Natural Language Processing in the Clinical Narrative

The Junior Research Group DE.xt

Luise Modersohn

Chair of Medical Informatics
Institute of AI in Medicine (AIIM)
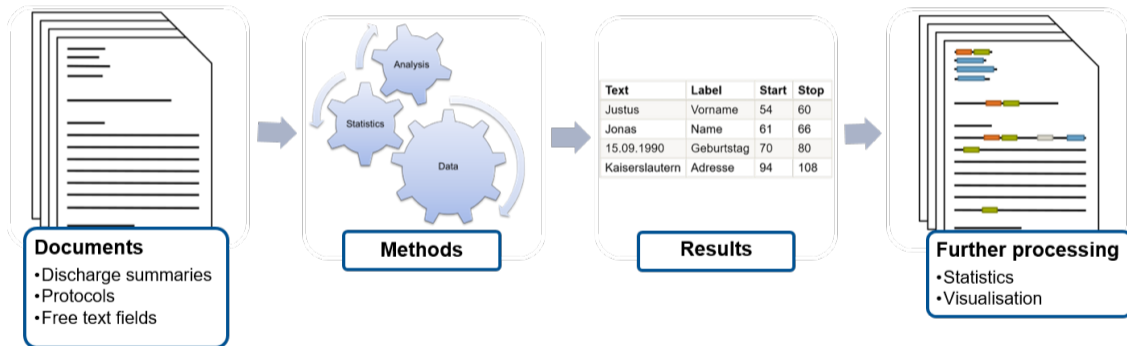Technical University of Munich

October 9th, 2023

Outline

(Clinical) text processing

Documents
- Discharge summaries
- Protocols
- Free text fields

Methods

Results

Text	Label	Start	Stop
Justus	Vorname	54	60
Jonas	Name	61	66
15.09.1990	Geburtstag	70	80
Kaiserslautern	Adresse	94	108

Further processing
- Statistics
- Visualisation

Outline

Creation of Gold Standards

Text data
- **Broad Consent** („genuine data")
- MockUp data
- Synthetic texts
- Data donations

De-Id

Annotation
- INCEpTION
- Active Learning

Ethics

Further use
- MII data use agreement
- Deployment of models
- Licence

Modersohn et al.: *GraSCCo - The First Publicly Shareable, Multiply-Alienated German Clinical Text Corpus*; GMDS 2022

Levels of Complexity

Type Level

- Entities, token-of-interest, etc.
- Token span or off-set (absolute start and end position)
- Small number of **independent** categories
- Examples:
  - PHI criteria (name, address, day of birth, etc.)
  - Temporal information (date, time, duration, etc.)
  - More abstract categories (type, specification, attribute, etc.)

Levels of Complexity

Connection Level

- Relations and other links
- Different levels of complexity:
  1. **X**[dose]   *is dose of*   **Y**[drug] (clear)
  2. **X**[protein] *has methylated* **Y**[gene] (mentioned)
  3. **X**[event]   *happened before* **Y**[date] (Contextual)

Levels of Complexity

Sense Level

- Mapping onto classification systems and terminologies (e.g. SNOMED CT, ICD 10)
- Normalisation of measurements to SI units
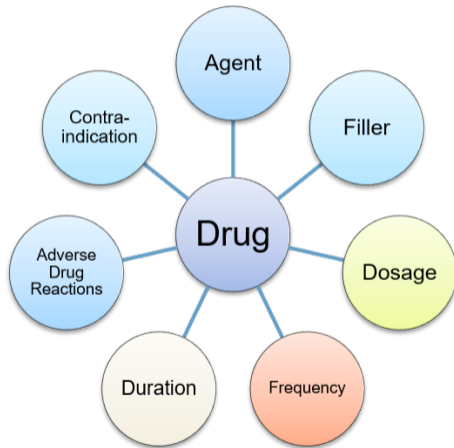- Conversion of dates and time into a standard format
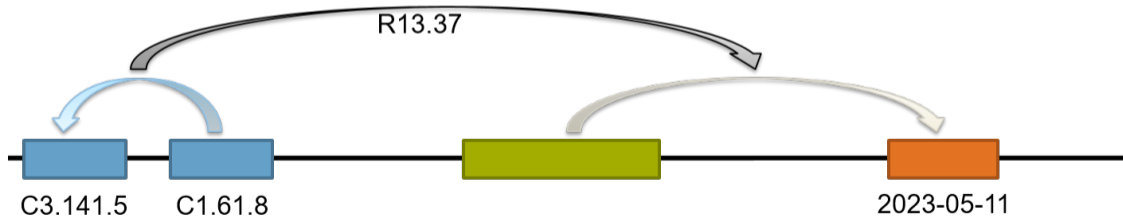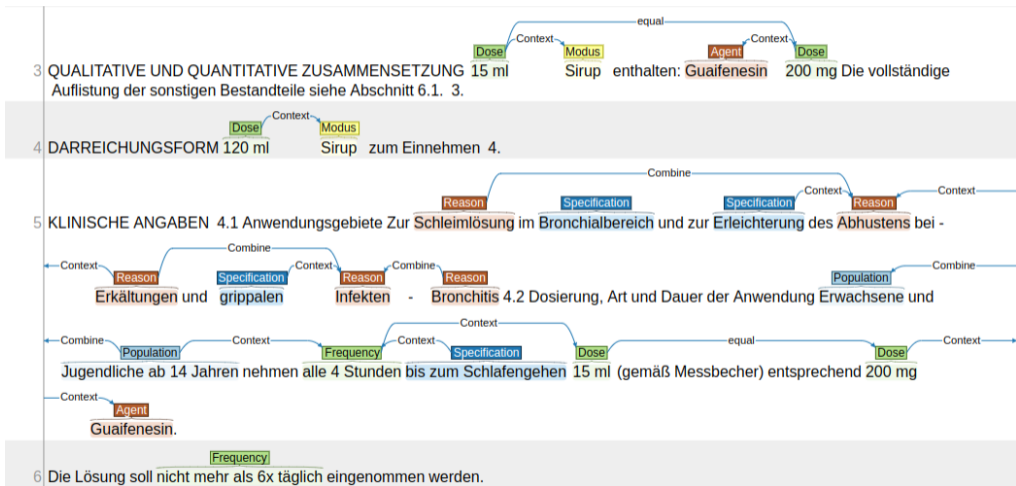
Annotation of German Package Leaflets
Till Sturm (MD thesis)

- Collecting information about drugs
- About 13.000 German package leaflets
- Information base for drug knowledge graph
- Conversion of unstructured information

- Difficulty: conditional information

Annotation of German Package Leaflets
## Till Sturm (MD thesis)

3 QUALITATIVE UND QUANTITATIVE ZUSAMMENSETZUNG 15 ml Sirup enthalten: Guaifenesin 200 mg Die vollständige Auflistung der sonstigen Bestandteile siehe Abschnitt 6.1. 3.

4 DARREICHUNGSFORM 120 ml Sirup zum Einnehmen 4.

5 KLINISCHE ANGABEN 4.1 Anwendungsgebiete Zur Schleimlösung im Bronchialbereich und zur Erleichterung des Abhustens bei - Erkältungen und grippalen Infekten - Bronchitis 4.2 Dosierung, Art und Dauer der Anwendung Erwachsene und Jugendliche ab 14 Jahren nehmen alle 4 Stunden bis zum Schlafengehen 15 ml (gemäß Messbecher) entsprechend 200 mg Guaifenesin.

6 Die Lösung soll nicht mehr als 6x täglich eingenommen werden.

Outline

TTM

1. Uses for NLP in Medicine

2. Corpus Development

3. Automated Information Extraction from Texts

4. Sharing is Caring

Finetuning Models for Downstream Tasks
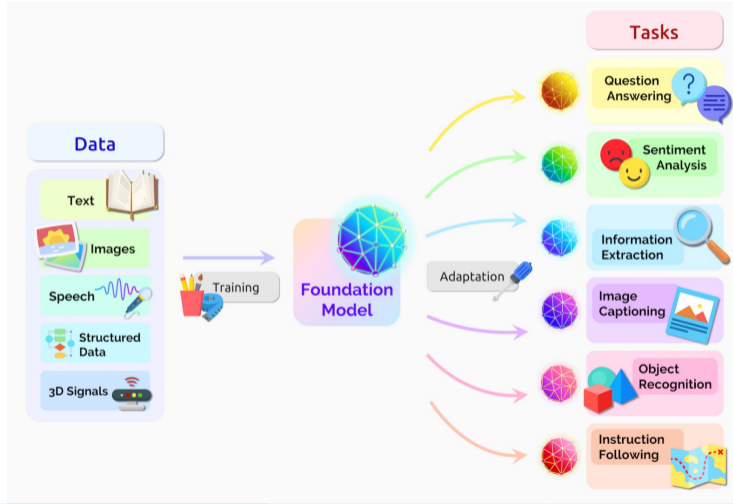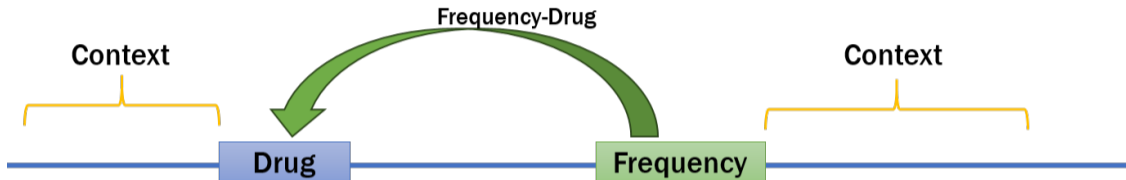
https://blogs.nvidia.com/wp-content/uploads/2022/03/Transformer-apps.jpg

Influence of Context on Relation Prediction

Modersohn et al.: *Influence of context in transformer-based medication relation extraction*; MedInfo 2023

Influence of Context on Relation Prediction

Annotated Datasets
- English
  - I2B2 2009
  - N2C2 2018
- German
  - 3000PA

Pre-processing
- IOB format for NER
- Sentence with named entity pairs for RE
- Train/dev/test split
- 65-10-25

Model training
- Huggingface library
- Based on BioBERT-pytorch implementation

Hyper-parameter optimization
- BOHB for optimization
- Microsoft/NNI implementation

Evaluation
- Test split
- Hyper-parameter from optimization step
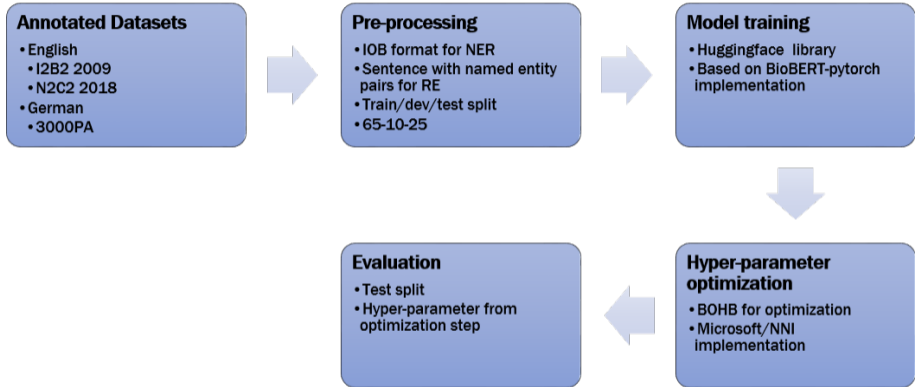
Modersohn et al.: *Influence of context in transformer-based medication relation extraction*; MedInfo 2023

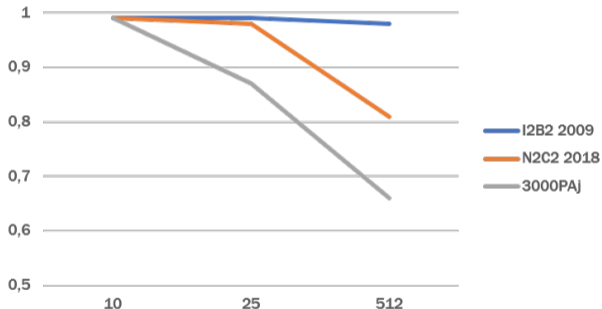# Influence of Context on Relation Prediction
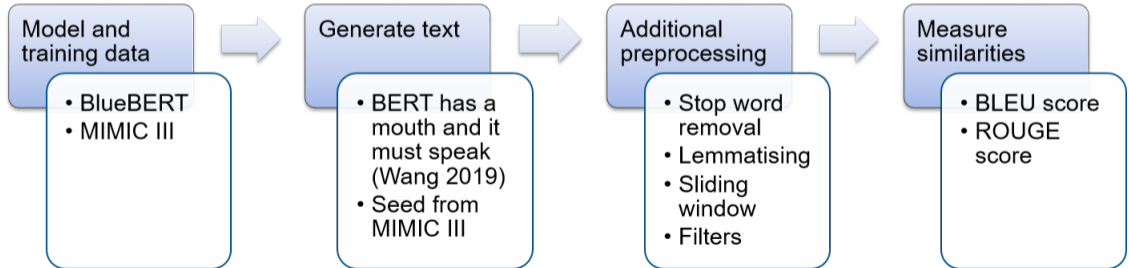
Datasets

- **English**
  - ☐ I2B2 2009
  - ☐ N2C2 2018
- **German**
  - ☐ 3000PA$_J$

F1-Score Medication Relation Extraction

Legend:
- I2B2 2009
- N2C2 2018
- 3000PAj

Modersohn et al.: *Influence of context in transformer-based medication relation extraction*; MedInfo 2023

Outline

ЛЛ

Security of BERT Models
Lukas Groß (students projects)

- Clinical data are sensitive data
- If clinical text data can not be shared - maybe models can
- Question: are BERT models privacy preserving

Model and training data
- BlueBERT
- MIMIC III

→

Generate text
- BERT has a mouth and it must speak (Wang 2019)
- Seed from MIMIC III

→

Additional preprocessing
- Stop word removal
- Lemmatising
- Sliding window
- Filters

→

Measure similarities
- BLEU score
- ROUGE score

Security of BERT Models
## Lukas Groß (students projects)

reference text:

his group home on the afternoon of his admission take to osh and note to be lethargic and febrile to 1034 also have diffuse maculopapular blanching rash to extremity lab be notable for cr 21 baseline unknown inr 36 ua cxr and ct head unremarkable in the ed he be give iv abx and fosphenytoin of note pt take to vascular physician 1d prior to admission and start on

generated text:

bwall hct 7 8 z hl stable c7 compression fx resorfer for c see by dr restart hctz and discuss management of elevated bp with pcp v medication adjust for p r prolongament per the follow service x5ms optho as note ct neck to assess an y defciency i d result pupillary funda be full d 0 5 s 32 7 8 z z 8 ed 6 head re examination this morningser family hear a friend hear you cry the last name on be the name the be on

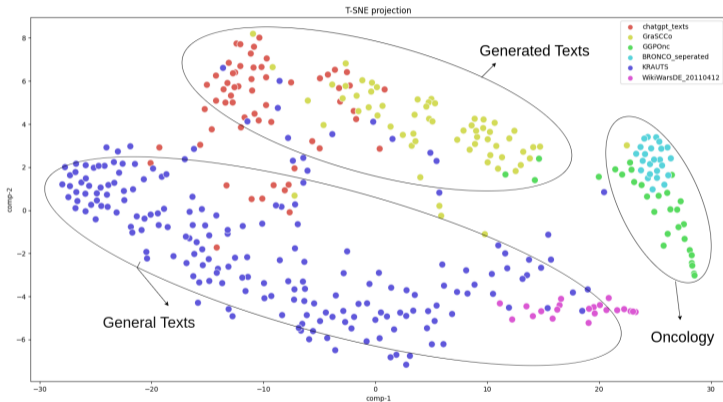# Generating Synthetic Discharge Summaries with LLM
Ipek Uyguner
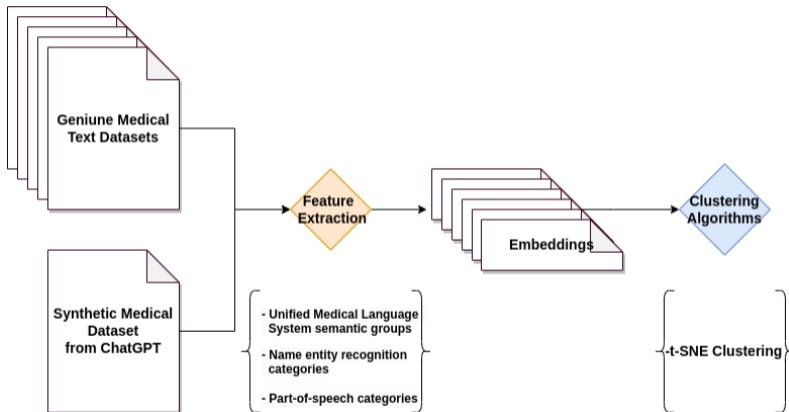
Uyguner et al.: *Potential of ChatGPT for generating synthetic German medical corpus: A comparison with real-world corpora*; GMDS 2023

Generating Synthetic Discharge Summaries with LLM
Ipek Uyguner

Uyguner et al.: *Potential of ChatGPT for generating synthetic German medical corpus: A comparison with real-world corpora*; GMDS 2023

Luise Modersohn
- Group leader
- Luise.modersohn@tum.de

Ingrid Martin
- PhD student
- Ingrid.martin@tum.de

Open Position
- PostDoc

Open Position
- DoktorandIn

We are hiring!

- **PostDoc** with experience in computational linguistics and/or language modelling
- **PhD** clinical text processing
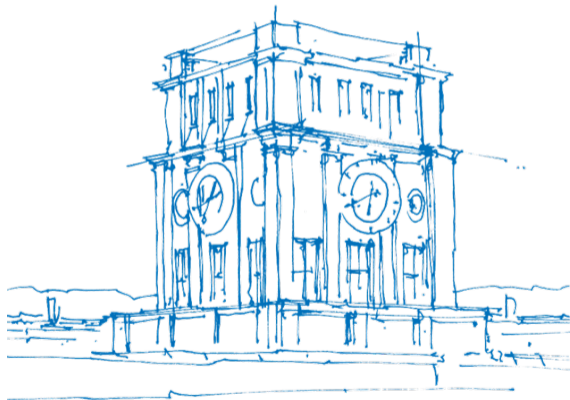- Annotation manager

Natural Language Processing in the Clinical Narrative

The Junior Research Group DE.xt

Luise Modersohn

Chair of Medical Informatics
Institute of AI in Medicine (AIIM)
Technical University of Munich

October 9th, 2023

Tum Uhrenturm