

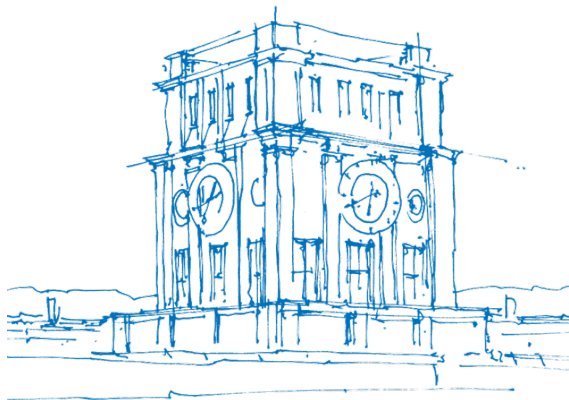
Natural Language Processing im klinischen Kontext

Die Nachwuchsgruppe DE.xt

Luise Modersohn

Lehrstuhl für Medizininformatik
Institut für KI und Informatik in der MEDizin
Technische Universität München

March 21st, 2023



TUM Uhrenturm

- 1 Eingrenzung der Fragestellung
- 2 Die Möglichkeiten von NLP
- 3 Das Ziel der Nachwuchsgruppe DE.xt
- 4 Gibt es da nicht schon was von <PRODUCT>?



- Klinische Dokumentation in Freitextform

- ☐ Entlass- und Verlegungsbriefe
- ☐ Befundberichte
- ☐ Pflegeprotokolle
- ☐ Freitextfelder in Formularen

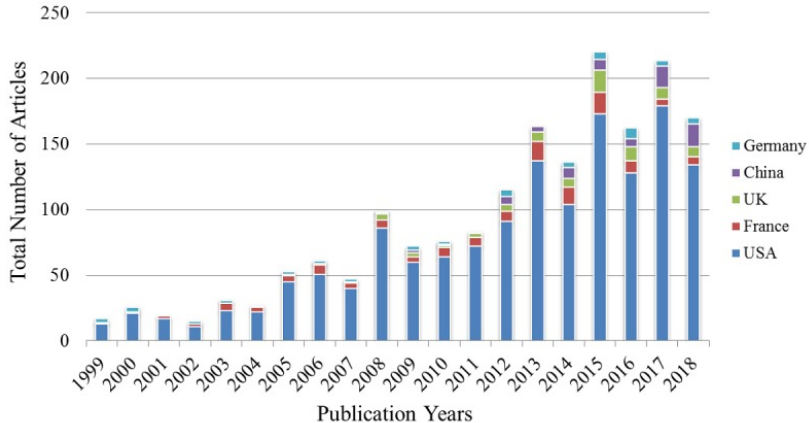
- Unstrukturierte Informationen

- Unerschlossene Datenquellen

- NLP als Werkzeug zur Überführung von Freitext zu strukturierten Daten

Internationale Erfolge

1. Boston Children Hospital – cTAKES
2. Mayo Clinic – NLPaaS
3. Harvard Medical School – N2C2/I2B2 challenges
4. Harvard Medical School – Health Natural Language Processing Center
5. SMITH et al.: *Natural Language Processing and Machine Learning to Enable Clinical Decision Support for Treatment of Pediatric Pneumonia*; AMIA 2020
6. Wen et al.: *Desiderata for delivering NLP to accelerate healthcare AI advancement and a Mayo Clinic NLP-as-a-service implementation*; njp digital medicine 2019
7. Rajkomar et al.: *Scalable and accurate deep learning with electronic health records*; njp digital medicine 2018



Wang et al.; *Systematic Evaluation of Research Progress on Natural Language Processing in Medicine Over the Past 20 Years*; JMIR 2020; Fig. 3

1. Zusammengesetzte Substantive
 - ☐ Methylprednisolonpulsstherapie
2. Klinischer Fachjargon
 - ☐ XY wurde punktiert
3. Telegrammstil (Twitterstil)
 - ☐ Pupillen mw, rund, isokor
4. Abkürzungen
 - ☐ Redurierter AZ, normaler EZ
5. (Temporal)er Kontext
 - ☐ Z.n. Appendektomie

Mehrdeutigkeiten

1. Schwalbe, ma(h)len, Schimmel, umfahren
2. Der Fenstersturz
3. Ein Fräulein ist eine Frau, der zum Glück ein Mann fehlt.
4. 1916 erkrankte Maurice Ravel an der Ruhr.

Arztbrief Beispiel

[...]

Diagnose

St.p. TE eines Malignen Melanoms (Level IV, TD 3,75mm)/ Oberbauch links

St.p. NE tumorfrei und Sentinel Node -Exstirpation (tumorfrei)

Psychose (ED 2018) seither Therapie mit Abilify + Cipralex

Histologie

Histologie vom 19.4.2023: Abtragung rechts: Dermal melanocytärer Naevus.

E-Nr.: 17217277: NE Oberbauch I: NE-präparat mit Narbengewebe.

E-Nr.: 37844677 Patho Universitätsklinikum Klagenfurt: Sentinel Node, Leiste links: Tumofr. Sentinel node.

E-Nr.: 17663757: TE Oberarm rechts: Angiolipom

Inspektion u. Palpation

Blande Narbe im Bereich des Oberbauchs. Ein sichtbarer kleiner subkutaner Faden wird entfernt.

Letzte Durchuntersuchungen: Thoraxröntgen, OB-Sono, Son axillär, ing.: unauffällig.

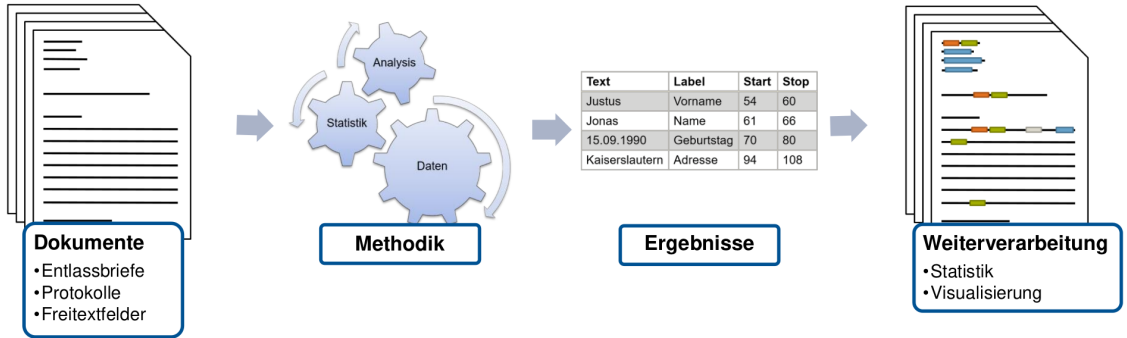
RR: 145/85 mmHg

Labor : incl. BB und Diff-BB, Elektrolyte (Na, K, Cl, Ca), Niere (Krea, Harnstoff, Harnsäure), Leber (Bili, AP, GGT, CHE, AST, ALT), Herz (CK, LDH), Pankreas (P-Amylase, Lipase), Glucose, CRP, Chol, Tri unauff. bis auf: Lipase 122, Cholesterin 219.

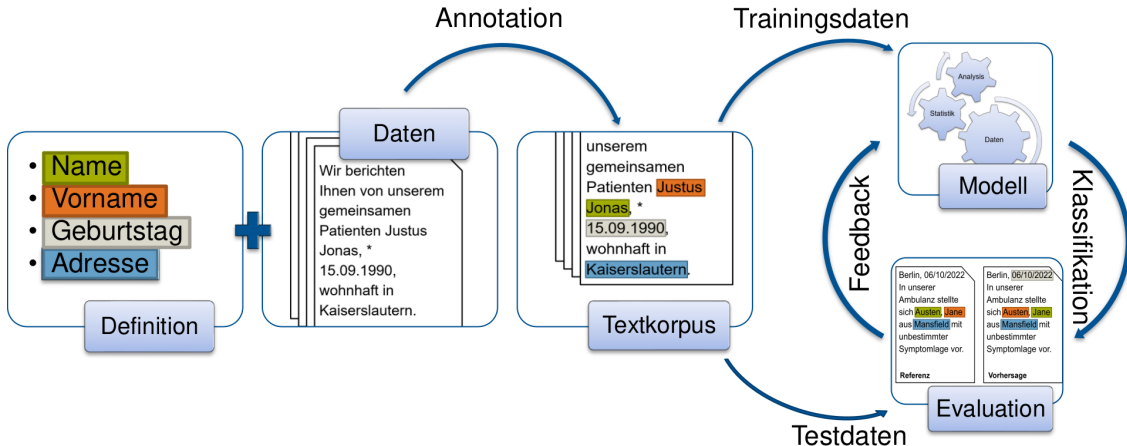
[...]

- 1 Eingrenzung der Fragestellung
- 2 Die Möglichkeiten von NLP**
- 3 Das Ziel der Nachwuchsgruppe DE.xt
- 4 Gibt es da nicht schon was von <PRODUCT>?

Vom Freitext zu strukturierten Daten



Automatische Erkennung personenbezogener Elemente



Kolditz et al.: *Annotating German Clinical Documents for De-Identification*; MedInfo 2019

Berlin, 06/10/2022

In unserer
Ambulanz stellte
sich **Austen**, **Jane**
aus **Mansfield** mit
unbestimmter
Symptomlage vor.

Referenz

Berlin, 06/10/2022

In unserer
Ambulanz stellte
sich **Austen**, **Jane**
aus **Mansfield** mit
unbestimmter
Symptomlage vor.

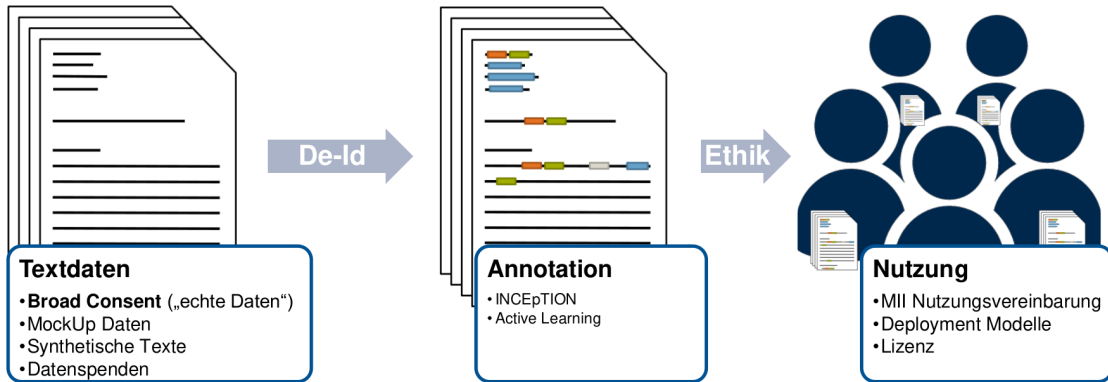
Vorhersage

Label:

- **Name**
- **Vorname**
- **Geburtstag**
- **Adresse**
- [kein Label]

Richtig ✓	12	80 %
Falsch !	3	20 %

Erstellung medizinischer Referenz-Korpora



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

- 1 Eingrenzung der Fragestellung
- 2 Die Möglichkeiten von NLP
- 3 Das Ziel der Nachwuchsgruppe DE.xt**
- 4 Gibt es da nicht schon was von <PRODUCT>?

Forschungsfragen und Projektaufgaben

Arbeitspaket		2021				2022				2023				2024				2025				2026				Sum
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	
AP 1	Bereitstellung von Ressourcen für die Sprachverarbeitung auf klinischen Texten																									56,2
AP 1.1	Aufbau eines de-identifizierten annotierten deutschsprachigen Korpus					2,0	2,0	2,0	2,0	M1	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	0,5	0,5		22,0
AP 1.2	Bereitstellung von klinischen Terminologien für NLP					2,0	2,0	2,0	1,2	M2	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,5	0,5		17,2
AP 1.3	Dissemination von Ressourcen für NLP auf klinischen Texten						1,0	1,0	1,0	M3	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	0,5	0,5		17,0
AP 2	Entwicklung von NLP-Methoden zur Nutzung an klinischen Daten im Routineeinsatz																									91,4
AP 2.1	Entwicklung von Werkzeugen für klinisches NLP				2,0	2,0	2,0	2,0	1,8	M4	1,5	1,5	1,5	1,5	1,5	1,5	1,5	1,5	1,5	1,5	1,5	1,5	1,5	1,5		32,2
AP 2.3	Modelle für die Verarbeitung klinischer Texte in deutscher Sprache				2,0	2,0	2,0	2,0	2,0	M5	2,0	2,0	2,0	2,0	2,0	M6	2,0	2,0	2,0	2,0	2,0	2,0	2,0	1,8	1,5	39,3
AP 2.4	Bereitstellung von klinischen Dokumenten und Aufbau von Recherche-Diensten				2,0	M7	2,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	M8	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0		20,0	
AP 3	Use Cases zur Umsetzung von NLP in der klinischen Anwendung																									39,4
AP 3.1	Use Case 1: NLP auf Radiologiebefunden								0,9		1,5	1,5	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0				13,9	
AP 3.2	Use Case 2: NLP auf Patient-reported Outcomes (PROs)										1,2	1,2	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0			13,4	
AP 3.3	Klinische Integration von NLP-basierten Methoden												0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	1,7	1,9	1,5	12,1
					6,0	9,9	9,9	9,9	9,9		9,9	9,9	9,9	9,9	9,9	9,9	9,9	9,9	9,9	9,9	9,9	9,9	6,7	6,0		187,0

Was wollen wir erreichen?

1. Methodenentwicklung
2. Generierung von Corpora
3. Anwendung
 - ☐ Demonstration der Methodik anhand von Use Cases
 - ☐ Verbesserung der Dokumentation
 - ☐ De-Identifizierung und Pseudonymisierung
 - ☐ Verbesserter Zugriff auf Informationen
4. Nachhaltigkeit
5. Aufbau einer Community

Aufbau einer deutschen hNLP/cNLP Community



- 1 Eingrenzung der Fragestellung
- 2 Die Möglichkeiten von NLP
- 3 Das Ziel der Nachwuchsgruppe DE.xt
- 4 Gibt es da nicht schon was von <PRODUCT>?**

Mögliche <PRODUCT> Kandidaten

Firmen und Software

- averbis
- ID-Berlin
- Empolis
- 3M Deutschland
- CGM MetaKIS

Einfache Methoden

- Reguläre Ausdrücke
- Listen von Wörtern
- Parsing Regeln und/oder Grammatiken

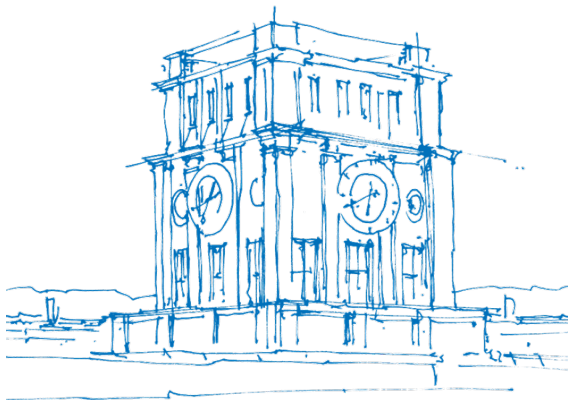
Natural Language Processing im klinischen Kontext

Die Nachwuchsgruppe DE.xt

Luise Modersohn

Lehrstuhl für Medizininformatik
Institut für KI und Informatik in der MEDizin
Technische Universität München

March 21st, 2023



TUM Uhrenturm