Caring is Sharing – Exploiting the Value in Data for Health and Innovation M. Hägglund et al. (Eds.) © 2023 European Federation for Medical Informatics (EFMI) and IOS Press. This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0). doi:10.3233/SHTI230280

Digital Health Data Capture with a Controlled Natural Language

Kristian Juha Ismo KANKAINEN^{a,1}, Inari LISTENMAA^b, Gunnar PIHO^a and Peeter ROSS^a ^aDepartment of Health Technologies, Tallinn University of Technology, Estonia ^bDigital Grammars AB, Sweden ORCiD ID: Kristian Juha Ismo Kankainen https://orcid.org/0000-0002-0551-927X, Inari Listenmaa https://orcid.org/0000-0001-5555-0251, Gunnar Piho https://orcid.org/0000-0003-4488-3389,

Peeter Ross https://orcid.org/0000-0003-1072-7249

Abstract. Written text has been the preferred medium for storing health data ever since Hippocrates, and the medical narrative is what enables a humanized clinical relationship. Can't we admit natural language as a user-accepted technology that has stood against the test of time? We have previously presented a controlled natural language as a human-computer interface for semantic data capture already at the point of care. Our computable language was driven by a linguistic interpretation of the conceptual model of the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT). This paper presents an extension that allows the capture of measurement results with numerical values and units. We discuss the relation our method can have with emerging clinical information modelling.

Keywords. Data capture, controlled natural language, semantic health data, SNOMED CT, information model

1. Introduction and Method

Modern clinical terminologies are undoubtedly expressive, essentially due to compositionality by post-coordination. For example, the SNOMED CT includes 432 person concepts and 52,181 causative agent concepts (as of release 2022-11-30). Consequently, the two concept groups can be composed to form more than 22 ½ million postcoordinated expressions. Nonetheless, a recent scoping literature review concluded that this expressiveness is rarely used and that "there is no easy solution for mapping free text to this terminology and to perform automatic post-coordination" [1].

We consider this lack of mapping to be due to the lack of a good user interface and have earlier proposed (detailed in [2]) a novel mapping in the form of a computable controlled natural language (CNL) that corresponds to a linguistic interpretation of SNOMED CT's *Situation With Explicit Context*. We insist a language-based user interface is appropriate for data capture as the written text has been the medium for health data ever since Hippocrates, and language is what enables a humanized clinical relationship [3].

¹ Corresponding Author: Kristian Juha Ismo Kankainen, E-mail: kristian.kankainen@taltech.ee.

We use Grammatical Framework (GF) [4] to program our CNL. Our GF grammar consists of one abstract application grammar and concrete syntaxes for Estonian, English, and SNOMED CT compositional grammar. The rule-based approach of GF supports explainability [5], which we see as crucial in healthcare.

2. Results and Discussion

A new concept *ObservedCondition* was added to our abstract grammar. This new type takes a numerical quantity, an evaluation procedure, and an observable entity as input parameters. Both latter are ordinary SNOMED CT objects. Although SNOMED CT has introduced support for numeric values for medications [6], it is not able to express numerical measurement results. As we cannot map numerals to SNOMED CT, we have instead started to explore mapping to clinical information models. Furthermore, our extended abstract model now complies with ISO 13940 ContSys as "a number of healthcare investigations reveal a number of observed conditions" [7].

Mapping to clinical information models makes sense to us in two ways. Grounding the necessary modeling work on the language used for data capture, can enable the consensual development of reusable domain information models [8]. As an example, our presented extension employs information about the unit encoded in the observable entity object. Although this encoding follows the Snomed CT concept model, it is little used, as only 23 terms specify a unit in the 2022-11-30 international release. Arguably, contextual information, like unit, could be managed in a national extension or repository.

Accommodating the data capturer's view can also improve user experience [9]: GF enables actionable text (data) entry. As GF translation functions in both ways, it can render already captured data as text, which can further allow multilingual authoring of the same health record by several actors across healthcare borders.

References

- Gaudet-Blavignac C, Foufi V, Bjelogrlic M, Lovis C. Use of the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) for Processing Free Text in Health Care: Systematic Scoping Review. Journal of Medical Internet Research. 2021 Jan;23.
- [2] Kankainen K, Klementi T, Piho G, Ross P. Using SNOMED CT as a Semantic Model for Controlled Natural Language Guided Capture of Clinical Data. CEUR Workshop Proceedings. 2022 Aug;Vol-3264.
- [3] Miranda-Bastidas C. Medical records: the medical writing of the patient's story to the medical narrative. Colombia Medica. 2020 Jun;51.
- [4] Ranta A. Grammatical framework: programming with multilingual grammars. Studies in computational linguistics. CSLI Publications; 2011.
- [5] Ranta A. Explainable machine translation with interlingual trees as certificates. In: Proceedings of the Conference on Logic and Machine Learning in Natural Language (LaML 2017); 2017. p. 63-78.
- [6] SNOMED International. SNOMED International delivers technical preview for enhancement of SNOMED CT with concrete domains in advance. 2021. Available from: https://www. snomed.org/news-and-events/articles/snomed-delivers-technical-preview-concrete-domains.
- [7] ISO 13940:2015(en), Health informatics System of concepts to support continuity of care.
- [8] Oemig F, Blobel B. The Value of Domain Information Models for Achieving Interoperability. pHealth 2020. 2020:75-82. IOS Press.
- [9] Jansson M, Liisanantti J, Ala-Kokko T, Reponen J. The negative impact of interface design, customizability, inefficiency, malfunctions, and information retrieval on user experience: A national usability survey of ICU clinical information systems in Finland. International Journal of Medical Informatics. 2022 Mar;159.