Galen-In-Use : Using artificial intelligence terminology tools to improve the linguistic coherence of a national coding system for surgical procedures

Jean Marie Rodrigues[a], Béatrice Trombert-Paviot[a], Robert Baud[b], Judith Wagner[b], Florence Meusnier-Carriot[a]

[a]*Dpt of Public Health and Medical Informatics University of Saint Etienne Jean Monnet France*
[b]*Dpt of Medical Informatics University Hospital of Geneva Switzerland*

Abstract

GALEN has developed a language independent common reference model based on a medically oriented ontology and practical tools and techniques for managing healthcare terminology including natural language processing. GALEN-IN-USE is the current phase which applied the modelling and the tools to the development or the updating of coding systems for surgical procedures in different national coding centres co-operating within the European Federation of Coding Centre (EFCC) to create a language independent knowledge repository for multicultural Europe.
We used an integrated set of artificial intelligence terminology tools named CLAssification Manager workbench to process French professional medical language rubrics into intermediate dissections and to the Grail reference ontology model representation. From this language independent concept model representation we generate controlled French natural language. The French national coding centre is then able to retrieve the initial professional rubrics with different categories of concepts, to compare the professional language proposed by expert clinicians to the French generated controlled vocabulary and to finalise the linguistic labels of the coding system in relation with the meanings of the conceptual system structure.

Keywords

Natural language processing, coding system, knowledge representation, artificial intelligence.

Introduction

The Galen artificial intelligence technology is used within the Galen-In-Use (HC 1018) current Health Care Telematics project of the 4 Th. R § D framework program to support the development of a new French coding system named NCAM. We present the methodology used and the results for urology rubrics concerning the increased linguistic coherence of this under development coding system.

Materials and methods

Materials

Galen

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 manipulable by computers. This scheme is named CORE (for COncept REference model) [1] and uses a compositional formalism named GRAIL (for Galen Representation and Integration Language) [2] which consists of a hierarchy subsumption of elementary entities and a set of sanctioning statements connecting these entities. 6000 primitive or composite categories were present in 1996 and 8000 at the beginning of 1997. Galen has developed an integrated set of tools named CLAssification Manager workbench to represent the knowledge of each rubric under the form of an intermediate semantic model (descriptors and links [3]) compliant with CEN ENV 1828 [4] and with the extended Galen model for surgical procedures GASP [5] and to map the intermediate dissection to the CORE knowledge representation [6]. This set allows as well retrieval based on the different concepts and development and management of the classifications linked to the CORE [7]. ALN (for Atelier Langage Naturel) is the natural language tool used to generate natural language sentences [8] from the Grail representation model.

Ncam

NCAM stands for Nomenclature Commune des Actes Médicaux. It is presently the provisional eponym for the new coding system for surgical procedures process initiated in 1994. The initial NCAM process was a traditional work based on the updating of flat list of existing labels coming from CDAM (for Catalogue des Actes Médicaux and used to code hospital discharge abstract) and validation by clinical specialists organisations named sociétés savantes. Among the

four national classification centres of Galen-In- Use (the current phase of Galen) from Italy, Sweden, The Netherlands and France the French demonstrator is applying Galen tools to the development of this new coding system as a step 4. The final product is the fifth stage.

NCAM is planned for 15 parts of 500 rubrics each : 8 000 rubrics for billing purposes including combined interventions. To day 7 parts have reached step 3 and 1 step 4. The final product is planned for 1998.

We are presenting the results concerning this step 4 on a provisional urology system version available in July 1996. It contains 522 labels. The labels of NCAM are short (50 to 250 letters or 3 to 30 words) for most of them but there are often several actions.

Methods

There are four stages :

Intermediate dissection

To enable dissemination of Galen utilisation the consortium developed an intermediate representation named intermediate dissection easy to use by end users clinician modellers. It uses descriptors and links available in the Galen knowledge base or generated by the modelling centres in coherence with the Galen ontology model.

Validation of the compliance with CEN / ENV 1828

We consolidated the previous stage by testing the compliance of the dissections with the 3 combinatorial rules of CEN / ENV 1828 [4].

Representation with the Grail CORE model

The Galen CORE model [1] aims to represent all and only sensible medical concepts which are believed to be shared across professional and national boundaries and not all medical knowledge. To perform the task automatic translation of the intermediate dissections into the Grail language of the CORE model is possible for the rubrics representation using descriptors and links already available in the Galen repository. On the other hand it was necessary to map the 771 new descriptors to the semantic network concepts of the CORE model.

Natural language generation

Once the formal knowledge representation of each rubric of NCAM built there are a lot of opportunities to retrieve, cross reference and restructure them but there is as well a strong need to come back to natural language expressions in French language to discuss with clinician experts and coders. There were two generations, one without any contraction which looks somehow awkward and one with contraction allowing to erase the concepts displayed below a specific concept.

Example of the Galen method :

- Source label

Résection endoscopique cervico-prostatique, avec urétrotomie à l'aveugle, urétrocystoscopie et mise en place d'une sonde urétrale.

- Intermediate dissection

RUBRIC "*Résection endoscopique cervico-prostatique, avec urétrotomie à l'aveugle, urétrocystoscopie et mise en place d'une sonde urétrale"*

PARAPHRASE " "

SOURCE "" CODE "U2.050"

MAIN resecting

ACTS_ON Anatomy: neck_of_urinary_bladder & prostate

BY_MEANS_OF endoscope

BY_TECHNIQUE inspecting

ACTS_ON Anatomy: urinary bladder & urethra

BY_MEANS_OF endoscope

WITH installing

ACTS_ON Device: catheter

HAS_DESTINATION Anatomy: urethra

WITH incising

ACTS_ON Anatomy: urethra

HAS_APPROACH closed approach

- Grail representation

[cl_SurgicalDeed,[rel_isMainlyCharacterisedBy([cl_performance,[rel_isEnactmentOf([cl_NLGCPT_Resecting,[rel_actsSpecificallyOn([cl_ArbitraryBodyConstruct,[rel_hasStructuralComponent(cl_UrinaryBladderNeck),rel_hasStructuralComponent(cl_ProstateGland)]]),rel_hasSpecificPhysicalMeans(cl_Endoscope),rel_hasSpecificSubprocess([cl_Inspecting,[rel_actsSpecificallyOn([cl_ArbitraryBodyConstruct,[rel_hasStructuralComponent(cl_UrinaryBladder),rel_hasStructuralComponent(cl_Urethra)]]),rel_hasSpecificPhysicalMeans(cl_Endoscope)]])]]]),rel_isCharacterisedBy([cl_performance,[rel_isEnactmentOf([cl_NLGCPT_InstallingProcess,[rel_LocativeAttribute([cl_Catheter,[rel_isActedOnSpecificallyBy([cl_Transport,[rel_hasSpecificConsequence([cl_Displacement,[rel_hasBetaConnection(cl_Urethra)]])]])]])]])]]),rel_isCharacterisedBy([cl_performance,[rel_isEnactmentOf([cl_NLGCPT_Incising,[rel_actsSpecificallyOn(cl_Urethra),rel_hasSpecificSubprocess([cl_SurgicalApproaching,[rel_hasSurgicalOpenClosedness([cl_SurgicalOpenClosedness,[rel_hasAbsoluteState(cl_surgicallyClosed)]])]])]])]])]]

- Natural language generation

FRENCH

French with_tc: *résection cervico-prostatique et endoscopique sous inspection urétro-vésicale et endoscopique*

avec mise en place d'un cathéter dans l'urètre et avec incision urétrale et fermée

French no_tc: *acte chirurgical de résection endoscopique d'une entité corporelle ayant pour partie le col de la vessie et ayant pour partie la prostate sous inspection endoscopique d'une entité corporelle ayant pour partie la vessie et ayant pour partie l'urètre avec mise en place d'un cathéter dans l'urètre et avec incision urétrale et fermée.*

ENGLISH

English with_tc: *endoscopic cervico-prostatic resection under control of endoscopic and urethro-vesical inspecting with installing process of a urethral catheter and with closed urethra incising*

English no_tc: *surgical deed of endoscopic resection of a body entity having as part the urinary bladder neck and having as part the prostate under control of endoscopic inspecting of a body entity having as part the urinary bladder and having as part the urethra with installing process of a urethral catheter and with closed urethra incising*

GERMAN

German with_tc: *Resektion des Blasenhalses und der Prostata durch Endoskop unter Inspektion der Harnblase und der Urethra durch Endoskop mit Legen eines Katheters in die Urethra und mit geschlossener Inzision der Urethra*

German no_tc: *chirurgische Behandlung Resektion des Blasenhalses und der Prostata durch Endoskop unter Inspektion der Harnblase und der Urethra durch Endoskop mit Legen eines Katheters in die Urethra und mit geschlossener Inzision der Urethra*

Results

We present two levels of results :the knowledge representation with the Galen CORE model and the natural language generation with the ALN tool.

Knowledge representation

overall evaluation

We have been able to represent 505 rubrics (97 %) to the 522.

For 17 (3 %) it was not possible to represent them for there was no main deed : for instance, *«traitement chirurgical d'une torsion du testicule ou de ses annexes avec fixation du testicule controlatéral» or «Cryothérapie de la prostate»*.

Or it was not possible to understand the anatomic location : for instance, *«drainage d'un lymphocèle par ponction percutanée»*.

These rubrics were not compliant with the combinatorial rules of CEN / ENV 1828.

Among the 505 representations 102 (20 %) needed additional knowledge to be modelled .

«Alignement endoscopique d'une rupture traumatique de l'urètre» means repositioning the two parts of urethra separated by a trauma using an endoscope and not repositioning the pathologic process of breaking (rupture) the urethra which is the consequence of a trauma which is not allowed by the Galen CORE model

771 additional descriptors were necessary for urology system rubrics.

surgical deed

The 505 urology rubrics are mapped to 42 different surgical deeds (marked with *) which are at different levels of the Grail CORE typology :

Arterialising *
CardiacSurgery
ClosedSurgicalDeed
EndoscopicProcedure
Microsurgery
OpenSurgicalDeed
SurgicalAccompanyingProcess
SurgicalAdaptingProcess
SurgicalBypassing
SurgicalClosingProcess *
SurgicalCompressing
SurgicalCreatingProcess *
    SurgicalConstructingProcess
        SurgicalConstructingProcess *
            Constructing
        SurgicalFistulaFormingProcess *
            SurgicalStomaFormingProcess *
SurgicalDecompressing
SurgicalDestroyingProcess..*
    FragmentingUSE *
    SurgicalLysing *
SurgicalDimensionDecreasingProcess
SurgicalDimensionIncreasingProcess
    SurgicalLengthIncreasing *
        DilatatingProcess *
SurgicalExaminationProcedure
SurgicalFasteningProcess
    SurgicalBinding
    SurgicalFixing *
    SurgicalConnectionProcess *
        SurgicalAnastomosisProcess *
        SurgicalFusing
        SurgicalShuntingProcess *
SurgicalFreeingProcess *
    DisconnectionProcess *
SurgicalInstallingProcess *
    SurgicalImplanting *
        SurgicalReimplanting *
    SurgicalGrafting *
        SurgicalTransplanting *
    SurgicalInserting *
    SurgicalInjecting *
    SurgicalReplacing *
SurgicalManipulating

SurgicalMoving *
SurgicalOpeningProcess *
    Marsupialising *
    Canalising *
    SurgicalBurring
    SurgicalDividing *
    SurgicalIncising *
    SurgicalPuncturing *
SurgicalPulling
SurgicalRemovingProcess *
    SurgicalResecting *
    SurgicalAbrading .
    SurgicalBiopsyingProcess *
    SurgicalEnucleating
    SurgicalEvacuating
    SurgicalExcising *
        Harvesting *
    SurgicalStripping
    SurgicalDraining *
    SurgicalExtracting *
SurgicalRepairProcess
    SurgicalFashioning *
SurgicalReshapingProcess
    SurgicalStraightening *
SurgicalRevising
SurgicalTensionForceApplyingProcess
    Suturing *

The same rubric 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 multicriteria retrieval is very useful to cluster the rubrics by their meaning proximity .

Anatomy

The same mapping can be done for any separated or related concept categories. This example for anatomy shows the power of the composite knowledge representation to generate automatically the meaning connections between elementary entities :

U0.151 *Pelvectomie antérieure avec urétérostomie cutanée transintestinale détubulée continente, y compris le curage ganglionnaire*

Internal Body Structure
    Internal Organe
        Urinary Tract Body Part
            Ureter
            Urinary bladder
        Genital Tract Body Part
            Female Genital Tract Body Part
                Uterus
                Vagin
Digestive System Component
    Gastrointestinal Tract
        Intestine
Solid Body Structure
    Lymphnode
Surface Region
    Skin Covering
Intrinsically Normal Body Structure
    Generic Body Internal Structure
        Lymphnode
Hollow Body Structure
    Tubular Body Part
        Ureter
Body Space
    Body Hole
        Pelvic cavity

Natural language generation

For the 505 rubrics the ALN tool was able to automatically generate the natural language expressions in less than one month. It was necessary nevertheless to finalise by hand the edition to smooth some linguistic awkwardness as singular / plural, conjunction and determiner.

Six months later a new urology version was produced. 176 rubrics were modelled, represented and generated in a shorter period of time : comparing the content with the previous version was very fast.

The outcome is a French controlled vocabulary which is more lengthy that the initial rubrics and is describing explicitly the surgical procedure. These «controlled» natural language expressions make easier for end users within classification centres the understanding than with the professional clinical language used by the expert clinicians. Examples of the improvement of the quality of the rubrics are given depending upon the introduction of additional knowledge.

Without additional knowledge

e.g. initial rubric : *Correction de reflux vésico-urétéral unilatéral, par voie endoscopique, chez l'homme, quelle que soit la substance injectée.*

generated label : *Injection unilatérale de substance dans la vessie sous endoscope pour normalisation de reflux vésico-urétéral chez l'homme.*

Here the deed is explicitly injecting TO correct or TO cure a backward flow regurgitation.

With additional knowledge

- On surgical deed :

e.g. initial rubric : *Lithotritie extracorporelle d'une lithiase unique pyélique ou calicielle inférieure, avec repérage échographique et drainage urétéral.*

generated label : *extraction d'un calcul unique du bassinet ou du calice inférieur avec fragmentation par un lithotripteur extra-corporel et avec imagerie par échographie pour repérage du calcul et avec évacuation du calcul par l'uretère.*

Lithotrypsy is replaced by extracting a lithiasis with 3 sub processes : imaging, fragmenting with a lithotriptor device and draining.

- On anatomy :

e.g. initial rubric : *Transplantation rénale.*

generated label : *réimplantation sur un receveur d'un rein prélevé sur un donneur.*

There is replacement of an overall procedure by the precise one knowing that it does not include harvesting of the donor kidney.

Discussion

All the recently developed coding systems for surgical procedures share some common features for the definition of labels or rubrics. The initial clinical input and the many reviews of the linguistic expressions are based on pragmatic professional language expertise from clinical consultants, coders and health care recorders. They used structured organisation for decisions and precise methods to deliver the coding structure. The French process is following the same main track but added a new step named step 4 where artificial intelligence terminology tools are used to represent the clinical knowledge with an ontology model and to compare the usual professional medical language with a controlled French vocabulary based on the conceptual model.

We have presented this precise step of the process for it uses language independent ontology tools as well as natural language processing and can be applied to any coding system of surgical procedures in at least 5 different languages to day : Dutch, English, French, German and Italian and more in the near future.

Two quality levels have been shown :

1. The coherence of linguistic clinical expressions with shared knowledge representation of sensible and only sensible medical concepts.

2. The distance between professional medical jargon and explicit and controlled vocabulary expression of semantic meanings.

Conclusion

This work is a practical example of the usefulness of artificial intelligence terminology tools developed by the European Galen consortium to increase the quality of the linguistic expressions for a minority national language coding system.

Artificial intelligence terminology products are much easier to be used by end users when they produce natural language outputs.

The increasing strategic use of coding systems is a strong support to the extended utilisation of Galen tools by other classification centres in the world to insure coherence with a shared ontology and advanced quality multi-lingual products.

Acknowledgements

The GALEN-IN-USE consortium includes University of Manchester (Co-ordinator), VAMP Health Ltd, UK ; National Research Council, Italy ; University of Saint-Etienne, France ; University of Nijmegen, Dutch National Classification Centre (WCC), Netherlands ; European Federation of Classification Centres ; Hôpital Universitaire de Genève, LNAT Associates, Switzerland ; Technical Research Centre of Finland (VTT), Oulu University Hospital, Medici Data Ltd, National Research & Development Centre of Welfare & Health, Finland ; Swedish Institute for Health Services Development (SPRI), University of Linköping, Sweden ; Office Line Engineering RAMIT vzw, Datasoft, University of Louvain, Belgium ; GSF Medis Institut, University of Hildesheim, ID Gesellschaft für Information und Dokumentation in Gesundheitswesen mbH, Germany ; University of Athens, Greece ; IASIST, Barcelona, Spain. Hewlett Packard and the Bavarian State Government are sponsoring partners.

References

[1] 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 l33-142.

[2] 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.

[3] GALEN/VUM03/96 version 1.1. Links and Templates Summary.

[4] CEN ENV 1828 Health care Informatics. Structure for classification and coding of surgical procedures. Brussels December 1995.

[5] GALEN/VUM01/96 version1.1. Co-operative Modelling.

[6] Rector AL, Solomon WD Nowlan WA Rush TW. A Terminology Server for Medical Language and Medical Information Systems. In Proceedings of IMIA WG6, Geneva, May 1994.

[7] Zanstra P het GALEN project. Classification manager. CBV Niews 1993 : (4).

[8] 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, pp 103-108.

[9] Rodrigues JM, Trombert-Paviot B, Baud R, Wagner J, Rusch P, Meusnier F. Galen IN Use : An EU Project applied to the development of a new national coding system for surgical procedures : NCAM. Proceedings of Medical Informatics in Europe MIE 97, IOS Press pp 897-901.

Address for correspondence

J.M Rodrigues, CHU Saint-Etienne, F 42650 Saint Jean Bonnefonds,
FRANCE, rodrigues@univ-st-etienne.fr