MEDINFO 2017: Precision Healthcare through Informatics
A.V. Gundlapalli et al. (Eds.)

doi:10.3233/978-1-61499-830-3-539

Unsupervised Abbreviation Expansion in Clinical Narratives

Michel Oleynik[a], Markus Kreuzthaler[a,b], Stefan Schulz[a]

[a] *Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria*
[b] *CBmed GmbH, Center for Biomarker Research in Medicine, Graz, Austria*

Abstract

Clinical narratives are typically produced under time pressure, which incites the use of abbreviations and acronyms. To expand such short forms in a correct way eases text comprehension and further semantic processing. We propose a completely unsupervised and data-driven algorithm for the resolution of non-lexicalised and potentially ambiguous abbreviations. Based on the lookup of word bigrams and unigrams extracted from a corpus of 30,000 pseudonymised cardiology reports in German, our method achieved an F_1 score of 0.91, evaluated with a test set of 200 text excerpts. The results are statistically significantly better ($p < 0.001$) than a baseline approach and show that a simple and domain-independent strategy may be enough to resolve abbreviations when a large corpus of similar texts is available. Further work is needed to combine this strategy with sentence and abbreviation detection modules, to adapt it to acronym resolution and to evaluate it with different datasets.

Keywords:

Natural Language Processing; Electronic Health Records.

Introduction

Clinical narratives are typically produced under time pressure. Besides incomplete sentences and typing errors, a typical manifestation of this is the tendency towards short forms like acronyms and abbreviations [1–3]. Their widespread use complicates further semantic processing and makes text understanding difficult, not only by laypersons, but also by clinicians from different disciplines or professional groups. In the context of large-scale concept mapping of medical texts for secondary use [4], the first processing steps are commonly devoted to text cleansing. This involves the correction of misspellings, the identification of sentence delimiters, and the expansion of short forms.

The use of abbreviations is especially pronounced in agglutinative languages such as German, in which single words are often formed by composition of morphemes, normally stems and affixes. The longer such a word grows, the more easily its completion can be guessed. Typing under time pressure therefore tends to stop close to the ending and an abbreviation marker is set. Thus, so-called ad-hoc abbreviations are created. Interestingly, the trailing period, required by German grammar to mark abbreviations, is mostly set, even in non-standard clinical narratives. It is also important to note that German requires capitalisation of all nouns, not only proper names, a rule commonly followed in clinical notes.

In combination, these characteristics lead to several NLP challenges related to the ambiguity of the period mark, which can be: (a) the trailing character of an abbreviation; (b) a sentence delimiter; or (c) both. Sentence delimitation, tokenization, and abbreviation resolution are, therefore, hardly separable tasks.

Regarding abbreviation resolution, three specific tasks are commonly listed: abbreviation detection, sense detection, and sense disambiguation. While the first is related to the disambiguation of the period character, the second and third relates to the sense expansion and its disambiguation, respectively.

Our recent work has focused on the first task, *viz.* abbreviation detection, first with a supervised approach based on support vector machines [5] and then with an unsupervised strategy that combined co-occurrence information with a large abbreviation-free domain dictionary [6]. The current paper will focus on the expansion (second task) and disambiguation (third task) of so-called period abbreviations, i.e. tokens that include a period mark as their rightmost character. The proposed approach is completely unsupervised.

Period abbreviations are clearly distinguished from acronyms, which consist mostly of upper case letters, never end with a period, and often represent multiword terms (e.g. "MI" for "myocardial infarction"). They are also distinct from other abbreviations that do not end with a period (e.g. "Ca", which stands both for "Cancer" and "Calcium").

Period abbreviations have, therefore, the following characteristics: (i) they abbreviate mostly single words; (ii) their first character always coincides with the first character of the word they abbreviate; (iii) in most cases, the string left of the period equals a string of characters from the left side of the abbreviated word.

Moreover, we distinguish between lexicalized period abbreviations and ad-hoc period abbreviations. Whereas the left substring rule can normally be taken for granted for the latter ones, lexicalized period abbreviations can be easily expanded via a lexicon. We also distinguish between unambiguous and ambiguous period abbreviations. General and domain specific dictionaries often list more than one sense for an abbreviation, especially for very short ones. Longer, ad-hoc period abbreviations are normally unambiguous. Conversely, single letter abbreviations like "E." in "E. coli" for "Escherichia coli" are almost always ambiguous.

The following work is limited to period abbreviations that cannot be unambiguously resolved by lookup in a general or medical domain lexicon. We also do not consider the special case where two period abbreviations are glued together thus forming one token, e.g. "St.p." = "St. p." = "Status post" (Latin for "history of"). It furthermore ignores the resolution of ordinal number expressions, for which in German the use of the period character is mandatory, such as "2." for "2nd". Consequently, the focus of our investigation is only on non-lexicalised, supposedly ad-hoc abbreviations.

Our hypothesis is that a high efficiency of abbreviation expansion can be obtained in a fully unsupervised fashion, i.e. without the (often considerable) effort of producing manually annotated training data. We thus hypothesize that if a clinical collection is sufficiently large, all knowledge needed is present therein.

Materials and Methods

Cardiology corpus

We tested our approach on a corpus of 30,000 pseudonymised discharge summaries from the cardiology department of the Graz University Hospital, the second largest hospital in Austria. The documents were written by German-speaking physicians. Figure 1 shows a typical sample from our corpus.

```
1) Stark vergrößerter li. Ventrikel mit
dünner Wand und höhergr. reduz. systol.
Fkt.; EF: ca. 25- 30%; diffuse
Kontraktilitätstörung; Schaukelbewegung des
LV bei LSB.
2) Li. Vorhof leicht vergrößert;
3) Re. Vorhof leicht- bis mittelgr.
vergrößert;
4) Re. Ventrikel normal groß, normale
Rechtsventrikelfunktion;
5) Aortenklappe morphol. und funktionell
normal;
6) Mitralklappe morphol. unauff.; mittelgr.
MINS; als Mechanismus der MINS besteht eine
Ringdilatation;
7) Trikuspidalklappe unauff.; leichte
TRINS; systol. Pulmonalisdruck ca. 45 mmHg;
```

Figure 1 – Example snippet with ad-hoc abbreviations, e.g. "systol." for "systolisch" and "Fkt." for "Funktion"

We first split the corpus into 90% for training and 10% for testing. Two hundred random substrings (100 characters) centred around a period character (followed by a single space) were extracted from the training corpus as source for building a validation set, in which 147 (73.5%) valid abbreviations were manually expanded by the third author, a physician. The period marks in the remaining substrings were considered out of scope for this work (see Introduction for our scope definition). Successive experiments were performed with this set. For the final evaluation, a set of 301 text snippets from the test set was used, of which 200 (66.4%) were considered valid abbreviations. The surrounding context, together with the original text corpus, allowed for unambiguous manual expansions in all cases. We report results from both sets.

N-gram lookup lists

Two frequency tables were created out of the tokenised training corpus, *viz.* a unigram list ***U*** and a bigram list ***B***. The tokenization process ignored period characters, so that they are included in the tokens. Also, as capitalization is an important feature in German and properly used in our corpus, the n-grams were *not* normalised. The lists were arranged in decreasing order of frequency. Corpus frequencies were calculated for all 155,801 token types and all 803,243 bigram types.

Abbreviation and expansion assumptions

An expansion E is a valid expansion of an abbreviation A (with its abbreviation mark, i.e. the final "." character, stripped) if:

- E does not end with an abbreviation mark (".");
- E is at least one character longer than A;
- E has only alphabetic characters.

Additionally, we define the relative gain $G(A,E)$ as the ratio of the length difference to the abbreviation length, as seen in Equation 1. The intuition behind this restriction comes from the observation that longer words are rarely abbreviated by overly short strings, as this may lead to ambiguity[1]. Nonetheless, they are seldom abbreviated by very long forms either, because the "economy" of using an abbreviation mark would be minimal.

$$G(A,E) = \frac{Length(E) - Length(A)}{Length(A)} \qquad (1)$$

Thus, we add extra assumptions regarding relative gains:

- $G(A,E)$ is greater than 0.01;
- $G(A,E)$ is lower than 6.

Finally, two matching types are distinguished, one of which must be true:

- Strict matching: A is a left-sided substring of E;
- Relaxed matching: all characters of A are contained in E in the same order. The first character of the abbreviation equals the first character of the full form.

While the strict matching accounts for the general case (e.g. "maximal" abbreviated as "max."), its relaxed version allows the correct matching of "*Tbl.*" to its expanded form "*Tablette*".

Resolution strategy

Our resolution strategy is based on the co-occurrence of adjacent tokens where one of the tokens appears both in shortened and expanded forms in the corpus. As an example, "A. subclavia" can be correctly resolved to "Arteria subclavia", as the latter is the most common expanded form that matches the abbreviation prefix (see Figure 2). To this end, the bigram frequency list is used. Only when the right or left context does not provide any valid expansions, the unigram frequency list is looked up.

#	Word pair
35	A. subclavia
23	Vena subclavia
12	Arteria subclavia
5	Art. subclavia
4	A subclavia

Figure 2 – Bigram lookup strategy

Considering an abbreviation A, its left context token L and its right context token R, we propose the following algorithm to guess its expanded form E. Every lookup is performed in sequence, hereafter denominated combined approach, until a valid expansion is found.

- Lookup in ***B*** for the first bigram $L\ E$ or $E\ R$ where E is a valid expansion of A with strict matching;
- Lookup in ***B*** for the first bigram $L\ E$ or $E\ R$ where E is a valid expansion of A with relaxed matching;
- Lookup in ***U*** for the first unigram E where E is a valid expansion of A with strict matching;
- Lookup in ***U*** for the first unigram E where E is a valid expansion of A with relaxed matching.

[1] For example, *Pathology* would preferably be abbreviated as *Path.*, as *Pat.* would be ambiguous with *Patient*.

In the special case where any of the contexts L or R is also an abbreviation itself (e.g. "St. p."), hereafter denominated pairwise abbreviations, two extra initial lookups are performed:

- Lookup in ***B*** for the first bigram E_L E_A or E_A E_R where E_i is a valid expansion of i with strict matching;
- Lookup in ***B*** for the first bigram E_L E_A or E_A E_R where E_i is a valid expansion of i with relaxed matching.

Similarity index

To overcome mismatches due to adjective inflection endings, which is typical for German, we consider guessed expansions E with similarity $S(A,E)$ to the abbreviation A greater than 70% as a correct match, as seen on Equation 2. The threshold was empirically found in a conservative way.

$$S(A,E) = 1 - \frac{Levenshtein(A,E)}{Max(Length(A),Length(E))} \quad (2)$$

Evaluation

We report precision, recall and F_1 score of our algorithm. Precision is defined as the ratio of correctly expanded tokens to all resolved abbreviations. Recall is defined as the ratio of correctly expanded tokens to all abbreviations [7]. To overcome high precision rates expected when the method gives only expansions with high confidence, F_1 score is defined as the harmonic mean between precision and recall, with equal weights.

Additionally, we define unigram lookup with relaxed matching as a baseline strategy, with which other strategies are compared. We perform a Chi-square test (Fisher's exact test) to verify if the differences are statistically significant and not by chance, with $\alpha = 0.05$.

Source code

The source code was made available under Version 2 of the Apache License at https://github.com/michelole/abbres.

Results

We report results for every lookup considered separately and in the combined approach in Table 1. Bigram strategies, as well as pairwise with strict matching, showed the highest precision rates, but low to moderate recall. When combined with the lower precision unigram approach, we boosted recall to 0.91, thereby achieving an F_1 score of 0.91 in the combined approach (measured in the test set). The combined approach is statistically significant better ($p < 0.001$) than the baseline approach (unigram with relaxed matching). Bigram with relaxed matching is statistically significant better ($p < 0.05$) than the baseline. Any pairwise approach is statistically significant worse ($p < 0.001$) than the baseline approach. Other approaches showed no statistically significant differences to the baseline.

Investigation of errors shows two main patterns: (i) abbreviations preceded or followed by a number (e.g. "52 *jähr. Patientin*", which should expand to "52 *jährige Patientin*", German for "52 years old patient"); and (ii) abbreviation pairs (e.g. "*Z. n.*", which should expand to "*Zustand nach*", German for "state after", i.e. "clinical history of"). We relate the former problem to the lack of any value normalisation (e.g. the conversion of "52" to "00"), which distributes frequencies in the bigram list over every number possibility. Even though we specifically addressed the latter problem in our pairwise resolution strategy, some abbreviation pairs are still ambiguous ("*Z. n.*" could also be expanded to "*Zeit normal*", German for "normal time") and a larger context window might be needed.

Discussion

Related work

Several works have been published on the problem of clinical abbreviation detection, expansion, and disambiguation, with different approaches, languages and types of data.

Pakhomov [8] used a semi-supervised Maximum Entropy model for disambiguation of different short forms in a given context. The system was evaluated exploiting a dataset of about 10,000 rheumatology notes from the Mayo Clinic. Extended context information as well as different models trained on disambiguating one specific acronym yielded about the same accuracy of 0.89 for acronym and abbreviation normalization compared to using one model for the normalization task. The hypothesis that similar context information of abbreviations and acronyms compared to their resolved form supports correct normalization in context could be exploited and confirmed in this work.

Joshi et al. [9] compared three different supervised machine learning approaches for acronym expansion: Naïve Bayes Classifier, Decision Trees and Support Vector Machines. All three models achieved an accuracy of over 0.90. The feature set consisted of part-of-speech tags, unigrams, bigrams and a combination of all of them. A flexible window chosen to catch a certain level of occurrence information of lexical features had a significant impact on the overall evaluation efficiency.

Suominen et al. [10] provided an overview of the ShARe/CLEF eHealth Evaluation Lab 2013, in which Task 2 was focused on normalization of abbreviations and acronyms to UMLS concept unique identifiers (CUIs). The challenge used data from the US intensive care publicly available in the MIMIC-II dataset, originally annotated to build the ShARe corpus. The corpus was further enhanced with annotations regarding abbreviations and acronyms text spans and mappings to UMLS codes. The best team obtained an accuracy of 0.72.

Later, Mowery et al. [11] compared the efficiency of participating systems to a majority baseline and with variable majority sense distribution. They observed that a majority approach performed second best (accuracy of 0.69), given that an estimate of around 81% of short forms have no ambiguity (one unique sense) or low ambiguity (two or more senses, with one incidence over 80%). The only winning system showed a slight improvement (accuracy of 0.72) with a hybrid technique that exploits the same differences in abbreviation frequency and ambiguity.

Siklósi et al. [12] addressed the abbreviation resolution problem in Hungarian ophthalmological notes, a language with less linguistic resources available. They specifically dealt with long series of abbreviations commonly found in their documents. The impact of an external lexicon (built out of 3,329 ICD descriptors), a handmade lexicon (with a size of 44 entries) and the corpus itself in the abbreviation expansion phase was evaluated. Additionally, the influence of the context window (from zero to three tokens) and corpus size (with a total of 2,008 documents) in the final expansion efficiency was measured. The expansion strategy leveraged regular expressions built out of the abbreviations and matched (i) again the corpus and (ii) against the lexicons. In the disambiguation phase, their system used a weighted ranking score based on features such as the size of the longest and shortest span covered. We calculated an F_1 score of 0.85 from the reported precision (0.93) and recall

Table 1 – Correct counts (C), precision (P), recall (R) and F_1 score with different strategies and in a combined approach (* denotes $p < 0.05$; ** denotes $p < 0.01$; and *** denotes $p < 0.001$)

Strategy	Training (n = 147)				Test (n =200)			
	C	P	R	F_1	C	P	R	F_1
Unigram								
Relaxed	91	0.62	0.62	0.62	146	0.73	0.73	0.73
Strict	105	0.76	0.71	0.74	158	0.81	0.79	0.80
Bigram								
Relaxed	119^{***}	0.91	0.81	0.86	165^{*}	0.92	0.83	0.87
Strict	108^{*}	0.94	0.73	0.82	157	0.95	0.79	0.86
Pairwise								
Relaxed[2]	24^{***}	0.63	0.16	0.26	49^{***}	0.80	0.25	0.38
Strict	29^{***}	0.91	0.20	0.32	47^{***}	0.90	0.24	0.37
Combined								
	136^{***}	0.93	0.93	0.93	182^{***}	0.91	0.91	0.91

(0.78) metrics. Moreover, their experiments showed that "considering the tokens without any context always performed worst" and "a context larger than one token […] has a positive effect only if the manually created lexicon is not used".

Wu et al. [13] applied three different neural word embedding models, viz. SBE (surrounding based embedding feature), LR_SBE and MAX_SBE as additional features for a support vector machine to disambiguate abbreviations in context. The investigation used annotated abbreviation datasets from Vanderbilt University Hospital's admission notes and narratives from the University of Minnesota-affiliated Fairview Health Services. The MIMIC-II corpus was used to initialise the word embeddings using the algorithm proposed by Collobert et al. [14]. About 42,000 sentences with resolved abbreviations were used for evaluation, achieving a maximum accuracy of 0.96.

Wu et al. [15] developed an open-source framework for clinical abbreviation recognition and disambiguation and evaluated it with a corpus from the Vanderbilt University Medical Center (VUMC) and in the ShARe/CLEF 2013 challenge corpus. They applied semi-supervised clustering methods for sense expansion and profile-based word sense disambiguation. While the former depends on manual review of around 20 sense clusters for each abbreviation to create a sense inventory, the latter builds upon feature vectors representing different senses in a vector space model. Their system achieved an F_1 score of 0.76 in the VUMC corpus, and a 0.29 F_1 score in the ShARe/CLEF dataset.

Our method differs from most of the works being totally unsupervised. Compared to Siklósi et al. [12], the main difference lies in the fact that our disambiguation strategy is also fully data driven and relies only on the frequencies of words and bigrams in a closely related corpus. Although we agree that a weighted ranking of features (e.g. combining the n-gram frequency to its type and relative gain) could improve the disambiguation process, it would need additional annotations to optimize the weight coefficients, thereby transforming our strategy into a supervised approach. Additionally, we did not observe the same incidence of abbreviation series in our corpus, which could explain our better results. Finally, we hypothesize that our much larger corpus (30,000 versus 2,008 documents) might have overcome the need of any lexicon.

Limitations

Our data-driven strategy might show suboptimal results in languages and subdomains with fewer data. It is also sensitive to spelling errors (e.g. "*bds.*", German abbreviation for "both sides", is at least once incorrectly written as "*bdsl*" in the corpus — note that the "l" character is found near the "." character in most keyboards). Although a minimum frequency could be enforced, preliminary experiments showed no improvement over training data.

Furthermore, we propose some basic writing guidelines that could further improve the automated processing of clinical notes. Considering the abbreviation expansion and disambiguation problem alone, guidelines should stress the importance of (a) correct capitalization (e.g. "*Patient*" instead of "*patient*"); (b) avoiding typos (e.g. "*bdsl*" instead of "*bds.*"); (c) marking abbreviations with a period mark consistently (e.g. "*Tbl.*" instead of "*Tbl*"); (d) standardising double abbreviations, preferably with a space character (e.g. "*St. p.*" instead of "*St.p*"); and (e) separating numbers from their units (e.g. "*10 mg.*" instead of "*10mg.*"). These simple measures would ease data-driven approaches like ours by increasing the signal-to-noise rate.

Future work

Future work could start exploring the impact of value normalisation (i.e. the transformation of all numbers to a standard value) in the results. At least in the training set, this was noted as a common error pattern.

Explorative work also suggests that our algorithm could be useful for the expansion of abbreviations without the period mark, as well as for expanding non-lexicalized and ambiguous acronyms. When uppercased and stripped of their abbreviation mark, abbreviations are similar to acronyms considering the relaxed matching strategy. However, many acronyms are rarely ever expanded, so that the clinical corpus might lack enough full forms. Therefore, other corpora may be necessary, thus complicating the disambiguation task. Our decision to focus on period abbreviations also makes it difficult to compare the results to other works that considered all kinds of short forms.

The proposed strategy should also be evaluated in the broader context of a natural language processing pipeline, in which sentence detection, tokenization, and abbreviation detection are performed together. The special case of glued (without a space mark) pairwise abbreviations (e.g. "St.p.") might be better addressed with the correct output of a tokenizer. Nonetheless, correct abbreviation detection might be fundamental to distinguish between sentence delimiters and real abbreviations, thereby avoiding false positives.

Apart from evaluation, the application of our method over a full sentence should be considered in the context of a Hidden Markov Model. Hence, the best sentence outcome might be obtained via a dynamic programming algorithm such as the Viterbi algorithm. Larger windows, e.g. trigrams, could be then

[2] The inclusion of the relaxed pairwise strategy did not improve results in the combined approach and is therefore excluded from it.

exploited to assure maximal context. Suavisation (e.g. Good-Turing) might be needed for cases unseen in the training set.

Finally, we would like to evaluate our strategy in other subdomains, other languages (e.g. English and Portuguese), and with different corpora sizes, both in publicly available and other restricted corpora. However, our approach would be difficult to run in challenge datasets, like ShARe/CLEF 2013, because it needs a large related corpus as a key resource. For its practical use in clinical language processing this should not be a problem because electronic health record systems do not lack large amounts of text and the choice of domain specific corpora (e.g. cardiology, intensive medicine, nursing, radiology) can be done with attached metadata.

Conclusions

We presented a completely unsupervised approach to the problem of abbreviation expansion. We focused on non-lexicalised and ambiguous abbreviations, commonly created ad-hoc and therefore abundant in clinical narratives. Our strategy is based on bigram and unigram lookup and yielded a 0.91 F_1 score when evaluated with a German cardiology corpus. The result is statistically significant better ($p < 0.001$) than a baseline approach. Our hypothesis that high efficiency rates can be obtained in an unsupervised fashion was therefore not rejected.

Hence, our work provides a successful and reusable method for abbreviation expansion. It improves text comprehension by non-experts and is supposed to improve processing of clinical texts, such as concept mapping and semantic analysis.

Acknowledgements

Our work is funded by the Brazilian National Research Council - CNPq (project number 206892/2014-4). It has been carried out as part of the IICCAB project (Innovative Use of Information for Clinical Care and Biomarker Research) within the K1 COMET Competence Center CBmed, which is funded by the Austrian Federal Ministry of Transport, Innovation and Technology (BMVIT); the Austrian Federal Ministry of Science, Research and Economy (BMWFW); the Austrian state of Styria (Department 12, Business and Innovation); the Styrian Business Promotion Agency (SFG); and the Vienna Business Agency. The COMET program is executed by the Austrian Research Promotion Agency (FFG).

References

[1] H. Xu, P.D. Stetson, and C. Friedman, A study of abbreviations in clinical notes, in: AMIA Annu. Symp. Proc., 2007: p. 821.

[2] S.M. Meystre, G.K. Savova, K.C. Kipper-Schuler, J.F. Hurdle, and others, Extracting information from textual documents in the electronic health record: a review of recent research, *Yearb Med Inf.* **35** (2008) 128–144.

[3] J.J. Berman, Pathology abbreviated: a long review of short terms, *Arch. Pathol. Lab. Med.* **128** (2004) 347–352.

[4] M. Wiesenauer, C. Johner, and R. Röhrig, Secondary use of clinical data in healthcare providers-an overview on research, regulatory and ethical requirements, *Stud Heal. Technol Inf.* **180** (2012) 614–618.

[5] M. Kreuzthaler, and S. Schulz, Detection of sentence boundaries and abbreviations in clinical narratives, *BMC Med Inf. Decis Mak.* **15 Suppl 2** (2015) S4. doi:10.1186/1472-6947-15-s2-s4.

[6] M. Kreuzthaler, M. Oleynik, A. Avian, and S. Schulz, Unsupervised Abbreviation Detection in Clinical Narratives, in: Proocedings Clin. Nat. Lang. Process. Work. (ClinicalNLP 2016), Osaka, Japan, 2016: pp. 91–98.

[7] T. Cohn, Performance metrics for word sense disambiguation, in: Proc. Australas. Lang. Technol. Work., 2003: pp. 86–93.

[8] S. Pakhomov, Semi-supervised maximum entropy based approach to acronym and abbreviation normalization in medical texts, *40th Annu. Meet. Assoc. Comput. Linguist.* (2002) 160–167. doi:10.3115/1073083.1073111.

[9] M. Joshi, S. Pakhomov, T. Pedersen, and C.G. Chute, A comparative study of supervised learning as applied to acronym expansion in clinical reports., *AMIA Annu. Symp. Proc.* (2006) 399–403.

[10] H. Suominen, S. Salanterä, S. Velupillai, W.W. Chapman, G. Savova, N. Elhadad, S. Pradhan, B.R. South, D.L. Mowery, G.J.F. Jones, and others, Overview of the ShARe/CLEF eHealth Evaluation Lab 2013, in: Int. Conf. Cross-Language Eval. Forum Eur. Lang., 2013: pp. 212–231.

[11] D.L. Mowery, B.R. South, L. Christensen, J. Leng, L.-M. Peltonen, S. Salanterä, H. Suominen, D. Martinez, S. Velupillai, N. Elhadad, G. Savova, S. Pradhan, and W.W. Chapman, Normalizing acronyms and abbreviations to aid patient understanding of clinical texts: ShARe/CLEF eHealth Challenge 2013, Task 2., *J. Biomed. Semantics*. **7** (2016) 43. doi:10.1186/s13326-016-0084-y.

[12] B. Siklósi, A. Novák, G. Prószéky, B. Siklosi, A. Novak, and G. Proszeky, Resolving abbreviations in clinical texts without pre-existing structured resources, in: Fourth Work. Build. Eval. Resour. Heal. Biomed. Text Process. Lr., 2014.

[13] Y. Wu, J. Xu, Y. Zhang, and H. Xu, Clinical Abbreviation Disambiguation Using Neural Word Embeddings, *Proc. BioNLP 15*. (2015) 171–176.

[14] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, Natural language processing (almost) from scratch, *J. Mach. Learn. Res.* **12** (2011) 2493–2537.

[15] Y. Wu, J.C. Denny, S. Trent Rosenbloom, R.A. Miller, D.A. Giuse, L. Wang, C. Blanquicett, E. Soysal, J. Xu, and H. Xu, A long journey to short abbreviations: developing an open-source framework for clinical abbreviation recognition and disambiguation (CARD), *J. Am. Med. Inform. Assoc.* **24** (2017) e79–e86. doi:10.1093/jamia/ocw109.

Address for correspondence

Michel Oleynik
Institute for Medical Informatics, Statistics and Documentation
Medical University of Graz, Austria
http://orcid.org/0000-0001-9099-6259
michel.oleynik@stud.medunigraz.at