Structuring Clinical Decision Support rules for drug safety using Natural Language Processing

George DESPOTOU[a,1], Ioannis KORKONTZELOS[b], Nicholas MATRAGKAS[c], Eda Bilici[a], and Theodoros N ARVANITIS[a]
[a]*Institute of Digital Healthcare, WMG, University of Warwick, UK.*
[b]*Department of Computer Science, Edge Hill University, UK.*
[c]*Department of Computer Science, University of Hull, UK.*

Abstract. Drug safety is an important aspect in healthcare, resulting in a number of inadvertent events, which may harm the patients. IT based Clinical Decision Support (CDS), integrated in electronic-prescription or Electronic Health Records (EHR) systems, can provide a means for checking prescriptions for errors. This requires expressing prescription guidelines in a way that can be interpreted by IT systems. The paper uses Natural Language Processing (NLP), to interpret drug guidelines by the UK NICE BNF offered in free text. The employed NLP component, MetaMap, identifies the concepts in the instructions and interprets their semantic meaning. The UMLS semantic types that correspond to these concepts are then processed, in order to understand the concepts that are needed to be implemented in software engineering for a CDS engine.

Keywords. Pharmacovigilance, drug safety, CDS, NLP

Introduction

Advances in interoperability and architecture of Health IT systems have enabled provision of new functions that allow integrating aspects of healthcare that were previously isolated. Examples include, integrating data sources under a single record, and provision of personalized care plans that can be accessed and evaluated by multiple stakeholders such as experts, family and the patients themselves [1]. Such health IT infrastructures, are based on collaboration amongst components, using messaging standards such as FHIR provide an overarching new capability.

Clinical Decision Support (CDS) modules often provide a means of checking the information exchanged between components for medical significance. For example improving drug safety, checking prescriptions for validity, as well as identifying adverse interaction between prescribed drugs, particularly to polypharmacy patients. Such check is the validation of a prescription against typical instructions, such as maximum dose during a time period. However, in order for this to be achieved by a health IT based system, these instructions need to be captured in a format that is executable by IT systems. The instructions need to be modelled in a way that an IT system can understand their

[1] Corresponding Author.

semantics, in order to be applied as rules when checking a prescription. The National Institute for Care Excellence (NICE) provides dose instructions for the drugs listed in their British National Formulary (BNF). The NICE BNF provides structured definitions of drug aspects such as drug to drug interactions, but uses free text for drug instructions. For example, "*500mg 3 times a day*", and "*300–900mg every 4–6hours as required; maximum 4g per day.*". Dose instructions are expressed using a number of (semantically) common concepts. For example, quantity of a substance, frequency, quantity checks (e.g., maximum 2g), as well as more complex logic including past or future events and checks. The work presented in this paper looks into the UK NICE BNF dose instructions, and using Natural Language Processing (NLP), identifies the types of semantic expression and investigates how we can generalize these types, and how they can be used to specify rules that can be applied to health IT systems that contribute to medication prescription. NLP has been used to extract semantic relations from clinical texts categorizing and mapping clinical knowledge [2, 3, 4, and 5].

1. Overview of the approach

Information about drugs are taken from the UK NICE section of the British National Formulary (BNF), a reference on pharmacology and prescribing. The drugs are stored in a structured way, as objects, in an Eclipse Modelling Framework model; this allows processing the elements of the model. The UK NICE BNF provides instructions for each combination of *patient category*, *drug administration route*, and *indication*. All instructions are exported in a format suitable for NLP analysis.

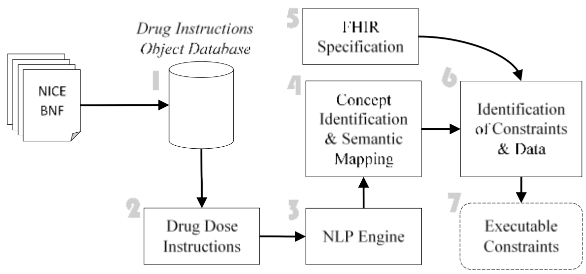

Figure 1. Overview of the Approach.

The instructions are then processed by MetaMap [6], an NLP tool that recognizes UMLS Metathesaurus concepts referred to in biomedical text entered in the NLP Engine. The UMLS, or Unified Medical Language System, is an interoperable collection of health and biomedical vocabularies. Concepts, i.e. vocabulary terms, are categorized in semantic types, i.e. broad subject categories, such as *Activity*, *Genetic Function* and *Fungus*.

The NLP engine annotates blocks of the text according to its semantics; for example *less than 200mg* consists of a quantitative concept (200mg) and a functional concept (*less than*) that is the operation on the quantity. A total of 15000 instructions for 1660 drugs have been processed using NLP, for identification of their semantic building blocks. This produces a set of concepts that will either represent a resource (i.e. an information about prescribing that will be available in the system), or a rule (an operation that will need to be performed on a resource which will be enforcing the instructions). In order to identify whether the identified resources will be available, we are looking at how they can be matched with the data elements of the *MedicationPrescription* data structure of the Fast Healthcare Interoperability Resource (FHIR) standard. FHIR is a messaging exchange standard that has improved interoperability of health IT systems, by offering a standardized means of exchanging information. Although not a storage standard, any component complying to FHIR would be expected to provide a specific set of information, structured in a specific way. Figure 2 shows the class diagram equivalent of the *MedicationPrescription* FHIR resource which is of interest. The last stage of the process (7) is outside the scope of this paper.

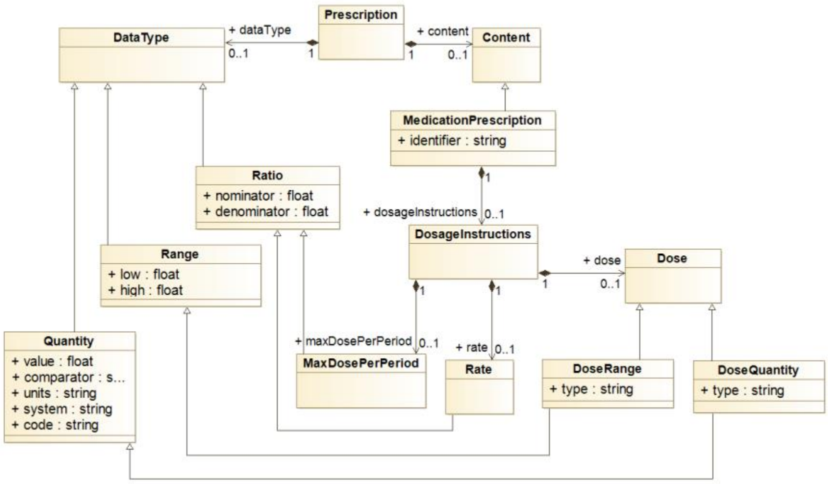

Figure 2. UML class diagram showing some of the typical information concepts expected in a prescription, according to the FHIR *MedicationPrescription* resource.

2. NLP Concepts

Table 1 presents the seven most frequent UMLS concepts that appear in the instructions, along with their associated semantic types.

Table 1. Frequency of the UMLS concepts in the NICE BNF drug instructions.

Position	Frequency	CUI	Semantic Type	Concept
1	4200	C0439210	Quantitative Concept	MILLIGRAM (milligram)
2	2888	C0178602	Quantitative Concept	DOSE (Dosage)
3	2757	C0332173	Temporal Concept	/day (Daily)
4	1289	C0205265	Temporal Concept	Initially
5	1015	C1883708	Temporal Concept	Then
6	974	C0442805	Functional Concept	Increased (Increase)
7	923	C0585361	Temporal Concept	Twice Daily (Twice a day)

A total of 1984 concepts were identified; however, the 80 most frequent concepts had 37,000 appearances in the instructions, and the rest 1904, 52,173. These 1984 concepts were identified to belong to 128 semantic types, with the 10 most frequent semantic types appearing 45,000 times.

3. Designing rules based on FHIR resources

Concepts and semantic types can be mapped to specific rules that will be implemented in an executable format in a CDS engine. This section shows an example of how such rules can be specified using the Object Constraint Language (OCL). OCL can then be applied to any object oriented environment and transformed into relational database query and rule. For example, three of the most common concepts appearing were *mg*, *dose* and *twice a day*. This respectively correspond to *Quantity*, *Dose*, and *Rate* in the *MedicationPrescription.* Figure 3 shows how such a rule would be specified in OCL.

```
context Prescription inv correctDosage:
  self.content.oclAsType(MedicationPrescription).dosageInstructions-
  >forAll( i | i.dose -> forAll(d | oclAsType(DoseQuantity).type = 'mg' and
  d.oclAsType(DoseQuantity).oclAsType(Quantity).value <200))
```

Figure 3. OCL rule using the *Quantity*, *Dose*, and *Rate*, FHIR resources.

4. Conclusions

The experiment offered two main conclusions. Firstly, there is a strong indication that drug instructions only need a limited number of concepts to be expressed. This can allow the specification of a controlled vocabulary language that can be used to define all potential instructions. Secondly, the concepts, and the even fewer semantic types they correspond to, allow their mapping to FHIR *MedicationPrescription* resources, which is the gold standard in health IT interoperability. This will allow the definition of CDS rules that can be imported and executed in any compliant module. Further work being planned entails a controlled vocabulary, as well as training of the NLP engine to recognize concepts specific to drug administration instructions. The latter will address a common limitation in NLP, which is the confidence with which the interpretation of the text is done.

References

[1] E.G.B. Laleci, M. Yuksel and T.N. Arvanitis, C3-Cloud: a federated collaborative care and cure cloud architecture for addressing the needs of multi-morbidity and managing poly-pharmacy, *Informatics for Health* 2017 / 24 – 26 April 2017.

[2] A. L. Minard and K. Kaiser, Supporting computer-interpretable guidelines' modeling by automatically classifying clinical actions. In *Process Support and Knowledge Representation in Health care* 2013 (pp. 39-52). Springer, Cham.

[3] W. G. El-Rab, O. R. Zaïane and M. El-Hajj, Formalizing clinical practice guideline for clinical decision support systems. *Health Informatics Journal*. 2017;**23**(2):146-56.

[4] Z. Luo, M. Yetisgen-Yildiz and C. Weng. Dynamic categorization of clinical research eligibility criteria by hierarchical clustering. *Journal of Biomedical Informatics*. 2011;**44**(6):927-35.

[5] N. Hong, D. Li, Y. Yu, Q. Xiu, H. Liu, and G. Jiang, A computational framework for converting textual clinical diagnostic criteria into the quality data model. Journal of biomedical informatics. 2016;**63**:11-21.

[6] A. R. Aronson and F. M. Lang, An overview of MetaMap: historical perspective and recent advances. *Journal of the American Medical Informatics Association*: JAMIA. 2010;**17**(3):229-236. doi:10.1136/jamia.2009.002733.