MEDIZIN INFORMATIK INITIATIVE

GeMTeX

GeMTeX – German Medical Text Corpus

MEDIZINISCHE TEXTE FÜR DIE FORSCHUNG AUTOMATISIERT ERSCHLIESSEN

Medizinische Texte aus der Routineversorgung enthalten hohe Mengen an komplexen und unstrukturierten Daten zu beispielsweise Krankheitsverläufen, Diagnosen und Therapien. Diese Daten können sehr nützlich für die Forschung und Patientenversorgung sein. Allerdings unterscheiden sich die Texte aus der klinischen Dokumentation zwischen den Einrichtungen in Struktur und Inhalt häufig sehr stark. Sie lassen sich deshalb nur schwer für die automatische Verarbeitung natürlicher Sprache nutzen, die wiederum die Grundlage für sämtliche Automatisierungsprozesse und Analysen bildet. Aufgrund der fehlenden Standardisierung medizinischer Freitexte kann das Potenzial dieses Datenschatzes nicht voll ausgeschöpft werden. Hier setzt die Methodenplattform GeMTeX an.

EINE GROSSE SAMMLUNG AN MEDIZINISCHEN TEXTEN IN DEUTSCHER SPRACHE ENTSTEHT

Die Methodenplattform GeMTeX ist ein konsortienübergreifendes Projekt der Medizininformatik-Initiative (MII) mit dem Ziel, medizinische Texte aus der Patientenversorgung wie z. B. Arzt- oder Entlassbriefe für Forschungsprojekte zugänglich zu machen. Durch die Zusammenarbeit der MII-Konsortien DIFUTURE, HiGHmed, MIRACUM und SMITH soll das größte medizinische Textkorpus in deutscher Sprache entstehen. Koordiniert wird das Projekt durch die Geschäftsstelle des SMITH-Konsortiums.

Mit dem Einverständnis der Patientinnen und Patienten werden Dokumente aus den elektronischen Gesundheitsakten (ePA) der sechs universitätsmedizinischen Standorte München, Leipzig, Essen, Berlin, Dresden und Erlangen gesammelt.

Anschließend werden die Dokumente mit Methoden der natürlichen Sprachverarbeitung (NLP) aufbereitet und in anonymisierter Form für Forschungszwecke zur Verfügung gestellt. Mit der entstehenden Datenbank können z. B. KI-Modelle trainiert und im klinischen Alltag erprobt werden.

ZIELE

- Eine breite Datenbasis für medizinische Forschungsprojekte und KI-Modelle mit dem Ziel der klinischen Anwendung schaffen.
- Das größte Textkorpus medizinischer Texte in deutscher Sprache erstellen.
- Texte aus der routinemäßigen Patientenversorgung maschinenlesbar aufbereiten und für die Forschung verfügbar machen.
- Technische und organisatorische Standards für die Abbildung von medizinischen Texten und deren Strukturierung etablieren.
- Den Kerndatensatz der Medizininformatik-Initiative (MII) erweitern.

GeMTeX ist am 1. Juni 2023 gestartet und wird durch das Bundesministerium für Bildung und Forschung (BMBF) bis zum 31. August 2026 mit rund sieben Millionen Euro gefördert.

MEDIZIN INFORMATIK INITIATIVE

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

→ **Mehr Informationen: https://www.smith.care/de/gemtex_mii/ueber-gemtex/**

MEDIZIN INFORMATIK INITIATIVE

GeMTeX

Projektpartner

KONSORTIALLEITUNG

München
Technische Universität München

KONSORTIALPARTNER

Berlin
- Charité – Universitätsmedizin Berlin
- ID GmbH & Co. KGaA

Darmstadt
- Technische Universität Darmstadt

Dresden
- Technische Universität Dresden

Erlangen
- Universitätsklinikum Erlangen

Essen
- Universitätsmedizin Essen

Freiburg
- Averbis GmbH

Hannover
- Medizinische Hochschule Hannover

Heidelberg
- Universitätsklinikum Heidelberg

Köln
- Deutsche Zentralbibliothek für Medizin (ZBMED)

Leipzig
- Universitätsklinikum Leipzig / Universität Leipzig

München
- Ludwig-Maximilians-Universität München

Münster
- Universität Münster

Potsdam
- Hasso-Plattner-Institut für Digital Engineering gGmbH

Tübingen
- Universitätsklinikum Tübingen

ASSOZIIERTE PARTNER

Graz
- Medizinische Universität Graz

KOORDINATIONSSTELLE

Berlin
- Geschäftsstelle TMF e. V.

DIFUTURE
Data Integration for Future Medicine

HiGHmed
Medical Informatics

miracum

SMITH
Smart Medical Information Technology for Healthcare

Kontakt

VERBUNDKOORDINATION

Prof. Dr. Martin Boeker
Verbundkoordinator GeMTeX
Konsortialleiter DIFUTURE

Technische Universität München /
Klinikum rechts der Isar

PROJEKTKOORDINATION

Universität Leipzig | Medizinische Fakultät
SMITH-Geschäftsstelle
Philipp-Rosenthal-Straße 27
04103 Leipzig

Telefon: +49 341 97-16720
E-Mail: info@smith.care
Web: www.smith.care

Stand: Februar 2024