UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Harvesting German clinical knowledge graphs from LLMs

Natalia Minakova

Master Thesis Presentation

Supervisor: Prof. Dr. Anette Frank

29th June 2023

Finalists' Colloquium

Overview

- Knowledge Graphs
  - Motivation
  - Clinical Domain
- Prior Work
- Research Questions
- Data and Models
- Prelimenary Experiments and Results
- Evaluation Strategies
- Potential Applications

Knowledge Graphs - Motivation

Definition:

> “a graph of data intended to accumulate and convey knowledge of the real world, whose nodes represent entities of interest and whose edges represent relations between these entities” (Hogan et al. 2021)

- Effective way of representing information in a structured manner: a network of interconnected entities, attributes, and relationships

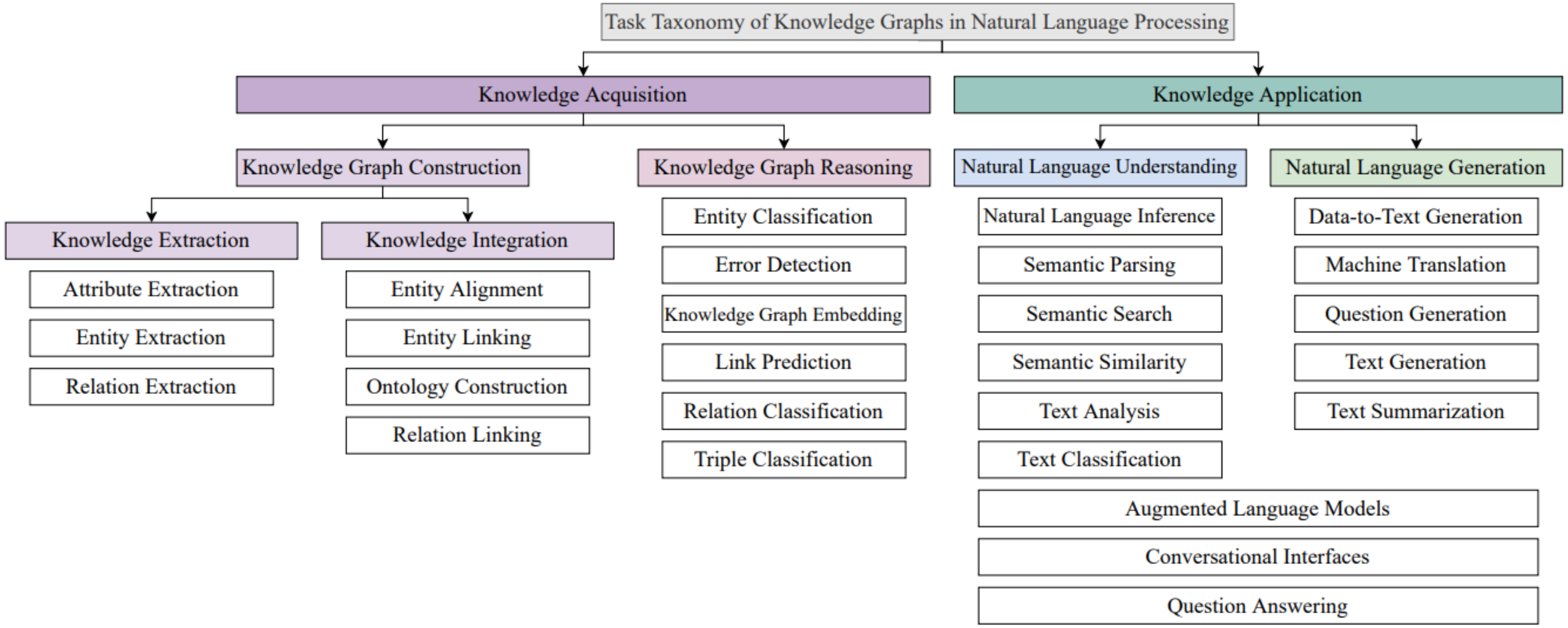


Fig 1. Taxonomy of tasks in the literature on KGs in NLP[1]

[1] Schneider et al. A Decade of Knowledge Graphs in Natural Language Processing: A Survey

Knowledge Graphs in Health Domain

Electronic Health Records contain massive quantities of unstructured patients data

Advantage of using KG in heathcare:

- Providing a unified view of patient's information
- Capturing the semantics and relationships between medical concepts leading to improvements in tasks like Natural Language Understanding
- Supporting clinical decisions:
  - fast, easy, and efficient access to health-related information, enables the physicians to make more informed decisions at the point of care
- Minimizing human errors
- Facilitating Precision Medicine and Personalized Healthcare
- Efficient Information Retrieval
  - more precise and context-aware search queries, for example, to access clinical guides

How can we construct a Knowledge Graph?

- Crowd sourcing

- Text mining pipelines to extract knowledge from text:
    - Entity extraction, coreference resolution, entity linking and relation extraction

What if there was another way?

Prior work

Language Models as Knowledge Bases?
(Petroni et al. 2019)

- Can a LM be queried for relational data the same as a knowledge graph?
- Language models:
  - No schema engineering
  - No need for human annotation
  - Support a open set of queries

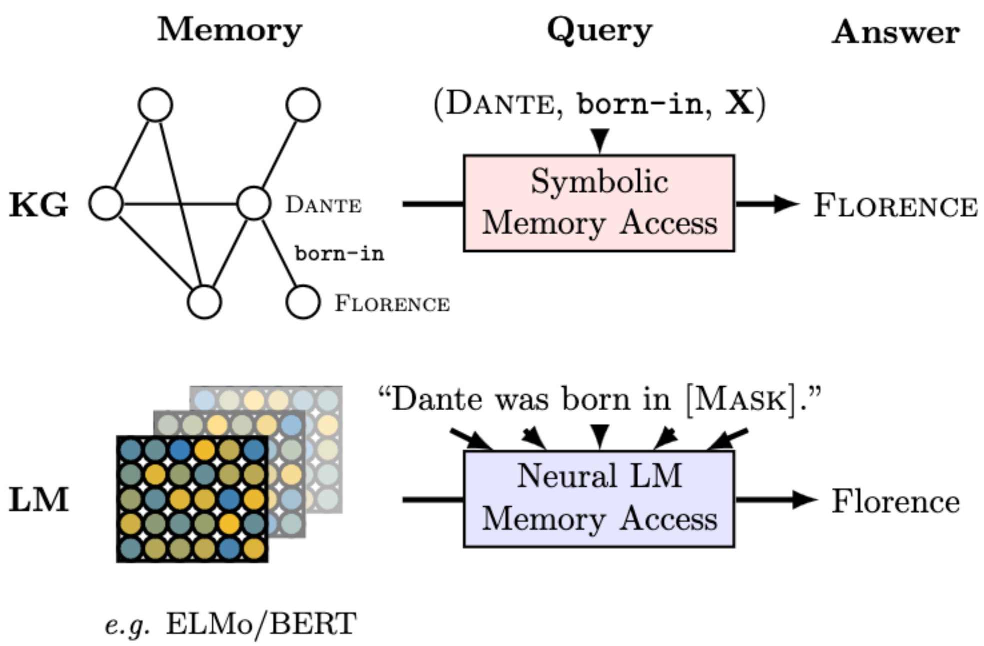


Fig 2. Querying KB and LM for factual knowledge[1]

[1] Petroni et al. 2019 Language Models as Knowledge Bases?

The LAMA (LAnguage Model Analysis) Probe

Questions:

1. Check how much knowledge do off the shelf models like ELMO and BERT store?

2. How does this differ for different types of knowledge such as facts about entities, common sense, and general question answering?

3. How does their performance without fine-tuning compare to symbolic knowledge bases automatically extracted from text?

A pretrained LM knows a fact (subject, relation, object) such as (Dante, born-in, Florence) if it can successfully predict masked objects in cloze sentences such as "Dante was born in [MASK]"

The LAMA (LAnguage Model Analysis) Probe

- Test for relations present in Wikidata and ConceptNet, and knowledge needed for SQUAD
- Evaluate each model based on how highly it ranks the ground truth token against every other word in a fixed candidate vocabulary
- For triples, find the sentence that contains both the subject and the object.
- Mask the object within the sentence and use the sentence as template for querying language models. i.e “[S] was born in [O]” for “place of birth”.

Findings

- (BERT-large) captures (accurate) relational knowledge comparable to that of a knowledge base extracted with an off-the-shelf relation extractor

- Actual knowledge can be recovered surprisingly well from pretrained language models, however, for some relations (particularly N-to-M relations) performance is very poor

- BERT-large consistently outperforms other language models in recovering factual and common-sense knowledge while at the same time being more robust to the phrasing of a query

- BERT-large achieves remarkable results for open-domain QA, reaching 57.1% precision@10 compared to 63.5% of a knowledge base constructed using a task-specific supervised relation extraction system.

Crawling The Internal Knowledge-Base of Language Models (Cohen et al. 2023)

- Crawling procedure: starts from the seed entity and recursively expands it to expose additional facts

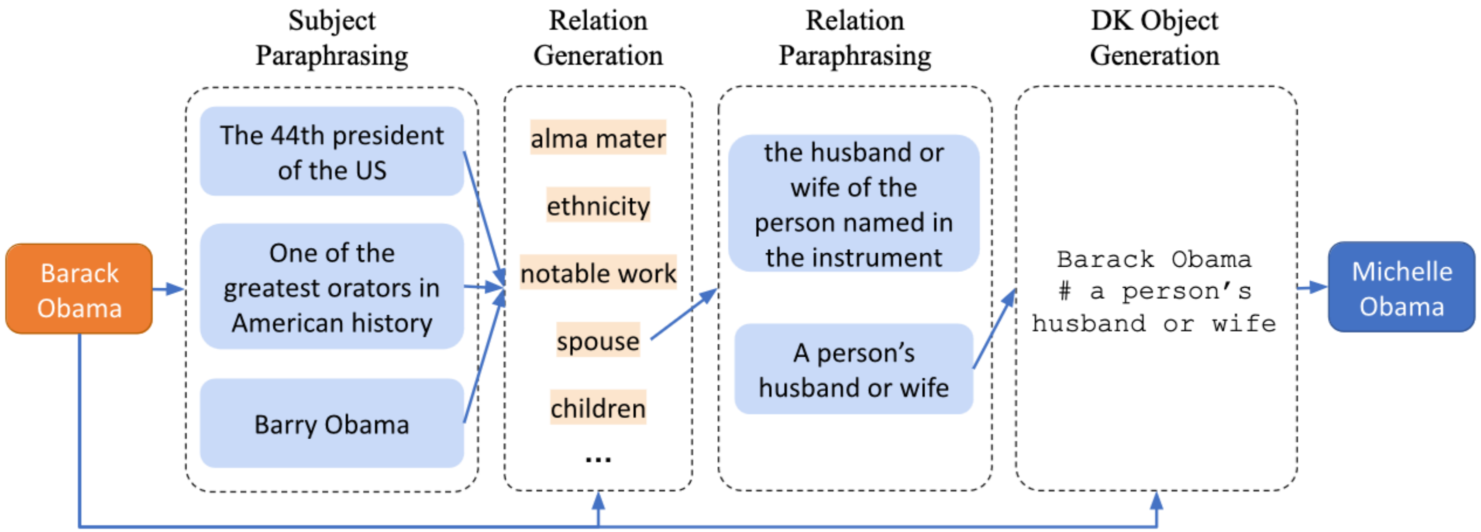


Fig 3. An illustration of the full method for crawling a subgraph [1]

[1] Cohen et al. 2023 "Crawling The Internal Knowledge-Base of Language Models"

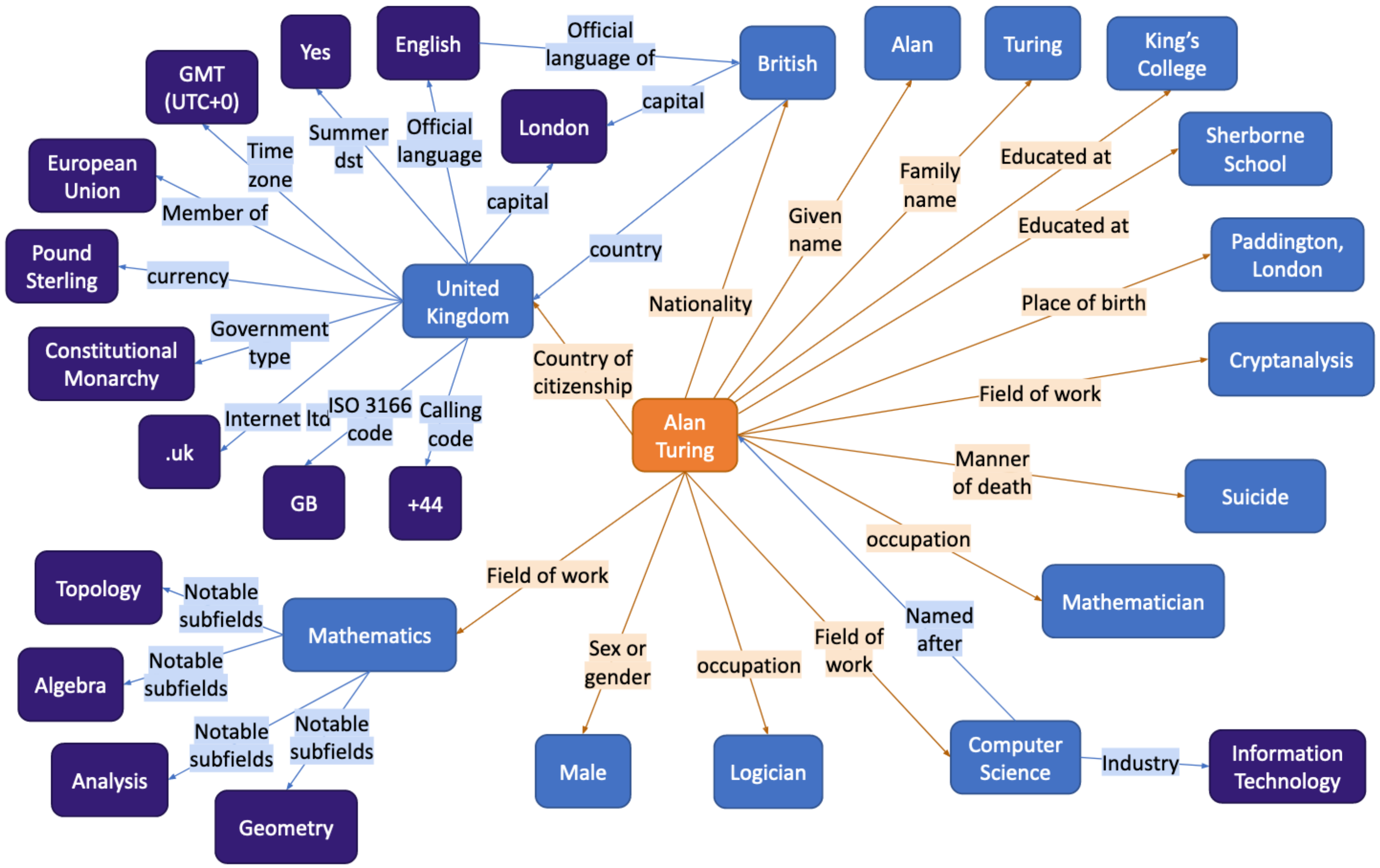


Fig 4. Example of generated depth-2 knowledge graph for entity Alan Turing

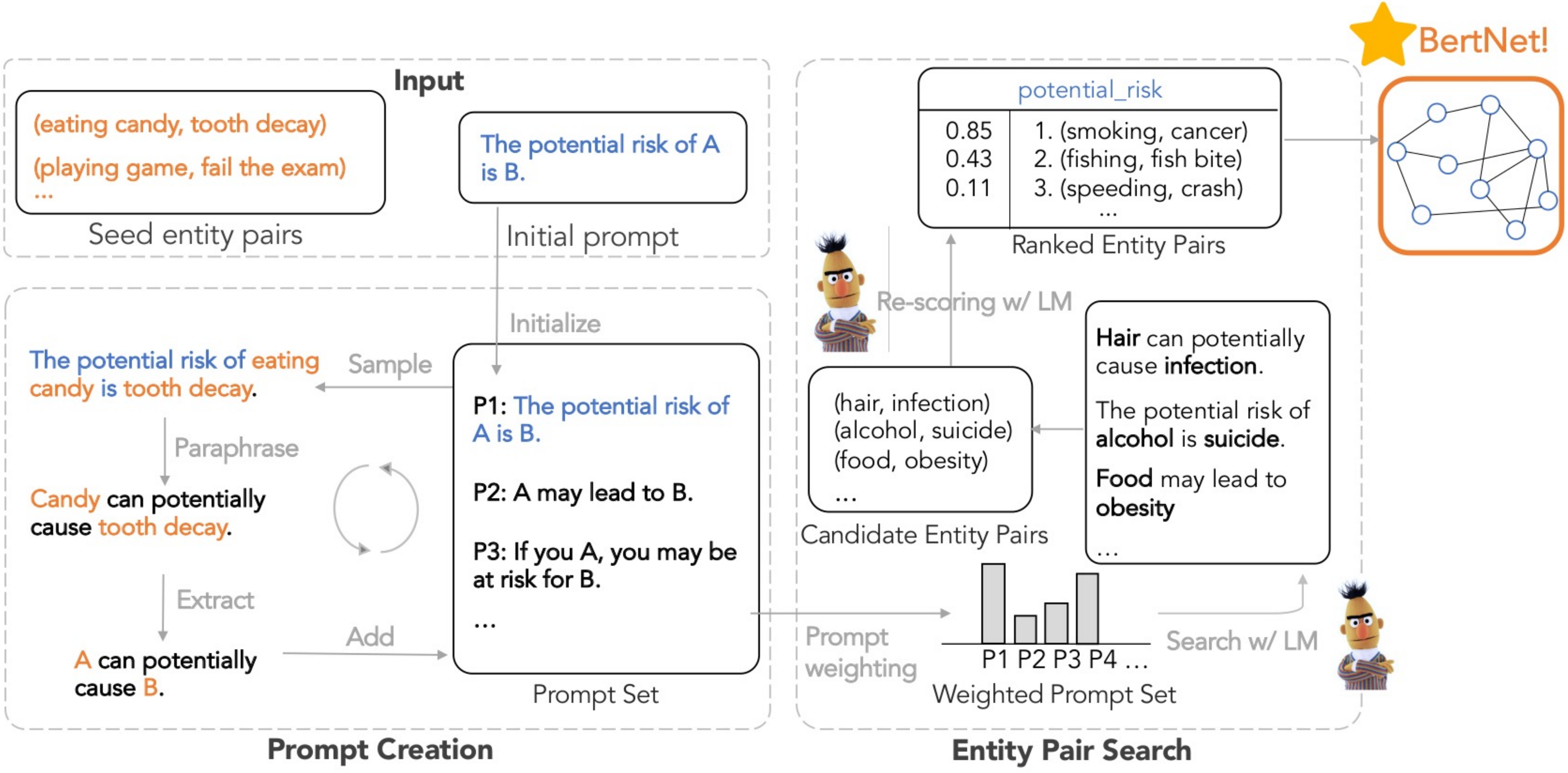
<u>Limitations</u>:

Error propagation:

Once wrong entity is predicted, the number of mistakes might increase

[1] Cohen et al. 2023 "Crawling The Internal Knowledge-Base of Language Models"

BertNet (Hao et al. 2022)



Hao et al. 2022 BertNet: Harvesting Knowledge graphs from pretrained language models

Difference to the previous methods

Only the relation is given as input to predict the head and the tail

The promt offers overall semantics and the seed entities eliminate possible ambiguities

Automatically create diverse weighted prompts describing the same relation

- edit distance to ensure diverse prompts
- Searching for pairs that have high likelihood under the LM
- average compatibility score of each prompt is computed over all seed entity pairs →weight of the prompt is defined as the softmax-normalized score across all prompts

Research Questions

Q1. Can German clinical knowledge be extracted from pre-trained language models?

Q2. What is the level of accuracy achieved by the resulting Knowledge Graph, and which data and models exhibit the highest performance?

Q3. What is the best evaluation strategy?

Resources:

Data, Models, Domain adaption

Don’t Stop Pre-training: Adapt Language Models to Domains and Tasks (Gururungan et al 2020)

Focused on adaptive pre-training:

- continue pre-training the model on a large corpus of unlabeled data
- domain and task adaptive

Findings:

even a model of hundred of millions of parameters struggles to encode the complexity of a single textual domain

pre-training the model towards a specific task or a small corpus can provide significant benefits.

Problems in acquiring clinical data

- Annotated corpora are not only expensive but also often unavailable for research due to patient privacy and confidentiality requirements

- The data is highly confidential and protected under the Health Insurance Portability and Accountability Act (HIPAA)
  - Protected health information (PHI) must be removed from medical records before the records can be shared outside of hospitals
- Anonymization – de-identification of names, date of birth, addresses and other personal information

German clinical data

- Cardiology Doctoral Letters:
  - CARDIO:DE – 500 cardio discharge letters
  - MieDEEP – 500 cardio discharge letters
- Oncology guidelines:
  - GGPONC 2.0 – 10,193 reports
- Oncology reports:
  - BRONCO150- 150 cancer patient reports
- Cardiology guidelines:
  - DGK: Leitlinien der Deutschen Gesellschaft für Kardiologie
  - 128 guidelines

		Guideline	Year	Files
1	·	Pancreatic cancer	2013	292
2	•	Penis cancer	2020	167
3	·	Psycho-oncology	2014	121
4	○	Oral cavity cancer	2021	132
5	·	Malignant ovarian tumors	2020	195
6	•	Anal cancer	2020	216
7	·	Chronic lymphocytic leukemia	2018	285
8	·	Laryngeal cancer	2019	189
9	•	Follicular lymphoma	2020	296
10	·	Oesophageal cancer	2018	172
11	○	Hodgkin lymphoma	2020	253
12	○	Hepatocellular and biliary cancer	2021	263
13	·	Testicular tumors	2020	315
14	·	Prevention of cervix cancer	2020	302
15	○	Renal cell carcinoma	2020	293
16	·	Endometrial cancer	2018	317
17	·	Stomach cancer	2019	246
18	•	Adult soft tissue sarcomas	2021	407
19	·	Actinic keratosis	2020	193
20	○	Malignant melanoma	2020	297
21	○	Cervical cancer	2021	415
22	·	Colorectal cancer	2019	546
23	○	Prostate cancer	2021	351
24	·	Supportive therapy	2020	819
25	·	Lung cancer	2018	665
26	○	Breast cancer	2021	685
27	○	Bladder cancer	2020	364
28	○	Prevention of skin cancer	2021	370
29	○	Palliative medicine	2020	700
30	•	Complementary medicine	2021	327
		Total		**10,193**

	Iteration			
	1a	**1b**	**2**	**3**
Number of annotators	3	7	7	7
Number of documents	5	5	6	3
Number of sentences	149	149	158	67
Number of tokens	4206	4206	3725	1814
IAA (γ)	.75	.89	.93	**.94**
Specification	.71	.87	**.91**	.89
Finding	.82	.93	.95	**.97**
Diagnosis/Pathology	-	.91	.94	**.96**
Other Finding	-	.85	.87	**.91**
Substance	.92	.99	.98	**.99**
Clinical Drug	-	.97	.98	**1.00**
Nutrient/Body Subs.	-	.99	**.99**	.98
External Substance	-	.96	-	**1.00**
Procedure	.82	.93	.96	**.96**
Therapeutic	-	.95	.96	**.96**
Diagnostic	-	.89	**.98**	.93
IAA ($_u\alpha$)	.56	.71	.79	**.85**

Fig 5. Overview of guidelines in the current GGPONC release[1]

[1] Borchert et al. GGPONC 2.0 — The German Clinical Guideline Corpus for Oncology: Curation Workflow, Annotation Policy, Baseline NER Taggers

Sample annotations – BRONCO150

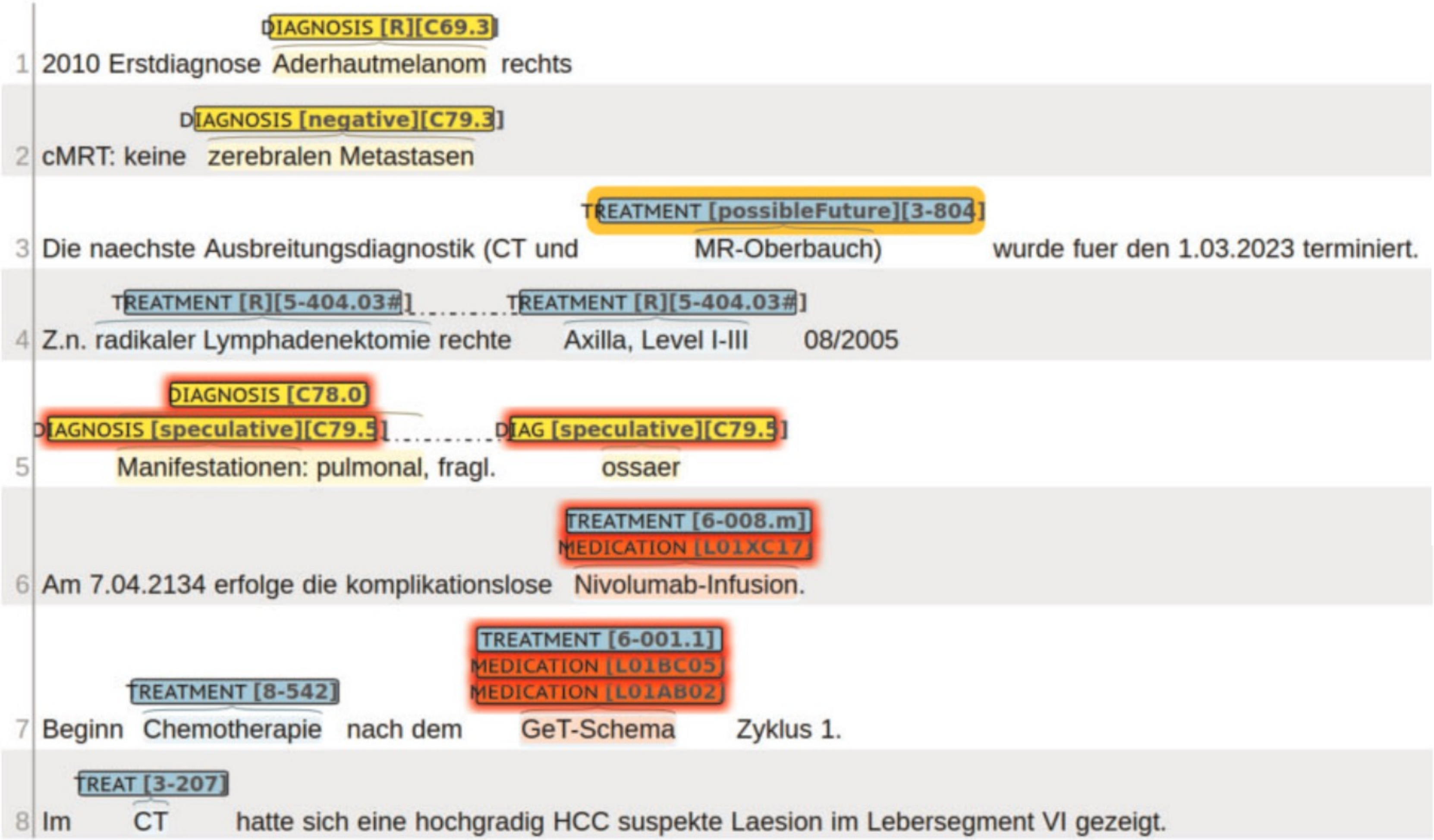


Fig 6a. Exemplary excerpts from original discharge summaries and annotations [1]

Annotation type	BRONCO150
Diagnosis	4080
Treatment	3050
Medication	1630
Total medical entities	8760
Laterality	1033
Negation	503
Speculation	474
Possible future	479
Total attributes	2489
#Documents	150
#Sentences	8976
#Tokens	70 572

Fig 6b. Annotations types[1]

[1] Kittner et al. 2021 Annotation and initial evaluation of a large annotated German oncological corpus

CARDIO:DE annotations

Universitätsklinikum Musterstadt Station Sowieso | Beispielstr. 12 | 12345 Musterstadt

Test-Klinik
Zentrum für Kardiologie

Klinik für Kardiologie
Station II

Dr. med. Muster
Ärztlicher Direktor
Station II

Station Sowieso
Besipiestr. 123
12345 Musterstadt
Tel +123 23 45 67
Fax +123 23 45 66

01.01.2010

Frau
Dr. med. Paul Beispiel
Musterplatz 1
56789 Beispielstadt

Nachrichtlich:
Herrn Max Mustermann, Beispielplatz 1, 12345 Musterstadt

Sehr geehrter Herr Kollege Muster,

wir berichten über Ihre Patientin Frau Maxima Musterfrau geboren am 01.01.1970, wohnhaft in 12345 Musterstadt, Beispielstr. 1, die sich vom bis in unserer stationären Behandlung befand.

Diagnosen:

- Schwerer Infarkt der … am 01.02
- Cvrf: **Hyperlipidämie, Nikotinkonsum seit 01.01.1980, 30 py.**
- Allergien: Hausstaub

Anamnese:

Die stationäre Übernahme von Frau Musterfrau erfolgte über die Chirurgie. Die Patientin klagt über **Tachykardien**. Auf gezielte Nachfrage eingeschränkte Belastbarkeit, **belastungsabhängiges thorakales Druck- und Engegefühl** außerdem **progrediente Belastungsdyspnoe**. Es bestehen **Ödeme bds.**, **kein Schwindelgefühl**, **keine Synkopen**.

Wir danken für die vertrauensvolle Zusammenarbeit und stehen bei Rückfragen selbstverständlich jederzeit gerne zur Verfügung.

Labor:

Bezeichnung	Wert	Datum
Abc	123	01.01.2010

Medikation:

ASS 50mg 1-0-0
Clexane 12mg 0-1-1 bis Mai 2011

Mit freundlichen Grüßen

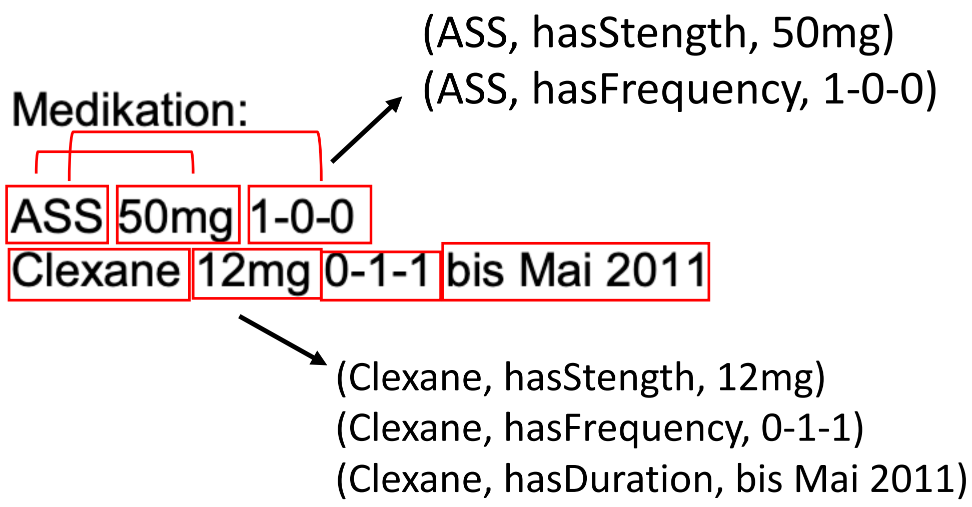
Dr. med. Muster
Ärztl. Direktor

Dr. Platzhalter
Oberarzt

Fig 7. German dummy doctor's letter from cardiology domain[1]

Medication information:
(ActiveIng, Dosage, Drug, Duration, Form, Frequency, Reason, Route, Strength)

Section types:
(Abschluss, Anamnese, Anrede, Diagnosen, AufnahmeMedikation, Befunde, EchoBefunde, AktuellDiagnosen, EntlassMedikation, KuBefunde, Labor, Anderes, RisikofaktorenAllergien, Zusammenfassung)



[1] Richter et al. 2023 a distributable German clinical corpus containing cardiovascular clinical routine doctor's letters

Models - German BERT

Training data:

- German Wikipedia(6GB of text files)
- OpenLegalData (2.4GB)
- News articles (3.6GB)

Model	germEval18Fine	germEval18Coarse	germEval14	CONLL03	10kGNAD
multilingual cased	0.441	0.71	0.834	0.792	0.888
multilingual uncased	0.461	0.731	0.823	0.786	0.901
German BERT cased (**ours**)	**0.488**	**0.747**	**0.84**	**0.804**	**0.905**

Table 1. Performance of the model[1]

[1] https://huggingface.co/bert-base-german-cased

medBERT.de:

A Comprehensive German BERT Model for the Medical Domain

- Trained on a diverse set of medical texts:
  - Scientific texts, medical books, and hospital data from various medical domains

Source	No. Documents	No. Sentences	No. Words	Size (MB)
DocCheck Flexikon	63,840	720,404	12,299,257	92
GGPONC 1.0 [2]	4,369	66,256	1,194,345	10
Webcrawl [24]	11,322	635,806	9,323,774	65
PubMed abstracts	12,139	108,936	1,983,752	16
Radiology reports	3,657,801	60,839,123	520,717,615	4,195
Spinger Nature	257,999	14,183,396	259,284,884	1,986
Electronic health records [22]	373,421	4,603,461	69,639,020	440
Doctoral theses	7,486	4,665,850	90,380,880	648
Thieme Publishing Group	330,994	10,445,580	186,200,935	2,898
Wikipedia	3,639	161,714	2,799,787	22
Summary	**4,723,010**	**96,430,526**	**1,153,824,249**	**10,372**

Table 2. Data Sources for medBERT.de[1]

Model	AUROC	Macro F1	Micro F1	Precision	Recall
Chest CT					
GottBERT	92.48	69.06	83.98	76.55	65.92
BioGottBERT	92.71	69.42	83.41	80.67	65.52
Multilingual BERT	91.90	66.31	80.86	68.37	65.82
German-MedBERT	92.48	66.40	81.41	72.77	62.37
medBERT.de	**96.69**	**81.46**	**89.39**	**87.88**	**78.77**
medBERT.de$_{dedup}$	96.39	78.77	89.24	84.29	76.01
Chest X-Ray					
GottBERT	83.18	64.86	74.18	59.67	78.87
BioGottBERT	83.48	64.18	74.87	59.04	78.90
Multilingual BERT	82.43	63.23	73.92	56.67	75.33
German-MedBERT	83.22	63.13	75.39	55.66	78.03
medBERT.de	**84.65**	**67.06**	**76.20**	**60.44**	**83.08**
medBERT.de$_{dedup}$	84.42	66.92	76.26	60.31	82.99
ICD-10 code classification on discharge notes					
GottBERT	77.23	18.32	51.23	38.30	14.27
BioGottBERT	78.01	17.96	50.56	35.97	13.95
Multilingual BERT	76.64	19.48	51.19	38.39	15.60
German-MedBERT	75.44	**23.41**	53.63	41.39	**18.94**
medBERT.de	80.78	**23.41**	**53.84**	**41.42**	18.75
medBERT.de$_{dedup}$	**80.84**	21.44	52.46	40.45	17.04

Table 3. Data Sources for medBERT.de[1]

[1] Bressem et al. 2023 MEDBERT.DE: A COMPREHENSIVE GERMAN BERT MODEL FOR THE MEDICAL DOMAIN

Relation List – (based on SNOMED CT)

Group 1:

- hasMedication
- HasActiveIngredient
- hasDrugForm
- hasSymptoms
- isRecommenedForIllness
- hasDose

Group 2:

- hasTreatment
- hasContraindications
- HasLabTest
- HasPharmacologicalPlan
- HasAdministrationProcess
- hasTreatmentDuration
- hasRiskFactors

Group 3:

- hasAssesment
- AffectedBodySite
- hasTreatmentTarget
- HasComplication
- OccursWith
- hasExplanation
- canBeCombinedWith

Prelimenary Experiments and Results

First Experiments

1. German BERT and medBERT.de
2. German BERT and medBERT.de further pre-trained
3. German BERT and medBERT.de further pre-trained + contextualization

Relations:

- hasMedication, HasActiveIngredient, hasDrugForm, hasSymptoms, isRecommenedForIllness, hasDose

Results for relation: **isRecommendForIllness**

Initial Promt:

<ENT0> wird für Krankheit <ENT1> empfohlen.


```
"seed_ent_tuples": [
    [
        "fasten",
        "akuten Cholezystitis"
    ],
    [
        "abschwellende Mittel",
        "Allergien"
    ],
    [
        "Chemotherapie",
        "Krebs"
    ],
    [
        "Betablocker",
        "Angina"
    ],
    [
        "Alendronate",
        "Osteoporose"
    ]
],
```


1a. German BERT:

(Fitness, Stress), (Alkohol, Krebs), (Wasser, Erkrankungen), (Zucker, Diabetes), (Training, Verletzungen), (Tourismus, reisen)

1b. German medBERT:

(Leber, Leberzirrhose), (Akupunktur, Schmerzen), (Sauerstoff, Asthma), (Alkohol, Depressionen), (Pflege, Demenz)

Results for relation: isRecommendForIllness

Initial Promt:

<ENT0> wird für Krankheit <ENT1> empfohlen.

2a. German BERT further pre-trained:

(Zahnarzt, Parodontitis), (Dopamin, Parkinson), (Antidepressiva, Depression), (Schlaf, Angst)

2b. German MedBERT further pre-trained:

(Tamoxifen, Brustkrebs) (Cisplatin, Krebs) (Ibuprofen, Migräne) (Paracetamol, Migräne)

Correct, but wrong relationship

(Ibuprofen, Schwindel) – side effect

(Cholesterin, Bluthochdruck) - causation

Results for relation: **isRecommendForIllness**

Initial Promt:

Kardiologen empfehlen die Verwendung von <ENT0> bei <ENT1>

<u>3b. German medBERT.de + contextualizing:</u>

- (Adrenalin, Herzinsuffizienz), (Magnesium, Herzrhythmusstörungen)
- (Antikoagulation, Vorhofflimmern), (Bettruhe, Herzrhythmusstörungen)
- (Tamoxifen, BrustKrebs)

Correct, but wrong relation:

- (Heparin, Blutung) – side effect
- (Koronararterien, Koronare Herzkrankheit) - location
- (Koronararterien, koronare Stenosen) - location


```
"seed_ent_tuples": [
    [
        "Betablocker",
        "Koronare Herzkrankheit"
    ],
    [
        "Implantation von Herzschrittmacher",
        "Arrhythmien"
    ],
    [
        "Betablocker",
        "Hypertonie"
    ],
    [
        "Betablocker",
        "Angina"
    ],
    [
        "Notfallmedizinische Versorgung",
        "Myokardinfarkt"
    ]
],
```

Next steps

- Continue working with contextualization:
  - add more context of the doctoral letters to the prompt
- Modify the framework to allow input of candidate pairs and a relation to return the score
- Experiment with exisiting positive and negative examples for enitities and relation to understand better the assigned scores and discover a thershold

Evaluation

- Some relations can get evaluated based on the annotated relations in CARDIO:DE
- Get relations from SNOMED CT and pass them to the model and see if the relations get the same output.
  - Check for threshold by passing positive and negative examples to evaluate the assigned score
- Check against the doctoral letters:
  - Give the disease name and ask for medication and check against the real medication
  - Give medication name and predict the dosage or frequency

Potential Applications

Medications

- Medication contraindication – > based on previous prescription
- Dosage/strength corrections –> the pharmasist is responsible to check the dosage before giving out the medication

Diagnosis

- Symptoms across patient visits --> diagnosis that might not be obvious

Thank you for your attention