MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190202

Building an Experimental German User Interface Terminology Linked to SNOMED CT

David Hashemian Nik[a], Zdenko Kasáč[a], Zsófia Goda[a], Anita Semlitsch[a], Stefan Schulz[a,b]

[a]*Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria*
[b]*Averbis GmbH, Freiburg, Germany*

Abstract

We describe the process of creating a User Interface Terminology (UIT) with the goal to generate a maximum of German language interface terms that are mapped to the reference terminology SNOMED CT. The purpose is to offer a high coverage of medical jargon in order to optimise semantic annotations of clinical documents by text mining systems. The first step consisted in the creation of an n-gram table to which words and short phrases from the English SNOMED CT description table were automatically extracted and entered. The second step was to fill up the n-gram table with human and machine translations, manually enriched by POS tags. Top-down and bottom-up methods for manual terminology population were used. Grammar rules were formulated and embedded into a term generator, which then created one-to-many German variants per SNOMED CT description. Currently, the German user interface terminology contains 4,425,948 entries, created out of 111,605 German n-grams, assigned to 95,298 English n-grams. With 341,105 active concepts and 542,462 (non FSN) descriptions, it corresponds to an average of 13 interface terms per concept and 8.2 per description. An analysis of the current quality of this resource by blinded human assessment terminology states equivalence regarding term understandability compared to a fully automated Web-based translator, which, however does not yield any synonyms, so that there are good reasons to further develop this semi-automated terminology engineering method and recommend it for other language pairs.

Keywords:

Natural Language Processing, Systematized Nomenclature of Medicine, Translations

Introduction

In medical documentation, user interface terminologies (UITs) bridge between the language in use by clinicians and the standardised language of reference terminologies. Interface terms vary between specialties, professional groups and dialects and undergo constant evolution due to medical progress and dynamics of language [1]. We present a cost-effective, manual, incremental approach to acquire and create interface terms with the goal to map them onto a reference terminology (SNOMED CT), both using bottom-up and top-down approaches.

Health care terminologies are crucial resources for semantic interoperability. Despite their differences in architecture and scope (e.g. ICD-10 as a classification system for diseases vs. SNOMED CT as an ontology for all aspects of electronic health record (EHR) content), there is a principal distinction between:

- Labels (fully specified names, reference terms), which aim at providing self-explaining descriptions of the meaning of concepts, often paralleled by formal definitions in ontologies or text definitions in thesauri.
- Interface terms (non-preferred synonyms), which represent the jargon used by clinical practitioners in their daily documentation and communication.

Take the SNOMED CT preferred term "*Primary malignant neoplasm of lung*" as an example. This term is precise but artificial. Screening the complete PubMed corpus yields no single occurrence of this term, and it would be very unlikely be found in clinical documents either. Its popular synonym "*lung cancer*", retrieves 120,682 documents from PubMed and is common in clinical documents and problem lists. Another example is "EKG", an acronym-term we retrieved 12,208 times in a corpus of cardiology discharge summaries, while the full term "Elektrokardiogramm" does not occur a single time. Reference terms are characterised by precision; interface terms by brevity. This highlights the need for interface terms, wherever content is automatically extracted from texts [2]. This has recently been emphasised by the European project ASSESS CT (Assessing SNOMED CT for Large Scale eHealth Deployments), which recommended broad efforts to be invested into UITs [3] linked to reference terminologies like SNOMED CT, rather than into translations proper. For humans, the possible fuzziness and ambiguity of interface terms (in this case, it may not be quite clear whether lung metastases are in the scope of "lung cancer") is a minor problem due to context and implicit understanding within a user group, whereas word sense disambiguation is still a major problem for machines.

A major use case for UITs is the provision of dictionary entries for natural language processing (NLP) systems. Other use cases are related to structured data entry using data acquisition forms, whenever the terms should be close to the user's language preferences. Limitations of UITs lie in the conceptual content of the underlying reference terminology (in this case SNOMED CT). There is still a substantial amount of interface terms that cannot be precisely mapped to terms of a corresponding domain terminology.

Interface terms can be single words like "pancreas" or "dermatology", compound words like "lymphangiosarcoma", acronyms like "ARDS" or multi word terms like "hereditary factor VIII deficiency disease". Possible sources of interface terms include:

- Automatic and manual term translations
- External, generally accessible corpora of the target language (e.g. books, publications, articles etc.);
- Institution-specific value sets and term collections;
- Clinical corpora constituted by EHR narratives (privacy protection must be taken into account).

Lexical ambiguity (e.g. "delivery" in "drug delivery" vs. "delivery of a baby") is characteristic for interface terms, and it rarely runs parallel between languages (e.g. in German, words for the delivery of goods, substances or babies are completely distinct). Especially short acronyms are prone to ambiguities (e.g. "MI" means "myocardial infarction" in cardiology, but "macular ischemia" in ophthalmology; "DM" may be expanded to "diabetes mellitus", "diameter", "disease management", or "dermatomyositis" according to its context. Only long acronyms like "NSTEMI" can be expected to have a unique meaning. Ambiguous acronyms are mostly unproblematic once they occur as constituents of a longer term, in which they are univocal (example: "type 2 DM").

In the following we describe the ongoing construction of a German UIT for SNOMED CT, which uses the English SNOMED CT description table for the semi-automated construction of German-language interface terms.

Material and Methods

SNOMED CT

SNOMED CT [4] is the most comprehensive healthcare terminology on the globe, with use in over 50 countries. The current international SNOMED CT version (July 2018) consists of 340,659 active representational units, known as *SNOMED CT concepts*. They have an exclusive meaning and a unique machine-readable identifier. *SNOMED CT Descriptions* include a *Fully Specified Name* (FSN) for each concept, together with one or more synonyms. FSNs represents the concept with an official, ideally self-explaining name. Synonyms share meaning with FSNs but find more common use in displaying or selecting a desired FSN. The meaning of SNOMED CT concepts is, in addition, delineated by text definitions and formal axioms in Description Logics. E.g., *Gastritis* is defined as logically equivalent to a disorder with inflammatory morphology that is located at some stomach structure.

SNOMED CT's increasing adoption in German-speaking countries (with Switzerland and Austria already being members of SNOMED International, whereas Germany is still negotiating) prioritises harmonisation with local language and terminologies. This has motivated the authors to develop a semi-automatic, resource-aware and pragmatic method for creating an experimental German UIT linked to the current international version of SNOMED CT. The methodological framework developed might be applied to other languages [5]. This UIT feeds the currently largest SNOMED CT use case in the German-speaking country, *viz.* semantic annotations of clinical texts within the hospital information system of the Austrian healthcare provider KAGes, using natural language processing [6].

Procedure

The starting point of the interfacing process was the set of over 700,000 English SNOMED CT terms ("Descriptions") from the international version. This list contains, apart from the official labels (Fully Specified Names), one preferred term (dependent on the release) and zero-to-many synonyms. SNOMED CT terms are characterised by many repetitive substrings, exemplified by the term "magnetic resonance imaging of hip". That there are 1,620 occurrences of "magnetic resonance imaging" and 1,239 occurrences of "of hip" in SNOMED CT shows the potential of a modular approach to term translation.

N-gram Table as Core Translation Resource

In order to harvest such repetitive passages, a language-specific rule set was created and implemented in Python 3 to chunk decomposed terms down into shorter units, constituting word *n*-grams with *n* ranging from *1* to *6*. Here "word" encompasses, to a minor extent, also sub-word entries needed for single-noun composition, which is particularly common in German. SNOMED CT terms therefore can be constituted by one to many n-grams, e.g. "Escherichia coli" is a 2-gram, which can equally stand alone, or occur in a longer term like "Escherichia coli antibody". In addition to the n-grams with $n > 1$, also all single words were added. In summary, n-grams include complete and partial constituents of the SNOMED CT labels and descriptions. The result was an .xlsx table with currently enclosing close to 550,000 English-language n-grams, ranked by their decreasing frequency in the source, i.e. the SNOMED CT's description table. It is structuredin the following collumns:

- *"ID"* – n-gram identifier code;
- "N-gram English": from words to n-grams, ($n < 7$);
- *"Length"*: number of tokens in the n-gram (n);
- *"Count"*: n-gram frequency in the source SNOMED CT description file;
- *"N-gram German 1"*, *"N-gram German 2"...* German translations of English n-grams.

We started populating this table with German terms, started four years ago with limited resources (one part-time terminologist and several medical students). Along time it has been subject to constant optimization and quality improvement heuristics. The description of some of these heuristics (mostly developed as a sequence of trial-and-error cycles) will occupy the remainder of this section. The main goal of this description is to outline a general methodology for modular and incremental CIT developments for new language with limited resources.

Translation Heuristics

The first step was to tackle large amounts of easily machine translatable content. Method of choice was Google Translate. For surprisingly many more common medical terms, for example *"arm", "status"*, or *"hormone",* but also *"cholecystectomy"* or *"tendon sheath",* useful translated content could be harvested. Several limitations were observed: (i) each translation only yielded one target n-gram (no synonyms are provided when inserting source terms in a batch), (ii) single English words were often undefined in terms of POS (part of speech, e.g. noun, adjective, verb), like *"set", "back", "general"* and therefore ambiguous, (iii) the translation of numerous words did not correspond to the medical meaning (e.g. *"delivery"* translated into *"Lieferung"* and not into the translations relevant to medical context like *"Entbindung"* or "Gabe"), (iv) less common medical or chemical terms were not translated, and the untranslated term was returned, instead (which then was preserved, waiting for manual correction).

Given SNOMED CT's richness in concepts for biological organisms, for which Latin terms are common and identical across languages, we harvested a large number of terms that occurred identically in the English and Spanish versions, assigned to the same concepts (e.g. *"Homo sapiens", "Ascaris lumbricoides", Angina pectoris", "Anorexia nervosa"*). Thus, they could be safely added to the German version.

For the left-over of untranslatable content, manual input was required. It required mainly addition of new German n-grams and modification of the existing human or machine generated ones, addition of synonyms (e.g. *"Leber- " to "hepatisches"*), as well as appending grammar tags to both manual and machine created words (e.g. *"hepatisches|JJ"* to mark it as an adjective and "-" to mark *"Leber-"* as a prefix-like morpheme). Humans also revised the machine generated terms by evaluating randomly sampled sets of the generated terms, often revealing major systematic errors in the machine translation. Mistranslations are deleted and revised, untranslated terms grouped by certain patterns (e.g. ending with "acid" and changed to the German ending "säure" by repetitive search / replace actions). Manual work was prioritized according to the guiding principles: (i) n-gram frequency; (ii) clinical relevance; and (iii) single words.

To ensure all SNOMED CT concepts get covered in one primary form, we set the goal of at least one translation for every single word n-gram (n=1), Such atomic units represent individual interface terms or parts of larger composite interface terms. Semantic composition is not robust multiword n-grams. Therefore, we considered multi word n-grams down to a frequency of eight. Less frequent n-grams were left out, because of the enormous amount of multiword n-grams. The decision in favour of manual translating multiword terms was positive whenever:

- The formation of a multiword term in the source language was not paralleled by a translation in the target language. E.g. "*malignant neoplasm*" translates not only into "maligne Neoplasie" but also into "*Malignom*"; and "oral solution" does not translate into "*mündliche Lösung*", but into "*Lösung zur oralen Einnahme*".
- One component of a multiword term was ambiguous and could be disambiguated in the composition. E.g. "*back pain*" in "*Rückenschmerzen*" ("back" = "Rücken"), and "*back door*" into "*Hintertür*" ("*back*" = "*Hinter-*").

The application of these rules often results in a labour-intensive walkthrough of the whole n-gram table.

Addressing German Language Features Like Inflection and Composition

German word inflections heavily depend on gender, tense, person, number, declination type and case. All entries, be they manual or automatically made, need therefore to be reviewed and enriched with grammatical information. This task relies mainly on human editing, using however machine support for pattern-based search-replace actions. Automated term creation with correct inflection suffixes requires a distinction between nouns, adjectives, verbs, prepositions and determiners, yielding the following tagging suffixes:

- Singular gender or plural tags for nouns (|NN|N, |NN|F, |NN|M, NN|P)
- Cases for verbs (|VV|N, |VV|G, |VV|D, |VV|A)
- Markers for adjectives (|JJ)
- Preposition tags depending on case (|PP|N, |PP|G, |PP|D, |PP|A)
- Definite and indefinite articles (|DET|D, |DET|I)
- Noun tags of Latin words or other nouns for which single-noun composition is not allowed (|NL|N, |NL|F, |NL|M, |NL|P)

Tags can be omitted in multi word entries where, e.g. due to a preposition, the inflection case is already determined.

In addition, all numbers were substituted by placeholders (ð, ðð, ððð etc. depending on number of digits and decimals) in the n-gram table. (While the decimal denominator in English is point in German comma is used.)

Single word composition is an important feature of the German language, e.g. with "[Fracture] of the big toe" translated into "*Großzehen[fraktur]*". Such composition patterns are characterised by inversion (with the head morpheme trailing) and by inconstant addition of infixes to the modifying morphemes. This requires the addition of translations to English prepositional phrases starting with "of", which not only translates to a structurally similar German genitive construction ("der Großzehe"), but also to morpheme-like prefixes like "Großzehen...", tagged by bracketing underscores ("_Großzehen_"). In certain instances, we use the additional tag "%VOID%". "%VOID%" is used as an empty word and can be suffixed with tags and thus gaining an inflection attribute. It also prohibits noun compositions (which are otherwise standard between neighbouring nouns). The German n-gram "zur Benutzung%VOID%|PP|G" (to use) illustrates how %VOID% can avoid unwanted word composition, while enforcing the next phrase set into genitive case. If "%VOID%" were absent, the German word "Benutzung" would potentially be attached to a noun right to it, resulting in a disallowed nominal compound. Other tags are %SWAP% and %RIGHT%, with the former swapping the right and the left portion of a phrase (due to different word order in German), and the latter putting a word to the end of a phrase, typically a separable verb prefix, e.g.: "stops" "hört mit|PP| %RIGHT% auf".

Top-down and Bottom Up Term Harvesting

Both top-down and bottom-up term harvesting has been used. Top-down means an intellectual effort of enriching English n-grams with additional translations. E.g., when encountering the English term "diabetes mellitus" (together with the identical German term) the editor might remember that there is a synonym "Zuckerkrankheit" to be added. One problem of this method is that it requires extensive manual editing, and is only performed when a manual reviewer goes through the list checking for missing synonyms (starting at very frequent and relevant terms), requiring an excellent command of medical terminology.

Bottom-up describes lexicon population starting with content from representative corpora (clinical texts, biomedical literature, existing terminologies in the target language), from which terms are extracted and mapped to corresponding n-grams in the source language. The advantage of this method is the adjustment of the interface term vocabulary to preselected document genres, covering specific jargon. N-gram hit lists (ordered by decreasing frequency) extracted from these corpora are excellent sources of commonly used terms in the context that is to be covered. One source prepared for our project was a hit list with 4,000 de-identified dermatology summaries, which resulted in 24,000 n-grams ($n \geq 3$). After filtering out entries already included in the master n-gram table (and their inflectional variants), terminologists walk through the hit list and try to identify a corresponding English term for each entry in the master n-gram table.

In cases of doubt, translations are marked for further analysis and discussion. Translation resources such as and offline dictionaries, online translation tools (web search for usage frequency and contexts, various online dictionaries, Linguee, Oxford German Dictionary, DUDEN Wissennetz), term clusters retrieved from the UMLS metathesaurus as well as German titles of Medline-indexed citations (marked as [tt] in the Medline records) have proven useful.

Tooling

Appropriate tooling for distributed, cooperative terminology work is still a desideratum. N-gram editing and maintenance is currently done via a Microsoft Excel table, shared via a repository, which assures that only one instance of this table can be edited at a time. This Excel n-gram was enriched by macros written in the VBA script language. E.g. the "exploration macro" and "modify and add macro". The former retrieves already existing word and sub-term translations for a given n-gram. The addition of synonyms is supported by a "*Modify and Add Macro*". It allows batch additions of synonyms. When entering an n-gram (e.g. "*Neoplasie|NN|F*") with a synonym (e.g. "*Neubildung|NN|F*") it iterates through the complete n-gram list and suggests additions (e.g. "*Basalzellen-Neubildung*" to "*Basalzellen-Neoplasie*"). Other macros are used for plausibility checking, e.g. identifying disallowed tags.

Term Generation

Term generation is done with a series of language specific Python 3 scripts. In a nutshell, they take the chunked English description table, translate each chunk by n-gram table lookups, thus collecting n-gram translations and assembling them into complete SNOMED CT description translations. In this process, the tagged words are interpreted by a variant generator, which generates inflectional variants, following an inbuilt noun phrase grammar. It also produces single word compounds, using specific tags and rules as explained above. The problem of this generative approach is its combinatorial explosion: if a SNOMED CT term consists of four n-gram chunks, each of which with 3 translations, $3^4 = 81$ terms are generated. To mitigate the growth of generated terms, as well as to improve quality we pursued two approaches. To increase translation of longer n-grams, in order to avoid uncommon term associations, and complete translations of very long terms, which are unlikely to occur in clinical texts, such as FSNs resembling textual definitions[1].

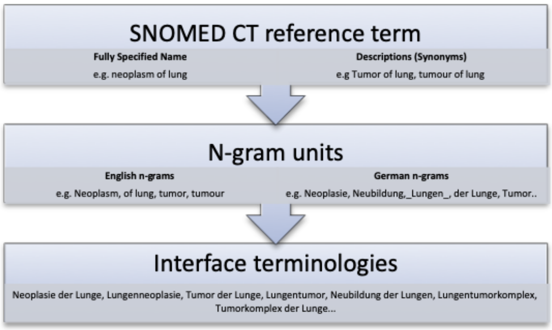

Figure 1 –Interface Term Generation

Quality Checks and Updates

Due to the large size of the UIT, only random samples are regularly quality-checked, assuming that frequent systematic errors also surface in these samples. Errors are traced back, either to the n-gram table or to the inflection and composition routines in the Python scripts. A challenge is also the biannual update of the n-gram table, together with the resulting interface term list, when a new international SNOMED CT version is released. Changes include addition, alteration and removal of concepts. During each version release, we have observed a growth of the n-gram resource. This opens the opportunity to fill in terms, e.g. more detailed pharmacological ingredients that previously were missing.

Validation Study

The last 2018 version of the UIT was used for a blinded validation study. 200 concepts were randomly selected, and for each of them, one of the German UIT entries – together with the English description in original from – was randomly selected. For each English description, an alternative translation was created using DeepL [7], a machine translation website based on neural networks. For each of three terminologists, a custom translation list with 200 term translations, 100 from DeepL, and 100 from the German UIT was generated. The selection was done by chance, and the terminologists were not informed about the source. For each translation four pieces of information were required: (i) assessment of *Content comprehensibility* (regardless of style and grammar), (ii) *Grammar* (regardless content errors or bad word choices), (iii) *Style & Spelling*, for each of which using a five-point Likert scale (1 - very bad, 5 very good). Finally, an "ideal" manual translation for the term was added by each terminologist.

Results

The German User Interface Terminology currently encompasses 4,425,948 entries, automatically generated from a core vocabulary of 111,605 German n-grams, assigned to 95,298 English n-grams. With 341,105 active concepts and 542,462 (non FSN) descriptions, this corresponds to an average of 13 interface terms per concept and 8.2 per description.

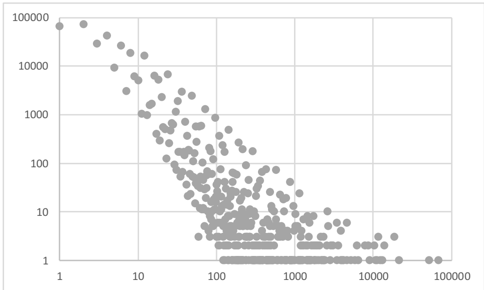

Figure 2 –Distribution of Synonym Sets in UIT. Logarithmic axes:x-Axis: size of synonym set, y-axis: frequency of sets of this size

Fig. 2 shows the distribution: most sets have one to 10 members, but there were outliers with more than 10,000 members, due to term compositionality. Table 1 provides the rating results.

Table 1 –Likert Scale Rating of Translations (Arithmetic Means of 300 Human Ratings per Category)

SNOMEC CT User Interface terminology (UIT)			**DeepL machine Translator**		
Content	Grammar	Spelling / Style	Content	Grammar	Spelling / Style
4.6	4.7	4.4	4.6	4.9	4.6

These results need to be interpreted in the light of the fact that DeepL provided only one translation per concept, whereas the UIT produced much more (mean: 12.1, median 3 translations per term). Whereas the DeepL result can be assumed to correspond to the most popular translations, our sampling

[1] Such as "pT2: Tumor invading two subsites in a single region or extending to involve an adjacent region within the nasoethmoidal complex, with or without bony invasion (nasal cavity and ethmoid sinus) (finding)"

algorithm had picked out the UIT completely randomly. However, the results show clear deficits regarding grammar, spelling and style issues of the current state UIT. Finally, a case insensitive, spelling variation tolerant match between the translation suggested by the users and the UIT entries occurred in 299 of 600 cases.

Discussion

Is it worthwhile investing in terminology acquisition as well as UIT maintenance, in the light of ever increasing performance of machine translation tools? Already in 2013, we had found rather surprising translation results proposed by Google Translate [8]. We suggested a combination of translations done by medical students and done by machine translations; the method we have been using in the described UIT project. Still, machine translation systems have problems with the generation of compound nouns, as well as with the production of sufficient numbers of term variants and combinations. Also, their power depends on the amount of training data they are fed; good results can therefore not be expected for rarely used terms or languages that have less content on the Web.

We are currently supporting the construction of an interface terminology for Portuguese, using the Spanish version from SNOMED International as source language. For this language pair, the power of web-based machine translation systems seems to be much poorer than for German / English.

The advantage of interface terminologies is obvious when re-structuring narrative content of the EHR in terms of SNOMED CT. This opens up the ability to be linked with aggregation terminologies, such as the International Classification of Diseases (ICD-10 or the upcoming ICD-11) [9]. Apart from classification systems of diseases and procedures, such semantic standards are rarely used in routine documentation. As long as interface terminologies act in the background, where they are only used by text mining software, their size and their tendency to over-generate term variants (most of which are never found in any text) do not constitute a problem. This changes if interface terms are to be used by humans; here, uncommon terms should be filtered out before use in order to limit the variety of terms per concept. Such filter criteria could use large clinical corpora together with a general corpus such as Wikipedia and require that a generated term was used at least once. This may, however, obviate the translation of long artificial terms like the one cited above, which are unlikely to be found in any clinical document.

Extracting them from a narrative can be thought of as a result of fuzzy matching between the narrative and multiple term candidates, an approach to be exploited, especially with new, powerful word embedding methods and deep learning.

Another route for translating medical terms capitalises on the fact that large parts of them derive from Latin or Greek roots and share a regular morphology. Such terms can be automatically translated by rewriting rules [10,11]. A machine learning process can learn from two language sets, identify letter-based patterns and apply these rules for the automatic translation of new terms. The word "bronchoscopy" ending with "-oscopy" would be such an example that could be detected by this method and automatically translated by inferring rewriting rules ("Bronchoskopie"). The accuracy of this method depends on how many pairs of terms get compared.

Conclusions

We presented an approach to acquire interface terms from SNOMED CT reference term translations and data from various clinical corpora and map these interface terms onto SNOMED CT reference terms. During the process we focused on frequent words both in SNOMED CT and in clinical texts with the goal to cover all commonly used medical terms and their synonyms, integrating them into our system. An analysis of the current quality of the UIT by blinded human assessment terminology states equivalence regarding term understandability compared to a fully automated Web-based translator. This tool, however, yields much less synonyms, so that there are good reasons to further develop our semi-automated technology and recommend it for other language pairs.

References

[1] S. Schulz, Building an experimental German user interface terminology linked to SNOMED CT, *SNOMED Expo*, Bratislava, Slovakia, October 2017. https://goo.gl/cHgvmo [accessed on March 31st, 2019]
[2] S. Schulz, "Using language technology for SNOMED CT localization (poster), *SNOMED Expo*, Montevideo, Uruguay, October 2015. https://goo.gl/z2aCX1 [accessed on March 31st, 2019]
[3] D. Kalra, S. Schulz, D. Karlsson, R.V. Stichele, R. Cornet, K.R. Gøeg, G. Cangioli, C. Chronaki, R. Thiel, S. Thun,and V. Stroetmann, *ASSESS-CT recommendations*, 2016, *http://assess-ct.eu/fileadmin/assess_ct/final_brochure/assessct_final_brochure.pdf* [accessed on March 31st, 2019]
[4] SNOMED International. *SNOMED CT Starter Guide*, 2018,
[5] S. Schulz, J.M. Rodrigues, A. Rector, and C. Chute, Interface terminologies, reference terminologies and aggregation terminologies: a strategy for better integration, *Studies in Health Technology and Informatics* **245** (2017), 940-944.
[6] M. Kreuzthaler, C. Martínez-Costa, P. Kaiser, and S. Schulz, Semantic technologies for re-use of clinical routine data, *Studies in Health Technology and Informatics* **245** (2017), 539-543.
[7] DeepL translator, https://www.deepl.com/en/translator [accessed on March 31st, 2019]
[8] S. Schulz, J. Bernhardt-Melischnig, M. Kreuzthaler, P. Daumke, and M. Boeker, Machine vs. human translation of SNOMED CT terms, *Studies in Health Technology and Informatics* **192** (2013), 581-584.
[9] M. Mamou, A. Rector, S. Schulz, J. Campbell, H. Solbrig, and J.M. Rodrigues, ICD-11 (JLMMS) and SCT Inter-Operation, *Studies in Health Technology and Informatics* **223** (2016), 267-272.
[10] V. Claveau, Translation of biomedical terms by inferring rewriting rules. *Information Retrieval in Biomedicine: Natural Language Processing for Knowledge Integration*, IGI - Global. 2009, Chap 6.
[11] J.A. Miñarro-Giménez, J. Hellrich J, and S. Schulz, Acquisition of character translation rules for supporting SNOMED CT localizations. *Studies in Health Technology and Informatics* **210** (2015), 597-601.

Address for correspondence

David Hashemian Nik, david.hashemian-nik@stud.medunigraz.at