PRUSKI Cedric, BONACIN Rodrigo, DA SILVEIRA Marcos « Towards the Formalization of Guidelines Care Actions using Patterns and Semantic Web Technologies ». Published in AIME 2011, 13th Conference on Artificial Intelligence in Medicine, 2-6 July, Bled, Slovenia.

Towards the Formalization of Guidelines Care Actions using Patterns and Semantic Web Technologies

Cédric Pruski[1], Rodrigo Bonacin[1,2] and Marcos Da Silveira[1]

[1]CR SANTEC, Centre de Recherche Public - Henri Tudor, 2A rue Kalchesbrück, L-1852 Luxembourg
{cedric.pruski, marcos.dasilveira}@tudor.lu
[2]CTI Renato Archer, Rodovia Dom Pedro I, km 143,6, 13069-901, Campinas, SP, Brazil
rodrigo.bonacin@cti.gov.br

Abstract. Computer Interpretable Guidelines (CIG) have largely contributed to the simplification and dissemination of clinical guidelines. However, the formalization of CIG contents, especially care actions, is still an open issue. Actually, this information, which is the heart of the guideline, is still expressed as free text and therefore prevents the development of intelligent tools for assisting physicians defining treatments. In this paper, we introduce a framework for formalizing care actions using natural language processing techniques, Semantic Web technologies and medical standards.

Keywords: CIG, Natural Language Processing, Ontologies, Semantic Web

1 Introduction

If CIGs have drastically improved the simplification and dissemination of clinical guidelines, there are still open issues. Actually, the expression of care actions denoting elements of the CIGs that intend to recommend something to be performed by health professionals is often done by using free text. In consequence, these expressions offer limited possibilities to be exploited in an efficient way by computers which in turn prevent the development of decision support tools for assisting healthcare professionals building the most appropriate treatment plans for their patients. In addition, there still exists a gap between some guidelines' specification framework and the application of the designed CIGs, mainly when patient's data is required [1]. Actually, existing medical information systems are using standard approaches like Health Level 7 (HL7), for building patients' health record (PHR) and stored it in the system. However, the current CIG execution tools are not able to automatically exploit this information because the connection between data and guidelines is not explicitly provided or in some cases this link is CIG specification language dependent [1].

This paper proposes, on the one hand, an approach to formalize care actions using linguistic patterns [2], [3] and Semantic Web technologies and, on the other hand, a (semi-)automatic way to build the link between the resulting formalization of the guidelines and the PHR. This is done through the consistent use of standards like the Unified Medical Language System (UMLS) and the Reference Information Model (RIM). Our approach is part of the global iCareflow framework [4]. It is guideline specification language independent and aims at deriving computer interpretable guidelines to personalized careflows using logic rules and Semantic Web technologies, and can be used in complement to the approach presented in [5].

The remainder of the paper is structured as follows: Section 2 introduces the conceptual framework for the formalization of care actions including the definition of the approach's elements. We wrap up with concluding remarks and outline future work in Section 3.

2 The Medical Action Formalization Framework

The Medical Action Formalization Framework (MedAForm) is basically the element of the iCareflow approach [4] that deals with the formalization of care actions contained in CIGs. This section is devoted to the description of the concepts composing the MedAForm framework.

2.1 Care Action Formalization

Our approach is performed in two steps (Fig 1): First, the care actions are extracted from CIGs and associated with standard vocabularies; second, these actions are associated to a standard data model via the generation of an ontology.

In the first step the text describing the care actions is used as input. This text, coming from a CIG independently of its specification language, is split into linguistic tokens in order to map them to a model representing medical knowledge like UMLS. It produces a set of *Actions Information* according to pre-defined selection's rules. It is basically all elements of the text that match elements of the knowledge model together with its associated formal concept inferred using the medical knowledge during the mapping process.

In the second step, the actions information is compared with a set of pre-defined patterns in order to identify the pattern group that each action belongs to. The output consists of an instance of a *reference model* (e.g. RIM), which serves as a link between the formalized guidelines and the patient health record for personalization and execution purposes in our global approach [4].

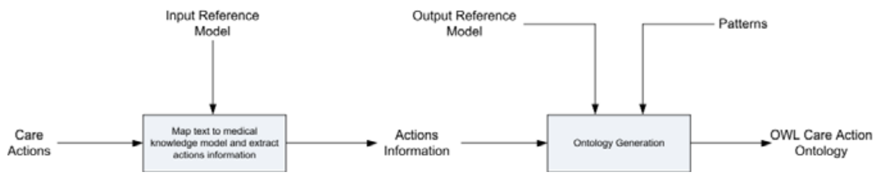

Fig. 1. The MedAForm Approach

The implementation of our approach is carried out through the definition and the use of patterns, ontology, and logic rules under the SWRL format which are described in the forthcoming sections.

2.1.1 Linguistic Pattern Matching

The selection of the pattern is done in a semi-automatic way. The initial set of patterns has been inspired by [3] and enriched after analyzing over 200 care actions extracted from guidelines implemented in SAGE [6] and PRO*forma* [7]. The goal is to generate an OWL ontology corresponding to the formal representation of the care actions contained in the guidelines, for adaptation purpose. The set of patterns we proposed contains information coming from the reference medical knowledge model and from the output model (see **Fig. 1**) in addition to the linguistic elements. The proposed set of patterns follows the format below:

(LinguisticElement, PatternOntology, Rules)

The *LinguisticElement* contained in the pattern indicates the type of action we are facing (i.e. a substance administration therapy, a radiological act, etc.). It is usually a verb which acts as a key to find the most appropriate pattern regarding the care action given as input. According to our study, 20 verbs cover more than 50% of the care actions and 5 verbs cover 24% of the selected actions information, hence we define:

$$\text{Linguistic element}=\left\{\begin{matrix}\text{Use, Give, Refer, Control, Set, Discharge, Account,}\\ \text{Return, Direct, Base, Treat, Supply,}\ldots\end{matrix}\right\}$$

The *PatternOntology* denotes the OWL ontology that contains concepts and relations to represent the care action given as input at a high level of abstraction. The ontology is populated with individuals whose labels are defined using the terms composing the care action, which have been mapped to the knowledge model. This ontology is common to all patterns and aligned with the UMLS semantic network.

Rules denotes a set of SWRL rules. It enables the definition of equivalent concepts of *PatternOntology* to concepts of the ontology representing the output reference model and which makes it possible to link the guideline with the patient's data.

Example: Consider the following care action *"Give oral activated charcoal 50g"* that has been extracted from the Tallis[1] set of guidelines.

[1] http://www.cossac.org/tallis

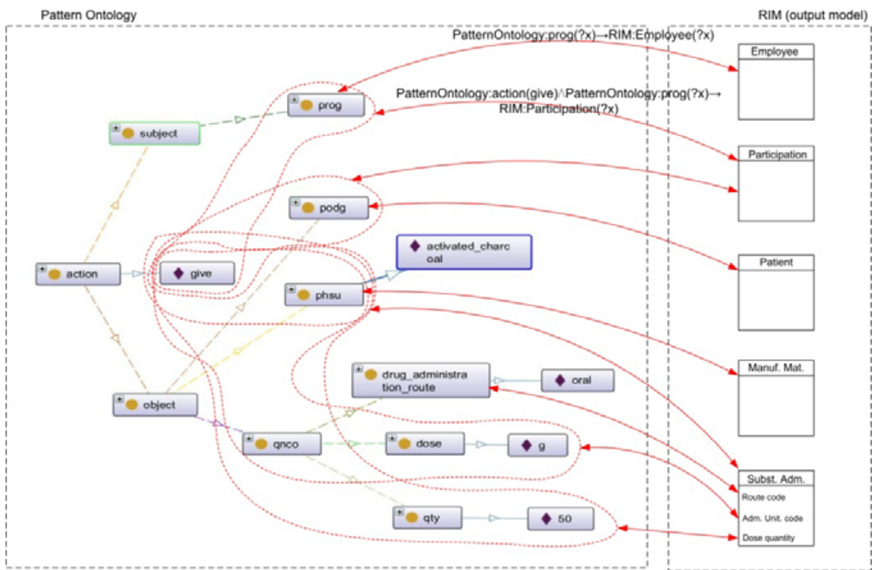

Fig. 2. Illustrative Example

In the above example, the left part corresponds to the *Pattern Ontology* describing the care action in the UMLS model (classes of this ontology are equivalent to those of the UMLS semantic network). The linguistic element is *Give*, boxes associated with circles represent the classes of the care action and boxes associated with diamonds denote individuals. These are created using the information extracted by the linguistic analyzer implemented at information extraction step (see **Fig. 1**). Their label corresponds to the terms of input care action user wants to formalize. The red arrows illustrate the mapping function defined through SWRL Rules. Since the set of rules to apply is selected according to the linguistic element of the pattern, there are no ambiguities and contradictions in the used rules. Elements inside of dashed lines serve as input to refine the representation of the action in the output model. In this example, sets in dashed line explain the construction of the SWRL rules[2].

At run time, the selection of the pattern is done in a semi-automatic way. The system starts by matching the *linguistic elements* of the existing patterns and those of the care actions. If several elements are detected, ambiguities are removed in a cyclic way by using the *PatternOntology*. As the *PatternOntology* is aligned with the medical knowledge model to extract the information from the care action, the system maps the classes of the *PatternOntology* with the information extracted based on the medical knowledge model. The system stops if the information can no longer be exploited. If several patterns remain, the user selects the most appropriate one.

---

[2] In **Fig. 2**, only two rules are mentioned for readability reasons.

2.1.2 Ontology Generation

Patterns serve as basis for the building of the ontology representing the initial guideline. The classes of the generated ontology are mapped to classes or a set of classes of the *PatternOntology*. The resulting ontology is expressed in OWL for homogeneity reasons. Then it is populated by individuals, inferred from using the individuals of the *PatternOntology* in combination with the SWRL rules.

In the example of **Fig. 2**, the instance of the class "Manuf. Mat." is "activated charcoal". It has been inferred from using the following SWRL rule:

PatternOntology:phsu(?x)→RIM:ManufMat(?x)

3 Conclusion

In this paper we have presented an approach to formally describe care actions in a way that can be linked to patients' medical records. The introduced framework, which is guideline specification language independent, implements linguistic patterns, Semantic Web technologies and is compatible with well-accepted standards. In our future tasks we will first consider creating a repository for linguistic patterns and we will improve the interface of the tool for facilitating both the capture of care actions and the construction of new linguistic patterns.

4 References

1. Peleg, M., Keren, S., Denekamp, Y.: Mapping Computerized Clinical Guidelines to Electronic Medical Records: Knowledge-Data Ontological Mapper (KDOM). Journal of Biomedical Informatics 41, 180-201 (2008)
2. Essaihi, A., Michel, G., Shiffman, R.N.: Comprehensive Categorization of Guideline Recommendations: Creating an Action Palette for Implementers. Proc AMIA Symp (American Medical Informatics Association), pp. 220-224, Washington D.C. (2003)
3. Serban, R., ten Teije, A., van Harmelen, F., Marcos, M., Polo-Conde, C.: Extraction and use of linguistic patterns for modelling medical guidelines. Artif. Intell. Med. 39, 137-149 (2007)
4. Bonacin, R., Da Silveira, M., Pruski, C.: From Medical Guidelines to Personalized Careflows: the iCareflow Ontological Framework. The 23[RD] IEEE International Symposium on Computer-Based Medical Systems. IEEE Computer Society, Perth, Australia (2010)
5. Kaiser, K., Akkaya, C., Miksch, S.: How can information extraction ease formalizing treatment processes in clinical practice guidelines?: A method and its evaluation Artificial Intelligence in Medicine 39, 151-163 (2007)
6. Beard, N., Campbel, J.R., Huff, S.M., Leon, M., Mansfield, J.G., Mays, E., McClay, J.C., Mohr, D.N., Musen, M.A., O'Brien, D., Rocha, R.A., Saulovich, A., Scheitel, S.M., Tu, S.W.: Standards-Based Sharable Active Guideline Environment (SAGE): A Project to Develop a Universal Framework for Encoding and Disseminating Electronic Clinical Practice Guidelines. Proc AMIA Symp, pp. 973 (2002)
7. Fox, J., Johns, N., Rahmanzadeh, A.: Disseminating medical knowledge: The PROforma approach Artificial Intelligence in Medicine 14, 157-181 (1998)