

# Detecting Named Entities and Relations in German Clinical Reports

Roland Roller<sup>1</sup>(✉), Nils Rethmeier<sup>1</sup>, Philippe Thomas<sup>1</sup>, Marc Hübner<sup>1</sup>,  
Hans Uszkoreit<sup>1</sup>, Oliver Staeck<sup>2</sup>, Klemens Budde<sup>2</sup>, Fabian Halleck<sup>2</sup>,  
and Danilo Schmidt<sup>2</sup>

<sup>1</sup> Language Technology Lab, DFKI, Berlin, Germany

{roland.roller,nils.rethmeier,philippe.thomas,marc.huebner,  
hans.uszkoreit}@dfki.de

<sup>2</sup> Charité Universitätsmedizin, Berlin, Germany

{oliver.staeck,klemens.budde,fabian.halleck,danilo.schmidt}@charite.de

**Abstract.** Clinical notes and discharge summaries are commonly used in the clinical routine and contain patient related information such as well-being, findings and treatments. Information is often described in text form and presented in a semi-structured way. This makes it difficult to access the highly valuable information for patient support or clinical studies. Information extraction can help clinicians to access this information. However, most methods in the clinical domain focus on English data. This work aims at information extraction from German nephrology reports. We present on-going work in the context of detecting named entities and relations. Underlying to this work is a currently generated corpus annotation which includes a large set of different medical concepts, attributes and relations. At the current stage we apply a number of classification techniques to the existing dataset and achieve promising results for most of the frequent concepts and relations.

## 1 Introduction

Within the clinical routine many patient related information are recorded in unstructured or semi-structured text documents and are stored in large databases. These documents contain valuable information for clinicians which can be used to, e.g., improve/support the treatment of long-term patients or clinical studies. Even today information access is often manual, which is cumbersome and time-consuming. This creates a demand for efficient and easy tools to access relevant information. Information extraction (IE) can support this process by detecting particular medical concepts and the relations between them to gather the context. Such structured information can be used to improve use-cases such as the generation of cohort groups or clinical decision support.

Generally, IE can be addressed in many different ways. If sufficient amounts of training instances are available, supervised learning is often the technique of choice, as it directly models expert knowledge. In context of detecting medical

concepts (named entity recognition; NER) and their relations (relation extraction; RE) conditional random fields (CRF) (Lafferty et al. 2001) and support vector machines (SVM) (Joachims 1999) have been very popular supervised methods that were frequently used for the last decade. In recent years neural network based supervised learning has gained popularity (see, e.g., Nguyen and Grishman (2015); Sahu et al. (2016); Zeng et al. (2014)).

In context of IE from German clinical data not much work has been done so far. One reason for that is the unavailability of clinical data resources in German language, as discussed in Starlinger et al. (2016). Only a few publications address the topic of NER and RE from German clinical data. Hahn et al. (2002) focus on the extraction of medical information from German pathology reports in order to acquire medical domain knowledge semi-automatically, while Bretschneider et al. (2013) presents a method to detect sentences which express pathological and non-pathological findings in German radiology reports. Krieger et al. (2014) present first attempts to analyzing German patient records. The authors focus on parsing and RE, namely: symptom-body-part and disease-body-part relations. Toepfer et al. (2015) present an ontology-driven information extraction approach which was validated and automatically refined by a domain expert. Their system aims to find objects, attributes and values from German transthoracic echocardiography reports.

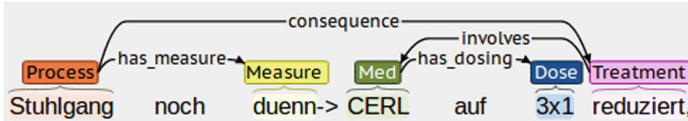
Instead, we focus on detecting medical concepts (also referred to as NE) and their relations from German nephrology reports. For both tasks, NER and RE, two different learning methods are tested: first a well established method (CRF, SVM) and later a neural method for comparison. However, the paper describes on-going work, both in terms of corpus annotations and classification methods. The goal of this paper is to present first results for our use case and target domain.

## 2 Data and Methods

The following section overviews our corpus annotations and the models we use. Note that, due to the (short) format of the paper, method descriptions are brief. We refer the reader to the corresponding publications for details.

### 2.1 Annotated Data

Our annotation schema includes a wide range of different concepts and (binary) relations. The most frequent concepts used in the experiments are listed in Tables 1 and 2, including a brief explanation. The ongoing annotations (corpus generation) include German discharge summaries and clinical notes from a kidney transplant department. An example of our annotations is presented in Fig. 1. Both types of documents are generally written by medical doctors, but have apparent differences. For more details on corpus generation please see Roller et al. (2016).



**Fig. 1.** Annotated sentence

For the following experiments 626 clinical notes are used for training and evaluation. Clinical notes are rather short and written during or shortly after a patient visit. Currently, only 267 of those documents contain annotated relations. The overall frequency of named entities and relations is included with the results in the experimental section (see Tables 3 and 4).

**Table 1.** Annotated concepts

Label	Explanation
Body_Part	Body parts; organs
Med_Con	Medical.Condition: symptom, diagnosis and observation
Process	Body’s own biological processes
Health_State	Positive, wanted finding; contrary to <i>Med_Con</i>
Treatment	Therapeutic procedures, treatments
Medication	Drugs, medicine
Med.Spec	Medical.Specification: closer definition; describing lexemes, often adjectives
Local.Spec	Local.Specification: anatomical descriptions of position and direction

## 2.2 Machine Learning Methods

*NER – Conditional Random Field (CRF)*. Conditional random fields have been used for many biomedical and clinical named entity recognition tasks, such as gene name recognition (Leaman and Gonzalez 2008), chemical compound recognition (Rocktäschel et al. 2012), or disorder names (Li et al. 2008). One disadvantage of CRFs is that the right selection of features can be crucial to achieving optimal results. However, for a different domain or language important features might change. In this work we are not interested in exhaustive feature engineering. Instead, we intend to re-use an existing feature setup as described by Jiang et al. (2015) who use word-level and part-of-speech information around the target concept. For our experiment we use the CRF++<sup>1</sup> implementation.

*NER – Character-level Neural Network (CharNER NN)*. In addition to the well established CRF for NER we also use a neural CRF implementation<sup>2</sup> as

<sup>1</sup> <https://taku910.github.io/crfpp/>.

<sup>2</sup> <https://github.com/ozanarkancan/char-ner>.

**Table 2.** Annotated relations

Label	Explanation
hasState	Describes the state of health (positive and negative) of different entities (e. g., <i>Process</i> , <i>Med_Con</i> , <i>Body_Part</i> )
involves	Describes a relation between <i>Treatment</i> and <i>Medication</i> : e. g., to use or to discontinue a medication
has_measure	Links a <i>Measurement</i> to a corresponding concept
is_located	Links a positional information ( <i>Body_part</i> , <i>Local_spec</i> ) to concepts such as <i>Med_con</i> or <i>Process</i>
is_specified	Links <i>Medical_Spec</i> to a corresponding concept (e. g., <i>Med_Con</i> , <i>Process</i> )

introduced by Kuru et al. (2016). The model uses a character-level Bidirectional-LSTM with a CRF objective. Using character level inputs has the advantage of reducing the unknown vocabulary word problem, as the vocabulary size and hence the feature sparsity are reduced compared to words allowing character models to compensate for words unseen during training, which helps on smaller datasets.

*RE – Support Vector Machine (SVM)*. SVMs are often the method of choice in context of supervised relation extraction (Tikk et al. 2010). Besides their advantages, SVMs also suffer from the issue of optimal feature/kernel selection. Other problems are related to the bias of positive and negative instances in training and test data which can significantly influence the classification results (Weiss and Provost 2001). Again, feature selection is not in our interest. For this reason we use the Java Simple Relation Extraction<sup>3</sup> (jSRE) (Giuliano et al. 2006) which uses a shallow linguistic kernel and bases on LibSVM (Chang and Lin 2011). jSRE provides reliable classification results and has been shown to achieve state-of-the-art results for various tasks, such as protein-protein extraction (Tikk et al. 2010), drug-drug extraction (Thomas et al. 2013) and extraction of neuroanatomical connectivity statements (French et al. 2012).

*RE – Convolutional Neural Network (CNN)*. Besides SVM, we also use a convolutional neural network for relation extraction. We employ a Keras<sup>4</sup> implementation of the model described by Nguyen and Grishman (2015) using a TensorFlow<sup>5</sup> backend and a modified activation layer. The architecture consists of four main layers: (a) lookup tables to encode words and argument positions into vectors, (b) a convolutional layer to create n-gram features, (c) a max-pooling layer to select the most relevant features and (d) a sigmoid activation layer for binary classification.

<sup>3</sup> <https://hlt.fbk.eu/technologies/jsre>.

<sup>4</sup> <https://keras.io>.

<sup>5</sup> <https://www.tensorflow.org>.

### 3 Experiment

In this section named entity recognition and relation extraction on German nephrology reports are carried out. Given a sentence (token sequence), the task of NER is to assign the correct named entity label to the given tokens in the test data. Relation extraction takes a sentence including the different named entity labels as input and determines for each entity pair whether one of our target relations exists. Both classification tasks are evaluated based on precision, recall and F1-Score. Note, due to space reasons, not all relations of the example in Fig. 1 are used for the experiment.

**Table 3.** Concept classification results

Label	Freq.	CRF			CharNER NN		
		Prec	Rec	F1	Prec	Rec	F1
Medical_Condition	2453	95.17	75.16	83.98	89.12	82.15	<b>84.93</b>
Treatment	1680	85.79	69.63	76.81	80.46	76.37	<b>78.33</b>
State_of_Health	1451	86.68	76.35	81.18	83.55	80.80	<b>82.14</b>
Medication	1214	92.28	68.56	78.55	90.37	82.39	<b>86.17</b>
Process	1145	90.53	60.56	72.57	84.74	66.29	<b>74.02</b>
Body_part	840	96.96	65.23	<b>77.90</b>	89.15	68.78	77.53
Medical_Specification	764	78.76	48.82	<b>60.20</b>	65.32	53.04	58.21
Local_Specification	189	95.83	31.94	45.87	81.84	49.77	<b>61.05</b>

**Table 4.** Relation classification results

Label	Freq.	SVM			CNN		
		Prec	Rec	F1	Prec	Rec	F1
hasState	388	86.86	86.86	<b>86.86</b>	81.96	88.10	84.64
involves	370	88.96	78.38	83.33	81.51	90.42	<b>85.58</b>
has_measure	427	80.25	88.52	<b>84.19</b>	81.47	76.61	78.97
is_located	162	46.96	85.80	60.70	65.14	64.12	<b>63.48</b>
is_specified	112	94.85	82.14	<b>88.04</b>	76.34	83.83	79.89

#### 3.1 Preprocessing

To carry out the experiment text documents are processed by a sentence splitter, a tokenizer, stemmer and Part-of-Speech (POS) tagger. The sentence splitting and tokenization are essential to split documents into single sentences and single word tokens. We use JPOS (Hellrich et al. 2015), to tag Part-of-Speech information, since the tool is specialized for German clinical data. POS tags are used for

both the CRF and SVM. Additionally, we stem words for jsRE using the German Snowball stemmer in NLTK. CharNER and CNN do not require additional linguistic features as input.

For both NER and RE the experiments are carried out multiple times – for each single named entity type and each single relation for two reasons: Firstly, in context of named entities tokens might be assigned to multiple labels which our classifiers can not handle directly. Secondly, jsRE does not handle multi-class classification. Hence, we use a One-vs. rest (OvR) classification to train separate models for each NER/RE type.

### 3.2 Named Entity Recognition

*Setup.* NER type evaluation uses the OvR setup to train a single classifier (CRF or CharNER) per class. The experiment run as a reduced 10-fold cross-validation on 3 out of 10 stratified dataset splits, since the CharNER model took a very long time to compute, despite using a GPU. Specifically, each split has a 80% training, a 10% validation and a 10% test part. To further save time, we determined the CharNERs optimal parameters on only one splits’ validation part for only one out of eight entity types (body part). Afterwards, we applied the found parameters to the other entity types and splits to produce average test part scores. Thus, the parameter settings may not be optimal for all entity types. In practice, the CRF trained in hours compared to days for the Bi-LSTM. Both models were evaluated using the 3-fold setup for comparison.

*Results.* The results for named entity recognition are shown in Table 3. Even though classifiers are not necessarily optimal (e.g., no feature engineering), the results are promising. All concepts with a frequency above 800 have an F1-Score above 70. Moreover, all concepts can be detected at a high level of precision. Both classifiers produce comparable results, with better F1 for the CharNER and a focus on precision for the CRF.

### 3.3 Relation Extraction

*Setup.* Our relation extraction task considers only entity pairs within the same sentence. While positive relation instances can be directly taken from the annotations, negative relation instances (used for training and testing) are generated by creating new (unseen) combinations between entities. The relation extraction experiment is then carried out within a 5-fold cross-validation using NE gold labels as input.

Due to the comparably smaller size of the dataset, hyperparameters of the CNN have been slightly modified in comparison to (Nguyen and Grishman 2015). As before, we used one relation type from one fold to find the optimal parameters and then applied those parameters to the other folds and types. This resulted in a reduced position embeddings dimensionality (from 50 to 5) compared to the original model. We also used pre-trained German word embeddings<sup>6</sup>.

<sup>6</sup> <https://devmount.github.io/GermanWordEmbeddings/>.

*Results.* The relation extraction results are presented in Table 4. Most relations can be detected at an F1-Score of 80. Only the relation *is\_located* produces a surprisingly low precision which results in a reduced F1. Overall, the results are very promising and leave space for further improvements using improved classification models.

## 4 Conclusion and Future Work

This work presented first results in context of detecting various named entities and their relations from German nephrology reports. For each task two different methods have been tested. Even though preliminary classification methods have been used (no feature engineering, sub-optimal tuning) and the relatively small size of training and evaluation data, the results are already very encouraging. Generally, the results indicate, that the classification of such information is not too complex. However, a more detailed analysis is necessary to support this assumption.

Future work will focus on increasing the corpus size, and extending/improving our classification models (e.g., elaborate hyperparameter search and selection of pre-trained embeddings). Then those models will be used for further use-cases such as general information access of clinical documents and cohort group generation.

**Acknowledgements.** This research was supported by the German Federal Ministry of Economics and Energy (BMWi) through the project MACSS (01MD16011F).

## References

- Bretschneider, C., Zillner, S., Hammon, M.: Identifying pathological findings in german radiology reports using a syntacto-semantic parsing approach. In: Proceedings of the 2013 Workshop on Biomedical Natural Language Processing, Sofia, Bulgaria, pp. 27–35. Association for Computational Linguistics, August 2013
- Chang, C.-C., Lin, C.-J.: LIBSVM: a library for support vector machines. *ACM Trans. Intel. Syst. Technol.* **2**, 1:27–27:27 (2011)
- French, L., Lane, S., Xu, L., Siu, C., Kwok, C., Chen, Y., Krebs, C., Pavlidis, P.: Application and evaluation of automated methods to extract neuroanatomical connectivity statements from free text. *Bioinformatics* **28**(22), 2963–2970 (2012). <https://doi.org/10.1093/bioinformatics/bts542>
- Giuliano, C., Lavelli, A., Romano, L.: Exploiting shallow linguistic information for relation extraction from biomedical literature. In: Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2006), Trento, Italy (2006)
- Hahn, U., Romacker, M., Schulz, S.: medSynDiKATe - a natural language system for the extraction of medical information from findings reports. *Int. J. Med. Inform.* **67**(1), 63–74 (2002)
- Hellrich, J., Matthies, F., Faessler, E., Hahn, U.: Sharing models and tools for processing German clinical texts. *Stud. Health Technol. Inf.* **210**, 734–738 (2015)

- Jiang, J., Guan, Y., Zhao, C.: WI-ENRE in CLEF eHealth evaluation lab 2015: clinical named entity recognition based on CRF. In: Working Notes of CLEF 2015 - Conference and Labs of the Evaluation forum, Toulouse, France, 8–11 September 2015 (2015)
- Joachims, T.: Making large-scale support vector machine learning practical. In: Schölkopf, B., Burges, C.J.C., Smola, A.J. (eds.) *Advances in Kernel Methods*, pp. 169–184. MIT Press, Cambridge (1999). ISBN 0-262-19416-3
- Krieger, H.-U., Spurk, C., Uszkoreit, H., Xu, F., Zhang, Y., Müller, F., Tolxdorff, T.: Information extraction from german patient records via hybrid parsing and relation extraction strategies. In: Calzolari, N., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., Piperidis, S. (eds) *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)*, Reykjavik, Iceland, May 2014. European Language Resources Association (ELRA) (2014). ISBN 978-2-9517408-8-4
- Kuru, O., Can, O.A., Yuret, D.: CharNER: character-level named entity recognition. In: *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, Osaka, Japan, December 2016, pp. 911–921. The COLING 2016 Organizing Committee (2016)
- Lafferty, J., McCallum, A., Pereira, F., et al.: Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: *Proceedings of the Eighteenth International Conference on Machine Learning, ICML, vol. 1*, pp. 282–289 (2001)
- Leaman, R., Gonzalez, G.: BANNER: an executable survey of advances in biomedical named entity recognition. In: *Proceedings of the Pacific Symposium on Biocomputing 2008*, Kohala Coast, Hawaii, USA, 4–8 January 2008, pp. 652–663 (2008)
- Li, D., Kipper-Schuler, K., Savova, G.: Conditional random fields and support vector machines for disorder named entity recognition in clinical texts. In: *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing, BioNLP 2008*, Stroudsburg, PA, USA, pp. 94–95. Association for Computational Linguistics (2008). <http://dl.acm.org/citation.cfm?id=1572306.1572326>. ISBN 978-1-932432-11-4
- Nguyen, T.H., Grishman, R.: Relation extraction: perspective from convolutional neural networks. In: *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*, pp. 39–48, Denver, Colorado, June 2015. Association for Computational Linguistics (2015)
- Rocktäschel, T., Weidlich, M., Leser, U.: ChemSpot: a hybrid system for chemical named entity recognition. *Bioinformatics* **28**(12), 1633–1640 (2012)
- Roller, R., Uszkoreit, H., Xu, F., Seiffe, L., Mikhailov, M., Staeck, O., Budde, K., Halleck, F., Schmidt, D.: A fine-grained corpus annotation schema of German nephrology records. In: *Proceedings of the Clinical Natural Language Processing Workshop*, vol. 28, no. 1, pp. 69–77 (2016)
- Sahu, S.K., Anand, A., Oruganty, K., Gattu, M.: Relation extraction from clinical texts using domain invariant convolutional neural network. In: *Proceedings of the 15th Workshop on Biomedical Natural Language Processing*, Berlin, Germany. Association for Computational Linguistics (2016)
- Starlinger, J., Kittner, M., Blankenstein, O., Leser, U.: How to improve information extraction from German medical records. *it-Inf. Technol.* **59**(4), 171–179 (2016)
- Thomas, P., Neves, M., Rocktäschel, T., Leser, U.: WBI-DDI: drug-drug interaction extraction using majority voting. In: *Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval)* (2013)



- Tikk, D., Thomas, P., Palaga, P., Hakenberg, J., Leser, U.: A comprehensive benchmark of kernel methods to extract protein-protein interactions from literature. *PLoS Comput. Biol.* **6**, e1000837 (2010). <https://doi.org/10.1371/journal.pcbi.1000837>
- Toepfer, M., Corovic, H., Fette, G., Klügl, P., Störk, S., Puppe, F.: Fine-grained information extraction from German transthoracic echocardiography reports. *BMC Med. Inform. Decis. Making* **15**(1), 1–16 (2015). <https://doi.org/10.1186/s12911-015-0215-x>. ISSN 1472-6947
- Weiss, G., Provost, F.: The effect of class distribution on classifier learning: an empirical study. Technical report, Rutgers Univ. (2001)
- Zeng, D., Liu, K., Lai, S., Zhou, G., Zhao, J.: Relation classification via convolutional deep neural network. In: *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, Dublin, Ireland, August 2014, pp. 2335–2344. Dublin City University and Association for Computational Linguistics (2014)

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

