Towards a rule-based support system for the coding of health conditions in the Patient Summary

Elena Cardillo[1], Maria Teresa Chiaravalloti[2], Claudio Eccher[3], Erika Pasceri[1], Vincenzo della Mea[4], Lucilla Frattura[5], Roberto Guarasci[6]

[1]Institute of Informatics and Telematics, Rende (CS), Italy
{elena.cardillo,erika.pasceri}@iit.cnr.it

[2]Institute for High Performance Computing and Networking, Rende (CS), Italy
chiaravalloti@icar.cnr.it

[3]Bruno Kessler Foundation, Trento, Italy
cleccher@fbk.eu

[4]Department of Mathematics and Computer Science, University of Udine, Italy
vincenzo.dellamea@uniud.it

[5]Central Health Directorate, Friuli Venezia Giulia Region, Udine, Italy
lucilla.frattura@regione.fvg.it

[6]Department of Languages and Education Sciences, University of Calabria, Rende (CS), Italy
roberto.guarasci@unical.it

Abstract. In the frame of federated and interoperable Electronic Health Records (EHRs), specific coding systems are mandatory for filling out healthcare documents such as the Patient Summary (PS). PS cannot be automatically generated from the patient's EHR data, because of the sensitivity of its content. For this reason it needs to be validated by a General Practitioner (GP), who is the sole responsible of this document. The literature shows that the practice of coding is recognized as a difficult task for GPs and it often generates coding errors and misspecifications of clinical data. To overcome this issue, a support system based on standardized and formalized coding rules for the domain of application is proposed, to facilitate a more accurate coding process without breaking the law.

Keywords: coding rules; patient summary; coding support systems; reference terminology; rule-based systems; ICD.

1 Introduction

In adopting the European Union (EU) directive on cross-border care and healthcare semantic interoperability, especially related to the Patient Summary (PS), most European Countries are regulating the coding systems use in the frame of federated and interoperable EHRs, making some of them mandatory for compiling healthcare documents. In compiling the PS, data related to health conditions cannot be automatically generated from those available in the GP's EHR, because the GP is fully responsible for its content and has to validate it. Nonetheless, since coding is proved to be a diffi-

cult task, an automated coding support system (CSS) can be of help without infringement of the law. The need for a centralized management of coding systems and processes by means of a rule-based supporting tool is motivated by a number of critical issues reported in the literature about the use of coding systems at different levels. In proposing a CSS, it is important to consider the harmonization and integration of medical terminologies used by domain experts to ensure information interoperability and the full understanding of the meaning conveyed, thus avoiding the proliferation of non-integrated and heterogeneous terminologies within EHRs. Furthermore, GPs massively use natural language to record health conditions [1], in particular comorbidities, thus generating unstructured and uncoded data, mainly because they do not know how to properly use coding systems and consider coding as an excessively time-consuming activity. This work proposes a methodology for the creation of a CSS that will be initially experimented for the Italian PS use case.

2 Related works

During the last twenty years much effort has been spent on the development of support systems for the semi-automatic editing of healthcare documents with structured data. Some of these tools have been tested on the coding of causes of death, which are generally coded from death certificates using the International Classification of Disease 10th revision (ICD-10). In particular, two software tools have been developed to help this type of coding: MICAR-ACME [2], developed by US National Center for Health Statistics, and more recently IRIS [3] developed by a European consortium. These tools served as the basis for the development of other national support systems for coding causes of death, such as the Italian one [4]. However, since death certificates already provide structured information, the issue of processing natural language is relatively trivial, although the mortality coding rules by themselves are complex.

In addition, automated coding tools based on Natural Language Processing (NLP) have been developed [5, 6]. Recently, a recommending system for ICD-10-CM (Clinical Modification) coding starting from SNOMED CT (Systematized Nomenclature of Medicine - Clinical Terms) annotated health records has been also developed as a consequence of the World Health Organization (WHO) – International Health Terminology Standards Development Organization (IHTSDO) harmonization effort [7]. Under the same framework, a further, ongoing evolution is the development of a common ontology between ICD-11 and SNOMED CT [8]. Nonetheless, only few systems solve coding tasks using a set of hand crafted expert rules, as in [9] which focused on the ICD-9-CM. Finally new methodologies use ontologies and automated reasoning to provide and support fast and incremental classification of medical terminologies or classification systems, as in [10] where the Snorocket reasoner has been developed to support SNOMED CT ultrafast classification.

3 A coding support system for the Patient Summary

According to the EU Guidelines, the PS is "the minimum set of information needed to assure healthcare coordination and the continuity of care" [11]. Member States adopted them often adding further clinical information, as in Italy, where a Prime Minister's Decree (to be issued) contains all the reference elements to be implemented to allow for interoperability among regional EHR systems, recognizing the critical role of the PS. According to it, PS reference elements, tagged as mandatory or optional, can be reported as free text or by using dedicated coding systems. The application of a CSS to ease the compilation of the PS allows for the selection of recommended codes to be assigned to the information required within the PS. Because of its highly structured content, the PS could be well coded using rules, reducing the variability of natural language free texts to interoperable codes.

To implement a challenging automated support system for coding health conditions in clinical documents such as the PS, a four-step methodology is proposed: (see Fig. 1 for an overview of the process):

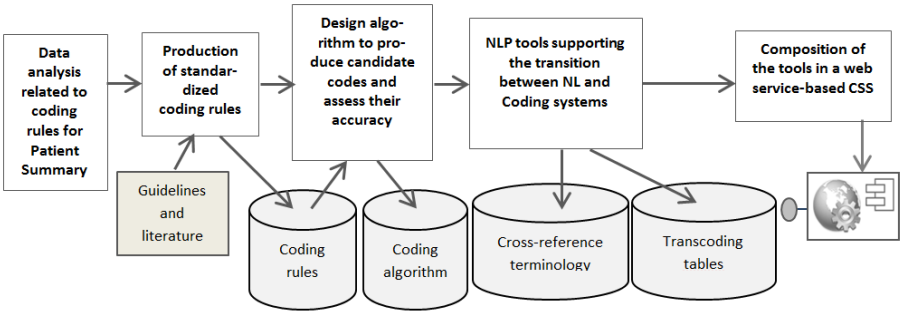

Fig. 1. Process overview

- Analysis of the epSOS project[1] results and specifications and study of the automated ICD-10 coding rules for morbidity and comorbidity[2], to verify features useful to guide the automated morbidity, procedures and interventions coding in the PS use case. This step will produce standardized *coding rules* based on general guidelines defined by qualified institutions (e.g. WHO) and described by the literature [12];
- Design of an *algorithm* that applies coding rules to produce candidate codes and assess their accuracy, and implemented in a suitable computable formal language for representing rules and the domain. The suitability of rule-based languages (e.g., OWL + SWRL)[3], and Task Network languages for the representation of guidelines (e.g., ASBRU)[4] will be analyzed;

---

1 epSOS Project: http://www.epsos.eu
2 WHO, ICD-10, vol. 2. Instruction Manual, 2010.
3 SWRL W3C recommendations: http://www.w3.org/Submission/SWRL/
4 http://www.openclinical.org/gmm_asbru.html

- Creation and use of complementary tools to support the transition from the specialized and natural language used by GPs in their EHRs to the coding language: *a cross reference terminology* of structured technical and lay terms, to be used as intermediate between the natural language and the concepts of the international coding systems; and finally *transcoding tables* to manage the different versions and revisions of a coding system (e.g. ICD-9-CM to ICD-10) or to map between different systems (e.g. SNOMED CT to ICD-10);
- Composition of the abovementioned tools to build a web service-based CSS.

The accuracy assessment of candidate codes proposed by the CSS is up to the GP, who, as mentioned above, has the full responsibility of PS clinical content.

In particular, the use of a rule-based approach with respect to other NLP ones (e.g. Support Vector Machines and Hidden Markov models) allows a better translation of rules for coding patient summaries using ICD9-CM. Those rules, similar to those defined by WHO for coding mortality, should be made explicit and translated also in a computable way. Furthermore, the creation of the cross reference terminology is based on existing terminological tools, such as the *ICD-10 Alphabetical Index*[5]; consumer-oriented medical vocabularies (e.g., the ICMV [13]); *ICD-11 narrower terms*[6] that include synonyms and quasi-synonyms; and a *dictionary* for NLP, created from a database of 295,000 EHRs [1]. Finally, a mapping to some major standardized coding systems will be performed. See Fig. 2 for an example of clinical data coding in the PS using the CSS.

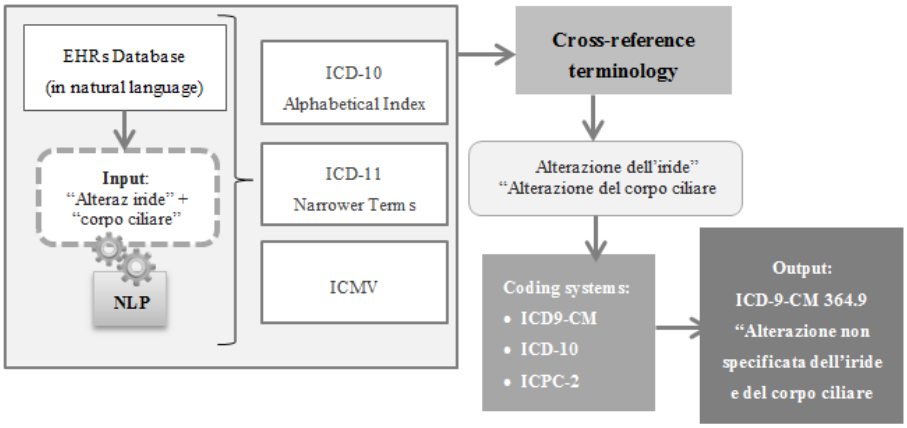

Fig. 2. A coding example using the CSS for PS

In the example above it is shown as the clinical information in input written in natural language by the GP is processed in order to arrive, as output, with the correct coding to be included in the PS. Another coding example is shown in Fig.3:

[5] WHO, ICD-10, vol. 3. Alphabetical Index, Italian version, 2014.
[6] ICD-11 beta draft available at: http://apps.who.int/classifications/icd11/browse/f/en

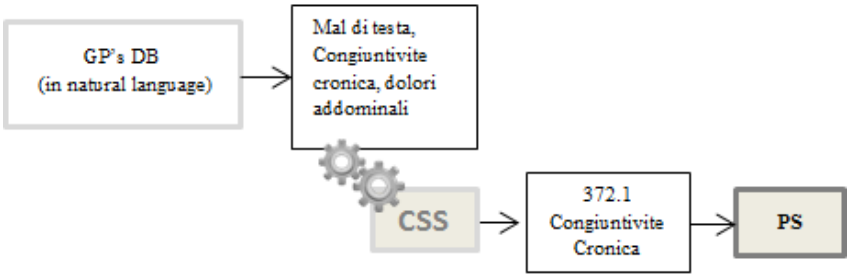

Fig. 3. Coding example 2

This use case shows how the CSS could filter the clinical information contained in GPs' database in order to include in the PS only the relevant information, (i.e. congiuntivite cronica "chronic conjunctivitis") those referring to a chronic disease, not only to single events (as "mal di testa" or "dolori addominali"). The PS, as stated above, contains only a standardized set of basic medical data including only the most important clinical facts required to ensure safe and secure healthcare [11].

4 Discussion and Conclusions

The present paper aims at proposing an experimental methodology for the development of a rule-based CSS, with an initial experimentation in Italy, that will allow to develop: i) a web service to directly support natural language text coding, and ii) a set of rules in an open format, to be embedded also in third-party software.

With respect to the limitations produced by manual coding, the use of a sound rule-based CSS presents consistent advantages (that are common to rule-based NLP methods): (i) it requires the adoption of internationally updated standard systems and standardized methodology for the accurate coding of health conditions; (ii) it could significantly reduce the coding time and costs, requiring GPs interaction only to choose among and validate the recommended codes; (iii) it improves the quality of coding by reducing the variability due to different subjective interpretations, especially in the case of comorbidities. By using the proposed CSS it is also possible to measure how often the GP is able to find the right candidates codes among those suggested by the system. On the other hand, some limitations need to be considered, mainly related to the complexity of the domain: (i) it may be necessary to formalize a huge amount of rules to represent the number of possible situations; (ii) maintenance and updating of the knowledge base and computational costs of the system can be high with the risk of inefficiency.

This pilot study will allow considerations also on the economic aspects, to be compared with the cost of training and maintaining up to date the large number of GPs that need to write and code patient summaries.

Although developed for the Italian PS, the proposed methodology could be further adapted to other European Countries.

Acknowledgements

This work is supported by the following projects: "Realizzazione di Servizi della Infrastruttura Nazionale per l'Interoperabilità per il Fascicolo Sanitario Elettronico" (prot. 7626) funded by the Agency for Digital Italy (AgID) and by the SemanticHealthNet Expert Agreement (prot. 0008991).

References

1. Cardillo, E., Chiaravalloti, M.T., Pasceri, E.: Assessing ICD-9-CM and ICPC-2 Use in Primary Care. An Italian Case Study. In: Kotsova, P., Grasso, F. (eds.), Digital Health 2015 (DH '15), pp. 95-102. ACM New York - USA (2015)
2. Israel, R. A.: Automation of mortality data coding and processing in the United States of America. World Health Stat Q. 43(4):259-62 (1990)
3. Johansson, L. A., Pavillon, G.: IRIS: A language-independent coding system based on the NCHS system MMDS. In: WHO-FIC NETWORK MEETING (2005)
4. Istituto Nazionale di Statistica – ISTAT: Metodi e software per la codifica automatica e assistita dei dati. Tecniche e Strumenti. N.4 (2007)
5. Chiaravalloti M.T., Guarasci R., Lagani V., Pasceri E., Trunfio R.: A Coding Support System for the ICD-9-CM standard, In: the IEEE International Conference on Healthcare Informatics (ICHI2014), pp. 71-78 (2014)
6. Friedman, C., Shagina, L., Lussier, Y., Hripcsak, G.: Automated Encoding of Clinical Documents Based on Natural Language Processing. JAMIA.11:392–402 (2004)
7. Campbell, J. R., Brear, H., Scichilone, R., White, S., Giannangelo, K., Carlsen, B, Solbrig, H., Fung, K.W.: Semantic interoperation and electronic health records: context sensitive mapping from SNOMED CT to ICD-10. Stud Health Technol Inform. 192:603-7 (2013)
8. Rodrigues J. M., Schulz S., Rector A., Spackman K., Üstün B., Chute C. G., Della Mea V., Millar J., Persson K. B.: Sharing ontology between ICD 11 and SNOMED CT will enable seamless re-use and semantic interoperability. Stud Health Technol Inform. 192:343-6 (2013)
9. Farkas, R., Szarvas, G.: Automatic construction of rule-based ICD-9-CM coding systems. BMC Bioinformatics. 9 (Suppl 3):S10 (2008)
10. Metke-Jimenez, A., Lawley, M.: Snorocket 2.0: Concrete Domains and Concurrent Classification. In: the OWL Reasoner Evaluation Workshop (ORE 2013), pp. 32-38 (2013)
11. eHealth Network of the European Union: Guidelines on minimum/non-exhaustive patient summary dataset for electronic exchange in accordance with the cross-border directive 2011/24/EU (2013)
12. Frattura, L., Gongolo, F., Munari, F.: Identification and coding of the main condition using ICD: suggested workflows. In: WHOFIC NETWORK Annual Meeting (2013)
13. Cardillo, E., Tamilin, A., Serafini, L.: A Methodology for Knowledge Acquisition in Consumer-Oriented Healthcare. In: Knowledge Discovery, Knowledge Engineering and Knowledge Management Communications in Computer and Information Science, Vol. 128, pp. 249-261 (2011)