Identifying Actions Described in Clinical Practice Guidelines Using Semantic Relations

Katharina Kaiser[1,2], Andreas Seyfang[1], and Silvia Miksch[1,3]

[1] Institute of Software Technology & Interactive Systems
Vienna University of Technology
Favoritenstraße 9-11/188, 1040 Vienna, Austria
[2] Center for Medical Statistics, Informatics, & Intelligent Systems
Medical University of Vienna
Spitalgasse 23, 1090 Vienna, Austria
[3] Department of Information & Knowledge Engineering
Danube University Krems
Dr.-Karl-Dorrek-Straße 30, 3500 Krems, Austria

Abstract. Clinical practice guidelines are important instruments to support clinical care. We have analyzed guidelines according to their semantic relations in order to generate a formal representation. We used the UMLS Semantic Network as a basis for our analysis. We defined relations that will be used to automatically identify the control flow described in guidelines for generating a computer-interpretable format. A study showed that by using semantic relations we are able to discover a large part of the control flow.

1 Introduction

Clinical practice guidelines (CPGs) are important instruments to provide state-of-the-art clinical practice for the medical and clinical personnel [1]. It has been shown that integrating CPGs into clinical information systems can improve the quality of care [2]. Among many services they can offer are: summarizing patient data; providing alerts and reminders; retrieving and filtering information which is relevant to a specific decision; and weighing up the pros and cons of clinical options in a patient-specific way [3]. For integrating CPGs into such systems, the CPGs (i.e., documents in free, narrative text) have to be translated into a computer-interpretable format. Various of such guideline representation formats have been developed (see [4] for an overview and comparison), but the translation process is still a large bottle-neck. It's a complex and challenging task requiring both medical and computer science expertise. Several methods have been developed that describe systematic approaches (e.g., [5,6,7]) for guideline modelling. Some of them use intermediate formats to break the complexity of the modelling task into manageable tasks (e.g., the GEM format [8], the *Many-Headed Bridge between Guideline Formats (MHB)* [9]).

A CPG can be translated into a computer-interpretable format by representing it as an instance of an ontology. Thereby, concepts are identified and relations between them are established. In this work we show what relations are necessary for identifying text representing the control flow of a guideline. We thereby utilize the UMLS Semantic

Network [10], which provides not only semantic types, but also relations between them. Our hypothesis is that by using these relations we can identify activities to be performed in a guideline document.

Modelling CPGs in a computer-interpretable and -executable format is a great challenge. Automatically identifying actions and processes that describe the control flow can facilitate the modelling process by providing important information necessary for the modelling task and by reducing the workload for the involved persons. Thereby, the modelling process will still require manual interaction (e.g., to solve ambiguities and inconsistencies and to add knowledge necessary for execution).

In the following section we give a short overview on knowledge-based methods for guideline modelling and on using relations for identifying information. In Section 3 we explain the methods developed and resources used. Section 4 describes the the evaluation of our methods and a discussion of its results. The final section contains our conclusions.

2 Background

First attempts to computerised CPGs have been implemented in the Protégé environment [11]. Thereby, the class structure has been defined according to the representation formalism (e.g., GLIF [12]). The according concepts are identified in the CPG and assigned to the classes. All implementations use a *model-centric approach* where no direct connection between the guideline text and the corresponding model is generated.

Other attempts have been made to support the modelling, maintenance, and shareability of computer-interpretable guidelines that rather stick on the original text representation. Moser and Miksch introduced prototypical patterns in clinical guidelines that can be used as means to reduce the gap between the information represented in clinical guidelines and the formal representation of these clinical guidelines in execution models [13]. They defined *structure patterns*, *temporal patterns*, and *element patterns*. Serban et al. [14] defined linguistic patterns found in CPGs that can be used to support both the authoring and the modelling process of guidelines. Furthermore, they defined an ontology out of these patterns and linked them with existing thesauri in order to use compilations of thesauri knowledge as building blocks for modeling and maintaining the content of a medical guideline [15]. Peleg and Tu developed visual templates that structure screening guidelines as algorithms of guideline steps used for screening and data collection and used them to represent the guidelines collected [16]. To support guideline implementers in standardising and implementing the action components of guideline recommendations, Essaihi et al. defined the Action Palette [17] — a set of medical action types that categorises activities recommended by clinical guidelines. The set of actions include: Prescribe, Perform therapeutic procedure, Educate/ Counsel, Test, Dispose, Refer/Consult, Conclude, Document, Monitor, Advocate, and Prepare. The intention is to develop commonly used services for each action type, thus easing workflow integration. In [18] we proposed a method to identify actions in otolaryngology CPGs using a subset of the *Medical Subject Headings* (MeSH) as a thesaurus.

In [19] SeReMeD is introduced. It is a method for automatically generating knowledge representations from natural language documents. The documents that were applied on this method were rather simple (i.e., X-ray reports). CPGs are substantial complex documents and the modelling task is particularly challenging.

3 Methods

In order to define the relations necessary we used the UMLS Semantic Network (SN) [10]. The SN offers amongst 135 semantic types also 54 semantic relationships between these types and therefore acts as an upper-level ontology. Together, there are 6,752 different relations. We then proceeded in four steps:

1. Analyzing CPG documents regarding actions and procedures to be performed within treatment.
2. Detection of relations that describe the actions and procedures.
3. Augmenting the initial relation set by specialism and broadening of semantic types.
4. Generation of a dictionary for identifying the correct relation type.

We started by analysing a protocol (i.e., a local adaption of a guideline) for natural birth that was derived from the guideline “Induction in labour” [20]. The document describes the labor and delivery management for risk-free births and consists of 120 sentences. 48 sentences include actions to be performed.

We used the MetaMap Transfer program (MMTx) [21] to find relevant concepts in the text and to assign a semantic type to each concept. In case an appropriate concept is assigned more than one semantic type, we chose the most specific one if more general were also applicable. If semantic types were not hierarchically linked, all remained as possible semantic types.

Then, we looked for appropriate relations between semantic types. In 47 sentences we detected 52 activities by means of 20 different semantic relations. For one sentence ('In case of abnormal FHR, monitoring should be continuous.') no relation could be identified although the text implies an action ('monitoring should be continuous'). Table 1 shows all found relations together with the number of occurrences.

All relations containing the semantic type *Professional or Occupational Group* on the left hand side are thereby incomplete relations. That means that not both semantic types that are linked by the relationship are mentioned in the text. This is the result from the fact that actions and activities in CPGs are related to tasks the health care personnel has to perform (in our specific case: gynaecologists and obstetricians, who are assigned the semantic type *Professional or Occupational Group*). As most of the actions address these users directly, they are only implicitly referred to in the text. Furthermore, passive voice is frequently used, which also allows omission of the agent of the action. See Figure 1 for an example.

We also identified complete relations. See Figure 2 for an example.

The last four relations displayed in Table 1 are not contained in the SN. We generated these relations in order to be able to represent activities such as ‘*Fetal heart rate should be auscultated*.’, where ‘fetal heart rate’ is assigned the semantic type *Finding*.

In the next step we augmented our initial set of relations by generalization, specialization, and broadening of semantic types. Thereby, we expanded a relation to semantic types of the same *Semantic Collection* [22]. A semantic collection is a partition of the

Table 1. Relations detected in the guideline document.

Semantic Type	relationship	Semantic Type	Quantity
Professional or Occupational Group	performs	Health Care Activity	8
Professional or Occupational Group	performs	Therapeutic or Preventive Procedure	14
Professional or Occupational Group	performs	Diagnostic Procedure	2
Professional or Occupational Group	performs	Research Activity	1
Population Group	performs	Therapeutic or Preventive Procedure	1
Population Group	performs	Diagnostic Procedure	1
Professional or Occupational Group	interacts_with	Population Group	4
Therapeutic or Preventive Procedure	affects	Mental Process	1
Health Care Activity	affects	Mental Process	1
Professional or Occupational Group	uses	Intellectual Product	1
Professional or Occupational Group	uses	Pharmacologic Substance	3
Pharmacologic Substance	isa	Pharmacologic Substance	1
Health Care Activity	isa	Health Care Activity	2
Therapeutic or Preventive Procedure	isa	Health Care Activity	2
Therapeutic or Preventive Procedure	isa	Therapeutic or Preventive Procedure	2
Diagnostic Procedure	measures	Clinical Attribute	1
Professional or Occupational Group	analyzes	Body Substance	3
Professional or Occupational Group	analyzes	Finding	1
Professional or Occupational Group	analyzes	Organism Function	2
Professional or Occupational Group	analyzes	Organism Attribute	1

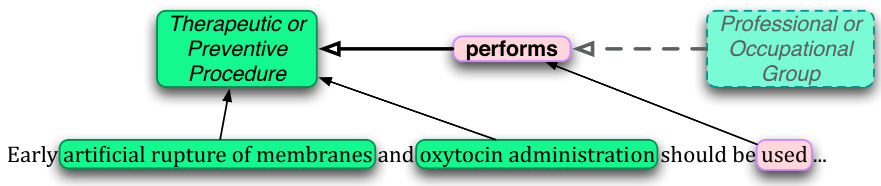

Fig. 1. An action describing an incompletely formulated semantic relation.

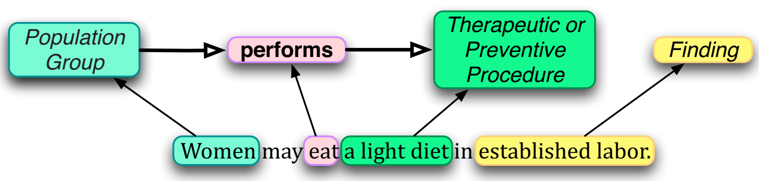

Fig. 2. An action describing a completely formulated semantic relation.

Table 2. Augmented relation set.

Semantic Type	relationship	Semantic Type
Group Professional or Occupational Group Population Group Family Group Age Group Patient or Disabled Group	performs	Health Care Activity Therapeutic or Preventive Procedure Diagnostic Procedure Laboratory Procedure Educational Activity Research Activity
Professional or Occupational Group	interacts_with	Age Group Population Group Family Group Group Patient or Disabled Group Professional or Occupational Group
Therapeutic or Preventive Procedure Health Care Activity Diagnostic Procedure Laboratory Procedure	affects	Mental Process
Professional or Occupational Group	uses	Intellectual Product Pharmacologic Substance Antibiotic Clinical Drug Drug Delivery Device Manufactured Object Medical Device
Therapeutic or Preventive Procedure	uses	Pharmacologic Substance Antibiotic Medical Device Clinical Drug Drug Delivery Device
Pharmacologic Substance	isa	Pharmacologic Substance
Antibiotic	isa	Pharmacologic Substance
Antibiotic	isa	Antibiotic
Diagnostic Procedure	isa	Health Care Activity
Laboratory Procedure	isa	Health Care Activity
Therapeutic or Preventive Procedure	isa	Health Care Activity
Health Care Activity	isa	Health Care Activity
Therapeutic or Preventive Procedure	isa	Therapeutic or Preventive Procedure
Diagnostic Procedure	measures	Clinical Attribute
Diagnostic Procedure	measures	Organism Attribute
Laboratory Procedure	measures	Clinical Attribute
Laboratory Procedure	measures	Organism Attribute
Professional or Occupational Group	analyzes	Body Substance Finding Organism Function Organism Attribute Clinical Attribute Sign or Symptom

semantic network consisting of semantic types with structural similarity. For instance, the semantic collection 'Group' consists of the semantic types 'Group', 'Age Group', 'Family Group', 'Patient or Disabled Group', 'Population Group', and 'Professional or Occupational Group'. All these semantic types have the same semantic relations. Altogether, there exist 28 semantic collections. We received 76 relations that can be used to cover activities and processes described in a CPG (see Table 2).

To identify a relation we need to know the relationship between two semantic types. As between two semantic types more than one relation can exist, we need to detect the appropriate relationship (i.e., the type of relation, e.g., performs, uses, interacts_with). This is accomplished by means of the verbs in the particular sentence or clause. We only have seven different relationships. We assigned the verbs appearing in the action sentences to their particular relationship. Furthermore, we expanded our relationship-verb assignment by synonymous verbs from an online thesaurus [23] (for an example see Table 3).

Table 3. Verbs indicating a relation for relation type 'perform'.

Verbs identified in the text	Synonyms from thesaurus
do, perform, eat, reserve, offer, register, maintain, evaluate, record, sign, use, consider, repeat, relieve, avoid, provide, insert	execute, discharge, accomplish, achieve, fulfill, start, complete, conduct, effect, dispatch, work, implement, require, undertake, order, arrange, keep, continue, discontinue, produce, install, load, push, drink, include, recommend, report, increase, reduce, treat, wait, initiate, carry out, carry off, ...

4 Evaluation

In order to show whether our semantic relations can be used also on other guideline documents, we performed an empirical study. We therefore used the guideline "Management of labor" [24]. Although the guideline covers the same application area, it is structured in a different way, uses different phrasing and wording. Furthermore, it contains lots of background information and literature references and it is almost twice as large (60 pages) as the document we used to define the set of semantic relations.

In a first step we had to identify all the actions and processes in the document that would be necessary to model the control flow. This was done by an expert for modelling guidelines using UMLS release 2010AA for identifying concepts and their semantic types. He also made an initial suggestion of the relation applicable to the action. For the whole document 95 sentences were identified containing actions or procedures to perform.

We then applied the MMTx program to the whole document to identify concepts and assign them semantic types. We further checked whether a relation between the semantic types exists that indicate an action and verified whether the verb that should indicate the relationship was also available in our dictionary.

With our relations we could correctly identify 67 actions. 33 actions could not be identified because the action was not expressed as a relation (e.g., grammatically incomplete sentences) or because no UMLS concept could be identified and therefore, no semantic type was assigned (see Table 5 for examples). In only two sentences actions were identified that were not modelled by our expert. Thus, our method has a recall of 67% and a precision of 97% (see Table 4).

Table 4. Evaluation results.

COR	INC	POS	ACT	REC	PRE
67	2	100	69	67%	97%

COR = number of correctly identified actions by the method
INC = number of incorrect identified actions by the method
POS = number of actions according to the key target template
ACT = number of actions identified by the method
REC = COR/POS
PRE = COR/ACT

Table 5. Examples for actions that could not be identified.

Sentence	Reason
Monitoring of fetal heart rate.	incomplete sentence
In the presence of preterm labor with bleeding, IV access is essential.	no relation identified for 'IV access is essential'
Do not start active pushing as soon as patient is fully dilated.	no UMLS concept found for 'active pushing'
Management of labor dystocia is especially important in the nulliparous woman to prevent unneeded Caesarean sections.	no relation identified for 'management ... is especially important ...'

In order to also include actions that were now erroneously omitted we have to consider incomplete sentences (e.g., headings, table items, lists) as well as sentences that contain an adjective instead of a verb to identify the relationship.

It is also to notice that the majority of relations appearing in the guideline are the same in the initial relation set. Only 15 actions were identified by nine relations that had been added to the initial relation set. These relations all contain semantic types that belong to the same Semantic Collection as the initial relation.

By using our dictionary for detecting the relationship we are able to omit the wrong identification of text that is not expressing an action. For instance, the sentence 'Vaginal examination is aimed at evaluation of cervical effacement ...' would be tagged with the semantic type *Health Care Activity* for both ‘vaginal examination’ and ‘evaluation of cervical effacement’. But the verb phrase ‘is aimed at’ could indicate an intention and not the relationship ‘isa’, which is the only relationship between *Health Care Activity* and *Health Care Activity*.

5 Conclusions and Further Work

This work presents a method to identify actions and procedures to be performed during a treatment which are described in a guideline document. We used relations of the UMLS Semantic Network [10] to identify these actions in a guideline document. We defined a set of relations that are relevant for these kind of documents. With these relations we are able to identify a large part of activities that are contained in the control flow of such documents.

By assigning semantic information to text and linking it using semantic relations a further processing is facilitated. This can be used for guideline modelling into a computer-interpretable formalism. Manual and semi-automatic modelling will benefit from the information gathered by this method. Furthermore, this method can also support the processing of new versions of guidelines by identifying new or different relations.

Our next steps will be (1) the expansion of using the Semantic Network for identifying other information dimensions, such as effects, intentions, or parameters; (2) the definition and categorization of further relations to be able to differentiate between a single action, a decomposition of an action, or a selection of one or more alternative actions; (3) defining special patterns to identify actions that are either not identifiable by a relation or where the relation is identifiable by adjectives; and (4) a further processing to (automatically) transform the guideline document towards a computer-interpretable guideline (e.g., into the *Many-Headed-Bridge* (MHB) [9] formalism). This can promote the application of computer-interpretable CPGs.

Acknowledgement. This work is partially supported by “Fonds zur Förderung der wissenschaftlichen Forschung FWF” (Austrian Science Fund), grant TRP71-N23, and the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 2161.

References

1. Field, M.J., Lohr, K.N., eds.: Clinical Practice Guidelines: Directions for a New Program. National Academies Press, Institute of Medicine, Washington DC (1990)
2. Quaglini, S., Ciccarese, P., Micieli, G., Cavallini, A.: Non-compliance with guidelines: Motivations and consequences in a case study. In Kaiser, K., Miksch, S., Tu, S.W., eds.: Computer-based Support for Clinical Guidelines and Protocols. Proceedings of the Symposium on Computerized Guidelines and Protocols (CGP 2004). Volume 101 of Studies in Health Technology and Informatics., Prague, Czech Republic, IOS Press (2004) 75–87
3. Fox, J., Patkar, V., Chronakis, I., Begent, R.: From practice guidelines to clinical decision support: closing the loop. Journal of the Royal Society of Medicine **102**(11) (2009) 464–473

4. Peleg, M., Tu, S.W., Bury, J., Ciccarese, P., Fox, J., Greenes, R.A., Hall, R., Johnson, P.D., Jones, N., Kumar, A., Miksch, S., Quaglini, S., Seyfang, A., Shortliffe, E.H., Stefanelli, M.: Comparing computer-interpretable guideline models: A case-study approach. Journal of the American Medical Informatics Association (JAMIA) **10**(1) (Jan-Feb 2003) 52–68
5. Tu, S.W., Musen, M.A., Shankar, R., Campbell, J., Hrabak, K., McClay, J., Huff, S.M., McClure, R., Parker, C., Rocha, R., Abarbanel, R., Beard, N., Glasgow, J., Mansfield, G., Ram, P., Ye, Q., Mays, E., Weida, T., Chute, C.G., McDonald, K., Mohr, D., Nyman, M.A., Scheital, S., Solbrig, H., Zill, D.A., Goldstein, M.K.: Modeling guidelines for integration into clinical workflow. In Fieschi, M., Coiera, E., Li, Y.C.J., eds.: Proceedings from the Medinfo 2004 World Congress on Medical Informatics. Volume 107 of Studies in Health Technology and Informatics., AMIA, IOS Press (2004) 174–178
6. Svátek, V., Růžička, M.: Step-by-step mark-up of medical guideline documents. International Journal of Medical Informatics **70**(2-3) (July 2003) 319–335
7. Seyfang, A., Kaiser, K., Miksch, S.: Modelling clinical guidelines and protocols for the prevention of risks against patient safety. In: Proceedings of the XXII International Conference of the European Federation for Medical Informatics. Volume 150 of Studies in Health Technology and Informatics., Sarajevo, Bosnia and Herzegovina, IOS Press (2009) 633–637
8. Shiffman, R.N., Karras, B.T., Agrawal, A., Chen, R., Marenco, L., Nath, S.: GEM: a proposal for a more comprehensive guideline document model using XML. Journal of the American Medical Informatics Association (JAMIA) **7**(5) (2000) 488–498
9. Seyfang, A., Miksch, S., Marcos, M., Wittenberg, J., Polo-Conde, C., Rosenbrand, K.: Bridging the gap between informal and formal guideline representations. In Brewka, G., Coradeschi, S., Perini, A., Traverso, P., eds.: European Conference on Artificial Intelligence (ECAI-2006). Volume 141 of Frontiers in Artificial Intelligence and Applications., Riva del Garda, Italy, IOS Press (2006) 447–451
10. McCray, A.T.: UMLS Semantic Network. In: Proc. of the 13th Annual Symposium on Computer Applications in Medical Care (SCAMC'89). (1989) 503–507
11. Gennari, J.H., Musen, M.A., Fergerson, R.W., Grosso, W.E., Crubézy, M., Eriksson, H., Noy, N.F., Tu, S.W.: The Evolution of Protégé: An Environment for Knowledge-based Systems Development. International Journal of Human Computer Studies **58**(1) (2003) 89–123
12. Boxwala, A.A., Peleg, M., Tu, S.W., Ogunyemi, O., Zeng, Q., Wang, D., Patel, V.L., Greenes, R.A., Shortliffe, E.H.: GLIF3: A representation format for sharable computer-interpretable clinical practice guidelines. Journal of Biomedical Informatics **37**(3) (June 2004) 147–161
13. Moser, M., Miksch, S.: Improving clinical guideline implementation through prototypical design patterns. In Miksch, S., Hunter, J., Keravnou, E., eds.: Proceedings of the 10th Conference on Artificial Intelligence in Medicine in Europe, AIME 2005. Volume 3581 of LNAI., Aberdeen, UK, Springer Verlag (2005) 126–130
14. Serban, R., ten Teije, A., van Harmelen, F., Marcos, M., Polo-Conde, C.: Extraction and use of linguistic patterns for modelling medical guidelines. Artificial Intelligence in Medicine **39**(2) (2007) 137–149
15. Serban, R., ten Teije, A.: Exploiting thesauri knowledge in medical guideline formalization. Methods Inf Med **48** (2009) 468–474
16. Peleg, M., Tu, S.W.: Design patterns for clinical guidelines. Artificial Intelligence in Medicine **47**(1) (2009) 1 – 24
17. Essaihi, A., Michel, G., Shiffman, R.N.: Comprehensive categorization of guideline recommendations: Creating an action palette for implementers. In: AMIA 2003 Symposium Proceedings, AMIA (2003) 220–224
18. 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**(2) (2007) 151–163

19. Denecke, K.: Semantic structuring of and information extraction from medical documents using the umls. Methods of Information in Medicine **47** (2008) 425–434
20. National Collaborating Centre for Women's and Children's Health: Induction of labour. Clinical guideline 70, National Institute for Health and Clinical Excellence (NICE), London (UK) (July 2008)
21. Aronson, A.R.: Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. In: Proceedings of the 25^{th} Annual Americal Medical Informatics Association Symposium (AMIA'01), Washington, D.C., AMIA, Washington, D.C. (Nov. 3-7 2001) 17–21
22. Chen, Z., Perl, Y., Halper, M., Geller, J., Gu, H.: Partitioning the UMLS Semantic Network. IEEE Transactions on Information Technology in Biomedicine **6**(2) (June 2002) 102–108
23. Lindberg, C.A., ed.: Oxford American Writer's Thesaurus. 2nd edn. Oxford University Press (2008)
24. Institute for Clinical Systems Improvement (ICSI): Management of labor. Clinical guideline, Institute for Clinical Systems Improvement (ICSI), Bloomington (MN) (May 2009)