Medical Informatics Europe 2002
G Surján et al. (eds)
IOS Press 2002

Design considerations for an Ontology in the domain of Organ Failure and Transplantation

Christian JACQUELINET [a], Anita BURGUN [b], Denis DELAMARRE [b], Nigel STRANG [a], Bernard BOUTIN [a], Pierre LE BEUX [b]

a Etablissement français des Greffes, 5, rue Lacuée, F-75012 Paris, France
b Laboratoire d'informatique Médicale, Faculté de Médecine, Rennes, France
christian.jacquelinet@univ-rennes1.fr

Abstract. The Etablissement français des Greffes (EfG) is a national agency dealing with Public Health issues related to organ, tissue and cell transplantation in France. The evaluation of organ retrieval and transplantation activities, one of its missions, is supported by a national information system (IS). In order to facilitate data recording, to improve the quality of information and to prepare semantic interoperability with other information systems, the existing thesaurus of the EfG was audited, leading to the design a new terminological module devoted to the support of the domain ontology.

1. Introduction

Transplant teams record patient data in the EfG-IS. This work is redundant with data recording in their hospital information system and other local, national or international registries. End-stage diseases requiring an organ transplantation may also be treated with alternative methods, such as mechanical heart assistant devices, dialysis and even medical treatment. The evaluation of transplantation results should not be disconnected from the evaluation of other therapeutic methods. Such an objective implies to organize the cooperation between a set of production databases and an exploitation-oriented data-warehouse [1]. This approach is under construction for renal insufficiency at the EfG, within the Renal Epidemiology and Information Network (REIN) project [2]. The evolution of the EfG-IS towards Electronic Data Interchange (EDI) is crucial, and requires semantic interoperability defined as the capability of IS to use the information that they exchange [3-5]. A preliminary audit showed that the existing thesaurus was split into a set of unconnected tables of values related to: (i) the registration of patient on the waiting list (initial diseases), (ii) the follow-up before and after the transplantation (complications, causes of death). The thesaurus consisted of a catalogue without hierarchy, differentiation principles or compositional rules. Terms themselves do not conform to medical standard terminologies. Only the terms for kidney diseases had been inspired from those of the ERA-EDTA register [6]. Some terms are inappropriate and are not used as demonstrated by a quantitative analysis of patients records. The qualitative analysis revealed that 25% of the terms show ambiguity, incompleteness, implicitness or inconsistency. Duplicates and overlaps were noticed. The granularity was ranging from a dozen terms for heart diseases to more than 60 detailed terms for renal diseases. This preliminary study prompt us to organise a medical terminology for organ failure and transplantation on sound ontological foundations. This paper reports the methodological approach, design considerations and some results that led to specify the requirements of a new terminological module for the EfG-IS and for the REIN-IS.

2. Materials and Methods

2.1. Terminological resources and Knowledge Representational Model

Selected terms were coming from : (i) international registries for transplantation, such as the International Society for Heart and Lung Transplantation registry, the European Liver Transplant Registry, the European Renal Association and European Dialysis and Transplantation registry; (ii) Medical Standards such as ICD-10, UMLS concepts and relations linked to transplantation; (iii) French accredited specialized thesauri such as thesauri of the French Société de Néphrologie or of the Société Nationale Francophone de Gastro-entérologie. The Conceptual Graphs (CG) was selected as underlying knowledge representational formalism [7-9] because it offers facilities for term description and existing operations on graphs. Its flexibility allows the comparison and improvement of various description structures.

2.2. Production of Diseases description frames by automated knowledge extraction from a test bed of medical terms

A knowledge extraction tool called RIBOSOME that parses short texts and medical terms into CGs was used to perform semantic analysis of the terms [10]. A test bed of medical terms used to describe diseases related to initial disease leading to the registration on the waiting list (10 items), complications (5 items) and causes of death (10 items) was selected. It was decided to restrict the possibilities offered by the CG model for diseases and terms description to a structure as simple as possible, so that it can be usable by clinicians and embedded in the EfG existing information system. In order to represent objects in a simple and stereotyped manner, it was decided to describe terms with CG based frame like structures using a limited set of slots. One interesting characteristics of such structures is to formulate what is expected, so that implicit information masked in some terms can be more easily detected. Frames also provide an easily readable surface representation guiding the description of terms and their validation by an clinician expert.

3. Results

3.1. Semantic Analysis of terms with RIBOSOME

The semantic analysis of a medical term with RIBOSOME is illustrated in figure 1. The input is firstly segmented into valid lexical entries: word, locution or entire label of a disease. Words composing locutions and disease labels are analysed at the same time. The next step is the contextual selection of conceptual structures, triggered by the lexical input and their combination in more complex primary conceptual structures. A set of post-processing functions is available: among them, one joins sub-graphs of the primary output into a frame like format close to the standard conceptual graph linear format (CGLF); another turns the description frames into a 'pretty ' format where conceptual relation labels are explicit, so that medical experts working with our group easily understand the graphs. CG's generated for medical terms support the semantic structure of the terms. Conceptual structures are inserted into a semantic network that is the support of the domain ontology and that permits the re-use of the acquired semantic knowledge in next analysis. The comparison of the semantic structures generated for the test bed allowed us to propose a general schema for the description of a disease whose core is given the next section.

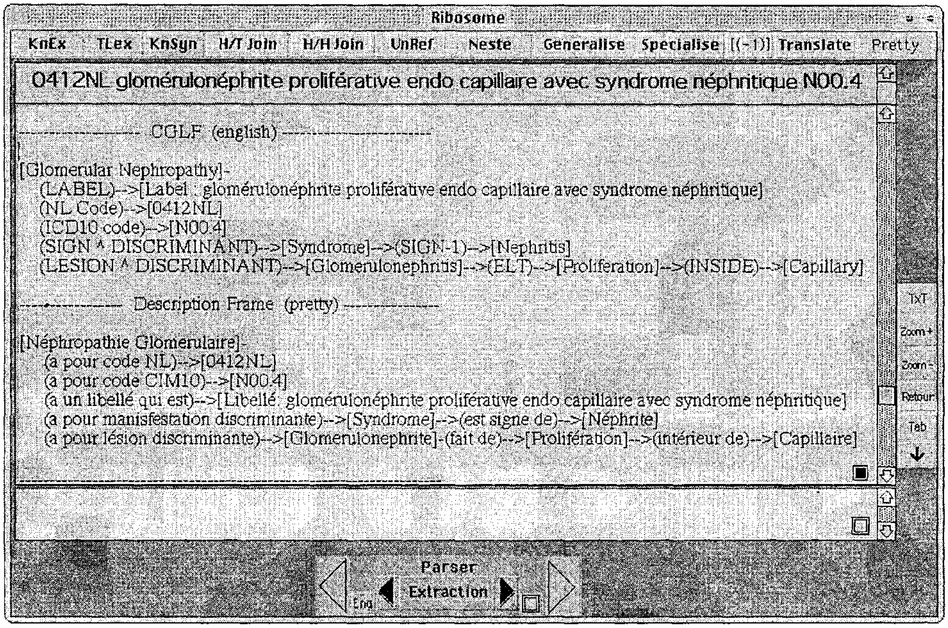

Figure 1: Knowledge Extraction with RIBOSOME

3.2. Terms and Diseases Description frames

One prominent result is that the formulation of terms can be specified according to a limited set of nosological discriminating slots such as etiology, semiology, pathology, evolution and associated diseases. Generic terms, as well as ambiguous or implicit terms, appeared as unspecified terms according to some slots. A description frame supports the type definition of a term: the head concept is the genus whereas discriminating slots act as differentiae. Not all clinical signs that can be encountered in a disease are to describe, nor all lesions, associated diseases or etiological processes, only those that constitute a sufficient condition. The description frame [Disease] is used to define a disease responding to a given term. It embeds others description frames such as Etiological process or Lesion.

[Disease]-
(label)→[Term]
(code)→[Code]
(discriminating etiological process)→[Etiological Process]
(discriminating clinical finding)→[Finding]
(discriminating pathological lesion)→[Lesion]
(discriminating associated disease)→[Disease]

[Etiological Process]-
(label)→[Term']
(discriminating agent)→[Agent]
(discriminating patient)→[Patient]

[Lesion]-
(label)→[Term'']
(discriminating location)→[Anatomical Component]
(discriminating evolution characteristic)→[Evolution]

A description frame can also be specialized if its specialization is often reused: [Liver Disease], [Renal Disease] for example.

3.3. Semantic Integration, Hierarchy and Granularity

One interesting rule related to conceptual refining is that a partially specified graph **G1** subsumes a more specified graph **G2**:
let **G1**: [C1]→(RC1)→[C2], G2: [C1]→(RC1)→[C3], if C2 > C3 then **G1** > **G2**.
As a corollary, because [C] > [C]→(RC)→[C'], a "more detailed" conceptual graph **G3**: [C1]→(RC1)→[C3]→(RC2)→[C4] is such that **G3** < **G2** < **G1**.
In view of interoperability, these rules find useful applications for the integration of medical terms coming from different terminologies. They also provide an efficient tool for the organisation of the hierarchy on formal cognitive foundations. As an example, let us consider the following description frames for Wilson's disease related Hepatopathies in two clinical presentations:

[Liver Disease]-	[Liver Disease]-
(label)→[**Wilson's disease Fulminant Hepatitis**]	(label)→[**Wilson's disease Chronic Hepatitis**]
(code)→[EfG#x]	(code)→[EfG#y]
(discriminating etiological process)→[Unspecified]	(discriminating etiological process)→[Unspecified]
(discriminating clinical finding)→[Unspecified]	(discriminating clinical finding)→[Unspecified]
(discriminating pathological lesion)→[**Necrosis**]-	(discriminating pathological lesion)→[**Hepatitis**]-
(discriminating evolution)→[**Hyper-Acuteness**]	(discriminating evolution)→[**Chronicity**]
(discriminating location)→[Liver]	(discriminating location)→[Liver]
(discriminating associated disease)→[**Wilson Disease**]	(discriminating associated disease)→[**Wilson's Disease**]

Both graphs have a common supertype that denotes Wilson's disease Hepatopathy with less specified lesions. With the same approach, the partial specification of the discriminating associated disease leads to the description frame of all metabolic liver diseases.

[Liver Disease]	[Liver Disease]-
(label)→[**Wilson's disease Hepatopathy**]	(label)→[**Metabolic Liver Disease**]
(code)→[EfG#z]	(code)→[EfG#w]
(discriminating etiological process)→[Unspecified]	(discriminating etiological process)→[Unspecified]
(discriminating clinical finding)→[Unspecified]	(discriminating clinical finding)→[Unspecified]
(discriminating pathological lesion)→[**Lesion**]-	(discriminating pathological lesion)→[Lesion]-
(discriminating evolution)→[**Unspecified**]	(discriminating evolution)→[Unspecified]
(discriminating location)→[Liver]	(discriminating location)→[Liver]
(discriminating associated disease)→[**Wilson's Disease**]	(discriminating associated disease)→[**Metabolic Disease**]

The same approach defines Wilson's disease as a subtype of metabolic disease:

[Metabolic Disease]-	[Disease]-
(label)→[**Wilson's disease**]	(label)→[**Metabolic disease**]
(code)→[EfG#k]	(code)→[EfG#j]
(discriminating etiological process)→[Metabolic Disorder]-	(discriminating etiological process)→[**Metabolic Disorder**]-
(discriminating patient)→[**Copper**]	(discriminating patient)→[**Unspecified**]
(discriminating clinical finding)→[Unspecified]	(discriminating clinical finding)→[Unspecified]
(discriminating pathological lesion)→[Unspecified]	(discriminating pathological lesion)→[Unspecified]
(discriminating associated disease)→[Unspecified]	(discriminating associated disease)→[Unspecified]

4. Discussion - Conclusion

This work shows that it is possible to model diseases using CG based description frames with a limited and stereotyped set of discriminating nosological slots. Our approach enables the integration of subsets of relevant medical terms coming from various terminologies in a constrained semantic network devoted to support the ontology of the domain of organ failure and transplantation, in a manner similar to the MAOUSSC project [11]. According to the typology proposed by [12], we are building a terminological system with two classical goals: standardization and communication [13-16]. The use of RIBOSOME as a knowledge extraction tool appeared as a important help to compare description frames and to build a terminological system for organ failure and transplantation on sound ontological foundations [10]. The implementation of a new terminological system is now in progress. It provides enough flexibility for the creation of new concepts, new relations and description frames so that the core model we have defined herein is extensible and demonstrates its robustness with a larger amount of terms.

Acknowledgments

We are in debt to the clinical experts who helped us in the validation of the semantic analysis validation process: A. Bisson, G Bobrie, K Boudjema, P Landais, M Stern.

References

[1] Sheth A, Larson J. Federated Systems for Managing Distributed, Heteregeneous and Autonomous Databases, ACM Computing Surveys, 22 (1990) 183-235

[2] Stengel B, Landais P. Data Collection about the case management of end-stage renal insufficiency. Nephrologie, 20 (1999) 29-40.

[3] IEEE Standard Computer Dictionary : A compilation of IEEE Standard Computer Glossaries, IEEE (1990)

[4] Hammond WE. The role of standards in creating a health information infrastructure. Int J Biomed Comput. 34 (1994) 29-44.

[5] Dudeck J. Aspects of implementing and harmonising healthcare communication standards. Int J Med Inf. 48 (1998) 163-71

[6] Mallick NP, Jones E, Selwood N. The European (European Dialysis and Transplantation Association-European Renal Association) Registry. Am J Kidney Dis. 1995 Jan;25(1):176-87.

[7] Sowa J.F. Conceptual Structures: Information Processing in mind and machine. Addison-Wesley, London, 1984.

[8] Sowa J.F. Conceptual Graph Summary , in Conceptual Structures: current research and practice ; Nagle, Nagle, Gerholz & Eklund Ed.; Ellis Horwood Workshops;1992.

[9] Volot F, Joubert M, Fieschi M. Review of biomedical knowledge and data representation with Conceptual graphs. Meth Inf Med. 37 (1998) 86-96.

[10] Jacquelinet C, Burgun A. Building the ontological foundations of a terminology from natural language to conceptual graphs with Ribosome, a knowledge extraction tool. ICCS proceedings. In Working with Conceptual Structures. G Stumme Ed, Shaker Verlag, Aachen 2000.

[11] Burgun A, Bodenreider O, Denier P, Delamarre D, Botti G, Oberlin P, Leveque JM, Bremond M, Fieschi M, Le Beux P. A collaborative approach to building a terminology for medical procedures using a Web-based application: from specifications to daily use. Medinfo. 1998; 9 Pt 1:596-9.

[12] De Keizer NF, Abu-Anna A, Zwetsloot-Shonk JHM. Understanding Terminological Systems I: Terminology and Typology. Meth Inform Med. 2000: 39; 16-21

[13] De Keizer NF, Abu-Anna A. Understanding Terminological Systems II: Experience with Conceptual and Formal Representation of Structure. Meth Inform Med. 2000: 39; 22-9

[14] Rossi Mori A et al. Exploiting the terminological approach from CEN/TC251 and GALEN to support semantic interoperability of healthcare record systems. Int J Med Inf. 48 (1998) 111-24.

[15] Mc Donald CJ et al. What is done, what is needed and what is realistic to expect from medical informatics standards. Int J Med Inf. 48 (1998) 5-12.

[16] Spackman K, Campbell KE, Compositional Concept Representation Using SNOMED: toward further convergence of clinical terminologies. In Proc AMIA symp 1998: 740-44.