Hence, this model provides the flexibility needed to tailor data entry to a variety of medical domains without the need to adapt the application.

Knowledge base for data entry

All clinical terms that can be stored in the data model are contained in ORCA's knowledge base. This knowledge base contains a semantic network of terms. This semantic network contains relationships and properties that are relevant for the input and presentation of patient data.

Relations

The most important relationships in the semantic network are: *has_specialization, has_feature, has_values,* and *refers_to.*

The *has_specialization* relationship between two terms is used to express that the child term specified has a narrower meaning than its parent term, or that the parent term is a more general term than the child term. This is in correspondence with the class-subclass concept in the sense that *has_feature* and *has_value* relationships are inherited by the *has_specialization* children.

Has_feature children represent aspects by which the parent can be described in a qualitative way. *Has_value* children enable a quantitative description of the parent. The children represent the units for the single value or range to be entered.

The *refers_to* relationship is used when a particular term may be reached from more than one viewpoint. For example, micturition may be important in a cardiovascular, urogenital, or endocrine context. There can, however, only be one micturition in the patient data. For coherent representation of all micturition findings, one of the parents is designated as main parent. When reaching a *refers_to* child, the path via the main parent is instantiated.

Properties

The properties only apply to concepts themselves. Properties provide the option to express whether abnormalities are *absent*, such as a stomach ulcer. In contrast, the heart can not be *absent*, but *normal.* When a normal statement denotes a set of findings, such as a normal examination of the heart, each physician can define the meaning of that normal statement and save if for future use.

The *multiple* property is defined for concepts that can be instantiated more than once, such as a tumor and multiple blood pressure measurements.

Some terms in the semantic network are created to improve the efficiency of browsing. These auxiliary terms are used to divide a large set of children into logical subsets of children. Each subset is applicable within a particular context, and depending on the context one of the subsets is automatically enabled.

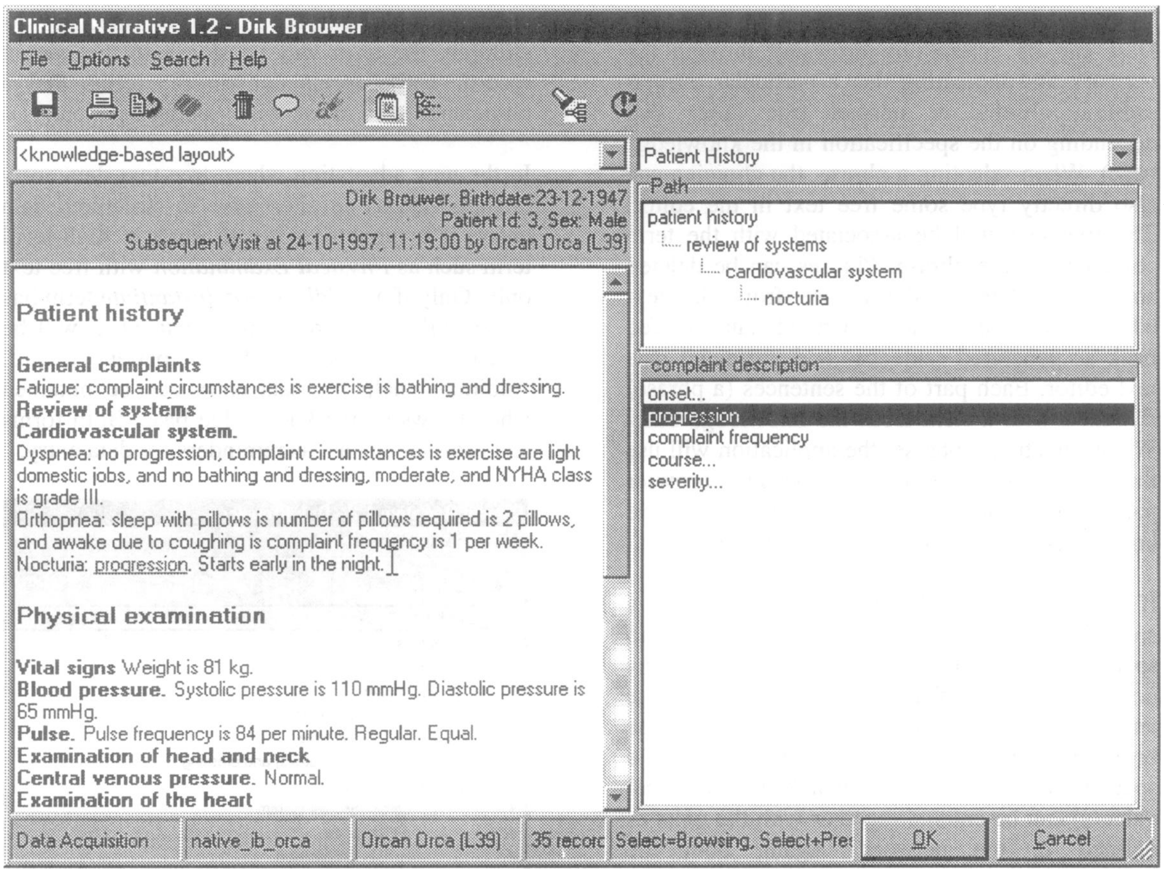

Figure 1 - ORCA's textual feedback editor (left pane) showing the sentences that have been generated from the structured data. In this first version, a simple algorithm has been used for text generation resulting in sometimes atrocious sentences. The underlined term in the left pane indicates the term that has the focus. The user has to focus a term to which he wants to add free text. The term with the focus is shown underlined and (on the screen) in red. The text following the underlined (focus) term has been entered as free text. The right pane contains the browser. The top dropdown menu contains the highest terms (domains), the *Path* window contains the path through the semantic network starting from the highest term, and the lowest window contains the children for the current term (decreased). On the screen, structured data is shown in blue and free text in black.

Since these terms are not necessary to represent the medical content, they have the *visibility* property disabled. As a result, these terms are used neither for data entry nor for data visualization.

In the initial application, the *childless-not-instantiated* property was introduced to avoid incomplete descriptions. For instance, it is not useful to store the term *Physical Examination* in the computer-based patient record has no meaning if no children have been instantiated.

The *sequence* property is used to order terms that have a common parent term. This property guarantees that siblings will always be shown at a fixed position in the computer-based patient record system, irrespective of the order of data entry.

RESULTS

Clinical Data Entry

In the I4C project, an application for clinical data entry that combines structured data entry and free text in a seamless way has been implemented. It has been assumed that all free text should be associated with a structured data entry term. A completely free text patient record will be associated with the two highest terms *patient history* and *physical examination*. The application consists of two main panes: a pane for the presentation of the patient record and a pane for browsing through the semantic network (see Figure 1).

Data can be entered by browsing through the network and indicating that a particular term is present, absent, or normal (the latter two depending on the specification in the knowledge base). When selecting a phrase, the clinician can also directly type some free text in the editor. The free text will be associated with the term contained in the phrase. Phrases can be deleted simply by deleting the phrase from the text editor. Data entered as structured data are fed back as generated sentences and presented in a text editor. Each part of the sentences (a phrase) is linked with a term from the knowledge base. When selecting a phrase, the application will use the associated term to directly focus the browser pane on that part of the semantic network in the knowledge base that is related to the term.

If a multiple or combination term is selected, a special window will pop up that can be used to specify which combinations coexist or how many occurrences exist of the finding specified by the term. When the clinician selects a term in the browser pane that is a quantitative feature, a special window pops up in which the numerical value(s) can be entered together with the units as defined in the network. Again, both the multiple/combination and quantitative terms are presented as text in the patient record text editor.

For data entry, flexibility is important so that each physician can follow his own strategy and his spontaneous associations. In contrast, when consulting existing patient data, it is convenient if these are presented in a predictable order. In ORCA, this order corresponds with the order defined in the semantic network. In practice this means that if a finding is not present at a certain place, the clinician does not need to search the rest of the progress notes. Items entered with structured data entry are linked with their corresponding terms in the semantic network browser. Access to a term for data entry is supported in two additional ways. A find function enables quick location of a term via a partial search string. The other for quick location of a term is the selection of one of the already entered terms.

When starting to capture data for a new visit, the clinician can decide to retrieve a predefined form. Such a form contains already a set of knowledge base terms. The clinician can directly type free text below the terms or use structured data entry to add new child terms. When the clinician adds text to a term or specifies its children, the term will be stored in the patient record. These forms contain the terms that a clinician is used to ask when seeing a patient.

In the new adaptation where free-text data entry is supported as an alternative to structured data entry, it is now possible to enter a high-level term such as *Physical Examination* with free text only. Only if a *childless_not_instantiate* term has either children or free text associated, it will be saved in the computer-based patient record. Additional help for data entry can be obtained when browsing the knowledge base. Both forms and browser can act as reminders for data entry.

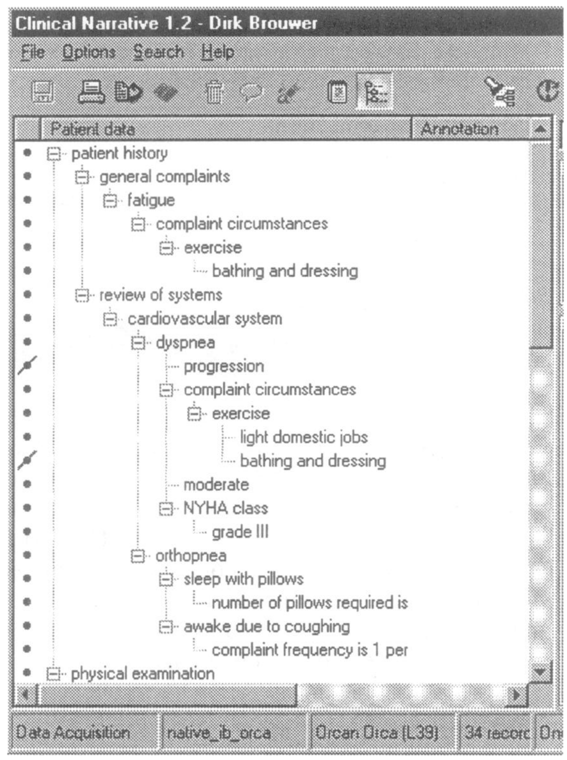

Figure 2 - Tree-wise presentation of patient data.

In addition to the narrative feedback style, a tree visualization style has also been implemented. This tree presentation shows more clearly the structure of the patient data tree as an instance of the partial network. The free text associated with a term is shown in a separate column. Icons before each tree node (see Figure 2) indicate presence, absence, and normality.

DISCUSSION

The clinical data entry application has been installed at various centers in Europe now. The first responses to this mixture of structured data entry and free text entry are positive. A number of actual patient paper records have been entered in ORCA by clinicians retrospectively. The contents of ORCA's knowledge base with the option to enter free text seems to be sufficient enough for storing patient records electronically. Although new interfaces have already been designed, the responses of clinicians to the current interface will first be investigated and analyzed. A formal evaluation has been planned for the end of June 1998. This formal evaluation will be based on a session with questionnaires and on quantitative parameters established during interaction by clincians with the application.

Regarding the atrocious sentences generated by the first prototype, it is clear that this text generation algorithm has to be improved. The inclusion of lexical descriptions to the semantic network will assist the application in generating more readable sentences.

Properties and relationships of the knowledge base can be effectively used by the clinical data entry application to improve the data capturing process and the data visualization. Additional functionality, such as quickly locating terms in the network, having directly an overview of the features that can be specified for a particular term, and the use of predefined forms further improve the quality of the clinical data entry.

The integration of decision support systems, quality assessment systems, and direct access to online biomedical literature, to the patient record will create benefits for clinicians and motivate them to structure and encode patient data and it is thought that clinicians will gradually switch from mainly free text to mainly structured data entry. Crucial will be the time necessary to find a term in the network and to create the data.

References

1. Van Ginneken AM. Structured Data Entry in ORCA: the strengths of two models combined. In: Cimino JJ, ed. J Am Med Inf Assoc, symposium supplement, 1996:797-83.
2. Van Ginneken AM, Stam H, Moorman PW. A multi-strategy approach for medical records of specialists. In: Greenes RA et al., eds. MedInfo 95. Edmonton: HC & CC, 1995:290-4.
3. Integration and Communication for the Continuity of Cardiac Care. Project Proposal for the Fourth Framework Programme on Health Telematics C/7 of DG XIII of the European Union
4. Van Ginneken AM, de Wilde M, van Mulligen EM, Stam H. Can data representation and interface demands be reconciled? Approach in ORCA. In: Masys DR, ed. J. Am Med Inf Assoc, symposium supplement, 1997:779-86.
5. Bell DS, Greenes RA. Evaluation of UltraSTAR: Performance of a collaborative structured data entry system. In: Ozbolt JG, ed. J Am Med Inf Assoc, symposium supplement, 1994:216-226.
6. Poon AD, Fagan LM. The design and evaluation of a pen-based computer system for structured data entry. In: Ozbolt JG, ed. J Am Med Inf Assoc, symposium supplement, 1994:447-51.
7. Moorman PW, van Ginneken AM, van der Lei J, van Bemmel JH. A model for structured data entry based on explicit descriptional knowledge. Meth Inf Med 1994;33(5):454-63.
8. Trace D, Naeymi-Rad F, Haines D, et al. Intelligent Medical Record-entry (IMR-E). J Med Syst 1993;17(3-4):139-51.
9. Dick RS, Steen EB, Detmer DE, eds. The computer-based patient record: an essential technology for health care. Revised Edition. Institute of Medicine. National Academy Press, 1997.
10. Nygren E, Johnson M, Henriksson P. Reading the medical record I. Analysis of physician's ways of reading the medical record. Comput. Biomed. 1992; 39:1-12.