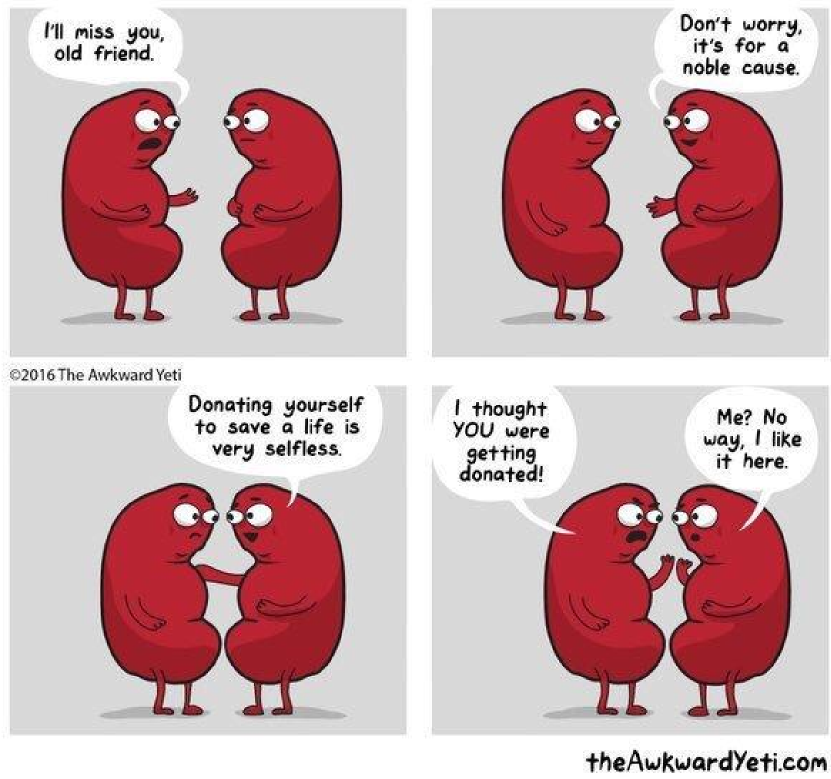
Detecting Named Entities and Relations in German Clinical Reports

Roland Roller[1], Nils Rethmeier[1], Philippe Thomas[1], Marc Hübner[1], Hans Uszkoreit[1], Oliver Staeck[2], Klemens Budde[2], Fabian Halleck[2] and Danilo Schmidt[2]

[1]Language Technology Lab, DFKI, Berlin, Germany
[2]Charité Universitätsmedizin, Berlin, Germany

GSCL 2017

- Background
- Clinical Data
- Annotations
- Methods
- Experiments
- Results
- Conclusion
- (small Demo)



Background

Supported by:

Federal Ministry
for Economic Affairs
and Energy

on the basis of a decision
by the German Bundestag

Smart
Service
Welt

MACSS:

Medical Allround-Care Service Solution

MACSS Medical Allround-Care Service Solutions

AOK
AOK Nordost - Die Gesundheitskasse

TK Techniker Krankenkasse
Gesund in die Zukunft.

medVISION
Medical Software Solutions

Chiesi

stakeholder

SAP

CHARITÉ

smart patient

consortium

dosing

DFKI

astellas

boso
BOSCH + SOHN
GERMANY

Pfizer

KfH

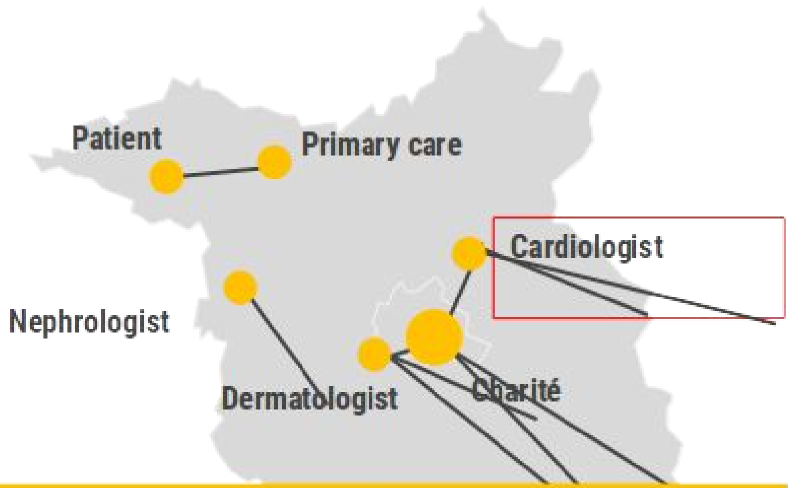
BUNDESVERBAND
NIERE E.V.

stakeholder

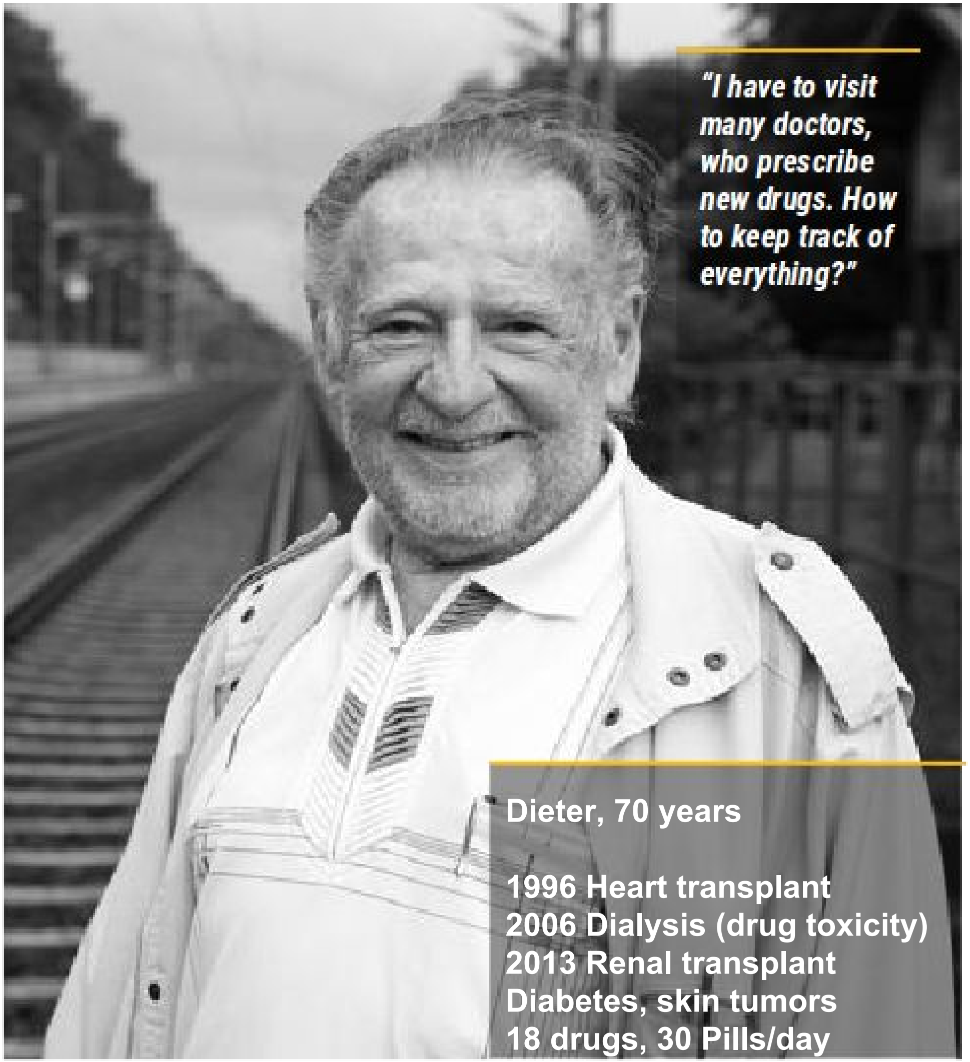
Background

need for an interconnected treatment for chronically ill patients with modern communication technology

Background



Need for a real-time and bidirectional communication within a single platform for optimized drug dosing and close surveillance of chronically ill patients



Background

My data:
The patient decides:

The patient carries his data. Aggregation of data only within
the data trust center at Charité.

Drug chart diary

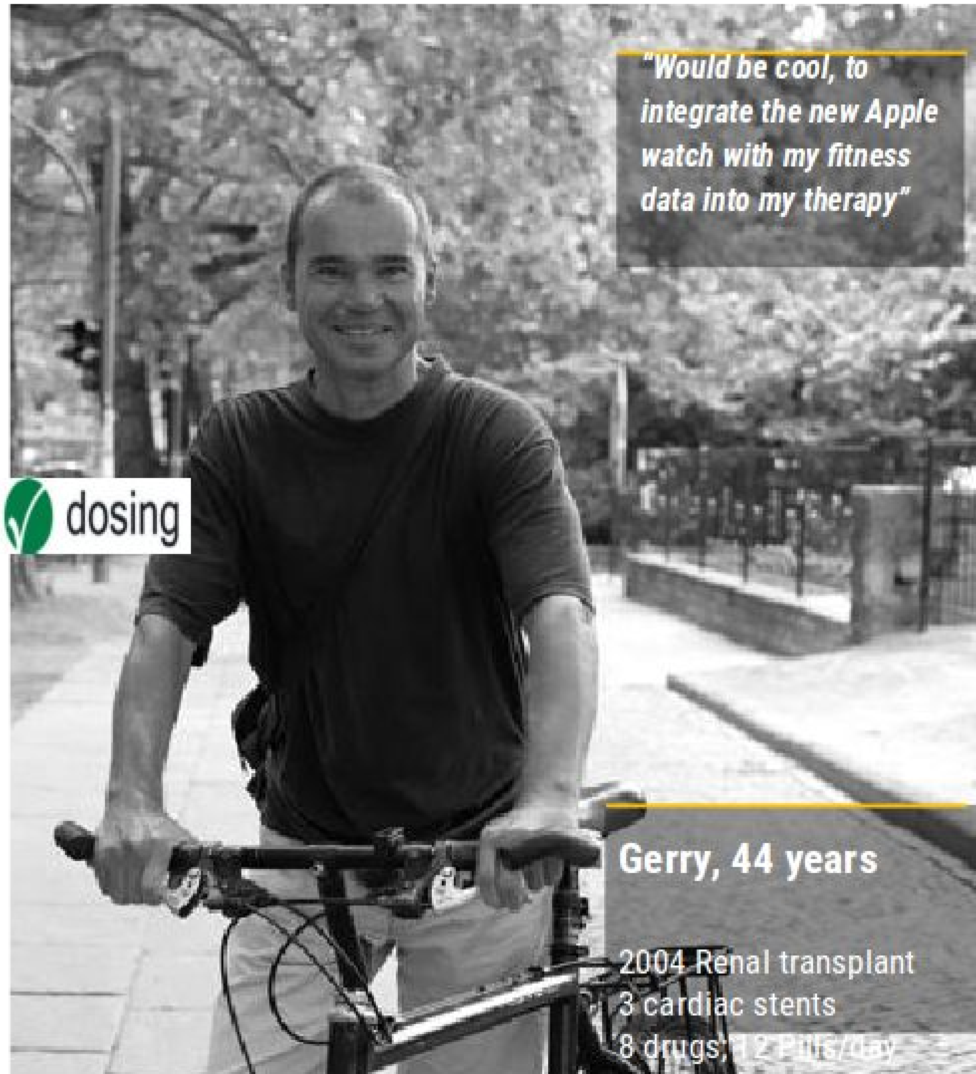
Messenger
Drug check

CHARITÉ

dosing

weight, activity, blood pressure...

MACSS Medical Allround-Care Service Solutions



Overview

- **Focus:** Nephrology reports - clinical notes
- **Goal:** Support physicians to access information
- **Problems:**
  - Availability & accessibility of data
  - Domain dependence
  - Less resources for German
- **This paper presents:**
  - Preliminary NER & RE results on German nephrology reports
  - Ongoing work

Clinical Data

- Clinical Notes (dt. Verlaufsnotizen)
  - Short notes written during a visit
  - Often just a few sentence
- Discharge Summaries (dt. Arztbriefe)
  - Written during stay in hospital
  - Semi-structured
  - Much longer than clinical notes
- Characteristics Clinical Data:
  - Information density
  - Technical language
  - Abbreviations, Negations, Uncertainty

Clinical Data

- Clinical Notes (dt. Verlaufsnotizen)
  - Short notes written during a visit
  - Often just a few sentence
- Discharge Summaries (dt. Arztbriefe)
  - Written during stay in hospital
  - Semi-structured
  - Much longer than clinical notes

	Discharge Summaries	**Clinical Notes**
#documents available	118	1607
#words (total)	89691	68480
#sentences (total)	16068	11871
avg. words per document (std. deviation)	760.09 (208.62)	42.61 (35.74)

Examples: Clinical Notes

Gutes Befinden. Keine Fieber, kein Infekt, keine Beschwerden bei Miktion. Keine Stuhlgangprobleme. Unklarer papillöser Hautbefund. Abklärung durch Dermatologie empfohlen. Soll ggf. vor Ort in der Kur geschehen.

MCP abgesetzt. Bei leicht erhöhten Blutdruckwerten Metrolopol in der Dosis erhöhten, ggf. weitere Dosissteigerung. Leberwerte weiter gestiegen. CyA reduziert. Montag Abklärung der Lebersituation.

Good condition. No fever, no infection, no micturition disturbances. No problems with defaecation. Unclear papillose structures on the skin. Dermatological examination recommended. Should be made during course of treatment, if necessary.

Discontinuation of MCP. Increase in Metrolopol dosage, the blood pressure being slightly increased, further dosage increases, if necessary. Further increased liver levels. Ciclosporin reduced. Examination of the liver on Monday.

Annotations

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 *Med_Con*
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

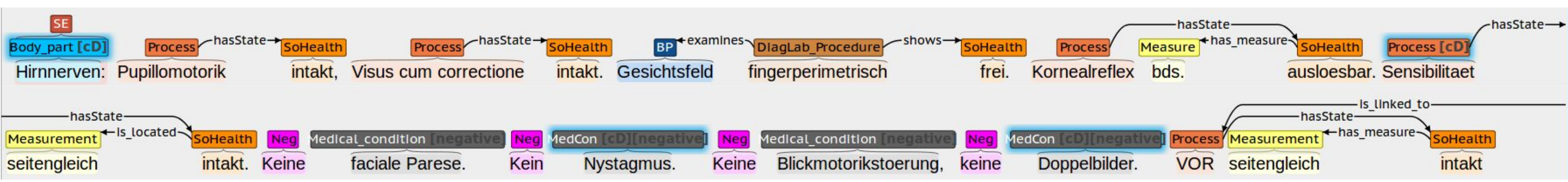
For more see Roller et al. (2016)

Annotations

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

Annotated Data

- Example Annotations



Methods

- Using existing methods to explore reliability
  - Named Entity Recognition:
    - Conditional Random Field (CRF)
      - based on Jiang et al., (2008)
    - Bidirectional LSTM (CharNER NN)
      - implementation of Kuru et al., (2016)
  - Relation Extraction
    - Support Vector Machine (SVM)
      - impl. of Giuliano et al., (2006)
    - Convolutional Neural Network (CNN)
      - Nguyen and Grishman (2015)

Experiment

- Pre-Processing
  - Sentence Splitting, Tokenization
  - POS Tagging (Hellrich et al. 2015)
  - Snowball Stemmer
- Setup
  - 626 clinical notes used for training & testing
  - Only 267 documents contain relations
  - Considering only a subset of concepts and relations for the experiments
  - Cross validation

Results

- Results: Concept Recognition

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

Results

- Results: Relation Extraction

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

Conclusion & Future Work

- First promising results:
  - Training dataset small
  - Classifier not optimized
- Future Work:
  - Building applications on top of NER and RE
  - Summarization
  - Cohort Group generation
  - Reduction of re-hospitalization

References

Roland Roller, Hans Uszkoreit, Feiyu Xu, Laura Seiffe, Michael Mikhailov, Oliver Staeck, Klemens Budde, Fabian Halleck, and Danilo Schmidt. 2016. *A fine-grained corpus annotation schema of German nephrology records.* Proceedings of the Clinical Natural Language Processing Workshop, 28(1):69–77.

Jingchi Jiang, Yi Guan, and Chao Zhao. 2015. *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, September 8-11, 2015.

Onur Kuru, Ozan Arkan Can, and Deniz Yuret. 2016. *CharNER: Character-Level Named Entity Recognition.* In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pages 911–921, Osaka, Japan, December. The COLING 2016 Organizing Committee.

Claudio Giuliano, Alberto Lavelli, and Lorenza Romano. 2006. *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.

Thien Huu Nguyen and Ralph Grishman. 2015. *Relation Extraction: Perspective from Convolutional Neural Networks.* In Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing, pages 39–48, Denver, Colorado, June. Association for Computational Linguistics.

Johannes Hellrich, Franz Matthies, Erik Faessler, and Udo Hahn. 2015. Sharing models and tools for processing German clinical texts. Studies in Health Technology and Informatics, 210:734–738.

Thank you.