Methods of Information in Medicine

R. Grütter[1], W. Fierz[2]

An Electronic Study Form to Support Collaborating Agents in the Management of Clinical Knowledge

[1]Institute for Media and Communications Management, University of St. Gallen,
[2]Institute for Clinical Microbiology and Immunology, St. Gallen, Switzerland

Abstract: When dealing with biological organisms, one has to take into account some peculiarities which significantly affect the representation of knowledge about them. These are complemented by the limitations in the representation of propositional knowledge, i.e. the majority of clinical knowledge, by artificial agents. Thus, the opportunities to automate the management of clinical knowledge are widely restricted to closed contexts and to procedural knowledge. Therefore, in dynamic and complex real-world settings such as health care provision to HIV-infected patients human and artificial agents must collaborate in order to optimize the time/quality antinomy of services provided. If applied to the implementation level, the overall requirement ensues that the language used to model clinical contexts should be both human- and machine-interpretable. The eXtensible Markup Language (XML), which is used to develop an electronic study form, is evaluated against this requirement, and its contribution to collaboration of human and artificial agents in the management of clinical knowledge is analyzed.

Keywords: Knowledge Representation, Artificial Intelligence, eXtensible Markup Language (XML), Clinical Trial

1. Introductory Remarks

The conceptual framework underlying this study is the knowledge medium, as being developed by Professor Dr. Beat F. Schmid and his coworkers at the Institute for Media and Communications Management of the University of St. Gallen, Switzerland [1]. Therefore, the terms used refer to the definitions of this framework.

We define the term *knowledge* as the internal state of an agent resulting from the input and the processing of information. This definition implies that knowledge, in a narrow sense, must be associated with an agent. Therefore, we avoid using the crude term for knowledge when it is related to a (stateless) carrier. Instead, we use "externalized knowledge", or simply "information", in such cases. Information is distinguished from data by its high-level semantics, referring to real-world objects, whereas low-level semantics of data, i.e., elementary data types, indicate only whether a given bitstream should be interpreted, e.g., as a character, an integer, etc.

The term *agent* is defined in a broad sense, including human as well as artificial agents. In this context, the latter refers to the concept of a virtual machine. A virtual (or abstract) machine implements an algorithm to process information. It has a set of valid inputs and, conversely, a set of well-defined outputs. It also has an internal state. Each input changes the internal state and produces an output. The term "virtual" refers to the black-box concept. Thus, the machine can also be a software program instead of a set of interoperating physical components. In the case of human beings, agents form a community with common beliefs, desires, and intentions. This community is organized in some way, either explicitly or implicitly. This means that agents act in well-defined roles and according to generally accepted protocols.

2. Application Area

2.1 The Swiss HIV Cohort Study

The Swiss HIV Cohort Study (SHCS) was initiated in 1987 (1) to collect clinical, laboratory, and socioeconomic data with the intention to analyze the dissemination and progression of the HIV-infection in Switzerland, as well as to promote and facilitate clinical research, and (2) to improve the health care services provided to HIV-infected patients [2]. SHCS involves the outpatient clinics of central hospitals (referred to as "Cohort centers") located in the cities of Basel, Berne, Geneva, Lausanne, Lugano, St. Gallen, and Zurich, as well as the

Coordination and Data Center in Lausanne. In recent years, more and more private practitioners joined the study to complement the Cohort centers.

Throughout the study, data are collected at the local Cohort centers and selected sets thereof are anonymously stored in a central database at the Coordination and Data Center. The whole process of filling in the preformatted study form, sending the form to the Coordination and Data Center, and entering the data into the database system is done paper-based and manually. Primarily because of the considerable time delay, the database has not been fully exploited thus far for daily clinical patient care. With the advances of antiretroviral (i. e., directed against retroviruses; the HI-virus is a retrovirus) therapy, information on past treatment and laboratory data is becoming increasingly important. To make use of this information, the technical infrastructure supporting the study has to be redesigned from scratch.

2.2 *Objectives of the Redesign Project*

In accordance with the objectives of SHCS, the redesign project aims at (1) improving the availability of data for timely research planning and (2) providing the clinicians with externalized knowledge for daily patient care. These two goals are further related: improving the availability of data reduces the duration of research projects, thereby speeding up the creation of new knowledge. Based on a redesigned technical infrastructure, this new knowledge can, in turn, be made available faster to the points of care.

Regarding the overall redesign project, this study focuses on aspect (2), that is, providing the clinicians with externalized knowledge for daily patient care. It concerns the collaboration between human and artificial agents and the contribution of an electronic study form to the management of clinical knowledge.

To provide a service of appropriate quality, clinicians must have immediate access to patient data and clinical knowledge at the time of decision making. As far as possible they should be assisted in their work by artificial agents. But how can this collaboration be best achieved? What are the requirements with respect to knowledge representation?

2.3 *Problems Associated with the Representation of Clinical Knowledge*

When dealing with knowledge about biological organisms, one has to take into account some special aspects that make appropriate processing difficult [3]:

First, a biological organism is to a large extent sub-divisible. Depending on the objectives of the documentation and the interests of the observer, any desired level of granularity can be accomplished and any classification scheme can be applied (e. g. a surgeon classifies a given disease along the locations affected, whereas a health insurance company tends to use a lump-sum per case classification).

The biological organism is also connected to the environment in many different ways. Abstraction from environmental conditions is often not possible as these conditions may essentially contribute to disease manifestation, thereby complicating cause/effect relationships.

Diseases often reflect predominantly or purely mental concepts with no material representation (e. g. our disease perception) which may vary considerably over space and time.

These aspects contribute to the insight that the documentation of diseases and the perception of illness are highly context-sensitive, often time-dependent and subjective. As clinical knowledge is at least partly a result of disease documentation and perception, these objections apply also to clinical knowledge, thereby hampering an appropriate representation. In this context, clinical knowledge is understood as the sum of knowledge required to properly apply the SOAP classification scheme (Subjective, Objective, Assessment, and Plan) [4]. The clinical examination starts with the subjective patient history and the current problem as perceived by the patient. These guide the physician in selecting the objective clinical and laboratory examination procedures. On the basis of the subjective and objective information, an overall assessment in terms of a diagnosis is made and a treatment plan is established.

The special aspects of biological organisms identified above, which significantly affect the representation of clinical knowledge, are complemented by limitations of artificial agents.

Whereas the representation of *procedural* knowledge (how to deal with incoming information to produce a well-defined output) is well established in applied informatics, the representation of *propositional* knowledge (what do the represented concepts and their relations mean) has barely developed from its infancy (these arguments will be discussed in Section 4). Clinical knowledge is primarily of the second type. In addition, the complexity of the information "objects" (i. e., data types) processed by artificial agents and hence the set of possible internal states (i. e., the represented knowledge) are very limited when compared to humans (for instance, during the patient's visit the experienced physician not only perceives and interprets the verbal information provided by the patient but also his or her gestures, the cadence of his or her voice, etc.) [5].

We conclude that the formal representation of clinical knowledge and, therefore, the possibilities to automate its management are largely restricted to confined contexts and to procedural knowledge. Thus, in dynamic and complex real-world settings, such as healthcare provision to HIV-infected patients, human and artificial agents must collaborate to optimize the time versus quality aspects of the services provided. If we apply this to the implementation of our project, it is required that the language used to model clinical contexts is both man- and machine-interpretable. This is further supported by the fact that we have to deal with a long lifecycle of information in health care, which might outlive many systems and technologies.

3. The Electronic Study Form

Based on an overall architectural concept, we developed an Electronic Study Form (ESF) as common middleware and made a first step towards encompassing computerization [6]. This was done using XML (eXtensible Markup Language) [7, 8]. XML is a sim-

```
<!DOCTYPE SHCS [
      ...
      <!ELEMENT LAB (REGULARTESTS,...)>
      <!ELEMENT REGULARTESTS (DATE,HEMATOLOGY,...)>
      <!ELEMENT DATE (#PCDATA)>
      <!ATTLIST DATE
            DD CDATA #REQUIRED
            MM CDATA #REQUIRED
            YYYY CDATA #REQUIRED>
      <!ELEMENT HEMATOLOGY (#PCDATA)>
      <!ATTLIST HEMATOLOGY
            LEUKOCYT CDATA #REQUIRED
            HEMOGLOB CDATA #REQUIRED
            PLATELET CDATA #REQUIRED>
      ...
]>
<SHCS>
      ...
      <LAB>
            <REGULARTESTS>
                  <DATE YYYY="1996" MM="12" DD="16"/>
                  <HEMATOLOGY LEUKOCYT="5700" HEMOGLOB="15"
                  PLATELET="217"/>
                  ...
            </REGULARTESTS>
            ...
      </LAB>
      ...
</SHCS>
```

Fig. 1 Excerpts of XML Document – Type Definition and Document.

plified subset of SGML (Standard Generalized Markup Language, ISO 8879), created particularly to support distributed computing on the World Wide Web (WWW). It was accepted as a recommendation by the WWW consortium (W3C) in February 1998 (the magazine Byte of January 1998 included XML as one of the top 25 technologies to watch that year). Besides conformance with the requirement specified above, XML was chosen to ensure independence on proprietary software systems in a distributed and heterogeneous environment. Furthermore, as a Web-based technology (cheap, short development cycles, ubiquitous access), it allows for easy integration of private practitioners with the system.

A particularity of XML is the explicit declaration of element types and representation of information on the structure of the document. These are represented in the Document Type Definition (DTD), a declarative part at the beginning of the document (Fig. 1). Alternatively, the DTD can be stored in a separate file and referenced in the document.

4. Results and Discussion

Referring to the definitions as given in the first section, the ESF is *not* an agent. It is rather a stateless carrier capable to exchange externalized knowledge between human and artificial agents. This carrier will be evaluated against the specified requirement and its contribution to collaboration between human and artificial agents in the management of clinical knowledge will be analyzed.

Table 1 Externalized knowledge represented by the Electronic Study Form.

Level of document	Type of externalized knowledge	Representation	Interpretable by
document carrier	propositional	markup and namespaces	human (artificial agents)
document carrier	procedural	DTD	(human) artificial agents
contents	propositional	elementary data types	human (artificial agents if elementary data types are specified with an additional layer)

We recall the requirement that the language used to model clinical contexts should be both man- and machine-interpretable. At a first glance, we can fulfill that requirement (Fig. 1). However, when going into more details, we have to distinguish between different levels (the document carrier, i. e., the meta-level, and the contents level) and different types of externalized knowledge (propositional and procedural).

At the meta-level we have to deal with XML as a meta-language. This meta-language includes a meta-syntax, i. e., a markup (a markup includes everything that is not content, which is a nondefinition, but helpful in either case), and semantics, i. e., the DTD that represents the declarations of element types and the hierarchical relationships among them, that is, the document structure.

Besides providing the logical framework at the carrier level, the meta-syntax assigns context information to the element contents (e. g., the tag <HEMATOLOGY...> provides the information that the content between the quotation marks is the result of a blood-cell count). This context information can be derived from so-called namespaces. Namespaces define the semantics of a markup, following a universal names approach, similar to an ontology [9]. They refer to predefined schemas. In the context of this paper, such a schema could be a clinical classification system. These schemas can be linked to a glossary with (nominal) definitions of the concepts used. (We pursue a similar approach with the NetAcademy project [10].) Markup and namespaces refer to the propositional part of externalized knowledge. They reduce the conceptualization of element contents to the syntactic level of the meta-language. Being syntactic constructs of a non-ambiguous formal language they can be processed by artificial agents. For instance, markup and namespaces allow search engines or other tools to operate over a range of documents that vary in many respects, but use common names for common element types.

However, the support of propositional knowledge by the XML meta-syntax is rather limited in anticipating the above-mentioned inherent problems associated with the representation of clinical knowledge. At this point, the artificial agents operating over XML have to be complemented by a different category of artificial reasoning, such as neural networks, and by humans.

The DTD relates the element types to each other, thereby imposing a tree structure on the document. It refers to the procedural part of externalized knowledge in that it sets the path for navigating from one node of the tree to another. This "structural semantics" can be used by artificial agents to browse through the document (e. g., by object-oriented database systems, implementing the structure of the document as a database structure). Alternatively, depending on the role-based privileges of the human agent and on the state of the process of health-care delivery, specific filters can be applied to allow for the retrieval of well-defined document components, based on the relationships between element types, while preventing access to confidential contents.

At the contents level, XML is rather poor in supporting elementary data types. Only the type of (parsed) character data (`#PCDATA`) are distinguished (thus far, further data types have to be specified with an additional layer). However, as shown, this shortcoming is more than compensated by the meta-syntactic concepts of markup and namespaces.

The levels and types of externalized knowledge represented by the ESF are summarized in Table 1. The ESF meets the specified requirement of man and machine interpretability. It represents different levels and different types of externalized knowledge allowing human and artificial agents to operate over it in many different ways. Therefore, the ESF provides a robust basis for collaboration between human and artificial agents. This collaboration is essential, as technology alone cannot fully anticipate the problems associated with the representation of clinical knowledge, but significant assistance by artificial agents allows to optimize the time versus quality aspects of the services provided. The latter applies particularly to dynamic and complex real-world settings, such as health-care provision to HIV-infected patients.

Acknowledgment

The project presented in this paper is generously supported by a grant from the Swiss HIV Cohort Study (SHCS).

REFERENCES

1. Schmid BF, ed. *Wissensmedien.* Wiesbaden: Gabler Verlag, 1999.
2. Ledergerber B, von Overbeck J, Egger M, Luthy R. The Swiss HIV Cohort Study: rationale, organization and selected baseline characteristics. Soz Präventivmed 1994; 39 (6): 387-94.
3. Moehr JR. Health informatics – a scientific challenge? Meth Inform Med 1994; 33 (3): 250-3.
4. Weed LL. *Medical Records, Medical Education and Patient Care.* Cleveland: Case Western Reserve University Press, 1969.
5. Grütter R. Die Konzepte der erweiterten Semantik – ein Beitrag zur Entwicklung von Wissensmedien. Zeitschrift Informatik 1998; 3: 30-3.
6. Fierz W, Grütter R. The use of a structured XML document to integrate distributed heterogeneous computer systems in a multicentered clinical study. In: Greiser E, Wischnewsky M, eds. *Methoden der Medizinischen Informatik, Biometrie und Epidemiologie in der modernen Informationsgesellschaft.* MMV, 1998.
7. Connolly D. *XML: Principles, Tools, and Techniques.* World Wide Web Journal 2 (4), Fall 1997.
8. Light R. *Presenting XML.* Indianapolis IN, United States of America: Sams.net Publishing, 1997.
9. http://www.w3.org/XML/, January 1999: W3C Recommendation: Namespaces in XML.
10. Handschuh S, Lechner U, Lincke DM, Schmid BF, Schubert P, Selz D, Stanoevska-Slabeva, K. The NetAcademy – a new concept for online publishing and knowledge management. In: Margaria T, Steffen B, Rückert R, Posegga J, eds. *Services and Visualization, Towards User-Friendly Design. Int. Workshop on Advanced Communication Services (ACoS'98).* Berlin/Heidelberg: Springer-Verlag, 1998: 29-43.

Address of the authors:
Dr. R. Grütter,
Institute for Media
and Communications Management,
University of St. Gallen,
Müller-Friedberg-Strasse 8,
CH-9000 St. Gallen,
Switzerland
E-mail: rolf.gruetter@unisg.ch