Medical Informatics Europe '99
P. Kokol et al. (Eds.)
IOS Press, 1999

Galen : a third generation terminology tool to support a multipurpose national coding system for surgical procedures

B Trombert-Paviot*, JM Rodrigues*, JE Rogers**, R Baud***, E van der Haring****, AM Rassinoux***, V Abrial*, L Clavel*, H Idir*
**Department of Public Health and Medical Informatics Fac de Médecine University of Saint Etienne France*
***Medical Informatics Group Victoria University of Manchester United Kingdom*
****Department of Medical Informatics University Hospital of Geneva Switzerland*
*****Department of medical Informatics Catholic University of Nijmegen Netherlands*

Abstract. GALEN has developed a new generation of terminology tools based on a language independent concept reference model using a compositional formalism allowing computer processing and multiple reuses. During the 4th framework program project Galen-In-Use we applied the modelling and the tools to the development of a new multipurpose coding system for surgical procedures (CCAM) in France. On one hand we contributed to a language independent knowledge repository for multicultural Europe. On the other hand we support the traditional process for creating a new coding system in medicine which is very much labour consuming by artificial intelligence tools using a medically oriented recursive ontology and natural language processing. We used an integrated software named CLAW to process French professional medical language rubrics produced by the national colleges of surgeons into intermediate dissections and to the Grail reference ontology model representation. From this language independent concept model representation on one hand we generate controlled French natural language to support the finalisation of the linguistic labels in relation with the meanings of the conceptual system structure. On the other hand the classification manager of third generation proves to be very powerful to retrieve the initial professional rubrics with different categories of concepts within a semantic network.

Keywords. knowledge representation, knowledge acquisition, classification, natural language processing

1. Introduction

The Galen artificial intelligence technology has been used within the Galen-In-Use (HC 1018) Health Care Telematics project of the 4 Th. R § D framework program to support the development of a new French coding system named CCAM. The work is going on within a contract between the French department of health and GOL (Galen Organisation Limited). We have presented in [1] the initial works on the increased linguistic coherence for urology rubrics by using Galen tools and techniques and natural language processing. We stressed here the power of the classification tool based on the language independent common

reference model as an example of a practical support given by the third generation terminology Galen.

2. Materials and methods

◆Materials

Galen stands for Generalised Architecture for Languages, Encyclopaedias and Nomenclatures in Medicine. The project has developed a scheme in which all and only sensible medical concepts are represented and which is accessible and processed by computers. This scheme is named CORE (for COncept REference model) [2]. It is based on an ontology (high level schemata) to model the categories within the universe of discourse and to classify them automatically being language independent and on a kernel of general medical knowledge. It is application independent and reusable for it uses a compositional formalism named GRAIL (for Galen Representation and Integration Language) [3]. The second outcome was the development of practical tools and techniques (the terminology server) for managing healthcare terminology including natural language processing. Galen-In-Use has been the application phase of Galen among nine national classification centres from Belgium, Finland, Germany, Greece, Italy, Sweden, United Kingdom, The Netherlands and France.
The French demonstrator is applying Galen tools to the development of a new coding system for surgical procedures initiated in 1994, named CCAM (for Classification Commune des Actes Médicaux) [4]. It is using an integrated set of tools named CLAW [5] for Classification workbench and ALN (for Atelier Langage Naturel) to generate controlled French natural language sentences [6]. It plans to merge two existing systems for billing the physician fees and for coding hospital discharge abstract in one only multipurpose coding system. The program is managed by a joint effort from the ministry of health and the national agency of the main mandatory health insurance system. CCAM is planned as part of a broader CCAP (Classification Commune des Actes Professionnels) for healthcare professionals. The initial work concerning surgery is planned for 15 parts of 500 rubrics each : 7 500 rubrics including combined interventions. To day 11 parts have reached step 3 and 6 step 4 with 3 000 rubrics. The final product is planned for 2001 for surgery and 2002 for an additional 15 parts of medical interventions.

◆Methods

There are four stages as described in [1] : an intermediate dissection, the validation of the compliance with CEN /ENV 1828, the representation with the Grail CORE model and the outputs.
Once the formal knowledge representation of each rubric of CCAM is built, there are a lot of opportunities for the outputs. First there is a strong need to come back to natural language expressions in French language for quality assurance : error detection, disambiguation or inconsistency [7] and to assess the process. The second opportunity is automatic indexing for retrieval and dynamic reclassification for cross reference to restructure an apparently first generation terminology (a flat list) in a de facto compositional (multi axial) second generation terminology [8] and with an unlimited range of possibilities. We are presenting some examples in the following part and more for the oral presentation.

3. Results

Overview

Tab 1 : Distribution of entities by classes by levels

Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Total
0.4 surgery in order to achieve repair	95	11	16	29	15	166
0.6 accompanying	8	44	105	6		163
0.7 closure	2	40	40	5	1	88
0.8 creation	32	7	16	7	4	66
0.9 destruction	2	28	7			37
0.10 cauterizing	0					0
0.11 fastening	5	36	32	8	2	83
0.12 freeing	14	11	1			26
0.13 tension	1					1
0.14 decompression	16	1				17
0.15 compression	1					1
0.16 installation	12	153	55	18	2	240
0.17 opening	9	120	66	16	4	215
0.18 manipulation	8	5	1	1		15
0.19 movement	29	26	13	1		69
0.20 removal	17	651	326	75	18	1087
Total	251	1133	678	166	46	2274

We are presenting the results for 3 clinical domains, Neurosurgery with 436 rubrics, Ear Nose and Throat surgery with 695 rubrics, oral surgery with 409 rubrics and a total of 1540 rubrics. They were mapped to 1926 different terminological phrases for some initial CCAM rubrics must be represented by more than a only dissection for ontology reason when for instance they describe different incompatible approaches within the same rubrics. Further on these 1926 different terminological phrases are classified in 2274 entities for the same sentence can have more than one parent and be classified in different classes. The granularity is summarised in Tab 1 where the 6 levels are presented starting from the 16 initial classes at level 1.

At level 1 they are 20 possible classes but only 16 are populated by our sample coming from only 3 clinical domains of surgical procedures. The classification is based first on the main deed, then on different concepts depending upon the main classes. There are 6 hierarchical levels clustering the 2274 entities. The main class (level 1) "removal" groups 47.8% of the entities with the most common surgical deed excising. The level 3 is the most frequent (49,8%) and the level 4 (29.81%) is more frequent than the level 2 (11.03%).

At level 2 the classes are based on different concepts. As an example :

- the deed, for class 0.11 "fastening" which is divided in 0.11.0 immobilization, 0.11.1 fixation, 0.11.2 fusion, 0.11.3 stabilization, 0.11.4 binding
- the object on which the deed is done e.g for 0.18 "manipulation".

Automatic indexing and multiple parents (Tab 2)

The same entity can be automatically retrieved several times depending upon the position of the concept in the hierarchy of the typology and its relation with other concepts in anatomy, device or approach. This multiple criteria retrieval is very useful to cluster the rubrics by their meaning proximity. One example for "alignment" shows the power of the composite knowledge representation to generate automatically the meaning connections between elementary entities.

Tab 2 : multiple indexing

0.4.0 alignment in order to achieve repair
0.4.0 alignment in order to achieve repair
0.4.0.0 reduction
0.4.0.0.0 reduction on fracture
0.4.0.0.0.5 reduction on fracture involves of Orbit & Skull
0.4.0.0.0.15 reduction on fracture of zygoma Transoral Approach
0.4.0.0.0.15.0 reduct. on fract. of zygoma, Transoral Approach with Incising of face
0.4.0.0.0.16 reduction on fracture of zygoma with Incising of Face
0.4.0.0.0.15.0 reduct. on fract. of zygoma, Transoral Approach with Incising of face
0.4.0.0.2 open reduction on bone of nose
0.4.0.0.3 closed reduction during postpartum period on nose
0.4.0.0.4 closed reduction on bone of nose
0.19 movement
0.19.9 alignment
0.4.0 alignment in order to achieve repair
0.19.9.1 alignment on segment of nose
0.19.9.2 primary bilateral alignment cartilagoalarismajor
0.19.9.3 primary unilateral alignment cartilagoalarismajor
0.19.9.4 alignment on orbit
0.4.0.0.0.5 reduction on fracture involves of Orbit & Skull
0.4.0.0.0.15 reduction on fracture of zygoma TransOralApproach
0.4.0.0.0.16 reduction on fracture of zygoma with Incising of Face
0.19.26 movement on nose
0.4.0.0.3 closed reduction during postpartum period on nose
0.19.9.1 alignment on segment of nose
0.4.0.0.2 open reduction on bone of nose
0.4.0.0.4 closed reduction on bone of nose

0.4.0 , 0.4.0.0.3 ,0.4.0.0.2 , 0.4.0.0.4 , 0.4.0.0.0.5 , 0.4.0.0.0.15 , 0.4.0.0.0.16 are classified within the level 1 class 0.4 "alignment to achieve repair" at levels ranging from 2 to 5 and reclassified within the level 1 class 0.19 "movement" at levels ranging from 3 to 4. This automatic indexing is very useful to show that the flat list first generation like CCAM coding system has a multiple parents structure which cannot be managed without computer processing.

Reclassification and multiple hierarchy

Another important opportunity with galen tools is to reclassify the coding system depending upon the different concepts categories which are represented with the core model. For instance within the 3 clinical domains there are 16 entities on oesophagus or 67 entities by endoscopy.

4. Discussion

The utilisation of a third generation terminology tool [8] like Galen for surgical procedure coding systems development must not be misinterpreted. Galen is not the new generation terminology to replace the existing coding systems in different countries but a set of tools and technique to support and to ease such developments at least on 3 points.

- to increase the coherence and the consistency of the linguistic (even in a minority language as French) expressions of the lists of rubrics (first generation terminology) which are still needed for some very important purposes (like the use for billing for CCAM). We have demonstrated in [1] the usefulness of generating controlled

vocabulary based on all the sensible and only the sensible knowledge to increase the quality of such first generation terminology which will be kept in use.
- to allow the creation of a second generation terminology starting from the apparent flat list of rubrics produced by the traditional consensus method by the multiple concepts categories representation allowing reclassification by different concepts hierarchies and cross references.
- to open the way to third generation terminology by automatic indexing based on the semantic network and the multiple parents structure of the GALEN concept model .The availability of all the acquired knowledge on surgical procedures and later on other clinical domains on a computer accessible form is a major enabling factor to shift from the utilisation by national or regional coding centres to the broader utilisation within the electronic healthcare record.

5. Conclusion

All the recently developed coding systems for surgical procedures share some common features for the process which can be named a consensus method. The clinical input and the many reviews of the linguistic expressions are scattered among clinical consultants, coders and health care recorders. The CCAM French process has introduced and added the use of artificial intelligence tools allowing to transform an apparent first generation terminology in a second generation one and opening the way to the advanced third generation use.
On the other hand the so called consensus method results in underestimated invisible costs and constant high cost maintenance. On the contrary the use of the third generation terminology tools have a clear defined cost and allow decreasing cost maintenance. The increasing strategic use of coding systems is a strong support to the extended utilisation of such tools.

Acknowledgements

To Galen-in-Use (HC 1018) project, funded by E.U Fourth R § D framework program .

References

[1] Rodrigues JM, Trombert-Paviot B, Baud R, Wagner J, Meusnier-Carriot F. Galen-In-Use :Using artificial intelligence terminology tools to improve the linguistic coherence of a national coding system for surgical procedures. In : Cesnik B and all, Eds. Proceedings of the ninth World Congress on medical informatics - Medinfo'98 ; Seoul; IOS Press. 1998, pp 623-627
[2] Rector AL Compositional Models of Medical Concept : Towards Re-Usable Application -Independent Medical Terminologies. In: P Barahona and JP Christensen (eds) Knowledge and Decisions in Health Telematics, IOS Press. 1994, pp 133-142.
[3] Rector AL, Bechofer S, Goble C, Horrocks I, Nowlan W, Solomon W. The GRAIL Concept Modelling Language for Medical Terminology.In: Medical Informatics Group, Department of Computer Science, University of Manchester, 1996.
[4] NCAM.Structure technique NCAPS Paris 02/01/1997.
[5] Zanstra P The Galen project. Classification manager. CBV Niews 1993 ; (4).
[6] Baud RH, Alpay L, Lovis C. Let's Meet the Users with Natural Language Understanding; In P Barahona and JP Christensen (eds) Knowledge and Decisions in Health Telematics, IOS Press. 1994, pp103-108.
[7] Rector AL, Baud R, Ceusters W, Claassen W, Rodrigues J-M, Rogers J, Rossi Mori A, van der Haring E, Solomon WD, Zanstra P. A Comprehensive Approach to Developing and Integrating Multilingual Classifications: GALEN's Classification Workbench. Theater-Style Demonstration, AMIA 1998 Annual Symposium (AMIA'98). Lake Buena Vista, FL, November 9, 1998.
[8] Rossi Mori A. Towards a new generation of terminologies and coding systems..Proceeding MIE 96 IOS Press Copenhaguen 1996