

GeMTeX

German Medical Text Corpus

Methodenplattform der MII in der Förderperiode 2023 - 2026

Martin Boeker
Medizinische Informatik, AIIM, TUM



2022-08-25 GMDS AG Terminologie

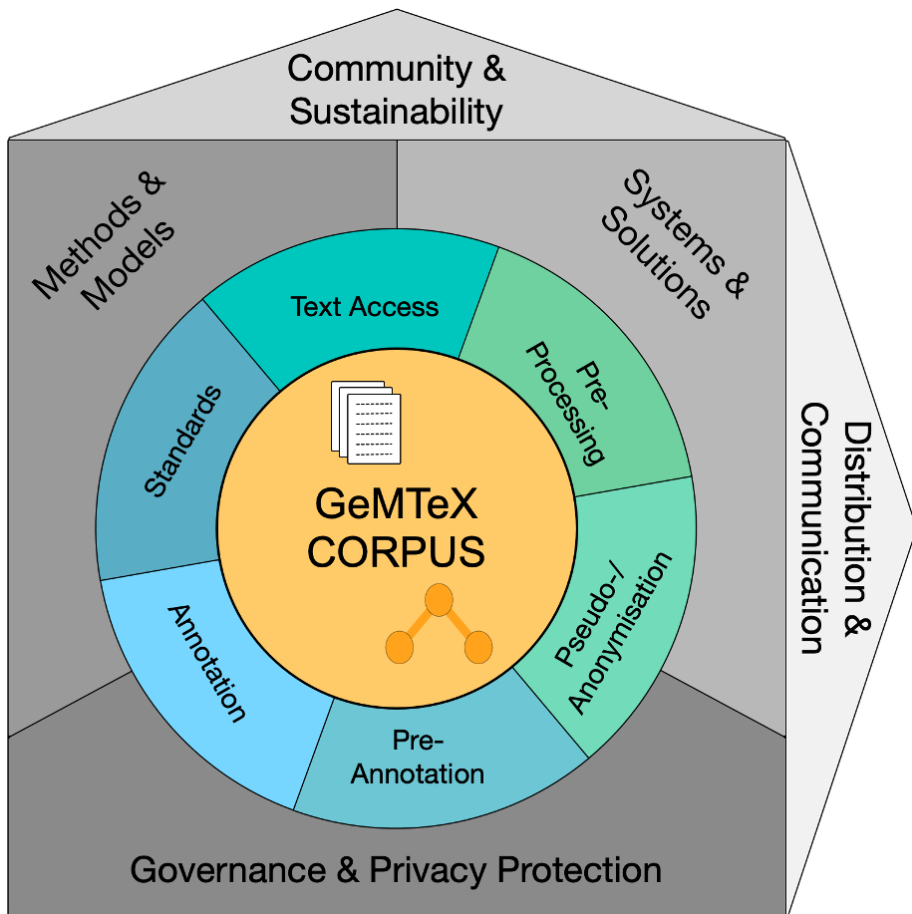
Hintergrund

- Vielzahl nicht strukturiert erschlossener Informationen in *klinischen* Texten
 - Arztbriefe
 - Befundberichte (z. B. Bildgebung, Pathologie)
 - Berichte über Prozeduren (z. B. OP)
 - Anamnesen

- Erfolgreiche Nutzung von NLP besonders im englischsprachigen Raum
- NLP an deutschsprachigen klinischen Texten: vergleichsweise geringe Fortschritte
- Größtes Hindernis in Deutschland und für die MII:
Fehlen von deutschsprachigen (annotierten) klinischen Texten (Textkorpus)

Zielsetzung GeMTeX

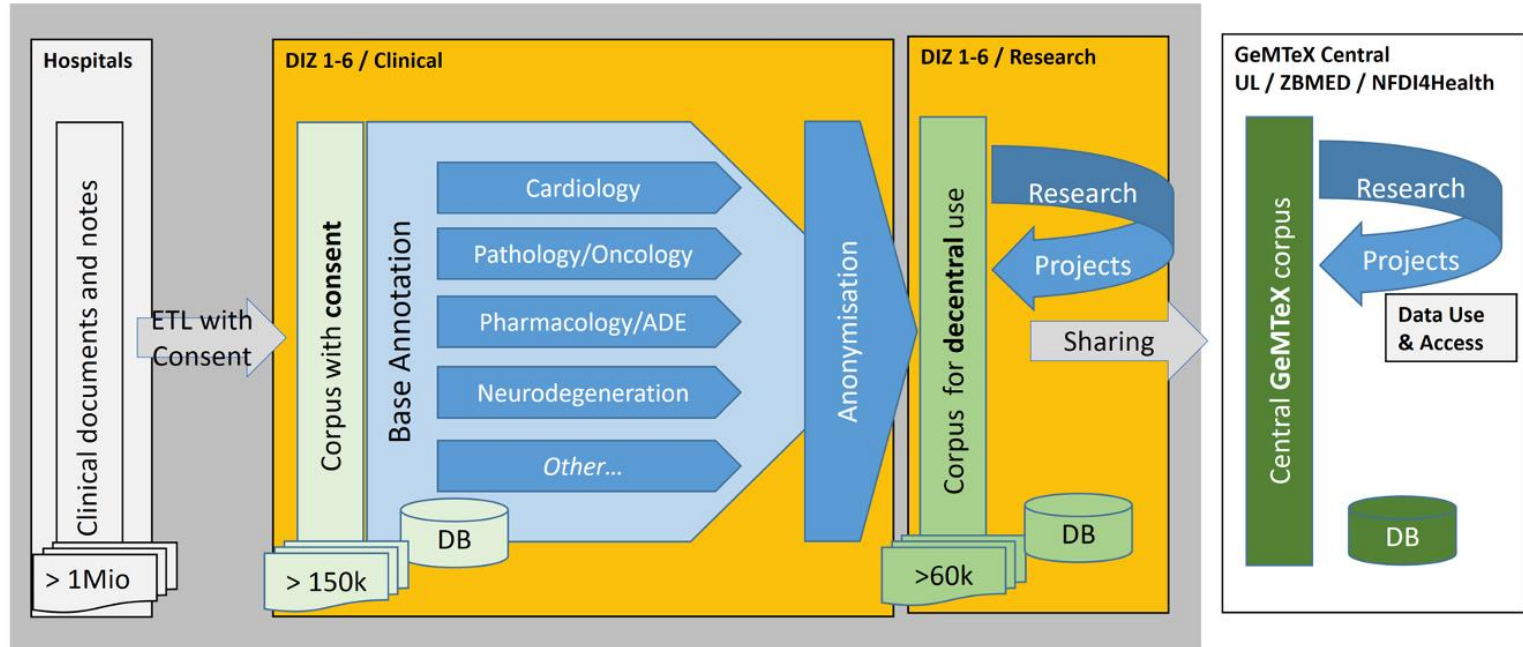
- Deutschen medizinischer (klinischer) Referenz-Korpus der MII
- Prospektive Textdaten als Ressource für die Forschung
 - Semantische Goldstandard Annotationen
 - Trainierte Sprachmodelle
 - Algorithmische Auswertung
- Nutzung von NLP im Rahmen der DIZ
- Initialisierung von Folgevorhaben
 - Demonstration von Vorteilen der semantischen Textanalyse für die Krankenversorgung



Anknüpfungspunkte GeMTeX an die MII


Methodenplattform mit

- Erweiterung des Datenschutzkonzeptes für (de-identifizierte) Texte
- Nutzung des Broad Consent und Erweiterung der Governance
- Definition unterschiedlicher Integrationskonzepte (Nutzungsszenarien)
- Standardisierung der Nutzung von Texten
- Bereitstellung von Methoden
 - Nutzung von Texten
 - Annotation
 - Training von Modellen
- Anwendungen von (förderierten) Verfahren der KI



 Annotation Project (DIC specific)

 Annotation Database

 summarized # of texts over all 6 sites

 DIC/Site area

Struktur des Konsortiums

München (TU)	DIFUTURE	
Leipzig		SMITH
Essen		SMITH
Erlangen	MIRACUM	
Dresden	MIRACUM	
Berlin		HIGHMed

Einbindung von Methoden-Partnern

Einbindung für DIZ-Standorte (data partner)

Firmenpartner: AVERBIS, ID Berlin

Internationale Partner: Medizinische Universität Graz

DIFUTURE
Data Integration for Future Medicine
Boeker und NWG (Modersohn)

SMITH
Smart Medical Information
Technology for Healthcare
Ganslandt

miracum
Medical Informatics in Research and Care in University Medicine
Sedlmayr
Prasser

HiGHmed
Medical Informatics

Aufbau eines deutschsprachigen klinischen Korpus

- Ausleitung von Textdokumente und textuelle Inhalte aus den KIS
 - in allen Formen
 - Tooling wird zur Verfügung gestellt
- De-Identifikation/Anonymisierung
- Annotation
- Nutzungs-Integration
 - lokal
 - zentral aggregiert
 - föderiertes Lernen und Modellintegration
- Governance und Privatheit
- Evaluation des Korpus
- Standardisierung der Prozesse und Repräsentation

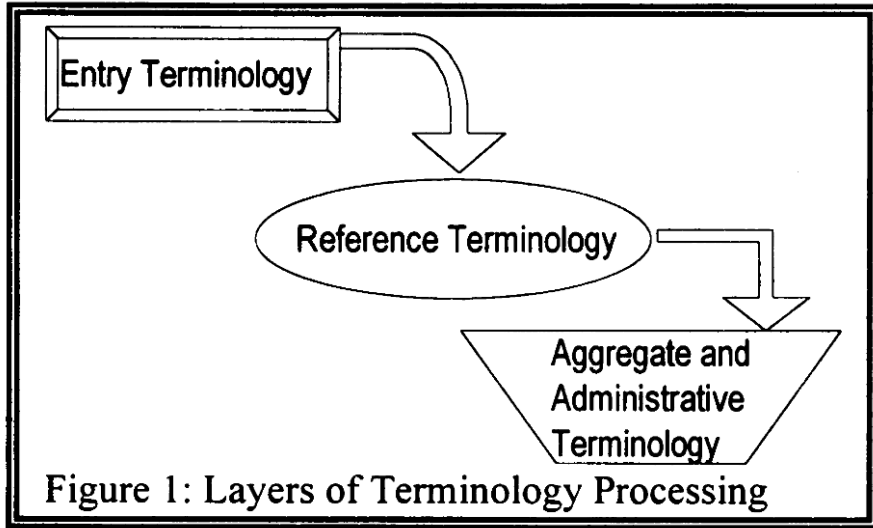
Annotationsmethodik

- Aufbauend auf bestehenden Methoden
 - Annotationsguideline(s)
 - Annotationsterminologie(n)
 - Automatisierte Vorannotation
 - Annotationseditor
- Annotation mit Medizinstudenten an den Standorten
- Integration von interaktivem Lernen

Bedeutung von Terminologien in GeMTeX

- Standardisierte Annotationsvokabularien (Referenzterminologien)
 - SNOMED CT
 - u. a.
- (teilweise) fehlende Übersetzungen
- Übersetzungen der pref. terms von Referenzterminologien reflektieren die klinische Sprache nicht
⇒ Nutzung von Interface Terminologien

Interface Terminology - Terminologie klinischer Sprache



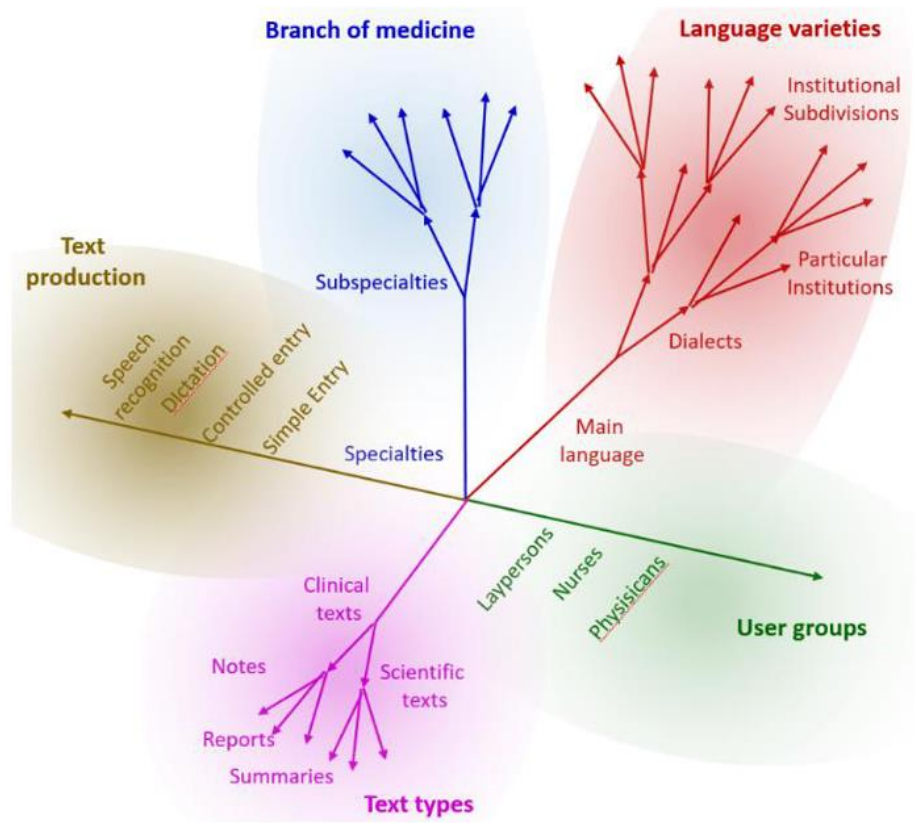
a systematic collection of health care–related phrases (terms) that supports clinicians’ entry of patient-related information into computer programs, ...

These terminologies generally embody a rich set of flexible, user-friendly phrases displayed in the graphical or text interfaces of specific computer programs.

Chute CG, Elkin PL, Sherertz DD, Tuttle MS. Desiderata for a clinical terminology server. Proc AMIA Symp. 1999;42–6.

Rosenbloom ST, Miller RA, Johnson KB, Elkin PL, Brown SH. Interface Terminologies: Facilitating Direct Entry of Clinical Data into Electronic Health Record Systems.

Journal of the American Medical Informatics Association. 2006 May 1;13(3):277–88.



Experimentelle deutsche Interface Terminology

- Mapping auf SNOMED CT
- Verfügbar: <https://user.medunigraz.at/stefan.schulz/mugit/>
 - <https://user.medunigraz.at/stefan.schulz/mugit/readme.txt>

GeMTeX

German Medical Text Corpus

Methodenplattform der MII in der Förderperiode 2023 - 2026

Martin Boeker
Medizinische Informatik, AIIM, TUM



2022-08-25 GMDS AG Terminologie