Diplomarbeit

Identifikation laiensprachlicher Ausdrücke durch Methoden der künstlichen Intelligenz

eingereicht von

Nina Reithofer

zur Erlangung des akademischen Grades

Doktorin der gesamten Heilkunde

(Dr. med. univ.)

an der

Medizinischen Universität Graz

ausgeführt am

Institut für Medizinische Informatik, Statistik und Dokumentation

unter der Anleitung von

Univ.-Prof. Dr.med. Stefan Schulz und

und

Ass.-Prof. Dipl.-Ing. Dr.scient.med. Markus Eduard Kreuzthaler

Graz, 02.08.2021

Eidesstattliche Erklärung

Ich erkläre ehrenwörtlich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst habe, andere als die angegebenen Quellen nicht verwendet habe und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am 02.08.2021 *Nina Reithofer eh*

Danksagungen

An dieser Stelle möchte ich mich bei all denjenigen bedanken, die mich während der Anfertigung dieser Diplomarbeit unterstützt und motiviert haben.

Zunächst gebührt mein Dank Univ.-Prof. Dr.med. Stefan Schulz, der meine Diplomarbeit als Erstbetreuer begutachtet hat. Für die hilfreichen Anregungen und konstruktive Kritik bei der Erstellung dieser Arbeit möchte ich mich herzlich bedanken.

Ein besonderer Dank gilt meinem Zweitbetreuer Ass.-Prof. Dipl.-Ing. Dr.scient.med. Markus Eduard Kreuzthaler, ohne seine Hilfe hätte diese Arbeit nie entstehen können. Mein Dank gilt besonders seiner Hilfestellung, den interessanten Beiträgen sowie der Unterstützung bei der technischen Umsetzung der Arbeit.

Abschließend möchte ich meine Familie, meine Freunde und meinen Partner dankend erwähnen. Vielen Dank für eure Unterstützung bei all meinen Entscheidungen und euren emotionalen Rückhalt.

Inhaltsverzeichnis

Abkürzungsverzeichnis

AIDS	Acquired Immune Deficiency Syndrome
AMIA	American Medical Informatics Association
CPU	Central Processing Unit
CT	Computertomographie
EKG	Elektrokardiogramm
GAP	Gut informierte Kommunikation zwischen Arzt und Patient
GPU	Graphics Processing Unit
MEDLINE	Medical Literature Analysis and Retrieval System Online
MRR	Mean Reciprocal Rank
NLP	Natural Language Processing
OOV	Out of Vocabulary
P	Precision
PubMed	Englischsprachige textbasierte Meta-Datenbank
R	Recall
SIDA	Syndrome Immuno-Déficitaire Acquis
SIDS	Sudden Infant Death Syndrome
SNOMED-CT	Systematized Nomenclature of Medicine Clinical Terms
TS	Trikuspidalstenose
t-SNE	t-Distributed Stochastic Neighbor Embedding
TT	Transliterated Title
UMLS	Unified Medical Language System

Glossar

Experte:
Der Term "Experte" wird für Personen des ärztlichen Berufs und Berufe des Gesundheitswesens wie beispielsweise Gesundheits- und Krankenpflege sowie Diätologie und weitere verwendet. Der Begriff wurde aufgrund der Eindeutigkeit gewählt. Da es hier kein zufriedenstellendes genderneutrales Wort gibt, wird darauf ausgewichen, es werden aber alle Geschlechter damit angesprochen.

Laie:
Das Wort "Laie" wird für Personen verwendet, welche keinem medizinischen Beruf oder vergleichbaren Berufen, wie etwa Krankenpflege, Diätologie nachgehen. Er wurde aufgrund der Geschlechtsneutralität und der Eindeutigkeit des Wortes gewählt.

Term:
Als Term werden sprachliche Ausdrücke bezeichnet, die aus einem Wort oder mehreren Wörtern bestehen können und denen eine definierte Bedeutung zugewiesen werden kann.

Abbildungsverzeichnis

Tabellenverzeichnis

Zusammenfassung

Hintergrund: In welchem Ausmaß und mit welchen Methoden lässt sich maschinelles Lernen zur Identifikation laiensprachlicher Terme wie Einzelwörter, Komposita, kurze Nominalphrasen, die für eine umschriebene Bedeutung in einem Fachgebiet stehen, im Kontext von Medizin und Gesundheit anwenden? Die Fragestellung ist von Bedeutung, da die Unterteilung die Grundlage für die Anpassung klinischer Texte an Leser ohne medizinisches Fachwissen, sogenannte Laien, darstellt.

Methoden: Durch die adäquate Klassifizierung anhand von Annotationsrichtlinien sollten objektive Kriterien gefunden werden, um medizinische Terme in Experten- und Laienterme zu unterteilen. Der theoretische Kern der Arbeit besteht darin, ein geeignetes aktuelles Verfahren des maschinellen Lernens zu trainieren, welches gut zwischen sprachlichen Mustern unterscheiden und anhand derer medizinische Terme binär klassifizieren kann.

Ergebnis: Der verwendete Algorithmus (fastText) konnte Laien- von Expertentermen ohne Verwendung eines vortrainierten Sprachmodells mit einem F_1-Score von 0,76 und unter Verwendung eines mit deutschen Texten vortrainierten Sprachmodells mit einem F_1-Score von 0,75 unterscheiden. Hierzu war basierend auf den Annotationsrichtlinien ein Goldstandard erstellt worden (Cohens Kappa κ = 0,71).

Diskussion: Indem Medizinterme automatisch als laiengerecht markiert werden, können medizinische AutorInnen bei der Erstellung patientengerechter Dokumente unterstützt werden. Des Weiteren kann ein durch den Algorithmus mit Information zur Laienverständlichkeit angereicherter Thesaurus in Anwendungen integriert werden, die das Lesen medizinischer Texte durch Laien erleichtern.

Abstract

Introduction: To what extent and with which methods can machine learning be applied to identify lay language terms (single words, compounds, short nominal phrases that have a precise meaning in a domain) in the context of medicine and health? This issue is important because the lay-expert distinction is the basis for adapting clinical texts to laypersons, i.e. readers without medical expertise.

Methods: Based on annotation guidelines, objective criteria should be found to classify medical terms into expert and lay terms.
As a result, an appropriate automatic classification of terms from medical texts into expert and lay terms is expected. The theoretical core of the work is to use an appropriate machine learning method in order to find out how well it can distinguish terminology according to linguistic patterns and classify them in binary terms.

Results: The fastText algorithm was able to distinguish lay and expert terms with an F_1-Score of 0.76 without a pre-trained language model, and with an F_1-Score of 0.75 using a language model pre-trained on German texts. Based on the annotation guidelines a gold standard had been created for which an inter-annotator agreement of 0.71 (Cohens Kappa κ) had been measured.

Discussion: By automatically tagging medical terms as understandable by laypersons, medical authors can be supported in the creation of patient-oriented documents. On the other hand, a thesaurus enhanced by the algorithm can be integrated into applications that simplify reading of medical texts by laypersons.

1. Einleitung

1.1. Motivation

Methoden der künstlichen Intelligenz werden im medizinischen Sektor bereits immer vielversprechender und zielgerichteter angewandt. Künstliche Intelligenz und maschinelles Lernen werden derzeit intensiv erforscht und bieten umfangreiches Potenzial zur Unterstützung und Automatisierung menschlicher Fähigkeiten. Im Bereich der Bilderkennung, beispielsweise zur Klassifizierung von Hautkrebs, zur Erkennung von Rhythmusstörungen im Elektrokardiogramm können auf künstlicher Intelligenz basierende Systeme die Entscheidungsfindung von ÄrztInnen unterstützen und damit zu einer effizienteren Qualitätssicherung beitragen (vgl. Esteva et al., 2017; Feeny et al., 2020).

Im Bereich der Texterkennung und der sprachlichen Unterscheidung von Wörtern, Phrasen und kurzen Sätzen sind insbesondere in der deutschsprachigen Literatur noch wenig wissenschaftliche Arbeiten verfügbar. Insbesondere die mögliche maschinelle Unterscheidung von Laien- und Expertentermen stellt hier einen Wissensgewinn für das Forschungsfeld dar. Der Mehrwert dieser Diplomarbeit soll dazu beitragen, diese Lücke zu füllen und in Zukunft Systeme zu unterstützen, welche Medizinterme binär mit den Etiketten "Experte" und "Laie" versehen können.

1.2. Zielsetzung

Daher ist die Zielsetzung dieser Arbeit, eine geeignete Methode des maschinellen Lernens zu trainieren, die Medizinterme binär in die zwei Kategorien Laie und Experte zuordnen kann. Die zu erwartende Performance soll dabei mit objektiven Gütekriterien aus dem Bereich der Evaluierung von maschinellen Lernverfahren geschätzt werden. Mögliche Unterschiede zwischen der Verwendung eines vorab trainierten Sprachmodells und der Anwendung des maschinellen Lernverfahrens ohne Verwendung eines Sprachmodells sollen dabei herausgearbeitet werden. Ein für das Experiment notwendiger Goldstandard von manuell annotierten Termen wird hierzu aufgebaut und die Qualität des aufgebauten Goldstandards bewertet. Der

Goldstandard wird unter anderem dazu verwendet, eine Aussage zu treffen, ob es der Maschine möglich ist, die manuell durchgeführte Klassifizierungsqualität für diese Aufgabenstellung zu erreichen.

1.3. Gliederung

Im Kapitel "Einleitung" wird ein Überblick über die Entwicklung medizinischer Fachsprache und deren Herausforderungen für die maschinelle Verarbeitung gegeben. Des Weiteren werden grundsätzliche Arten des maschinellen Lernens in der Medizin unterschieden, gefolgt von einer Übersicht von Arbeiten, die sich mit einer ähnlichen Aufgabenstellung beschäftigt haben und so den Mehrwert der Arbeit hervorheben.

Das Kapitel "Material und Methoden" erläutert die Herangehensweise zur Auswahl des maschinellen Lernverfahrens und der damit verbundenen methodischen Konzepte. Verwendete Datensätze werden beschrieben und statistische Evaluierungsmaße, die zur Beurteilung der Klassifizierungsmethode notwendig sind, hervorgehoben.

Das Kapitel "Ergebnisse" beinhaltet die Evaluierung des verwendeten maschinellen Lernverfahrens mit statistischen Kennzahlen, einer visualisierten Dimensionsreduktionsmethode des Datensatzes und Abbildungen der generierten Vierfeldertafeln (Confusion Matrix).

Das Kapitel "Diskussion" betrachtet die generierten Ergebnisse und diskutiert diese im Rahmen der wissenschaftlichen Fragestellung und Zielsetzung. Eine selektive Fehleranalyse wird durchgeführt und Limitationen werden beschrieben. Der Abschluss der Arbeit bildet eine Zusammenfassung und einen Ausblick auf weiterführende Arbeiten.

1.4. Medizinische Fachsprache

1.4.1. Allgemeine Informationen

Die medizinische Terminologie nimmt im 5. Jahrhundert vor Christus ihren Ursprung. Platon beschreibt, dass bereits Hippokrates Medizin von der Philosophie, also der damaligen Wissenschaft, abgetrennt habe. Dementsprechend blickt die spezielle Terminologie auf ein bereits sehr hohes Alter zurück (vgl. Michler und Benedum, 1981, s. 8).

Ein typisches Charakteristikum der medizinischen Sprache ist die unverhältnismäßig hohe Anzahl an fachsprachlichen Ausdrücken, die unter anderem aus dem Lateinischen und Griechischen in viele neue Sprachen übernommen wurden (vgl. Michler und Benedum, 1981, s. 5).

Ein Vorteil des griechischen Sprachgebrauchs ist die Fähigkeit, beliebig viele Wörter zu langen Komposita zusammenzuführen. Beispielsweise bedürfte der Term “Pneumoperikard” in anderen Sprachen lange Absätze um den Sachverhalt darzustellen (vgl. Michler und Benedum, 1981, s. 4).

In der frühen griechischen Medizin mussten Ärzte sich aus dem allgemeinen Sprachschatz bedienen, um Teile des menschlichen Körpers zu beschreiben. Da aber Wörter wie “Kopf”, “Arm” oder “Bein” für genauere Beschreibungen ungeeignet erschienen, griffen die damaligen Ärzte auf Metaphern zurück. Als Beispiel dient hier “Tarsos”, damals für Geflecht, Reusen oder Korbgeflecht verwendet. Dieser Term wurde später zur Beschreibung des Fußskeletts entlehnt. Für das Korbflechten war die parallele Längsanordnung, welche regelmäßig von Quergeflechten unterbrochen wurde, ein Charakteristikum (siehe Abbildung 1.) (vgl. Michler und Benedum, 1981, s. 12).

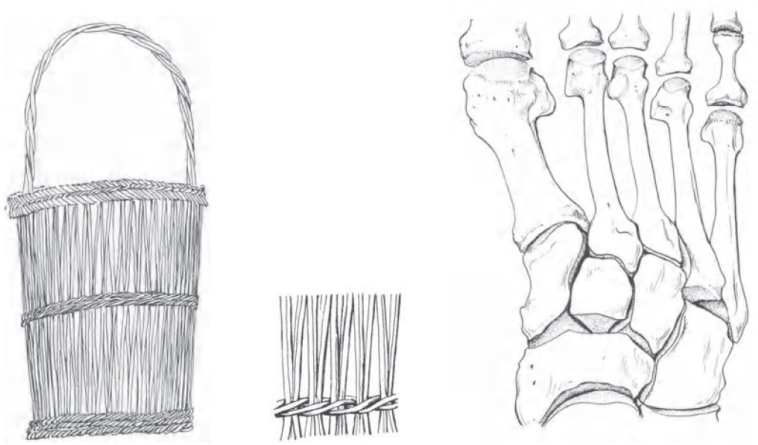

Abbildung 1. Korb in uralter Flechtart mit Ausschnitt von Quer- und Längsstreben im Vergleich zu Fußskelett (Michler und Benedum, 1981, s. 13)

Die Entstehung des Terms "Achillessehne" sei wie folgt: Als Kind wurde Achilles von seiner göttlichen Mutter Thetis im Wasser des Styx gebadet, um ihn unverwundbar zu machen. Die Ferse blieb jedoch vom Wasser unberührt und somit verletzlich. In der Schlacht um Troja starb Achill an einem Pfeilschuss des Paris in jene Stelle. In der Neuzeit wurde dem Anatom Philippe Verheyen (1648-1710) das Bein amputiert. Anschließend sezierte er es und gab der Sehne, die heute als Achillessehne bekannt ist, den Namen "Tendo Achillis". Auch in der heutigen Medizinsprache wird dieser Term noch verwendet (vgl. Michler und Benedum, 1981, s. 22). Kürze, Präzision, Einfachheit und Ausdruckskraft sind einige der vielen Vorteile des lateinischen Sprachgebrauchs für die Wissenschaft. Im Deutschen müssten wir das lateinische Wort "Divertikel" als "blind endigende Ausstülpung umschriebener Wandteile eines Hohlorgans" umschreiben (vgl. Michler und Benedum, 1981, s. 4). Ein weiterer Vorteil der lateinischen Sprache gegenüber modernen Sprachen ist die Tatsache, dass es sich um eine "tote" Sprache handelt. Ihr Nutzen besteht darin, dass Wörter und daraus zusammengesetzte Ausdrücke konstante Bedeutung bieten, sofern nicht neue Erkenntnisse in der Wissenschaft einen Bedeutungswechsel herbeiführen (vgl. Michler und Benedum, 1981, s. 22).

Das humoralpathologische Krankheitskonzept der antiken Medizin spiegelt sich in Ausdrücken wie "Schlagfluß" (Apoplex), "Herabfließen" (Katarrh) oder auch "Durchfluß" (Diarrhoe) wider. Dabei erfolgte die Interpretation der Krankheitsphänomene und die Erfassung und Behandlung von Krankheiten unter

anderen theoretischen Vorbedingungen als es heute in der westlichen Medizin der Fall ist. Um manche Ausdrücke verstehen zu können, reichen selbst sehr umfangreiche Latein- oder Griechischkenntnisse eventuell nicht aus (Beddies et al., 2008, s. 3–4).

1.4.2. Historischer Exkurs

1.4.2.1.Entwicklung der medizinischen Fachsprache bis zur Renaissance

Die Schriften von Galen (129 bis ca. 210) enthielten bereits einheitliche anatomische und physiologische Konzeptionen der Humoralpathologie mit systematischen Darstellungen des seinerzeit bekannten medizinischen Wissens. Auch als Römer schrieb er in griechischer Sprache, der damaligen Wissenschafts- und Fachsprache. Demgemäß erlangten Galens Werke über die Gebiete des Römischen Reiches hinaus an Bekanntheit. Viele Schriften wurden nach dem Fall des Römischen Reiches übersetzt. Die griechische Tradition wurde aber trotz dessen über das Oströmische Reich sowie im östlichen Mittelmeerraum weiterverfolgt (vgl. Beddies et al., 2008, s. 4–5).

So fanden seine Werke sowie dessen fachsprachlichen Wendungen seit dem 9. Jahrhundert direkt und indirekt auch im syrischen und persischen Raum Einzug, wo sie weiterentwickelt wurden. Als wichtigstes Beispiel dient hier die Zusammenstellung, Systematisierung, Bereicherung und Latinisierung der galenischen Schriften durch Ali al-Husain ibn Abd Allah ibn Sina, bekannt als Avicenna (980-1037). Dieses Werk bildete die wichtigste Grundlage der frühneuzeitlichen und mittelalterlichen Medizin. Da seit dem Mittelalter Latein als verbindliche Verkehrs- und Wissenschaftssprache (lingua franca) galt, sind in der heutigen medizinischen Fachsprache lateinische Elemente zu finden. Es wurden auf diesem Wege auch medizinische und naturwissenschaftliche Schriften aus der arabischen Welt im europäischen Raum bekannt (vgl. Beddies et al., 2008, s.4–5). Durch verschiedenste Übersetzungstraditionen im 15. sowie 16. Jahrhundert, in denen viele Schriften aus dem Griechischen ins Lateinische übersetzt wurden, entstanden so auch krude Wortbildungen. Ein Beispiel dafür wäre der Ausdruck "Pia

Mater", welcher wörtlich übersetzt "Fromme Mutter" bedeutet. In der medizinischen Fachsprache wird der Term zur Beschreibung der weichen Hirnhaut verwendet. Auch in der Renaissance blieb man der lateinischen Sprache treu. Die Werke des Andreas Vesalius, welche für die anatomische Fachsprache wichtig waren, als auch William Harvey's Werke über den Blutkreislauf wurden im klassischen Latein verfasst (vgl. Beddies et al., 2008, s. 5).

1.4.2.2. Entwicklung der medizinischen Fachsprache nach der Renaissance

Nach der französischen Revolution 1789 erreichte auch deren Heilkunde ihren Höhepunkt. Damals hat sich in den Pariser Krankenhäusern, durch Einrichtung von verschiedenen Gebäudekomplexen, welche einer Krankheit zugeordnet waren, bereits die eigentliche "Krankenhausmedizin" entwickelt. Dies prägte den medizinischen Sprachgebrauch bis heute. "Bandagieren", "Bougieren" und "Kürettage" sind Beispiele für den Einfluss des Französischen (vgl. Murken, 2009, s. 14–15). Doch bis weit in die Neuzeit blieb Latein weiterhin die Sprache der Gelehrten, so wurden an der Berliner Universität bis zur Mitte des 19. Jahrhunderts klinische Vorlesungen in Latein gehalten. Gegen Ende des 19. Jahrhunderts ging diese Kommunikationsbasis aufgrund des allmählichen Wechsels von der humanistischen zugunsten der mathematisch-naturwissenschaftlichen Schulbildung verloren (vgl. Beddies et al., 2008, s. 5). In Abbildung 2. werden Überlieferungswege der griechischen medizinischen Texte abgebildet.

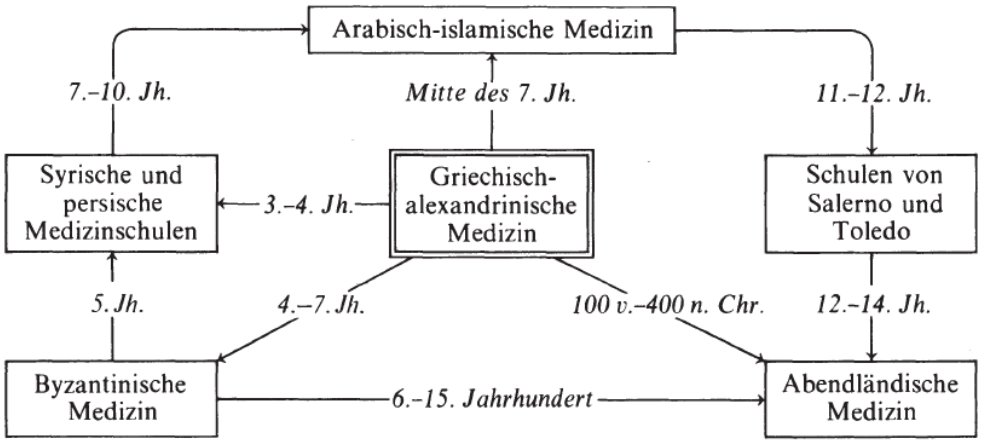

Abbildung 2. Diagramm von den Überlieferungswegen der griechischen medizinischen Texte (Michler und Benedum, 1981, s. 9)

Die Vereinigten Staaten blieben im Westen nach dem Zweiten Weltkrieg die einzige Großmacht, deren Wissenschafts- und Bildungsinfrastruktur verschont wurden. Dies förderte die Hinwendung zur englischen Sprache in den meisten Forschungsgebieten (vgl. Hamel, 2007; Popilka, 2014, s. 8–9; Roelcke, 1999).

Mit Ende des Zweiten Weltkrieges, der Gründung der Vereinten Nationen, der Erfindung des Computers sowie auch den Entwicklungen im wissenschaftlichen und technischen Bereich setzte sich Englisch als vorherrschende Sprache endgültig durch (vgl. Hamel, 2007; Popilka, 2014, s. 9). Aufgrund der Entwicklung der englischen Sprache zur Sprache von Wissenschaft und Technik sind auch Publikationen auf Deutsch und Französisch international gesehen immer seltener (vgl. Hamel, 2007; vgl. Popilka, 2014, s. 9; Roelcke, 1999). ÄrztInnen in englischsprachigen Ländern erlangten außerdem durch innovative Operationstechniken, die Entwicklung biomedizinischer Geräte und weitere Innovationen immer mehr an Beachtung. So haben sich Wörter wie "Bypass", "Pacemaker", "Scanner" auch international durchgesetzt (vgl. Murken, 2009; Popilka, 2014, s. 9).

Während die jeweilige Landessprache im Gespräch mit PatientInnen, KollegInnen vor Ort, in der klinischen Dokumentation und weitgehend in der medizinischen Lehre verwendet wird, kann die Kommunikation in der medizinischen Wissenschaft folgendermaßen charakterisiert werden: Ein in englischer Sprache gehaltener Kern mit den weltweit bedeutendsten Fachzeitschriften, Debatten und wichtigen Forschungsfragen, welcher von einem Saum umgeben wird, welcher in der jeweiligen Landessprache gehalten wird (vgl. Baethge, 2008).

Ein Problem bei englischen Entlehnungen stellt die Mehrdeutigkeit dar. "Bypass" könnte Überbrückungsgefäß aber auch Überbrückungsoperation bedeuten. "Compliance" könnte man zur Beschreibung der Dehnbarkeit der Lunge verwenden. Es kann aber auch die Bereitschaft der zu behandelnden Person zur Mitarbeit bezüglich medizinischer Maßnahmen beschrieben werden. Mischkompositionen können gezwungen wirken, wie etwa im Beispiel "Kammerstiffness"(vgl. Karenberg, 2014, s. 24), ebenso wie die Anwendung morphologischer Regeln der Zielsprache wie in "Bypässe".

1.4.3. Eigenheiten der medizinischen Fachsprache

1.4.3.1. Eponyme

Die Benennung von Gegenständen und Sachverhalten nach anerkannten Persönlichkeiten hat auch in der Medizin einen großen Stellenwert erlangt. Symptome, Krankheiten, Reaktionen, Behandlungsverfahren und vieles mehr werden häufig nach der Person der Erstentdeckung benannt (siehe Tabelle 1.) (vgl. Karenberg, 2014, s. 19–20).

Eponym	**Zuordnung**	**Geschichte**
Alzheimer-Krankheit	Erkrankung des Gehirns	Erstbeschreibung 1906 durch Alois **Alzheimer** 1906
Apgar-Index	Skala zur Beurteilung von Neugeborenen	Entwickelt 1953 von Virginia **Apgar**
Billroth-I-Operation	Operation am Magen-Darm-Kanal	Erstmals durchgeführt 1881 von Theodor **Billroth**
Eustachische Röhre	Ohrtrompete	1550 von Bartolomeo **Eustachio** erstmals beschrieben
Rickettsia prowazekii	Erreger des Fleckfiebers	1900 von Howard **Ricketts** und **Prowazek**

Tabelle 1. Beispiele zu Eponymen (in Anlehnung an Karenberg, 2014, s. 20)

Vorteile dieser sogenannten Eponyme sind die Prägnanz und Präzision des Ausdrucks. Ein Ersatzausdruck würde langwierig und kompliziert werden. Der Nachteil eines Eponyms ist die fehlende Verständlichkeit, und oft werden Eponyme auch zusätzlich international variierend benutzt (vgl. Karenberg, 2014, s. 20). Deswegen bestehen Tendenzen, auf Eponyme zu verzichten, zum Beispiel in den Guidelines der SNOMED-CT-Terminologie (Høy und Howarth, 2012), wenngleich diese durch die weiterhin enorme Popularität bestimmter Eponyme nur teilweise umgesetzt werden.

1.4.3.2. Akronyme

Werden die Anfangsbuchstaben mehrerer Wörter eines Fachausdrucks zu einer Abkürzung zusammengefügt, bildet sich ein neues künstlich erzeugtes Wort, ein Akronym. Häufig entstehen international übliche medizinische Akronyme aus ursprünglich englischsprachigen Termen (siehe Tabelle 2.) (vgl. Karenberg, 2014, s. 20).

Akronym	**Auflösung**	**Definition**
AIDS	Acquired Immune Deficiency Syndrome	erworbenes Immunschwäche Syndrom
SIDS	Sudden Infant Death Syndrome	Syndrom des plötzlichen Kindstodes

Tabelle 2. Beispiele für Akronyme (Karenberg, 2014, s. 20)

Ein Nachteil der Akronyme ist auch hier die variierende internationale Verwendung. Der französische Sprachgebrauch verwendet anstatt "AIDS" das Akronym "SIDA" (Syndrome Immuno-Déficitaire Acquis). Die Knappheit und Genauigkeit sind hingegen ein Vorteil in der Beschreibung (vgl. Karenberg, 2014, s. 20).

Das große Problem der Mehrdeutigkeit von Akronymen illustriert folgendes verheerendes Beispiel: Der amerikanische Kardiologe und Friedensnobelpreisträger Bernhard Lown hatte im Rahmen einer Visite mit seinem Chefarzt eine Patientin besucht. Der Chefarzt teilte den Kollegen mit, dass es sich hier um eine "TS" handeln könnte. Er meinte damit eine Trikuspidalstenose, die Patientin aber hatte fälschlicherweise die Abkürzung als terminale Situation interpretiert. Lown hatte die Patientin aufgeklärt und versuchte sie zu beruhigen, ihr Zustand verschlechterte sich aber dennoch weiter. Als der Chefarzt die Patientin später besuchen wollte, war sie bereits gestorben. Die Patientin hatte in ihrer letzten Lebensspanne große Angst und war aufgeregt (vgl. Hüllemann, 2013, s. 28–29).

1.4.3.3. Synonyme

In allen Sprachen finden sich Wörter oder Ausdrücke mit nahezu gleicher Bedeutung. Mumps, Ziegenpeter oder auch Bauernwötzel sind je nach Gebiet umgangssprachliche Ausdrücke für eine virusbedingte Entzündung der

Ohrspeicheldrüse. Besonders im medizinischen Sprachgebrauch gibt es für viele Erkrankungen mehrere Bezeichnungen (siehe Tabelle 3.) (vgl. Karenberg, 2014, s. 21).

Synonym	**Inhaltlicher Schwerpunkt der Begriffsbildung**
Pfeiffer-Drüsenfieber	Eigenname – erkrankte Körperteile – Leitsymptom
Mononucleosis infectiosa	befallene Körperzellen – Art der Erkrankung (lateinisch)
Infektiöse Mononukleose	Art der Erkrankung – befallene Körperzellen (eingedeutscht)
Monozyten-Angina	befallene Körperzellen – Leitsymptom
Lymphoidzell-Angina	befallene Körperzellen – Leitsymptom (veraltet)
Knutsch-Krankheit	häufiger Übertragungsmodus (Slang)
kissing disease	häufiger Übertragungsmodus (amerikanische Form, Slang)
Teenager-Fieber	bevorzugt betroffene Altersgruppe (Slang)

Tabelle 3. Beispielabbildung: Synonyme für infektiöse Mononukleose (Karenberg, 2014, s. 21).

Obige Tabelle gibt verschiedene bedeutungsgleiche Ausdrücke für dieselbe Virusinfektion wieder. Die Darstellung erlaubt, die konkurrierenden Möglichkeiten in der Benennung zu erkennen. Durch die vielen Alternativen ist es also möglich zwischen verschiedenen Termen auszuwählen. Die daraus resultierende Unübersichtlichkeit und die fehlende globale Verbindlichkeit sind gerade im Zeitalter der elektronischen Datenverarbeitung ein großer Nachteil (vgl. Karenberg, 2014, s. 21).

1.4.3.4. Antonyme

Sprachliche Ausdrücke mit entgegengesetzter Bedeutung werden als Antonyme bezeichnet. Sie spielen vor allem in der Orientierung am menschlichen Körper eine bedeutende Rolle. Auch in der Klinik finden sich solche Oppositionswörter wieder. In der Tumorbeschreibung zum Beispiel "benigne", was gutartig bedeutet, in Opposition zu "maligne", was bösartig bedeutet (vgl. Karenberg, 2014, s. 21).

1.4.3.5. Metonyme

Metaphorische Begriffsbildungen und Bildübertragungen nennt man Metonyme. Der Vorteil liegt darin Namen für beispielsweise anatomische Strukturen durch Formanalogien zu veranschaulichen. Ohr"muschel" und "Rabenschnabel"fortsatz sollen als Beispiele dienen (vgl. Karenberg, 2014, s. 21).

1.4.3.6. Hybridbildungen

Hybridbildungen durch Vermischung von Wortbestandteilen unterschiedlicher Sprachen sind möglich. Als Beispiel dient hier der Ausdruck “Hämoglobin”, er besteht aus dem griechischen Wortbestandteil "haima" (Blut) als auch dem lateinischen Wortbestandteil "globus" (Kugel) (vgl. Michler und Benedum, 1981, s. 5).

1.4.3.7. Anglizismen und Amerikanismen

In den letzten Jahrzehnten entstanden sprachliche Neubildungen aus dem britischen und amerikanischen Wortschatz. Als Beispiele dienen: “Pacemaker” für Herzschrittmacher oder “Rooming in” für die Unterbringung der Mutter und ihres Neugeborenen (vgl. Karenberg, 2014, s. 24).

1.4.4. Unterschiede und Schwierigkeiten der medizinischen Fachsprache und der Laiensprache

Die medizinische Fachsprache hebt sich mit ihrem spezialisierten und technischen Wortschatz von der laienverständlichen allgemeinen Sprache ab. Darin spiegelt sich auch das umfangreiche Wissen auf diesem Gebiet wider. Eine daraus resultierende Schwierigkeit ist, dass das Vokabular in medizinischen Dokumenten für Laien, die keine Spezialisten auf diesem Gebiet sind, oft unverständlich beziehungsweise undurchsichtig bleibt (vgl. Grabar, 2019, s. 1).

Befundberichte oder andere medizinische Texte enthalten Fachsprache, Akronyme und Terme, die das medizinische Personal kennt und versteht. Bevor die

betreffenden Personen mit medizinischen Ausdrücken vertraut werden, gibt es häufig eine flache Lernkurve. Damit Laien schneller und besser verstehen, benötigt es verständlichere Wörter und Phrasen, um medizinische Bedeutungen darzustellen. Inhalte von Terminologiedatenbanken, Thesauri und Lexika, die für medizinisches Personal erstellt wurden, sind oft schwer in laienverständliche Sprache zu übersetzen (vgl. Åhlfeldt et al., 2006).

Auch gestaltet sich die Kommunikation zwischen Experten und Laien umso komplexer, wenn es um die eigene Gesundheit geht. Der innere Bewertungsprozess des Laien in der Kommunikation mit dem Experten hat den Nachteil, dass emotionale Bewertungen ungenau sind und zu falschen Assoziationen sowie Handlungen führen können (vgl. Hüllemann, 2013, s. 13,19).

Im Deutschen gibt es beispielsweise auch keine Abgrenzung von "krank ‘sein’" zu "Krank’heit’". Die betreffende Person wird sich bei Gebrauch des Wortes Krankheit auch immer krank fühlen. Im Englischen beispielsweise gibt es diese Abgrenzung. Krank"sein" wird mit "illness" übersetzt, Krank"heit" mit "disease" (vgl. Hüllemann, 2013, s. 24).

In einer psycho-onkologischen Studie wurde das Verständnis von TeilnehmerInnen in Bezug auf das Verständnis von allgemeinen Ausdrücken verbunden mit Diagnosen, Prognosen und der Behandlung geprüft. Der Einfluss von medizinischem Fachjargon im Vergleich zum allgemeinen Sprachgebrauch auf die wahrgenommene Wirksamkeit in Interaktionen mit dem medizinischen Fachpersonal, als auch die Teilnahme an medizinischen Entscheidungsprozessen und dem zwischenmenschlichen Vertrauen wurden untersucht. Das medizinische Fachpersonal kommunizierte einmal im medizinischen Fachjargon und einmal in laienverständlicher Sprache. Im Durchschnitt konnte die Mehrheit krebsbezogene Terme verstehen, jedoch waren nur 2,2% im Stande, alle Terme korrekt zu verstehen. Korrektes Verständnis geht im Gegenzug zu falschem Verständnis im Allgemeinen mit höherem Vertrauen in das Verständnis einher. Die Steigerung der Komplexität der Sprache hatte keinen signifikanten Einfluss auf Messungen der wahrgenommenen Wirksamkeit oder des Vertrauens. Die Ergebnisse der Studie deuten darauf hin, dass das Verständnis der Laien der in der Onkologie

gebräuchlichen Fachsprache suboptimal ist. Das Vertrauen in das Verständnis der TeilnehmerInnen hing vor allem mit der wahrgenommenen Wirksamkeit bei der Beteiligung der Konsultation zusammen. Das teilnehmende medizinische Fachpersonal sollte versuchen, Missverständnisse zu korrigieren, um unnötige Sorgen zu verringern und damit möglicherweise die Teilnahme zu erleichtern (vgl. Pieterse et al., 2013).

In einer weiteren Studie wurden verschiedene medizinische Fachtermini evaluiert. Es konnte festgestellt werden, dass die Verwendung medizinischer Fachtermini in Verbindung mit der Kommunikation zwischen behandelnder und behandelter Person einer weiteren Klärung bedarf (vgl. Conti, 2013).

1.5. Maschinelles Lernen in der Medizin

An der Schnittstelle zwischen Statistik, die aus Daten Zusammenhänge ableitet, und der Informatik, mit Schwerpunkt auf Rechenalgorithmen, hat sich maschinelles Lernen (ML) als wissenschaftliche Disziplin etabliert, die das Lernen aus Daten durch Computersysteme zum Gegenstand von Forschung und Entwicklung hat. Durch die rechnerischen Herausforderungen beim Aufbau statistischer Modelle mit außerordentlich großen Datensätzen zeigt sich der Nutzen von maschinellem Lernen. Computer können immer komplexere Lernaufgaben erfüllen. Das liegt an den Fortschritten von Rechenleistung und Speicher, ebenso wie der Verfügbarkeit einer immensen Datenfülle. Beliebte Pokerspiele oder Gesetze der Physik aus experimentellen Daten lernen, sollen hier nur als Beispiele für Aufgaben dienen, welche Computer heutzutage berechnen können. Diese galten vor nicht allzu langer Zeit noch als unmöglich zu lösen. Die Zahl der Unternehmen, die in diesem Bereich agieren, wächst stetig, darunter auch Firmen, die mit Methoden des ML medizinische Daten analysieren und sich so Problemen des Gesundheitswesens widmen (vgl. Deo, 2015).

1.5.1. Arten des maschinellen Lernens

Prinzipiell kann man maschinelles Lernen in Supervised Learning (überwachtes Lernen), und in Unsupervised Learning (unüberwachtes Lernen) unterteilen (vgl. Deo, 2015; Handelman et al., 2018).
Ersteres befasst sich aus technischer Sicht mit Klassifizierungs- und Regressionsproblemen. Letzteres kann man vor allem zur Clusterbildung und Reduktion der Dimensionalität anwenden. Erhaltene Muster durch unüberwachtes Lernen werden in der Regel von Menschen oder durch Anwendung auf deren Nutzen evaluiert (vgl. Handelman et al., 2018).

1.5.1.1. Supervised Learning

Beim Supervised Learning wird der Rechner mit Funktionen ausgestattet, die sich auf ein Ziel und die gewünschten zu erreichenden Zielvorgaben beziehen, um Verbindungen zwischen diesen beiden im Datensatz zu identifizieren. Beispielsweise lässt sich ein Modell zur Unterscheidung von Äpfeln, Orangen und Zitronen etablieren, indem ein Algorithmus zunächst auf Farbe, Größe, Gewicht und Form trainiert wird und dabei die verschiedenen Früchte zu unterscheiden lernt. Wird dann eine neue "nicht etikettierte" Frucht präsentiert, sollte das Modell in der Lage sein, die Frucht der richtigen Art zuzuweisen (vgl. Handelman et al., 2018). Anwenden kann man dies zum Beispiel auf die automatische EKG-Interpretation, indem die Erkennung des Musters mittels eines vorannotierten Diagnosedatensatzes trainiert wird. In der Bildgebung kann ein vergleichbarer Ansatz die automatische Detektion von Lungenknoten in einem Thorax-Röntgenbild unterstützen (vgl. Deo, 2015). Auch bei der umgekehrten Bildsuche wurde oben genannte Methode verwendet; man verwendete Bilder von Präparaten und nutzte eine Suchmaschine, um ähnliche Bilder zu finden (vgl. Mamrosh und Moore, 2015). Häufige Anwendungen von überwachtem Lernen umfassen die Risikoeinschätzung. Das Modell kann hier mehr als die bloße Annäherung an die Fähigkeiten eines/r Arztes/Ärztin leisten. Es kann neue Zusammenhänge finden, die für die jeweilige Person nicht ohne weiteres ersichtlich wären (vgl. Deo, 2015). Ein Beispiel dafür ist der Framingham-Risiko-Score, ein geschlechtsspezifischer Algorithmus, der eine

Einschätzung des Risikos zur Entwicklung einer koronaren Herzerkrankung in den nächsten 10 Jahren voraussagt (vgl. Jahangiry et al., 2017).

Bei einer Studie zur antithrombotischen Therapie bei Vorhofflimmern kommen Modelle des Supervised Learning zum Einsatz (vgl. Lip et al., 2010). Ein weiteres Beispiel für die Anwendung des Supervised Learning ist bei der Implantation von automatisierten implantierbaren Defibrillatoren bei hypertropher Kardiomyopathie (vgl. O'Mahony et al., 2014).

1.5.1.2. Unsupervised Learning

Anders als beim überwachten Lernen werden beim unüberwachten Lernen keine vorhergesagten Ergebnisse ausgegeben. Es wird versucht, latente Muster oder Gruppierungen innerhalb von Daten herauszubekommen (vgl. Deo, 2015). Dabei kommen oft Fragen als auch Antworten auf, die primär nicht von den Entwicklern des Modells als solche erdacht wurden (vgl. Handelman et al., 2018). Anhand der Leistung in nachfolgenden beaufsichtigten Lernaufgaben kann die Nützlichkeit der Gruppen oder erstellten Muster bewertet werden (vgl. Deo, 2015).

Eine Möglichkeit für unüberwachtes Lernen in der Präzisionsmedizin besteht in der Erstellung eines Wissensnetzwerks für Krankheiten, welches auf der Definition von Krankheitsveranlagungen und pathogenen Prozessen auf molekularer Ebene verschiedener Individuen basiert. Daraus wird eine genaue, molekular basierte Taxonomie von Krankheiten vorgeschlagen. Umgesetzt wurde dieser Ansatz in der Identifizierung eines eosinophilen Subtyps von Asthma (vgl. Woodruff et al., 2009).

1.6. Verwandte Arbeiten

In einer 2019 veröffentlichten Studie wurde versucht, ein verständliches Vokabular für Laien zu erstellen, um die Lücke zwischen Experten- und Laienvokabular zu schließen. Dazu wurde eine Methode entwickelt, nach der man Terme identifizieren und auch neue Terme hinzufügen kann, um mit dem ständig erweiternden medizinischen Fachwissen sprachlich Schritt zu halten. Hierbei kam ein Vektorraum-Modell zur Anwendung. Zuerst wurden Terme über einen großen

Textkorpus erlernt, danach wurde eine "überwachte" Methode mit bereits existierendem Laienvokabular adaptiert, um entsprechendes Feature-Engineering zu betreiben. Medizinische Fachterme sowie deren Laien-Äquivalente einschließlich Varianten wurden über deren semantische Entfernung im Vektorraum identifiziert. Die Ergebnisse wurden mittels "Mean reciprocal rank" (MRR) ermittelt. Die Ergebnisse zeigten, dass ein Fine-Tuning des Vektorraum-Modells ein besseres Ergebnis als die nicht-getunte Variante lieferte. Das Modell konnte außerdem auch häufige Abkürzungen und Tippfehler erfassen. Da eine große Menge an Textinformationen und bereits bestehendem Laienvokabular integriert wurde, hat deren Vektormodell laut Studie mehrere Basis-Ranking-Methoden in der Leistung übertroffen (vgl. Gu et al., 2019).

In einer weiteren Studie wurde aufgrund von Diskrepanzen zwischen der Sprache der Laien und der Fachsprache im semantischen und lexikalischen Bereich ein laienverständliches Vokabular erarbeitet. Es wurden Beziehungen zwischen allgemeinen Gesundheitsausdrücken von Laien und professionellem Personal untersucht und dann ein laienverständliches Vokabular erstellt. Dabei wurden auch die Grenzen, beispielsweise die Nichtberücksichtigung psychosozialer und kultureller Faktoren beschrieben und vorgeschlagen, weiter explorativ zu forschen und zu entwickeln, um die Gesundheitsinformationen für Laien besser zu gestalten (Zeng und Tse, 2006).

Vydiswaran et al. (2014) haben Fachterme und deren alternative Varianten mittels musterbasierten Text-Mining aus Medline-Abstracts und allen Beiträgen des Online-Gesundheitsforums MedHelp identifiziert und verknüpft. Die Ergebnisse dazu zeigten, dass der Ansatz im Stande war, bedeutungsgleiche Paare mit hoher Genauigkeit zu identifizieren und diese auch als Laien- beziehungsweise Expertenterm zuzuordnen.

Das Ziel einer weiteren Studie war es, Terme nach Gebrauch durch Laien und Gesundheitspersonal zu identifizieren und charakterisieren. Außerdem wurde auch geprüft, ob und wieweit die Terme mit dem medizinischen Fachvokabular übereinstimmen. Es wurden Laien damit beauftragt, circa 100.000 Terme aus

Beiträgen von Online-Diskussionsforen und Artikeln aus Printmedien zu identifizieren. Mehr als 75 Prozent konnten als echte oder nahe Synonyme dem Metathesaurus des Unified Medical Language System (UMLS), einer großen Sammlung von Medizintermen aus über 100 kontrollierten Vokabularien, zugeordnet werden. Überschneidungen der Terme von oben genannten Personengruppen konnten in 80 Prozent gefunden werden, während dabei aber die Hälfte dieser Terme Laienterme enthielten, die sich von den Fachtermen unterschieden. Es soll also auch die theoretische und praktische Implikation für die Überwindung der Kommunikationsbarriere zwischen den beiden Personengruppen nicht außer Acht gelassen werden (vgl. Zeng und Tse, 2006).

2. Material und Methoden

Aufgrund der geringen Datenlage zur oben genannten Thematik im deutschsprachigen Raum wurde eine Pilotstudie als notwendig erachtet. Da es sich hier um keine geschützten Daten von Personen handelt, war kein positives Ethikvotum erforderlich.

2.1. Erarbeitung der Annotationsrichtlinien

Als Grundlage diente eine Zufallsauswahl von 196 Termen aus MEDLINE-Datensätzen mit deutschen Titeln sowie der deutschen Wikipedia. Diese wurde von jeweils drei medizinischen ExpertInnen und drei medizinischen Laien in "E" und "L" unterteilt. "E" steht für Expertenterm und "L" für Laienterm. Übereinstimmungen und Diskrepanzen zwischen den Expertenurteilen waren Grundlagen für die Erarbeitung von Annotationsrichtlinien. Von guten Annotationsrichtlinien wird erwartet, dass ihre Befolgung zu einer hohen Übereinstimmung zwischen BewerterInnen führt.

2.1.1. Beispiele zur Erarbeitung der Annotationsrichtlinien

"Anaphylaxie" wurde beispielsweise von allen sechs teilnehmenden Personen als Expertenterm gekennzeichnet. Auch "Sulfonylharnstoffe" lieferte ein eindeutiges Ergebnis mit der sechsmaligen Kennzeichnung als Expertenterm.
Der Term "Hepatitis B-Impfung" wurde zweimal als Expertenterm sowie viermal als Laienterm bezeichnet. "Coolpack" lieferte auch uneindeutige Ergebnisse, er wurde dreimal als Expertenterm und dreimal als Laienterm gekennzeichnet. Der Term "Arzneimittel" wurde sechsmal als Laienterm gekennzeichnet, ebenso der Term "Harnblase".
Die oben genannte Liste wurde im Team (bestehend aus zwei medizinischen Experten und einem medizinischen Laien) mehrmals bearbeitet. Aus den dadurch gewonnenen Erfahrungen wurden die folgenden Annotationsrichtlinien formuliert.

2.1.2. Annotationsrichtlinien

Richtlinie 1. Wird ein Term aus mehreren Wörtern oder Teilwörtern zusammengesetzt und ist mindestens eines davon nur Experten verständlich, so ist der gesamte Ausdruck als Expertenterm zu klassifizieren. Als Beispiel dient hier der Expertenterm "Magenerosionen". Das Wort "Magen" ist dem Laien verständlich, das Wort "Erosionen" jedoch nicht, woraus folgt, dass der aus Laien- und Expertenwortteilen zusammengesetzte Term als Expertenterm zu klassifizieren ist.

Richtlinie 2. Ein Term, der laienverständlich ist, aber auch von Experten häufig verwendet wird, ist dennoch als Laienterm zu klassifizieren. Dies lässt sich gut an dem laienverständlichen Wort "Blut" zeigen, da dieses von Experten ebenso verwendet wird (während das lateinische Wort "Sanguis" ungebräuchlich ist).

Richtlinie 3. Wird ein Ausdruck aus zwei laienverständlichen Teilen zusammengesetzt und ergibt das Wort sinngemäß eine nur dem Experten verständliche Bedeutung, so ist das Wort als Expertenterm zu klassifizieren. Als Beispiel dient hier das Wort "knochenmarksinfiltrierend". Das Wort "Knochenmark" und der Teil "infiltrierend" sind beide dem Laien verständlich, jedoch zusammengesetzt bildet das Wort eine nur Experten begreifliche Bedeutung.

Richtlinie 4. Ein Wort, das einem Laien nicht geläufig ist, aber dessen Bedeutung leicht aus dem Kontext zu erschließen ist, gilt als Laienterm. Zum Beispiel das Wort "Biokohle", denn "Biokohle" setzt man erfolgreich zur Therapie von Durchfallserkrankungen ein. Hier erschließt sich dem Laien die Bedeutung des Wortes durch den Kontext.

Richtlinie 5. Kann ein Wort sinngemäß durch einen eindeutigen Laienterm ersetzt werden, gilt es als Expertenterm. Als Beispiel wird das Wort "Thrombose" angeführt, hier wäre ein laienverständlicher Term das Wort "Blutgerinnsel".

Richtlinie 6. Kann man einem Expertenterm keinen laienverständlichen Term zuordnen, gilt er als Expertenterm. Als Beispiel dient hier das Wort "Lymphknotenreaktion".

2.2. Erstellung des Trainings- und Testdatensatzes

Nachdem die Annotationsrichtlinien erstellt wurden, war es notwendig, zur weiteren Bearbeitung einen Trainingsdatensatz als auch einen Testdatensatz zu generieren. Es wurden aus drei unterschiedlichen Quellen Stichproben gezogen. Um eine Unterteilung in einen Trainings- sowie Testdatensatz zu gewährleisten wurde der gesamte Datensatz in 80 % Testdaten sowie 20 % Trainingsdaten unterteilt. Mittels der folgenden genannten Methoden wurden die Stichproben selektiert und bereinigt.

2.2.1. Datensatz-Generierung

2.2.1.1. Erster Datensatz

Es wurde eine Stichprobe aus deutschen Titeln in der Literaturdatenbank MEDLINE gezogen. Mittels PubMed wurden alle Datensätze mit Dokumentensprache "deutsch" heruntergeladen und die Inhalte des Feldes "TT" (Transliterated Title) herausgefiltert. Danach wurden Wort-N-Gramme ($1 \leq N \leq 4$), also Einzelwörter bis Sequenzen aus vier Wörtern) generiert.

2.2.1.2. Zweiter Datensatz

Innerhalb des Projekts GAP[1] ("Gut informierte Kommunikation zwischen Arzt und Patient") war vom Institut für Medizinische Biometrie und Statistik der Universität Freiburg, Deutschland, ein Korpus aus öffentlich verfügbaren Webseiten (Schwerpunkt auf medizinische Informationen für Laien) erstellt worden. Aus diesem Korpus wurden ebenfalls Wort-N-Gramme ($1 \leq N \leq 4$) generiert.

2.2.1.3. Dritter Datensatz

Aus der deutschen Wikipedia wurde ein dritter Datensatz generiert. Auch hier wurden Wort-N-Gramme ($1 \leq N \leq 4$) erhoben.
Pro Datensatz wurden Zufallsstichproben von 3000 N-Grammen ausgewählt.

[1] https://innovationsfonds.g-ba.de/projekte/neue-versorgungsformen/gap-gut-informierte-kommunikation-zwischen-arzt-und-patient.114

2.2.2. Selektion und Bereinigung der Datensätze

2.2.2.1. Prüfung der Terme auf Relevanz

Beispiel: Wörter wie "Klimafaktoren", "Vollpappe" und "Zeuge", die eindeutig im Medizinkontext irrelevant sind, wurden ignoriert. Ebenfalls ignoriert wurden englische Terme, die im Deutschen nicht verwendet werden.

2.2.2.2. Bereinigen der Terme

Es wurden Umlaute nachgetragen, sofern diese in MEDLINE fehlten. Außerdem wurde der Term in die übliche lexikalische Zitierform gebracht, nämlich Nominativ Singular bei Substantiven, Infinitiv bei Verben. Auch Endungen von Adjektiven wurden in die Normalform gebracht.

2.2.3. Beschreibung des Datensatzes

Nach der Bereinigung und Selektion enthält der Datensatz 1697 Laienterme und 2229 Expertenterme. Der Datensatz wurde in weiterer Folge in einen Trainings- (1237 Laienterme, 1904 Expertenterme) und Testdatensatz (460 Laienterme, 325 Expertenterme) unterteilt, im Verhältnis 80:20.

Dieser Datensatz wurde von zwei Experten anhand der Annotationsrichtlinien annotiert bei einer resultierenden Interrater-Übereinstimmung von κ = 0,71 (Cohens Kappa siehe Abschnitt 2.7.7).

2.3. Worteinbettungen - Vector-Space-Modell

Eine zunehmende Rolle in der Verarbeitung der natürlichen Sprache (Natural Language Processing, NLP) sind Wortvektordarstellungen (vgl. Gurunath Shivakumar und Georgiou, 2019). Solche Vektorraum-Modelle können Wörter als kontinuierlich bewertete Vektoren darstellen und messen Ähnlichkeit basierend auf deren Abstand oder den Winkeln zwischen diesen Vektoren. Sie bieten eine

effiziente Schätzung aus sehr großen Datenmengen (vgl. Gurunath Shivakumar und Georgiou, 2019). Ein Text wird in einem Vektorraum abgebildet, welcher anschließend für numerische Berechnungen und Verarbeitungsaufgaben verwendet wird (vgl. Zhu et al., 2018). Eine Schwierigkeit liegt darin, menschliche Ähnlichkeitsurteile und deren vielen Eigenschaften mit den geometrischen Einschränkungen, die eine räumliche Darstellung erfordert, zu verknüpfen (vgl. Gurunath Shivakumar und Georgiou, 2019).

2.4. t-Distributed Stochastic Neighbor Embedding

t-Distributed Stochastic Neighbor Embedding (t-SNE) ist ein statistisches Verfahren zur Visualisierung hochdimensionaler Daten, bei dem jeder Datenpunkt über eine nichtlineare Dimensionsreduktionstechnik in einen niedrig dimensionalen (zwei- oder dreidimensionalen) Raum projiziert wird. Jedes hochdimensionale Objekt wird dabei im dimensionsreduzierten Raum so platziert, dass ähnliche Objekte durch nahe Punkte in ihrer Umgebung über eine Wahrscheinlichkeitsverteilung modelliert werden. Die daraus entstehende Punktverteilung gibt Aufschluss über Objektgruppierung mit ähnlichen Eigenschaften (vgl. van der Maaten und Hinton, 2008).

2.5. Feature-Engineering

Mit Hilfe von Fachwissen werden beim Feature-Engineering bestimmte Merkmale aus Rohdaten isoliert. Diese isolierten Merkmale können verwendet werden, um das Ergebnis von Algorithmen für maschinelles Lernen zu verbessern (vgl. "Feature engineering, " 2021). Für die Fragestellung in dieser Arbeit wurden Wortbestandteile untersucht (Prä- und Suffixe), die einen Hinweis auf einen Experten- oder Laienterm geben. Eine selektive Auswahl der erkannten Muster, die einen Hinweis für eine Klassenzugehörigkeit geben, ist in der Tabelle 4. dargestellt.

Wortteil	Beispiel	Bewertung
-nese	Genese	Experte
-karzinom	Plattenepithelkarzinom	Experte
-oid	Steroid	Experte
-itis	Hepatitis	Experte
-broncho	bronchopulmonal	Experte
-alter	Säuglingsalter	Laie

Tabelle 4. Wortbestandteile, die auf eine Klassenzugehörigkeit schließen lassen

2.6. Auswahl des maschinellen Lernverfahrens

Aufgrund der Analyse aus den vorhergehenden Abschnitten wird ein Verfahren gesucht, das Wortbestandteile (Character-N-Gramme) in der Modellgenerierung im Rahmen einer Parametrierung berücksichtigt. Die Zerlegung soll in Summe zu einem Beitrag für eine repräsentative Semantik eines Embedding-Vektors des zu klassifizierenden Eintrags führen. Das Verfahren soll gegenüber Out of Vocabulary Vorkommnissen robust sein. Auch sollen vortrainierte Sprachmodelle und deren Einfluss in die Klassifizierungsaufgaben miteinbezogen werden können. Die Modellierung soll auch auf CPUs ohne die Verwendung von Graphikprozessoren (GPUs) in einem vernünftigen Rahmen durchführbar sein (vgl. Bojanowski et al., 2017; Joulin et al., 2016).

Nach dieser Anforderungsanalyse an die Algorithmik ergab sich fastText als Textklassifizierungsengine der Wahl, welche alle Punkte der oben angeführten Bedarfserhebung an das maschinelle Lernverfahren für die Textklassifikation erfüllt (vgl. Bojanowski et al., 2017; Joulin et al., 2016). FastText besteht aus einer neuronalen Netzwerkarchitektur, die entweder bestehende, auf fastText basierende Sprachmodelle auf die Klassifikationsaufgabe feinjustiert, beziehungsweise Wortrepräsentationen als Embeddingvektoren im Rahmen des Trainings neuer Sprachmodelle aufbaut.

Für die Durchführung des Experiments wurde ein vortrainiertes deutsches Sprachmodell verwendet (vgl. Grave et al., 2018), um den Einfluss auf die Klassifikationsqualität, im Gegensatz zur Verwendung des maschinellen Lernverfahrens ohne Verwendung eines vortrainierten Modells, zu beurteilen.

2.7. Maße zur Beurteilung der Ergebnisse

2.7.1. Confusion Matrix

Die Basis für die weitere Einteilung der Ergebnisse stellt die Confusion Matrix dar. Es handelt sich hierbei in ihrer einfachsten Form um eine Vierfeldertafel, die zwischen "richtig" und "falsch" klassifiziert (vgl. Manning et al., 2008, s. 307–308). In folgender Tabelle werden die Bezeichnungen "True positive", "False positive", "True negative" und "False negative" eingeführt.

True positive	Das Objekt wurde abgerufen und ist relevant.
False positive	Das Objekt wurde abgerufen und ist nicht relevant.
True negative	Das Objekt wurde nicht abgerufen und ist nicht relevant.
False negative	Das Objekt wurde nicht abgerufen und ist relevant.

Tabelle 5. Kernelemente der verwendeten Evaluierungskennzahlen (Schütze et al., 2008, s. 359)

Die häufigsten und grundlegendsten Maße für die Effektivität von Information-Retrieval-Systemen sind Precision und Recall (vgl. Manning et al., 2008).

2.7.2. Recall

Die Formel für den Recall (R) lautet: Anzahl der gefundenen relevanten Objektedividiert durch die Anzahl der relevanten Objekte in der Datenbank. Dadurch erhält man eine Abschätzung wie viele der relevanten Objekte gefunden wurden (vgl. Manning et al., 2008, s. 155).

2.7.3. Precision

Die Formel für Precision (P), der Genauigkeit, lautet: Anzahl der gefundenen relevanten Objekte dividiert durch die Anzahl aller gefundenen Objekte. Somit wird dadurch der Anteil der gefundenen und relevanten Objekte im Verhältnis zu den gefundenen Objekten angegeben (vgl. Manning et al., 2008, s. 155).

Werte für Recall und Precision liegen immer zwischen 0 und 1. Ein Recall von 1 bedeutet, dass alle relevanten Objekte gefunden wurden. Eine Precision von 1 bedeutet, dass alle gefundenen Objekte relevant sind (vgl. Manning et al., 2008, s. 155).

2.7.4. Accuracy

Da das Information-Retrieval-System als Zwei-Klassen-Klassifikator betrachtet werden kann, wäre eine offensichtliche Alternative zur Beurteilung der Genauigkeit das Maß Accuracy. Die Formel für Accuracy lautet: Richtig positive Objekte summiert mit richtig negativen Objekten geteilt durch die Summe aller vorkommenden Objekte (vgl. Schütze et al., 2008, s. 155).

Da aber bei einem Information-Retrieval-System eine der beiden Klassen meist über- oder unterrepräsentiert ist (relevant versus nicht relevant), ist dies kein gutes Maß zur Beurteilung (vgl. Manning et al., 2008, s. 155).

2.7.5. F_1-Score

Es handelt sich hierbei um ein gleich gewichtetes harmonisches Mittel aus Precision und Recall. Es ermöglicht die Gesamtgenauigkeit der Vorhersagen eines Klassifikators zu beschreiben. Ein gewisser Grad an Recall, der dabei nur einen bestimmten Prozentsatz an falsch positiven Ergebnissen liefert, ist im Allgemeinen das Ziel (vgl. Manning et al., 2008, s. 155-156).

2.7.6. Macro-average und Weighted-average

Der Macro-average berechnet klassenunabhängig die Metrik für jede einzelne Klasse und nimmt dann den Durchschnitt. Sind die Daten aber unausgewogen, muss einigen Vorhersagen, basierend auf ihrem Anteil, mehr Bedeutung beigemessen werden. Dann wird der "gewichtete Durchschnitt", also der Weighted-average verwendet (vgl. Manning et al., 2008, s. 281).

2.7.7. Cohens Kappa

Es handelt sich um ein Maß für die Übereinstimmung zweier kategorialer Stichproben. Es können dabei sowohl die zweifache Bewertung einer beurteilenden Person sowie die einfache Bewertung zweier verschiedener beurteilenden Personen verglichen werden. Der Wert kann zwischen 0 und 1 liegen, wobei 1 vollständige Übereinstimmung ausdrückt. Cohens Kappa (κ) ist ein wichtiges Maß für die Zuverlässigkeit der Methodik (vgl. Hripcsak und Wilcox, 2002). Cohens Kappa wird durch Division von p_0 minus p_e dividiert durch 1 minus p_e errechnet. p_0 stellt die empirische Wahrscheinlichkeit der Übereinstimmung dar, p_e wird mit einem empirischen Prior pro Annotator über die Klassenlabels geschätzt (vgl. Di Eugenio und Glass, 2004). Interrater-Übereinstimmungsindizes bewerten das Ausmaß, in dem die Antworten von zwei oder mehr unabhängigen Ratern übereinstimmen. Interrater-Reliabilitätsindizes bewerten das Ausmaß, in dem Rater konsistent zwischen den verschiedenen Antworten unterscheiden (vgl. Gisev et al., 2013).

3. Ergebnisse

Die Ergebnisse[2] beider Modellansätze wurden mit der folgenden Parametrierung erreicht, wobei die Character-N-Gramm-Zerlegung für $2 \leq N \leq 5$ variiert wurde, um den optimalen Punkt für das Experiment zu finden. Die übrigen Parameter wurden von den Default-Einstellungen der fastText Library übernommen:

"Without pretrained model": Lernrate = 1,0; Anzahl der Trainingsepochen = 50; beste Char-N-Gramm-Spanne: min=2, max.=5; Embedding-Dimensionalität = 300.

"With pretrained model": Lernrate = 1,0; Anzahl der Trainingsepochen = 50; beste Char-N-Gramm-Spanne: min=2, max.=4; Embedding-Dimensionalität = 300.

3.1. Kennzahlen

Without pretrained model	Precision	Recall	F_1-score	Support
Expert	0.70	0.71	0.71	325
Layman	0.80	0.78	0.79	460
Macro-average	0.75	0.75	0.75	785
Weighted-average	0.76	0.76	0.76	785

Abbildung 3. Ergebnisse der Klassifikation medizinischer Terme in Laien- und Expertenterme ohne Verwendung des vortrainierten Sprachmodells fastText.

With pretrained model	Precision	Recall	F_1-score	Support
Expert	0.69	0.71	0.70	325
Layman	0.79	0.78	0.79	460
Macro-average	0.74	0.74	0.74	785
Weighted-average	0.75	0.75	0.75	785

Abbildung 4. Ergebnisse der Klassifikation medizinischer Terme in Laien- und Expertenterme mit Verwendung des vortrainierten Sprachmodells fastText.

[2] https://gitlab.com/imi-theses/lay-expert-language/lay-expert-classification

3.2. Visualisierung mittels t-Distributed Stochastic Neighbor Embedding

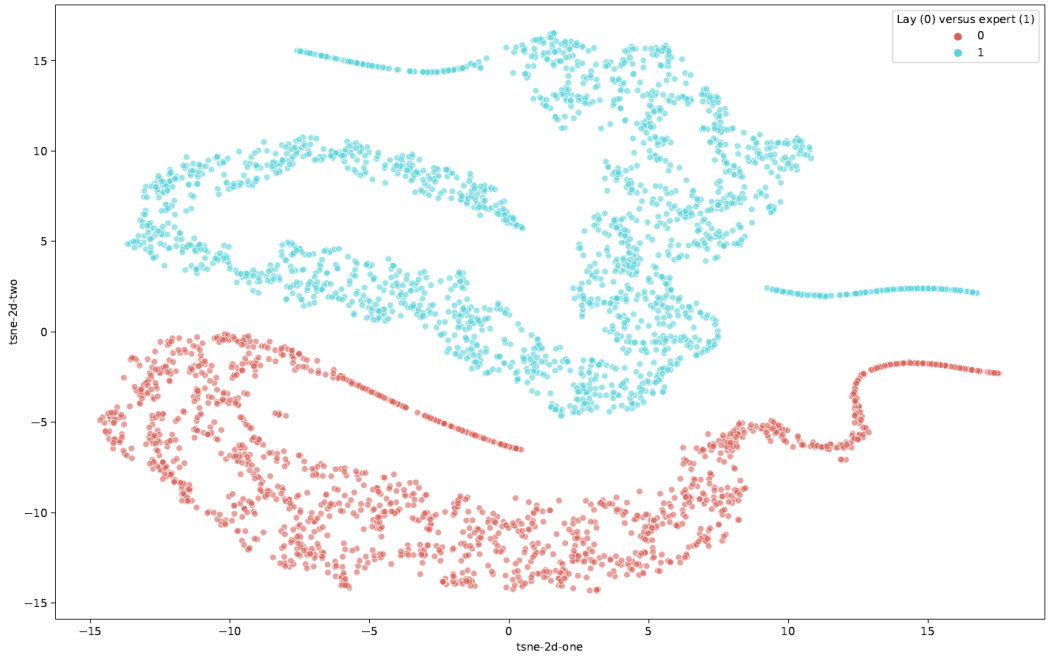

Abbildung 5. t-SNE Visualisierung der Embedding-Repräsentation ohne Verwendung des vortrainierten Sprachmodells.

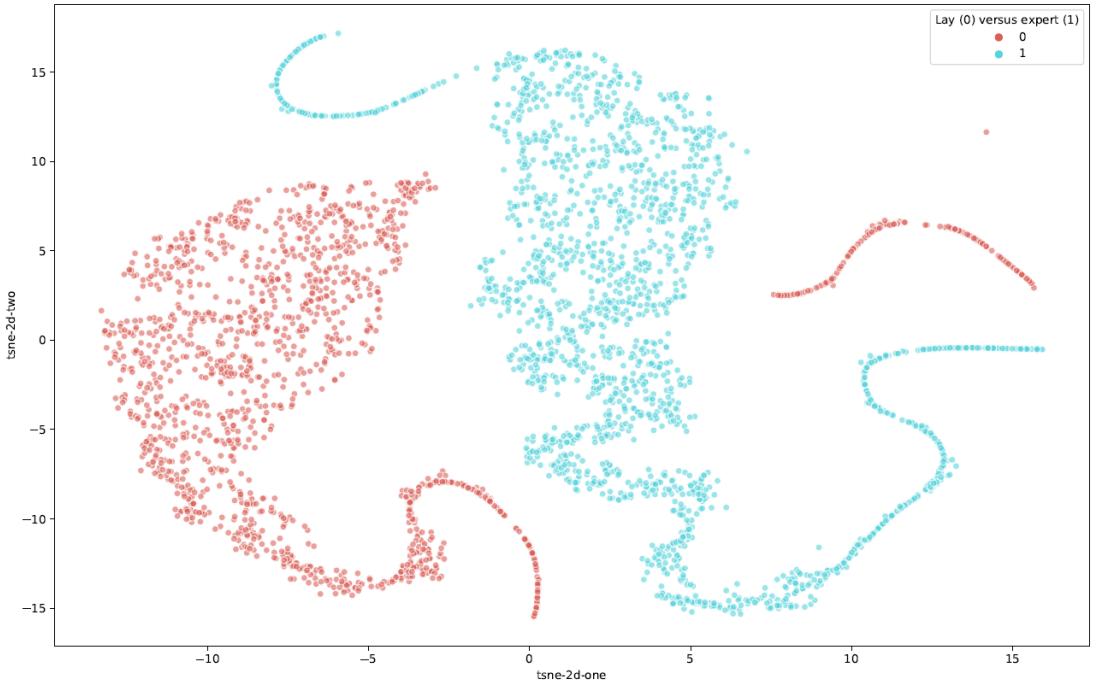

Abbildung 6. t-SNE Visualisierung der Embedding-Repräsentation unter Verwendung des vortrainierten Sprachmodells.

3.3. Confusion Matrix

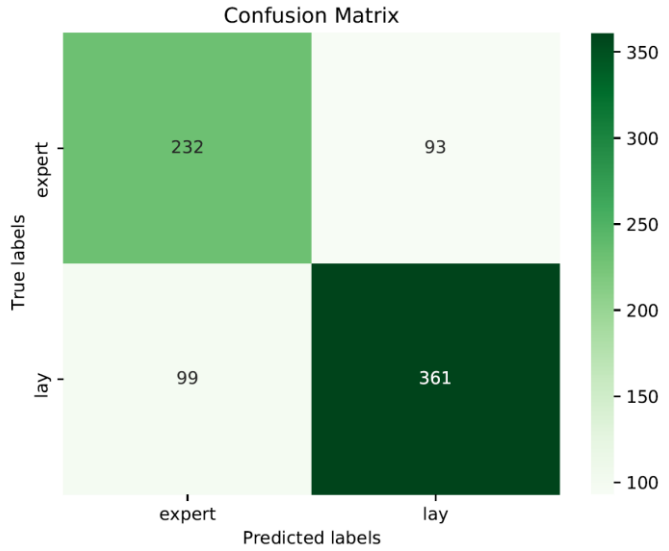

Abbildung 7. Visualisierung der Confusion Matrix ohne Verwendung des vortrainierten Sprachmodells.

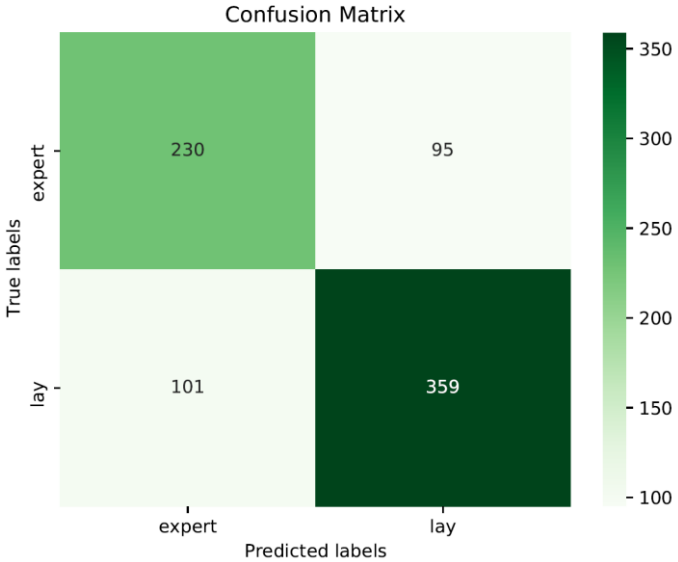

Abbildung 8. Visualisierung der Confusion Matrix mit Verwendung des vortrainierten Sprachmodells.

4. Diskussion

Laien- und Expertenterme können mittels F_1-Score von 0,76 ohne Verwendung eines vortrainierten Sprachmodells und einem F_1-Score von 0,75 mit Verwendung eines vortrainierten Sprachmodells unterschieden werden (siehe Abbildung 3. und 4.). Die mögliche Unterscheidbarkeit zwischen Laien- und Expertentermen wurde auch über eine Visualisierung im zweidimensionalen Raum (t-SNE) angedeutet (siehe Abbildung 5. und 6.).

Mit einem Cohens Kappa von 0,71 zeigt sich, dass die verwendete Methode des maschinellen Lernens das Niveau von zwei menschlichen Annotatoren erreicht. Auf Basis der erstellten Annotationsrichtlinien kann das System daher die Unterscheidung zwischen Laien- und Expertenterm so gut wie menschliche Experten durchführen, was so interpretiert werden kann, dass die Richtlinie durch das Modell abstrahiert wurde. Die Methode sollte sich daher im praktischen Einsatz bewähren, um laiensprachliche von expertensprachlichen Inhalten zu unterscheiden, um so beispielsweise in Arztbriefen oder Befundberichten Expertenterme durch laienfreundliche Terme zu ersetzen.

Ein Chi-Quadrat-Test ($X^2(1) = 0{,}055$, $p = {,}815$) zeigt, dass die Verwendung des vortrainierten Sprachmodells für diese Klassifikationsaufgabe keinen signifikanten Einfluss auf die Trennschärfe zwischen Laien- und Expertentermen in diesem Experiment hat. Die dafür verwendeten Vierfeldertafeln sind in Abbildung 7. und 8. visualisiert.

4.1. Fehleranalyse

Es erfolgt eine Auswahl anhand der an die Annotationsrichtlinien angepassten Analyse von falsch positiven Objekten und falsch negativen Objekten anhand des erstellten Goldstandards. Für die Fehleranalyse wurden die Ergebnisse der Klassifikationsmethode ohne Verwendung eines vorab trainierten Sprachmodells verwendet.

"Arteriosklerose" wurde im Goldstandard anhand der Annotationsrichtlinien richtig als Expertenterm bezeichnet, von der Maschine aber als Laienterm. "Medizinethiker" wurde im Goldstandard als Expertenterm klassifiziert, von der Maschine aber als Laienterm.

"Spezielle Lebenssituation" wurde anhand des Goldstandards als Expertenausdruck bezeichnet, von der Maschine aber richtig als Laienterm klassifiziert. Hier handelt es sich um eine Unschärfe im Goldstandard, was bei einem Interrater-Agreement von über 0,71 nicht verwunderlich ist.

"Zahnschonend" wurde anhand des Goldstandards richtig als Laienterm klassifiziert, von der Maschine aber als Expertenterm interpretiert. "Dynamisch" und "Netzhaut" sind laut Goldstandard Laienterme, von der Maschine wurden sie als Expertenterme klassifiziert.

Grundsätzlich besteht bei den ausgewählten Beispielen die Schwierigkeit, menschliche Ähnlichkeitsurteile und deren Eigenschaften im n-dimensionalen Raum abzubilden. Ein erweiterter Kontext, in dem das Wort eingebettet ist, könnte zusätzliche Informationen für die Unterteilung als Laien- oder Expertenterm geben.

4.2. Limitationen

Ein spezieller Algorithmus (fastText) wurde anhand von Anforderungen ausgewählt und angewandt, man könnte aber auch verschiedene Algorithmen des maschinellen Lernens heranziehen und diese miteinander vergleichen. Außerdem hätte der Wortkontext noch weiter mit eingebunden werden können. Da sich diese Arbeit grundsätzlich nur mit der Unterscheidbarkeit zwischen Laien- und Expertentermen befasst, kann dies nur eine Basis für weitere anwendungsorientierte Betrachtungen sein.

4.3. Zusammenfassung und Ausblick

Es wurde mittels Methoden des maschinellen Lernens überprüft, ob Laien- und Expertenterme erfolgreich unterschieden werden können. Hierzu wurden in einem ersten Schritt Annotationsrichtlinien erstellt. Ein Goldstandard konnte daraus erarbeitet werden (Cohens Kappa κ = 0,71), um als ausgewähltes maschinelles Lernverfahren fastText anzuwenden. Dieser Algorithmus wurde anhand von einer Anforderungsliste selektiert, wobei vor allem eine Character-N-Gramm Zerlegung für die Berechnung der Embedding-Repräsentation der zu untersuchenden Terme als relevant betrachtet wurde. Der fastText-Algorithmus konnte Laien- und Expertenterme gut unterscheiden, und zwar mit einem F_1-Score von 0,76 ohne Verwendung eines vortrainierten Sprachmodells und mit einem F_1-Score von 0,75 mit Verwendung eines auf deutschen Texten vortrainierten Sprachmodells. So ist hervorzuheben, dass das in einem frei verfügbaren deutschen Sprachmodell vorhandene Wissen für unsere Klassifikationsaufgabe keinen Mehrwert bot.

Eine Anwendung des erstellten Algorithmus könnte in der Charakterisierung deutschsprachiger Einträge bei medizinischen Thesauri sein. Dazu müsste der beschriebene Klassifikator auf eine möglichst große Menge von medizinischen Termen angewendet werden. Sofern dieser Thesaurus Synonymbeziehungen enthält, könnten Laienterme zu Expertentermen gefunden werden.
Dies kann einerseits medizinische AutorInnen bei der Erstellung patientengerechter Dokumente unterstützen. Bei der Formulierung "Der Patient leidet unter einer Thrombose." würde dann zusätzlich zu dem Term "Thrombose", der Term "Blutgerinnsel" angeboten. Dies könnte eine Unterstützung für das Schreiben von laiengerechten Inhalten sein.
Andererseits könnte ein durch den Algorithmus aufgewerteter, also mit Markierungen wie "L" und "E" versehener Thesaurus in Anwendungen integriert werden, die Laien beim Lesen medizinischer Texte unterstützen, in dem diese Texte automatisch mit Laientermen angereichert werden.

5. Literaturverzeichnis

1. Åhlfeldt, H., Borin, L., Grabar, N., Hallett, C., Hardcastle, D., Kokkinakis, D., Mancini, C., Markó, K., Merkel, M., Pietsch, C., Power, R., Scott, D., Silvervarg, A., Gronostaj, M., Williams, S., Willis, A., 2006. Literature Review on Patient-Friendly Documentation Systems. http://mcs.open.ac.uk/sw6629/Publications/TR2006_04.pdf
2. Baethge, C., 2008. Die Sprachen der Medizin. Dtsch Arztebl Int. 105, 37–40. https://doi.org/10.3238/arztebl.2008.0037
3. Beddies, Th., Brinkschulte, E., Brumme, M., Hess, V., Marz, I., Müller, Th., Scholl, A., 2008. Medizinische Terminologie. Berlin.
4. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T., 2017. Enriching Word Vectors with Subword Information. Trans. Assoc. Comput. Linguist. 5, 135–146. https://doi.org/10.1162/tacl_a_00051
5. Conti, A.A., 2013. Medical terminology and lay users. A quali-quantitative survey of a group of young motivated graduates. Clin. Ter. 164, e297-300. https://doi.org/10.7417/CT.2013.1592
6. Deo, R.C., 2015. Machine Learning in Medicine. Circulation 132, 1920–1930. https://doi.org/10.1161/CIRCULATIONAHA.115.001593
7. Di Eugenio, B., Glass, M., 2004. Squibs and Discussions - The Kappa Statistic: A Second Look. Comput. Linguist. 30, 95–101. https://doi.org/10.1162/089120104773633402
8. Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S., 2017. Dermatologist-level classification of skin cancer with deep neural networks. Nature 542, 115–118. https://doi.org/10.1038/nature21056
9. Feature engineering, 2021. . Wikipedia. https://en.wikipedia.org/wiki/Feature_engineering
10. Feeny, A.K., Chung, M.K., Madabhushi, A., Attia, Z.I., Cikes, M., Firouznia, M., Friedman, P.A., Kalscheur, M.M., Kapa, S., Narayan, S.M., Noseworthy, P.A., Passman, R.S., Perez, M.V., Peters, N.S., Piccini, J.P., Tarakji, K.G., Thomas, S.A., Trayanova, N.A., Turakhia, M.P., Wang, P.J., 2020. Artificial Intelligence and Machine Learning in Arrhythmias and Cardiac Electrophysiology. Circ. Arrhythm. Electrophysiol. 13, e007952. https://doi.org/10.1161/CIRCEP.119.007952
11. Gisev, N., Bell, J.S., Chen, T.F., 2013. Interrater agreement and interrater reliability: key concepts, approaches, and applications. Res. Soc. Adm. Pharm. RSAP 9, 330–338. https://doi.org/10.1016/j.sapharm.2012.04.004
12. Grabar, N., 2019. Adaptation de documents techniques pour les locuteurs non spécialisés. Universität Paris-Süd. https://docplayer.fr/158999817-Adaptation-de-documents-techniques-pour-les-locuteurs-non-specialises.html
13. Grave, E., Bojanowski, P., Gupta, P., Joulin, A., Mikolov, T., 2018. Learning Word Vectors for 157 Languages, in: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). European Language Resources Association (ELRA), Miyazaki, Japan. https://arxiv.org/abs/1802.06893v2

14. Gu, G., Zhang, X., Zhu, X., Jian, Z., Chen, K., Wen, D., Gao, L., Zhang, S., Wang, F., Ma, H., Lei, J., 2019. Development of a Consumer Health Vocabulary by Mining Health Forum Texts Based on Word Embedding: Semiautomatic Approach. JMIR Med. Inform. 7, e12704. https://doi.org/10.2196/12704
15. Gurunath Shivakumar, P., Georgiou, P., 2019. Confusion2Vec: towards enriching vector space word representations with representational ambiguities. PeerJ Comput. Sci. 5, e195. https://doi.org/10.7717/peerj-cs.195
16. Hamel, R., 2007. The dominance of English in the international scientific periodical literature and the future of language use in science. AILA Rev. 20, 53–71. https://doi.org/10.1075/aila.20.06ham
17. Handelman, G.S., Kok, H.K., Chandra, R.V., Razavi, A.H., Lee, M.J., Asadi, H., 2018. eDoctor: machine learning and the future of medicine. J. Intern. Med. 284, 603–619. https://doi.org/10.1111/joim.12822
18. Høy, A., Howarth, J., 2012. Guidelines for Translation of SNOMED CT® 37. https://www.snomed.org/SNOMED/media/SNOMED/documents/IHTSDO_Translation_Guidelines_v2_02_20121211-(1).pdf
19. Hripcsak, G., Wilcox, A., 2002. Reference Standards, Judges, and Comparison Subjects. J. Am. Med. Inform. Assoc. JAMIA 9, 1–15.
20. Hüllemann, K.-D.-, 2013. Patientengespräche besser gestalten, 1.Auflage, Carl-Auer Verlag. Heidelberg.
21. Jahangiry, L., Farhangi, M.A., Rezaei, F., 2017. Framingham risk score for estimation of 10-years of cardiovascular diseases risk in patients with metabolic syndrome. J. Health Popul. Nutr. 36, 36. https://doi.org/10.1186/s41043-017-0114-0
22. Joulin, A., Grave, E., Bojanowski, P., Mikolov, T., 2016. Bag of Tricks for Efficient Text Classification. https://arxiv.org/abs/1607.01759v3
23. Karenberg, A., 2014. Fachsprache Medizin im Schnellkurs, 3.Auflage Schattauer GmbH, Stuttgart.
24. Lip, G.Y.H., Nieuwlaat, R., Pisters, R., Lane, D.A., Crijns, H.J.G.M., 2010. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. Chest 137, 263–272. https://doi.org/10.1378/chest.09-1584
25. Mamrosh, J.L., Moore, D.D., 2015. Using Google Reverse Image Search to Decipher Biological Images. Curr. Protoc. Mol. Biol. 111, 19.13.1-19.13.4. https://doi.org/10.1002/0471142727.mb1913s111
26. Manning, Ch.D., Raghavan, P., Schütze, H., 2008. An Introduction to Information Retrieval. Cambridge University Press, Cambridge.
27. Michler, M., Benedum, J., 1981. Einführung in die Medizinische Fachsprache: Medizinische Terminologie für Mediziner und Zahnmediziner auf der Grundlage des Lateinischen und Griechischen, 2. Auflage. Springer-Verlag, Berlin Heidelberg. https://doi.org/10.1007/978-3-642-68015-1
28. Murken, A.H., 2009. Lehrbuch der Medizinischen Terminologie. Wissenschaftliche Verlagsgesellschaft Stuttgart.
29. O'Mahony, C., Jichi, F., Pavlou, M., Monserrat, L., Anastasakis, A., Rapezzi, C., Biagini, E., Gimeno, J.R., Limongelli, G., McKenna, W.J., Omar, R.Z., Elliott, P.M., Hypertrophic Cardiomyopathy Outcomes Investigators, 2014. A

novel clinical risk prediction model for sudden cardiac death in hypertrophic cardiomyopathy (HCM risk-SCD). Eur. Heart J. 35, 2010–2020. https://doi.org/10.1093/eurheartj/eht439
30. Pieterse, A.H., Jager, N.A., Smets, E.M.A., Henselmans, I., 2013. Lay understanding of common medical terminology in oncology. Psychooncology. 22, 1186–1191. https://doi.org/10.1002/pon.3096
31. Popilka, T., 2014. Paralleltexte und andere Hilfsmittel in der translatorischen Praxis. Universität Wien.
32. Roelcke, T., 1999. Fachsprachen. Schmidt, Berlin.
33. van der Maaten, L., Hinton, G., 2008. Visualizing Data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605.
34. Vydiswaran, V.G.V., Mei, Q., Hanauer, D.A., Zheng, K., 2014. Mining Consumer Health Vocabulary from Community-Generated Text. AMIA. Annu. Symp. Proc. 2014, 1150–1159.
35. Woodruff, P.G., Modrek, B., Choy, D.F., Jia, G., Abbas, A.R., Ellwanger, A., Koth, L.L., Arron, J.R., Fahy, J.V., 2009. T-helper type 2-driven inflammation defines major subphenotypes of asthma. Am. J. Respir. Crit. Care Med. 180, 388–395. https://doi.org/10.1164/rccm.200903-0392OC
36. Zeng, Q.T., Tse, T., 2006. Exploring and Developing Consumer Health Vocabularies. J. Am. Med. Inform. Assoc. 13, 24–29. https://doi.org/10.1197/jamia.M1761
37. Zhu, W., Jin, X., Ni, J., Wei, B., Lu, Z., 2018. Improve word embedding using both writing and pronunciation. PLoS ONE 13, e0208785. https://doi.org/10.1371/journal.pone.0208785

Stand der Weblinks: 12.7.2021