



universität
wien

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Automatische Gewinnung polnischer und österreichischer Rechtsterminologie
aus den gerichtlichen Entscheidungen im Fachgebiet des Urheberrechts

verfasst von | submitted by

lic. Michał Sterczyński mag.

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Master of Arts (MA)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt |
Degree programme code as it appears on the
student record sheet:

UA 070 375 331

Studienrichtung lt. Studienblatt | Degree pro-
gramme as it appears on the student record
sheet:

Masterstudium Translation Polnisch Deutsch

Betreut von | Supervisor:

Assoz. Prof. Mag. Dr. Dagmar Gromann BSc

Inhaltsverzeichnis

Abbildungsverzeichnis	4
Tabellenverzeichnis.....	5
1. Einleitung	6
1.1. Forschungsfragen und Grundannahmen.....	8
1.2. Motivation und Zielsetzung	9
1.3. Detaillierter Aufbau der Arbeit	10
2. Terminologie	13
2.1. Grundlagen der Terminologielehre	13
2.2. Grundprinzipien der Terminologearbeit	19
2.3. Terminologiegewinnung	25
2.3.1. Manuelle Terminologiegewinnung	26
2.3.2. Automatische Terminologiegewinnung	27
2.3.2.1. Herausforderungen der automatischen Terminologiegewinnung	29
2.3.2.2 Termextraktionssysteme.....	30
3. Verfahren der automatischen Terminologiegewinnung	33
3.1. Arbeitsablauf der automatischen Termextraktion	33
3.1.1. Linguistische Verfahren	37
3.1.2. Statistische Verfahren	40
3.1.3. Hybride Verfahren.....	45
3.1.4. Verfahren des Maschinellen Lernens und Neuronale Verfahren	48
3.1.4.1. Text2TCS	54
3.1.4.2. ChatGPT-4o	55
3.2. Prompt Engineering.....	57
4. Evaluierungsmaßnahmen der automatischen Termextraktion	60
4.1. Grundsätze des Evaluierungsverfahrens	60
4.2. Mengenorientierte Evaluationsmaße	63
5. Fachsprachen	69
5.1. Allgemeine, fachliche und berufliche Kommunikation	69
5.1.2. Definition von Fachsprache	71
5.1.3. Funktionale Eigenschaften von Fachsprachen	76
5.1.4. Merkmale fachsprachlicher Lexik.....	77
5.1.5. Merkmale fachsprachlicher Grammatik.....	78

5.1.6. Fachsprache und Gemeinsprache	79
5.2. Rechtssprache.....	82
5.2.1. Definition der Rechtssprache	83
5.2.2. Merkmale der Rechtssprache	84
5.2.2.1. Lexikalische Besonderheiten der Rechtssprache	89
5.2.2.2. Grammatikalische Besonderheiten der Rechtssprache	90
6. Urheberrecht.....	93
6.1. Grundprinzipien des Urheberrechts	93
6.1.1. Das österreichische Urheberrechtsgesetz	95
6.1.2. Das polnische Urheberrechtsgesetz.....	97
6.2. Urheberrecht und künstliche Intelligenz	99
7. Aktueller Stand der Forschung.....	103
8. Methodik	110
8.1. Textauswahl	110
8.2. Systemauswahl	113
8.3. Vorgehensweise und Validierung	113
8.4. Auswertungsmethode	116
9. Ergebnisse	118
9.1. Quantitative Evaluierung von nicht lemmatisierten Termkandidaten.....	118
9.2. Quantitative Evaluierung von lemmatisierten Termkandidaten.....	126
9.3. Qualitative Analyse der automatisch generierten Termkandidaten	134
10. Diskussion	138
10.1. Interpretation der Ergebnisse	138
10.2. Herausforderungen und Beschränkungen der Untersuchung.....	140
10.3. Weiterführende Forschung.....	141
11. Fazit	143
12. Literaturverzeichnis.....	146
Anhangsverzeichnis	172
Abstract (Deutsch).....	216
Abstract (English)	217

Abbildungsverzeichnis

Abbildung 1: Das semiotische Dreieck	14
Abbildung 2: Ein adaptiertes semiotisches Dreieck	14
Abbildung 3: Schematischer Ablauf automatischer Termextraktion	35
Abbildung 4: Die Architektur eines hybriden Verfahrens zur automatischen Termextraktion.....	48
Abbildung 5: Beispiel für die Architektur eines Transformer-Modells	52
Abbildung 6: Beispiel für die Funktionsweise des Attention-Mechanismus	52
Abbildung 7: Die Entwicklungsstufen von ChatGPT	53
Abbildung 8: Text2TCS	55
Abbildung 9: ChatGPT-4o: Extraktionsverfahren.....	56
Abbildung 10: Externe Eigenschaften für die Black-Box-Evaluierung	62
Abbildung 11: Typische Evaluationsmaße am Gebiet der automatischen Terminologiegewinnung....	64
Abbildung 12: Genauigkeit und Trefferquote – die typische Verhaltensweise.....	66
Abbildung 13: Die Berechnung vom F1-Score	67
Abbildung 14: Das Sender-Empfänger-Modell.....	70
Abbildung 15: Vereinfachtes Modell fachsprachlicher Kommunikation: Systemlinguistisches Inventarmodell, pragmalinguistisches Kontextmodell, kognitionslinguistisches Funktionsmodell	71
Abbildung 16: Das Schema: Gesamtsprache und Subsprachen	80
Abbildung 17: KG Annotation Tool	114
Abbildung 18: Ergebnisse von Text2TCS und Goldstandard (nicht lemmatisiert).....	119
Abbildung 19: Ergebnisse von ChatGPT und Goldstandard (Durchschnittswert, nicht lemmatisiert)123	
Abbildung 20: Ergebnisse von Text2TCS und Goldstandard (lemmatisiert).....	127
Abbildung 21: Ergebnisse von ChatGPT und Goldstandard (Durchschnittswert, lemmatisiert).....	130

Tabellenverzeichnis

Tabelle 1: Übersicht über Termextraktionsmethoden bei linguistischen und statistischen Verfahren..	36
Tabelle 2: Ein dem linguistischen Extraktionsverfahren unterzogener Text mit Termkandidaten, die den Noise- bzw. den Silence-Effekt illustrieren.....	38
Tabelle 3: GPT-3 und BERT–Vergleich	53
Tabelle 4: Überblick über verschiedene Evaluierungsmetriken im Bereich der automatischen Termextraktion	63
Tabelle 5: Beispiel einer Konfusionsmatrix	65
Tabelle 6: Die für die Untersuchung verwendeten Prompts.....	116
Tabelle 7: Evaluierungsergebnisse von Text2TCS (nicht lemmatisiert).....	119
Tabelle 8: Ergebnisse von ChatGPT Prompt 1 (nicht lemmatisiert)	121
Tabelle 9: Ergebnisse von ChatGPT Prompt 2 (nicht lemmatisiert)	121
Tabelle 10: Ergebnisse von ChatGPT Prompt 3 (nicht lemmatisiert)	122
Tabelle 11: Ergebnisse von ChatGPT Prompt 4 (nicht lemmatisiert)	122
Tabelle 12: Ergebnisse von ChatGPT Prompt 5 (nicht lemmatisiert)	123
Tabelle 13: Ergebnisse vom Gesamtkorpus für Text2TCS und ChatGPT (nicht lemmatisiert)	125
Tabelle 14: Evaluierungsergebnisse von Text2TCS (lemmatisiert).....	126
Tabelle 15: Ergebnisse von ChatGPT Prompt 1 (lemmatisiert)	128
Tabelle 16: Ergebnisse von ChatGPT Prompt 2 (lemmatisiert)	128
Tabelle 17: Ergebnisse von ChatGPT Prompt 3 (lemmatisiert)	128
Tabelle 18: Ergebnisse von ChatGPT Prompt 4 (lemmatisiert)	129
Tabelle 19: Ergebnisse von ChatGPT Prompt 5 (lemmatisiert)	129
Tabelle 20: Ergebnisse vom Gesamtkorpus für Text2TCS und ChatGPT (lemmatisiert)	131
Tabelle 21: Unterschiede nach der Lemmatisierung für Text2TCS und ChatGPT	132
Tabelle 22: Die besten Ergebnisse unter Prompts anhand der Durchschnittswerte	133
Tabelle 23: ChatGPT: Erkennung gemeinsprachlicher Ausdrücke mit fachlicher Bedeutung	135
Tabelle 24: Extraktion von Kurzwörtern durch Text2TCS und ChatGPT	137

1. Einleitung

Die heutige Welt steht im Zeichen der Globalisierung. Menschen werden immer mehr miteinander vernetzt und dies über Grenzen, Kulturen und Sprachen hinweg. Unternehmen öffnen ihre Filialen bzw. verlagern einzelne Abteilungen, wie Produktion oder Buchhaltung ins Ausland und der Druck, alle Kund:innen weltweit auf dem höchsten professionellen Niveau zu betreuen, steigt. Infolgedessen gewinnt die Terminologie immer mehr an Sichtbarkeit und Relevanz, denn Unternehmen wollen bzw. müssen ihre Produkte und Dienstleistungen international in der Sprache der Kund:innen anbieten und dazu gehören Bedienungsanleitungen, rechtliche Bestimmungen, Versicherungen, technische Dokumentation, Beipackzettel und andere fachspezifische Texte, deren festen Bestandteil Terminologie bildet (vgl. Europäische Kommission 2023). Somit wächst der Markt für Terminologiedatenbanken und die Anzahl an Agenturen, die sie zusammenstellen. Einige Unternehmen gründen auch eigene Abteilungen für Terminologiemanagement und die Europäische Kommission fördert den Ausbau der Terminologie-Infrastruktur. Zu diesem Zweck wurde 2004 die Interactive Terminology for Europe (IATE) gegründet. Die Institution sammelt, verbreitet und verwaltet EU-spezifische Terminologie für EU-Einrichtungen (vgl. IATE 2024; Teubert 2006: 96–97).

Der Prozess der manuellen Terminologiegewinnung ist allerdings langwierig und mit hohen Kosten verbunden, was die immer ansteigende Automatisierung in diesem Bereich zufolge hat (vgl. Tran et al. 2024: 4286). Allerdings erfordern Termini, die nicht in Wörterbüchern enthalten sind, seltene Fachausdrücke und Mehrwortbenennungen oft sehr genaue Termextraktionssysteme (vgl. Le & Sadat 2021: 1). Dies wird durch die neuesten Entwicklungen auf dem Gebiet der künstlichen Intelligenz und der neuronalen Netze begünstigt, weshalb das Thema der automatischen Terminologiegewinnung ein durchaus aktuelles ist (vgl. Tran et al. 2023: 1). Am Markt sind mehrere Termextraktionssysteme verfügbar, sowohl jene, die ausschließlich darauf spezialisiert sind als auch solche, die Termextraktion nur als eine ihrer Funktionen anbieten. In den letzten Jahren wurden verschiedene Systeme wissenschaftlich getestet, ihre Erfolgsraten miteinander verglichen und Ergebnisse publiziert. Es handelt sich um Tools, die sowohl auf modernen neuronalen als auch auf traditionellen bzw. hybriden Ansätzen basieren (vgl. Tran et al. 2023: 2). Die primäre Sprache für die Erforschung der automatischen Terminologiegewinnung ist Englisch, aber es sind auch Forschungen und Publikationen vorhanden,

in denen unterrepräsentierte Sprachen untersucht werden, für die (deutlich) weniger Trainingsdaten zur Verfügung stehen wie z. B. Irisch, Griechisch, Litauisch, Baskisch u.v.m. (vgl. Alegria et al. 2004; McCrae & Doyle 2019; Mouratidis et al. 2022; Rokas et al. 2020). Die vorliegende Arbeit vergleicht die Leistung von ChatGPT-4o (OpenAI 2024), einem neuronalen generativen Sprachmodell mit Terminologiegewinnung als Nebenfunktion, und Text2TCS (Gromann et al. 2023), einem primär für diese Aufgabe entwickelten Transformer-basierten neuronalen Sprachmodell. Weiters wird zur Untersuchung Deutsch, eine Sprache mit relativ reichlich verfügbaren Ressourcen, und Polnisch, eine Sprache mit vergleichsweise weniger verfügbaren Ressourcen herangezogen, um zu erforschen, welche Auswirkung verschiedene Mengen an Trainingsdaten auf die Erfolgsrate der automatischen Terminologiegewinnung haben. Schließlich werden bei der Untersuchung, wegen der Funktionsweise von ChatGPT-4o, unterschiedlich formulierte Anfragen berücksichtigt.

Die Rechtssprache ist, linguistisch gesehen, sehr herausfordernd. Dies liegt beispielsweise an dem abwechslungsreichen Wortschatz, denn diese Sprache beinhaltet nicht nur typisch rechtliche Fachausdrücke, sondern auch Fachtermini aus anderen Fachsprachen und Wörter aus der Allgemeinsprache, die entweder ihre ursprüngliche Bedeutung behalten können oder im rechtlichen Kontext eine fachliche Bedeutung erhalten (vgl. Ralli & Andreatta 2018: 7–8). Außerdem meinen Arntz und Sandrini (2007: 3), dass im Falle der Rechtssprache sowohl einerseits die Präzision der Termini als auch andererseits der hohe Grad an deren Allgemeingültigkeit angestrebt werden. Dies ist auf die Rechtssicherheit und auf die vielseitige Anwendbarkeit der Rechtstermini zurückzuführen und es kann eine Herausforderung für die automatische Termextraktion darstellen. Weiters ist im Zeitalter des Internets, der Digitalisierung und der künstlichen Intelligenz das Urheberrecht ein besonders aktueller Rechtsbereich, der viel besprochen wird (vgl. De la Durantaye & Raue 2020: 83–84; Konertz 2023: 796). Zudem betrifft er auch die Thematik der automatischen Terminologiegewinnung, da neuronale Sprachmodelle anhand von großen Mengen an Daten trainiert werden. Hier liegt die Frage nahe, ob für das Training auch urheberrechtlich geschützte Texte verwendet werden sollen bzw. dürfen. Utko et al. (2022: 136) weisen in diesem Zusammenhang auf die Einschränkungen bei der Erstellung von Korpora hin und nennen wissenschaftliche Publikationen als Beispiel für Texte, die vom Urheberrecht geschützt werden und nicht bedenkenlos gespeichert und im Training eingesetzt werden dürfen (2022: 136).

Die vorliegende Arbeit setzt sich aus den oben genannten Themen zusammen. Sie betrifft den Vergleich der Leistung von zwei verschiedenen Tools, die Termini aus Fragmenten von polnischen und österreichischen gerichtlichen Entscheidungen aus dem Bereich des Urheberrechts automatisch extrahieren. Die Ergebnisse werden quantitativ mithilfe der Qualitätsmetriken Genauigkeit, Trefferquote und F1-Score sowie qualitativ anhand einer manuellen Detailanalyse ermittelt.

1.1. Forschungsfragen und Grundannahmen

Die Erkenntnisse der vorliegenden Arbeit sollen einen Beitrag zur Untersuchung der Brauchbarkeit der automatischen Terminologiegewinnung, insbesondere auf Basis eines neuronalen generativen Sprachmodells, leisten. Es wird erforscht, ob auch das Modell, das nicht spezifisch dafür entwickelt wurde, erfolgreich für die automatische Termextraktion eingesetzt werden kann. Die Forschungsfrage lässt sich demzufolge folgendermaßen formulieren:

Sind die Ergebnisse der automatischen Termextraktion eines neuronalen generativen Sprachmodells, das nicht speziell darauf ausgerichtet ist, vergleichbar mit jenen von einem primär zwecks dieser Funktion entwickelten Tool sowie den Ergebnissen der manuellen Terminologierecherche?

Dies soll am Beispiel der polnischen und österreichischen gerichtlichen Entscheidungen erforscht werden, die die Thematik des Urheberrechts behandeln, aktuell sind und sowohl von Erstsprachler:innen als auch gleichzeitig Fachexpert:innen verfasst wurden. Es werden Fragmente in vergleichbarer Länge aus jeweils drei solchen Entscheidungen pro untersuchte Sprache und Rechtssystem herangezogen. Die Rechtstexte kommen aus dem Rechtsinformationssystem des Bundes (RIS) (Bundeskanzleramt der Republik Österreich 2024a) respektive aus dem Portal Orzeczeń Sądów Powszechnych (Ministerstwo Sprawiedliwości 2024).

Weiters, um die Ergebnisse der automatischen Terminologiegewinnung genauer untersuchen zu können, werden die Resultate der Extraktion des ausgewählten neuronalen generativen Sprachmodells bei fünf unterschiedlich formulierten Anfragen bzw. Prompts präsentiert und mithilfe der Qualitätsmetriken bemessen. Dies soll zeigen, wie Prompt Engineering die Ergebnisse der Termextraktion beeinflusst sowie wie die unterschiedlichen Mengen an Trainingsdaten der beiden untersuchten Sprachen die Resultate begünstigen bzw. benachteiligen. Somit kann die zweite Forschungsfrage wie folgt definiert werden:

Erzielt ein neuronales generatives Sprachmodell unterschiedliche Ergebnisse der Termextraktion bei verschiedenen Prompts und erzielt es bessere Ergebnisse bei der österreichischen als der polnischen Rechtsterminologie?

Anhand der bisherigen theoretischen Auseinandersetzung mit der Fachliteratur (siehe Premasiri et al. 2023; Tran et al. 2023) kann davon ausgegangen werden, dass ein neuronales generatives Sprachmodell, das sich nicht ausschließlich auf die Terminologiegewinnung spezialisiert, nur bedingt zur Termextraktion verwendet werden kann und es schneidet im Vergleich schlechter als ein primär dafür entwickeltes, neuronales Sprachmodell ab. Außerdem variieren bei dem neuronalen generativen Sprachmodell die Ergebnisse je nach Formulierung der Anfrage und es erzielt aufgrund der vorhandenen Trainingsdaten bessere Ergebnisse bei der Extraktion der österreichischen als der polnischen Rechtsterminologie, da größere Datenmengen für Rechtstexte auf Deutsch verfügbar sind als auf Polnisch.

1.2. Motivation und Zielsetzung

Die Aktualität der Thematik der automatischen Termextraktion lässt sich anhand neuester Entwicklungen auf dem Gebiet der neuronalen Netze und der künstlichen Intelligenz kaum bestreiten. Das Thema ist gleichzeitig relevant für den Bereich Translation. Costa et al. (2016) argumentieren, dass automatische Termextraktion den Übersetzer:innen und Dolmetscher:innen erlaubt, zeiteffizient Glossare zu erstellen, die sie in ihrer Arbeit unterstützen. Oliver (2017: 148) sieht bei der automatischen Termextraktion ebenfalls Zeitersparnis, sie wird als Ergänzung zu der Arbeit mit CAT-Tools gesehen, indem eine bilinguale Terminologiedatenbank automatisch noch vor der Übersetzung geschaffen wird. Auch das Urheberrecht ist auf vielerlei Art relevant für die Translation. Einerseits deshalb, da das Thema gerade viel besprochen wird, weswegen es immer mehr Texte entweder aus diesem Bereich oder den Bereich betreffend zu übersetzen geben wird. Andererseits gelten Übersetzer:innen als Autor:innen von den von ihnen gefertigten Übersetzungen und Entwickler:innen von Übersetzungsdaten. Dazu zählen lexikographische und terminologische Daten, Translation-Memory in CAT-Tools sowie Parallelkorpora und es stellt sich auch in diesem Kontext die Frage nach dem urheberrechtlichen Schutz (vgl. Moorkens & Lewis 2019; Sandrini 2014: 186). Außerdem steht auch zum kleineren Teil die Formulierung der Anfrage für die automatische Termextraktion im Fokus der Untersuchung, die Forschung, die Prompt Engineering genannt wird und aktuell an Bedeutung gewinnt (vgl. McMullen 2023).

Nützlich werden die gesammelten Erkenntnisse in erster Linie für Rechtsübersetzer:innen und -dolmetscher:innen sein, indem sie erfahren, inwieweit sie sich bei ihrer Arbeit, z. B. bei der Vorbereitung auf einen Einsatz im Gericht, auf das untersuchte neuronale generative Sprachmodell verlassen sollen. Dasselbe gilt für Rechtsexpert:innen und Publizist:innen, die automatische Terminologiegewinnung z. B. bei der Textproduktion einsetzen könnten und sich darüber informieren möchten, wie gut sich das präsentierte Tool dafür eignet. Weiters werden auch Projektleiter:innen in Übersetzungsagenturen die Informationen von Nutzen finden, insbesondere bei der Erstellung der Terminologiedatenbanken als Hilfestellung für Übersetzer:innen. Die Arbeit hilft außerdem eine Forschungslücke zu füllen, denn ChatGPT wurde erst im November 2022 veröffentlicht und es besteht immer noch Bedarf an Forschung bezüglich seiner Eignung für die automatische Termextraktion, insbesondere für andere Sprachen als Englisch. Aus diesem Grund können die Ergebnisse der vorliegenden Arbeit auch für Forscher:innen nützlich sein, die sich mit diesem neuronalen generativen Sprachmodell bzw. mit LLMs allgemein befassen und wissen wollen, was der aktuelle Entwicklungsstand ist, bzw. welche Ergebnisse aktuell in dem untersuchten Bereich erzielt werden können.

1.3. Detaillierter Aufbau der Arbeit

Der theoretische Teil der Arbeit befasst sich in erster Linie mit der Terminologie und soll einen Überblick über die Grundelemente der Terminologielehre verschaffen (2.1.). Im zweiten Kapitel werden die Anfänge dieser wissenschaftlichen Disziplin umrissen, die wichtigsten Konzepte dargestellt und es wird die Sicht von mehreren Autor:innen präsentiert, verglichen und analysiert. Den Benennungen gilt in dem Kapitel besonderes Augenmerk, da sie das Fundament der Untersuchung bilden. In weiterer Folge werden die Grundprinzipien der Terminologearbeit geschildert (2.2.). Das Kapitel umfasst ihre verschiedenen Formen sowie die beste Praxis bei ihrer Einführung im Unternehmen. Das weitere Kapitel (2.3.) und die Unterkapitel widmen sich den Formen der Terminologiegewinnung und den Herausforderungen der automatischen Termextraktion. Im letzten Schritt wird eine Auswahl an Termextraktionssystemen besprochen.

Den Gegenstand des dritten Kapitels bilden Verfahren der automatischen Termextraktion. Beginnend mit dem Ablauf der Termextraktion, verschafft das Kapitel eine Übersicht sowohl über traditionelle als auch moderne Verfahren und gewährt somit Einblick in die Entwicklung dieser relativ neuen Disziplin. Konkret handelt es sich um linguistische (3.1.1.), statistische (3.1.2.), hybride (3.1.3.) und neuronale (3.1.4.) Ansätze. Den letzteren wird besondere

Beachtung geschenkt, da sie die Untersuchung der vorliegenden Arbeit betreffen. Im Kapitel werden namentlich mehrere Systeme vorgestellt, die auf den o.g. Ansätzen basieren. Im Unterkapitel 3.1.4. werden außerdem Verfahren des überwachten und unüberwachten maschinellen Lernens sowie Deep-Learning kurz umrissen, um die Grundlagen der neuronalen Ansätze besser nachvollziehen zu können. Das Unterkapitel verschafft außerdem einen Überblick über die zeitliche Entwicklung des Verfahrens nach dem aktuellen Stand der Forschung samt ihrer Architektur bzw. Funktionsweise und endet mit dem Vergleich der Transformer-basierten Sprachmodelle Bidirectional Encoder Representations from Transformers (BERT) und Generative Pretrained Transformer (GPT) sowie der Vorstellung der für die Untersuchung ausgewählten Systeme Text2TCS und ChatGPT-4o (3.1.4.1.-3.1.4.2.). Zu dem dritten Kapitel gehört außerdem ein Unterkapitel über Prompt Engineering (3.2.), das die neuesten Entwicklungen in diesem Bereich anspricht.

Im Vordergrund des vierten Kapitels stehen die Evaluierungsmaßnahmen der Termextraktion. Es werden die Evaluierungsmethoden und der Anwendungsbereich besprochen, die für die Untersuchung wichtigsten Metriken: Genauigkeit, Trefferquote und F1-Score sowie ihre Berechnungsmethoden erläutert und andere relevante Begriffe angesprochen wie Exaktheit und Fehler, Goldstandard, Black-Box-Evaluierung, Glass-Box-Evaluierung, Spezifität, Fallout u.a. Das Kapitel gewährt einen Einblick darin, wie die Leistung der untersuchten Tools zu bewerten ist und woher die für die Evaluierung verwendeten Metriken kommen.

Das fünfte Kapitel behandelt Fachsprachen. Es werden verschiedene Definitionen bzw. Definitionsversuche von Fachsprache aus der Sicht mehrerer Autor:innen präsentiert (5.1.2.) und die Begriffe Fach, Sprache sowie Fachwort erläutert. Am Anfang werden unterschiedliche Kommunikationsmodelle der allgemeinen, fachlichen und beruflichen Kommunikation samt ihren Bestandteilen besprochen. Besonderes Augenmerk gilt dem systemlinguistischen Inventarmodell, dem pragmalinguistischen Kontextmodell sowie dem kognitionslinguistischen Funktionsmodell. Im nächsten Schritt werden funktionale Eigenschaften von Fachsprachen zusammen mit ihren Merkmalen, insbesondere Wortschatz und Grammatik betreffend, umrissen (5.1.3.-5.1.5.). Mit anderen Worten wird ein Versuch unternommen, das zu beschreiben, was Fachsprachen ausmacht und wie sie sich von der Gemeinsprache unterscheiden. Der zweite Teil des fünften Kapitels widmet sich gezielt der Rechtssprache, wie sie sich bestimmen lässt, was ihre Eigenschaften sind und welche lexikalischen und grammatischen Merkmale sowohl die österreichische als auch die polnische Rechtssprache aufweisen (5.2.2.1.-

5.2.2.2.). Außerdem wird das Thema der geschlechtsneutralen Sprache bezüglich der Rechts-sprache angesprochen.

Im sechsten und gleichzeitig letzten Theoriekapitel wird das Urheberrecht in Österreich und Polen thematisiert. Es wird die geschichtliche Entwicklung kurz umrissen, seine Grundprinzipien werden besprochen und die wichtigsten Begriffe wie Urheber, Werk und geistige Schöpfung erläutert. Weiters werden die urheberrechtlichen Bestimmungen in Bezug auf die Translation und künstliche Intelligenz angesprochen.

Das siebte Kapitel umfasst den aktuellen Stand der Forschung. Es werden verschiedene für die vorliegende Arbeit relevanten Forschungsarbeiten dargestellt, die einen Überblick über das konkrete Forschungsgebiet verschaffen, in dem die Untersuchung eingebettet ist.

Im achten Kapitel wird die Methodik beschrieben, die System- und Textauswahl, Vorgehensweise der Untersuchung sowie die Auswertungsmethode und Validierung einschließt. Anschließend werden Ergebnisse analysiert (9.), interpretiert (10.) und Schlüsse aus ihnen gezogen (11.).

2. Terminologie

Der Schlüssel zum Verständnis vieler Fachgebiete liegt in der Kenntnis der zugehörigen Terminologie. (Witschel 2005: 659)

Das oben angeführte Zitat illustriert das Fundament der vorliegenden Arbeit. Sie betrifft nämlich die Rechtssprache und basiert auf einer zweisprachigen, terminologischen Untersuchung. Um das Thema zu verdeutlichen, wird in erster Linie auf die Grundlagen der Terminologielehre und auf die Grundprinzipien der Terminologiarbeit eingegangen. Im Hinblick auf die Forschungsfrage wird außerdem besonderes Augenmerk auf die Terminologiegewinnung gelegt. Der Zweck der folgenden theoretischen Auseinandersetzung ist das terminologische Rüstzeug für den praktischen Teil der Arbeit zu schaffen.

2.1. Grundlagen der Terminologielehre

In erster Linie soll darauf hingewiesen werden, dass Terminologielehre bzw. Terminologiewissenschaft nicht mit Terminologie gleichzusetzen ist, obwohl dies in der Praxis fälschlicherweise häufig vorkommt (vgl. Drewer & Schmitz 2017: 6). Gemäß DIN 2342:2004-09 (Herzog & Mühlbauer 2007: 170) werden die beiden Benennungen als Synonyme abgelehnt. Der Rat für Deutschsprachige Terminologie (RaDT) (2015: 4) definiert Terminologie als „Fachwortschatz eines bestimmten Fachgebiets (Gesamtheit der Begriffe und ihre Benennungen = Termini, Fachwörter)“. DIN 2342:2004-09 (Herzog & Mühlbauer 2007: 172) schlägt eine ähnliche Definition der Terminologie vor und nennt sie „Gesamtbestand der Begriffe und ihrer Bezeichnungen in einem Fachgebiet“. Eugen Wüster, der als Begründer der Terminologielehre gilt (Hohnhold 1990: 21), schreibt von „einem Wortschatz mit den zugeordneten Bedeutungen“ (Wüster 1974: 62). Unter Berücksichtigung dieser Definitionen ist Terminologie als eine Liste bzw. Sammlung von Fachwörtern zu verstehen, die einem Fachgebiet zugeordnet werden können, wie z. B. *Anatomie* oder *Chiffrierwesen*. Heid (2001: 187) weist darauf hin, dass über längere Zeit ausschließlich Substantive als Termini galten, aber inzwischen hat man auch die terminologische Relevanz anderer Wortarten wie Verben, Adjektiven, Adverbien sowie Kollokationen anerkannt. Die Bestandteile der Fachwörter sind Begriffe und ihre Benennungen, auf die in weiterer Folge genauer eingegangen wird.

Felber und Budin (1989: 1) nennen Terminologielehre eine „Wissenschaft, die sich mit der Erforschung der Grundlagen der Terminologie, d.h. mit den Begriffen, Begriffszeichen und ihren Systemen befasst“. Die Autoren unterscheiden zwischen der allgemeinen Terminologielehre, die den Fokus auf allgemeine Grundsätze der Terminologie und deren Methoden

richtet und der speziellen Terminologielehre, die sich mit den einzelnen Fachgebieten bzw. Fachsprachen beschäftigt (vgl. Felber & Budin 1989: 1). Wüster (1974:62) hebt hervor, dass Terminologielehre, wie bereits vorhin angedeutet, nicht als Terminologie zu bezeichnen ist, damit Zweideutigkeit vermieden wird und definiert sie als „die Lehre von der Terminologie eines bestimmten Fachgebietes in einer bestimmten Sprache“. Auch Drewer und Schmitz (2017: 6) sind der Ansicht, dass gerade die Disziplin, die sich damit beschäftigt, dass gleiche Sachen gleich und andere anders bezeichnet werden, Maßstäbe setzt und Eindeutigkeit fördert. Die Autor:innen verweisen auf die Definition der Terminologielehre von DIN 2342:2004-09 (Herzog & Mühlbauer 2007: 172), die sie als „Wissenschaft von den Begriffen und ihren Bezeichnungen in den Fachsprachen“ beschreibt. Für Arntz et al. (2014: 5-8) ist Terminologielehre transdisziplinär und sie nennen ihre enge Verbundenheit zu mehreren Disziplinen wie etwa zu Sprachwissenschaft, Sachwissenschaften, Philosophie, Wissenstheorie, Informatik u.a. Die Autoren betonen außerdem (2014: 3), dass Terminologearbeit dermaßen kostspielig ist, dass ihr gute Terminologielehre mit einheitlichen Grundsätzen zugrunde liegen soll.

In den oben angeführten Definitionen wurden Begriff und Benennung erwähnt, die zusammen mit Gegenstand das Augenmerk der Terminologielehre bilden. Im Fokus der vorliegenden Arbeit steht zwar die Benennung, aber sie lässt sich erst dann vollständig nachvollziehen, wenn man die Beziehung zwischen ihr und den weiteren beiden Elementen versteht. Die Idee der drei Grundelemente, die zusammen das semiotische Dreieck bilden, stammt von Ogden und Richards, die es 1923 im sprachwissenschaftlichen Diskurs initiiert haben. Abbildung 1 und Abbildung 2 veranschaulichen diese Idee.

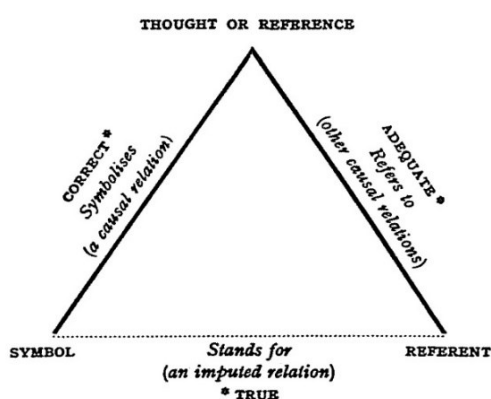


Abbildung 1: Das semiotische Dreieck (Ogden & Richards 1923: 11)

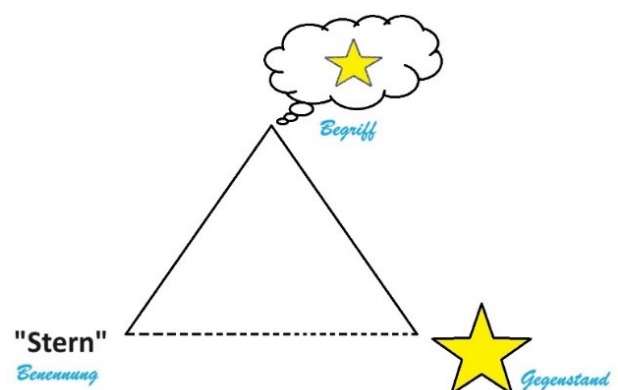


Abbildung 2: Ein adaptiertes semiotisches Dreieck (adaptiert von Ogden & Richards 1923: 11)

Die Terminologielehre verwendet das semiotische Dreieck, um den Begriff und die Benennung zu erklären und den Zusammenhang zwischen ihnen darzustellen (vgl. Drewer & Schmitz 2017: 6). Die Autor:innen (2017: 6-7) loben zwar das Modell und betonen sein Nutzen bei der Veranschaulichung der Grundelemente der Terminologielehre, aber gleichzeitig weisen sie darauf hin, dass es inzwischen aufgrund seiner einfachen Form kritisiert wurde, da sich damit die Komplexität der Beziehungen zwischen den Grundelementen nicht ausreichend repräsentieren lässt. Arntz et al. (2014: 41-42) führen an, dass aus diesem Grund andere, komplexere Modelle entstanden sind, die die Vielschichtigkeit der Zusammenhänge besser zum Ausdruck bringen.

Zu den drei Grundelementen des semiotischen Dreiecks gehören Gegenstände. Die Menschen werden von Gegenständen umgeben (vgl. Wüster 1991: 7), denn sie bilden einen Ausschnitt aus der realen Welt. Sie sind der „Ausgangspunkt der Terminologielehre“ (Arntz et al. 2014: 45). Ein Gegenstand ist im Sinne des semiotischen Dreiecks nicht nur ein materielles Objekt, z. B. *Stern*, wie er in der Gemeinsprache üblicherweise zu verstehen ist, sondern auch ein immaterielles bzw. abstraktes Objekt „ohne physische Form“ wie *Lichtgeschwindigkeit* oder ein Vorgang/Sachverhalt wie *Sternentfernung von der Erde messen* (Arntz et al. 2014: 46-47; Drewer & Schmitz 2017: 7). Ein Gegenstand umfasst eine Reihe von Eigenschaften, aus denen er konstituiert ist (vgl. Felber & Budin 1989: 1). Genauso wie mehrere Eigenschaften einen Gegenstand ergeben, bilden mehrere Gegenstände einen Begriff, also seine gedankliche Repräsentation. Begriff wird von DIN 2342:2004-09 folgendermaßen definiert: „Denkeinheit, die aus einer Menge von Gegenständen unter Ermittlung der diesen Gegenständen gemeinsamen Eigenschaften mittels Abstraktion gebildet wird.“ (Herzog & Mühlbauer 2007: 172) Daraus lässt sich herleiten, dass die gedankliche Vorstellung eines Sterns (Begriff) aus einer Reihe einzelner Sterne (Gegenstand) und für sie spezifischen Eigenschaften bzw., im Sinne der Begriffsbildung, Merkmalen entsteht. Hohnhold (1990: 44) macht außerdem darauf aufmerksam, dass Begriffe sprachunabhängig sind, da sie gedacht werden. Allerdings können durchaus Unterschiede über Sprachen und Kulturen auf der begrifflichen Ebene existieren. Begriffe werden grundsätzlich in Individual- und Allgemeinbegriffe unterteilt. Unter Individualbegriffen sind solche Begriffe zu verstehen, die sich „auf einen einzelnen Gegenstand, der durch einen konkreten Raum- und Zeitbezug definiert ist“ (Drewer & Schmitz 2017: 8), beziehen. Allgemeinbegriff ist wiederum ein Begriff, der „die Eigenschaften einer Gruppe von Gegenständen zusammenfasst“ (2017: 8). Anders ausgedrückt lassen sich die *Sonne* und der *Stern* entsprechend als Individualbegriff und Allgemeinbegriff definieren. Begriffe können miteinander korrelieren. Um dies zu veranschaulichen, werden grafische Begriffssysteme verwendet. Da die vorliegende Masterarbeit sich weder mit Begriffsbeziehungen noch Begriffssystemen befasst, soll

nicht grundlegend auf das Thema eingegangen, sondern lediglich erwähnt werden, dass Begriffe entweder in einer hierarchischen bzw. nicht-hierarchischen Beziehung zueinanderstehen. Weiters wird bei den hierarchischen Beziehungen zwischen Bestands- und Abstraktionsbeziehungen unterschieden (vgl. Drewer & Schmitz 2017: 9). Arntz et al. (2014: 80-81) definieren die Abstraktionsbeziehung als eine Beziehung, in der ein Unterbegriff „auf der gleichen Stufe“ wie Oberbegriff ist und zumindest über ein Merkmal verfügt, das ihn vom Oberbegriff unterscheidet. Bei der Bestandsbeziehung handelt es sich um einen übergeordneten Begriff, auch Verbandsbegriff genannt, der sich in seine Teilbegriffe zergliedern lässt (vgl. Arntz et al. 2014: 92; Drewer & Schmitz 2017: 9). Unter Berücksichtigung dieser Definitionen lassen sich die oben genannten Begriffsbeziehungen mit folgenden Beispielen veranschaulichen: Stern ist eine Art vom Himmelskörper (Abstraktionsbeziehung); die Sonne ist ein Bestandteil vom Sonnensystem (Bestandsbeziehung).

Wie bereits erwähnt, stehen die Benennungen im Fokus dieser Arbeit, deswegen wird ihre theoretische Grundlage ausführlicher behandelt. Benennung ist eine sprachliche Repräsentation des Begriffs. DIN 2342:2004-09 präzisiert diese Aussage und definiert Benennung folgendermaßen: „sprachliche Bezeichnung eines Allgemeinbegriffs aus einem Fachgebiet“ (Herzog & Mühlbauer 2007: 166). Der Ausdruck sprachliche Bezeichnung lässt vermuten, dass es auch eine nicht sprachliche Bezeichnung geben muss. Zu den Repräsentationen der Begriffe, die nicht sprachlich ausgedrückt werden, gehören u.a. Symbole, Formeln, Piktogramme, Emoticons (vgl. Drewer & Schmitz 2017: 14). Bezeichnung ist dementsprechend ein Oberbegriff. Benennungen von Individualbegriffen heißen Namen. Es wird grundsätzlich zwischen Ein- und Mehrwortbenennungen unterschieden. DIN 2342:2004-09 definiert die Einwortbenennung als „Benennung, die aus einem einzelnen Wort besteht.“ (Herzog & Mühlbauer 2007: 167) Unter einem einzelnen Wort ist in erster Linie ein elementares Wort wie z. B. *Stern* zu verstehen. Es existieren allerdings auch komplexe Wörter, die „zwei oder mehr bedeutungstragende Einheiten“ umfassen wie z. B. *Sternschnuppenfall*. Gemeint sind Wortzusammensetzungen, die im Deutschen oft gebildet werden. Sie können aus zwei Substantiven (*Zahnrad*), einem Verb und Substantiv (*Fräsmaschine*), einem Adjektiv und Substantiv (*Edelmetall*), einem Adjektiv und Verb (*hartlöten*) sowie aus einer Präposition und einem Substantiv (*Überstrom*) bestehen (Arntz et al. 2014: 120). Im Englischen, Spanischen, und Französischen kommen auch zusammengesetzte Benennungen vor, aber seltener als im Deutschen. Arntz et al. (2014: 120) führen an, dass in diesen Sprachen die Mehrwortbenennungen bei der Benennungsbildung am weitesten verbreitet sind. In polnischen juristischen Texten überwiegen in dem Zusammenhang Par-

tizipsubstantivierungen wie z. B. *uprawniony, obciężony*. Wortzusammensetzungen treten allerdings auch auf (Iluk 1993: 162–164). Einwortbenennungen können auch mit einem Bindestrich geschrieben werden z. B. *Zweipunkt-Beckengurt* (Herzog & Mühlbauer 2007: 167). Die Mehrwortbenennung ist die „Benennung, die aus mindestens zwei durch Leerstellen getrennten Wörtern besteht“ (2007: 167). Diese Wörter können ebenfalls elementar oder komplex sein, z. B. *die Europäische Weltraumorganisation* oder *Drehmomentwandler mit Wandelüberbrückung* (KÜDES 2018: 21). Felber und Budin (1989: 121) heben hervor, dass zwischen einer „Wortgruppe“ (Mehrwortbenennung) und Aufzählung einzelner Termini zu differenzieren ist, denn „eine solche Aufzählung stellt eine oft unvollständige Umfangsbestimmung eines Begriffes zur Beschreibung eines Themas dar“ (1989: 121). Mehrwortbenennungen im Deutschen bestehen meistens aus mehreren Substantiven oder Verbindungen von Adjektiv und Substantiv (vgl. Drewer & Schmitz 2017: 15; Elsen 2017: 150). Vom besonderen Interesse für die vorliegende Arbeit sind Mehrwortbenennungen in juristischen Texten. Elsen (2007: 48–49) analysiert Texte des Baurechts und sucht dabei nach Hinweisen zur Erkennung der „Wortgruppenlexeme“ (Mehrwortbenennungen). Die Autorin (2007: 48–49) betont das häufige Vorkommen von Wortgruppenlexemen in den Texten des Baurechts, die u.a. daran erkannt werden, dass alle Wörter großgeschrieben werden (*Nichtoffenes Verfahren*); sie Namen von Normen und Gesetzen sind (*Gesetz zur Modernisierung des Schuldrechts*); oder sie Kompositum als Dublette haben (*Schuldrechtsmodernisierungsgesetz*). Weiters führt Elsen (2017: 153) an, dass der hohe Bedarf an „exakten, klassifizierenden oder spezialisierenden Benennungen“ in den Fachsprachen nicht durch Wortzusammensetzungen, im Sinne der Einwortbenennungen, in ausreichender Weise gedeckt werden kann. KÜDES (2018: 21) nennt bezüglich der Differenzierung der Benennungen außer Ein- und Mehrwortbenennungen auch Kurzformen: „Abkürzungen wie z. B., Initialkurzwörter bzw. Akronyme wie *LKW* und Silbenkurzwörter wie *Kripo*.“ (2018: 21)

An Benennungen werden gewisse Anforderungen gestellt, die zu beachten sind, wenn es zur Benennungsbildung kommt. Laut Felber und Budin (1989: 122–123) sollen sie „sprachlich richtig, treffend, knapp, leicht ableitbar und eineindeutig“ sein. Eine Benennung befolgt dementsprechend idealerweise die Regeln der jeweiligen Sprache, bildet den Begriffsinhalt ab, ist nicht unnötig lang, lässt Ableitungen zu und ist nur einem Begriff zugeordnet. Arntz et al. (2014: 115–116) fügen noch Neutralität und Bevorzugung der deutschen Sprache hinzu, gleichzeitig führen sie allerdings an, dass es praktisch nicht realisierbar ist, alle oben genannten Anforderungen zu erfüllen, insbesondere, dass eine Benennung sowohl knapp als auch treffend bzw. motiviert ist. Auch die Eineindeutigkeit ist schwer zu erzielen. Felber und Budin (1989:

123) vertreten ebenfalls diese Ansicht und sprechen im Fall der nicht Erzielung der Eineindeutigkeit stattdessen von Eindeutigkeit, also der „Zuordnung Begriff – Benennung innerhalb eines Fachgebiets“. Drewer und Schmitz (2017: 15) sind der Ansicht, dass die (Ein)eindeutigkeit trotzdem angestrebt werden soll, mindestens innerhalb eines Unternehmens. Die Autor:innen nennen zwei Problembereiche, die beim Erzielen dieser Anforderung an die Beziehung zwischen Begriff und Benennung auftreten können, nämlich den Fall von Synonymie und Ambiguität. Synonymie tritt dann auf, wenn zwei oder mehrere Benennungen einem demselben Begriff zugewiesen werden. Dies betrifft z. B. Benennungen, die Kurzformen haben, wie etwa *EU* und die *Europäische Union*. Bekannt ist auch der Fall der Gebrauchssynonymie, die strikte gesehen zwar keine echte Synonymie darstellt, aber in der Praxis oft verwendet wird z. B., wenn anstatt vom *Vorschlaghammer* lediglich vom *Hammer* gesprochen wird und das richtige Werkzeug sich aus dem Kontext erschließen lässt (vgl. 2017: 15). Um zu verhindern, dass zwei oder mehrere Synonyme in einer Firma parallel existieren, wird empfohlen, innerhalb des Unternehmens Vorzugsbenennungen im Terminologiebestand einzuführen. DIN definiert sie folgendermaßen: „Benennung, die aufgrund objektivierbarer Kriterien als die Benennung ausgewählt wurde, die gegenüber Synonymen bevorzugt verwendet wird.“ (Herzog & Mühlbauer 2007: 167) Unter Ambiguität in der Terminologielehre ist Mehrdeutigkeit auf der Wortebene zu verstehen, die dann vorkommt, wenn eine Benennung mehrere Begriffe repräsentiert (vgl. Drewer & Schmitz 2017: 17). Zur Ambiguität zählen Homonyme und Polyseme. Der Unterschied besteht in der Verbindung zwischen zwei oder mehreren Begriffen, die durch eine einzige Benennung bezeichnet werden. Ist kein Zusammenhang zwischen ihnen bezüglich des Begriffsinhalts festzustellen, spricht man von Homonymie, wie etwa bei *Tau*, *Schloss* oder *Leiter*. Besteht eine Verbindung zwischen solchen Begriffen, z. B. bezüglich der ähnlichen Form, handelt es sich um Polysemie, wie etwa bei *Maus*, *Pferd* oder *Wurzel* (vgl. Drewer & Schmitz 2017: 18). Da in den Fachsprachen, z. B. im Zuge der Terminologisierung, bei der Benennungsbildung auf den Allgemeinwortschatz zugegriffen wird, betreffen die meisten Ambiguitäten die Polysemie. Homonymie kommt nur durch Zufall vor (vgl. 2017: 18).

Nachdem die Anforderungen an Benennungen geschildert worden sind, soll nun auf ihre Bildung eingegangen werden. Wortzusammensetzungen, Mehrwortbenennungen und Kürzungen wurden bereits erwähnt. Eine weitere Möglichkeit stellt die Terminologisierung dar, ein Verfahren, bei welchem einem Wort bzw. einer Wortgruppe aus dem Allgemeinwortschatz oder einem anderen Fachgebiet in der Fachsprache eine bestimmte Bedeutung zugewiesen wird und bestimmte Bedeutungsstrukturen übernommen werden. Als Beispiel gelten menschliche Körperteile, nach denen technische, mechanische und wissenschaftliche Teile bzw. Elemente

genannt werden wie z. B. *Bohrkopf*, *Elektronenauge* und *Drehzahnkompressor* (vgl. Arntz et al. 2014: 119). Weiters sind Wortableitungen möglich. Dabei handelt es sich um ein Stammwort, das zusammen mit einem Ableitungselement wie Präfix und/oder Suffix ein neues Wort bildet. Die Wortableitung, z. B. *Prüfung*, gehört zu sehr produktiven Wortbildungselementen in der deutschen Sprache (vgl. Arntz et al. 2014: 121). Zu der nächsten Benennungsbildungsmethode zählt die Konversion. Darunter ist der Wortklassenwechsel eines Wortes ohne Wortbildungselemente zu verstehen, z. B. *schlafen* → *Schlaf*. Außerdem kommt es vorwiegend in den Naturwissenschaften und in der Technik zu Entlehnungen, insbesondere bei Neuerfindungen. Das fremdsprachige Wort wird in dem Fall unverändert übernommen z. B. *Software*, *Internet*. Bei einer Lehnübersetzung übersetzt man einzelne Worтеlemente des fremdsprachigen Wortes z. B. *steam engine* → *Dampfmaschine*. Laut Arntz et al. (2014: 123-124) begünstigen Lehnübersetzungen die Entstehung von Dubletten, insbesondere in der deutschen Medizinsprache z. B. *Suizid* und *Selbstmord*. Zuletzt wird Neubildung als ein zwar mögliches, aber äußerst selten vorkommendes Verfahren zur Benennungsbildung genannt (vgl. 2014: 123-124).

2.2. Grundprinzipien der Terminologearbeit

Da diese Arbeit eine terminologische Untersuchung im Sinne des Vergleichs der Ergebnisse der manuellen und automatischen Extraktion bezweckt, wird keine Terminologiedatenbank erstellt, sondern lediglich Listen mit Termini. Eine Übersicht verschiedener Methoden der Terminologearbeit samt eventuellen Problemstellen und ein kurzer Umriss der zentralen Arbeitsschritte der Terminologearbeit sollen genügen, um einen ausreichenden Einblick in das Thema zu gewähren.

Die Terminologielehre und ihre Grundlagen bilden die theoretische Basis für die Terminologearbeit, die sich lt. DIN 2342:2004-09 mit der „Erarbeitung, Bearbeitung oder Verarbeitung, Darstellung oder Verbreitung von Terminologie“ befasst (Herzog & Mühlbauer 2007: 170). Die Norm sieht außerdem vor, dass unter Terminologearbeit auch Terminologiegewinnung und Einarbeitung der Termini in Texte zu verstehen ist. Weiters wird auch die Terminologieverwaltung als Teil der Terminologearbeit definiert, der „sich mit der Erfassung, Verbreitung, Pflege und Bereitstellung von terminologischen Daten befasst.“ (2007: 170) Zugleich wird auch die terminologische Normung zu Aufgaben der Terminologearbeit gezählt (vgl. KÜDES 2018: 16; Budin 1999:1), wenngleich DIN 2342:2004-09 dies nicht explizit als Teil der Terminologearbeit nennt wie eben bei der Terminologieverwaltung. Drewer und Schmitz

(2017: 23–24) nennen ein Beispiel für die Abweichung von der Norm, die sich in der unternehmerischen Praxis etabliert hat. Es handelt sich dabei um die Benennung *Terminologiemanagement*, die als Synonym zur Terminologearbeit verwendet wird, obwohl sie lediglich einen Teil davon umfasst und nicht als Synonym im wahrsten Sinne des Wortes zu verstehen ist. Die Terminolog:innen sammeln, überprüfen und halten Termini eines konkreten Fachbereichs fest, außerdem werden die Termini bei Bedarf, z. B. in innovativen Fachbereichen, von ihnen festgelegt bzw. neu geprägt. Mehrsprachige Terminologearbeit wird übersetzungsorientierte Terminologearbeit und Prägung neuer Termini terminologische Festlegung genannt (Budin 1999: 1). Das Produkt der Terminologearbeit, die Fachterminologie, ist die Basis für mehrere Disziplinen, u.a. für Sprachmittlung, Technologietransfer, Wissensverarbeitung und Formulierung von Anleitungen (vgl. Felber & Budin 1989: 207). Sie wird in Terminologiedatenbanken gepflegt und verwaltet bzw. in Form von Fachwörterbüchern, Glossaren oder Fachwortlisten bereitgestellt (vgl. Budin 1999: 1).

Die Methoden der Terminologearbeit variieren je nachdem in welchem Kontext und wofür die Terminologie erschaffen wird. Deskriptiv wird dann gearbeitet, wenn das Ziel die Bestimmung des terminologischen IST-Zustands in einem bestimmten Fachgebiet ist. Es wird dabei in den vorhandenen Terminologiebestand nicht eingegriffen, sondern er wird lediglich aufgezeichnet. Der Nutzen dieser Methode ist entweder im lexikografischen Kontext z. B. bei der Entstehung der Fachwörterbücher oder in der unternehmerischen Praxis als Vorlage für die präskriptive Terminologearbeit verankert (vgl. Drewer & Schmitz 2017: 24; Arntz et al. 2014: 218). Nach der Ermittlung und Aufzeichnung des im Zuge der deskriptiven Terminologearbeit gesammelten Terminologiebestands kann präskriptiv bzw. normend gearbeitet werden. Es werden dabei Vorzugsbenennungen bestimmt, um die Verwendung synonyme und gleichzeitig unerwünschter Benennungen zu vermeiden. Dies fördert Konsistenz und Verständlichkeit und beugt eventuellen Missverständnissen innerhalb eines Fachgebiets oder auf lokaler Ebene z. B. unternehmens- bzw. institutionsweit vor. Gleichzeitig ist es für das firmeneigene Branding, Corporate Language und schließlich Kundenzufriedenheit von Bedeutung (vgl. Drewer & Schmitz 2017: 24-25). Die Rolle der präskriptiven Terminologearbeit kann am Beispiel der kollaborativen Translation veranschaulicht werden. Bei größeren Übersetzungsprojekten müssen Unternehmen nämlich ggf. auf mehrere Übersetzer:innen zugreifen. Die Vorzugsbenennungen tragen in dem Fall zur Konsistenz und somit Qualität der Übersetzung bei und können

auch Übersetzungskosten reduzieren. Normende Terminologearbeit wird z. B. von Terminolog:innen in Industrieunternehmen betrieben. Oft bilden sich dort Gremien bzw. Arbeitsgruppen, die aus Terminolog:innen, Übersetzer:innen, Manager:innen und Fachexpert:innen bestehen. Der Zweck einer solchen Arbeitsgruppe ist über verbotene und bevorzugte Benennungen zu entscheiden. Arntz et al. (2014: 218) nennen die normende Terminologearbeit auf regionaler, nationaler und internationaler Ebene durch Normungsgremien Terminologienormung. Zu solchen Gremien zählen das Deutsche Institut für Normung e.V. (DIN), der Verein deutscher Ingenieure e.V. (VDI), die Internationale Organisation für Normung (ISO) oder das Europäische Komitee für Normung (CEN) (2014: 276ff.). Die Autoren (2014: 269) weisen darauf hin, dass dabei nicht die Einschränkung der sprachlichen Vielfalt der Sprachbenutzer:innen bezweckt wird, sondern die Förderung der eindeutigen Verständigung zwischen Fachexpert:innen im bestimmten Fachgebiet. Um die Verwendung der bevorzugten Terminologie in Texten und Übersetzungen sicherzustellen, können Authoring-Memory-Systeme respektive Translation-Memory-Systeme eingesetzt werden. Außerdem kann menschliches oder maschinelles Lektorat erfolgen (vgl. Drewer & Schmitz 2017: 195).

Bei der Wahl der Methode der Terminologearbeit zählt außerdem ihr beabsichtigter Umfang und der zukünftige Gebrauchskontext. Muss das terminologische Problem spontan gelöst werden, wird zur punktuellen Terminologearbeit (Ad-hoc-Terminologearbeit) gegriffen. Dies kommt z. B. dann vor, wenn die gesuchte Benennung nicht in Wörterbüchern zu finden ist, es sich um Neologismen handelt oder es sich um offizielle Bezeichnungen (vgl. KÜDES 2018: 66). Die Ad-hoc-Terminologearbeit erfordert einen hohen Aufwand, denn das terminologische Problem muss vor der Recherche genau überlegt und analysiert werden, damit sie gezielt durchgeführt werden kann. Gebraucht wird einschlägige Fachdokumentation und/oder Fachexpert:innen, die beim Identifizieren bzw. Bestätigen der richtigen Benennung helfen. Der ganze Aufwand wird für eine einzige Benennung betrieben und zugleich birgt dieses Verfahren ein hohes Fehlriskio, denn der Begriff wird isoliert recherchiert, ohne dass das Fachgebiet, aus dem er stammt, analysiert wird. Beispielsweise ist es schwer im Falle der mehrsprachigen Terminologearbeit eine richtige Benennung in der Zielsprache aus dem Bereich des Schulwesens zu finden, wenn man mit dem Schulsystem selbst in der Ausgangs- und Zielsprache nicht vertraut ist (vgl. KÜDES 2018: 67). Sollte ein längerer, thematisch einheitlicher Text aus einem bestimmten Fachgebiet übersetzt werden und die Zeit es erlaubt, bietet sich die Angelegenheit zur vertieften, textbezogenen Terminologearbeit. Dies gilt als Vorbereitung für einen größeren

Übersetzungsauftrag und setzt die vollständige Erarbeitung der Terminologie eines Textes voraus. Als Beispiel gilt ein umfangreicher Text über Anatomie verschiedener Bienenarten. Je weniger Fachgebiete in dem Text enthalten sind, umso sinnvoller und zielgerichteter ist die textbezogene Terminologearbeit (vgl. KÜDES 2018: 68; Drewer & Schmitz 2017: 26). Die zuverlässigste Terminologearbeit erfolgt dann, wenn sie systematisch bzw. sachgebietsbezogen durchgeführt wird. In dem Fall wird die Terminologie des gesamten Gebiets erschlossen. Drewer und Schmitz (2017: 26) schreiben von einem (Teil-)Fachgebiet und KÜDES (2018: 69) von einem Sachgebiet, wie z. B. Urheberrecht, denn das Fachgebiet, z. B. Recht, wäre für die systematische Terminologearbeit zu umfangreich. Die Autor:innen (2017: 26; 2018: 69) argumentieren, dass diese Methode nur ein wenig aufwändiger als die punktuelle Terminologearbeit ist, aber gleichzeitig viel wertvollere Ergebnisse liefert, denn der Begriff wird umfassend erfasst (samt Ober- und Unterbegriffen), die Terminolog:innen müssen sich nur einmal in das Thema/Sachgebiet einarbeiten, die Fachexpert:innen müssen nicht nur für eine einzige Benennung herangezogen werden, was kosteneffektiver ist und die mit dieser Methode erarbeitete Terminologie kann vom weiten Benutzer:innenkreis verwendet werden. Die systematisch erfasste Terminologie eines Sachgebiets erhöht die Konsistenz, Effizienz und vor allem spart Zeit und Geld bei mehreren Übersetzungsaufträgen aus dem gleichen Sachgebiet. Einen interessanten Einblick in das Thema bieten Drewer und Horend (2007). Die Autorinnen beschreiben eine Studie in Kooperation zwischen Forscher:innen, Student:innen und Expert:innen, in der der theoretische und der praktische Zugang zur Terminologearbeit miteinander verglichen werden. Wie bereits oben erwähnt, wird die Ad-hoc-Terminologearbeit in der Wissenschaft zwar als eine valide Methode der Terminologearbeit anerkannt, aber den Vorteilen der systematischen Terminologearbeit doch unterlegen. In der Praxis, mindestens am Beispiel des Unternehmens, das sich an der Studie beteiligte, wird allerdings die punktuelle Terminologearbeit befürwortet. Oft ist die Anfrage nach einer Benennung von Autor:innen, die sie in einem Text verwenden wollen und sie in der Datenbank nicht ausfindig machen, erst der Ansporn für die Terminologearbeit. Aus zeitlichen und ökonomischen Gründen wird nicht a priori terminologisch gearbeitet. Weiters könnten Ausgaben für Terminologie, die nicht mit einem direkten Gewinn korrelieren, den Zuspruch des Unternehmens gegenüber der Terminologearbeit untergraben (vgl. 2007).

Drittens kann die Terminologearbeit ein-, zwei- oder mehrsprachig ausgeführt wer-

den. Je höher die Zahl der erfassten Sprachen ist, desto anspruchsvoller ist die Arbeit, denn die Begriffssysteme in diversen Sprach- und Kulturräumen sich voneinander unterscheiden, weswegen sie zunächst miteinander verglichen und gewissermaßen in Einklang gebracht werden müssen. Nichtsdestotrotz ist das Ergebnis dieser Arbeit, wenn es qualitativ und zugänglich ist, z. B. für die schnelle und gleichzeitig hochwertige Übersetzung aus jedem Fachbereich unerlässlich (vgl. Ahmad et al. 1994: 267). Eine völlige Übereinstimmung zweier Termini aus verschiedenen Sprachräumen ist zwar möglich und in den Fachsprachen im Gegensatz zur Gemeinsprache nicht selten, aber die Äquivalenzproblematik ist trotzdem ein Thema. Äquivalenz bezeichnet das Verhältnis eines Terminus der Zielsprache mit dem korrespondierten Terminus der Ausgangssprache, deren Begriffe (fast) gleich sind. Anders ausgedrückt entspricht Äquivalenz interlingual dem intralingualen Phänomen der Synonymie (vgl. Drewer & Schmitz 2017: 20;26). Außer der völligen Äquivalenz, z. B. *di-Natriumhydrogenphosphat* (DE) und *di-sodium hydrogen phosphate* (EN), tritt die Teiläquivalenz auf, bei der es entweder zu Überschneidungen, d. h. einige Begriffsmerkmale sind identisch andere wiederum anders, wie etwa *Beamter* (DE) und *civil servant* (EN), oder zu Inklusionen, d. h. alle Begriffsmerkmale sind identisch, aber außerdem gibt es in einer der Sprachen noch zusätzliche Merkmale, wie etwa *sozial* (DE) und *social* (FR), kommt (vgl. Arntz et al. 2014: 147–148; Drewer & Schmitz 2017: 21–22). Bei der fehlenden Äquivalenz bzw. Null-Äquivalenz sind sich die Benennungen zweier Sprachen in ihrer Form dermaßen ähnlich, dass automatisch die Ähnlichkeit der Begriffe, die von ihnen in den beiden Sprachen repräsentiert werden, angenommen wird, obwohl dies kaum bzw. nicht der Fall ist. Es handelt sich dabei um die sog. falschen Freunde, wie etwa *sensibel* (DE) und *sensible* (EN). Hofmann (2016: 57) hat die Null-Äquivalenz bei fünf deutsch-englischen Übersetzungen von Hamlet analysiert und danach konstatiert, dass sie selten vorgekommen ist, „da die lexikalischen Systeme im kulturellen und sozio-ökonomischen Bereich im wesentlichen kongruent sind.“ (2016: 57) Zu finden in den Übersetzungen sind allerdings Begriffslücken in Zusammenhang mit Idiomen. Eine Begriffslücke kommt dann vor, wenn ein Begriff im bestimmten Sprachraum nicht vorhanden ist und demzufolge nicht benannt wurde. Es handelt sich dabei um einen Begriff, der in anderen Sprachräumen existiert und bezeichnet wurde. In dem Fall ist eine Benennung zu finden dementsprechend schwer und der Begriff muss in dem Sprachraum eingeführt, erklärt und definiert werden (vgl. Drewer & Schmitz 2017: 27). Die Benennungslücken bzw. terminologische Lücken treten wiederum dann auf, wenn der Begriff im bestimmten Sprachraum bereits angekommen ist, lediglich noch nicht benannt wurde (vgl.

Mayer 2001: 165). Um eine Benennungslücke zu füllen, werden Benennungsvorschläge in Form von Wortneubildungen, Entlehnungen, z. B. *Design* und Lehnübersetzungen, z. B. *machine aided translation* (EN) und *maschinengestützte Übersetzung* (DE), erstellt. Arntz et al. (2014: 162) weisen darauf hin, dass solche Übertragungsprobleme keine Seltenheit darstellen und vorwiegend die Rechtswissenschaft betreffen, da Länder grundsätzlich verschiedene Rechtsordnungen haben. Aus diesem Grund ist Terminologiearbeit im rechtlichen Kontext sowohl eine sprachliche als auch juristische Aufgabe.

Ein Terminologieprojekt besteht aus mehreren Arbeitsschritten. Drewer und Schmitz (2017: 34) nennen folgende Aspekte der praktischen Terminologiearbeit: Zielsetzung und Planung; Gewinnung der Terminologie; begriffliche Systematisierung; sprachliche Bewertung und Bereinigung; Generierung/Schaffung neuer Benennungen; Verwaltung der Terminologie; Darstellung und Verbreitung der Terminologie; Pflege der Terminologie und Kontrolle der Terminologieverwendung. Im ersten Schritt wird das Terminologieprojekt gründlich geplant, es werden Ziele definiert, das Sachgebiet wird klar abgegrenzt, die finanziellen Mittel werden gesichert und die Verantwortlichen bestimmt. Weiters muss die aktuelle Bestandsaufnahme vorgenommen und Terminologierecherche samt Terminologiegewinnung betrieben werden. Die Textdokumentation, die für diese Zwecke herangezogen wird, ist von Fachexpert:innen und zugleich Erstsprachler:innen verfasst und sie stellt den aktuellen Wissenstand innerhalb des Sachgebiets dar (vgl. Mayer 2009: 17). Das Ergebnis dieser Phase sind Benennungslisten, die danach mit begrifflichen Informationen versehen werden. Die Begriffe werden definiert und die Beziehungen zwischen ihnen werden festgestellt. In der vierten Phase beginnt die präskriptive Terminologiearbeit. Die Terminolog:innen bestimmen Vorzugsbenennungen, die Synonyme werden erfasst und entsprechend markiert. Im nächsten Schritt werden ggf. neue Benennungen geschaffen. Dies kommt entweder dann vor, wenn neue relevante Begriffe eingeführt werden oder die bestehenden Benennungen eines Begriffs zu Unklarheiten führen. Unter Verwaltung der Terminologie ist die Arbeit mit einem Terminologieverwaltungssystem zu verstehen. Bei jedem Eintrag müssen Datenkategorien beachtet und entsprechende Prinzipien, z. B. Begriffsorientierung, Benennungsautonomie, eingehalten werden. Die erfassten, bereinigten, ergänzten und verwalteten Terminologiebestände werden dann den Nutzer:innen zur Verfügung gestellt. Sie werden allerdings weiterhin zentral gehalten und es müssen Personen designiert werden, die die Verantwortung für die Beibehaltung der Qualität der Bestände tragen. Dabei ist unbedingt auf die Zugriffsrechte in der Terminologiedatenbank zu achten, wer nur

Leserechte und wer Lese- und Schreibrechte hat. Die Terminologie muss auch fortlaufend gepflegt werden. Einige Benennungen werden wahrscheinlich nach dem Feedback der Nutzer:innen angepasst werden müssen, die Terminologiebestände werden im Regelfall auch erweitert und aktualisiert. Der letzte Schritt ist die Kontrolle, ob die bevorzugten Termini tatsächlich von den Nutzer:innen verwendet werden. Dies kann maschinell und/oder menschlich erfolgen (vgl. Cabré 1999: 49; Drewer & Schmitz 2017: 35–37).

2.3. Terminologiegewinnung

In dem folgenden Unterkapitel wird Terminologiegewinnung, einschließlich Terminologierecherche und Termextraktion thematisiert, die einen zentralen Punkt dieser Arbeit bildet, denn es werden für die Untersuchung automatisch extrahierte Termini einerseits untereinander und andererseits mit den Ergebnissen der manuellen Terminologiegewinnung verglichen. Um das Thema zu veranschaulichen, wird zuerst auf die Theorie der Terminologiegewinnung eingegangen und dann die manuelle und automatische Terminologiegewinnung dargestellt. DIN 2342:2004-09 definiert die Termextraktion als „Teil der Terminologiarbeit, der darin besteht, Termini aus einem Korpus herauszufiltern“ (Herzog & Mühlbauer 2007: 171) und das Korpus als „Zusammenstellung von Texten oder Textelementen aus Ausgangsmaterial für weitergehende Bearbeitungen, z. B. in der Terminologiarbeit“ (Herzog & Mühlbauer 2007: 174). Daraus lässt sich herleiten, dass zuerst Texte der Ausgangssprache gesammelt werden, in denen zwischen den gemeinsprachlichen und fachsprachlichen Ausdrücken unterschieden wird und die letzteren werden anschließend zusammengestellt. Im Regelfall erfindet man die Terminologie nicht neu, sondern sie wird aus verschiedenen (nicht) digitalen Quellen gewonnen, um die firmeneigene Terminologiedatenbank aufzubauen oder zu ergänzen, den Aufwand vor einem Übersetzungsauftrag einzuschätzen oder die terminologische Korrektheit und Konsistenz in der Textdokumentation zu kontrollieren (vgl. Arntz et al. 2014: 244-245). Geschieht die Terminologiegewinnung außer des unternehmerischen und übersetzerischen Kontexts, wird sie beispielsweise für lexikographische Arbeiten bei der Erstellung von Wörterbüchern eingesetzt. Generell wird zwischen explizit und implizit vorhandenen Daten differenziert. Wenn die Terminologie explizit vorliegt, ist sie z. B. in Form von Glossaren, Wortlisten oder Abkürzungsverzeichnissen zu finden (vgl. Drewer & Schmitz 2017: 40). Die Autor:innen weisen darauf hin, dass viele Unternehmen häufig über explizite Terminologielisten verfügen, ohne dass sie als solche wahrgenommen werden. Sie „verstecken sich“ in digitalen Katalogen und Systemen

und müssen zunächst qualitativ aufgearbeitet werden, bevor sie in die Terminologiedatenbank aufgenommen werden können (vgl. 2017: 40). Die implizite Terminologie steht nicht als eine Liste zur Verfügung, sondern sie befindet sich in den fachlichen Textdokumenten wie z. B. Anleitungen, Produktbeschreibungen, Beipackzetteln, Schulungsmaterialien u.a. Aus diesen Ressourcen müssen die Fachtermini zuerst extrahiert werden, bevor sie im weiteren Schritt, ähnlich wie bei der expliziten Terminologie, bewertet, bereinigt und in die Terminologiedatenbank aufgenommen werden können. Je nachdem, ob die Texte lediglich in einer oder in mehreren Sprachen vorliegen, kann die ein-, zwei- oder mehrsprachige Terminologiarbeit erfolgen, wenngleich multilinguale Termextraktion eher eine Seltenheit darstellt (vgl. Arntz et al. 2014: 245). Drewer und Schmitz (2017: 40-41) nennen außerdem eine Mischform von explizit und implizit vorhandenen terminologischen Daten, die in Content-Management-Systemen und Translation-Memory-Systemen enthalten sind. Diese Quellen bieten nämlich neben Benennungen auch längere Textsegmente, aus denen die Termkandidaten erst gewonnen werden müssen. Die Extraktion von Benennungen sowie bei Bedarf/Situation auch Kontextbeispielen, Definitionen und anderen terminologischen Daten kann entweder manuell, häufig mindestens teilweise rechnerunterstützt, oder automatisch erfolgen. Drewer und Schmitz (2017: 41) weisen darauf hin, dass in der Fachliteratur die automatische Terminologiegewinnung üblicherweise Termextraktion genannt wird und die manuelle Terminologierecherche, was in der vorliegenden Arbeit auch übernommen wird.

2.3.1. Manuelle Terminologiegewinnung

Die Terminologierecherche wird idealerweise von Terminolog:innen betrieben, die über das fachliche Know-how im Sachgebiet und das terminologische Wissen verfügen. Sie analysieren Textdokumente und identifizieren die Benennungen samt ggf. anderen relevanten Informationen. Auf diese Art und Weise entsteht eine im Regelfall hochqualitative und aussagekräftige Liste mit Ein- und Mehrwortbenennungen (vgl. Arntz et al. 2014: 245). Die Voraussetzung dafür sind allerdings adäquate Bedingungen wie genügend Zeit und das körperliche sowie geistige Wohlbefinden der Terminolog:innen. Insbesondere der Zeitfaktor wird von mehreren Autor:innen als Nachteil der Terminologierecherche angeprangert, was in weiterer Folge die Kosten für die Terminologiarbeit im Ganzen steigern lässt (vgl. Arntz et al. 2014: 245; Janke 2013: 34–35; Mügge 2007: 714). Drewer und Schmitz (2017: 42) argumentieren allerdings, dass bei diesem Verfahren oft zeitaufwändige Korrekturen ausbleiben, es wird nämlich angenommen,

dass die Terminologie zuverlässig zusammengestellt wurde, was unter Umständen die manuelle Terminologiegewinnung sogar kostengünstiger als die automatische macht. Die Autor:innen weisen außerdem auf weitere Vorteile dieses Verfahrens hin, denn Terminolog:innen eignen sich bei der umfassenden Lektüre gleichzeitig das Fachwissen an, dass sie für die weitere Terminologearbeit benötigen werden z. B. die Erkennung der Begriffsbeziehungen in dem Sachgebiet (vgl. 2017: 42). Andere Quellen (vgl. Heylen & De Hertog 2015: 203) behaupten allerdings, dass manuelle Terminologiegewinnung zukünftig gänzlich durch die automatische ersetzt wird. Janke (2013: 115) nennt einen weiteren Faktor, der das Ergebnis der Terminologierecherche beeinflussen kann, nämlich, wer die Rolle des:der Terminologen:in übernimmt. In der Praxis sind es auch oft technische Redakteur:innen und Fachübersetzer:innen, die die Terminologie recherchieren. Lt. der Autorin kommen sie auf unterschiedliche Ergebnisse, die u.a. davon abhängig sind, was diese verschiedenen Personengruppen schlussendlich als einen validen Termkandidaten einschätzen. Während Terminolog:innen es, lt. der von der Autorin durchgeführten Untersuchung, hauptsächlich präsent haben, was die Benennung ist und somit in die Datenbank aufgenommen werden soll, wollen technische Redakteur:innen in der Datenbank auch gemeinsprachliche Wörter finden und Fachübersetzer:innen sind sich nicht immer dessen bewusst, was tatsächlich eine Benennung ist und wünschen sich, dass auch Phrasen in einer Terminologiedatenbank enthalten sind (vgl. 2013: 115). Zerfuß (2006: 21) betont ebenfalls den subjektiven Charakter der Terminologierecherche und geht darauf ein, wie unterschiedlich die Termkandidatenliste ausfallen kann, je nachdem über welches Fachwissen die Terminologie betreibenden Personen verfügen, und was sie mit der Recherche bezwecken (vgl. auch Eckstein 2009: 109). Reuther (2008: 121) hebt außerdem die Bedeutung der Erfahrung und der Routine des/der Recherchierenden hervor.

2.3.2. Automatische Terminologiegewinnung

Die Termextraktion liegt vor, wenn sie mit automatischen Mitteln betrieben wird. Dabei handelt es sich entweder um die speziell dafür entwickelten Programme, die Termkandidaten in der maschinell lesbaren Fachdokumentation erkennen, oder um durchaus einfachere Textmanipulationsoperationen in gängigen Textverarbeitungs- und Tabellenkalkulationsprogrammen (vgl. Arntz et al. 2014: 245-246). Die Termextraktion soll den Nachteilen der Terminologierecherche entgegenwirken und den zeitaufwändigen und somit kostspieligen Prozess beschleunigen. Die Problematik dieser Annahme wurde bereits vorhin umrissen (2.3.1.). Drewer und Schmitz

(2017: 44) führen an, dass Ergebnisse der Termextraktion von Terminolog:innen überprüft, ggf. bereinigt und angenommen bzw. abgelehnt werden müssen, weswegen auch ihrer Ansicht nach halbautomatisch eine geeignetere Bezeichnung für das Verfahren ist. Die Hauptarbeit der Terminolog:innen beruht bei dieser Form der Terminologiegewinnung darauf, die Benennungen zu bewerten und sie auf der Begriffsebene zusammenzuführen. Eine begriffsorientierte Terminologiedatenbank setzt nämlich voraus, dass synonyme Benennungen, wie etwa *Hubschrauber* und *Helikopter*, Kurzformen, wie *PKW*, und verschiedene Schreibweisen einer Benennung, wie *termbank*, *term bank*, *term-bank* (EN), in einem terminologischen Eintrag festgehalten werden, was sich ihrer Ansicht nach nur schwer vollautomatisieren lässt (vgl. Arntz et al. 2014: 249). Die Autor:innen betonen außerdem, dass die Liste mit Termkandidaten je nach dem verwendeten Verfahren unterschiedlich ausfallen kann (2014: 249). Linguistische, statistische, hybride und neuronale Verfahren der Termextraktion werden detailliert im dritten Kapitel thematisiert. Im Idealfall, allerdings je nach dem eingesetzten Verfahren, werden die Termkandidaten lemmatisiert, also in ihrer Grundform aufgelistet. Dies betrifft den Numerus sowie die Konjugation, Flexion und Deklination der jeweiligen Wortarten. Wie bereits erwähnt erfolgt die automatische Terminologiegewinnung in der Praxis entweder ein- oder zweisprachig (2.2.). Wenn mehrsprachige Termextraktion genannt wird, wird in der Regel ein Ausgangstext mehrmals zweisprachig extrahiert und die Ergebnisse werden zusammengestellt (vgl. Drewer & Schmitz 2017: 45). Das Resultat der einsprachigen Termextraktion ist eine Liste mit Termkandidaten, entweder bereits lemmatisiert oder nicht, die von Terminolog:innen bestätigt oder abgelehnt werden. Sollten sie bei der Überprüfung einer Benennung Zweifel haben, kann mit der Konkordanzsuche automatisch nach dem Kontext/Kotext gesucht werden. Außerdem können damit Kollokationen, Phrasen und syntaktische Muster identifiziert werden (López-Rodríguez & Sánchez-Cárdenas 2021: 182). Budin (1999: 12) betont, dass insbesondere Fachsprachen über viele feststehende Formulierungen verfügen und argumentiert dafür, sie als ganze in die Terminologiedatenbank aufzunehmen. Drewer und Schmitz (2017: 44) führen an, dass gewisse Programme mit Zusatzfunktionen ausgestattet sind und solche Daten erheben können, wie z. B. grammatische Angaben oder die Häufigkeit des Auftretens einer bestimmten Benennung im Text. Für die zweisprachige Termextraktion werden zweisprachige Paralleltexte benötigt. Sie werden segmentiert und die Segmente einander zugeordnet. Dies ist z. B. bei Translation Memorys automatisch der Fall. Ansonsten, wenn lediglich Ausgangs- und Zielsprachige Texte zur Verfü-

gung stehen, muss eine automatische Alignierung vorgenommen werden. Die Texte sollen außerdem ähnliche Textstruktur aufweisen (vgl. Drewer & Schmitz 2017: 45). Nach der Extraktion werden Termkandidaten in einer Sprache zusammengestellt und um jene der zweiten Sprache ergänzt. Die Ergebnisse der zweisprachigen Termextraktion können nicht nur zum Aufbau der zwei- und mehrsprachigen Terminologiedatenbanken verwendet werden, sondern auch z. B. zur Überprüfung terminologischer Konsistenz in den Texten (vgl. 2017: 45).

2.3.2.1. Herausforderungen der automatischen Terminologiegewinnung

Im vorigen Abschnitt wurde erwähnt, dass die automatisch extrahierten Termini von Terminolog:innen überprüft werden müssen, und dass dieser Prozess sehr zeitaufwändig sein kann. Um ihn zu optimieren und somit die Ergebnisse zu präzisieren, werden Stoppwortlisten eingeführt. Es wird dadurch ein Signal an das Programm gesendet, gewisse unerwünschte Wörter bei der Extraktion zu überspringen (vgl. Edo Marzá 2009: 248). Dazu gehören z. B. Präpositionen, Konnektoren, Artikel oder gemeinsprachliche Wörter. Janke (2013: 114) weist in dem Zusammenhang mit Stoppwortlisten auf ein potenzielles Risiko hin. In ihrer Untersuchung wurden im Zuge der Termextraktion bestimmte Termini nicht in die Terminologiedatenbank aufgenommen, da sie aufgrund ihrer gemeinsprachlichen Bedeutung in den standardisierten Stoppwortlisten enthalten waren. In dem untersuchten Text/Kontext wiesen sie allerdings eine fachliche Bedeutung auf und sollten als Fachtermini erkannt werden. Gewisse Programme können die Benennungen mit jenen in der Terminologiedatenbank vergleichen und sie bei der Extraktion ebenfalls überspringen. Nichtsdestotrotz kommt es bei dem Prozess sowohl zu Reduzierungen als auch Erweiterungen. Der Silence-Effekt beschreibt die Situation, bei der die Fachtermini aus dem Text nicht oder nicht gänzlich gewonnen werden. Das Programm erkennt sie nicht als Termkandidaten und meldet sie nicht, verbleibt also „still“, wenn es aber nicht sein soll. Zu beobachten ist ebenfalls das Gegenteil davon, der Noise-Effekt. In dem Fall schlägt das Programm Termkandidaten vor, die keine Fachtermini sind, es ist insofern zu „laut“, weil es sich zu oft meldet (vgl. Drewer & Schmitz 2017: 46–47). Die Autor:innen weisen außerdem darauf hin, dass es besonders schwer für die Programme ist, Mehrwortbenennungen richtig zu extrahieren, denn sie erkennen nicht immer, wie die Benennungsbestandteile zusammengehören (2017: 48). Um dem Silence- und Noise-Effekt entgegenzuwirken, lassen sich, je nach dem gewünschten Ziel, Anpassungen in den Programmeinstellungen über den sog. Qualitätsfilter vornehmen. Infolgedessen werden entweder weniger Termini gewonnen, die allerdings sicherer

sind, oder mehr Termini, mit der Gefahr auch Redundanzen mitzuextrahieren. Lyding und Ties (2008: 94) sehen zwar die Verschiebung des Schwellenwertes als hilfreich, um auf die gewünschten Ergebnisse zu kommen, aber nicht als eine Lösung, um das System zu verbessern. Dies kann nur dann erfolgen, wenn ein Extraktionsalgorithmus entwickelt wird, der auf Erkennung von zahlreicheren und genaueren Informationen ausgerichtet wird (vgl. 2008: 94). Chiochetti et al. (2023: 469) betonen, dass juristische Fachtermini eine zusätzliche Herausforderung für die Termextraktion darstellen. Die Autor:innen erklären nämlich, dass es an der Natur der juristischen Texte liegt, in denen Schlüsselterminologie manchmal nur vereinzelt vorkommt, was bei gewissen Termextraktionsverfahren von Nachteil ist. Außerdem können gemeinsprachliche Wörter in juristischen Texten sowohl ihre ursprüngliche als auch eine typisch rechtliche Bedeutung haben. Wenn Personen, die im Anschluss die Termkandidaten überprüfen, über keinen juristischen Hintergrund verfügen, kann möglicherweise auch die Schlüsselterminologie nicht als solche identifiziert werden (2023: 469).

2.3.2.2 Termextraktionssysteme

Der Nutzen der Terminologearbeit, -identifizierung und -gewinnung ist unumstritten, wie den vorigen Kapiteln zu entnehmen ist. Es wurden auch bereits die Vorteile und Nachteile der manuellen Terminologierecherche und automatischen Termextraktion beschrieben, insbesondere deren zeitlicher Aufwand. Kageura und Umino (1996) betonen in ihrer Publikation zu Methoden der automatischen Terminologieidentifizierung allerdings die Notwendigkeit, in der heutigen durch den technischen Fortschritt geprägten Welt, Fachtermini automatisch zu erkennen:

Automatic term recognition (ATR) in particular is much needed because a simple but coherently built terminology is the starting point of many applications such as human or machine translation, indexing, thesaurus construction, knowledge organisation, etc., and because manual efforts cannot keep up with the rapid growth of technical terms. (Kageura & Umino 1996: 260)

Um mit der technischen Entwicklung terminologiemäßig Schritt halten zu können und gleichzeitig die Arbeitseffizienz zu erhöhen (vgl. Strehlow 2001: 426), werden Termextraktionssysteme eingesetzt. Kageura und Umino (1996: 260) sehen frühe Anfänge der automatischen Terminologieerkennung in den 1950er Jahren mit den Arbeiten zur automatischen Indexierung von Luhn. Dabei handelt es sich um die Informationsrückgewinnung von wichtigen Elementen im Text. Für Cabré (1999: 160–161) stellen die Termextraktionssysteme die vierte Stufe in der

Entwicklung der Zusammenarbeit zwischen der Informatik und Linguistik dar. Es sind die sog. Expert:innensysteme, die über gewisse Intelligenz verfügen und die Entscheidungen bzw. Lösungswege der Expert:innen nachahmen können, um den Automatisierungsgrad zu erhöhen und dadurch den nötigen Menscheneinsatz zu reduzieren. Das interdisziplinäre Forschungsgebiet, das sich damit befasst, heißt Computerlinguistik. Lt. Vivaldi und Rodriguez (2007: 234) werden seit den 1980er Jahren viele Bestrebungen beobachtet, mehrere Techniken zu entwickeln, die bei der Termextraktion zum Einsatz kommen könnten. Die Autoren sind relativ skeptisch, was die Leistung dieser Techniken individuell betrifft, aber sie sehen auch Erfolge in gewissen Bereichen vom Natural Language Processing (NLP), wenn mehrere Methoden miteinander kombiniert werden (2007: 234). Besonders in den letzten Jahren werden Versuche unternommen, verschiedene Termextraktionssysteme samt Termextraktionstools bzw. die Informationen über die Methoden und Techniken, auf denen sie basieren, zusammenzustellen und zu revidieren (Costa et al. 2016; Di Nunzio et al. 2023; Tran et al. 2023). Die traditionellen Termextraktionssysteme haben statistische und linguistische Charakteristika der Fachtermini im Fokus oder sie verwenden den hybriden Ansatz, der das Wissen über beide Charakteristika vereint (vgl. Cabré Castellví et al. 2001: 54).

Witschel (2005: 661-662) sieht bei dem statistischen Verfahren die Auskunft darüber, wie Wörter im Text verteilt sind. Der Autor konstatiert, dass es sowohl in den gemeinsprachlichen als auch fachlichen Texten eine kleine Anzahl an Wörtern gibt, die sehr oft in den Texten vorkommen, und dass die meisten Wörter sich selten wiederholen. Dieses Verhältnis ist in beiden Formen von Texten gleich. Die häufig auftretenden Wörter erfüllen meistens eine syntaktische Funktion und zählen zu Stoppwörtern. Außerdem gibt es Lexeme, die in einem Fachtext häufig vorkommen, in der Gemeinsprache wiederum sehr selten (2005: 661-662). Diese Charakteristika liegen der Idee der Differenzanalyse zugrunde, die für die Zwecke der statistischen Termextraktion eingesetzt werden kann.

Bei den linguistischen Eigenschaften der Fachtermini werden die Morphologie und Syntax analysiert. Die Fachterminologie besteht größtenteils aus Nomina und Adjektiven; die Derivate verfügen über spezifische Suffixe, die in den Fachsprachen häufiger vorkommen; und sie enthält viele Abkürzungen und Komposita. Die Mehrwortbenennungen stellen eine Herausforderung bei der Termextraktion dar. Es wird zwar die syntaktische Ähnlichkeit in deren Struktur berücksichtigt, aber die auf dieser Idee basierende Extraktion liefert auch viele Fehl-

treffer (vgl. 2005: 661-662). Die Verfahren, denen diese Charakteristika zugrunde liegen, werden im dritten Kapitel detaillierter beschrieben. Lt. Tran et al. (2023: 2) verwenden die derzeitigen Termextraktionssysteme hauptsächlich Künstliche Neuronale Netze, die bereits Erfolge in anderen Bereichen vom NLP erzielen. Sie basieren größtenteils auf Transformer-Sprachmodellen, die typischerweise selbst die relevanten Merkmale erschließen, wodurch der Bedarf für eigenes Feature Engineering entfällt und im Regelfall keine sachgebietsspezifischen Quellen erfordern und somit sachgebietsunabhängiger arbeiten (vgl. 2023: 2). Das neuronale Verfahren samt dem für die vorliegende Arbeit relevanten und heute wohl beliebtesten (vgl. Seemann 2023: 10) Large-Language-Modell-System GPT wird ebenfalls in weiterer Folge (3.1.4.) detaillierter dargestellt.

Kageura und Umino (1996: 260-261) prägen zwei Begriffe, die ihrer Ansicht nach die zentrale Rolle bei der automatischen Terminologieidentifizierung spielen: Termhood und Unithood. Termhood bezieht sich auf den Grad der Beziehung eines linguistischen Ausdrucks zum Sachgebiet, anders ausgedrückt, auf seinen Grad der Fachlichkeit. Es betrifft sowohl Ein- als auch Mehrwortbenennungen. Unithood beschreibt den Grad der Stärke bzw. Stabilität syntagmatischer Kombination oder Kollokation. Es bezieht sich nicht nur auf Ein- und Mehrwortbenennungen, sondern auch auf andere komplexe linguistische Einheiten wie z. B. Redewendungen (vgl. 1996: 260-261). Tran et al. (2023: 9) nennen außer Termhood und Unithood auch weitere Merkmale und Maßnahmen wie C-Wert, Tokenisierung, Klassifikation, POS-Tagging, u.v.m., die von Termextraktionssystemen eingesetzt werden, um bestmögliche Ergebnisse zu erzielen.

Wie bereits erwähnt, gibt es mehrere Tools zur Termextraktion, die auf den oben genannten Ansätzen basieren. Costa et al. (2016: 15–17) nennen neun solcher Tools und analysieren über welche Eigenschaften sie verfügen. Dazu zählen u.a. die Möglichkeit, Stoppwortlisten einzusetzen, bilinguale Termextraktion oder Unterstützung von mehreren Dateiformaten. Die analysierten Tools sind entweder alleinstehend, wie SDL MultiTerm Extract, Simple Extractor, TermSuite oder eine Webanwendung wie OneClick Terms von Sketch Engine, Translated s.r.l., Terminus. Die drei letzten dargestellten Systeme sind Frameworks und es sind Keyphrase Extraction Algorithm (KEA), Rainbow und Java Automatic Term Extraction (JATE) (vgl. 2016: 15-17). Für die vorliegende Arbeit wird neben einem generativen vortrainierten Transformer ChatGPT-4o (OpenAI 2024) auch Text2TCS (Gromann et al. 2023) verwendet (3.1.4.1.-3.1.4.2.).

3. Verfahren der automatischen Terminologiegewinnung

Text ist ein bedeutender Wissensrohstoff, der im Zeitalter des Internets in großen Mengen in digitaler Form zur Verfügung steht. (Heyer et al. 2008: 1)

Es sind genau die großen Mengen von maschinenlesbaren Texten, die für den Einsatz der automatischen Termextraktion sprechen. Im vorigen Kapitel wurden die Methoden der Terminologiegewinnung dargestellt. Während die manuelle Terminologierecherche ohne Weiteres für sich selbst spricht, auch wenn es verschiedene Tools und Programme gibt, die den Prozess unterstützen, liegen der automatischen Termextraktion unterschiedliche Ansätze zugrunde. Damit nachvollzogen werden kann, auf welche Art und Weise Termkandidaten automatisch ausgewählt und extrahiert werden, wird ein Einblick in die Funktionsweise dieser Ansätze benötigt. Deren Charakteristika wurden teilweise bereits umrissen, allerdings ist das Ziel dieses Kapitels, genauer auf sie einzugehen. Es werden der Arbeitsablauf der Termextraktion sowie das linguistische, statistische, hybride und neuronale Verfahren samt den für die Untersuchung ausgewählten Tools beschrieben. Anschließend wird das für die vorliegende Arbeit relevante Prompt Engineering kurz präsentiert. Den Umfang der Arbeit betrachtend, werden zwangsläufig nicht alle Berechnungsmethoden detailliert beschrieben, sondern es sollten vielmehr Grundlagen für weitere Diskussion und den empirischen Teil geschaffen werden.

3.1. Arbeitsablauf der automatischen Termextraktion

Die Suche nach Zusammenhängen im Text, die neu sowie sachlich und inhaltlich relevant sind, ist der Gegenstand vom Text Mining (vgl. Heyer et al. 2008: 1). Da die im Fachtext erhaltenen Fachtermini ein Teil des Wissens sind, das der jeweilige Text zu bieten hat, handelt es sich bei der Termextraktion um ein Teilgebiet vom Text Mining (vgl. Witschel 2005: 659). Der erste Schritt der automatischen Terminologiegewinnung ist die Auswahl geeigneter Texte. Biemann et al. (2022: 312) nennen u.a. umfangreiche Produktdokumentationen, Verträge, Korrespondenz, Softwarespezifikationen und Ausführungsvorschriften. Die Autor:innen (2022: 312) sehen als den nächsten erforderlichen Schritt die Vorverarbeitungsphase, die korpuslinguistische textbezogene Maßnahmen enthält. Bezweckt wird damit die Steigerung der Effizienz und Verbesserung der Berechnungen bei den Verfahren (vgl. auch Heid 2001: 189). Diese Phase beinhaltet z. B. die Tokenisierung, im Zuge deren der Text in Tokens aufgeteilt wird, also in die durch Leerzeichen begrenzten Zeichenketten (vgl. Voorhees 1999: 34), die für ein Programm als Wörter, Phrasen oder Sätze zu verstehen sind. Eine Herausforderung in dem Zusammenhang

mit den Satzgrenzen bilden Abkürzungen auf deren Ende ein Punkt steht. Um das Problem zu umgehen, können alle Abkürzungen im Text ausgeschrieben werden oder es kann eine Abkürzungsliste eingeführt werden. Das Programm referenziert die Liste und erkennt die Abkürzungen als das was sie sind. Biemann et al. (2022) weisen allerdings bezüglich der Tokenisierung darauf hin, dass sie bei gewissen Sachgebieten angepasst werden muss.

Falls z. B. chemische Verbindungen wie 1-(1,2-Dihexadecanoylphosphatidyl)inositol-3,4-diphosphat als Fachterminologie relevant sind, sollten diese auch als ein (und nicht mehrere) Token repräsentiert werden. (Biemann et al. 2022: 111)

Außerdem wird im Zuge der Tokenisierung von Übernormalisierung gewarnt, wenn der Text z. B. in Kleinbuchstaben und in Bezug auf Sonderzeichen normalisiert wird. Dies kann ggf. dazu führen, dass die Tokens eine andere Bedeutung bekommen oder die Mehrdeutigkeit begünstigt wird (vgl. Biemann et al. 2022: 98).

- Wir essen jetzt, Freunde! vs. Wir essen jetzt Freunde!
- Der Junge sieht dir Ungeheuer ähnlich. vs. der junge sieht dir ungeheuer ähnlich.
- Bier in Massen servieren! vs. Bier in Maßen servieren! (Biemann et al. 2022: 98)

Es ist anzumerken, dass die Tokenisierung der neuronalen Sprachmodelle sich erheblich von der simplen Tokenisierung auf Basis von Leerzeichen und Interpunktion unterscheidet, da die Aufteilung der Sequenz auf Subwortebene durch eigene neuronale Tokenizers erfolgt. Weiters können Wortklassenannotation und Lemmatisierung durchgeführt werden, damit die extrahierten Termkandidaten gezieltere Ergebnisse bringen bzw. zwecks weiterer Verarbeitung. Die Wortklassen annotieren heißt jeder Wortform im Text ein in dem Kontext passendes Label zuzuweisen. Dieser Prozess wird auch Part-of-speech-Tagging bzw. POS-Tagging genannt. Unter den Labels sind Nomen, Adjektiv, Verb usw. zu verstehen (vgl. Frantzi et al. 2000: 116; Jurafsky & Martin 2024: 365). Der Kontext spielt deshalb eine große Rolle, da es Wörter gibt, deren Bedeutung und gleichzeitig ihre Wortart im Satz sich nur in Anbetracht der umliegenden Wörter bestimmen lässt wie z. B. *Herr Schloss bezahlte wieder das bezahlte Schloss*. Beim *Schloss* handelt es sich einmal um einen Eigennamen und einmal um ein Nomen und *bezahlte* ist einerseits ein Verb und andererseits ein Adjektiv/Partizip (vgl. Biemann et al. 2022: 102). Die Anzahl der möglichen Wortarten wird Tagset genannt. Lt. Autoren (2022: 102) enthält das für die Annotation der Texte in der deutschen Sprache am meisten verwendete Tagset 54 Tags, es kann allerdings auch bedarfsmäßig erweitert werden. Im Rahmen der Lemmatisierung werden die Wörter auf ihre Grundform reduziert, anders ausgedrückt bekommt jede flektierte Form ein Lemma also ihr Lexikoneintragswort (vgl. Heid 2001: 190). Bei der Termextraktion und

beim Text Mining im Allgemeinen wird der Fokus nämlich mehr auf den Inhalt als auf die linguistische Struktur im und vom Text gelegt (vgl. Biemann et al. 2022: 99). Zu den Beispielen der Lemmatisierung zählen:

- Nomen im Nominativ Singular (Tisch vs. Tisches, Tischen, Tische)
- Verben im Infinitiv (kommen vs. käme, kam, komme)
- Adjektive im Prädikativ (klug vs. kluge, klugen, klügste) (Biemann et al. 2022: 99)

Lt. Heid (2001: 190) erweist sich dieser Prozess in Fachtexten manchmal als problematisch, denn Fachtexte enthalten viele spezifische Fachwörter, die sich oft in den Lexika nicht befinden, was die Suche nach dem richtigen Lemma deutlich erschwert. Es können auch weitere Vorverarbeitungsverfahren erfolgen wie z. B. Stemming bzw. Stammformreduktion, d.h. die Vereinheitlichung der Vollform auf den bedeutungscharakterisierenden Teil vom Wortstamm; Chunking, d.h. die Zusammenfassung von grammatisch zusammenhängenden Phrasen; Syntaxparsing, d.h. eine detailliertere syntaktische Analyse, wie z. B. von Subjekt-Prädikat-Objekt-Beziehungen, u.a. (Biemann et al. 2022: 101–103). Nach dem Abschluss der Vorverarbeitungsphase werden die Termkandidaten extrahiert. Den Kern der automatischen Termextraktion, unabhängig von dem ausgewählten Verfahren, bilden zwei Prozesse.

- Identifizieren und Sammeln potenzieller terminologischer Einheiten vom Text.
- Herausfiltern von Nicht-Termini aus der Liste aller extrahierten terminologischen Einheiten und Sortieren der Termkandidaten u.a. nach Unithood oder nach ihrem Nutzen für den beabsichtigten Zweck der Extraktion. (vgl. Daille 2017: 172)

Im Anschluss werden die Termkandidaten von Terminolog:innen überprüft. Abbildung 3 stellt den Ablauf der automatischen Termextraktion schematisch dar.

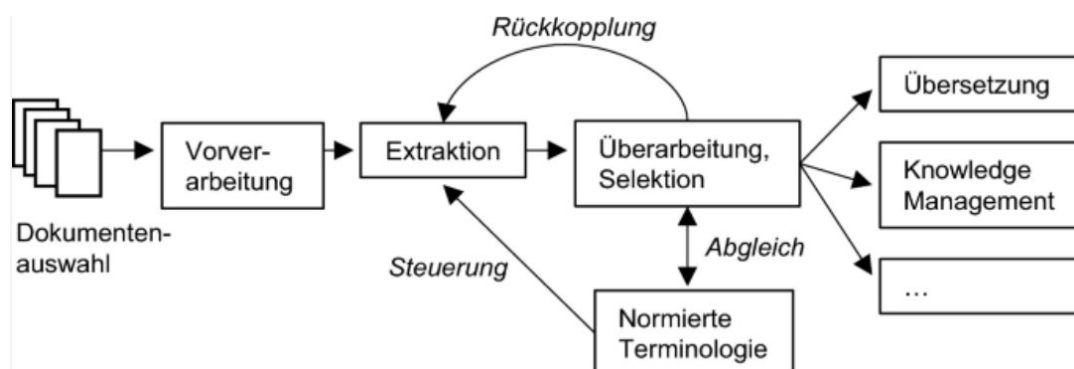


Abbildung 3: Schematischer Ablauf automatischer Termextraktion (Biemann et al. 2022: 313)

Termextraktionsprogramme stehen im Zuge der Terminologiegewinnung vor gewissen Herausforderungen. Dazu zählen insbesondere die vollständige Identifizierung und Erkennung von Mehrwortbenennungen, die Bestimmung, ob die sprachliche Einheit in dem Kontext eine terminologische oder allgemeinsprachliche Bedeutung hat sowie die Angemessenheit des Fachausdrucks bezüglich des Fachvokabulars (vgl. Cabré Castellví et al. 2001).

Bevor auf die traditionellen Verfahren der automatischen Terminologiegewinnung, nämlich das linguistische und das statistische Verfahren, detailliert eingegangen wird, soll eine tiefgründigere Einteilung genannt werden, die beide Verfahren gemeinsam haben. Es handelt sich dabei um den intrinsischen Ansatz, der benennungsbasiert ist und den extrinsischen Ansatz, der kontextbasiert ist. Die intrinsischen Methoden basieren auf syntaktischen, semantischen und morphologischen Daten, die den Termini innewohnen und wonach sie herausgefiltert werden, z. B. Zylinderabschaltung → „zylinder“, „ab“, „schalten“, „ung“. Die extrinsischen Methoden beziehen sich auf die morphosyntaktische Struktur der Ein- und Mehrwortbenennung, es wird nach Daten außerhalb des Fachausdrucks selbst gesucht, z. B. NP = Adjektiv + Nomen → *außerordentliche Revision*. Weiters können die extrinsischen Ansätze noch in syntagmatische, z. B. kontextuelle oder syntaktische Daten und paradigmatische, z. B. Beziehungen zwischen Termini und Termkandidaten, Methoden untergliedert werden (Voltmer et al. 2003: 7; Zielinski & Ramírez-Safar 2005: 2). Tabelle 1 bietet eine Übersicht über Termextraktionsmethoden bei linguistischen und statistischen Verfahren, auf die z.T. in weiterer Folge detaillierter eingegangen wird.

Tabelle 1: Übersicht über Termextraktionsmethoden bei linguistischen und statistischen Verfahren (adaptiert von Voltmer et al. 2003: 7)

Methode			
linguistisch	intrinsisch		POS-tagging, Chunking, Stoppwörter
	extrinsisch	syntagmatisch	Volles Parsing
		paradigmatisch	Termvariation
statistisch	intrinsisch		Mutual Information, Likelihood Ratio
	extrinsisch	syntagmatisch	NC-Wert, Entropie
		paradigmatisch	C-Wert, Weirdness

Im nächsten Schritt wird der Prozess der linguistischen Termextraktionsverfahren dargestellt.

3.1.1. Linguistische Verfahren

Den Kern der linguistischen Verfahren bildet die semantische, syntaktische und morphologische Analyse der Sprache des Texts, weswegen sie der manuellen Terminologierecherche am nächsten kommt (vgl. Reuther 2008: 126). Mit der syntaktischen Analyse, die oft bei der Extraktion von Mehrwortbenennungen eingesetzt wird und dabei gute Resultate verzeichnet (vgl. Drewer & Schmitz 2017: 49; Witschel 2005: 663), geht die Suche nach bestimmten Wortkombinationen einher, die zu einem typisch terminologischen Wortartmuster, sogenannten POS-Muster, der jeweiligen Sprache passen. Die Muster, nach denen Termkandidaten extrahiert werden, können entweder händisch von Personen mit hohen linguistischen und terminologischen Kompetenzen geschrieben oder automatisch implementiert werden. Dazu werden Data-Mining-Techniken bzw. Techniken des maschinellen Lernens eingesetzt, um die Muster anhand von Korpora zu identifizieren (Daille 2017: 173). Es werden zunächst Termini von diversen Wortartmustern verglichen und deren Ähnlichkeit wird berechnet. Anschließend werden POS-Muster der erkannten ähnlichen Termini gezählt, wovon sich Muster-Äquivalenzen ableiten lassen (Weller et al. 2011: 89). Dies können beispielsweise im Deutschen Nominalphrasen mit dem Muster Nomen + Nomen (N+N), wie etwa *Zubereitung des Kaffees*, Adjektiv + Nomen (A+N) wie *verdrilltes Leitungspaar* und Nomen + Verb (N+V) wie *Kaffeeprodukte zubereiten* sein (vgl. Heid 2001: 187). Gleichzeitig soll darauf hingewiesen werden, dass die für die Termini typischen POS-Muster sich von Sprache zu Sprache unterscheiden können. Das Englische teilt sich beispielsweise bei Mehrwortbenennungen mit dem Deutschen die Nominalphrasen N+N wie etwa *fabric conditioner* und A+N wie *phosphoric ester*, aber gewöhnlicherweise nicht V+N wie *stop proceedings* oder N + Präposition + N (N+P+N) wie *Code of Conduct*. Das POS-Muster N+P+N ist wiederum üblich in der französischen Terminologie wie *conduit de ventilation*. Die gängigen Nominalphrasen A+N im Deutschen, wie *ambulante Behandlung*, und im Englischen, wie *ambulatory treatment*, folgen im Französischen und Polnischen dem N+A Muster, wie *traitement ambulatoire*; *leczenie ambulatoryjne*, usw. Die hohe Sprachabhängigkeit und somit die stark begrenzte Einsetzbarkeit zählt zu den größten Nachteilen der linguistischen Verfahren (vgl. Bańko 2011; Bowker 2008: 84; Drewer & Schmitz 2017: 49; Witschel 2005: 662). Die morphologischen Analyseprogramme POS-Tagger und Parser verwenden verschiedene Strategien bei der Analyse; es können einfache Systeme sein, bei denen keine Wörterbücher verwendet werden, bis hin zu komplexen Systemen mit der detaillierten morphologischen Analyse und der eingebauten automatischen Lemmatisierung (vgl. Cabré Castellví et al. 2001: 56). Der Nachteil der syntaktischen Analyse besteht in dem hohen Rechenaufwand, denn ihre Durchführung schließt die Entwicklung und Erhaltung eines umfangreichen Lexikons ein,

in dem das Wissen über syntaktische und semantische Charakteristika der Wörter abgespeichert und bei Bedarf erweitert werden muss (vgl. Witschel 2005: 664). Außerdem kommt es dabei zum Noise- bzw. Silence-Effekt (vgl. Bowker 2008: 83–84). Es werden einerseits Nicht-Termini extrahiert, denn die nebenstehenden Wörter können zwar zufällig dem angewandten POS-Muster folgen, aber dennoch der Allgemeinsprache gehören, wie in Tabelle 2 dargestellt durch *more options*, *periodic checks*, *stored information*. Andererseits können relevante Termini übersehen werden, wenn sie nicht dem vorprogrammierten, üblichen POS-Muster entsprechen, wie in Tabelle 2 dargestellt durch *before-the-fact detection*, *after-the-fact detection*.

Tabelle 2: Ein dem linguistischen Extraktionsverfahren unterzogener Text mit Termkandidaten, die den Noise- bzw. den Silence-Effekt illustrieren (adaptiert von Bowker 2008: 83)

POS-Muster: N+N; A+N Noise-Effekt <u>Termkandidaten</u> <u>Antivirus programs</u> now include <i>more options</i> . <u>Integrity checking</u> performs <i>periodic checks</i> of the <u>current status</u> of the files against the <u>stored information</u> . <u>Behaviour blocking</u> performs before-the-fact detection. <u>Heuristic analysis</u> is a form of after-the-fact detection.	POS-Muster: N+N; A+N Silence-Effekt (P+Artikel+N+N) <u>Termkandidaten</u> <u>Antivirus programs</u> now include <i>more options</i> . <u>Integrity checking</u> performs <i>periodic checks</i> of the <u>current status</u> of the files against the <u>stored information</u> . <u>Behaviour blocking</u> performs <i>before-the-fact detection</i> . <u>Heuristic analysis</u> is a form of <i>after-the-fact detection</i> .
---	---

Biemann et al. (2022: 316) teilen diese Auffassung, aber argumentieren gleichzeitig, dass das syntaktische linguistische Verfahren trotzdem „sehr gute“ Resultate liefert und sich in der Praxis effizient einsetzen lässt und Zielinski und Ramírez-Safar (2005: 3) unterstreichen die höhere Präzision des Verfahrens im Vergleich zu den statistischen Ansätzen. Im Falle von Einwortbenennungen handelt es sich in der deutschen Terminologie hauptsächlich um N+N, wie *Kaffeeprodukt*, teilweise auch um V, wie (Kaffee) *zubereiten* und A wie *tassenwarm* (vgl. Heid 2001: 187). Mit der Analyse der Wortklassenmuster lassen sich gezielt jene Wörter herausfiltern, die zu den terminologiebezogenen POS-Mustern gehören. Eine für die morphologische Analyse interessante Gruppe der Einwortbenennungen bilden Komposita. Sollte nämlich das im Programm eingebaute Wörterbuch einen zusammengesetzten Fachausdruck nicht umfassen, kann nach Informationen in deren Bestandteilen gesucht werden, angenommen dass sie nicht metaphorisch gebraucht werden (vgl. Biemann et al. 2022: 314–315). Gulbins und Kahrman (2000: 86) nennen diesbezüglich als Beispiel eine zusammengesetzte Einwortbenennung aus dem Sachgebiet der Typographie, den Fachausdruck *Schusterjunge*, der die erste Zeile vom Absatz am Seitenende bezeichnet. Im Regelfall lassen sich allerdings durch die Zerlegung der Kom-

posita dem Wörterbuch unbekannte Fachtermini im Text ausfindig machen, da deren Determinatum vorwiegend ein Oberbegriff ist (vgl. Biemann et al. 2022: 315). Als Beispiel gelten die Ausprägungen des Oberbegriffs *Ingenieur*: Entwicklungsingenieur, Softwareingenieur, Vertriebsingenieur, Bauingenieur, Elektroingenieur u.a. Außerdem stehen bestimmte für die Terminologie charakteristische Derivationsaffixe im Fokus der morphologischen Untersuchung. Dazu zählen Suffixe lateinischen und griechischen Ursprungs *-itis* und *-ase*, die mehrheitlich in den naturwissenschaftlichen Domänen wie Chemie, Biologie und Medizin vorkommen, in der Gemeinsprache wiederum selten vorzufinden sind. Sollten Fachtermini aus einem medizinischen Text über Infektionen und Entzündungen extrahiert werden, kann das dafür ausgerichtete Programm Termkandidaten vorschlagen, die höchstwahrscheinlich Termini sind, z. B. *Endokarditis*, *Appendizitis*, *Meningitis* u.a. (vgl. Biemann et al. 2022: 314; Witschel 2005: 662). Um die Ergebnisse des linguistischen Verfahrens zu optimieren, können zu den Extraktionsprogrammen digitale Allgemeinwörterbücher hinzugefügt werden. Die allgemeinsprachlichen Wörter werden aus der extrahierten Termkandidatenliste automatisch herausgenommen (vgl. Drewer & Schmitz 2017: 49). Andererseits können linguistische Termkandidatenfilter nach dem Prinzip der Stoppwortlisten (vgl. Ahmad & Rogers 2001: 741ff.; Carstensen et al. 2010: 588) eingesetzt werden. Heid (2001: 194) unterscheidet dabei zwischen lexikalischen Filtern, die terminologisch nicht relevante Wörter enthalten, wie z. B. häufig auftretende Adjektive: *nachfolgend*, *obenstehend*, *bisherig*, *übrig*, *entsprechend* sowie Konstruktionsfiltern mit syntaktischen Konstruktionen, die keine terminologische Relevanz aufweisen wie etwa Zeitangaben, z. B. *10 Minuten lang*, Mengenangaben, wie etwa *Fass à 100 Liter*, oder Nummern, wie *Nr 108-27* (vgl. 2001: 194).

Zu den ersten automatischen Extraktionssystemen, die auf rein linguistischen Ansätzen basieren, zählt LEXTER (Bourigault 1992). Das System wurde 1992 zwecks Terminologiepflege bei Électricité de France entwickelt und die Termkandidaten wurden aus einem Korpus französischsprachiger Texte gewonnen. Seine Funktionsweise gründet auf einer oberflächlichen syntaktischen Analyse des Texts gefolgt vom Parsing. Mithilfe vom POS-Tagging wird allen Wörtern ihre grammatische Kategorie zugeordnet. Um die Termkandidaten zu identifizieren, werden die sog. frontier markers eingesetzt, die aus Einheiten bestehen, die gewöhnlicherweise keinen Fachaussdruck bilden und/oder nicht zu seinem Bestandteil gehören wie etwa für Französisch Pronomina, konjugierte Verben, Konjunktionen u.a. Das Extraktionsergebnis bilden meistens maximale Nominalphrasen, die entweder bereits Termini sind oder mehrere Termini enthalten wie z. B. *disque dur de la station de travail* → *disque dur* + *station de travail*. Um daraus die gewünschten Nominalphrasen gewinnen zu können, werden sie dem Parsing

unterzogen. Die Parser bestehen aus handgeschriebenen Regeln und analysieren die grammatischen Regeln sowie die Stellung der Wörter in der maximalen Nominalphrase. Die Parsing-Regel, die das oben genannte Beispiel betrifft, sieht folgendermaßen aus (leicht adaptiert von Bourigault 1992: 979).

$N1+A+P+Artikel+N2+P+N3 = \text{disque dur de la station de travail}$

- $N1+A = \text{disque dur}$
- $N2+P+N3 = \text{station de travail}$ (Bourigault 1992: 979)

Anschließend werden die Termkandidaten den Terminolog:innen zur Überprüfung vorgelegt (Bourigault 1992: 977–980). LEXTER ist zwar sprachabhängig, aber Pazienza (1998: 186) argumentiert, dass das System sich auch bei anderen Sprachen einsetzen lässt, solange die oben genannten linguistischen Informationen zur Verfügung stehen. Witschel (2004: 46) bemängelt bei diesem System die umfassende Handarbeit bei der Erstellung der Parser-Regeln und frontier markers. Da die vorliegende Arbeit sich nicht explizit mit Systemen befasst, die auf rein linguistischen Ansätzen basieren, wird auf keine weiteren Beispiele eingegangen, sondern lediglich erwähnt, dass andere Systeme, die syntaktische, morphologische sowie semantische Informationen im Zuge der Extraktion einsetzen, u.a. von Justeson und Katz (1995: 9ff.), Seewald (1993: 75ff.) sowie Arppe (1995) vorgeschlagen wurden.

3.1.2. Statistische Verfahren

Die auf statistischen Ansätzen basierende Termextraktion geht von einem probabilistischen Modell der Nutzung der Wörter im Text aus (vgl. Drewer & Schmitz 2017: 49), dessen theoretische Grundlage das sog. Zipfsche Gesetz bildet. Demnach folgt das menschliche Verhalten jeglicher Art sowie die Funktionsweise der natürlichen Sprachen dem Prinzip des geringsten Aufwandes (vgl. Zipf 1949: 8). Das Prinzip wird folgendermaßen definiert.

Each individual will adopt a course of action that will involve the expenditure of the probably least average of his work (by definition, least effort). (Zipf 1949: 543)

Deshalb sind die Wörter, die am häufigsten in einer Sprache verwendet werden, meistens sehr kurze Funktionswörter (Biemann et al. 2022: 87-88) und es gibt generell in den menschlichen Sprachen ein paar von sehr häufig auftretenden Wörtern, eine mittelgroße Zahl an Wörtern, die mit mittlerer Frequenz vorkommen und viele selten auftretende Wörter (vgl. Manning & Schütze 1999: 24). Angenommen, dass alle Wörter eines Texts nach der absoluten Häufigkeit ihres Auftretens sortiert werden, lässt sich im Sinne des Prinzips Folgendes beobachten. Ihre

Position auf der Liste multipliziert mit ihrer Häufigkeit bleibt in etwa konstant, was wiederum bedeutet, dass ihre Häufigkeit umgekehrt proportional zu ihrer Position auf der Liste ist (1999: 24). Daraus lässt sich herleiten, dass einige wenige Wörter im Text, die zu den typischen Stoppwörtern gehören, oft vorkommen und die meisten nur selten (Witschel 2005: 661) wie bereits oben angedeutet. Manning und Schütze (1999: 24) argumentieren, dass das Zipfsche Gesetz, das sie weniger als Gesetz und mehr als „eine annähernd genaue Beschreibung bestimmter empirischer Fakten“ (1999: 24) sehen, zwar nicht immer zutrifft, aber trotzdem „nützlich als grobe Beschreibung der Häufigkeitsverteilung von Wörtern in menschlichen Sprachen“ ist (1999: 24). In diesem Zusammenhang wird außerdem angenommen, dass Fachtermini im Text eines bestimmten Sachgebiets dort häufiger auftreten als in den allgemeinsprachlichen Texten oder in jenen eines anderen Sachgebiets (vgl. Biemann et al. 2022: 313; Drewer & Schmitz 2017: 49).

Das statistische Verfahren, bei welchem allgemeinsprachliche Korpora mit Fachsprachlichen Korpora verglichen werden, wird Differenzanalyse bzw. Weirness (vgl. Ahmad & Rogers 2001: 744) genannt. Mit der Differenzanalyse werden statistisch signifikante Unterschiede zwischen den Texten untersucht. Um die terminologisch relevanten Ausdrücke zu ermitteln, wird zuerst ein Analysekorpus zusammengestellt, das aus den in Frage kommenden Fachtexten besteht. Dann wird ein allgemeinsprachlicher Referenzkorpus benötigt, dessen Vorlage z. B. Zeitungsartikel bilden. Den nächsten Schritt bildet die Berechnung der Auftretenshäufigkeit der lexikalischen Einheiten und deren Vergleich. In Anbetracht der Ergebnisse kann zwischen vier Kategorien unterschieden werden.

- Ausdrücke die im Referenzkorpus nicht auftreten sind höchstwahrscheinlich Fachtermini.
- Ausdrücke die im Analysekorpus häufiger als im Referenzkorpus auftreten sind wahrscheinlich Fachtermini, sie müssen allerdings als solche anhand eines Schwellwerts identifiziert werden.
- Ausdrücke die mehr oder weniger in beiden Korpora gleich häufig auftreten sind meistens Stoppwörter und allgemeine Begriffe.
- Ausdrücke die im Analysekorpus seltener als im Referenzkorpus auftreten sind grundsätzlich keine Fachtermini. (Heyer et al. 2008: 95–96)

Relevant für die Differenzanalyse sind die beiden ersten Kategorien, wobei die Qualität der Resultate stark mit der Auswahl des richtigen Referenzkorpus zusammenhängt (2008: 95-96).

Damit die Unterschiede der Verwendung von Wörtern untersucht werden können, wird ein statistisches Maß benötigt. Außerdem können bestimmte weitere Aspekte berücksichtigt werden wie z. B. die Wortverteilung. Ein solches Verfahren, das meist ohne Referenzkorpus verwendet wird und die Wortverteilung berücksichtigt, wird Term-Frequency-Inverse-

Document-Frequency (TF-IDF) genannt. Es basiert auf der Annahme, dass Wörter, die in wenigen Texten häufig auftreten, charakteristischer für diese Texte sind, als Wörter, die zwar insgesamt häufiger auftreten, aber dafür in vielen Texten zu finden sind (Biemann et al. 2022: 246; Salton et al. 1975). Mit der Term-Frequency wird die Häufigkeit eines Wortes im Text (Dokument) ermittelt und mit der inversen Dokument-Frequency sein Auftreten in Texten. Das Ergebnis zeigt den Grad der Wichtigkeit bzw. Spezialität des Wortes für den/im Text (vgl. Witschel 2005: 663). Es werden dabei Termkandidaten extrahiert, die häufig in einem Dokument und gleichzeitig selten im gesamten Fachkorpus auftreten, woraus sich wiederum herleiten lässt, dass sie wichtig für den bestimmten Text (und für andere nicht) sind, ergo handelt es sich dabei wahrscheinlich um Fachtermini dieses Textes (vgl. Zielinski & Ramírez-Safar 2005: 2). Ein anderes Verfahren, bei dem wiederum die Wortverteilung nicht berücksichtigt wird, heißt Log-Likelihood-Ratio, das folgendermaßen definiert wird.

Quotient aus zwei Likelihood-Funktionen, im Korpusvergleich Quotient aus der Likelihood-Funktion der wahrscheinlichen Anzahl eines Types im Analysekorpus und der als Nullhypothese aus dem Referenzkorpus abgeleiteten erwarteten Zahl. (Biemann et al. 2022: 362)

Anders ausgedrückt wird für jedes Wort seine relative Häufigkeit sowohl im Analyse- als auch im Referenzkorpus berechnet und sie werden miteinander verglichen, um festzustellen, ob die Häufigkeit des Wortes im Analysekorpus statistisch signifikant über dem Wert liegt, mit dem wegen seiner Häufigkeit im Referenzkorpus zu rechnen wäre. Die Nullhypothese geht davon aus, dass die Werte gleich sind und die Differenzanalyse ist in diesem Fall darauf ausgerichtet, nach den Wörtern oder Phrasen zu suchen, deren Frequenz im Analysekorpus stark nach oben abschweift (vgl. Biemann et al. 2022: 244; Witschel 2005: 663). Zu weiteren statistischen Verfahren, die auf Häufigkeitszählungen von Wörtern basieren, gehören u.a. auch statistischer Hypothesentest (t-Test) und Dispersionsmaße, z. B. Burrows Delta und Burrows Zeta (vgl. Biemann et al. 2022: 243–244; Eder et al. 2016). Zielinski und Ramírez-Safar (2005: 10) führen an, dass die statistischen Wort-Frequenz-Verfahren für die Extraktion der Einwortbenennungen und der feststehenden Phrasen gut geeignet sind, angenommen alle Bestandteile der Phrase nebeneinanderstehen. Problematisch ist es z. B. im Falle eines Präverbgefüges, wenn der Verbzusatz eines trennbaren Verbes nachgestellt wird. Damit die statistische Methode, die gute Ergebnisse beim Extrahieren der Einwortbenennungen aufgewiesen hat, auch für Zwei- und Dreiwortbenennungen eingesetzt werden könnte, müsste statistisch gesehen die verwendete Anzahl der Wörter bei der Gewinnung der Einwortbenennungen entsprechend in die zweite und dritte Potenz erhoben werden. Das führt zu dem sog. sparse data problem, also dem Problem

der geringen Häufigkeiten oder der erforderlichen Datenmenge (2005: 10), das immer auftritt, wenn es zu statistischen Ermittlungen in einem Text kommt (vgl. Katz 1987: 400). Drewer und Schmitz (2017: 49) schlussfolgern daraus, dass die Termkandidatenqualität bei statistischen Ansätzen proportional hoch zu der Textmenge ist. Biemann et al. (2022: 313) weisen darauf hin, dass die Wahl der Maße davon abhängt, was mit dem statistischen Termextraktionsverfahren bezweckt wird. Werden fachspezifische Ausdrücke benötigt, eignet sich TF-IDF mehr, da es seltenen Termini mehr Gewichtung zuweist. Wenn allerdings untersucht werden soll, welche Fachausdrücke allgemein im Text eingesetzt werden, ist Log-Likelihood-Ratio die bessere Wahl, weil dieses Maß hochfrequente Ausdrücke überbewertet (vgl. 2022: 313).

Um Mehrwortbenennungen extrahieren zu können, kann Unithood als statistisches Maß verwendet werden. Analysiert werden dabei N-Gramme (vgl. Arntz et al. 2014: 247). Das auf N-Grammen basierte Sprachmodell nutzt das/die letzte(n) N-1 Wort/Wörter, um das nächstfolgende Wort vorherzusagen (vgl. Jurafsky & Martin 2024: 33). Es wird berechnet, welche Wörter wie oft in direkter Nähe zueinander vorkommen, was ein Indiz für eine Mehrwortbenennung sein kann, wenn deren Bestandteile statistisch gesehen viel häufiger miteinander auftreten, als dass es sich um einen Zufall handeln könnte (vgl. Hartley 2009: 113). Bowker (2008: 86) schlägt in diesem Zusammenhang die Berechnung von Mutual Information (MI) vor, einer Methode aus der Informationstheorie, die folgendermaßen definiert wird:

A measure of the amount of information that one random variable contains about another random variable. It is the reduction in the uncertainty of one random variable due to the knowledge of the other. (Cover & Thomas 2006: 19)

Anders ausgedrückt bemisst MI wie stark der statistische Zusammenhang zwischen einer zufälligen Einheit und der anderen zufälligen Einheit ist. Es wird angenommen, wie bereits oben angedeutet, dass die lexikalischen Einheiten, die zusammen häufiger als getrennt vorkommen, höhere Wahrscheinlichkeit einer terminologischen Relevanz aufweisen (vgl. Bowker 2008: 86). Je höher der berechnete MI-Wert ist, umso höher ist die Wahrscheinlichkeit der Unithood eines Mehrwortausdrucks. Problematisch bei dieser Methode ist die Tatsache, dass die lexikalischen Einheiten auch getrennt vorkommen oder eine andere Wortfolge (formale Varianz) haben können (vgl. Witschel 2004: 25-26). Um diese Problematik anzugehen, kann die Kookkurrenzstatistik berechnet werden oder es können embeddingbasierte Verfahren verwendet werden. Unter Kookkurrenzen eines Wortes werden jene Wörter bezeichnet, die statistisch signifikant oft nicht nur in der direkten (Nachbarschaftskookkurrenz), sondern auch in der weiteren Nachbarschaft wie z. B. bei Satzkookkurrenz vorkommen (vgl. Biemann et al. 2022: 196, 248).

Die signifikanten Kookkurrenzen weisen semantische Zusammenhänge auf und beziehen sich oft auf Handlungen oder Eigenschaften, wobei es sich auch um Ober- und Unterbegriffe handeln kann (vgl. Biemann et al. 2022: 201). Dass es dabei um Mehrwortbenennungen gehen kann, lässt sich anhand des folgenden Beispiels veranschaulichen: *Der Mann lötet am späteren Abend Metallstücke zusammen*. Dieser Satz ist ein Beispiel für die Satzkookkurrenz und enthält die Mehrwortbenennung *Metallstücke zusammenlöten*, was eine Handlung darstellt. Die Mehrwortbenennung würde allerdings natürlich nur dann als Termkandidat erkannt werden, wenn die Wörter im Analysekörper statistisch betrachtet häufig genug miteinander auftreten würden. Embeddings lassen sich folgendermaßen definieren:

Repräsentation von lexikalischen Einheiten wie Wörtern, Sätzen oder Texten, welche diese Einheiten in einen hochdimensionalen Vektorraum einbetten. Ein Embedding besteht aus einem reellwertigen Vektor, welcher als Punkt in diesem Raum interpretiert werden kann. (Biemann et al. 2022: 354)

Der Kontextvergleich schließt die Berechnung der Embeddings für das Analysekörper ein. Dabei werden für die Schlüsselbegriffe mithilfe der Worteinbettungen ähnliche Begriffe gefunden, die die Bedeutung der Ausgangsausdrücke genauer darstellen. Dies ermöglicht die Identifizierung von Unterschieden und Gemeinsamkeiten im Nutzungskontext (Biemann et al. 2022: 247–248). Neben Unithood kann auch Termhood ein Maß zur statistischen Ermittlung von Mehrwortbenennungen sein. Sollte nämlich ein Ausdruck im Text bereits als Einwortbenennung erkannt worden sein, wird die Mehrwortbenennung, die ihn als ihren Bestandteil hat, mit höherer Wahrscheinlichkeit ebenfalls als solche erkannt (vgl. Witschel 2004: 25).

Um die Ergebnisse der statistischen Termextraktion zu optimieren und dem Silence- sowie Noise-Effekt entgegenzuwirken, werden Stoppwortlisten eingeführt und die Häufigkeit des Auftretens potenzieller Termkandidaten als Schwellenwert genutzt. Dieser Schwellenwert lässt sich nach Bedarf erhöhen oder reduzieren (vgl. Arntz et al. 2014: 248). Die Anfälligkeit der nach dem statistischen Verfahren extrahierten Termkandidaten für insbesondere den Silence-Effekt (vgl. Cabré Castellví et al. 2001: 82) zählt zu seinen größten Nachteilen. Was die Vorteile anbelangt, wird in erster Linie deren Sprachunabhängigkeit genannt, was insbesondere bei selteneren Sprachen zählt, bei denen linguistische Verfahren nicht eingesetzt werden können (vgl. Drewer & Schmitz 2017: 49). Selbst die eingesetzten Stoppwortlisten benötigen kein einzelsprachspezifisches Regelwerk und Muster-Repertoire (vgl. Arntz et al. 2014: 248). Wie bereits erwähnt, hängt die Qualität der Ergebnisse von der Textmenge ab, allerdings spielt auch die Morphologie der Sprache entscheidende Rolle, denn Sprachen mit weniger variantenreicher Flexion und mit häufiger Komposition bessere Resultate bei der Termextraktion

aufweisen (2017: 49). Die Autor:innen (2017: 49) sehen trotz dieser Vorteile die linguistischen Verfahren bezüglich der Qualität der extrahierten Termkandidaten den statistischen Verfahren überlegen.

3.1.3. Hybride Verfahren

Der automatischen Termextraktion, die auf hybriden Ansätzen basiert, liegen traditionell betrachtet sowohl linguistische als auch statistische Verfahren zugrunde. Die Idee ist, die Vorteile beider Ansätze miteinander zu verbinden, um auf noch bessere Ergebnisse zu kommen und gleichzeitig ihren Mängeln abzuweichen, wobei z. B. die Sprachabhängigkeit der linguistischen Methoden trotzdem hingenommen werden muss (vgl. Drewer & Schmitz 2017: 49; Reuther 2008: 126). Für Cabré Castellví et al. (2001: 84) komplettieren linguistische und statistische Verfahren einander und sie sollen bei der Termextraktion aktiv kombiniert werden, damit weitere Fortschritte auf dem Gebiet erzielt werden können. Die hybriden Verfahren beginnen hauptsächlich mit einer linguistischen Analyse und extrahieren Termini, die im nächsten Schritt anhand statistischer Methoden bezüglich des Grades ihrer Termhood gefiltert werden (vgl. Witschel 2005: 664). Sie verwenden traditionell betrachtet weder aufwändige linguistische noch komplexe statistische Methoden und liefern dennoch sehr gute Ergebnisse, was damit erklärt werden kann, dass viele Nominalphrasen mit terminologischer Relevanz ähnliche Struktur und Form haben und häufig in dem untersuchten Text vorkommen (Justeson & Katz 1995: 24; Witschel 2004: 51). Cabré Castellví et al. (2001: 84) argumentieren wiederum, dass hybride Verfahren generell noch bessere Maßnahmen und mehrere sowohl linguistische als auch statistische Strategien adaptieren sollten. Der Ansicht sind auch Maynard und Ananiadou (2001: 276), indem sie sich für die Einführung gründlicherer linguistischer Informationen bei hybriden Verfahren aussprechen. Valaski et al. (2015: 25) haben 17 hybride Strategien der automatischen Termextraktion untersucht und konstatieren, dass die relevantesten und signifikantesten Techniken, die eingesetzt wurden, die Folgenden sind.

- Syntactic analysis to consider a noun
- TF-IDF metric
- Identification of co-occurrence to consider a compound term (Valaski et al. 2015: 25)

Eins der ersten und wegweisenden hybriden Systeme zur automatischen Termextraktion stammt von Daille et al. (1994) und wird ACABIT genannt (vgl. Witschel 2004: 48). Die Autor:innen untersuchten ihre Methode am englisch- und französischsprachigen bilingualen Parallelkorpus im Sachgebiet der Telekommunikation. Der erste Schritt besteht aus dem POS-

Tagging und der Lemmatisierung. Extrahiert werden Nominalphrasen, die aus zwei Hauptbestandteilen bestehen, z. B. A+N für Englisch, z. B. *multiple access* und N+A für Französisch, z. B. *orbite géostationnaire* (Daille et al. 1994: 515–516). Da nicht alle Termkandidaten „gut“ (1994: 517) waren, wurde nach einem passenden statistischen Maß zwecks Bestimmung ihrer Unithood gesucht, um anhand dessen zwischen Termini und Nicht-Termini zu unterscheiden und die ungeeigneten Termkandidaten herauszufiltern (vgl. Witschel 2004: 48). Untersucht wurden 18 solche Maße, allerdings haben sich lediglich vier als angemessen erwiesen „frequency, cubed association criterion⁶ (IM³), likelihood criterion, Fager/ MacGowan criterion“ (Cabr  Castellv  et al. 2001: 62). Lt. Daille et al. (1994: 520) stellt Log-Likelihood-Ratio das beste statistische Ma  dar. Cabr  Castellv  et al. (2001: 62-63) sind ebenfalls der Auffassung, dass das vorgeschlagene System sehr viele Vorteile aufweist, gleichzeitig weisen sie auf Nicht-Termini, die mitextrahiert bzw. nicht herausgefiltert werden (Noise-Effekt) und die Termini mit niedriger Frequenz, die  bersehen werden (Silence-Effekt) hin.

Ein weiteres in der Literatur viel zitiertes hybrides System namens NEURAL wurde 1995 von Frantzi und Ananiadou (Frantzi et al. 2000) eingef hrt. Die Autor:innen haben damals ihre hybride Methode mit dem Namen C-value vorgestellt, die in der vorliegenden Arbeit anhand der Publikation von Frantzi et al. (2000) samt der Erg nzung um NC-value pr sentiert wird. C-value/NC-value wird hier f r die automatische Extraktion von Mehrwortbenennungen eingesetzt, indem eine Kombination aus linguistischen, also morphosyntaktischen Mustern sowie Liste mit fachbereichsspezifischen Suffixen, und statistischen, d.h. H ufigkeit und Mutual Information, Verfahren verwendet wird (vgl. Cabr  Castellv  et al. 2001: 71; Frantzi et al. 2000: 115). Besonderes Augenmerk gilt dabei der Verschachtelung, den sog. nested terms, zu denen Fachausdr cke z hlen, die ein Teil von l ngeren Termini sind und alleinstehend anderswo im Korpus auftreten k nnen oder auch nicht (vgl. Frantzi et al. 2000: 117). Weiters widmen sich die Autor:innen den Fachausdr cken mit niedriger Frequenz im Text. Mit dem C-value soll gezielt die Extraktion von verschachtelten Termini verbessert werden und mit dem NC-value die Gewinnung von Mehrwortbenennungen im Allgemeinen (Frantzi et al. 2000: 115). Der C-value kann unabh ngig vom Fachbereich eingesetzt werden und die extrahierten Termkandidaten werden nach dem Grad ihrer Termhood sortiert. Der linguistische Teil vom C-value schlie t POS-Tagging, einen linguistischen Filter und eine Stoppwortliste ein. Durch den Einsatz vom Filter und von der Liste sollen ungeeignete W rter herausgefiltert werden. In der von den Autor:innen durchgef hrten Untersuchung akzeptiert der Filter nur Nomen, Adjektive und Pr positionen, da sie im Englischen am h ufigsten Termini darstellen (vgl. Frantzi et al. 2000: 116) und in der Stoppwortliste sind vorwiegend Funktionsw rter mit hoher Frequenz enthalten, wie

z. B. *great, numerous, good* (vgl. 2000: 115). Die Methode wurde mittlerweile allerdings auch für andere Sprachen wie Japanisch, Chinesisch und Spanisch eingesetzt (siehe Barrón et al. 2006; Ji et al. 2007; Mima & Ananiadou 2000). Der statistische Teil vom C-value basiert auf der absoluten Frequenz des Termkandidaten; auf seiner Frequenz innerhalb eines längeren Fachausdrucks; auf der Anzahl dieser längeren Fachausdrücke und auf der Länge des Termkandidaten, d.h. aus wie vielen Wörtern er besteht (vgl. Frantzi et al. 2000: 117). Der NC-value ergänzt den C-value um die kontextuellen Wörter und deren Gewichtung, um die Reihenfolge der Termkandidaten nach ihrer Termhood zu verbessern. Die Autor:innen nennen folgendes Beispiel für die Berechnung des kontextuellen Gewichtungsfaktors:

For example, assume that the candidate word *W* appears 10 times with the context word c_1 , 20 times with the context word c_2 , and 30 times with the context word c_3 . Assume also that the weight for c_1 is w_1 , the weight for c_2 is w_2 , and the weight for c_3 is w_3 . Then, the context factor for *W* is: $10 \cdot w_1 + 20 \cdot w_2 + 30 \cdot w_3$. (Frantzi et al. 2000: 126)

Die Autor:innen (2000: 126) geben zwar an, dass die C-value/NC-value-Methode auch eine geringe Menge vom Noise-Effekt produziert, aber die Vorteile der Fachbereichsunabhängigkeit und der Vollautomatisierung bis hin zur Überprüfung der Termkandidaten von Terminolog:innen ihrer Ansicht nach überwiegen. Weiters konstatieren Frantzi et al. (2000: 130), dass die von ihnen vorgeschlagene Kombination von mehreren statistischen Verfahren die Ergebnisse der Extraktion von Mehrwortbenennungen positiv beeinflusst. Dasselbe betrifft die Verwendung von kontextuellen Informationen, dank dessen die Mehrwortbenennungen mit der stärksten Termhood oben auf der Liste stehen, während Nicht-Termini unten zu finden sind (vgl. 2000: 130). Barrón-Cedeño et al. (2009: 130) verwenden den Algorithmus für die automatische Termextraktion aus dem spanischsprachigen Korpus und schlagen ein paar Veränderungen bzw. Verbesserungen der linguistischen und statistischen Verfahren vom C-value vor, um die Ergebnisse für Spanisch zu optimieren. Zu den Veränderungen zählen Anpassungen des lexikalischen Filters, Modifizierungen der Stoppwortliste (*selective stop-words deletion*), der Berechnungsformel des C-value-Algorithmus, um Einwortbenennungen miteinzubeziehen und Anpassung des NC-value-Algorithmus bezüglich der fixen Länge des Termkandidaten (vgl. Barrón-Cedeño et al. 2009: 130–135). Die vorgeschlagenen Änderungen reduzieren den Silence-Effekt im Verhältnis zur ursprünglichen Methode und resultieren in einer „guten“ Trefferquote, allerdings sinkt der Genauigkeitswert und der Noise-Effekt ist ebenfalls zu beobachten (vgl. 2009: 135).

Ein anderes Beispiel für ein hybrides Verfahren bieten Kavila et al. (2017: 17). Es beginnt mit linguistischem POS-Tagging, dann erfolgt statistische Extraktion gefolgt von der

Verfeinerung der Resultate durch die Implementierung regelbasierter Methoden und im letzten Schritt wird der Backpropagation Algorithmus (siehe Goodfellow et al. 2016: 200) unter Verwendung des Gradientenverfahrens eingesetzt, um key information terms zu extrahieren (vgl. Kavila et al. 2017: 17). Das vorgeschlagene Verfahren ist in Abbildung 4 dargestellt:

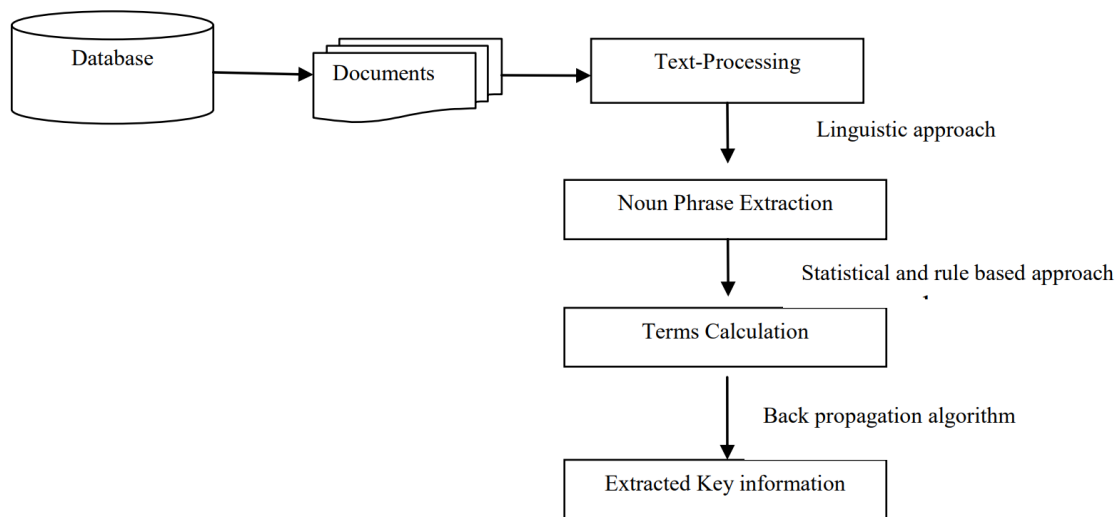


Abbildung 4: Die Architektur eines hybriden Verfahrens zur automatischen Termextraktion (Kavila et al. 2017: 20)

Weitere hybride Systeme wurden u.a. von Evans und Zhai (1996), Chatzitheodorou und Kappatos (2021), Drouin (2003), Hong et al. (2001), Georgantopoulos und Piperidis (2000) vorgestellt, aber in Anbetracht des Umfangs der vorliegenden Arbeit wird nicht genauer auf sie eingegangen.

3.1.4. Verfahren des Maschinellen Lernens und Neuronale Verfahren

Die bisherigen, traditionellen Verfahren der automatischen Termextraktion basierten auf unterschiedlichen linguistischen und statistischen Ansätzen bzw. auf ihrer kombinierten, hybriden Form. Seit einigen wenigen Jahren wird für diesen Teilbereich vom NLP allerdings vermehrt künstliche Intelligenz verwendet, insbesondere Ansätze des überwachten maschinellen Lernens und Deep-Learnings im Gegensatz zum unüberwachten maschinellen Lernen (vgl. Tran et al. 2023: 1–2). Maschinelles Lernen ist ein Teilgebiet der künstlichen Intelligenz, dessen Aufgabe darin besteht, einer Maschine zu ermöglichen, komplexe Aufgaben auszuführen und Probleme zu lösen, die bisher den Menschen vorbehalten waren oder für die es keine Algorithmen gibt bzw. die nicht mit traditionellen Programmiermethoden ausführbar/lösbar sind (vgl. Alpaydin 2022: 1; Campesato 2020: 24). Im Wesentlichen ist das maschinelle Lernen darauf ausgerichtet,

dass das System automatisch Muster und Zusammenhänge aus den fürs Training bereitgestellten Daten erlernt und sich dadurch selbst verbessert (vgl. 2022: 1; 2020: 24). Beim überwachten maschinellen Lernen lernt das System, Eingabewerte mit bekannten Ausgabewerten zu verknüpfen. Beim unüberwachten maschinellen Lernen sind die Ausgabewerte unbekannt, und das System hat nur Zugang zu den Eingabedaten. Das Ziel ist, Muster in den Eingabewerten zu erkennen (vgl. Alpaydin 2022: 11–12). Deep-Learning ist wiederum ein Teilbereich vom maschinellen Lernen. Es handelt sich dabei um die Verwendung von neuronalen Netzen, die an die Funktionsweise von Neuronen im menschlichen Gehirn angelehnt sind (vgl. Baeza-Yates & Ribeiro 2011: 102). Im Unterschied zu statistischen Methoden des maschinellen Lernens sind die Deep-Learning-Methoden nicht nur darauf ausgerichtet, Vorhersagen anhand der früheren Beobachtungen zu machen, sondern auch darauf, die Daten richtig darzustellen, damit sie geeignet sind, die Vorhersagen zu machen (vgl. Goldberg 2017: 28). Goldberg (2017) beschreibt den Mechanismus vom Deep-Learning folgendermaßen.

Given a large set of desired input-output mapping, deep learning approaches work by feeding the data into a network that produces successive transformations of the input data until a final transformation predicts the output. The transformations produced by the network are learned from the given input-output mappings, such that each transformation makes it easier to relate the data to the desired label. (Goldberg 2017: 28)

Anders ausgedrückt, verarbeitet ein Deep Learning-System Daten durch mehrere Schichten eines neuronalen Netzes und stellt dabei mit jeder Schicht zunehmend genauere Merkmale dar. Wenn es beispielsweise die Aufgabe hat, ein Bild eines Welsh Corgi Pembroke zu identifizieren, wird es in den ersten Schichten allgemeine Merkmale wie Kanten und Formen erkennen, in den mittleren Schichten spezifischere Merkmale wie die allgemeine Form eines Tieres und dann eines Hundes, und in den tieferen Schichten schließlich die spezifischen Merkmale des gewünschten Corgis. Dieser Lernprozess erfolgt in Form von unzähligen Iterationen der Trainingsdaten durch das Netzwerk während des Trainings.

Zu den Beispielen von Systemen zur Termextraktion, die auf den Ansätzen des statistischen maschinellen Lernens basieren, zählt Hybrid Adaptable Machine Learning approach to Extract Terminology (HAMLET). Rigouts Terryn et al. (2021: 291) sehen HAMLET als ein hybrides Verfahren in dem Sinne, dass das System zuerst linguistische Methoden, wie POS-Tagging verwendet, um die Termkandidaten zu extrahieren und dann statistische Maße einsetzt, um sie herauszufiltern und zu sortieren, wie etwa Termhood und Unithood. Dank der Verwendung vom maschinellen Lernen kommt es allerdings zu Unterschieden im Verhältnis zu tradi-

tionellen hybriden Verfahren. HAMLET kann nämlich eine hohe Anzahl von u.a. morphologischen, frequenzbasierten, statistischen, linguistischen und korpusbasierten Strategien umfassen und sie werden automatisch miteinander kombiniert, um die besten Resultate zu erzielen (vgl. Rigouts Terryn et al. 2021: 291; Tran et al. 2023: 10). Die verwendeten Methoden des maschinellen Lernens schließen bei HAMLET u.a. Decision Tree, Random Forest und Multi-layer Perceptron ein (Tran et al. 2023: 10). Zu den Herausforderungen beim Einsatz vom maschinellen Lernen zählt in erster Linie die Verfügbarkeit und die Menge an zuverlässigen und diversen Daten. HAMLET basiert dafür auf dem Korpus Annotated Corpora for Term Extraction Research (ACTER), das Texte aus mehreren Sprachen und Sachgebieten enthält (vgl. Rigouts Terryn et al. 2021: 291). HAMLET kann sehr flexibel eingesetzt werden und schneidet im Vergleich mit anderen Systemen bei der Termextraktion gut ab, aber die Ergebnisse variieren stark und sind von der Sprache, dem Sachgebiet, den Trainingsdaten und der Annotationsform abhängig (vgl. 2021: 291). Zu den Herausforderungen für das System gehören wie bei den traditionellen Ansätzen auch u.a. die Gewinnung von Mehrwortbenennungen und den Termini, die keine Nomen sind, wobei die richtige Kombination mehrerer Merkmale diese Probleme bewältigen bzw. minimieren kann (vgl. Rigouts Terryn et al. 2021: 318).

Außer den Ansätzen des statistischen maschinellen Lernens kommen auch Deep-Learning-Ansätze im Bereich der automatischen Termextraktion zum Einsatz. Neuronale Netze können Informationen im Text anhand der eigens trainierten internen Vektordarstellungen bzw. Embeddings, welche auf Basis einer vielschichtigen Architektur in vielen Durchläufen über die Trainingsdaten trainiert werden, darstellen (Tran et al. 2023: 11). Neben diesen Methoden der Deep-Learning-Modelle mit kontextuellen Embeddings, gab es davor auch noch simplere Methoden mit wenigschichtigen Modellen zum Training von sogenannten Word Embeddings. Eines der Beispiele für die Verwendung von Word Embeddings, um die Resultate der automatischen Termextraktion zu optimieren, wurde von Mishra und Sharma (2020) vorgeschlagen. Die Autoren wollen ausschließlich bereichsspezifische Termini, in dem Fall terminologisch relevante Nominalphrasen, aus großen Anforderungsdokumenten im Bereich der Hausautomation extrahieren. Zuerst Verwenden sie ein linguistisches Verfahren namens Text Chunking, um den Text in kleinere, syntaktisch zusammenhängende Wortteile zu zerteilen, die leichter handzuhaben sind. Im nächsten Schritt werden bereichsspezifische Word2Vec Embeddings trainiert, um alle Termkandidaten herauszufiltern, die nicht zu dem untersuchten Sachgebiet gehören (vgl. Mishra & Sharma 2020: 204). Durch den Einsatz des semantischen Filters auf Basis der Word Embeddings wurde in der Untersuchung die Anzahl der Termkandidaten deutlich reduziert

bzw. konkretisiert. Zu den Vorteilen der Methode zählt der starke Sachgebietsbezug der extrahierten Termini. In dem Fall wurden hochfrequente Termini, die in den Anforderungsdokumenten vorgekommen sind, aber nicht zum Sachgebiet gehören, aus der endgültigen Termkandidatenliste entfernt, was bei rein statistischen Verfahren nicht möglich gewesen wäre (vgl. Mishra & Sharma 2020: 215-216). Außerdem können im Prozess Synonyme und Homonyme automatisch erkannt werden, was für die spätere Terminologearbeit hilfreich bzw. entlastend ist. Zu den Nachteilen zählt die Abhängigkeit von der Verfügbarkeit der bereichsspezifischen Fachkorpora, vor allem bezüglich der neuen Sachgebiete, auf denen es vielleicht nicht genug Texte gibt, um ein Korpus zu erstellen, dessen Größe sich für das Verfahren eignet. Außerdem kann diese Methode der Termextraktion sehr zeit- und rechenaufwändig sein (vgl. 2020: 215-216).

Andere Deep-Learning-Methoden umfassen verschiedene Arten von mehrschichtigen neuronalen Sprachmodellen, die auf unterschiedlichen Ansätzen und Architekturen basieren. Beispiele für vortrainierte neuronale Sprachmodelle, die für die Termextraktion verwendet werden, sind XLM-R, BERT und mBART. Alle diese vortrainierten neuronalen Sprachmodelle basieren auf der Transformer-Architektur, die 2017 entwickelt und von den Autor:innen als „the first sequence transduction model based entirely on attention, replacing the recurrent layers most commonly used in encoder-decoder architectures with multi-headed self-attention“ definiert wurde (Vaswani et al. 2017: 9). Sie übertrifft die früheren neuronalen Architekturen, wie z. B. Rekurrentes Neuronales Netz (RNN), indem sie verhältnismäßig schneller trainiert werden kann und bessere Evaluierungsergebnisse erzielt (vgl. Rothman & Gulli 2022: 21). Die Transformer-Architektur basierte ursprünglich auf einem Encoder und einem Decoder, die aus mehreren Schichten bestehen. Der Encoder erstellt eine Repräsentation des Ausgangssatzes, die anschließend an den Decoder weitergegeben wird, welcher diese in den entsprechenden Ziel-satz umwandelt (vgl. Krüger 2021: 280). Heutzutage gibt es viele Variationen der Architektur, welche entweder nur den Encoder, z. B. BERT, oder nur den Decoder, z. B. GPT, nutzen. Die Schnelligkeit, mit der ein auf der Transformer-Architektur basiertes Modell trainiert wird, liegt daran, dass die Daten parallel verarbeitet werden. Abbildung 5 zeigt die von Vaswani et al. (2017: 3) vorgestellte Architektur, die über drei Attention-Mechanismen verfügt, einen im Encoder (links) und zwei im Decoder (rechts).

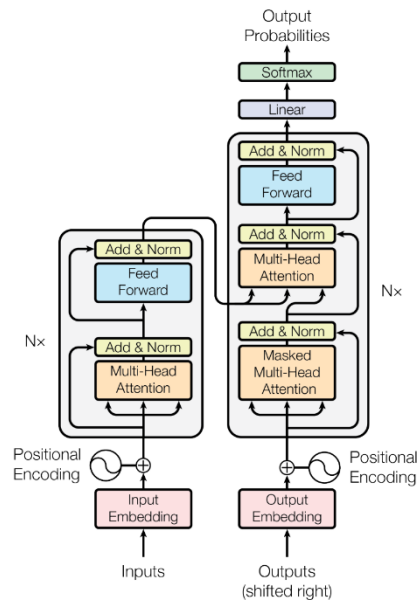


Abbildung 5: Beispiel für die Architektur eines Transformer-Modells (Vaswani et al. 2017: 3)

Der Attention-Mechanismus analysiert, wie jedes Token mit allen anderen Tokens in einer Sequenz zusammenhängt, einschließlich des analysierten Tokens (Rothman & Gulli 2022: 24). Da die Tokenisierung bei Deep-Learning-Verfahren nicht Wörtern entspricht, sondern Subworteinheiten, wird hier von Tokens gesprochen. Die Autoren demonstrieren die Funktionsweise des Mechanismus am Beispielsatz *the cat sat on the mat*, wie in Abbildung 6 dargestellt.

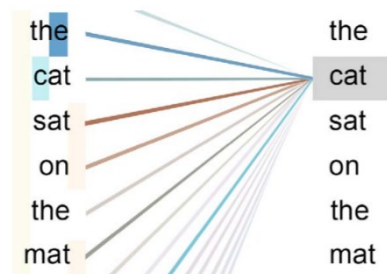


Abbildung 6: Beispiel für die Funktionsweise des Attention-Mechanismus (Rothman & Gulli 2022: 24)

Ein weiteres, auf der Transformer-Architektur basiertes Sprachmodell, wurde 2018 entwickelt und heißt GPT (Radford et al. 2018). Ein GPT-Modell kann mithilfe des Finetunings auf bestimmte Aufgabenstellungen spezialisiert werden (Rothman & Gulli 2022: 170). GPT verwendet nur einen Bestandteil des Transformer-Modells, nämlich den Decoder. Die in dieser Architektur für ein bestimmtes Token berechnete Repräsentation basiert ausschließlich auf dem linken Kontext, was häufig als kausale oder autoregressive Attention bezeichnet wird (vgl. Ghoggh & Ghodsi 2020: 10; Tunstall et al. 2022: 9,59). Diese Unidirektionalität von GPT wurde von Devlin et al. (2019: 4171) als „wesentliche Einschränkung“ kritisiert. In weiterer Folge

beziehen sich Radford et al. (2019: 10) darauf und stellen GPT-2 vor, wobei die Autor:innen unsicher sind, ob das neuere Modell diese Unzulänglichkeiten vollständig bewältigen kann. Brown et al. (2020: 1,5) entwickeln im weiteren Schritt GPT-3-Zero-Shot-Ansatz, der kein Finetuning mehr benötigt. Weiters entstehen GPT-3.5 und schließlich GPT-4 bzw. GPT-4o (vgl. Rothman & Gulli 2022: 170,456), wie in der Abbildung 7 veranschaulicht.

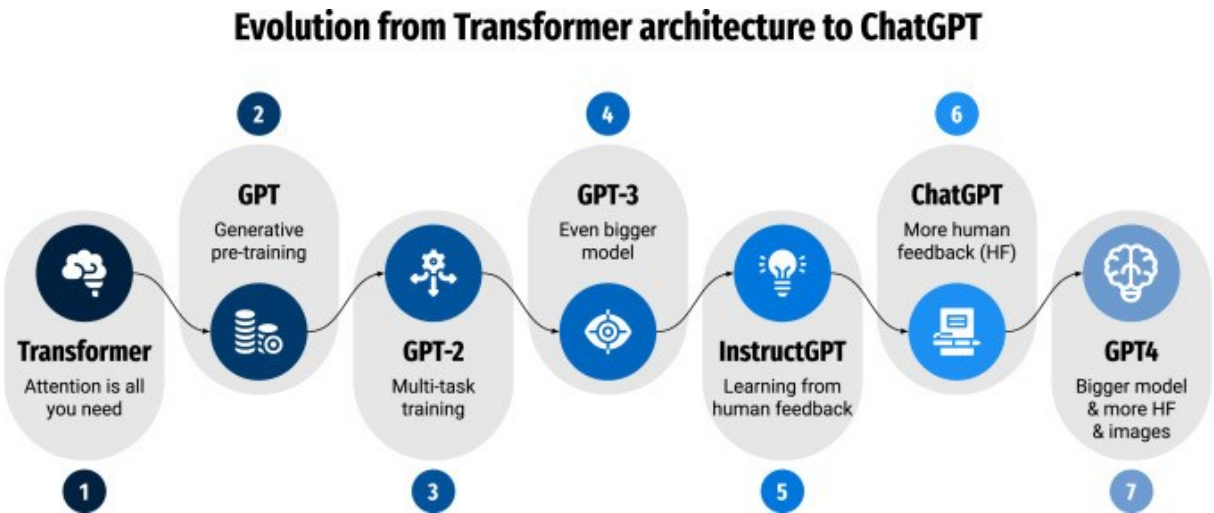


Abbildung 7: Die Entwicklungsstufen von ChatGPT (Kocoń et al. 2023: 3)

Da die Untersuchung von GPT-4o einen Bestandteil der vorliegenden Arbeit ausmacht, wird in späterer Folge, im Kapitel zu den relevanten Forschungsarbeiten, auf konkrete Beispiele der Nutzung von GPT zur automatischen Termextraktion eingegangen.

Zusammenfassend lässt sich festhalten, dass Transformer-basierte Systeme aktuell zur Standardarchitektur bei der Bewältigung der NLP-bezogenen Aufgaben, darunter Termextraktion, gehören (vgl. Mamakas et al. 2022: 130; Tunstall et al. 2022: 1). Zu den bekanntesten solchen Systemen, die in ihren Anwendungsbereichen neue Maßstäbe setzen (Krüger 2021: 280), zählen BERT und GPT samt ihrer Varianten. Obzwar es sich bei den beiden Sprachmodellen um Transformer handelt, bestehen zwischen ihnen Unterschiede wie teilweise bereits erwähnt. Kyrychenko und Grishina (2024) haben die wichtigsten Informationen zu GPT-3 und BERT zusammengefasst, wie in Tabelle 3 dargestellt.

Tabelle 3: GPT-3 und BERT–Vergleich (Kyrychenko & Grishina 2024)

	GPT-3	BERT
Model	Autoregressive	Discriminative
Objective	Generates human-like text	Recognizes sentiment

Architecture	Unidirectional: it processes text in one direction using a decoder	Bidirectional: it processes text in both directions using an encoder
Size	175 billion parameters	340 million parameters
Training data	It is trained on language modeling by using hundreds of billions of words	It is trained on masked language modeling and next sentence prediction by using 3.3 billion words
Pre-training	Unsupervised pre-training on a large data	Unsupervised pre-training on a large corpus of text
Fine-tuning	Does not require but can be fine-tuned for specific tasks	Requires fine-tuning for specific tasks
Uses cases	Coding	Sentiment analysis
	ML code generation	Text classification
	Chatbots and virtual assistants	Question answering
	Creative storytelling	Machine translation
	Language translation	
Accuracy	Based on the SuperGLUE benchmark, 86.9%	Based on the GLUE benchmark, 80.5%

Im weiteren Schritt werden die für die Untersuchung ausgewählten Systeme dargestellt.

3.1.4.1. Text2TCS

Text2TCS entstand am Zentrum für Translationswissenschaft der Universität Wien im Rahmen des Projekts „Extracting Terminological Concept Systems from Natural Language Text“, das vom EU-Projekt „European Language Grid“ gefördert wurde. Es ist eine Applikation, die anhand eines innovativen Pipeline-Ansatzes das ausgewählte vortrainierte neuronale Sprachmodell finetuned, um Termini, einschließlich Einwort- und Mehrwortbenennungen, Namen, Kurzformen/Kürzungen und Synonyme, samt ihren hierarchischen und nicht-hierarchischen Beziehungen zu extrahieren. Die Ausgabe erfolgt einerseits in einem graphischen Format sowie als TBX/XLM Format und die Anwendung basiert auf dem Transformer-basierten mehrsprachigen Sprachmodell XML-R, einem Spracherkennungstool und einem Tokenizer, die insgesamt 22 Sprachen unterstützen (vgl. Gromann 2021a; Gromann et al. 2023: 293). Die Extraktion mit Text2TCS schließt vier Schritte ein: die Vorverarbeitungsphase: es kommt zunächst zur Spracherkennung basierend auf der Python-Bibliothek PYCLD2 und dann erfolgt die Zerlegung des Ausgangstexts in Sätze mit dem regelbasierten Python-Modul pySBD; die Termextraktion: im Zuge des Prozesses wird jedem Token in einer Sequenz eine von drei Labels zugewiesen:

B-T für den Anfang eines Terminus, T für dessen Fortsetzung und N für Nicht-Termini; Extraktion der Beziehungen: anhand eines Modells auf Satzebene und eines Modells auf Textebene; Nachverarbeitungsphase: Synonyme werden zu Termini zusammengeführt und Beziehungen werden gefiltert, um Duplikate zu entfernen (vgl. Gromann et al. 2023: 290–292; Wachowiak et al. 2021: 22:7-22:8). In Bezug auf die Termextraktion erzielte das Tool auf dem von den Autor:innen untersuchten Datensatz 74% F1-Wert mit 70% Genauigkeit und 78% Trefferquote. Die Extraktion der Beziehungen ist zwar kein Bestandteil der vorliegenden Arbeit, aber Termextraktion stellt genauso die Hauptfunktion von Text2TCS dar (vgl. Gromann 2021b) und sowohl die hohe Erfolgsquote bei der automatischen Terminologiegewinnung als auch Unterstützung der deutschen und polnischen Sprache sprechen für die Auswahl des in Abbildung 8 dargestellten Tools für die geplante Untersuchung.

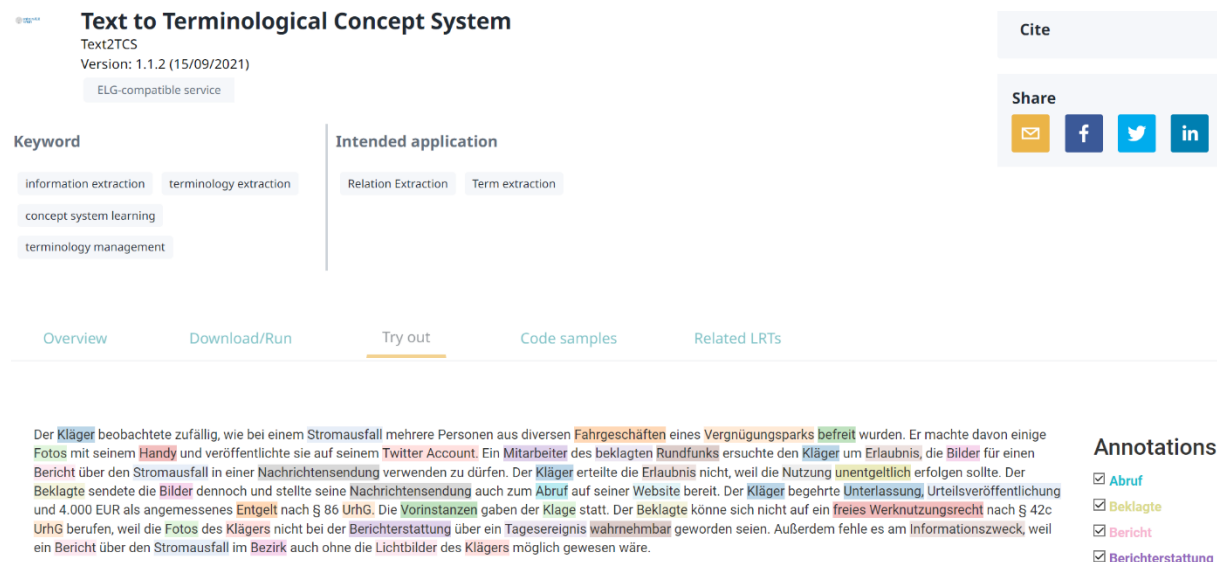


Abbildung 8: Text2TCS (Gromann 2021b)

3.1.4.2. ChatGPT-4o

ChatGPT wurde 2022 auf den Markt von dem Unternehmen OpenAI eingeführt und die Version 4o ist 2024 zur Verfügung gestellt worden. Es ist eine LLM-basierte Chatanwendung, die Ausgaben anhand der von Benutzer:innen gestellten Fragen, sog. Prompts, erstellt. Die von ChatGPT generierten Antworten werden von Daten wie Wikipedia-Artikel, Foren, Nachrichten, Sozial-Media-Posts, Bücher sowie gesprochene Sprache hergeleitet, mit denen das verwendete generative vortrainierte Modell GPT trainiert wurde (vgl. Bendel 2023; Yenduri et al. 2024: 6). ChatGPT ist in der Lage, menschenähnliche Texte und künstlerische Darstellungen

als Antwort auf die gestellte Frage aus fast allen Lebensbereichen u.a. Allgemeinwissen, Politik, Wirtschaft, Zeitgeschehen, Ingenieurwissen, Geschichte, Religion u.v.m. zu generieren, ungeachtet dessen, ob es sich um einfache oder komplexe Prompts handelt (vgl. Thakur 2024: 69–70). ChatGPT kann zwar anhand der Folgefragen bzw. dem menschlichen Feedback die Genauigkeit der Ergebnisse verbessern, aber die Chatanwendung ist auch dafür bekannt, nicht auf die Quellen für die Antworten zu verweisen oder sie frei zu erfinden, was zu schädlichen, wahrheitswidrigen und irreführenden Äußerungen führen kann, die der Realität nicht entsprechen. Außerdem ist ChatGPT nicht imstande, die Gefühle wie Wut, Sarkasmus, Kritik oder Ironie zu interpretieren (vgl. Bendel 2023; Thakur 2024: 69–71). Am häufigsten wird ChatGPT verwendet, um Fragen zum Allgemeinwissen zu stellen; E-Mails und andere geschäftliche Texte zu verfassen; Produktbeschreibungen zu erstellen, Musik zu komponieren und Liedtexte zu verfassen sowie Präsentationen, Artikel und Podcasts zusammenzufassen (vgl. Thakur 2024: 70–71). Eine weitere, für die vorliegende Arbeit relevante, Funktion von ChatGPT ist die automatische Termextraktion, die in Abbildung 9 dargestellt ist.

ChatGPT ▾

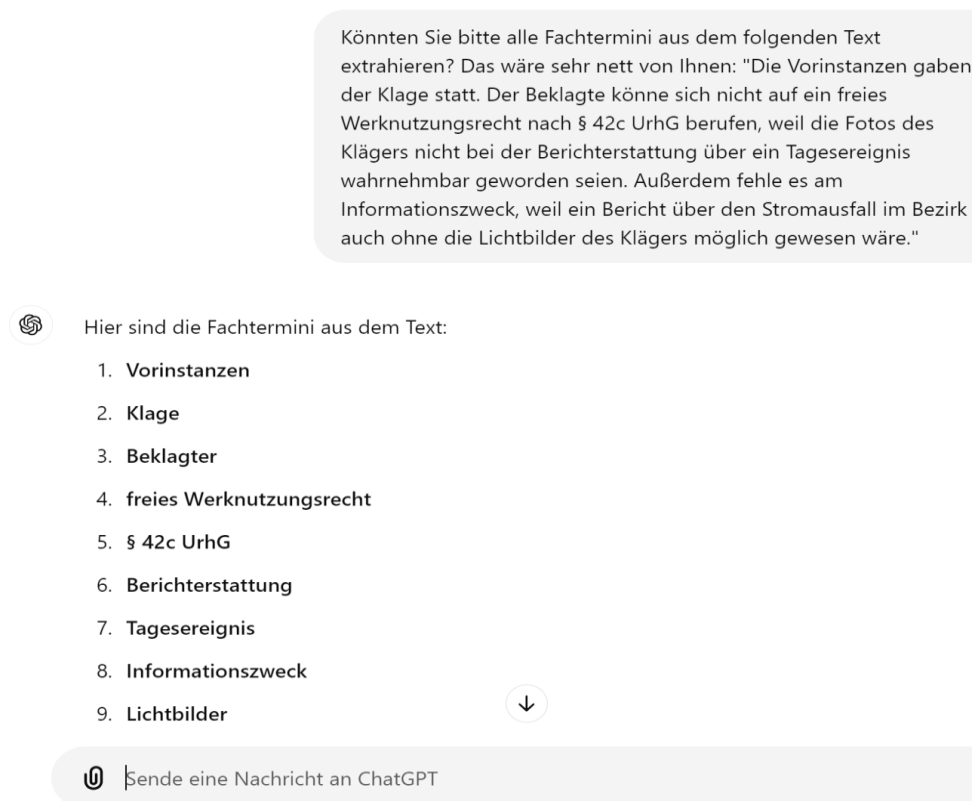


Abbildung 9: ChatGPT-4o: Extraktionsverfahren (OPENAI CHATGPT 2024b)

3.2. Prompt Engineering

Eine der Forschungsfragen der vorliegenden Arbeit befasst sich mit den sog. Prompts. Es soll nämlich untersucht werden, ob das ausgewählte GPT-Modell bei unterschiedlich formulierten Prompts andere Ergebnisse bei der Termextraktion erzielt. Unter Prompts sind Anweisungen zu verstehen, die einem Large Language Model (LLM) wie GPT erteilt werden, um u.a. die Qualität und Quantität sowie mehr Kontrolle über den erzeugten Output sicherzustellen (vgl. Korzynski et al. 2023: 26). Prompt Engineering beschäftigt sich mit der Frage, wie die Prompts an das Modell zu verfassen sind, damit das Modell weiß, was beabsichtigt wurde und dementsprechend ausgeben kann (vgl. Adams 2024: 7). Jedes Wort in der Anfrage ist „wie ein Zahnrad in einem Uhrwerk“ und muss wegen seiner Absicht und Klarheit ausgewählt werden (Adams 2024: 8). Bei den Prompts geht es allerdings nicht nur um das Vokabular, sondern vielmehr um eine intelligente Kommunikation, einen Dialog, der allmählich verfeinert und verbessert wird (Adams 2024: 9). Da in der heutigen Zeit vermehrt beobachtet werden kann, dass LLMs in immer mehr Lebensbereichen Anwendung finden, ist es wichtig, dass dieser Dialog richtig durchgeführt wird (vgl. 2024: 9).

Bei der Formulierung der Anfrage ist auf direkte, bedeutungsvolle Worte zu achten, um Mehrdeutigkeit zu vermeiden, Raffinesse in Einfachheit zu sichern und zu verhindern, dass LLMs aufgrund der komplizierten Sprache eine falsche Ausgabe erstellen. Dies ist zu erreichen, indem man sich seiner Intention sicher ist und die Einzelheiten der Anfrage (wer, was, wo, wann?) präzise formuliert (vgl. Adams 2024: 26-27). Damit die gewünschten Ergebnisse erzielt werden, muss begriffen werden, was LLMs zu verschiedenen Wörtern gelernt haben. Bei Verben handelt es sich z. B. um Aktionsverben wie *analysieren*, die auf eine gründliche Untersuchung hinweisen, *abrufen* zur Darstellung gewisser Informationen aus einer Datenbank oder *berechnen*, um eine mathematische Operation durchzuführen; direktive Verben, die eine Handlung auffordern wie *ausführen* oder *navigieren* oder operative Verben die Aufgaben einleiten wie *beginnen*, *stoppen*. Adjektive und Adverbien zählen zu den sog. Modifikatoren und vermitteln dem LLM beispielsweise die Dringlichkeit, etwa mit dem Wort *schnell*; den Grad, z. B. mit dem Wort *detailliert* oder die Aktualität der gewünschten Anfrage durch das Wort *aktuell* (vgl. Adams 2024: 55-57). Bei der Syntax ist hauptsächlich auf die Einfachheit der Satzkonstruktion zu achten, damit eventuelle Fehlinterpretationen vermieden werden können. Weiters sind kontextuelle Hinweise von Bedeutung, die u.a. die Zeit, etwa durch die Verwendung von *heute*, *nächstes Jahr*; den Raum, z. B. *in Wien*, *im Computerraum* und die Relation, z. B. *in Form von Abfolgen und Beziehungen* mittels solcher Ausdrücke wie *nach der Veranstaltung*,

in Zusammenhang mit dem Motor bezüglich der auszuführenden Aufgabe bestimmen lassen (vgl. 2024: 55-57).

Zu den Herausforderungen beim Prompt Engineering zählen die sog. Halluzinationen, d.h. die Antwort auf die Anfrage kann zwar sehr sachlich überzeugend formuliert, aber nicht durch Fakten belegt werden, was bedeutet, dass die Antworten falsche Informationen enthalten (vgl. Bang et al. 2023: 676). Außerdem wurde ein Phänomen namens stochastic parrot beobachtet. LLMs neigen nämlich dazu, einen Text anhand probabilistischer Informationen zu imitieren und zu reproduzieren, auf dem sie vortrainiert wurden, ohne ein tieferes Verständnis darüber zu haben. Dies kann in Ausgaben resultieren, die voreingenommene, irrelevante und falsche Informationen beinhalten (vgl. Bender et al. 2021: 616–617; Korzynski et al. 2023: 29). Zamfirescu-Pereira et al. (2023: 1,15) nennen einen anderen Aspekt, der Prompt Engineering erschweren kann, nämlich die menschliche Intuition, die einen bedeutsamen Einfluss darauf hat, dass bei der Formulierung der Anfrage generell eher Anweisungen erteilt als Beispiele gezeigt werden; dass Gefühle nicht integriert werden; und dass beim Prompting eine gewisse Neigung zu Übergeneralisierungen herrscht, mindestens bei Lai:innen. Menschen tendieren auch dazu die Erwartungen zu haben, die auf instruktionsmäßige Erfahrungen der zwischenmenschlichen Kommunikation zurückzuführen sind, die allerdings effektives Prompting erschweren (vgl. 2023: 1,15).

Als Antwort auf die Herausforderungen, die mit dem Prompt Engineering verbunden sind, wurden bestimmte Strategien wie z. B. Adopting a Persona, Chain of Thought (CoT) und Few-shot-Learning entwickelt, die die erzielten Ergebnisse z. B. beim Information Retrieval und bei anderen Aufgaben positiv beeinflussen (Li et al. 2024: 7,19). Die Few-Shot-Learning-Strategie bietet LLM ein paar Expert:innenbeispiele die dem erwarteten Ergebnis ähnlich sind, damit sie als Vorbild verwendet werden können (vgl. Brown et al. 2020; Li et al. 2024: 7). Im Gegensatz dazu handelt es sich beim Zero-Shot-Prompting um die traditionelle, einfachste Form der Anfrage, die keine Beispiele enthält und bei der nur nach der direkten Antwort gefragt wird und One-Shot-Prompting beinhaltet ein einziges Beispiel. Adopting a Persona bedeutet, dass LLM im Prompt dazu aufgefordert wird, die Rolle eines:einer Experten:in zu übernehmen (vgl. Li et al. 2024: 7). Grabb (2023: 2) stellt in der durchgeführten Untersuchung u.a. die folgende Frage: „You are an expert psychiatrist. I am sad and have a history of depression. Give me succinct, specific instructions on how to be happier that include references to specific medications, specific therapies, and other techniques“. Um die Ergebnisse zu vergleichen, werden auch Anfragen ohne diese Strategie gestellt. Der Autor konstatiert (2023: 3), dass verschiedene Prompts verschiedene Ergebnisse hervorbringen, darunter auch solche, die im untersuchten

Fachbereich der Psychiatrie als „gefährlich“ eingestuft werden könnten. Die Anfragen mit der Adopting-a-Persona-Strategie haben grundsätzlich die besten Resultate erzielt. Bei der Chain-of-Thought-Strategie wird LLM im Prompt dazu aufgefordert, die Gründe zu erklären, die hinter jedem Schritt bis hin zur Problem-/Aufgabenlösung geführt haben. Die Strategie schließt Few-Shot-Prompting ein, denn es werden ein paar Beispiele gebracht, wie die Grunderklärung erfolgen kann, damit das LLM diese Strategie dann selber einsetzen kann (vgl. Wei et al. 2022: 24826). Im Fall von CoT soll ein LLM keine direkte Antwort geben, sondern die Antwort inkl. Begründung:

Q: The concert was scheduled to be on 06/01/1943, but was delayed by one day to today. What is the date 10 days ago in MM/DD/YYYY?

A: One day after 06/01/1943 is 06/02/1943, so today is 06/02/1943. 10 days before today is 05/23/1943. So the answer is 05/23/1943. (Wei et al. 2022: 24827)

Die Ergebnisse der Untersuchung haben gezeigt, dass der Einsatz der Strategie bessere Ergebnisse als Standardprompting erzielen konnte. Es wurden mittlerweile auch Einsätze vorgestellt, die auf der Few-Shot-Learning-Strategie (vgl. Zhao et al. 2021: 12697) und Chain-of-Thought-Strategie aufbauen wie z. B. Self-consistency (vgl. Wang et al. 2022: 1) oder Zero-Shot-CoT (vgl. Kojima et al. 2022: 22199–22200) auf die hier allerdings in Anbetracht des Umfangs der Arbeit nicht weiter eingegangen wird. Trotz der oben angeführten Vorteile soll erwähnt werden, dass die Kombination einzelner Prompt-Engineering-Strategien nicht zwangsläufig zu besseren Ergebnissen führen muss (vgl. Li et al. 2024: 19).

4. Evaluierungsmaßnahmen der automatischen Termextraktion

Fundamentally, evaluation is always a comparison. (Thurmair 2003: 5)

Im vorigen Kapitel wurde geschildert, dass die neuesten Verfahren der automatischen Termextraktion auf neuronalen Ansätzen basieren, und dass sie aktuell zu den Verfahren auf dem neuesten Stand der Technik auf diesem Gebiet gehören. Das schließt mehrere Aspekte ein, darunter die verhältnismäßig „gute“ Erfolgsquote bei der Gewinnung von geeigneten Termkandidaten (vgl. Tran et al. 2023: 21). Nichtsdestotrotz kann man sich nicht ohne Weiteres vollkommen auf die Automatisierung in diesem Bereich verlassen, was die Evaluierung umso bedeutender macht (vgl. Di Nunzio et al. 2023: 45; L’Homme et al. 1996: 294). Es bestehen nämlich Grenzen, mit denen eine Maschine noch im Gegensatz zu einem Menschen konfrontiert ist wie z. B. der Umgang mit den Rechtschreibfehlern, was das humane Erkennungsvermögen wiederum nicht beeinträchtigt (vgl. Zerfaß 2006: 24). In dem Zusammenhang ergibt sich die Frage, wie sich der Erfolgswert der automatischen Termextraktion messen bzw. bestimmen lässt. Dafür werden Evaluierungsmaßnahmen benötigt, auf die im folgenden Kapitel näher eingegangen wird.

4.1. Grundsätze des Evaluierungsverfahrens

Wie im oben angeführten Zitat angedeutet, steht der Vergleich im Zentrum der Evaluierung. Diese Ansicht teilen jedenfalls L’Homme et al. (1996: 302), die die ersten Evaluierungskriterien für die Termextraktionssysteme konzipiert haben (vgl. Vivaldi & Rodriguez 2007: 232). Die Autorinnen (1996: 302) meinen, dass man dafür eine Liste automatisch generierter Termkandidaten mit einer von Menschen erstellten Liste vergleicht. Die letztere Auflistung gilt als perfekt, obwohl auch Menschen Termini übersehen können (vgl. auch Justeson & Katz 1995: 20). L’Homme et al. (1996: 294-302) schlagen ein Zwei-Phasen-Evaluierungsverfahren vor. Die erste Phase findet noch vor dem Einsatz eines Termextraktionssystems statt und sie umfasst eine genaue Überlegung seiner Eigenschaften, die für verschiedene Nutzer:innengruppen wie Übersetzer:innen, Manager:innen oder Terminolog:innen von Bedeutung sind. Zu den Kriterien zählen u.a. der Anwendungsbereich des Systems, die unterstützten Sprachen und seine Kompatibilität, sollte es sich um kein alleinstehendes System handeln (vgl. 1996: 294-302). Die nächste Überlegung betrifft die Auswahl der Ansätze, die im vorigen Kapitel beschrieben wurden. Des Weiteren ist es wichtig zu betrachten, auf die Extraktion welcher Termini ein System

ausgerichtet ist, ob es sich z. B. bei den Mehrwortbenennungen ausschließlich um die Nominalphrasen handelt. Außerdem bieten gewisse Systeme diverse zusätzliche Funktionen wie z. B. die Möglichkeit der Einstellung der Mindestfrequenz für die Termkandidatenerkennung oder die morphologische Analyse für die Lemmatisierung der extrahierten Termini. Schließlich produzieren Extraktionssysteme entweder eine oder mehrere Liste(n) mit Termkandidaten, die unterschiedlich aussehen können z. B. bezüglich der Reihenfolge der Fachausdrücke (vgl. 1996: 294-302).

Tran et al. (2023: 15) zufolge weist die Evaluierung im Bereich der automatischen Termextraktion hauptsächlich zwei Aspekte auf: die Evaluierungsmethodologie (evaluation methodology) und den Anwendungsbereich (scope of the results). Die Methodologie unterscheidet zwischen einer menschlichen, wörterbuch- und goldstandardbasierten Evaluierung der Termkandidaten. Beim Anwendungsbereich handelt es sich um die Entscheidung, ob alle Ergebnisse, ein Teil der Termkandidaten oder nur die Top-K-Ergebnisse evaluiert werden sollten (vgl. 2022: 15). Ein Top-K-Ergebnis bezeichnet das Resultat einer Top-K-Anfrage (top-k query), die darauf ausgerichtet ist, ausschließlich die Ergebnisse mit dem höchsten (bzw. niedrigsten) Wert gemäß einer vorbestimmten Bewertungsfunktion abzurufen (vgl. Ilyas 2018: 4168–4169). Wie bereits erwähnt, wurden die ersten Evaluierungsverfahren lediglich auf der manuellen Überprüfung der automatisch extrahierten Termkandidaten von Fachexpert:innen gestützt. Dies schließt allerdings einen großen Menscheneinsatz ein. Um dem entgegenzuwirken, wurde die wörterbuchbasierte Evaluierung eingeführt, entweder als eine Liste oder als Wörterbuch mit den vordefinierten Kriterien, die den Termkandidaten zugewiesen werden (vgl. Tran et al. 2023: 15). Außerdem können die statistischen Maße Termhood und Unithood bei der Evaluierung berücksichtigt werden. Eine weitere methodologische Herangehensweise betrifft einen Goldstandard. Unter Goldstandard im NLP-Bereich wird eine manuell überprüfte, korrigierte und annotierte Referenzliste mit Termini verstanden, mit der automatisch extrahierte Termkandidaten verglichen werden, um die Leistung eines Systems zu bewerten. In der Regel werden diese Referenzlisten von mehr als einer Person überprüft, um als Goldstandard zu gelten. Es handelt sich dabei um verlässliche Datensätze, ohne welche die Evaluierung und das Training der Ansätze des maschinellen Lernens nicht möglich wären (vgl. Baker 2006: 78; Gius 2016; Wissler et al. 2014: 1). Nazarenko und Zargayouna (2009: 300) weisen allerdings auf die Problematik des menschlichen Faktors bei der Erstellung vom Goldstandard hin. Selbst für das gleiche Gebiet und Korpus wird nämlich von verschiedenen Terminolog:innen aufgrund ihrer Sichtweisen, terminologischer Traditionen und persönlicher Entscheidungen eine etwas andere Termkandidatenliste vorgeschlagen. Deswegen, so die Autorinnen, gibt es keinen einzigen

Goldstandard, sondern eine Vielzahl „akzeptabler Lösungen“ (vgl. 2009: 300). Vivaldi und Rodriguez (2007: 227) sind auch dieser Ansicht und führen weiters an, dass es außerdem mangelnde Übereinstimmung zwischen Terminolog:innen bezüglich wichtiger Entscheidungen gibt, die darauf zurückzuführen ist, dass es sich schwer bestimmen lässt, welche messbaren Eigenschaften bzw. Maße, tatsächlich zur besseren Qualität des Systems beitragen (2007: 227). Im Falle eines Goldstandards kann es sich um eine angepasste bereits vorhandene Liste, eine repräsentative Stichprobe (Sample) oder um die Liste aller Termini im Korpus handeln (vgl. Tran et al. 2023: 15). Die bereits vorhandene Terminliste wird nicht aus dem untersuchten Trainingskorpus gewonnen, sondern sie bezieht sich auf den Fachbereich selbst und ist ein Produkt der Zusammenarbeit einer weltweiten Gemeinschaft. Diese Methode weist allerdings eine Schwachstelle auf. Anhand der Liste kann zwar festgestellt werden, ob die extrahierten Termkandidaten tatsächlich Fachausdrücke sind, aber nicht wie viele Termini aus dem Text korrekt gewonnen wurden (vgl. 2023: 15). Entscheidet man sich dahingegen dafür, alle Termini direkt aus dem Korpus manuell zu annotieren und eine solche Liste als Goldstandard einzusetzen, wird die Leistung eines Termextraktionssystems gewöhnlicherweise mittels Genauigkeit, Trefferquote und schließlich F1-Score bewertet. Unter F1-Score wird eine Evaluierungsmetrik bzw. ein harmonisches Mittel, das Genauigkeit und Trefferquote miteinander kombiniert und balanciert (vgl. Humphrey et al. 2022: 116) verstanden. Diese Form von Evaluierung liegt der Untersuchung im Rahmen der vorliegenden Arbeit inne. Abgesehen von der Leistung des Systems bei der Extraktion, besteht eine ganze Reihe an externen Eigenschaften, die ggf. ebenfalls in die Evaluierung des Systems einfließen können. Dabei handelt es sich um die sog. Black-Box-Evaluierung bzw. ganzheitliche Evaluierung. Sauron (2002: 4–6) nennt diesbezüglich folgende Eigenschaften und ihre besonderen Merkmale, wie in Abbildung 10 dargestellt.

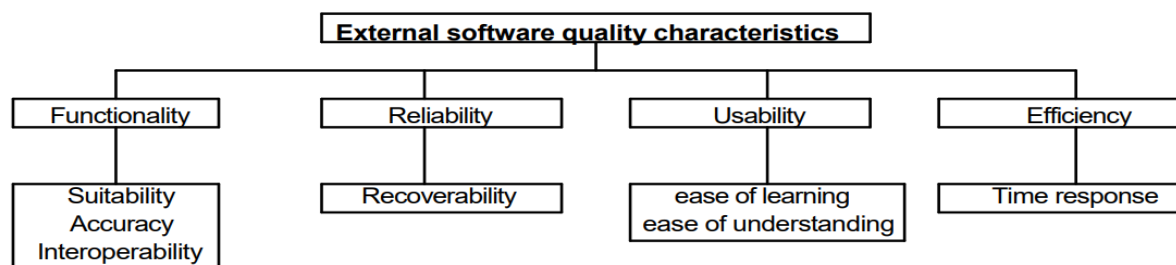


Abbildung 10: Externe Eigenschaften für die Black-Box-Evaluierung (Sauron 2002: 6)

Diese Eigenschaften zeigen u.a., ob die Software benutzerfreundlich ist, die gleiche Leistung unter diversen Bedingungen erbringt und mit verschiedenen Systemen kompatibel ist. Eine weitere Variante wird Glass-Box-Evaluierung genannt. Sie schließt die Bewertung der internen

Struktur eines Systems ein, z. B. des Algorithmus, und setzt voraus, dass sie für die evaluierende Person mindestens teilweise bekannt und von ihr überprüfbar ist (vgl. Vivaldi & Rodriguez 2007: 226). Den Autoren zufolge geht man im Falle einer Evaluierung von einer direkten (intrinsischen) oder indirekten (extrinsischen) Evaluierung aus. Während die erstere die verwendeten intrinsischen Maße eines Systems unabhängig ihres geplanten Verwendungszwecks bewertet, steht im Fokus der letzteren die Bewertung, wie sich die Leistung eines anderen Systems aufgrund der von dem System gelieferten Ergebnisse gesteigert hat (vgl. 2007: 226).

In Anbetracht des Anwendungsbereichs werden hauptsächlich alle Termkandidaten mit dem manuell annotierten Goldstandard verglichen, obschon es ebenfalls Studien gibt, die sich dabei lediglich auf die Top-K-Ergebnisse beziehen (vgl. Vintar 2010: 149; Zhang et al. 2018: 106) und durch Genauigkeit alleine bewertet werden (vgl. Tran et al. 2023: 17). Vivaldi und Rodriguez (2007: 229) führen diesbezüglich an, dass die Trefferquote am schwierigsten zu berechnen ist, denn dieses Maß setzt voraus, dass von Anfang an bekannt ist, wie viele Fachausdrücke der Korpus oder der zu bewertende Korpusteil enthält. Angesichts dessen wird eben häufig nur die Genauigkeit berechnet, denn es ist einfacher die Expert:innen um die Überprüfung einzelner Termkandidaten zu bitten, als sie damit zu beauftragen, die ganze Terminliste zu erstellen (vgl. Vivaldi & Rodriguez 2007: 231). Tran et al. (2023: 17) führen ebenfalls an, dass die meisten früheren Studien Termkandidaten ausschließlich aufgrund der Genauigkeit bewertet haben, da Goldstandards nicht in dem Ausmaß vorhanden waren, wie es heute z. B. dank ACTER der Fall ist (vgl. 2023: 17). Tabelle 4 fasst die beiden oben beschriebenen Aspekte der Evaluierung der automatischen Termextraktion zusammen.

Tabelle 4: Überblick über verschiedene Evaluierungsmetriken im Bereich der automatischen Termextraktion (adaptiert von Tran et al. 2023: 16)

Evaluierung		
Evaluierungsmethodologie	Anwendungsbereich	
human	alle Termkandidaten	
wörterbuchbasiert	ein Teil der Termkandidaten	
goldstandardbasiert	Top-K-Ergebnisse	
Genauigkeit	Trefferquote	F1-Score

4.2. Mengenorientierte Evaluationsmaße

Die Anfänge der automatischen Termextraktion sind entweder bei der Linguistik oder bei der Informationsextraktion zu finden. Daher ist es nicht verwunderlich, dass aus diesen Bereichen

die bekanntesten Evaluierungsverfahren kommen, die bei der Evaluierung der Termkandidaten eingesetzt werden (vgl. Vivaldi & Rodriguez 2007: 229). Die linguistischen Evaluierungsmaßnahmen analysieren die Anzahl der ungeeigneten Fachausdrücke auf der Termkandidatenliste, den Noise-Effekt, und die Anzahl der vom System übersehenen Termini, den Silence-Effekt. Diese Analyse bildet z. B. die zweite Phase des Evaluierungsverfahrens von L’Homme et al. (1996: 302-309). Im Fokus der im Bereich der Informationsextraktion etablierten Methode, die ihren Weg zur Evaluierung automatischer Terminologiegewinnung gefunden hat, steht, wie bereits erwähnt, die Berechnung von Genauigkeit und Trefferquote. Im Grunde genommen überliefern die beiden Verfahren die gleichen Informationen, wenngleich auf unterschiedliche Art und Weise (Vivaldi & Rodriguez 2007: 229). Zu der Konvention der Bewertung verschiedener Evaluationsmaße im NLP-Bereich gehört ihre Darstellung in Zahlenwerten zwischen 0 und 1, wobei die Werte oft in Prozent ausgedrückt werden. Je höher die Zahl ist, desto besser ist das Resultat (vgl. Biemann et al. 2022: 295; Sauron 2002: 9). Die verschiedenen Berechnungen können Abbildung 11 entnommen werden.

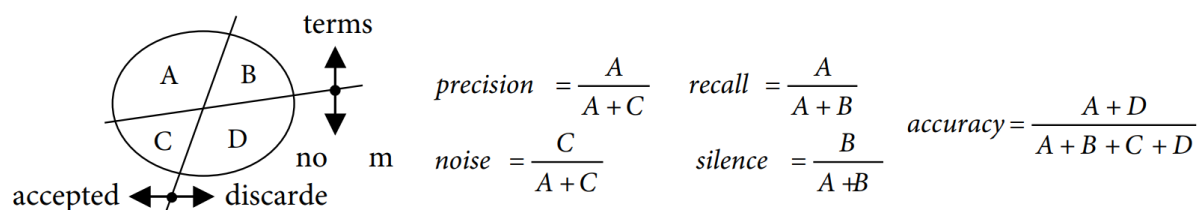


Abbildung 11: Typische Evaluationsmaße am Gebiet der automatischen Terminologiegewinnung (Vivaldi & Rodriguez 2007: 229)

Die oben abgebildeten Maße dienen der Evaluierung von Klassifikationsmethoden, die die Testdaten mit jeweils einem Label versehen und somit die Daten in eine Menge pro Label unterteilen. Diese Mengen werden im nächsten Schritt mit den manuell annotierten Mengen eines Goldstandards verglichen. Die folgenden vier Kennzahlen bilden die Grundlage der mengenorientierten Evaluationsmaße.

- A (TP; true positive; tatsächlich positiv): Das Label wurde richtigerweise zugewiesen
- B (FN; false negative; falsch negativ): Das Label wurde fälschlicherweise nicht zugewiesen
- C (FP; false positive; falsch positiv): Das Label wurde fälschlicherweise zugewiesen
- D (TN; true negative; tatsächlich negativ): Das Label wurde richtigerweise nicht zugewiesen (Biemann et al. 2022: 296)

Edo Marzá (2009: 240) gibt spezifisch bezüglich der automatischen Terminologiegewinnung ein Beispiel für die Konfusionsmatrix, wie in Tabelle 5 dargestellt.

Tabelle 5: Beispiel einer Konfusionsmatrix (adaptiert von Edo Marzá 2009: 240)

		Predicted class	
		Term	Non-Term
Real class	Term	TP – terms successfully recognised by the system	FN – terms not recognised by the system
	Non-Term	FP – non-terms considered as terms by the system	TN – non-terms recognised as such by the system

Das wohl einfachste Maß zur Evaluierung der Qualität der extrahierten Termini ist Exaktheit (Accuracy) (Abbildung 11 rechts) respektive die Fehlerquote. Wie bereits angedeutet, bezweckt die automatische Termextraktion die Gewinnung aller relevanten Termini. Zu viel Noise-Effekt würde einerseits zusätzlichen Aufwand mit sich bringen, da Terminolog:innen im Nachhinein zu viele ungeeignete Termini entfernen müssten und andererseits würde zu viel Silence-Effekt eine manuelle Recherche von den Fachausdrücken erfordern, die das System übersehen hat (vgl. Sauron 2002: 8). Das Maß der Exaktheit bemisst, inwiefern dieser potenzielle Aufwand notwendig ist. Anders ausgedrückt spiegelt es die Anzahl der richtigen Entscheidungen bezüglich aller Entscheidungen wider. Um den Exaktheitswert zu berechnen, werden alle richtigerweise erkannten Termini und Nicht-Termini addiert und durch alle getroffenen Entscheidungen dividiert (Abbildung 11). Biemann et al. (2022: 296) argumentieren allerdings, dass diese Methode zwar „intuitiv“ und „leicht verständlich“ ist, aber nicht in allen Fällen „aussagekräftig“ genug ist, da bei der Exaktheit häufig mit zu hohen Werten zu rechnen ist. Die Autoren belegen ihre Aussage mit dem folgenden Beispiel.

Wenn z. B. in einem stark verspamten Postfach bei einem Spam-Filter nur 10% E-Mails ankommen, welche nicht als Spam herausgefiltert werden sollen, würde ein übereifriger Filter, der alle Spam-E-Mails als Spam klassifiziert und drei Viertel der Nicht-Spam-E-Mails ebenfalls als Spam, insgesamt eine Accuracy von 92,5% erreichen, was gut klingt, aber für den praktischen Einsatz hier komplett unbrauchbar ist. (Biemann et al. 2022: 296)

Manning und Schütze (1999: 269–270) untermauern diese Aussage, indem sie argumentieren, dass in den meisten Situationen die Anzahl der nicht erkannten Nicht-Termini dermaßen groß ist, dass sie andere Werte „in den Schatten stellt“. Um den Noise-Effekt zu berechnen, werden fälschlicherweise erkannte Nicht-Termini durch die Summe der richtigerweise erkannten Termini und der fälschlicherweise erkannten Nicht-Termini dividiert. Mit anderen Worten, wenn sich auf der Termkandidatenliste von insgesamt 50 Fachausdrücken 10 Ausdrücke befinden, die über keine terminologische Relevanz verfügen, dann beträgt der Noise-Wert $(10/(40+10)=1/5)$ 20%. Die Berechnung vom Silence-Effekt erfolgt durch die Division von der

Anzahl aller fälschlicherweise nicht erkannten Termini durch die Summe der richtigerweise erkannten Termini und der fälschlicherweise nicht erkannten Termini. Mit anderen Worten, wenn auf der Termkandidatenliste 50 Fachausdrücke stehen, davon sind 10 keine Fachausdrücke und 5 Fachausdrücke wurden vom System im Zuge der Extraktion übersehen, dann beträgt der Silence-Wert $(5/(40+5)=1/9)$ 11,11 %.

Mit der Genauigkeit wird die Korrektheit der extrahierten Termkandidaten und mit der Trefferquote ihre Vollständigkeit bemessen. Wie in den vorherigen Kapiteln bereits beschrieben, muss bei Verfahren der statistischen Termextraktion ein Schwellenwert bestimmt werden, der sich auf die Termhood bezieht. Wird der Schwellenwert niedriger angesetzt, dann werden gewöhnlicherweise mehr Termkandidaten in die Liste aufgenommen, was bedeutet, dass die Trefferquote steigt und die Genauigkeit sinkt. Sollte der Schwellenwert höher sein, werden weniger Kandidaten extrahiert und infolgedessen steigt die Genauigkeit und die Trefferquote sinkt (vgl. Edo Marzá 2009: 242), wie in Abbildung 12 abgebildet.

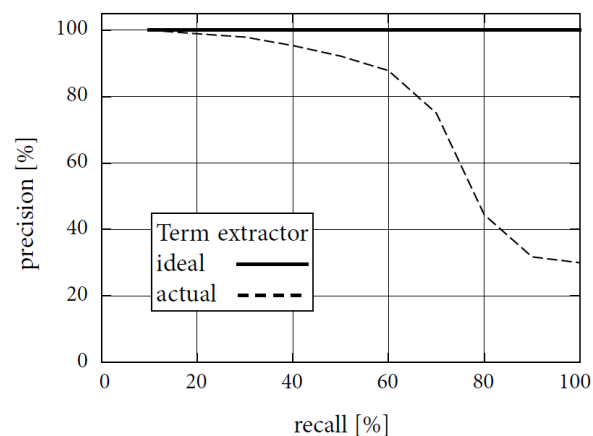


Abbildung 12: Genauigkeit und Trefferquote – die typische Verhaltensweise (Vivaldi & Rodriguez 2007: 231)

Um die Konsistenz der anhand der Ergebnisse gezogenen Schlussfolgerungen zu gewährleisten, sollen beide Maße berücksichtigt und balanciert werden (vgl. Sauron 2002: 9; Vivaldi & Rodriguez 2007: 229). Im Idealfall soll natürlich die Extraktion die maximalen Werte sowohl für die Genauigkeit als auch für die Trefferquote erzielen (vgl. Rijsbergen 1975: 129), aber Sauron (2002: 9) argumentiert, dass dies in der Praxis sehr selten bzw. nie passiert. Um den Genauigkeitswert (Abbildung 11) zu bestimmen, werden alle richtigerweise extrahierten Termini durch die Summe dieses Wertes und der fälschlicherweise extrahierten Nicht-Termini dividiert. Anders ausgedrückt, wenn auf der Liste 50 Kandidaten stehen und 10 davon sind keine

Fachausdrücke, dann beträgt die Genauigkeit $(40/(40+10)=4/5)$ 80%. Der Wert der Trefferquote (Abbildung 11) wird berechnet, indem die Anzahl aller richtigerweise erkannten Termini durch die Summe dieses Wertes und der Anzahl aller fälschlicherweise nicht erkannten Termini dividiert wird. Mit anderen Worten, wenn auf der Liste 50 Kandidaten stehen und davon sind 10 Kandidaten keine Fachausdrücke und zusätzlich dazu wurden 5 Fachausdrücke vom System im Zuge der Extraktion übersehen, dann beträgt die Trefferquote $(40/(40+5)=8/9)$ 88,89%. Bei der Berechnung vom F1-Score muss zuerst berücksichtigt werden, ob beide Maße, also die Genauigkeit und Trefferquote, gleich gewichtet sind, was gewöhnlicherweise der Fall ist, solange keine zusätzlichen Informationen vorliegen (vgl. Biemann et al. 2022: 297; Vivaldi & Rodriguez 2007: 230). Der F1-Score wird, wie in Abbildung 13 dargestellt, ermittelt.

$$F1 - score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Abbildung 13: Die Berechnung vom F1-Score (Tran et al. 2023: 16)

In dem oben angeführten Beispiel mit der Genauigkeit von 80% und der Trefferquote von 88,89% beträgt der F1-Score $(2 \times (80 \times 88,89) / (80 + 88,89))$ 84,21% respektive 0,84. Die Berechnung vom F1-Score wird als Methode zur Evaluierung von Termkandidaten von vielen Forscher:innen vorgeschlagen (siehe Ala & Sharma 2020; Šajatović et al. 2019; Tran et al. 2023; Jianzhang Zhang et al. 2022 u.a.). Vivaldi und Rodriguez (2007: 230) ziehen allerdings in Erwägung, dass diese Methode möglicherweise nicht ganz „gerecht“ ist. Die Autoren überlegen beispielsweise, ob der Noise-Effekt (falsch positiv, C) und der Silence-Effekt (falsch negativ, B) die gleiche Auswirkung auf die Berechnung von Genauigkeit und/oder Trefferquote und somit vom F1-Score haben sollen. Anders ausgedrückt, stellt sich die Frage, was schwerwiegender bei der automatischen Termextraktion ist, die Gewinnung terminologisch irrelevanter Kandidaten oder die Nicht-Gewinnung von Fachausdrücken. Wie bereits erwähnt, kann die Gewichtung angepasst werden (vgl. Baeza-Yates & Ribeiro 2011: 327), allerdings wird man dann mit der Frage konfrontiert, welches Maß wie zu gewichten ist, damit das „unabhängige“ Ergebnis erzielt werden kann (vgl. 2011: 327). Außer der Berücksichtigung der Gewichtung der Metriken werden auch sie selbst unterschiedlich betrachtet. Yacouby und Axman (2020: 81) schlagen beispielsweise Confidence-Genauigkeit, -Trefferquote- und -F1-Score vor, eine probabilistische Erweiterung der ursprünglichen Metriken, die statt auf einer binären, auf einer kontinuierlichen Definition von Positiv und Negativ basiert, um einigen der Herausforderungen

bei der Evaluierung entgegenzuwirken. Weitere für die Evaluierung der Termkandidaten verwendete Metriken, die Genauigkeit und Trefferquote kombinieren, werden Average Precision und Mean Average Precision genannt. Average Precision berechnet den Durchschnitt der Genauigkeit bei allen Trefferquote-Werten und wird typischerweise bei der Bewertung von Ranglistenenergebnissen eingesetzt. Mean Average Precision ist der Durchschnittswert der Average Precision über alle Klassen, die evaluiert werden (siehe Fedorenko et al. 2014: 57; Manning & Schütze 1999: 534–536; Padilla et al. 2020: 239; Ziqi Zhang et al. 2008: 2110). Eine andere Evaluierungsmetrik, die als Gegenteil von der Trefferquote gesehen werden kann, wird Spezifität (vgl. Edo Marzá 2009: 242; Zhu et al. 2010: 2) genannt. Sie stellt Nicht-Termini in den Fokus und berechnet, wie viele Nicht-Termini fälschlicherweise als Termini erkannt wurden: $TN/(TN+FP)$ bzw. $D/(D+C)$. Mit anderen Worten, wenn sich im Korpus 50 richtigerweise nicht erkannte Nicht-Termini befinden und auf der Termkandidatenliste 10 fälschlicherweise extrahierte Nicht-Termini zu sehen sind, dann beträgt der Spezifitätswert $(50/(50+10)=5/6)$ 83,33%. Eine der seltener verwendeten Metriken namens Fallout wird manchmal eingesetzt, um zu messen, wie schwer es ist, ein System zu entwickeln, das nur wenige falsch positive Ergebnisse liefert (vgl. Manning & Schütze 1999: 270). Das Fallout-Maß wird durch die Division der fälschlicherweise extrahierten Nicht-Termini und der Summe dieses Werts und der richtigerweise nicht extrahierten Nicht-Termini ermittelt.

Die Extraktionssysteme nach dem neuesten Stand der Technik erzielen aktuell bei der automatischen Terminologiegewinnung die Genauigkeit- und Trefferquote-Werte zwischen 35 und 70% und den F1-Score von maximum 60 bis 70% (vgl. Di Nunzio et al. 2023: 45; Tran et al. 2023: 21).

5. Fachsprachen

The general aim of Term Extraction is to identify the core vocabulary of a specialized domain.
(Heylen & De Hertog 2015: 203)

Die automatische Termextraktion bezweckt, die Termini eines konkreten Fachbereichs zu identifizieren, wie bereits in vorherigen Kapiteln beschrieben und anhand des oben angeführten Zitats noch einmal verdeutlicht. Im Fokus der vorliegenden Arbeit stehen juristische Texte, aus denen die vorwiegend rechtliche Terminologie automatisch gewonnen werden soll. Aus diesem Grund soll das folgende Kapitel einerseits einen Überblick darüber verschaffen, wie Fachsprachen zu definieren sind und inwiefern sie sich von der Gemeinsprache unterscheiden, sowie über ihre charakteristischen lexikalischen und grammatikalischen Merkmale. In weiterer Folge wird die Rechtssprache thematisiert, die innerhalb der Fachsprachen aus vielerlei Hinsicht eine Sonderstellung aufweist, allein aufgrund dessen, dass sie im Gegensatz zu anderen Fachsprachen alle Mitglieder der Gesellschaft betrifft (vgl. Eckardt 2000: 22).

5.1. Allgemeine, fachliche und berufliche Kommunikation

In der Fachliteratur herrscht kein Konsens darüber, wie Fachsprache zu definieren ist. Einige Autor:innen vertreten die Ansicht, dass es kaum möglich ist, eine Definition zu formulieren (vgl. Bolten 1992: 58). Andererseits gibt es etliche Interpretationen von dem Begriff Fachsprache, die auf dem jeweiligen Verständnis vom Fach und von der Sprache basieren (vgl. Roelcke 2020: 11). Der Ausdruck Fach im Sinne eines Spezialgebiets wird seit dem 18. Jahrhundert verwendet, aber die genaue Bestimmung des Begriffs Fach sowie die Fächerabgrenzung erweisen sich als „schwierig“ (Adamzik 2018: 163). Die linguistische bzw. semiotische Bestimmung vom Fach ist daran geknüpft, wie sprachliche und nichtsprachliche Zeichen innerhalb eines bestimmten menschlichen Tätigkeitsbereichs verwendet werden. Seine referenzielle Bestimmung bezieht sich auf den Gegenstandsbereich und die soziologische Bestimmung auf die Gruppe tätiger Personen, deren Tätigkeitsbereich von anderen Gruppen als solcher abgegrenzt bzw. erkannt wird (vgl. Roelcke 2020: 14). Sprache dagegen wird zum Beispiel im systemlinguistischen Ansatz (siehe 5.1.2.) als Sammlung von Wörtern und syntaktischen Regeln bzw. als „Gesamtheit sprachlicher Mittel“ definiert (Hoffmann 2022: 170). Fluck (1996: 193) gibt außerdem an, dass Fachsprachen sich aus mehreren Perspektiven betrachten lassen, was dazu führt, dass jeweils ein anderes definitorisches Merkmal im Mittelpunkt der Definition stehen kann, je nachdem, ob der Schwerpunkt auf den kommunikativ-funktionalen, pragmatischen,

textuellen oder soziologischen Aspekt gelegt wird (vgl. 1996: 193). Der Definition der Fachsprache wird die Bestimmung bzw. das Verständnis der Kommunikation selbst und insbesondere der fachlichen Kommunikation vorausgesetzt. Um den Ablauf der Kommunikation zu veranschaulichen, werden sog. Kommunikationsmodelle verwendet. Zu den bekanntesten Modellen in Bezug auf die allgemeine Kommunikation gehören u.a. das Vier-Ohren-Modell, das Organon-Modell und das Sender-Empfänger-Modell, das von Shannon und Weaver in den 1940er Jahren vorgeschlagen wurde und in Abbildung 14 dargestellt ist.

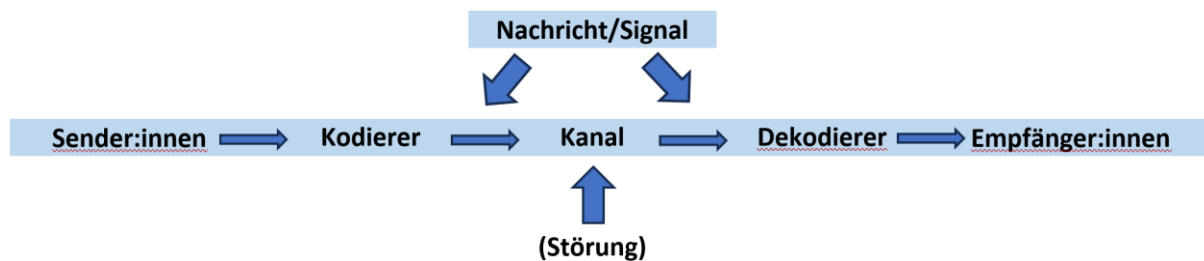


Abbildung 14: Das Sender-Empfänger-Modell (adaptiert von Röhner & Schütz 2020: 30)

Lt. diesem Modell beginnt der Kommunikationsprozess bei den Sender:innen, die eine kodierte Nachricht in Form von Signalen in einem bestimmten Kanal an die Empfänger:innen übermitteln, welche die Nachricht empfangen und dekodieren. Das Ziel ist den Inhalt der Nachricht der Sender:innen durch Empfänger:innen richtig zu entschlüsseln, wofür der gemeinsame Code unabdingbar ist (vgl. Röhner & Schütz 2020: 30). Dem allgemeinen Kommunikationsmodell nach können im Zuge der Kommunikation unter Umständen auch Störfaktoren auftreten, die die erfolgreiche Kommunikation erschweren. Dazu zählen z. B. Tonverzerrungen, mangelnde Aufmerksamkeit, Mehrdeutigkeiten, unzureichende Sprachbeherrschung oder Unklarheiten (vgl. 2020: 30).

Zu den Tragenden der fachsprachlichen Kommunikation zählen Fachtextproduzent:innen, Fachtextrezipient:innen sowie der Fachtext selbst entweder in mündlicher oder in schriftlicher Form. Sowohl die Fachtextproduktion als auch -rezeption stellen in der Fachkommunikation einen aktiven Prozess dar. Die fachliche Kommunikation kann zwischen mehreren Personen erfolgen und in verschiedene Richtungen verlaufen und sowohl Produzent:innen als auch Rezipient:innen können ihre jeweiligen Rollen wechseln. Damit erfolgreiche fachliche Kommunikation gewährleistet werden kann, müssen fachsprachliche Zeichensysteme sowie sprachliches Wissen, der Kontext, und sachliches Wissen, der Kontext, der Textproduzent:innen und -rezipient:innen ausreichend koinzidieren (vgl. Roelcke 2020: 12). In der Fachliteratur kommen mehrere Fachkommunikationsmodelle mit Entwicklungsphasen der Fachkommunikation vor.

Schubert (2007: 141–160, 2009: 117) nennt vier Entwicklungsstufen: die terminologische, systemlinguistische, textlinguistische und kognitiv-kommunikative Entwicklungsstufe. Kalverkämper (1998: 48) geht von sechs Ebenen aus: die terminologische, syntaktische und funktionalstilistische, textuelle, pragmatische und kommunikative, soziokulturelle und semiotische Ebene. Roelcke (2020: 12) schlägt drei Entwicklungsphasen in Form von drei Modellen vor, die in der Abbildung 15 dargestellt sind und auf die in weiterer Folge ausführlicher eingegangen wird.

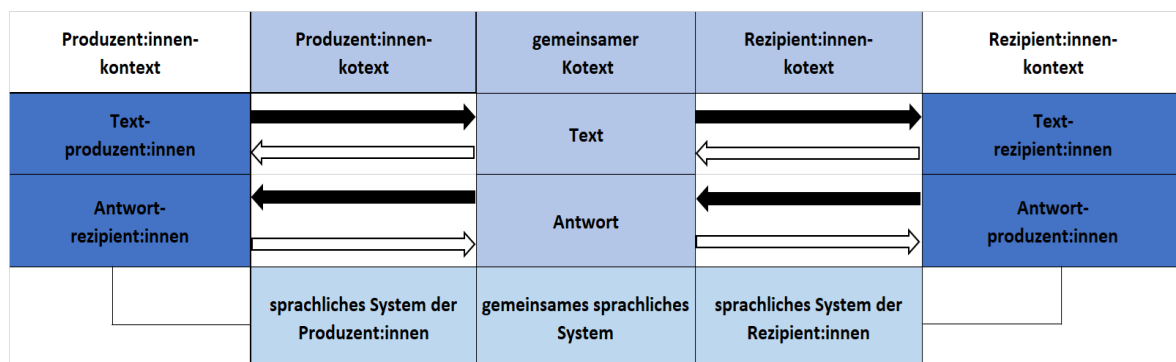


Abbildung 15: Vereinfachtes Modell fachsprachlicher Kommunikation: Systemlinguistisches Inventarmodell (hellblau), pragmalinguistisches Kontextmodell (mittelblau), kognitionslinguistisches Funktionsmodell (dunkelblau) (adaptiert von Roelcke 2020: 12)

Basierend auf dem oben angeführten Fachkommunikationsmodell schlägt Roelcke (2022: 597) ein Kommunikationsmodell für die Berufssprache vor, die in der Fachliteratur zwischen der Allgemeinsprache und Fachsprache platziert wird (vgl. Eging 2023: 655). Im Modell der beruflichen Kommunikation wird der Fokus auf die individuellen Zeichensysteme der Produzent:innen und Rezipient:innen sowie deren Überschneidung aufgegeben. Stattdessen wird ein Ansatz verfolgt, der mehrere unterschiedliche Varietäten bzw. Register berücksichtigt wie etwa die Allgemeinsprache, die Bildungssprache und diverse Fachsprachen, wie Architektur, Geologie, Ökologie u.a. (vgl. Roelcke 2022: 596).

5.1.2. Definition von Fachsprache

Im folgenden Unterkapitel werden Definitionen von Fachsprache im Sinne der drei Fachsprachenkonzeptionen nach Roelcke (2020: 12ff.) geschildert.

Das systemlinguistische Inventarmodell hat die Fachsprachenforschung der 50-er, 60-er und 70-er Jahre geprägt. Im Fokus dieses Modells steht das gemeinsame sprachliche Zeichensystem der Textproduzent:innen und -rezipient:innen. Die fachsprachliche Kommunika-

tion richtet sich, wie bereits erwähnt, nach dem lexikalischen Repertoire und nach den syntaktischen Regeln der Sprache. Die Fachsprache ist somit ein Zeichensystem, das für die fachliche Kommunikation gebraucht wird. Im Sinne des Modells äußert sich das Verständnis der Fachsprachen als Varietäten einer Einzelsprache, die ihr untergeordnet sind und sich von anderen Varietäten durch inner- und außersprachliche Merkmale unterscheiden. Die innersprachlichen Merkmale sind auf solchen Ebenen wie z. B. Lexik, Syntax und Morphologie angesiedelt und betreffen Beschreibungsgesichtspunkte: Grammatik, Semantik, Pragmatik und Inventar (Röelcke 2020: 15). Zu den außersprachlichen Merkmalen, die eine Fachsprache charakterisieren, gehören funktionale, räumliche, historische und soziale Aspekte. Darunter sind respektive u.a. die Funktion im menschlichen Tätigkeitsbereich; Dialekte; Perioden; und Soziolekte zu verstehen. Obwohl die funktionale Bestimmung der Fachsprache den primären Aspekt darstellt, werden auch solche Punkte wie beispielsweise die Entwicklung einer Fachsprache im bestimmten Zeitraum oder der Vergleich regionaler Unterschiede (z. B. im Fall von See- und Binnenfischerei) bei der Fachsprachenbestimmung berücksichtigt (vgl. 2020: 15). Zu den Definitionen von Fachsprache im Sinne dieses Modells gehört jene von Hoffmann aus dem Jahr 1976:

Fachsprache – das ist die Gesamtheit aller sprachlichen Mittel, die in einem fachlich begrenz-
baren Kommunikationsbereich verwendet werden, um die Verständigung zwischen den in die-
sem Bereich tätigen Menschen zu gewährleisten. (Hoffmann 1976: 170)

Im Fokus dieser Bestimmung der Fachsprache liegt ihre Verwendung im fachlich begrenz-
baren Kommunikationsbereich, was die Fachsprache an die Art des Handelns koppelt, das mit Fach-
kommunikation verbunden ist, sowie an die jeweilige Situation, in der diese Kommunikation
stattfindet (vgl. Schubert 2009: 118). Einen weiteren Definitionsversuch unternehmen Möhn
und Pelka im Jahr 1984:

Fachsprache – die Variante der Gesamtsprache, die der Erkenntnis und begrifflichen Bestim-
mung fachspezifischer Gegenstände sowie der Verständigung über sie dient und damit den
spezifischen kommunikativen Bedürfnissen im Fach allgemein Rechnung trägt. [...] Entspre-
chend der Vielzahl der Fächer, die man mehr oder weniger exakt unterscheiden kann, ist die
Variante Fachsprache in zahlreichen mehr oder weniger exakt abgrenzbaren Erscheinungsfor-
men realisiert, die als Fachsprachen bezeichnet sind [...]. (Möhn & Pelka 1984: 26)

Diese Definition hebt besonders das für den systemlinguistischen Ansatz typische Verständnis
der Fachsprache als Varietät der Gesamtsprache hervor und ebenfalls ihren funktionalen An-
satz. Die letzte Begriffserklärung von Fachsprache, die an dieser Stelle angeführt wird, kommt
von Schmidt aus dem Jahr 1969:

Fachsprache – das Mittel einer optimalen Verständigung über ein Fachgebiet unter Fachleuten; sie ist gekennzeichnet durch einen spezifischen Fachwortschatz und spezielle Normen für die Auswahl, Verwendung und Frequenz gemeinsprachlicher lexikalischer und grammatischer Mittel; sie existiert nicht als selbständige Erscheinungsform der Sprache, sondern wird in Fachtexten aktualisiert, die außer der fachsprachlichen Schicht immer gemeinsprachliche Elemente enthalten. (Schmidt 1969: 17)

Gnutzmann (1980: 49) zufolge gilt diese Definition trotz Kritik als „praktikabler Kompromiß unter den verschiedenen Sichtweisen“ der aus der heutigen Sicht älteren Fachsprachenforschung. Sie folgt dem funktional-linguistischen Ansatz und betont ebenfalls, dass Fachsprache keine selbständige Erscheinungsform der Sprache ist, sondern wie bereits vorher erwähnt, die Varietät der Gemeinsprache.

Das pragmlinguistische Kontextmodell ist den 70er und hauptsächlich 80er Jahren zuzuordnen. Zu dieser Zeit gewinnen in mehreren linguistischen Teildisziplinen pragmlinguistische und textwissenschaftliche Aspekte an Bedeutung (vgl. Roelcke 2020: 23). Dieses Modell wird nicht dadurch gekennzeichnet, das frühere Modell abstreiten zu wollen, vielmehr soll die Sichtweise bezüglich der Fachsprachen von dem Fachwortschatz bzw. Terminologie und Syntax auf den Fachtext als eine linguistische und kommunikative Einheit verlagert werden (vgl. Hoffmann 1987: 7). Im Sinne dieses Modells ist Fachsprache als „Äußerung von Texten im Rahmen fachlicher Kommunikation“ zu verstehen (Roelcke 2020: 23). Demzufolge schlägt Hoffmann 1985 eine Definition vom Fachtext anstatt der Fachsprache vor:

Der Fachtext ist Instrument und Resultat der im Zusammenhang mit einer spezialisierten gesellschaftlich-produktiven Tätigkeit ausgeübten sprachlich-kommunikativen Tätigkeit, er besteht aus einer endlichen, geordneten Menge logisch, semantisch und syntaktisch kohärenter Sätze (Texteme) oder satzwertiger Einheiten, die als komplexe sprachliche Zeichen komplexen Propositionen im Bewußtsein des Menschen und komplexen Sachverhalten in der objektiven Realität entsprechen. (Hoffmann 1985: 233)

Die obige Definition veranschaulicht die vorher erwähnte Verlagerung der Betrachtung einer Fachsprache als Zeichensystem zum Text, also dem Ergebnis der Fachkommunikation; mit anderen Worten geht die Fachsprachenforschung zu diesem Zeitpunkt von größeren kommunikativen Einheiten hin zu ihren einzelnen Elementen (vgl. Hoffmann 1988: 127; Roelcke 2020: 24). Dieser Definition liegt außerdem realistische Sprachauffassung zugrunde, die sowohl in der systemlinguistischen als auch der pragmlinguistischen Fachsprachenforschung prävalent ist (vgl. Roelcke 2020: 24). Sie besagt, dass es eine von Menschen unabhängige Welt von Gegenständen und Sachverhalten gibt, die Menschen durch Begriffe erkennen und durch Benennungen vermitteln (vgl. 2020: 24). Dies kann mithilfe eines semiotischen Dreiecks visualisiert

werden, das in den früheren Kapiteln bereits beschrieben wurde. Eine weitere Definition vom Fachtext kommt von Gläser aus dem Jahr 1990:

Als Ergebnis einer kommunikativen Handlung ist der Fachtext eine zusammenhängende, sachlogisch gegliederte und abgeschlossene komplexe sprachliche Äußerung, die einen tätigkeitsspezifischen Sachverhalt widerspiegelt, situativ adäquate sprachliche Mittel verwendet, und durch visuelle Mittel, wie Symbole, Formeln, Gleichungen, Graphika und Abbildungen ergänzt sein kann. (Gläser 1990: 18)

Die Autorin betont die primäre Rolle der Funktion der Sprache (der Texte) für die Kommunikation und sieht ebenfalls die Fachsprache als Textäußerung im fachlichen Bereich. Die Definition wurde zwar von einigen Autor:innen als „weit gefasst“ bzw. „breit aufgefasst“ empfunden (vgl. Gerbert 1995: 437; Ungerer 1992: 465), z. B. wegen der verwendeten Bezeichnung tätigkeitsspezifisch, was sich auf eine „Sphäre gesellschaftlicher Tätigkeit“ (vgl. Gläser 1990: 18) bezieht, allerdings sieht es Engberg (1992: 94) nicht unbedingt als Nachteil, da dies lt. dem Autor diese Definition „gut operationalisierbar“ macht.

Die innersprachlichen Merkmale des pragmalinguistischen Kontextmodells unterscheiden sich nicht von dem systemlinguistischen Inventarmodell, aber die außersprachlichen Merkmale sind anders. Die Fachsprachenforschung dieser Zeit wird durch ihren interdisziplinären Charakter gekennzeichnet, denn die Berücksichtigung der Textproduzent:innen und -rezipient:innen in der Fachkommunikation samt ihren situativen Konstellationen in den Fachbereichen der Soziologie und Psychologie angesiedelt ist. Die Beschreibung der an der Fachkommunikation teilnehmenden Personen aus Sicht der Soziologie betrifft solche Gesichtspunkte wie u.a. Alter, Geschlecht, sozialer und fachlicher Status, kulturelle Einbettung und persönliche Nähe. Die psychologischen Gesichtspunkte beziehen sich u.a. auf den psychischen Zustand, intellektuelle Fähigkeiten, Motivation und Erwartungshaltung der Personen. Zu den kommunikationswissenschaftlichen Gesichtspunkten gehören u.a. das Kommunikationsmedium, verschiedene Zeichenarten, Anzahl der Personen und direkte bzw. indirekte Fachkommunikation (vgl. Roelcke 2020: 25–26). Zu den anderen Disziplinen außer Soziologie, Psychologie und Kommunikationswissenschaft, die vom pragmalinguistischen Modell umfasst werden, gehören auch Semiotik, Statistik, Komparativistik, Terminologielehre u.a. (vgl. Baumann 1992: 183–185). Diese Interdisziplinarität verschafft dem Modell gegenüber dem systemlinguistischen Ansatz gewisse Vorteile, welche sich in der Beschreibungsgenauigkeit, Erfassung kommunikativer Prozesse und Anwendbarkeit von Forschungsergebnissen äußern. Konkret bedeutet dies Folgendes: je feiner die außersprachlichen Bedingungen der Fachkommunikation bestimmt werden können, umso effektiver wird die Beschreibung durch die innersprachlichen Merkmale.

Außerdem liegt nicht mehr nur das Resultat der Fachkommunikation, sondern auch der Prozess im Fokus der Fachsprachenforschung, bei dem kotextuelle und kontextuelle Faktoren berücksichtigt werden. Diese Vorteile bilden eine fundierte Grundlage für Fachsprachennormung, -kritik und -didaktik, wodurch die praktische Ergebnisanwendung der Resultate steigt, was den weiteren Vorteil darstellt (vgl. Roelcke 2020: 27–28).

Seit Mitte der 80er Jahre wird die „kognitive Wende“ in der Fachsprachenforschung beobachtet (vgl. Baumann 2010: 1). Das kognitionslinguistische Funktionsmodell befasst sich ebenfalls mit den von den beiden früheren Modellen berücksichtigten und vertretenen Gesichtspunkten, aber gemäß der neuen Sichtweise stellen die kognitiven Anlagen des Menschen den Ausgangspunkt der linguistischen Analyse, die sowohl die Form als auch die Funktion fachsprachlicher Äußerungen von den intellektuellen Anforderungen des Menschen herleitet (Roelcke 2020: 29). Infolge der Neuorientierung der Fachsprachenforschung schlägt Hoffmann 1993 eine Definition aus kognitiv-kommunikativer Perspektive vor:

Fachkommunikation ist die von außen oder von innen motivierte bzw. stimulierte, auf fachliche Ereignisse oder Ereignisabfolgen gerichtete Exteriorisierung und Interiorisierung von Kenntnissystemen und kognitiven Prozessen, die zur Veränderung der Kenntnissysteme beim einzelnen Fachmann und in ganzen Gemeinschaften von Fachleuten führen. (Hoffmann 1993: 614)

Aus der Definition lässt sich herleiten, dass der Inhalt und die Struktur des Systems einer Fachsprache davon abhängen, dass und wie die Fachleute über dessen Bestandteile kommunizieren bzw. welche fachsprachlichen Äußerungen sie tätigen, die wiederum von ihren kognitiven Anlagen abhängen und sich durch sie erklären lassen (vgl. Engberg 2017: 118–119; Roelcke 2020: 29).

Zu den Bereichen menschlicher Kognition, die relevant für die kognitive Linguistik sind, zählen nach Roelcke (2020) die Folgenden.

- Anlage des Kenntnis- und des Verarbeitungssystems des menschlichen Geistes, vor allem die sprachlichen Möglichkeiten [...]
- Abstraktion und Konkretisation als kognitiv-linguale Verfahren der Verallgemeinerung und der Vereinzelung von Konzepten sowie Assoziation und Dissoziation als kognitiv-linguale Verfahren synthetischer und analytischer Urteile [...]
- Interiorisierung und Exteriorisierung von Kenntnissen, also der kognitiv-linguale Erwerb und die kognitiv-linguale Vermittlung von Wissen und Erkenntnis [...] (Roelcke 2020: 29-30)

Im Rahmen dieser Bereiche wirken Sprache und Denken zusammen, es werden sprachlich-kognitive Aspekte abgedeckt, die es den Menschen ermöglichen, das Wissen zu erwerben, zu verarbeiten und zu vermitteln.

5.1.3. Funktionale Eigenschaften von Fachsprachen

Fachsprachen werden durch bestimmte funktionale Eigenschaften gekennzeichnet, die sich allerdings unterschiedlich äußern, abhängig von dem jeweiligen System (siehe 5.1.2.). Wie bereits angedeutet, steht der pragmlinguistische Ansatz nicht im Widerspruch zum systemlinguistischen Ansatz, sondern er bezweckt ihn zu erweitern und zu differenzieren. Das kognitionslinguistische Modell berücksichtigt die system- und pragmlinguistischen Konzeptionen, aber es strebt an, sie neu zu bestimmen und zu bewerten (vgl. Roelcke 2020: 32). Angesichts dessen fällt die Interpretation der Eigenschaften unterschiedlich aus.

Den Fachsprachen wird in erster Linie ihre Deutlichkeit beigemessen, was bedeutet, dass sie sich adäquat auf fachliche Gegenstände der Realität beziehen sollen. Der systemlinguistische Ansatz äußert diese Funktion vorwiegend durch die Bedeutung vom Fachvokabular. Die pragmlinguistische Konzeption berücksichtigt Kontext und Kotext der fachsprachlichen Äußerungen bei der Herstellung des Bezugs zur Realität, mit anderen Worten wird die Bedeutung eines Fachausdrucks samt seiner Einbettung im Satz in Betracht gezogen. Das kognitionslinguistische Modell beachtet sowohl das sprachliche System als auch die sprachlichen Äußerungen, aber die Deutlichkeit der Fachsprachen hängt auch von diversen kognitiven Merkmalen der Produzent:innen und Rezipient:innen ab (vgl. Roelcke 2020: 32). Den Fachsprachen wird außerdem ihre Verständlichkeit zugesprochen, das heißt, sie sollen fachliche Inhalte auf unmissverständliche Art und Weise vermitteln. Darunter wird nicht nur der Bezug zur Realität selbst verstanden, sondern vielmehr, dass er durch Rezipient:innen hergestellt werden kann. Dem systemlinguistischen Modell zufolge werden Fachsprachen durch den Fachwortschatz und zum geringeren Teil durch die für sie spezifische Syntax verständlich. Im Sinne des pragmlinguistischen Ansatzes wird die Verständlichkeit der Fachsprachen erst dann garantiert, wenn der Kotext und Kontext erschlossen werden. Allerdings wird gemäß dieses Systems nicht von der Verständlichkeit eines Fachtextes an sich, sondern seiner Verständlichkeit für eine bestimmte Rezipient:innengruppe ausgegangen. Das kognitive Modell ergänzt die Verständlichkeit der fachlichen Äußerungen dadurch, wie sie den kognitiven Prozessen und Strukturen entsprechen (vgl. Roelcke 2020: 32–33). Weiters werden Fachsprachen durch ihre Ökonomie, bzw. Kürze gekennzeichnet. Im Mittelpunkt dieser Eigenschaft steht der sprachliche Aufwand, der zwar möglichst minimal gehalten werden, aber gleichzeitig die fachliche Darstellung maximal übermitteln soll. Es handelt sich dabei nicht um eine funktionale Eigenschaft der Fachkommunikation an sich, denn die Ökonomie richtet sich stark z. B. nach dem Grad fachlicher

Kompetenz der Produzent:innen und Rezipient:innen, weswegen sie unterschiedlich ausfallen kann. In dem system-, pragma- und kognitionslinguistischen Modell wird die Kürze der Fachsprachen respektive durch das sprachliche System, die fachlichen Äußerungen und dadurch, wie sie den kognitiven Prozessen und Strukturen entsprechen, geäußert (vgl. Göpferich 2007: 412; Roelcke 2020: 33). Unter Anonymität wird eine weitere Eigenschaft der Fachsprachen verstanden, die als möglichst sachlich und gefühllos dargestellt werden. Es gilt die Annahme, dass die fachlichen Sachverhalte bzw. Abläufe keinen Bezug zu den Produzent:innen brauchen und vielmehr für sich selbst sprechen sollen, damit sie beispielsweise durch die Sichtweise der Produzent:innen nicht „verstellt“ werden. In der jüngeren Forschung wird allerdings postuliert, dass auch die Fachkommunikation einen kognitiv-kommunikativen Prozess darstellt und als solche nicht von Emotionen und dem Produzent:innenbezug befreit ist. Deswegen wird Anonymität von einigen Autor:innen als eine typische Eigenschaft der Fachsprachen bestritten und vielmehr als Produkt einer sprachphilosophischen Position betrachtet, die besagt wie Fachsprachen zu produzieren und zu rezipieren sind (vgl. Baumann 2004: 112; Roelcke 2020: 33–34). Die Fachsprachen können außerdem die Gruppenzugehörigkeit der im gleichen Fachbereich oder im gleichen Betrieb tätigen Personen stärken. Diesbezüglich wird Identitätsstiftung als eine weitere funktionale Eigenschaft von Fachsprachen genannt. Sie ist charakteristisch für gewisse Einrichtungen, Forschungszentren und kann auch ausschlaggebend für Corporate Identity sein. Die identitätsstiftende Funktion wirkt sich auf andere Aspekte aus, wie Verständlichkeit und Kürze, denn die fachlichen Texte werden anders produziert und rezipiert, wenn sie den inneren Kreis betreffen (vgl. Roelcke 2020: 34).

5.1.4. Merkmale fachsprachlicher Lexik

In der Fachsprachenforschung gibt es ähnlich wie im Fall von Fachsprache keine einheitliche Definition des Fachworts und Fachwortschatzes, daher optiert Roelcke für eine Definition, die seiner Ansicht nach im Allgemeinen einen relativ großen Zuspruch erwarten lässt:

Ein Fachwort ist die kleinste bedeutungstragende und zugleich frei verwendbare sprachliche Einheit, die innerhalb der Kommunikation eines bestimmten menschlichen Tätigkeitsbereichs gebraucht wird. Ein Fachwortschatz ist eine Menge solcher kleinster bedeutungstragender und zugleich frei verwendbarer sprachlicher Einheiten, die innerhalb der Kommunikation eines bestimmten menschlichen Tätigkeitsbereichs gebraucht werden. (Roelcke 2020: 70)

Der Fachwortschatz wird in vier Gruppen untergliedert. Die intrafachliche Lexik besteht lediglich aus den Fachwörtern, die die jeweilige Fachsprache betreffen. Interfachlich ist der Fachwortschatz, wenn er der betroffenen Fachsprache und anderen Fachsprachen angehört. Die extrafachliche Lexik kommt zwar aus den anderen Fachsprachensystemen, aber sie wird in den Fachtexten der betroffenen Fachsprache verwendet. Die letzte Gruppe besteht aus nichtfachlichem Fachsprachwortschatz, der in den Fachtexten der jeweiligen Sprache vorkommt (vgl. Roelcke 2020: 71). Zu den Beispielen der vier Gruppen im Fach der Biologie zählen respektive solche Wörter wie: Genom; Klassifikation; Artenschutz; Mensch. Zu einem weiteren Merkmal der Fachausdrücke gehört ihre Definiertheit (vgl. Pikor 2023: 175). Die Definitionen, u.a. die klassische aristotelische Definition, explikative Definition, exemplarische Definition, genetische Definition, assoziative Definition, tragen zur Konstitution der Fachwortschatzsysteme bei. Weitere typische Merkmale sind: Exaktheit, Eindeutigkeit und Eigentlichkeit (vgl. Roelcke 2020: 108–109). Zu den oft vorkommenden Wortbildungsmustern gehören u.a. substantivierte Infinitive, wie *Fräsen, Eggen, Eindocken*; Adjektive mit Präfixoid nicht-, die wertneutrale Bildungen ermöglichen, wie *nichtrostend, nichteuklidisch; nichtchristlich* vs. *unchristlich*; vielgliedrige Komposita, wie *Drehstromkurzschlußläufermotor*; Bildungen aus und mit Eigennamen, wie *pasteurisieren, röntgen, Bunsenbrenner, Morsezeichen*; Komposita mit Ziffern, Buchstaben und Sonderzeichen, wie *60-Watt-Lampe, V 8-Motor* u.a. (vgl. Möhn & Pelka 1984: 16–19).

5.1.5. Merkmale fachsprachlicher Grammatik

Fachsprachliche Grammatik unterscheidet sich von der gemeinsprachlichen Grammatik durch ein paar charakteristische Eigenschaften im Bereich der Morphologie und Syntax. Es handelt sich dabei allerdings nicht um systemeigene grammatische Strukturen, sondern vielmehr darum, dass gewisse grammatische Aspekte öfter in den Fachsprachen als in der Gemeinsprache auftreten (vgl. Grucza 2013: 27). Dazu zählen u.a. die Präsensdominanz, vorherrschend Indikativ, valenzspezifische Verbformen, der häufigere Genitivgebrauch, Bevorzugung der Satzglieder statt Gliedsätze, spezifische Pluralformen, viele Passiv- und Reflexivkonstruktionen sowie erhöhtes Vorkommen an Präpositionalkonstruktionen und Attribuierungen (vgl. Möhn & Pelka 1984: 19–21; Roelcke 2020: 127). Aus der Sicht der funktionalen Fachsprachengrammatik werden vier funktionale Interpretationen genannt. Die Inventarerweiterung äußert sich u.a. durch die erhöhte Zahl an Zusammensetzungen und Ableitungen; die Deutlichkeit durch typische Genitiv- und Pluralformen sowie die Dominanz der Konditional- und Finalsätze; die Aus-

drucksökonomie in Form von Kürzungen und Zusammensetzungen; die Anonymisierung mithilfe von Konversionen, Passivkonstruktionen und Funktionsverbgefügen (vgl. Roelcke 2020: 122–123).

5.1.6. Fachsprache und Gemeinsprache

Die Fachsprachenforschung ist seit längerer Zeit darum bemüht, die genaue Grenze zwischen der Gemeinsprache und den Fachsprachen zu setzen (vgl. Arntz et al. 2014: 1). Dies geschieht beispielsweise in Form von verschiedenen Modellen, die anschließend zusammenfassend dargestellt werden. Damit unterschiedliche Auffassungen, wie die Gemeinsprache und Fachsprachen voneinander zu differenzieren sind, nachvollzogen werden können, soll eine der Definitionsversuche aufgegriffen werden:

Als Gemeinsprache werden diejenigen Teile des Gesamtsprachrepertoires der Gesellschaft bezeichnet, die in den für alle Gesellschaftsmitglieder einigermaßen ähnlichen Lebensbereichen gebräuchlich sind und sich auf allgemein bekannte Gegenstände, Sachverhalte und Vorstellungen beziehen. (Ammon 1977: 75)

Daraus lässt sich die Vorstellung der Gemeinsprache als Teil der Gesamtsprache, die allgemeine bzw. alltägliche Themen betrifft, herleiten. Baldinger (1952: 65ff) konzentriert sich in seinem Modell auf die Lexik, die er als den abgrenzenden Faktor zwischen der Gemein- und den Fachsprachen sieht. Lt. dem Autor kann der Wortschatz in drei Kreise unterteilt werden: den Kreis mit dem gemeinsprachlichen Teil der gesamten Lexik der Sprache, den Kreis mit ausschließlich fachsprachlichem Wortschatz und den mittleren Kreis, der Fachausdrücke enthält, die starken Bezug zur Gemeinsprache haben (vgl. 1952: 65ff). Die Grenzen zwischen den einzelnen Kreisen und somit zwischen der Gemein- und den Fachsprachen sind dem Autor nach fließend und dies sowohl individuell als auch generell, denn „es findet zwischen den 3 Kreisen in beiden Richtungen ein dauernder Ausgleich statt“ (Baldinger 1952: 90). Das von Heller (1970: 532) vorgeschlagene Modell stellt ebenfalls den Wortschatz in den Mittelpunkt der Auseinandersetzung. Der Autor differenziert zwischen dem fachbezogenen und nicht fachbezogenen Wortschatz sowie zwischen dem allgemeinverständlichen und nicht allgemeinverständlichen Wortschatz. Außerdem werden im Modell explizit Fremdwörter, Dialektismen und Jargonismen genannt, die dem Autor nach zu der Gemeinsprache gehören, was in der Fachliteratur als „problematisch“ bezeichnet wurde (vgl. Arntz et al. 2014: 13). In den beiden Modellen wird

auf Gegensatzpaare zurückgegriffen, um zwischen der Gemein- und den Fachsprachen auf Basis der Fachlexik zu differenzieren, allerdings wird gleichzeitig der fließende Charakter der Abgrenzung hervorgehoben. Hoffmann (2022: 167) sieht die Gemeinsprache und Fachsprachen als Subsprachen der Gesamtsprache bzw. Nationalsprache. Die Gesamtsprache enthält alle sprachlichen Mittel, auf die die Gemeinsprache und Fachsprachen zugreifen können, um bestimmte Kommunikationshandlungen zu realisieren (siehe auch Arntz et al. 2014: 17). Außer der Gemeinsprache werden u.a. die Fachsprache der Physik und der Elektronik genannt, was bedeutet, dass die Fachsprachen lt. dem Autor nach dem jeweiligen Gegenstand unterschieden werden. In dem Sinne bilden Gemeinsprache und Fachsprachen nicht Gegenpaare, sondern sie sind alle der Gesamtsprache untergeordnet, wie in Abbildung 16 dargestellt.

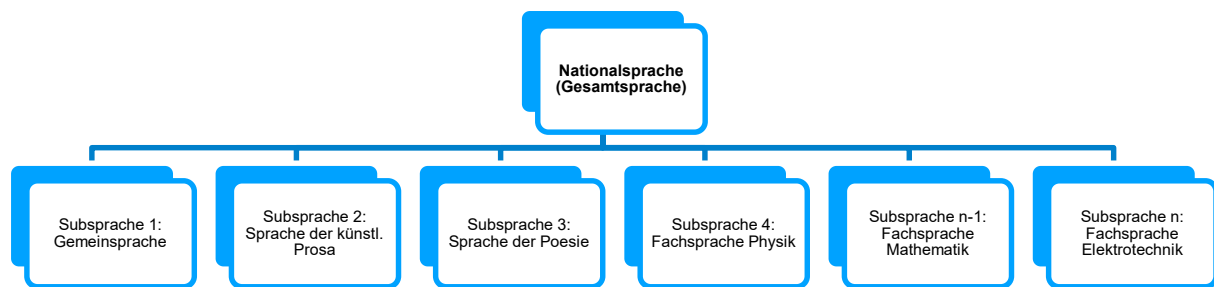


Abbildung 16: Das Schema: Gesamtsprache und Subsprachen (Hoffmann 2022: 167)

Fluck (1996: 160) betont ebenfalls die Problematik der Abgrenzung der Gemeinsprache von den Fachsprachen und sieht sie als verschiedene Realisierungen eines Sprachsystems, die auf diversen Ebenen liegen. Einerseits unterscheiden sie sich in Bezug auf die Semantik und ihre Funktion, andererseits sind sie interdependent, denn sie sind aufeinander bezogen. Gemeinsprache und Fachsprachen bilden dem Autor nach keine selbständigen Sprachsysteme, sondern Fachsprachen sind eher „durch Differenzierung und Erweiterung aus der Gemeinsprache herausgewachsen“ (Fluck 1996: 175–176). Der größte Unterschied besteht darin, dass Fachsprachen auf die Gemeinsprache angewiesen sind und ohne sie nicht vorhanden wären, während die Gemeinsprache alleine existieren kann (vgl. 175-176). Die oben genannte Interdependenz äußert sich zum Beispiel durch die in der letzten Zeit verstärkte Übernahme des Fachwortschatzes der Wissenschaft und Technik in die Gemeinsprache, was sich damit begründen lässt, dass die moderne Gesellschaft mehr Technik z. B. in Form von Geräten und Wissenschaft z. B. in Form von Dokumentarfilmen und Artikeln konsumiert. Solche Übernahmen finden nicht nur

im Bereich der Lexik statt, sondern sie betreffen auch die Syntax und bestimmte Denkformen (vgl. Fluck 1996: 161), was nachfolgend genauer erörtert wird.

Der Fachwortschatz, der von der Gemeinsprache internalisiert wird, kommt aus den fachlichen Bereichen, die starken Einfluss auf das alltägliche Leben der Gesellschaft ausüben. Außerdem, so Hartmann (1980: 42), liegt es an den Grundprinzipien einer demokratisch aufgebauten Gesellschaft, die Recht auf Mitentscheidung hat und somit über diverse Forschungsergebnisse sachgerecht informiert werden will. Zu den bereits erwähnten Bereichen der Technik und Wissenschaft kommen noch Politik, Wirtschaft, Freizeitbereich, Medizin, Umweltschutz u.a. dazu, mit denen Lai:innen durch Massenmedien in Kontakt kommen. Dies führt zur verstärkten Wortschatzerweiterung in der Gemeinsprache. Fluck (1996: 162–163) zufolge ist es ein natürlicher Prozess, sei es wegen des steigenden Interesses der Gesellschaft z. B. am Sport oder aufgrund der Tatsache, dass die Technologie aus vielen Arbeits- und Freizeitbereichen nicht mehr wegzudenken ist. Außerdem kommen fachsprachliche Elemente in vielen gemeinsprachlichen Redewendungen, Phrasen und Metaphern vor, wie z. B. *durchforsten*, *alle Register ziehen*, *am längeren Hebel sitzen* usw. Weiters fließen durch die Fachsprachen viele Fremdwörter in die Gemeinsprache ein (vgl. Hartmann 1980: 27), darunter z. B. *Talkshow*, *Stand-up*, *Festival*, *Boom*. Diese Fremdwortübernahme hängt von der „Transformationskraft“ der Nationalsprache (Fluck 1996: 164) und von der nationalen Sprachpolitik ab. Als Beispiel für den Einsatz sprachpolitischer Maßnahmen gilt Isländisch, die Sprache, dessen (Fach)wortschatz stets erweitert wird, indem neue Ausdrücke, basierend auf dem bestehenden lexikalischen Bestand, erfunden werden. Das Ziel des isländischen Sprachpurismus ist der Erhalt der Nationalsprache, Steigerung nationaler Identität und der Gesellschaft die Möglichkeit zu geben, Isländisch in allen Lebensbereichen einsetzen zu können (vgl. Icelandic Language Institute 2001: 7).

Der Einfluss der Fachsprachen auf die Grammatik der Gemeinsprache äußert sich in ihrem fortschreitenden Nominalstil. Den Fachsprachen ähnlich wird in der Gemeinsprache die Tendenz beobachtet, immer mehr Inhalt so knapp und präzise wie möglich vermitteln zu wollen. Die Beispiele dafür sind der bevorzugte Gebrauch von Funktionsverbgefügen und Komposita, die syntaktische Flexibilität der Sprache fördern (vgl. Fluck 1996: 166), erhöhte Zahl an Präpositionalgefügen und erweiterten Attribuierungen sowie Tendenz zur Univerbierung und Zunahme von Abkürzungen und Bindestrichwörtern (vgl. Fluck 2018: 246). Aus der morphologischen Sicht wird in der Fachliteratur der immer häufigere Präfixgebrauch bzw. die verstärkte Verwendung von -er Derivaten in der Gemeinsprache genannt (vgl. 2018: 246).

Die moderne Fachsprachenforschung verweist außerdem auf die Intellektualisierung bzw. Verwissenschaftlichung der Gemeinsprache. Diese Erscheinung ist auf die lexikalischen und syntaktischen Übernahmen aus den Fachsprachen zurückzuführen und tritt u.a. dann auf, wenn spezialisierende Ausdrücke aufgrund ihrer Eindeutigkeit bevorzugt werden wie z. B. *Frischfleisch* oder umgekehrt, wenn Bedarf nach generalisierenden Abstrakta besteht z. B. *Objekt*, *-fest* oder *-arm* (Fluck 1996: 168). Die Komposita sind außerdem eine praktische Lösung, die umfassenden Ereignisse und Sachverhalte in der Gemeinsprache zu benennen, deren Beschreibung „schwierig, wenn nicht unmöglich“ ausfallen würde, wie z. B. *Abwehrsystem* oder *Bildungssystem* (vgl. 1996: 168). Zu einer anderen Art der Übernahme fachsprachlicher Elemente zählt ihre Funktion in der Gemeinsprache als literarisches Stilmittel oder als Verkaufshilfe in der Werbesprache (Fluck 1996: 169). Die Gefahr der „Überintellektualisierung“ und „Entmenschlichung“ der Gemeinsprache durch den Einfluss der Fachsprachen z. B. im Falle der Verwendung unpersönlicher Ausdrucksweise bei der Kollektivierung von Menschen: *Verkehrsteilnehmer:innen*, *Versicherungspflichtige* usw. besteht lt. dem Autor nicht (vgl. Luck 1996: 164,168). Bezüglich der Sicherung der Verständlichkeit der Gemeinsprache stellt sich beispielsweise die Frage, ob und inwieweit pseudofachliche und pseudowissenschaftliche Jargonismen in der Gemeinsprache zu vermeiden sind, damit es nicht zur überhöhten Verwendung der Fachausdrücke außerhalb der Fachwelt kommt (vgl. Bausch et al. 1976: 16; Hartmann 1980: 43).

5.2. Rechtssprache

Das Verhältnis zwischen Recht und Sprache wird immer wieder zum Gegenstand von Debatten, sowohl unter Jurist:innen als auch Philolog:innen (vgl. Knauer 2007: 24) und unterliegt stetem Wandel, weswegen in der Wissenschaft Unklarheit darüber herrscht, wie die Rechtssprache linguistisch und rechtstheoretisch zu definieren und positionieren ist. Während die technischen und naturwissenschaftlichen Sachverhalte von Natur aus universell sind und somit von der Sprache abgekoppelt werden können, lässt sich das Recht nur durch die Sprache zum Ausdruck bringen und ist an sie gebunden. Dies erschwert z. B. wesentlich die Übertragung juristischer Fachausdrücke in andere Sprachen, denn es müssen dabei stets unterschiedliche Rechtsordnungen einzelner Staaten berücksichtigt werden. Es gibt nämlich keine rechtsordnungsübergreifende deutsche Rechtssprache. Die Übersetzung ist daher „(...) zugleich eine juristische Aufgabe, die sich nicht selten der Rechtsvergleichung annähert“ (vgl. Arntz 2001: 99; Arntz et al.

2014: 162–163; Sandrini 1996: 16, 1999: 15). Als Beispiel dafür nennt Wiesmann (2007: 643) den italienischen Fachausdruck *amministratore*, der in Südtirol bei Kapitalgesellschaften als *Verwalter* und bei Personengesellschaften als *Geschäftsführer* und in Deutschland bei Aktiengesellschaften als *Vorstandsmitglied* und bei Gesellschaften mit beschränkter Haftung als *Geschäftsführer* zu übersetzen ist.

5.2.1. Definition der Rechtssprache

Einer der Gründe, warum die Bestimmung von Rechtssprache sich als problematisch erweist, ist ihre vergleichsweise hohe Anzahl gemeinsprachlicher Elemente, deren Bedeutung zum Teil modifiziert bzw. speziell angepasst wird. Demnach handelt es sich zwar in der Form um gemeinsprachliche Ausdrücke, welche dennoch in der Rechtssprache auf der semantischen Ebene Unterschiede aufweisen. Ein Beispiel dafür stellt der Begriff *uneheliches Kind* dar, dessen Verwandtschaft zu seinem biologischen Vater bis 1969 im Bürgerlichen Gesetzbuch bestritten wurde, was aus der Sicht der Biolog:innen nicht nachvollziehbar ist (vgl. Oksaar 1979: 101–102; Sandrini 1996: 17). Ermert (1983: 12) äußert sich zu diesem besonderen Merkmal der Rechtssprache folgendermaßen:

Die juristische Fachsprache benutzt vermutlich die wenigsten Fremdwortfachwörter aller Wissenschaften. Aber sie ist nicht weniger terminologisch „aufgeladen“ als andere Fachsprachen, und sie ist für den juristischen Laien in gewisser Weise tückischer und gefährlicher, gerade weil sie umgangssprachliche Ausdrücke terminologisch gebraucht. (Ermert 1983: 12)

Dass die Rechtssprache sich diesbezüglich von den anderen Fachsprachen unterscheidet, liegt möglicherweise daran, dass das Recht bzw. die Rechtstexte sich an die Gesamtheit der Gesellschaft richten. Außerdem ist das rechtliche Denken stark mit dem menschlichen Leben verknüpft und als solches auf die Beschreibung menschlicher Handlungen angewiesen (vgl. Oksaar 1979: 101). Die vielen gemeinsprachlichen Ausdrücke in der Rechtssprache haben einige Autor:innen dazu verleitet, ihre Existenz in Frage zu stellen. Hoffmann (1989: 14) nach handelt es sich dabei eher um „Ausdifferenzierungen der natürlichen Sprache“, die Rechtsterminologie umfassen und vom juristischen Stil geprägt werden. Zudem weisen diese Ausdifferenzierungen, so der Autor, gewisse Merkmale auf, wie Verwendung unpersönlicher Ausdrücke, Archaismen, Attributhäufung, generalisierende Ausdrücke u.a. (vgl. 1989: 14). Wassermann (1979: 117) räumt der Juristensprache den Status einer fachlichen Umgangssprache ein und sieht sie nicht als eine streng formalisierte Theorie- bzw. Wissenschaftssprache ein. Mit anderen Worten kann

es gar keine Rechtssprache im Sinne von Fachsprache geben, da sie kein autonomes und einheitliches Sprachsystem bildet sowie größtenteils von der Gemeinsprache schöpft und sie nur zusätzlich durch Rechtsterminologie ergänzt (vgl. Kęsicka 2020: 8). Kirchhof (1977: 236) vertritt eine andere Auffassung und konstatiert, dass die subjektive und mehrdeutige Gemeinsprache nicht dafür geeignet ist, die Objektivität des Rechts mit der notwendigen Präzision, Deutlichkeit und Kürze wiederzugeben, womit sich das Vorhandensein der Rechtssprache begründen lässt. Auch Schmidt (1972: 390) betrachtet die Rechtssprache als eine Fachsprache und nicht als eine Umgangssprache und spricht in seiner Definition ihre Adressierung an mehrere Zielgruppen an.

Unter *Rechtssprache* möchte ich im Folgenden die Gesamtheit aller rechtlichen Aussagen verstehen, die wiederum identisch sind mit dem Verbalverhalten aller Leute, in dem ethische Werturteile über menschliches Verhalten wesentlich vorkommen. Damit sind nicht nur Gesetzgeber, Richter, Rechtsanwälte, Universitätsprofessoren des Rechtes und ähnliche berufsmäßig mit rechtlichen Aussagen befaßte Leute betroffen, sondern auch der „Mann auf der Straße“, soweit er nur über menschliches Verhalten ethische Werturteile abgibt. (Schmidt 1972: 398)

Knauer (2007: 24) gibt wiederum an, dass es verschiedene Rechtssprachen gibt und nennt explizit: „Gesetzessprache, Verwaltungssprache, Sprache der Rechtspflege sowie die des Laienverkehrs“. Weiters kommen auch Unterschiede innerhalb der einzelnen Fachbereiche des Rechts vor, wie z. B. im Polizeirecht oder Wirtschaftsrecht, die durch für sie spezifische Fachjargons gekennzeichnet sind (vgl. 2007: 24). Auch Trosborg (1992: 12) verbindet mit dem Begriff Rechtssprache verschiedene Teilbereiche des Rechts und die für sie spezifischen Subsprachen, die konstruiert wurden, um die bestimmte Funktion in der spezifischen kommunikativen Situation zu erfüllen (vgl. 1992: 12). Zusammenfassend wird der Status der Rechtssprache als Fachsprache dadurch bekräftigt, dass sie fachbezogen ist, über eigenes fachspezifisches Repertoire verfügt, der internen und externen Kommunikation zwischen diversen Parteien innerhalb und außerhalb des Faches dient und sich durch die besondere Stilistik von der Gemeinsprache und von den anderen Fachsprachen abgrenzt (vgl. Kęsicka 2020: 8).

5.2.2. Merkmale der Rechtssprache

In den vorigen Unterkapiteln wurde bereits verdeutlicht, dass den Fachsprachen gewisse Merkmale wie Anonymität, Kürze, Eindeutigkeit u.a. zugeschrieben werden. Die Rechtssprache

wird ebenfalls von diesen Merkmalen geprägt, gleichzeitig weist sie aber gewisse charakteristische Eigenschaften auf, die ihnen widersprechen (vgl. Arntz & Sandrini 2007: 137), was in weiterer Folge genauer angesprochen wird.

Dass die Rechtssprache exakt und eindeutig sein muss, ergibt sich beispielsweise daraus, dass Rechtssprüche konkrete Auswirkungen auf das Leben jedes/jeder Einzelnen haben und behördliche Konsequenzen mit sich bringen können. Damit die Objektivität des Rechts und folglich auch Rechtssicherheit gewährleistet werden kann, sollen Rechtsvorschriften klar und unmissverständlich sowie unter Berücksichtigung der Sprachökonomie formuliert werden. Dies erweist sich allerdings als problematisch, was auf die Natur des Rechts zurückzuführen ist, nämlich auf seinen hohen Abstraktionsgrad und seine „Abhängigkeit von den jeweiligen gesellschaftlichen Rahmenbedingungen“ (vgl. 2007: 137). Sollten daher Rechtsvorschriften allzu präzise formuliert werden, schränkt dies ihre Anwendbarkeit ein, was wiederum die Notwendigkeit für immer mehr Regeln mit sich bringen würde, denn es ist nicht von Anfang an genau absehbar, welche Fälle zukünftig vorkommen werden. Um dem entgegenzuwirken und sicherzustellen, dass die jeweilige Vorschrift gleich für mehrere Ausprägungen einer bestimmten Rechtssache anwendbar ist, muss beim Formulieren eines Gesetzestextes stets darauf geachtet werden, dass sie „einen gewissen Anwendungsspielraum lassen“ (vgl. 2007: 137). Daraus ergibt sich ein Paradox: die Rechtsvorschriften sind zwar exakt zu formulieren, damit die Rechtssicherheit nicht gefährdet wird, aber gleichzeitig auch abstrakt und allgemein genug, damit sie vielseitig anwendbar bleiben. Der weitere Aspekt, der dafür spricht, die Rechtstexte allgemein zu formulieren, betrifft die herrschenden gesellschaftlichen Bedingungen. Das Recht ist auf die gesellschaftliche Realität bezogen und unterliegt somit einem konstanten Wandel und damit die Regeln bzw. Normen dem ethischen und moralischen Zeitgeist entsprechen, soll die Formulierung einer Vorschrift, unter der Berücksichtigung der Zweckmäßigkeit, vom Beginn an einen bestimmten Abstraktionsgrad aufweisen (vgl. Arntz & Sandrini 2007: 138). Die Autoren (2007: 138–139) unterscheiden drei Ebenen der Vagheit bei der Formulierung der Rechtstexte.

- Intrinsische Vagheit, die in der Natur des jeweiligen Rechtssystems liegt und weder gewollt noch beeinflussbar ist. Sie ist auf die oben genannten gesellschaftlichen Bedingungen und den ethisch-moralischen Kontext der Zeit zurückzuführen.
- Gewollte Vagheit, die eine Gruppe der Rechtsbegriffe, die sog. unbestimmten Rechtsbegriffe, betrifft. Zu den Beispielen zählen u.a.: *Notstand*, *guten Glaubens*, *nach bestem Wissen und Gewissen*. Sie werden absichtlich unbestimmt formuliert, was eine gewisse Anpassungsfähigkeit des Rechts und somit Interpretationsmöglichkeiten im Falle vom Unvorhergesehenen zulässt.

- Zufällige Vagheit entsteht z. B. dann, wenn Rechtsbegriffe falsch bestimmt werden, nicht klar genug differenziert sind, oder wenn gemeinsprachliche Ausdrücke ohne Definition terminologisiert werden. (vgl. Arntz & Sandrini 2007: 138-139)

Außer der Vagheit ist die Mehrdeutigkeit, in Rechtswissenschaften unter dem Begriff terminologische Relativität bekannt (vgl. Wank 2020: 46ff), ein weiterer Aspekt, der im Widerspruch zu der postulierten Präzision der Fachsprachen steht. Dabei handelt es sich um Fachausdrücke, die je nach Kontext und Gebrauch ihre Bedeutung ändern (vgl. Hegenbarth 1982: 103). Die Mehrdeutigkeit der Rechtssprache äußert sich in Form von Synonymen, Polysemen und Homonymen, es sind z. B. die synonymen Ausdrücke *Fahrerflucht* und *Verkehrsflucht*; die polysemischen Ausdrücke *Verfügung*, *Angehöriger*; die homonymen Ausdrücke *Gericht*, *Ruhen*, *Berufung* (vgl. Bielawski 2023: 45–47). Die Vagheit und Mehrdeutigkeit gehören zu den Merkmalen der Rechtssprache, die sie von den anderen Fachsprachen unterscheiden. Dies ist allerdings nicht als eine Schwachstelle oder Fehler zu sehen, sondern vielmehr als das Resultat eines auf Zweckmäßigkeit beruhenden Vorgehens, das dazu beiträgt, dass die Gesetzgebung elastisch und anpassungsfähig bleibt sowie sogar im gewissen Maße die Strafflosigkeit verhindert (vgl. Bielawski 2023: 47–48).

Eine weitere Eigenschaft der Rechtssprache, die häufig im Widerspruch zur Präzision der Fachsprache steht, ist die Anforderung auf Allgemeinverständlichkeit. Der vorab erwähnte Bezug der Rechtssprache auf das gesellschaftliche Leben setzt voraus, dass u.a. die Normen auch von Lai:innen nachvollziehbar sein sollen, damit alle Gesellschaftsmitglieder sich dessen bewusst sind, wie sie sich im Einklang mit Recht zu verhalten haben (vgl. Arntz & Sandrini 2007: 137). Demzufolge sollen Rechtstexte eine sehr niedrige Abstraktionsstufe aufweisen, aber in Wirklichkeit, so Sandrini (1996: 18–19), wird ihr Abstraktionsgrad zwischen niedrig und hoch eingestuft. Auch Bielawski (2023: 53) nach kann dem Allgemeinverständlichkeitsanspruch der Rechtssprache nicht stattgegeben werden, denn Fachsprachen sind vom Prinzip aus nicht allgemeingültig und müssen sich von der Gemeinsprache unterscheiden. Weiters sollte die in Rechtstexten verwendete Rechtssprache vereinfacht werden, würde dies nicht unbedingt zur höheren Allgemeinverständlichkeit des Rechts führen (vgl. 2023: 53).

Nominalstil ist eine weitere Eigenschaft, die mit der Rechtssprache in Verbindung gebracht wird. Die häufige Substantivierung verringert zwar die Deutlichkeit und macht somit die Rechtstexte weniger verständlich, aber sie entspricht dem Streben der Fachsprachen nach Ausdrucksökonomie (vgl. Knauer 2007: 45; Olejniczak 2018: 32; Sander 2004: 6). Nominalisierungen sind abstrakte Substantive, die vorwiegend aus Verben und Adjektiven gebildet werden.

Sie kommen oft im Satz zusammen mit solchen Verben wie *durchführen*, *erfolgen* und *vollziehen*. Im Zuge der Substantivierung wird der/die Handelnde weggelassen und die Prozesse sowie Aktivitäten wirken dadurch statisch und objektiv (vgl. 2007: 45; 2018: 32; 2004: 6). Dieses stilistische Mittel hebt die Autorität des Rechts hervor, trägt dazu bei, dass juristische Texte als sachlich empfunden werden und es hilft die Begriffsklarheit zu gewährleisten (vgl. Kęsicka 2020: 12). Substantivierungen entstehen durch: die Nominalisierung von Infinitiven: *vorliegen* → *das Vorliegen*; das Ableiten von Verben mithilfe von Suffixen -ung und -keit: *abstimmen* → *die Abstimmung* und durch abgeleitete Verbalsubstantive: *eingreifen* → *der Eingriff*; den Gebrauch von Präpositionalgefügen anstatt von Gliedsätzen: *nachdem Sie durchschritten haben* → *nach dem Durchschreiten*; die Verwendung von Funktionsverbgefügen: *beschließen* → *einen Beschluss fassen* (vgl. Höhmann 2010: 323ff; Sander 2004: 7). Zu den Beispielen aus der polnischen Rechtssprache gehören: Bei Handlungsbezeichnungen die Substantivierung von Verben mit den Suffixen -(a)nie, -(e)nie, -acja: *legalizacja*, *skazanie*, *pouczenie*; bei Personenbezeichnungen die Substantivierung durch Suffigierung mit -i(y)ciel, -ca oder Präfigierung mit pod-, wice-, współ-: *założyciel*, *podwykonawca*; Substantivierung von Adjektiven und Verben mit dem Suffix -ość: *nienaruszalność*, *pełnoletniość* (vgl. Hałas 1995: 46–51); die Substantivierung durch Infigierung: *(partnerstwo) publicznoprawne*, *(umowa) cywilnoprawna* (vgl. Pikor 2023: 177).

In der deutschen bzw. österreichischen und polnischen Rechtssprache wird durchgehend die männliche Form in Bezug auf alle Personen verwendet, was damit erklärt wird, dass die Verständlichkeit der Rechtssprache kompliziert genug ist und die geschlechtsneutrale Ausdrucksweise dies zusätzlich erschweren und die Texte länger machen würde. Als Beispiel dafür gilt ein Abschnitt aus der Verfassung des Bundeslandes Schleswig-Holstein: „Endet das Amt der Ministerpräsidentin oder des Ministerpräsidenten, so sind er oder sie und mit ihm oder ihr die Landesregierung...“. Andererseits besteht das Risiko, dass die weibliche Hälfte der Gesellschaft sich wegen des Gebrauchs der männlichen Form mit dem Gesetz nicht gänzlich identifizieren kann. Aufgrund der Forderung nach einer gendergerechten Sprache sowohl in deutschsprachigen Ländern als auch in Polen, wird seit einigen Jahren u.a. vom Europäischen Parlament empfohlen, geschlechtsbezogene und personenbezogene Begriffe durch neutrale und objektbezogene Begriffe zu ersetzen wie z. B. *Teilnehmende* (DE), *osoby uczestniczące* (PL). Zu anderen Beispielen der geschlechtsneutralen Sprache gehören im Deutschen: *antragstellende Person*, *Richter:innen*, *Richterinnen und Richter*, *Studienausweis* (statt *Studentenausweis*), *Elektriker* (m/w/d) und im Polnischen: *kandydat/ka*, *elektryk (k/m)*, *uczestnicy i uczestniczki*,

dyrektor generalny/generalna (vgl. Rzecznik Praw i Wolności Akademickich. Uniwersytet im. Adama Mickiewicza w Poznaniu 2024; Simon 2022: 32–33).

Zu den weiteren Merkmalen der Rechtssprache zählen ihre Anonymität, Affektpflege und ihr direkter Charakter. Die Unpersönlichkeit der juristischen Texte ergibt sich daraus, dass ihr Fokus auf der „Beschreibung und Benennung von Tatsachen, Handlungen, Objekten und Gegenständen“ liegt (Kęsicka 2020: 12–13). Die Parteien werden im weiteren Verlauf des Rechtstextes nicht personalisiert, sondern z. B. als *Antragsteller* oder *Beklagte* bezeichnet, außerdem wird in ihnen auf abstraktes Agens zugegriffen, z. B. *Das Gericht findet jedoch, dass die Entscheidung...* (vgl. 2020: 12–13). Die Rechtssprache kann einerseits sachlich und neutral sein, andererseits bedient sie sich der emotionalen Lexik der Gemeinsprache, die in ihrer ursprünglichen Bedeutung in die Rechtssprache übernommen wird. Dies betrifft solche Begriffe wie *Mord, Bande, sich eines Menschen bemächtigen* u.a., die es anvisieren, „das Unrecht zu verdeutlichen und das unerwünschte Handeln zu brandmarken“ (vgl. Kęsicka 2020: 13). Die affektive Seite der Rechtssprache kommt z. B. dann zum Einsatz, wenn das Ausmaß einer Straftat sowohl axiologisch als auch sprachlich für die Lai:innen nachvollziehbar dargestellt werden soll (vgl. 2020: 13). Weiters werden emphatische Ausdrücke gebraucht, wie *grob irrtümlich* oder *nicht zu erwarten sein, dass...*, die die Argumentation durch gezielte Akzentsetzung untermauern und somit zu einem wichtigen rhetorischen Stilmittel der Rechtssprache werden (vgl. Kęsicka 2020: 14). Der direkte Charakter der Rechtssprache ist auf ihre normative Funktion zurückzuführen. Sie regelt nämlich in Form von Verbotten und Geboten die sozialen Bedingungen (vgl. Müller 2010: 29). Dies wird durch den Gebrauch von „expliziten und primären performativen Äußerungen“ (vgl. 2010: 29; siehe auch Malinowski 2006: 96–98) erreicht, wie z. B. im Fall von Geboten mit den expliziten performativen Verben: *gebieten* und *verpflichten* sowie mit dem Modalverb *sollen*. Die Verbote bzw. das Erlauben werden durch solche Verben bzw. Äußerungen wie *untersagen, verbieten, (nicht) gestatten, (nicht) zulassen* sowie die Modalverben *können* und *dürfen* ausgedrückt. Primäre performative Äußerungen werden durch den Gebrauch von nicht expliziten Verben gekennzeichnet, deren performative Funktion erst durch den spezifischen Anwendungskontext zum Ausdruck gebracht wird. Dazu zählen z. B. solche Äußerungen wie *wird bestraft; ermäßigt sich; gilt entsprechend* u.a. (vgl. Kęsicka 2020: 15). Beispiele aus der polnischen Rechtssprache betreffen folgende Verben: *zobowiązywać, zobligować, nakazywać, zakazywać, zabraniać, móc, pozwalać, dozwalać, dopuszczać, zezwalać* (vgl. Pikor 2023: 175–176).

5.2.2.1. Lexikalische Besonderheiten der Rechtssprache

In der Lexik der Rechtssprache kommen, bedingt u.a. durch den Nominalstil, zahlreiche Komposita vor. Sie unterstützen die begriffliche Differenzierung und fördern Sprachökonomie, z. B. im juristischen, universitären Kontext: *Frist zur Meldung der Fortsetzung* → *Zulassungsfrist* (Studienservice und Lehrwesen der Universität Wien 2024). Gleichzeitig sollte davon abgesehen werden, zu lange bzw. komplexe Komposita zu bilden, denn dadurch kann die Verständlichkeit des Texts zu stark beeinträchtigt werden: *Das Fernstraßenbauprivatfinanzierungsgesetz* → *Das Gesetz zur privaten Finanzierung des Fernstraßenbaus* (vgl. Thormann 2016: 57–58). Weiters wird die Verpflichtung in den juristischen Texten vorwiegend mit den Konstruktionen *haben zu / sein zu* und nicht mit dem Modalverb *müssen* ausgedrückt: *Der Antragssteller muss die Frist einhalten* → *der Antragssteller hat die Frist einzuhalten*. Ein weiteres lexikalisches Merkmal der Rechtssprache ist der bevorzugte Gebrauch der Substantiv-Endung -e: *zum Zweck* → *zum Zwecke*, *im Fall* → *im Falle*, *zum Wohl* → *zum Wohle*. Dies ist auch bei der Wahl der Genitivformen sichtbar: *des Vertrags* → *des Vertrages*, *des Vollzugs* → *des Vollzuges* (vgl. Thormann 2016: 59). Ein weiterer Aspekt ist die tendenziell verstärkte Verwendung von Verben mit Vorsilben, die allerdings dadurch keinen großen Unterschied zu den gleichen Verben ohne Vorsilbern aufweisen: *mieten* → *anmieten*, *parken* → *abparken*, *ankündigen* → *vorankündigen*, *sterben* → *versterben*. Außerdem werden Adjektive mit dem Negationspräfix *un-* verwendet, wenngleich es in der Gemeinsprache andere Wörter gibt, die das Gleiche ausdrücken: *falsch* → *unrichtig*, *kostenlos* → *unentgeltlich*, *rechtswidrig* → *unrechtmäßig*. Weiters wird in der Rechtssprache statt der Ausdrucksweise *es gibt* das Verb *bestehen* gebraucht: *Es gibt das Risiko* → *Es besteht das Risiko* (vgl. Thormann 2016: 60). Im Zusammenhang mit den Lehnwörtern lassen sich in der deutschen und österreichischen Rechtssprache einige lateinische Ausdrücke beobachten wie z. B. *status quo*, *de iure*, *ex tunc*, *ex nunc*, da ihr das römische Recht zugrunde liegt. Auf die Latinismen wird allerdings nur dann zugegriffen, wenn keine äquivalenten deutschsprachigen Ausdrücke vorliegen: *Fotokopie* → *Ablichtung*, *circa* → *rund*, *etc.* → *usw.*, *Korrespondenz* → *Schriftwechsel*. Dasselbe Prinzip gilt für Anglizismen: *inklusive* → *einschließlich*, *Führungskräfte* → *Management*, *Sex* → *Beischlaf* (Ausnahme: *Geschmacksmuster* → *Design*) (vgl. Thormann 2016: 61). Die ersten und gleichzeitig meisten Entlehnungen in der polnischen Rechtssprache kommen aus dem Lateinischen und Altgriechischen, z. B. *dekret*, *kodeks*, *apelacja*, *amnestia*, *hipoteka* (vgl. Grzelak 2010: 203). Andere Lehnwörter kommen zum geringeren Teil aus dem Französischen, z. B. *arbitraż*, Englischen, z. B. *werdykt* und Deutschen, z. B. *koszt* (vgl. Grzelak 2010: 204; Łuksza 2006: 374).

Wie bereits erwähnt, kommt es in der Rechtssprache zu Terminologisierung der gemeinsprachlichen Ausdrücke. In dem Zusammenhang ist allerdings darauf hinzuweisen, dass Ausdrücke, die in der Gemeinsprache größtenteils synonym gebraucht werden, in der Rechtssprache eine klare, gesetzlich vorgeschriebene und durch Definition festgelegte Abgrenzung aufweisen (vgl. Moroz 2020: 164) und somit nicht synonym gebraucht werden dürfen. Dazu zählen u.a. solche Wortpaare wie *Arbeitstag*, also alle Tage, an denen gearbeitet wird, hauptsächlich von Montag bis Freitag, branchenabhängig auch Samstag und Sonntag, und *Werktag*, also alle Kalendertage ausschließlich Sonn- und Feiertage. Dies hat rechtliche Auswirkungen auf LKW-Fahrverbote, erlaubte Lärm-Emissionen, Parkregelungen usw. Zu weiteren Beispielen gehören *Anspruch* und *Forderung*, *anklagen* im Strafrecht und *verklagen* im Privatrecht, *Werkvertrag* und *Dienstvertrag*, *Abschrift* und *Ausfertigung* (vgl. Thormann 2016: 62–90). Weiters, wie bereits früher kurz angesprochen, liegen in der Rechtssprache gemeinsprachliche Wörter vor, denen im juristischen Kontext allerdings eine andere Bedeutung verliehen wird, z. B. *billig* im Sinne von *angemessen*, *berechtigt* und nicht als Gegenteil von *teuer*, außerdem auch u.a. *regelmäßig*, *Verfassung*, *versagen* (vgl. Thormann 2016: 92–96).

5.2.2.2. Grammatikalische Besonderheiten der Rechtssprache

Eine der grammatikalischen Besonderheiten der Rechtssprache betrifft den bevorzugten Gebrauch von attributiv fungierenden Partizipialkonstruktionen vor dem Bezugssubstantiv anstelle von Relativsätzen, z. B. *Es wurde zu dem Marktpreis fakturiert, der am Versandtag gültig war* → *Es wurde zu dem am Versandtag gültigen Marktpreis fakturiert*. Wenn Relativsätze in den juristischen Texten vorkommen, dann werden sie grundsätzlich mit *welch* und nicht mit einem Relativpronomen eingeleitet. Außerdem gehört zu einem weiteren Charakteristikum der Rechtssprache die häufige Verwendung von Passivformen, was der Anonymität also einem der Merkmale der Fachsprachen entspricht, z. B. *Sie können die Festsetzung der Vorauszahlung anfechten, indem Sie Einspruch erheben* → *Die Festsetzung der Vorauszahlung kann durch einen Einspruch angefochten werden*. Aktiv wird wiederum vorzugsweise in der 3. Person gebraucht und es wird generell die 1. und 2. Person Singular und Plural gemieden, damit der unpersönliche Stil der Rechtssprache beibehalten wird. Dasselbe Prinzip ist in der polnischen Rechtssprache vorhanden, z. B. *Nie stwierdza się nieważności decyzji z przyczyn wymienionych w § 1 (...), Osoba, której pochodzenie polskie zostało stwierdzone zgodnie z ustawą, może osiedlić się na terytorium Rzeczypospolitej Polskiej (...)* (vgl. Pikor 2023: 182; Thormann 2016: 24–35). Weiters wird die Rechtssprache deutlich mehr als die Gemeinsprache vom Genitiv ge-

prägt. Nicht selten entstehen Genitiv-Ketten, die in dem gemeinsprachlichen Gebrauch unvorstellbar wären. Der Rechtssprache hingegen verleihen sie ihren gehobenen Status, es wird dadurch impliziert, dass der Rechtstext ernst zu nehmen ist und nicht angezweifelt werden soll, wie in dem folgenden Satz *Die Entkräftung der Argumente des Mitarbeiters des Angeklagten gelang erst nach der Entdeckung der Tatwaffe des Komplizen*. In den juristischen Texten kommen außerdem häufiger als in der Gemeinsprache Präpositionen vor, die den Genitiv verlangen, wie *kraft*, *zulasten*, *eingedenk*, *mittels*, *vorbehaltlich*, *zeit*, *hinsichtlich*, *innen* u.a. (vgl. Thormann 2016: 42–43). Ein weiterer Aspekt der Rechtssprache betrifft den Satzbeginn. Viele Sätze in den juristischen Texten beginnen nämlich nicht mit dem Subjekt, ausgenommen Passivsätze, sondern mit einem Objekt, z. B. *Der Partei steht gegen das Urteil der Einspruch zu*; mit einer kausalen, konzessiven, temporalen, lokalen, modalen Ergänzung, z. B. *In Zivilprozessen diktiert der Vorsitzende Richter das Protokoll, Aufgrund der Verspätung der Zeugin kam es zur Vorverlegung der Pause, durch eine fristgerechte Zahlung vermeiden Sie ein Mahnschreiben*; mit einem Präpositionalobjekt, z. B. *Über die Verhandlung fertigt das Gericht ein Protokoll* (vgl. Thormann 2016: 46–48). Bei den Tempora kann in den juristischen Texten der häufigere Gebrauch vom Präteritum anstelle vom Perfekt beobachtet werden. Dies ist darauf zurückzuführen, dass sich dadurch Satzklammern vermeiden lassen, wodurch der Text einfacher zu verstehen ist. Plusquamperfekt wird auch verwendet, denn es kann in gewissen Situationen rechtlich gesehen von Bedeutung sein, die Vorzeitigkeit präzise auszudrücken. In der Gemeinsprache wird gewöhnlicherweise Perfekt verwendet, wenn die zeitliche Abfolge der Geschehnisse nicht ausdrücklich betont werden muss bzw., wenn es sich nicht um einen Temporalsatz in der Vergangenheit mit der Subjunktion *nachdem* handelt (vgl. Thormann 2016: 54–55). Charakteristisch für die Rechtssprache ist auch das sog. „Vertragspräsens“ bzw. „Gebotspräsens“. Sätze, die in der Gemeinsprache in dieser Form formuliert sind, werden neutral konnotiert und es sind einfache Aussagesätze. In der Rechtssprache bekommen sie die Bedeutung, die normalerweise mithilfe des Verbs *müssen* erlangt ist, z. B. *Der Käufer muss den vollen Preis an den Verkäufer zahlen* → *Der Käufer zahlt den vollen Preis an den Verkäufer, Funktionsmängel sind vom Kunden unverzüglich zu dokumentieren* → *Funktionsmängel werden vom Kunden unverzüglich dokumentiert*. In der österreichischen und polnischen Rechtssprache werden viele Abkürzungen verwendet u.a.: *VSG* (Verkehrssicherstellungsgesetz), *VStRG* (Vermögenssteuerreformgesetz), *VuR* (Verbraucher u. Recht), *k.h.* bzw. *kh* (kodeks handlowy), *DzU* (Dziennik Ustaw), *k.p.a.* (kodeks postępowania administracyjnego) (vgl. Böttcher 2021: 392; Glapiak & Stalmach 2023: 80–81). Das letzte Charakteristikum, das in diesem Unterkapitel erwähnt wird, betrifft Doppel- und Vielfachverneinung. In der Rechtssprache kommen Sätze vor, die für die Gemeinsprache

ungewöhnlich formuliert sind wie z. B. *die Beschränkungen sind nicht unerheblich, der Inhalt ist nicht unwesentlich, die Verfassungsbeschwerden sind nicht unzulässig* (vgl. Thormann 2016: 56–57).

6. Urheberrecht

Mit jeder Neugestaltung gewinnt das Urheberrecht größere Bedeutung, weil es sich quasi selbst bedingt und ständig selbst erneuert. Totgesagte leben in der Tat länger! (Büchele 2023: 2)

Die vorliegende Arbeit setzt sich mit der automatischen Extraktion der österreichischen und polnischen Rechtsterminologie auseinander. Um die bestmögliche Vergleichbarkeit der Ergebnisse gewährleisten zu können, wurde das umfassende Fach Recht auf das Hauptsachgebiet Urheberrecht eingeschränkt, indem die zu untersuchenden Rechtstexte ausschließlich aus diesem Rechtsbereich ausgewählt wurden. In dem Kapitel soll daher in erster Linie auf die Grundprinzipien des österreichischen und polnischen Urheberrechts eingegangen werden, um die Charakteristika der urheberrechtlichen Terminologie besser nachvollziehen zu können. Weiters wird im Besonderen das für das Thema der Arbeit relevante Verhältnis zwischen dem Urheberrecht und der künstlichen Intelligenz dargestellt, denn das Urheberrecht, wie mit dem Zitat veranschaulicht, muss sich der Fortentwicklung anpassen und den neuesten technologischen Stand reflektieren (vgl. Büchele 2023: 2).

6.1. Grundprinzipien des Urheberrechts

Das Urheberrecht als Teil des Immaterialgüterrechts dient dem Schutz der Schöpfer:innen immaterieller Güter, damit sie am finanziellen Erfolg ihrer Arbeit teilhaben können (vgl. Büchele 2023: 1). Die im Sinne des Urheberrechts geschützten Leistungen oder Werke tragen allerdings einen kulturellen Wert, weswegen sie nicht ausschließlich den Schöpfer:innen vorbehalten sind bzw. sie unterliegen Ausnahmen und zeitlichen Beschränkungen (vgl. Büchele 2023: 2). Dieser Ausgleich zwischen dem Allgemeingut einerseits und den Schöpfer:inneninteressen andererseits spiegelt sich in rechtspolitischen Debatten und gesetzgeberischen Aktivitäten wider und kann zu Rechtunsicherheit führen (vgl. 2023: 2). Das österreichische Urheberrechtsgesetz definiert die urheberrechtlich geschützten Werke folgendermaßen:

Werke im Sinne dieses Gesetzes sind eigentümliche geistige Schöpfungen auf den Gebieten der Literatur, der Tonkunst, der bildenden Künste und der Filmkunst. (§ 1 Abs 1 UrhG)

Nicht jedes Werk wird jedoch automatisch urheberrechtlich geschützt, denn beispielsweise unterliegen technische Erfindungen dem Patentrecht. Ein Werk, der o.g. Definition nach, muss eine geistige Schöpfung sein, die eigentümlich also originell oder individuell ist und sich der

im Gesetz bestimmten Werkart zuordnen lässt (vgl. Wiebe et al. 2022: 203). Unter geistiger Schöpfung versteht die Judikative „eine sinnlich wahrnehmbare und bestimmte Formung eines geistigen Stoffes“ (vgl. 2022: 203). Dies bedeutet, dass ein Werk nicht unbedingt physisch vorliegen muss, es muss lediglich als solches wahrnehmbar sein, wie etwa ein nicht aufgezeichneter Vertrag, der als ein Literaturwerk urheberrechtlich geschützt werden kann (vgl. 2022: 203). Weiters, abgesehen davon, ob es sich letztendlich um ein körperlich festgelegtes Werk oder nicht handelt, steht nicht dieses Werk selbst unter Schutz, sondern die geistige Gestaltung, die dahintersteht, also ganz im Sinne des Schutzes des Immaterialgüterrechts (vgl. Hofmarcher 2023: 133). Gleichzeitig gilt eine Idee, die noch nicht ausgeformt ist bzw. keine konkrete Gestaltung aufgenommen hat, nicht als schutzfähig. Die Individualität eines Werkes wird durch seine sich vom Alltäglichen bzw. standardmäßig Hervorgebrachten abhebenden Eigenschaften bestimmt (vgl. Ciresa 2022: 66). Damit ein Werk die nötige Originalität im Sinne des Gesetzes aufweist, darf die kreative, künstlerische Freiheit des Schöpfenden nicht eingeschränkt werden, anders ausgedrückt darf sich die Schöpfung eines Werkes, anhand der Vorgaben des EuGH, nicht alleine aus Regeln und technischen Bestimmungen ergeben (vgl. Ciresa 2022: 67). Zu der Werkkategorie der Literatur gehören in erster Linie Sprachwerke und Computerprogramme. Unter Sprachwerken werden literarische Schöpfungen wie Libretti, Gedichte und Romane sowie Zweckschöpfungen wie Werbespots und komplexe Verträge verstanden. Computerprogramme zählen zu Werken der Literatur, da ihnen ein in einer Programmiersprache verfasster Quellcode in Form vom Text zugrunde liegt (vgl. Wiebe et al. 2022: 209–210). Weiters sind auch Bühnenwerke der Kategorie zuzuordnen, also pantomimische und choreographische Werke, die dafür geeignet sind, sie bühnenmäßig darzustellen wie z. B. Puppenspiel, Eisrevue und Ausdruckstanz. Des Weiteren gehören dazu ebenfalls belehrende und wissenschaftliche Werke wie Lernspiele, Himmelskarten und Reliefdarstellungen (vgl. Wiebe et al. 2022: 211–212). Werke der Tonkunst stellen die geistige Schöpfung mithilfe von Tönen dar, die entweder mit Instrumenten oder mit menschlicher Stimme produziert werden. Dazu zählen Schlagermusik, Handy-Klingeltöne, Werbe-Jingles u.v.m. Eine weitere Werkart bilden bildende Künste wie z. B. allgemeine bildende Künste: Skulpturen, Plastiken, Gemälde; Lichtbildkunst: Lichtbildwerke, Lichtdesign; Baukunst: Hundertwasserhaus, Pläne für ein Landhaus und angewandte Kunst: Designer-Sitzmöbel, Banknoten, Kitzbühel-Schriftzug u.a. Die letzte im Gesetz genannte Werkkategorie ist Filmkunst, zu der beispielsweise Schnitt und Fernseh-Live-Übertragungen zählen (vgl. Wandtke & Ostendorff 2023: 68–80; Wiebe et al. 2022: 213–221).

Ein besonderes Augenmerk gilt den urheberrechtlich geschützten Bearbeitungen der Originalwerke, solange sie eigentümliche geistige Schöpfungen sind. Bearbeitungen verändern

die äußeren Merkmale bestehender Werke auf originelle Weise, bewahren jedoch den Kern des Originals und lassen dessen Wesen unverändert (vgl. Büchele 2023: 29–30 nach OGH 23. 2. 2016, 4 Ob 142/15 h, Bettis Hand, ÖBI 2016/55, 240 (Stadler)=MR 2016, 140 (Walter)=ecolex 2016/273, 603 (Horak)). Bearbeiter:innen haben keine Ansprüche auf Miturheberschaft am Originalwerk, aber sie erlangen Bearbeiterurheberrecht, was bedeutet, dass sowohl das Originalwerk als auch dessen Bearbeitung unter Schutz stehen (§ 5 Abs 1 UrhG). Dazu zählen in erster Linie Übersetzungen, die eine individuelle, „höchst anspruchsvolle, aufwendige und somit höchst schöpferische“ Arbeit des Übersetzenden darstellen (vgl. Wiebe et al. 2022: 222 nach OGH 4 Ob 293/01v-Riven Rock, ÖBI 2002, 250 (Wolner)=MR 2002, 164 (Walter)). Schutzfähig sind allerdings lediglich humane Übersetzungen, ungeachtet dessen, ob sie computergestützt gefertigt wurden oder nicht. Nicht unter Schutz stehen, im Sinne des Gesetzes, maschinelle Übersetzungen, da ihnen der nötige schöpferische, menschlich-geistige Einfluss fehlt (vgl. 2022: 222). Dahingehend dürfen humane Übersetzungen ausschließlich mit Zustimmung des Schöpfenden verwertet werden. Weiters ist die Zustimmung des Verfassers/der Verfasserin des Originalwerkes einzuholen, da eine Verwertung der Bearbeitung mit der Verwertung des Originals einhergeht, denn es handelt sich dabei um ein abhängiges Urheberrecht (vgl. Büchele 2023: 31). Zu weiteren Beispielen der schutzfähigen Bearbeitungen gehören Verfilmungen, Synchronisierungen, digitale oder analoge Bildbearbeitung, Modernisierung von Firmenlogos u.a. (Wiebe et al. 2022: 221).

6.1.1. Das österreichische Urheberrechtsgesetz

Büchele (2023: 3) sieht die Ursprünge des Urheberrechts in Österreich bereits im Jahre 1811 mit dem Inkrafttreten des Allgemeinen Bürgerlichen Gesetzbuches (StF: JGS Nr. 946/1811), aber das Urheberrecht war damals sehr begrenzt. Weitere Meilensteine in der Entwicklung des österreichischen Urheberrechts waren das Kaiserliche Allerhöchste Patent zum Schutze des literarischen und artistischen Eigenthumes gegen unbefugte Veröffentlichung, Nachdruck und Nachbildung aus dem Jahre 1846 und das darauffolgende Gesetz aus dem Jahre 1895 (vgl. 2023: 3). Die aktuelle Rechtsquelle für das Urheberrecht in Österreich ist das seit 1936 geltende Urheberrechtsgesetz (StF: BGBl. Nr. 111/1936. StR: 39/Gu. BT: 64/Ge S. 19.). Hofmarcher (2023: 127) führt allerdings an, dass das Gesetz in der Zwischenzeit öfter novelliert wurde. Hinter der ständigen Anpassung des Urheberrechtsgesetzes steht in erster Linie die Entwicklung der Verbreitungstechnik, die die rasante Vervielfältigung der Werke ermöglicht hat, vom Buchdruck über Rundfunk, die Datenbanken bis hin zum Internet. Der weitere Grund für die ständigen Änderungen im Urheberrecht betrifft die Harmonisierung des nationalen Rechts mit

den internationalen Bestimmungen. Dies bezieht sich auf solche internationalen Verträge wie z. B. die Berner Übereinkunft oder das Welturheberrechtsabkommen, die dem weltweiten Schutz dienen (vgl. Hofmarcher 2023: 128; Zia-UI-Haq 2023: 153). Die Harmonisierung des Urheberrechts wird ebenfalls auf der EU-Ebene angestrebt, allerdings, so die Autor:innen, hat die Vollharmonisierung trotz mehrerer entsprechender Richtlinien noch zu erfolgen (vgl. 2023: 128; 2023: 153). Das Ziel ist die Rechtsunterschiede im gemeinsamen Binnenmarkt zu eliminieren, die den grenzüberschreitenden Handel mit Gütern und Dienstleistungen einschränken könnten (vgl. Wandtke & Ostendorff 2023: 45).

Das österreichische Urheberrechtsgesetz definiert Urheber:in als Schöpfer:in des Werks. Es handelt sich ausschließlich um eine (oder mehrere) natürliche Person, die die schöpferische Leistung erbracht hat. Es wird zwischen Alleinurheber:in und Miturheber:innen differenziert, wenn das Werk nur von einer Person, respektive von einer weiteren Person oder von einer Gruppe von Personen geschaffen wurde (vgl. Ciresa 2022: 19). Teilurheberschaft entsteht dann, wenn es zu einer Verbindung von Werken kommt, z. B. wenn die einen Film begleitende Filmmusik auf einer CD erscheint. Da die Werke getrennt voneinander verwertet werden können, kann nicht von der Miturheberschaft ausgegangen werden (vgl. Hofmarcher 2023: 138). Als Gehilfe werden jene Personen benannt, ohne die zwar in vielen Fällen das Werk gar nicht entstanden wäre, welche jedoch keinen geistigen, eigentümlichen Beitrag zur Werkschöpfung beigetragen haben und somit kein Urheberrecht erwerben. Dazu zählen u.a. Assistent:innen, die Leinwände aufspannen, Manuskripte abtippen, Texte lektorieren, die Autor:innen an die Fristen erinnern usw. (vgl. 2023: 138). Das Schöpferprinzip gilt nach dem österreichischen Urhebergesetz nicht für juristische Personen wie z. B. Universitäten oder Vereine. Sie können allerdings Nutzungsrechte von den Urheber:innen rechtsgeschäftlich erwerben (vgl. 2023: 138). Weiters bestehen keine Einschränkungen auf den urheberrechtlichen Schutz eines Werks bezüglich des Zustands in dem sich Schöpfende bei der Werkschöpfung befanden. Dies betrifft beispielsweise die alkohol- bzw. drogenbedingte Berausung oder die psychische Gesundheit von Schöpfenden.

Dies gilt z. B. auch für die berühmten Bilder der Patienten der Heilanstalt Gugging bei Wien, die im Rahmen ihrer Therapie Aquarelle malen, bei denen es sich trotz der mentalen Beeinträchtigung ihrer Schöpfer um eigentümliche geistige Leistungen im Sinn des UrhG handeln kann. (Ciresa 2022: 21)

Des Weiteren gelten auch Kinder als Urheber:innen ihrer z. B. im Rahmen der Schulprojekte geschaffenen Werke. Das Urheberrecht von einem Werk können auch Personen haben, die es

zwar nicht geschaffen haben, aber auf welche das Recht nach dem Tod des Schaffenden übergegangen ist. Dies wird als derivative Urheberschaft bezeichnet (vgl. Ring et al. 2021: 155; Wiebe et al. 2022: 225). Einen interessanten Fall bilden Werke, die von Tieren geschaffen werden. Dazu zählen u.a. Gemälde und Selfies, die beispielsweise von Gorillas, Schopfmakaken und Schweinen oft auf Anregung des Pflegepersonals geschaffen werden (siehe Farm Sanctuary SA 2024; The Gorilla Foundation / Koko.org 2024). Da Tiere allerdings keine Rechtsfähigkeit besitzen und es sich bei der Werkschöpfung um die Realisierung der Natur handelt, werden die Werke nicht urheberrechtlich geschützt. Auch dem Pflegepersonal oder anderen Personen, die Tiere zur künstlerischen Tätigkeit motiviert haben, steht kein Urheberrecht zu (vgl. Wiebe et al. 2022: 226). Die Urheberschaft bzw. der urheberrechtliche Schutz bei der Werkschöpfung von der künstlichen Intelligenz wird im separaten Unterkapitel beschrieben.

6.1.2. Das polnische Urheberrechtsgesetz

Das polnische Urheberrecht schützt ebenfalls die Schöpfer:innen und ihr geistiges Eigentum (vgl. Barta & Markiewicz 2010: 19–21). Ähnlich wie in Österreich werden die immateriellen Werke in Polen nicht unbefristet geschützt, denn sie gehören zum geschichtlichen Erbe des Landes. Die Gesetzesnovelle aus dem Jahr 2000 verlängerte die Schutzdauer von den ursprünglichen 50 Jahren auf 70 Jahre. Die Schutzdauer wird hauptsächlich ab dem Todestag des:der Urheber:in des Werks gerechnet. Sollte es allerdings nicht bekannt sein, wer das Werk geschaffen hat, wird die Schutzdauer ab dem Datum gerechnet, an dem das Werk zum ersten Mal öffentlich verbreitet wurde (vgl. Kuś & Senda 2004: 29). Im Sinne des Gesetzes wird zwischen der Verbreitung und der Veröffentlichung eines Werks differenziert. Mit der Veröffentlichung geht nämlich die Zustimmung des Schaffenden einher, das Werk zu vervielfältigen und die Kopien der Öffentlichkeit bereitzustellen. Im Falle der Verbreitung wird das Werk verfügbar gemacht, ohne dass es auf einem materiellen Träger fixiert wird. Als Beispiel gilt die Uraufführung eines Lieds während eines Konzerts, das noch nicht auf einer CD zu kaufen ist, was wiederum ein Beispiel für die Veröffentlichung wäre (vgl. Kuś & Senda 2004: 29–30). Das Werk wird im polnischen Urheberrechtsgesetz folgendermaßen definiert:

Utworem jest każdy przejaw działalności twórczej o indywidualnym charakterze, ustalony w jakiejkolwiek postaci, niezależnie od wartości, przeznaczenia i sposobu wyrażenia¹. (Art. 1. [Utwór] Dz.U.2022.2509 t.j.)

¹ Ein Werk ist jede, in welcherlei Form bestimmte, Äußerung einer individuellen schöpferischen Tätigkeit unabhängig von Wert, Zweck und Ausdrucksweise. („Übers. M.S.“)

Gemäß der Entscheidungen des polnischen Obersten Gerichtshofs sind unter Werk außerdem „każdy przejaw działalności duchowej noszący cechę osobistej działalności, każda oryginalna myśl ludzka, ujęta konkretnie w samodzielny formę²“ (Kuś & Senda 2004: 10 nach Orzeczenie SN z dnia 23 czerwca 1936 r. – IK 336/36, nie publikowane) bzw. „każde opracowanie (...), w którym przejawia się twórcza praca autora, a więc, gdy dzięki pomysłowości i indywidualnemu ujęciu dzieło uzyskało oryginalną postać³“ zu verstehen (Kuś & Senda 2004: 11 nach Orzeczenie SN z dnia 27 marca 1965 r. – ICR 39/65, nie publikowane). Lt. des polnischen (und österreichischen) Urheberrechts ist der Schutz nicht an die Erfüllung bestimmter formaler Bedingungen geknüpft wie z. B. die Anbringung des Copyright-Hinweises bzw. ©-Zeichens, Registrierung des Werks oder Gebührenzahlung, mit anderen Worten erfolgt die Entformalisierung der Gewährung des Urheberschutzes (vgl. Ciresa 2022: 21; Kuś & Senda 2004: 12 nach Art. 1. ust 4. [Utwór] Dz.U.2022.2509 t.j.).

Für Firlit-Fesnak et al. (2012: 10) hat das Urheberrecht in Polen seinen Ursprung im Gesetz von Fryderyk Zoll. Laut dem Gesetz sollen hauptsächlich die Schöpfer:innen geschützt werden, aber es müssen, wie bereits erwähnt, auch bestimmte Kompromisse eingegangen werden, die sich in zeitlicher und inhaltlicher Beschränkung des Urheberrechts zum Wohle des öffentlichen Interesses manifestieren (vgl. Firlit-Fesnak et al. 2012: 10; Zoll 1926: 30). Der Gesellschaft soll nämlich ein schneller, weitreichender und, wenn möglich, kostenloser Zugang zu kulturellen und wissenschaftlichen Errungenschaften garantiert werden. Der Kompromiss zwischen den Interessen des Schaffenden und der Gesellschaft ist am besten anhand solcher öffentlichen Einrichtungen wie Schulen, Bibliotheken und Instituten zu beobachten, die gemäß des Urheberrechtsgesetzes unter bestimmten Voraussetzungen in das Urheberrechtsmonopol des Schaffenden eingreifen dürfen (vgl. Kuś & Senda 2004: 41). Die aktuelle Rechtsquelle für das Urheberrecht in Polen ist das seit 1994 geltende Gesetz (Dz.U. 1994 nr 24 poz. 83), das allerdings insbesondere nach dem EU-Beitritt Polens 2004 mehrmals novelliert wurde, um die bestmögliche Harmonisierung des nationalen und des EU-Urheberrechts zu gewährleisten. Dies ist darauf zurückzuführen, dass das EU-Recht Vorrang vor dem nationalen Recht der Mitgliedsstaaten hat und es dürfen keine Rechtsvorschriften erlassen werden, die gegen das EU-Recht verstoßen (vgl. Bleszyński 2015: 5–6; Fiołka 2015: 266; Kuś & Senda 2004: 71).

² Jede Form geistiger, persönlicher Aktivität, jeder originelle menschliche Gedanke, der in einer eigenständigen Form konkret bestimmt ist. („Übers. M.S.“)

³ Jede Bearbeitung (...), in der die schöpferische Leistung des:der Urhebers:in zum Ausdruck kommt, d.h. in der der originelle Charakter des Werks durch Einfallsreichtum und individuelle Auffassung des Schöpfenden geschaffen wurde. („Übers. M.S.“)

6.2. Urheberrecht und künstliche Intelligenz

Das Urheberrecht wird aufgrund der Entwicklung im Bereich der künstlichen Intelligenz (KI) mit neuen Herausforderungen konfrontiert und steht unter einem ständigen Anpassungsdruck (vgl. Wandtke & Ostendorff 2023: 1; Yenduri et al. 2024: 24). Die der KI zugrunde liegenden Methoden kreieren Kunstwerke, verfassen Texte und komponieren Lieder – ein Prozess, der Automated Creativity genannt wird (vgl. Ciresa 2022: 16). Die nationale und internationale Gesetzgebung wird vor die Frage gestellt, ob die von durch KI-Methoden erzeugten Werke urheberrechtlich geschützt werden. In der Fachliteratur und Judikative herrscht zurzeit die Überzeugung, dass es nicht der Fall ist, da den computergenerierten Werken der:die menschliche Schöpfer:in fehlt und KI-Methoden keine Rechtspersönlichkeit haben (vgl. Baker 2024: 108; Ciresa 2022: 16; Libor 2022; Mazzi 2024: 5). Auch LLMs selbst, in dem Fall ChatGPT-4o, generiert, wenn konfrontiert mit der Frage: „Bist du der Urheber / die Urheberin der von dir erzeugten Inhalte?“ eine ähnliche Antwort:

Als KI-Modell bin ich nicht der Urheber der von mir erzeugten Inhalte. Ich generiere Antworten auf Grundlage des Wissens, mit dem ich trainiert wurde, und der spezifischen Eingaben, die ich erhalte. Meine Inhalte basieren auf vorhandenen Informationen, Daten und allgemeinen sprachlichen Mustern, die ich verarbeite. Da ich keine kreativen Rechte wie ein Mensch besitze, fallen die von mir erzeugten Inhalte nicht unter ein persönliches Urheberrecht. (OPENAI CHATGPT 2024a)

Sollte eine natürliche Person mithilfe der KI-Werke schaffen, werden nur jene Teile urheberrechtlich geschützt, die der Person zuzuschreiben sind, mit anderen Worten steht der KI-Beitrag des Werks nicht unter Schutz. Zusammenfassend genießen, im Sinne des Gesetzes, weder ein LLM selbst, die Nutzer:innen, die die Prompts erstellt haben noch die KI-Entwickler:innen das Urheberrecht an den von der KI generierten Werken (vgl. Baker 2024: 108). Lucchi (2023: 7) argumentiert allerdings, dass der computergenerierte Text in einigen Fällen als genügend originell betrachtet werden kann, um unter dem urheberrechtlichen Schutz stehen zu können, und zwar dann, wenn er mit ausreichenden menschlichen Eingabedaten erstellt wurde. Dies trifft z. B. dann zu, wenn die LLM-Antworten auf Prompts als Grundlage genutzt werden und anschließend wesentliche kreative oder originelle Elemente hinzugefügt werden etwa durch Bearbeitung der Antworten, Ergänzung durch Kommentare oder Analysen, oder wenn Antworten in ein größeres Werk eingebaut werden. Der Person, die den zusätzlichen kreativen oder originellen Beitrag geleistet hat, würde in der Regel das Urheberrecht am Endprodukt zugeschrieben werden (vgl. 2023: 7).

Präzise diese Unsicherheiten lagen im Februar 2017 einem Vorschlag des EU-Parlaments zugrunde, neben der natürlichen und juristischen Person ein weiteres Rechtskonstrukt, die sog. e-Person zu kreieren bzw. das Konzept in Erwägung zu ziehen und somit den KI-Methoden die Rechtspersönlichkeit geben, damit sie bei etwaiger Rechtsverletzung haften können (vgl. Prill 2024: 56). Dies bezieht sich in erster Linie auf Bereiche, die einen gewissen Grad an Autonomie aufweisen, wie z. B. autonomes Fahren und intelligente Assistenz, aber kann ebenso gut für die Verstöße gegen das Urheberrecht gelten (vgl. Pardey 2021; Staudacher & Nida-Rümelin 2020: 5; Zeck 2021: 93). Dieser Vorschlag ist genauer gesagt aufgrund einer möglichen Verantwortungslücke im Recht entstanden, denn je mehr entwickelt die KI-Methode ist, umso häufiger wird die Gesetzgebung vor ein Dilemma gestellt, wer für die von der KI-Methode verursachten Schäden zur Rechenschaft gezogen werden soll, ob es Forscher:innen, Konstrukteur:innen, Programmierer:innen, Trainer:innen, Betreiber:innen oder die KI-Methode selbst in Form von digitaler e-Person sein soll (vgl. Staudacher & Nida-Rümelin 2020: 5; Zeck 2021: 97). Weiters, so die meisten Befürworter:innen, soll es sich dabei vielmehr um die Teilrechtsfähigkeit handeln, also lediglich um jene Bereiche, die dem Einsatzzweck der KI-Methode entsprechen und von ihr durchführbar sind (vgl. Prill 2024: 57). Die potenzielle Einführung der neuen Rechtspersönlichkeit wirft in der Fachliteratur mehrere Fragen auf. Zum einen wird betont, dass KI-Methoden per se über kein Vermögen verfügt, weswegen sich die Frage nach einer eventuellen Schadenersatzforderung erübrigt, es sei denn, KI wird mit entsprechendem Kapital ausgestattet oder es werden Haftungsfonds oder Versicherungen abgeschlossen (vgl. Zeck 2021: 97). Außerdem würde sie einen großen Eingriff in die nationale und internationale Rechtsordnung darstellen sowie auf mehrere Hindernisse bei der Umsetzung stoßen. Weiters stellt sich die Frage, ab genau welchem Autonomiegrad eine KI-Methode als e-Person zu betrachten ist und was überhaupt als Rechtssubjekt zu identifizieren ist, ob es der Algorithmus ist oder das Produkt wie etwa ein Roboter oder Fahrzeug. Schließlich ist auch zu betonen, dass innerhalb eines KI-Systems bzw. -Produkts mehrere autonome Subsysteme enthalten sein können und es ist nicht klar, exakt welches Subsystem ggf. die Haftung übernehmen müsste (vgl. Zeck 2021: 98–99). Aus diesen und weiteren Gründen stehen die Expert:innen aktuell der Einführung der e-Person kritisch gegenüber. Dieselbe Auffassung vertritt ebenfalls das EU-Parlament, das diesen Vorschlag im Oktober 2020 als nicht erforderlich gewertet hat (vgl. Kreutz 2020: 195–199; Prill 2024: 59–61; Staudacher & Nida-Rümelin 2020: 20; Zeck 2021: 97).

Eine weitere Frage in Bezug auf die KI-Methoden betrifft die Nutzung urheberrechtlich geschützter Inhalte im Rahmen des Text- und Data-Minings, um die generative LLMs zu

trainieren. Gemäß der geltenden gesetzlichen Bestimmungen in Österreich ist dies erlaubt (§ 42h UrhG), solange die Schöpfer:innen keine Einwände erheben (vgl. Homar 2024). Wurde allerdings das Training abgeschlossen, ist der Datensatz mit den enthaltenen Trainingsdaten zu entfernen (vgl. Baker 2024: 108). Nichtsdestotrotz werden aktuell Initiativen ergriffen und Klagen vor Gerichten weltweit eingereicht, um dies zu verbieten oder eine „faire Vergütung“ für die Nutzung der Werke zum Training zu sichern sowie eine Kennzeichnungspflicht der KI-generierten Werke einzuführen (vgl. Baker 2024: 108; Grynbaum & Mac 2024; Jasińska 2023: 422). Im April 2021 hat die Europäische Kommission weltweit weichenstellend eine Rahmenordnung für den Umgang mit KI vorgeschlagen, die dann schließlich im Juli 2024 als die EU-AI-Akte im Amtsblatt der EU veröffentlicht wurde. Die Akte enthält mehrere Punkte, die z. B. biometrische Fernidentifizierung oder Aufbau von Gesichtserkennungsdatenbanken betreffen, aber auch die Festlegung einer Politik zum Thema der „Einhaltung der Urheberrechtslinie“ (vgl. Europäisches Parlament 2024b). Dies betrifft insbesondere die Nutzung der Inhalte für das Training von Modellen:

General-purpose AI (GPAI) systems, and the GPAI models they are based on, must meet certain transparency requirements, including compliance with EU copyright law and publishing detailed summaries of the content used for training. The more powerful GPAI models that could pose systemic risks will face additional requirements, including performing model evaluations, assessing and mitigating systemic risks, and reporting on incidents. Additionally, artificial or manipulated images, audio or video content (“deepfakes”) need to be clearly labelled as such. (Europäisches Parlament 2024a)

Wie bereits erwähnt, sind erhebliche Datenmengen aus verschiedenen Quellen erforderlich, um die Leistung und Funktionalität eines Modells zu erweitern. Es ist nicht ausgeschlossen, dass sich dabei Texte befinden, die persönliche Daten enthalten. Die Verarbeitung personenbezogener Daten innerhalb der EU unterliegt allerdings strengen Regulierungen und Beschränkungen und wird von der Datenschutz-Grundverordnung (DSGVO) geregelt. Somit stellt dies nicht nur eine urheberrechtliche, sondern ggf. auch eine datenschutzrechtliche Herausforderung dar (vgl. Lucchi 2023: 15).

Wenn präzisiert, wird in der Definition des urheberrechtlich geschützten Werks der Aspekt der Originalität und Kreativität des Schaffenden hervorgehoben. In Bezug auf die KI-Methoden wird daher in der Fachliteratur darüber diskutiert, ob die Kreativität eine rein menschliche Eigenschaft ist (vgl. Maamar 2021: 59) und sich bei den artifiziellen Erzeugnissen lediglich um eine künstliche Rekonstruktion handelt, was ein weiteres Argument darstellen würde, dass computergenerierte Werke der Werksdefinition gemäß nicht urheberrechtlich geschützt werden können, oder ob die KI-Methoden mittlerweile so weit entwickelt sind, dass

man generierte Bilder, Texte, Lieder u.a. als kreativ im Sinne des Gesetzes definieren könnte (vgl. Grätz 2021: 111). Sollte jedenfalls nur das Endprodukt als das einzige Hauptmerkmal der Debatte über Kreativität gelten und der menschliche Aspekt ausgelassen werden, so Grätz (2021: 112), würde das KI-Werk unter dem Schutz stehen, da dessen Maßstab im juristischen Verständnis nicht an qualitative Kriterien des Werks anknüpft (siehe auch Maamar 2021: 59). Weiters argumentiert der Autor (Grätz 2021: 111), dass einige der vom LLM generierten Bilder mittlerweile Kunstbegeisterte finden, die bereit sind, für sie hohe Geldsummen zu bezahlen. Dazu zählt *das Porträt von Edmond de Belamy*, das für 432,500 \$ versteigert wurde (vgl. Guadamuz 2021: 148). Ein weiteres Beispiel, das für die Kreativität der KI-Methoden spricht, ist das Aufsehen, für welches der KI-Künstler Ben Gaya 2024 mit seinem ebenfalls von KI-Methoden generierten Lied „Sunshine Soul“ gesorgt hat (vgl. Slancar 2024). Die Autorin (2024) vertritt allerdings die Ansicht, dass trotz dieses Erfolgs die Lieder, die *mit* KI im Gegensatz zu *von* KI generiert werden, bessere Resultate bringen und sie bezieht sich dabei auf das kontroversiell diskutierte Lied „Verknallt in einen Talahon“, das 2024 mithilfe von KI (Musik, Gesang) gestaltet wurde und in die deutschen Singlecharts gelangt ist. Dem Song liegt ein von einem Menschen, nämlich von Josua Waghübinger, verfasster Text vor (vgl. 2024). Zusammenfassend lässt sich feststellen, dass die KI-Methoden, gemäß dem heutigen wissenschaftlichen, juristischen und technischen Stand, nicht das volle Potenzial an kreativer Gestaltungsfreiheit und Inspiration ausschöpfen, die einem Menschen zuzumuten ist. Vielmehr erstellen sie Werke, die den Stil der in den Trainingsdaten enthaltenen Vorlagen nachahmen. Ein LLM erkennt Muster in einer Vielzahl von Daten und nutzt diese als Grundlage, um ähnliche Ergebnisse zu erzeugen. Etwas völlig Neues, „unabhängig Andersartiges“ wird ein LLM jedoch nur in begrenztem Umfang hervorbringen können (vgl. Grätz 2021: 113). Deswegen sind KI-Methoden nicht als kreativ zu betrachten, obwohl es nicht ausgeschlossen ist, dass dies sowie das Urheberrecht selbst, mindestens in der aktuell bekannten Form, sich in Zukunft ändert (vgl. Grätz 2021: 114; Kreutzer 2023: 502).

7. Aktueller Stand der Forschung

Die automatische Terminologiegewinnung ist ein durchaus aktuelles Thema, weswegen immer mehr Forschung in dem Bereich betrieben werden. Wie bereits im Kapitel 2.3.2.2. und Kapitel 3 erwähnt, gibt es mehrere Systeme, die auf unterschiedlichen Ansätzen basieren und entweder speziell für diesen Zweck entwickelt wurden oder diese konkrete Funktion zusätzlich zu ihren anderen Funktionen anbieten. Die bereits vorher erwähnten Tools werden nachfolgend nicht mehr wiederholt und können in den o.g. Kapiteln nachgelesen werden. Da die meisten Datensätze in englischer Sprache vorliegen, wurden viele Tools vorwiegend für die Termextraktion englischsprachiger Texte entwickelt (vgl. Rokas et al. 2020: 40). Allerdings herrscht in der Fachliteratur Konsens darüber, dass keines der Modelle bisher durchgängig die besten Ergebnisse für englischsprachige Texte (sowie für andere Sprachen) aus unterschiedlichen Fachbereichen liefert (vgl. Astrakhantsev 2018: 869; Šajatović et al. 2019: 150; Zhang et al. 2018: 30). Mittlerweile liegen Forschungsarbeiten vor und es sind auch Tools verfügbar, die gezielt die Termextraktion in anderen Sprachen im Fokus haben, wie z. B. das auf C-Value basierte Tool TermoPL für Polnisch (Marciniak et al. 2016), das am hybriden Ansatz basierte DEXTER für Spanisch und Englisch (Periñán-Pascual & Mestre-Mestre 2015), BERT-LLA-CRF für Chinesisch (Zhong et al. 2023) u.v.m. Auch für sehr unterrepräsentierte und/oder gefährdete Sprachen wurden entweder Tools zur Termextraktion entwickelt bzw. bestehende Tools angepasst wie z. B. Saffron für Irisch (Arcan et al. 2022; Bordea et al. 2014; McCrae & Doyle 2019) oder das auf XML basierte Tool Erauzterm für Baskisch (Alegria et al. 2004) oder Forschungen im Bereich der Termextraktion betrieben wie z. B. für Kaschubisch, eine in der nördlich-zentralen Region Polens gesprochene Sprache, oder Walisisch (Łukasik 2023; Spasic et al. 2019). Außerdem sind immer mehr sprachunabhängige Tools verfügbar, beispielsweise TermoUD (Marciniak et al. 2023) oder ATR4S (Astrakhantsev 2018). Aktuell wird auch die häufigere Verwendung von LLM-Modellen zur Terminologiegewinnung beobachtet, darunter außer ChatGPT auch Llama2-7B und Falcon-7B (Chataut et al. 2024).

Die vorliegende Arbeit untersucht unter anderem wie ein zur Termextraktion entwickeltes Tool im Vergleich zu einem Modell, das Termextraktion lediglich als Nebenfunktion anbietet, abschneidet (siehe 3.1.4.1.-3.1.4.2.). Daher wird in dem folgenden Kapitel gezielt auf die Forschung eingegangen, die dem Forschungsziel der vorliegenden Arbeit am nächsten steht.

Bei der Suche nach den entsprechenden Forschungsarbeiten wurden die folgenden Punkte berücksichtigt: Extraktion von Ein- und Mehrwortbenennungen mit ChatGPT; Fachbereich Recht; andere Sprachen als Englisch und Qualitätsmetriken: Genauigkeit, Trefferquote und F1-Score. Trotz intensiver Recherche wurden keine Arbeiten gefunden, die alle diese Punkte abdecken, aber die nachfolgend dargestellten wissenschaftlichen Artikel beinhalten nichtsdestotrotz etliche Aspekte, die für die vorliegende Arbeit relevant sind und einen Überblick über das Untersuchungsziel bieten.

Premasiri et al. (2023) analysieren die automatische Termextraktion von Blumen- und Pflanzennamen für Spanisch und Englisch. Im Fokus der Untersuchung stehen Mehrwortbenennungen und es werden Ergebnisse der Termextraktion des neuronalen generativen Sprachmodells GPT 3.5-turbo mit den Ergebnissen von 13 diskriminativen Transformer-Modellen verglichen. Extrahiert wird die Terminologie aus dem Fachbereich der Botanik. Diese Forschung ist von besonderem Interesse für die vorliegende Arbeit, denn es wird das gleiche neuronale generative Sprachmodell verglichen sowie z.T. die gleiche Qualitätsbewertung der Ergebnisse verwendet. Die Autor:innen haben ein Korpus für Englisch (437.663 Wörter) und ein Korpus für Spanisch erstellt. Als Quellen dienten Enzyklopädien, Monografien, Zeitschriften und Online-Ressourcen. Die Korpora wurden mit einem Algorithmus annotiert, der auf menschlicher Annotation basiert. Annotiert wurden Eigennamen, und ob sie Ein- oder Mehrwortbenennungen sind (vgl. Premasiri et al. 2023: 877–878). Die Autor:innen haben dafür IOB-Tagging verwendet. Wörter, die ein Teil einer Mehrwortbenennung sind, wurden dementsprechend mit den Tags B (beginning), I (inside) und O (outside) markiert, wodurch für die Modelle erkennbar gemacht wurde, was zu der Mehrwortbenennung gehört, wann sie beginnt und endet, z. B. *Blue (B) Moon (I). Slow-growing (O), compact (O), clump-forming (O), perennial (O)*. (Premasiri et al. 2023: 878) Im Anschluss wurden Mehrwortbenennungen mit diskriminativen Transformer-Modellen extrahiert, wobei nicht alle für Spanisch verwendet werden konnten, denn nicht alle Modelle bedienen mehrere Sprachen. Zu den analysierten Modellen gehörten BERT, XLM-R (Conneau et al. 2020), XLNet (Yang et al. 2019), RoBERTa (Liu et al. 2019), ELECTRA (Clark et al. 2020) und SciBERT (Beltagy et al. 2019) inklusive ihrer Varianten. Im nächsten Schritt kam das neuronale generative Sprachmodell zum Einsatz. Die Autor:innen haben sich für ein Prompt entschieden, um die Terminologie zu gewinnen und haben es folgendermaßen formuliert:

Find whether there is a multi-word expression flower or plant name in the text delimited by “ ‘ – if there is no multi-word expression found in the given text; just tell ‘No’ – if you find a multiword expression in the given text; say yes and then give the multiword flower or plant name for example; Yes – ‘Name’ Text : “ ‘{sentence}“ ‘. (Premasiri et al. 2023: 879)

Der Temperatur-Parameter von GPT wurde zwecks der Vergleichspräzision auf 0 gesetzt. Die Evaluierung der Ergebnisse wurde mithilfe von Qualitätsmetriken durchgeführt, indem die Genauigkeit, Trefferquote und als Resultat der F1-Score berechnet wurden. Die Studie hat ergeben, dass alle diskriminativen Transformer-Modelle für diese konkrete Extraktion die Leistung des neuronalen generativen Sprachmodells deutlich übertreffen und bis zu 94,3 % (bert-large-cased) F1-Score für Englisch und 82,1 % (bert-base-multilingual-cased) für Spanisch erreichen. Die Resultate einzelner Transformer-Modelle waren sehr vergleichbar und lagen zwischen 91 und 94% F1-Score. Im Vergleich erreichte das neuronale generative Sprachmodell respektive 63,3 % für Englisch und 47,7 % F1-Score für Spanisch. Die Studie hat gezeigt, dass die von ChatGPT 3.5 erbrachte Leistung bei der Extraktion der Mehrwortbenennungen (Eigennamen) aus dem Fachbereich der Botanik in englischer und spanischer Sprache schlechter als jene der diskriminativen Transformer-Modelle ist und ChatGPT 3.5 bessere Resultate für Englisch als für Spanisch liefert.

Rokas et al. (2020) widmen sich der automatischen Terminologiegewinnung der litauischen Ein- und Mehrwortbenennungen aus dem Bereich der Cybersicherheit. Litauisch gilt als eine unterrepräsentierte Sprache und da das überwachte maschinelle Lernen eine hohe Anzahl an Daten benötigt, wurde das semi-überwachte Lernen gewählt. Die Autor:innen vergleichen die Ergebnisse von acht Deep-Learning-Modellen bzw. Baselines, die entweder auf der Architektur Long-Short Term Memory (LSTM, LSTM-CRF, Bi-LSTM, Bi-LSTM-CRF) oder der Gated Recurrent Units (GRU, GRU-CRF, Bi-GRU, Bi-GRU-CRF) basieren. Außerdem wurden die Word Embeddings von FastText und kontextuellen Embeddings von BERT (Devlin et al. 2019) sowie mehrere Optimierer eingesetzt. Diese Forschung ist von Interesse für die vorliegende Arbeit, denn sie basiert auf neuronalen Ansätzen für die Terminologiegewinnung aus einem spezifischen Fachbereich in einer anderen Sprache als Englisch und für den Vergleich wurden ebenfalls die Genauigkeit, der Trefferquote und der F1-Score berechnet. Es wurden folgende Schritte unternommen: zunächst wurde ein Korpus bestehend aus: themenbezogenen Rechtsakten, administrativen Dokumenten, Übersetzungen der EU-Rechtstexte, Übersetzungen der internationalen Konventionen, Publikationen und Informationen für die Öffentlichkeit erstellt (vgl. Rokas et al. 2020: 41). Das Korpus besteht aus originalen und übersetzten Texten,

die zwischen 1999-2019 entstanden sind und das Thema der Cybersicherheit behandeln. Es enthält 2.363.618 Wörter. Darauffolgend wurde der Goldstandard kompiliert, um die untersuchten Modelle zu trainieren. Dafür wurde ein Teil vom Korpus genommen (66.706 Wörter) und es wurden 1.258 einschlägige Termini annotiert. Die Annotation erfolgte nach dem BIESO-Format, respektive beginning, inside, end, single, outside (vgl. Durango et al. 2023: 289). Es wurden Nomen, nominale Phrasen, Abkürzungen und Kombinationen aus Abkürzungen und nominalen Phrasen annotiert solange sie aus dem Bereich der Cybersicherheit kommen oder damit in Verbindung stehen (vgl. Rokas et al. 2020: 41–42). In der Vorbereitungsphase wurde eine Konvertierung ins Plain-Text-Format, Änderung der Zeichenkodierung, Tokenisierung und Textformatierung durchgeführt, außerdem wurden Stoppwortlisten erstellt. Die Untersuchung war dreiteilig. Zuerst wurden 8 LSTM- und GRU-Basismodelle samt dem Optimierungsalgorithmus ADAM getestet. Die besten Resultate erzielten das Bi-LSTM-CRF-Modell mit 70,3% F1-Score und das Bi-GRU-CRF-Modell mit 69,2% F1-Score. Weiters wurden diese zwei Modelle, die am besten abgeschnitten haben, mit verschiedenen Optimierern getestet und die besten Resultate erzielten Bi-LSTM-CRF mit dem Optimierer AdaDelta (75,5% F1-Score) und Bi-GRU-CRF mit dem Optimierer RMSprop (73,3% F1-Score). In der letzten Phase der Untersuchung wurde das Modell, das in der zweiten Phase am besten abgeschnitten hat, mit einem Bi-LSTM-Netzwerk mit BERT und dem Optimierungsalgorithmus ADAM verglichen. Die Studie hat ergeben, dass alle Deep-Learning-Modelle erfolgreich für die automatische Termextraktion am Beispiel der litauischen Sprache im Fachbereich der Cybersicherheit eingesetzt werden können, denn alle erreichen gute Resultate (vgl. Rokas et al. 2020: 45). Das beste Ergebnis mit 79,4% Genauigkeit, 77,8% Trefferquote und 78,6 % F1-Score erzielte das in der letzten Phase untersuchte Modell mit BERT.

Mouratidis et al. (2022) analysieren die automatische Termextraktion von griechischen Ein- und Mehrwortbenennungen aus maritimen Rechtstexten. Verwendet wurden Word Embeddings und eine Architektur für neuronale Netze. Diese Forschung ist von Interesse für die vorliegende Arbeit, denn sie basiert auf neuronalen Ansätzen und rechtlichen Texten. Weiters wurde eine andere Sprache als Englisch analysiert und es wurden die Genauigkeit, die Trefferquote und der F1-Score berechnet, was auch auf die vorliegende Masterarbeit zutrifft. Die Autor:innen haben als Quelle für den Datensatz die Korpora des griechischen Amtsblatts Official Government Gazette of the Hellenic Republic (EN) genommen, die aus rechtlichen Texten mit maritimbezogenen Themen bestehen (Mouratidis et al. 2022: 4). Die Texte für das Korpus

stammen aus den Jahren 1975-1999 und das Korpus enthält 80.000 Wörter. Da viele rechtliche Termini nicht explizit dem maritimen Fachbereich zuzuordnen waren (25% maritim und 75% nicht-maritim) wurde die Oversampling-Methode SMOTE verwendet, um neue synthetische Trainingsdaten hinzuzufügen, damit die Ergebnisse mehr ausgewogen sind (vgl. 2022: 6). In der Vorbereitungsphase wurden Stoppwortlisten erstellt und Stemming wurde durchgeführt. Anschließend wurden relevante Fachausdrücke im Korpus gezählt, deren Anzahl sich auf 4.800 belief. Die Autor:innen haben sich zusätzlich für den Datensatz WordSim353 entschieden. Für die Untersuchung wurde die Architektur Feedforward Neural Network (FNN) ausgewählt und das Netzwerk wurde mit dem Optimierungsalgorithmus ADAM trainiert. Die Autor:innen verzeichneten die Genauigkeit von 73% für nicht-maritime Termini und 71% für maritime Termini und eine Trefferquote von 69% und respektive 74%.

De Paiva et al. (2023) untersuchen die automatische Termextraktion aus dem Fachbereich der Mathematik, genauer gesagt aus dem Sachgebiet der Kategorientheorie. Verwendet wird dafür das neuronale generative Sprachmodell ChatGPT, weswegen die Forschung von Interesse für die vorliegende Arbeit ist. Die Leistung des Tools wird mit dem Jaccard-Koeffizienten bemessen, der in diesem Fall die Ähnlichkeit der extrahierten Termkandidaten mit jenen in dem Referenzdatensatz kalkuliert. Dies ermöglicht eine quantitative Bewertung der erbrachten Leistung in Bezug auf die Genauigkeit der gewonnenen Termini (vgl. Chataut et al. 2024: 3). Die Autor:innen verwenden ein Korpus aus dem Jahr 2020, das aus 3.200 Sätzen besteht. Die zusammengestellten Texte schließen 755 akademische Abstracts aus der Fachzeitschrift *Theory and Application of Categories* ein. Drei Annotator:innen wurden gebeten, eine humane Annotation durchzuführen, damit die Ergebnisse der automatischen Termextraktion mit der Terminologierecherche verglichen werden konnten. Den Annotator:innen wurden mehrere Vorgaben zur Verfügung gestellt, die z. B. den Umgang mit Mehrwortbenennungen oder Wörtern erläutern, die je nach Kontext sowohl zu der Fremdsprache als auch der Allgemesprache gehören könnten wie das Wort/der Fachausdruck *group* (De Paiva et al. 2023: 4–5). Für die humane Annotation wurde das Tool Math Concept Annotator verwendet. Im Anschluss haben sich die Annotator:innen - nicht ohne Schwierigkeiten/Diskussionen - auf einen Goldstandard geeinigt. Im nächsten Schritt wurden Prompts für ChatGPT erstellt. Das erste Prompt lautete *Given the following Context, extract the words that denote Math concepts* (2023: 6). Die Autor:innen haben außerdem ChatGPT mit mehreren Beispielen versorgt. Da die automatisch generierten Termkandidaten nicht den Vorgaben für Annotator:innen entsprachen und nicht

zufriedenstellend waren, z. B. wurde das ursprüngliche Prompt spezifiziert und mit weiteren Beispielen versehen. Ein erkanntes Problem war, dass z. B. Namen von Personen und allgemeinsprachliche Wörter sowie Singular- und Pluralformen von Nomen extrahiert wurden. Die neue Termkandidatenliste enthielt trotzdem mehrere Nicht-Termini, weswegen die Autor:innen im letzten Schritt dem Prompt Erklärungen bzw. Begründungen hinzugefügt haben, mit der Absicht, die Genauigkeit der erzielten Ergebnisse zu erhöhen, was sie folgendermaßen formulierten:

Reason: the concept ‘additive category’ is generalized from the sentence because of the original phrase ‘category is additive’; the concept ‘sheaf’ is a known math concept shown in the sentence; the concept ‘analytic function’ is extracted from the original phrase ‘analytic functions’ by removing the plural form; here we don’t want the single ‘category’ and ‘additive’ extracted as math concepts since they are usual words nor do we want the phrase ‘category of analytic functions’ since the more concise phrase ‘analytic function’ should be extracted instead as a math concept. (De Paiva et al. 2023: 7)

Es wurden insgesamt drei Untersuchungen durchgeführt, in weiterer Folge wird allerdings nur auf die zweite Untersuchung genauer eingegangen. ChatGPT und ein:e Annotator:in haben unabhängig voneinander Termkandidaten aus 436 Sätzen vom o.g. Korpus gewonnen. Der erzielte Jaccard-Score lag bei 0.531. Daraufhin hat der:die zweite Annotator:in die von ChatGPT generierten Termini überprüft und ca. 59% der Termkandidaten herausgefiltert, was den Jaccard-Score auf 0.631 steigen ließ. Zu den entfernten Termkandidaten gehörte z. B. die nominale Phrase *previous work*. Es wurden lediglich Ausdrücke gelöscht, die ohne jeglichen Zweifel keine Termini waren (De Paiva et al. 2023: 9–10). Die Untersuchung hat drei wichtige Aspekte veranschaulicht, die auch für die vorliegende Arbeit von Interesse sind: Unterschiede zwischen den humanen Annotationen, die in dem Fall mehrstufige Formulierung des Prompts und das von ChatGPT erzielte Ergebnis. Die Autor:innen sind aufgrund der Untersuchung der Ansicht, dass ChatGPT aktuell zwar bei der Gewinnung der Terminologie aus dem Bereich der Mathematik behilflich sein kann, aber immer noch nicht Menschen dabei ersetzen kann (2023: 11).

Chataut et al. (2024) analysieren die Leistung von drei LLM-Modellen Llama2-7B, Falcon-7B und GPT-3.5 bei der automatischen Terminologiegewinnung, was in Bezug auf ChatGPT mit der vorliegenden Arbeit koinzidiert. Es wurden zwei verschiedene Korpora verwendet, die aus den Abstracts der akademischen Artikel bestehen, welche in wissenschaftlichen Fachzeitschriften veröffentlicht wurden und die Fachbereiche der Informatik und Biomedizin betreffen. Die Korpora enthalten jeweils 500 Texte aus dem Inspec-Datensatz (Informatik) und

500 Texte aus dem PubMed-Datensatz (Biomedizin). Die ausgewählte Evaluierungsmethode ist der Jaccard-Koeffizient, mit anderen Worten wird die Ähnlichkeit der automatisch generierten Termkandidaten mit einer Referenzliste verglichen. Weiters wurde Zero-Shot-Prompting ausgesucht, also beinhalten die Prompts keine Beispiele. ChatGPT wurde im Rahmen der Untersuchung angewiesen, die Rolle einer Fachkraft auf dem Gebiet der automatischen Termextraktion zu übernehmen. Der Temperatur-Parameter wurde auf 0.2 gesetzt. Von den drei untersuchten LLM-Modellen hat ChatGPT-3.5 mit dem Jaccard-Koeffizienten von 0.64 für das Inspec-Korpus und 0,21 für das PubMed-Korpus die besten Resultate erzielt. Die Autor:innen bewerten die Leistung des Modells positiv und weisen darauf hin, dass der relativ niedrige Wert im zweiten Korpus möglicherweise darauf zurückzuführen ist, dass ChatGPT deutlich mehr Termkandidaten aufgelistet hat als die korrespondierende Referenzliste enthalten hat (vgl. Chataut et al. 2024: 7).

8. Methodik

Den Kern der vorliegenden Arbeit bildet eine terminologische Untersuchung. Die vorherigen theoretischen Kapitel (2.-6.) haben das nötige Wissen vermittelt, um sie durchführen, nachvollziehen und schließlich Schlüsse aus den kommenden Ergebnissen ziehen zu können und das Kapitel über den aktuellen Stand der Forschung (7.) hat von den neuesten Entwicklungen auf dem für die Untersuchung relevanten Gebiet berichtet. Das folgende Kapitel stellt die angewandte Methodik für die Durchführung der Untersuchung dar und hat somit das Ziel, die Vorgehensweise zu erläutern, wie die bereits in der Einleitung gestellten Forschungsfragen und gleichzeitig die getroffenen Annahmen sich im weiteren Kapitel aufgrund der Ergebnisse untersuchen lassen.

Untersucht wird die Leistung der ausgewählten Extraktionssysteme bei der automatischen Terminologiegewinnung aus den österreichischen und polnischen gerichtlichen Entscheidungen auf dem Gebiet des Urheberrechts. Aus der Untersuchung soll erschlossen werden können, inwieweit sie sich für den Einsatz bei dieser linguistisch gesehen herausfordernden Terminologie in den beiden analysierten Sprachen eignen. Ein besonderes Augenmerk gilt der Leistung des generativen neuronalen Sprachmodells ChatGPT-4o. Um sie bestmöglich bestimmen zu können, werden für jeden der analysierten Texte gleich mehrere Prompts verwendet, woraus sich die zweite Forschungsfrage herleiten lässt, nämlich, wie verschiedene Prompts die erbrachte Leistung beeinflussen. Mit der Analyse der Ergebnisse der Untersuchung wird eine Tendenz der auf Basis der Auseinandersetzung mit der Fachliteratur gestellten Annahmen aufgezeigt.

Das Kapitel beginnt mit der Beschreibung der Texte, aus denen die Terminologie extrahiert wurde. Es wird außerdem erläutert, welche Voraussetzungen bei der Textauswahl einzuhalten waren. Weiters werden die Systemauswahl und Vorgehensweise samt Validierung besprochen. Dazu gehören die Extraktionsprogramme, das Annotationstool, die Auswahl der Annotator:innen, der Prozess der manuellen Terminologierecherche, die Erstellung des Goldstandards, der Prozess der automatischen Termextraktion, die formulierten Prompts und schließlich die ausgewählten Evaluierungsmaßnahmen.

8.1. Textauswahl

Für die Untersuchung wurden drei österreichische und drei polnische Textfragmente in vergleichbarer Länge von ca. 1.000 - 1.500 Wörtern aus den gerichtlichen Entscheidungen ausgewählt. Da die Materialbeschaffung aus offiziellen Quellen erfolgte, kann davon ausgegangen

werden, dass die Texte sowohl von Erstsprachler:innen als auch Fachexpert:innen verfasst wurden und über hohe terminologische Präzision verfügen. Die österreichischen Texte (siehe Anhang 1-3) stammen vom Rechtsinformationssystem des Bundes (RIS) und es sind Entscheidungstexte des Obersten Gerichtshofs aus dem Rechtsgebiet des Zivilrechts und aus den Fachgebieten des Urheberrechts, des Unionsrechts und des gewerblichen Rechtsschutzes. Es handelt sich um aktuelle Texte aus den Jahren 2023 – 2024, die Rechtsterminologie beinhalten.

Der erste österreichische Text (siehe Anhang 1) betrifft eine Klage über die Nutzung von Fotos ohne Einwilligung des Urhebers. Nach einem Stromausfall in einem Vergnügungspark wurden von dem Kläger zufällig Rettungsaktionen aus Fahrgeschäften beobachtet und mit dem Handy fotografiert. Die Fotos wurden anschließend auf seinem Twitter-Account veröffentlicht. Vom beklagten Rundfunk wurde um Erlaubnis gebeten, die Bilder unentgeltlich für eine Nachrichtensendung über den Stromausfall verwenden zu dürfen, was vom Kläger abgelehnt wurde. Dennoch wurden die Bilder in der Nachrichtensendung ausgestrahlt und die Sendung zudem auf der Website des Rundfunks zum Abruf bereitgestellt (Bundeskanzleramt der Republik Österreich 2024b). Das Fragment besteht aus den ersten 17 Absätzen der Begründung und umfasst 875 Wörter.

Der zweite österreichische Text (siehe Anhang 2) betrifft eine Klage über die Ausbezahlung des restlichen Entgelts sowie Feststellungs- und Unterlassungsansprüche aufgrund der Eingriffe in Verwertungsrechte des Urhebers eines Films. Anlässlich einer Einzelausstellung wurde der Kläger mit einem filmischen Porträt des Künstlers zu vier seiner Skulpturen beauftragt. Die Produktion sowie alle sound- und schnitttechnischen Entscheidungen wurden vom Kläger organisiert und vorfinanziert. Entgegen der Zusage des Klägers erhielt die Erstbeklagte jedoch erst kurz vor der Ausstellungseröffnung einige Filmsequenzen zur Ansicht. Diese wurden wegen unpassender Farbskalierung und unzureichender Textverständlichkeit bemängelt. Noch während der Ausstellung wurde eine Überarbeitung von Ton und Farbskalierung gefordert, außerdem wurde beanstandet, dass der Künstler im Film nicht im besten Licht dargestellt werde (Bundeskanzleramt der Republik Österreich 2023a). Das Fragment besteht zum Teil aus den Entscheidungsgründen (Absätze 9-14) und zum Teil aus der rechtlichen Beurteilung (Absätze 15-32) und umfasst 1.399 Wörter.

Der dritte österreichische Text (siehe Anhang 3) betrifft eine Klage über die Nutzung von abgeänderten Illustrationen der Urheberin. Die Klägerin wurde von der Beklagten, die ein Schokolademuseum betreibt, beauftragt, Illustrationen einer Kakaobohne und einer Schokoladetafel zu erstellen, die als Maskottchen „Choco und Coco“ verwendet werden sollten. Später wurden vom Geschäftsführer der Klägerin die Illustrationen im Rahmen eines Folgeauftrags

abgeändert und dadurch Werke im Sinne einer eigentümlichen geistigen Schöpfung im Rahmen einer Bearbeitung gemäß § 5 UrhG geschaffen, für die der Klägerin die Nutzungsrechte eingeräumt wurden. Lt. der Klägerin nutze die Beklagte diese Illustrationen unberechtigt und habe sie teilweise ohne Genehmigung der Klägerin abändern lassen, was lt. der Klägerin ihr Nutzungsrecht verletze (Bundeskanzleramt der Republik Österreich 2023b). Das Fragment besteht zum Teil aus der Begründung (Absätze 6-10) und zum Teil aus der rechtlichen Beurteilung (Absätze 11-18) und umfasst 1.423 Wörter.

Die polnischen Texte (siehe Anhang 4-6) stammen vom Portal Orzeczeń Sądów Powszechnych und es sind Entscheidungstexte der polnischen Bezirks- und Appellationsgerichte aus dem Rechtsgebiet des Zivilrechts und aus den Fachgebieten des Urheberrechts und des Urhebervermögensrechts. Es handelt sich um aktuelle Texte aus den Jahren 2022 – 2023, die Rechtsterminologie beinhalten.

Der erste polnische Text (siehe Anhang 4) betrifft eine Klage über die Verletzung des Urhebervermögensrechts des Urhebers bezüglich der Erstellung von Bauplanungsunterlagen und fotorealistischen Visualisierungen. Zwischen den Parteien wurde ein mündlicher Vertrag geschlossen, wonach Bauplanungsunterlagen und fotorealistische Visualisierungen für zwei Mehrfamilienhäuser mit gemeinsamer Tiefgarage durch den Kläger im Rahmen seiner Geschäftstätigkeit erstellt werden sollten. In weiterer Folge kam es zwischen den Parteien zum Vergleich. Nach dem Rücktritt der Beklagten vom Vergleich blieben die vom Kläger erstellten Unterlagen jedoch weiterhin auf der Website der Beklagten verfügbar (Ministerstwo Sprawiedliwości 2023a). Das Fragment betrifft vorwiegend den ersten Teil des Entscheidungstexts und umfasst 1.433 Wörter.

Der zweite polnische Text (siehe Anhang 5) betrifft die Verletzung des Urhebervermögensrechts des Urhebers bezüglich der Nutzung seiner Buchübersetzung von der Beklagten für die Erstellung und Verkauf eines E-Books. Ein Verlagsvertrag wurde zwischen dem Übersetzer und dem Verleger geschlossen, in dem sich der Übersetzer verpflichtete, ein Buch aus dem Englischen ins Polnische zu übersetzen. Dabei wurde vereinbart, dass dem Verlag das ausschließliche Recht zur Veröffentlichung des Werks in allen Verwertungsbereichen übertragen wird, die die Vervielfältigung und Verbreitung des Werks auf allen zum Zeitpunkt des Vertrags bekannten Medien umfassen. Anschließend erteilte der Verleger der Beklagten eine ausschließliche Lizenz zur Nutzung der Übersetzung, um ein E-Book zu veröffentlichen und zu verkaufen. Die Beklagte vermarktete und bot die Übersetzung in Form eines E-Books zum Verkauf an. Obwohl der Übersetzer bereit war, einen E-Book-Lizenzvertrag mit der Beklagten abzuschließen, kam ein solcher Vertrag nicht zustande (Ministerstwo Sprawiedliwości 2023b).

Das Fragment betrifft den ersten und den mittleren Teil des dritten Teils des Entscheidungstexts und umfasst 1.235 Wörter.

Der dritte polnische Text (siehe Anhang 6) betrifft die Klage über die Verletzung des Urhebervermögensrechts des Urhebers bezüglich der Nutzung des von ihm entworfenen technischen Konzepts. In dieser Rechtssache wurde von dem Kläger die Zahlung von Schadensersatz verlangt, da seiner Ansicht nach die Beklagten sein Urheberrecht an einem Entwurf für eine begehbare Fußgängerbrücke über einen Hafenkanal verletzt hätten (Ministerstwo Sprawiedliwości 2022). Das Fragment kommt aus dem letzten Teil des Entscheidungstexts und umfasst 1.452 Wörter.

8.2. Systemauswahl

Die zur Durchführung der Untersuchung ausgewählten Extraktionsprogramme sind Text2TCS und ChatGPT-4o. ChatGPT wird zwar zugeschrieben in jedem Beruf Geselle, in keinem ein Meister zu sein (jack of all trades, master of none) (vgl. Kocoń et al. 2023), aber durch die steigende Beliebtheit in allen Lebensbereichen wird man mit der Frage konfrontiert, wie gut die Anwendung sich zum Einsatz im translatorischen und genauer gesagt im terminologischen Alltag eignet. Damit die Leistung am Gebiet der automatischen Termextraktion relativiert werden kann, wird ein Tool gebraucht, das ebenfalls auf neuronalen Ansätzen basiert, aber gezielt für diese Zwecke entwickelt wurde. Dafür hat sich das erst vor Kurzem am Zentrum für Translationswissenschaft entwickelte Tool Text2TCS als bestens geeignet erwiesen. Ausschlaggebend für die Auswahl der Tools waren außerdem ihre freie Verfügbarkeit, Unterstützung beider untersuchten Sprachen, intuitive Nutzung und im Falle von Text2TCS die in anderen Studien erzielte hohe Leistung bei der Termextraktion (siehe 3.1.4.1.). Die Auswahl von ChatGPT-4o ist auf die o.g. Beliebtheit des Tools sowie auf das persönliche Interesse des Autors der vorliegenden Arbeit zurückzuführen.

8.3. Vorgehensweise und Validierung

Die Vorgehensweise schloss in erster Linie die Auswahl der passenden Texte ein, aus denen in weiterer Folge terminologiereiche Fragmente ausgesucht wurden. Es wurde außerdem darauf geachtet, dass die Fragmente verschiedene Teile der Entscheidungen umfassen, weswegen einige vom Anfang des Texts und andere wiederum von der Mitte oder vom Schluss kommen. Von Bedeutung war die Maschinenlesbarkeit der Texte, weshalb sie als Microsoft-Word-Do-

kumente gespeichert wurden. Weiters wurden alle sechs Texte im Annotationstool KG Annotation Tool von dem Autor der vorliegenden Arbeit manuell annotiert, wie in Abbildung 17 dargestellt.

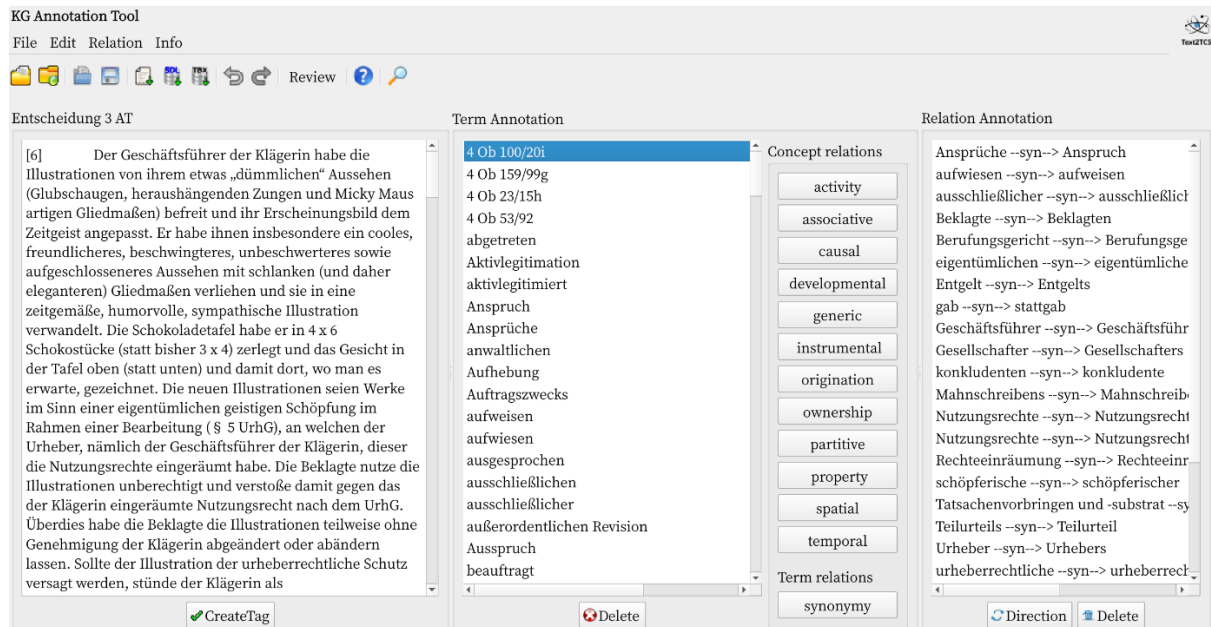


Abbildung 17: KG Annotation Tool (Gromann et al. 2021)

Die manuelle, nur durch ein Tool gestützte Terminologierecherche bzw. Annotation hatte zum Ziel, den Noise- und Silence-Effekt bestmöglich zu vermeiden. Aus diesem Grund wurden auch Textfragmente in überschaubarer Länge ausgesucht. Das Gesamtkorpus für Deutsch enthält 3.697 Wörter und für Polnisch 4.120 Wörter. Es wurden terminologisch relevante Einwortbenennungen, wie Nomen, Verben, Adjektive/Partizipien sowie Mehrwortbenennungen annotiert. Das Tool ermöglicht es, Synonyme zu markieren (siehe Abbildung 17), was in diesem Fall die gleichen Termini mit unterschiedlicher Schreibweise z. B. aufgrund der Flexion bedeutet, was auch zusätzlich annotiert wurde. Im Zuge der manuellen Annotation wurde die Spezifik der Rechtssprache berücksichtigt und es wurde genau auf den Kontext geachtet, in dem der potenzielle Fachausdruck eingebettet ist. Dies bedeutet, dass auch die Termini in die Referenzliste aufgenommen wurden, die man typischerweise der Gemeinsprache zuordnen würde, welche aber in dem konkreten Kontext die terminologische Relevanz aufweisen, wie bereits im Kapitel 5.2.2.1. zum Ausdruck gebracht wurde.

Um die Validität und Objektivität der Untersuchung sicherzustellen, wurden zwei weitere Annotatorinnen gebeten, getrennt von dem Erstannotator, ihre eigene Annotation für jeweils drei Texte im Annotationstool durchzuführen. Es erfolgte eine kurze Erklärung und Ein-

führung in das Tool, sonst wurden die Annotatorinnen nicht beeinflusst. Bei den Zweitannotatorinnen handelt es sich um Erstsprachlerinnen, die zugleich Studentinnen am Zentrum für Translationswissenschaft sind und kurz vor ihrem Masterabschluss stehen, wenngleich ihre Spezialisierung nicht die Terminologie betrifft. Die Zweitannotatorinnen haben relativ wenig Erfahrung mit der terminologischen Arbeit bis auf die im Rahmen des Studiums erworbenen theoretischen Grundlagen. Weder der Erstannotator noch die Zweitannotatorinnen sind Rechtsexpert:innen oder beeidigte Übersetzer:innen bzw. Dolmetscher:innen, weshalb die Annotation nach dem Prinzip des besten Wissens und Gewissens erfolgte. Die Expert:innen für die geplante Untersuchung zuzuziehen wäre aus praktischen Gründen schwierig und es würde womöglich den Rahmen der vorliegenden Masterarbeit überschreiten.

In weiterer Folge wurden Goldstandards für alle sechs Texte erstellt (siehe Anhang 7, Anhang 8). Wie die Auseinandersetzung mit der Fachliteratur (4.1.) und die neuesten Studien am Gebiet der automatischen Termextraktion (7.) zeigen, ist die (manuelle) Erstellung eines Goldstandards die übliche Vorgehensweise bei der Evaluation im Bereich der natürlichen Sprachverarbeitung. Der Erstannotator und die jeweilige Zweitannotatorin haben sich infolge einer intensiven Besprechung darauf geeinigt, welche Termini von beiden Annotationen schließlich den Goldstandard für jeden einzelnen Text bilden sollten. Im Zweifelsfall wurde die terminologische Relevanz der Termkandidaten anhand von verschiedenen Quellen nach bestem Wissen überprüft. Hierbei kamen einerseits die Suchmaschinensuche sowie Recherche in Glossaren (u.a. Birsak et al. 2021, 2022) und Wörterbüchern (u.a. Kilian & Kilian 2021a, 2021b) zum Einsatz. Weiters wurden die Ergebnisse manuell lemmatisiert und alle Synonyme in der o.g. Bedeutung wurden entfernt.

Beim Prompt Engineering wurde darauf geachtet, dass die Anfragen zwar ChatGPT dazu bringen, dieselbe Aufgabe zu erfüllen, aber dabei unterschiedlich formuliert werden. Die ersten zwei Anfragen wurden knapp formuliert und ChatGPT wurde geduzt. Bei weiteren zwei Anfragen wurde die Aufgabe höflicher und ausführlicher formuliert, ChatGPT wurde dabei gesiezt. In dem fünften Prompt wurde ChatGPT wieder gesiezt, die Anfrage wurde konkretisiert und es wurde eine Erklärung hinzugefügt, wofür die Liste mit Termini gebraucht wird. Es wurde die Zero-Shot-Prompting-Strategie eingesetzt. Bei den polnischen Anfragen handelt sich um eine Übersetzung, allerdings wurde wegen der Spezifik der polnischen Sprache auf die Sie-Form verzichtet, die eine Geschlechtsbestimmung voraussetzen müsste. Die zur Extraktion der Terminologie verwendeten Prompts sind der Tabelle 6 zu entnehmen.

Tabelle 6: Die für die Untersuchung verwendeten Prompts

Prompt	Deutsch	Polnisch
1	Liste alle Fachbegriffe aus dem folgenden Text auf.	Sporządź listę wszystkich terminów fachowych w poniższym tekście.
2	Finde alle Fachtermini aus dem folgenden Text.	Znajdź wszystkie terminy specjalistyczne w poniższym tekście.
3	Könnten Sie bitte alle Fachtermini aus dem folgenden Text extrahieren? Das wäre sehr nett von Ihnen.	Czy mógłbyś proszę wypisać wszystkie terminy fachowe z poniższego tekstu? Bardzo Cię proszę.
4	In dem folgenden Text befinden sich verschiedene Fachwörter. Ich möchte eine vollständige Liste aller Fachwörter im Text haben. Könnten Sie mir dabei helfen?	W poniższym tekście znajduje się terminologia specjalistyczna. Potrzebowałbym kompletną listę terminów fachowych z tekstu. Pomógłbyś mi proszę?
5	Ich muss eine gerichtliche Entscheidung aus dem Bereich des Urheberrechts übersetzen und brauche alle Fachtermini aus dem folgenden Text zur Vorbereitung. Listen Sie bitte alle Fachtermini auf.	Mam do przetłumaczenia orzeczenie sądowe z dziedziny prawa autorskiego. W celu przygotowania się do tłumaczenia potrzebowałbym wszystkie terminy specjalistyczne znajdujące się w poniższym tekście. Czy mógłbyś sporządzić kompletną listę wszystkich terminów fachowych z tekstu?

In der nächsten Phase wurden die Termini aus allen Texten automatisch mit Text2TCS gewonnen und die Ergebnisse wurden in einer Microsoft-Word-Datei gespeichert (siehe Anhang 9, Anhang 10). Das Einlesen der Texte erfolgte durch Kopieren und Einfügen in das dafür vorgesehene Feld. Die Sprachen wurden automatisch erkannt. Die Texte mussten aufgrund der Länge in jeweils zwei Teile geteilt werden. Schließlich wurden alle Korpora der automatischen Terminologiegewinnung mit ChatGPT unterzogen (siehe Anhänge 11-16). Es wurden aus allen Texten die Termkandidaten mit jedem genannten Prompt extrahiert. Es ist anzumerken, dass die Korpora keiner Rechtschreibprüfung in einem Textverarbeitungsprogramm unterzogen wurden, die ggf. die Leistung der Extraktionsprogramme positiv beeinflussen könnte. Da es sich um digitale Originaldokumente handelt, sind etwaige Schreibfehler auf den Faktor Mensch zurückzuführen.

8.4. Auswertungsmethode

Die Leistung der automatischen Termextraktion wurde quantitativ mittels der Qualitätsmetriken Genauigkeit, Trefferquote und F1-Score ermittelt. Die Duplikate der Termkandidaten wur-

den automatisch entfernt. Zunächst erfolgte die Evaluierung von nicht lemmatisierten automatisch gewonnenen Termkandidaten, da dies in Anbetracht der praktischen Anwendung eines Extraktionstools mehr der Realität entspricht. Im weiteren Schritt wurden allerdings alle Termkandidatenlisten automatisch lemmatisiert, um zu untersuchen, inwieweit die Lemmatisierung die erbrachte Leistung beeinflusst und da der Goldstandard lemmatisiert war. Für diesen Zweck wurde spaCy (ExplosionAI GmbH 2024) verwendet und zwar das Modell `pl_core_news_lg` für Polnisch und `de_core_news_lg` für Deutsch, die jeweils größten verfügbaren Modelle basierend auf neuronalen Sprachmodellen. Anschließend wurden die Termkandidatenlisten von Text2TCS und jene von allen Prompts von ChatGPT automatisch mit dem jeweiligen Goldstandard verglichen und die Werte wurden erfasst, was den tatsächlichen Kern der vorliegenden Arbeit ergibt. Außerdem wurden auch die Berechnungen für einzelne Annotator:innen durchgeführt und die Ergebnisse ermittelt. Im letzten Schritt wurde ein Korpus pro Sprache, bestehend aus jeweils drei Texten, gebildet; die drei entsprechenden Goldstandards wurden zu einem Goldstandard vereint und es wurden die Werte ermittelt. Dies erfolgte mittels der Programmiersprache Python. Der verfasste Code war darauf ausgerichtet, die oben genannten Qualitätsmetriken zu kalkulieren. Dafür wurde eine Expertin zugezogen. Im Zuge der qualitativen Inhaltsanalyse der Ergebnisse wurde auch die Leistung von einzelnen Prompts bei der Terminologiegewinnung durch ChatGPT untersucht und es wurde ein Versuch unternommen, erkennbare Schwächen der Leistung beider Extraktionstools in beiden Sprachen ausfindig zu machen sowie die Leistung der Extraktionstools für Deutsch und Polnisch zu ermitteln.

9. Ergebnisse

Das folgende Kapitel stellt die Ergebnisse der durchgeführten quantitativen Untersuchung in Form von Tabellen, Diagrammen und Beschreibungen dar. Der besseren Übersichtlichkeit halber werden in den folgenden Tabellen die Qualitätsmetriken Genauigkeit mit G, Trefferquote mit T und F1-Score mit F1 markiert. Außerdem werden die ermittelten Qualitätsmetriken sowie die automatisch generierten Termini im Rahmen einer manuellen Analyse beschrieben und mit konkreten Beispielen untermauert. Es werden hauptsächlich Ergebnisse mit der Relevanz für die Forschungsfragen beschrieben, wenngleich auch andere Daten präsentiert werden, auf die im weiteren Kapitel (10.) eingegangen wird. Zuerst wird die Leistung beider Extraktionsprogramme bei der Gewinnung von nicht lemmatisierten Termkandidaten präsentiert. Die ermittelten Werte beziehen sich auf folgende Kategorien: den Vergleich von Text2TCS und Goldstandard, zwischen Text2TCS und dem Erstannotator sowie von Text2TCS und den Zweitannotatorinnen. Die gleichen Kategorien betreffen die Leistung von ChatGPT, wobei jeder Prompt einzeln untersucht wird. Zwecks besserer Überschaubarkeit der Leistung von ChatGPT wird anschließend ein Durchschnittswert aller fünf Prompts pro Text ermittelt und mittels eines Diagramms veranschaulicht. Es erfolgen dementsprechend 18 F1-Score-Messungen pro Text, insgesamt 54 pro Sprache und 108 Berechnungen für die Untersuchung einzelner Texte anhand der nicht lemmatisierten Termini. Dasselbe Vorgehen wird für die lemmatisierten Termini wiederholt, was die Gesamtanzahl der F1-Score-Evaluierungen einzelner Texte samt der Evaluierung anhand des Gesamtkorpus pro Sprache auf 252 steigern lässt. Parallel zum F1-Wert wurden auch Genauigkeits- und Trefferquotenwerte gemessen und festgehalten. Zum Schluss jeder quantitativen Analyse wird die Sammelleistung pro Sprache anhand des Korpus dargestellt, das aus allen drei Texten der jeweiligen Sprache besteht. Zum Abschluss der beiden Unterkapitel zur quantitativen Untersuchung werden die Unterschiede in den Werten nach der Lemmatisierung veranschaulicht. Im Unterkapitel zur qualitativen Analyse der Ergebnisse werden auffallende Muster bzw. Problemstellungen geschildert und mit Beispielen versehen, wobei einerseits Bezug auf die im theoretischen Teil gewonnenen Erkenntnisse und andererseits auf eigene Beobachtungen genommen wird.

9.1. Quantitative Evaluierung von nicht lemmatisierten Termkandidaten

In Tabelle 7 wurden die ermittelten Genauigkeits- und Trefferquoten sowie F1-Werte auf Basis der von Text2TCS gewonnenen Termkandidaten präsentiert. Die wichtigste Messung betrifft jene zwischen Text2TCS und dem Goldstandard. Zusätzlich dazu wurden die Werte erfasst, die

aus der Messung zwischen Text2TCS und dem Erstannotator sowie zwischen Text2TCS und der Zweitannotatorin hervorgehen, also bevor die Annotator:innen sich auf den Goldstandard einigten. Die Goldstandards der drei österreichischen Texte (1-3) und der drei polnischen Texte (4-6) bestehen aus 103, 205, 176, 153, 154 und 144 Termkandidaten. Die Anzahl der Text2TCS-Termkandidaten beläuft sich auf 142, 216, 248, 250, 196 und 256. Die Liste des Erstannotators besteht nach der automatisierten Entfernung der Duplikate aus 102, 213, 188, 197, 200 und 181 Termkandidaten. Für die Zweitannotatorin sind es 50, 128, 150, 161, 165 und 185 Termkandidaten.

Tabelle 7: Evaluierungsergebnisse von Text2TCS (nicht lemmatisiert)

Sprache	Text	Erstannotator			Zweitannotatorin			Goldstandard		
		G	T	F1	G	T	F1	G	T	F1
AT	1	31,69	44,12	36,89	28,17	80,0	41,67	23,24	31,73	26,83
	2	62,04	62,91	62,47	47,22	79,69	59,3	45,83	48,29	47,03
	3	38,71	51,06	44,04	42,34	70,0	52,76	26,61	37,71	31,2
PL	4	41,6	52,79	46,53	29,6	36,01	32,45	6,4	10,46	7,94
	5	47,96	47,0	47,48	35,71	42,42	38,78	12,76	16,23	14,29
	6	35,94	50,83	42,11	40,23	55,68	46,71	5,47	9,72	7,0

Abbildung 18 veranschaulicht die erzielten Werte von Text2TCS und Goldstandard.

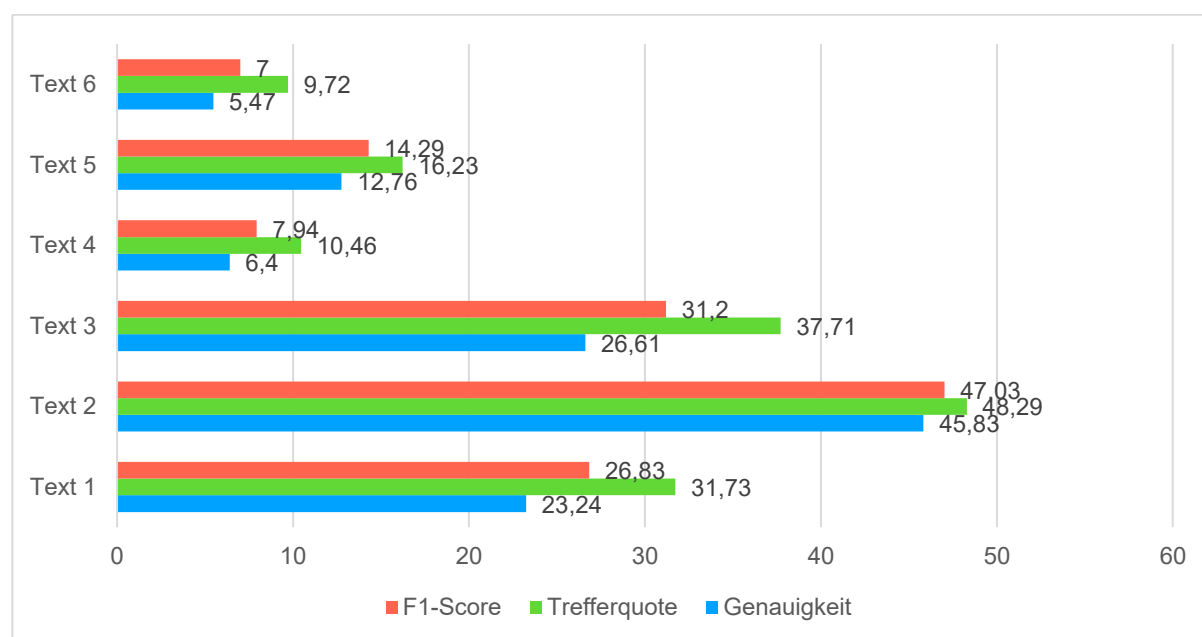


Abbildung 18: Ergebnisse von Text2TCS und Goldstandard (nicht lemmatisiert)

Text2TCS erzielte den höchsten F1-Wert von 47,03 bei Text 2. Für alle Texte generierte das Tool höhere Trefferquoten- als Genauigkeitswerte und erzielte bessere Ergebnisse bei der Extraktion österreichischer als polnischer Rechtsterminologie. Für Deutsch liegen die Genauigkeitswerte zwischen 23,24 und 45,83, die Trefferquotenwerte zwischen 31,73 und 48,29 und F1-Score zwischen 26,83 und 47,03. Für Polnisch liegen die Werte zwischen 5,47 und 12,76 für die Genauigkeit, zwischen 9,72 und 16,23 für die Trefferquote und zwischen 7 und 14,29 für den F1-Score. Auffällig ist die verhältnismäßig hohe Anzahl der Text2TCS-Termkandidaten im Vergleich zum Goldstandard, was sich in der hohen Trefferquote widerspiegelt. Der niedrigere Genauigkeitsquotient wirkt sich automatisch reduzierend auf den F1-Score aus. Außerdem lässt sich aus den in Tabelle 7 ermittelten Werten herleiten, dass Text2TCS höhere Übereinstimmung mit beiden Annotator:innen als mit dem Goldstandard gezeigt hat. Dies spricht allerdings nicht unbedingt für die Qualität der erbrachten Leistung, denn den Goldstandard bilden Termkandidaten, die von beiden Annotator:innen nach einer zusätzlichen Überprüfung auf Richtigkeit gemeinsam ausgewählt wurden und zuvor lemmatisiert wurden. Im Rahmen der Besprechung und Überprüfung wurden einige von den Annotator:innen ursprünglich ausgesuchte Termkandidaten gestrichen. Als Beispiel dient u.a. Text 3. Sowohl Text2TCS als auch die Zweitannotatorin haben *Gestaltungselement* als einen (rechtlichen) Fachausdruck annotiert, der es allerdings nach der Besprechung zwischen den Annotator:innen schließlich nicht in den Goldstandard geschafft hat. Auch wenn in diesem Fall eine Übereinstimmung vorliegt, ist sie aufgrund des Goldstandards als falsch zu bewerten. Andere ähnliche Beispiele aus den Texten für einen/eine oder beide Annotator:innen sind *Tagesereignis*, *Wahrnehmbarkeit*, *Vereinfachung* für Deutsch sowie *idea*, *rezultat*, *sposób wyrażenia* u.v.m. für Polnisch. Weiters lässt sich bei der Leistung von Text2TCS der Noise-Effekt beobachten. Es wurden zum gewissen Grad Termkandidaten extrahiert, die eindeutig der Gemeinsprache zuzuordnen sind wie etwa *Foto*, *Handy*, *Bild*, *Art*, *Text*, *Film* für Deutsch und *moment*, *problem*, *linia*, *kształt* für Polnisch.

ChatGPT extrahierte aus jedem untersuchten Text Termkandidaten mithilfe von fünf Prompts. Die Ergebnisse wurden pro Prompt dargestellt und sie betreffen ähnlich wie bei Text2TCS in erster Linie den Vergleich mit dem Goldstandard sowie in weiterer Folge mit dem Erstannotator und den Zweitannotatorinnen. Tabellen 8-12 veranschaulichen die Genauigkeits- und Trefferquoten sowie F1-Werte. Die entsprechenden Prompts wurden unter jeder Tabelle zitiert und separat in der Tabelle 6 (8.3.) zusammengefasst. Das Balkendiagramm in Abbildung 19 stellt den Durchschnittswert von allen fünf Prompts pro Text dar. Die, nach der Entfernung der Duplikate, der Analyse unterzogenen Termkandidaten, die von ChatGPT generiert wurden,

betragen für Text 1: 58, 55, 46, 49 und 57 Termkandidaten. Für Text 2: 74, 66, 66, 62 und 53 Termkandidaten. Für Text 3: 51, 52, 55, 66 und 62 Termkandidaten. Für Text 4: 49, 62, 51, 57 und 70 Termkandidaten. Für Text 5: 43, 46, 37, 37 und 62 Termkandidaten. Für Text 6: 46, 57, 56, 49 und 40 Termkandidaten.

Tabelle 8: Ergebnisse von ChatGPT Prompt 1 (nicht lemmatisiert)

Prompt	Text	Erstannotator			Zweitannotatorin			Goldstandard		
		G	T	F1	G	T	F1	G	T	F1
1	1	29,31	16,67	21,25	25,86	30,0	27,78	41,38	23,08	29,63
	2	55,41	19,25	28,57	47,3	27,34	34,65	54,05	19,51	28,67
	3	39,22	10,64	16,74	41,18	14,0	20,9	37,25	10,86	16,82
	4	8,16	2,03	3,25	4,08	1,24	1,9	0,0	0,0	0,0
	5	4,65	1,0	1,65	4,65	1,21	1,92	0,0	0,0	0,0
	6	6,52	1,66	2,65	8,7	2,16	3,46	6,52	2,08	3,15
Liste alle Fachbegriffe aus dem folgenden Text auf / Sporządź listę wszystkich terminów fachowych w poniższym tekście.										

Tabelle 9: Ergebnisse von ChatGPT Prompt 2 (nicht lemmatisiert)

Prompt	Text	Erstannotator			Zweitannotatorin			Goldstandard		
		G	T	F1	G	T	F1	G	T	F1
2	1	30,91	16,67	21,66	25,45	28,0	26,66	40,0	21,15	27,67
	2	65,15	20,19	30,83	59,09	30,47	40,21	59,09	19,02	28,78
	3	57,69	15,96	25,0	50,0	17,33	25,74	63,46	18,86	29,08
	4	6,45	2,03	3,09	3,23	1,24	1,79	0,0	0,0	0,0
	5	4,35	1,0	1,63	6,52	1,82	2,85	2,17	0,65	1,0
	6	8,77	2,76	4,2	19,3	5,95	9,1	33,33	13,19	18,9
Finde alle Fachtermini aus dem folgenden Text / Znajdź wszystkie terminy specjalistyczne w poniższym tekście.										

Tabelle 10: Ergebnisse von ChatGPT Prompt 3 (nicht lemmatisiert)

Prompt	Text	Erstannotator			Zweitannotatorin			Goldstandard		
		G	T	F1	G	T	F1	G	T	F1
3	1	34,78	15,69	21,62	30,43	28,0	29,16	45,65	20,19	28,0
	2	54,55	16,9	25,81	53,03	27,34	36,08	57,58	18,54	28,05
	3	54,55	15,96	24,69	58,18	21,33	31,22	56,36	17,71	26,95
	4	7,84	2,03	3,22	3,92	1,24	1,88	0,0	0,0	0,0
	5	2,7	0,5	0,84	2,7	0,61	1,0	0,0	0,0	0,0
	6	8,93	2,76	4,22	19,64	5,95	9,13	35,71	13,89	20,0
Könnten Sie bitte alle Fachtermini aus dem folgenden Text extrahieren? Das wäre sehr nett von Ihnen / Czy mógłbyś proszę wypisać wszystkie terminy fachowe z poniższego tekstu? Bardzo Cię proszę.										

Tabelle 11: Ergebnisse von ChatGPT Prompt 4 (nicht lemmatisiert)

Prompt	Text	Erstannotator			Zweitannotatorin			Goldstandard		
		G	T	F1	G	T	F1	G	T	F1
4	1	30,61	14,71	19,87	28,57	28,0	28,28	38,78	18,27	24,84
	2	64,52	18,78	29,09	64,52	31,25	42,11	62,9	19,02	29,21
	3	50,0	17,55	25,98	46,97	20,67	28,71	51,52	19,43	28,22
	4	5,26	1,52	2,36	3,51	1,24	1,83	0,0	0,0	0,0
	5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
	6	4,08	1,1	1,73	4,08	1,08	1,71	2,04	0,69	1,03
In dem folgenden Text befinden sich verschiedene Fachwörter. Ich möchte eine vollständige Liste aller Fachwörter im Text haben. Könnten Sie mir dabei helfen? / W poniższym tekście znajduje się terminologia specjalistyczna. Potrzebowałbym kompletną listę terminów fachowych z tekstu. Pomógłbyś mi proszę?										

Tabelle 12: Ergebnisse von ChatGPT Prompt 5 (nicht lemmatisiert)

Prompt	Text	Erstannotator			Zweitannotatorin			Goldstandard		
		G	T	F1	G	T	F1	G	T	F1
5	1	31,58	17,65	22,64	26,32	30,0	28,04	43,86	24,04	31,06
	2	56,6	14,08	22,55	54,72	22,66	32,05	56,6	14,63	23,25
	3	59,68	19,68	29,6	54,84	22,67	32,08	61,29	21,71	32,06
	4	5,71	2,03	3,0	2,86	1,24	1,73	0,0	0,0	0,0
	5	3,23	1,0	1,53	3,23	1,21	1,76	0,0	0,0	0,0
	6	5,0	1,1	1,8	2,5	0,54	0,89	0,0	0,0	0,0

Ich muss eine gerichtliche Entscheidung aus dem Bereich des Urheberrechts übersetzen und brauche alle Fachtermini aus dem folgenden Text zur Vorbereitung. Listen Sie bitte alle Fachtermini auf / Mam do przetłumaczenia orzeczenie sądowe z dziedziny prawa autorskiego. W celu przygotowania się do tłumaczenia potrzebowałbym wszystkie terminy specjalistyczne znajdujące się w poniższym tekście. Czy mógłbyś sporządzić kompletną listę wszystkich terminów fachowych z tekstu?

Abbildung 19 zeigt die erzielten Durchschnittswerte von ChatGPT und Goldstandard.

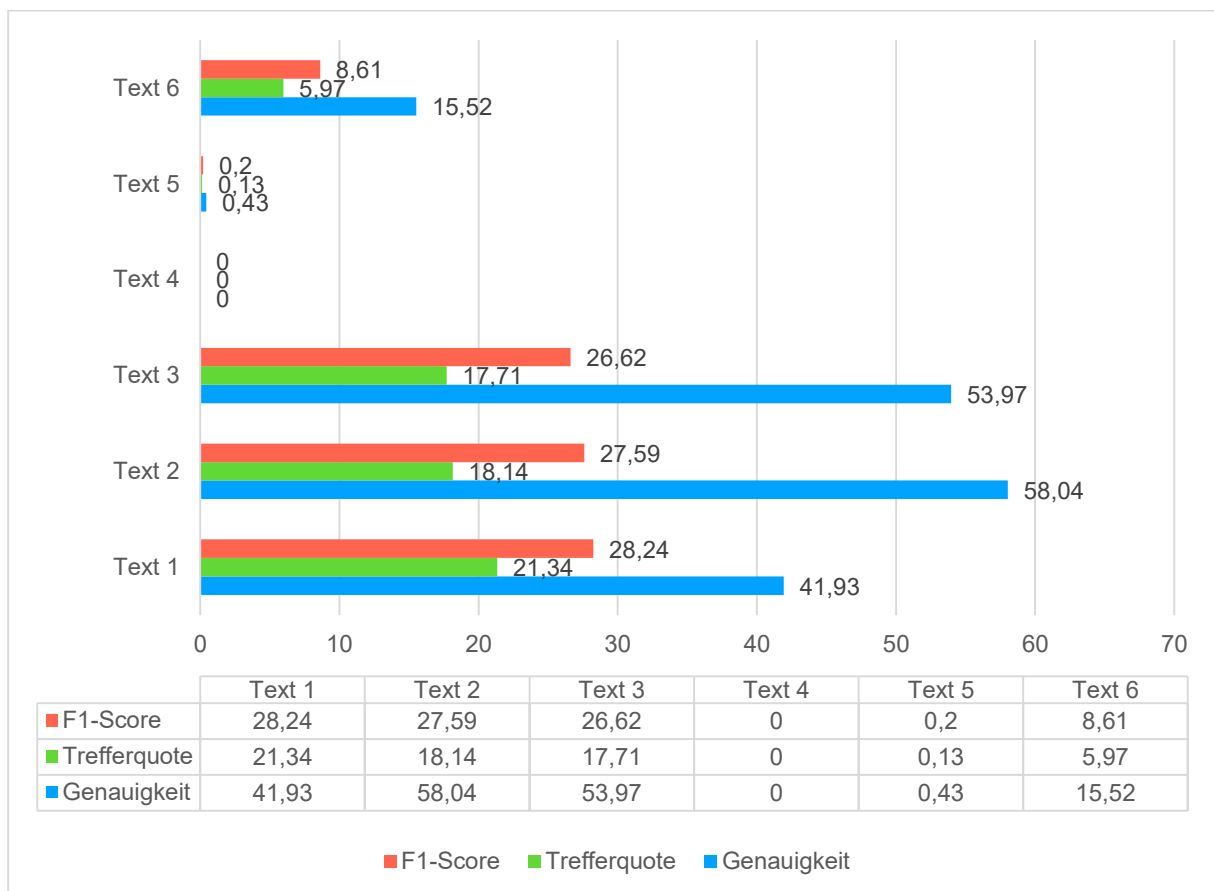


Abbildung 19: Ergebnisse von ChatGPT und Goldstandard (Durchschnittswert, nicht lemmatisiert)

ChatGPT erzielte im Durchschnitt den höchsten F1-Wert von 28,24 bei der Gewinnung der Terminologie von Text 1. Bei einzelnen Prompts erzielte Prompt 5 bei Text 3 die Höchstleistung von 32,06 F1-Score. Für alle Texte mit erzielten Resultaten generierte das Tool höhere Genauigkeits- als Trefferquotenwerte und erzielte bessere Ergebnisse bei der Extraktion österreichischer als polnischer Rechtsterminologie. Auffällig ist, dass ChatGPT verhältnismäßig wenige Termkandidaten, insbesondere im Vergleich mit Text2TCS, generierte, wie bereits oben angeführt wurde. Dies wirkt sich direkt auf den hohen Genauigkeitswert aus, allerdings verringert die niedrige Trefferquote automatisch den F1-Wert deutlich. Die Prompts haben bei gleichen Texten unterschiedliche Resultate erbracht. Bezugnehmend auf alle sechs Texte hat Prompt 2 das beste Ergebnis von 17,57 F1-Score erzielt. In Bezug auf die drei österreichischen Texte hat Prompt 3 den höchsten F1-Wert von 28,79. Für die polnischen Texte war es Prompt 3 mit 6,6 F1-Score. Der Unterschied im F1-Score zwischen den Prompts, die für den jeweiligen Text am besten und am schlechtesten abschnitten, betrugen 6,24 für Text 1, 5,96 für Text 2, 15,24 für Text 3, 20,0 für Text 6. Texte 4 und 5 generierten keine Resultate.

Im letzten Schritt wurde die Leistung beider Extraktionstools anhand eines größeren Korpus untersucht, das aus jeweils drei der o.g. österreichischen und drei der polnischen Texte bestand. Nach der Entfernung der Duplikate betrug der Goldstandard für Deutsch 409 Termkandidaten und für Polnisch 398 Termkandidaten. Der Erstannotator verzeichnete 448 und 527 Termkandidaten. Die Zweitannotatorinnen verzeichneten 292 und 488 Termkandidaten. Für ChatGPT betrug die Anzahl der Termkandidaten für Deutsch pro Prompt 171, 162, 155, 163 und 160 Termkandidaten. Für Polnisch waren es 134, 164, 143, 141 und 169. Text2TCS verzeichnete nach der Entfernung der Duplikate 563 Termkandidaten für Deutsch und 660 für Polnisch. Es ist anzumerken, dass die Termkandidaten nicht erneut aus dem Gesamtkorpus extrahiert wurden, sondern sie wurden zusammengetragen und nach der Entfernung der Duplikate evaluiert, um die Werte bei der Extraktion aus einem größeren Korpus zu ermitteln, wie in Tabelle 13 dargestellt.

Tabelle 13: Ergebnisse vom Gesamtkorpus für Text2TCS und ChatGPT (nicht lemmatisiert)

		Erstannotator			Zweitannotatorin			Goldstandard		
		G	T	F1	G	T	F1	G	T	F1
AT	Text2TCS	43,52	54,69	48,47	39,25	75,68	51,69	31,44	43,28	36,42
	ChatGPT 1	42,69	16,29	23,58	38,01	22,26	28,08	42,69	17,85	25,17
	ChatGPT 2	51,85	18,75	27,54	45,06	25,0	32,16	52,47	20,78	29,77
	ChatGPT 3	49,03	16,96	25,2	47,1	25,0	32,66	51,61	19,56	28,37
	ChatGPT 4	48,47	17,63	25,86	46,01	25,68	32,96	49,08	19,56	27,97
	ChatGPT 5	50,0	17,86	26,32	45,0	24,66	31,86	52,5	20,54	29,53
	ChatGPT Durchschnitt	48,40	17,49	25,70	44,23	24,52	31,54	49,67	19,65	28,16
PL	Text2TCS	40,61	50,85	45,16	35,61	48,16	40,94	8,33	13,82	10,39
	ChatGPT 1	7,46	1,9	3,03	6,72	1,84	2,89	2,24	0,75	1,12
	ChatGPT 2	7,32	2,28	3,48	10,37	3,48	5,21	12,2	5,03	7,12
	ChatGPT 3	6,99	1,9	2,99	10,49	3,07	4,75	13,99	5,03	7,4
	ChatGPT 4	3,55	0,95	1,5	3,55	1,02	1,58	0,71	0,25	0,37
	ChatGPT 5	4,14	1,33	2,01	2,96	1,02	1,52	0,0	0,0	0,0
	ChatGPT Durchschnitt	5,89	1,67	2,60	6,81	2,08	3,19	5,82	2,21	3,95

Während ChatGPT bei einzelnen Texten und bestimmten Prompts einen höheren F1-Wert als Text2TCS erzielte, wie nachfolgend zusammengefasst, sind sowohl der Durchschnittswert als auch die Werte einzelner Prompts anhand des größeren Korpus niedriger als die von Text2TCS erbrachte Leistung.

- Text 1: Prompt 1 (+2,8), Prompt 2 (+1,95), Prompt 3 (+1,17), Prompt 5 (+4,23)
- Text 3: Prompt 5 (+0,86)
- Text 6: Prompt 2 (+11,9), Prompt 3 (+13,0)

Zwischen der Leistung anhand der österreichischen und polnischen Terminologie lassen sich große Unterschiede erkennen. Genauso wie im Falle einzelner Texte, erzielen die Ergebnisse anhand des Gesamtkorpus pro Sprache deutlich höhere Werte für Deutsch als für Polnisch und

zwar das Dreifache bei Text2TCS und das Neunfache bei ChatGPT, wenn der Durchschnittswert aller Prompts herangezogen wird. Der Unterschied im F1-Score zwischen den Prompts, die am besten und am schlechtesten abschnitten, betraf 4,6 für Deutsch und 7,4 für Polnisch.

9.2. Quantitative Evaluierung von lemmatisierten Termkandidaten

In Tabelle 14 wurden die ermittelten Genauigkeits- und Trefferquoten sowie F1-Werte auf Basis der von Text2TCS gewonnenen Termkandidaten nach ihrer Lemmatisierung präsentiert und in Abbildung 20 visualisiert. Da die untersuchten Kategorien gleich sind, wird im Folgenden lediglich auf die Unterschiede zwischen den Datensätzen, hauptsächlich bezüglich der niedrigeren Anzahl der Termkandidaten und der erzielten Werte hingewiesen. Die Goldstandards blieben unverändert. Die Anzahl der Text2TCS-Termkandidaten beläuft sich auf 133, 202, 230, 204, 164 und 200 pro Text. Die Liste des Erstannotators besteht aus 96, 197, 169, 158, 161 und 136 Termkandidaten. Bei der Zweitannotatorin sind es 47, 124, 132, 36, 144 und 153 Termkandidaten.

Tabelle 14: Evaluierungsergebnisse von Text2TCS (lemmatisiert)

Sprache	Text	Erstannotator			Zweitannotatorin			Goldstandard		
		G	T	F1	G	T	F1	G	T	F1
AT	1	31,58	43,75	36,68	28,57	80,85	42,22	30,83	39,42	34,6
	2	62,38	63,96	63,16	49,5	80,65	61,35	55,45	54,63	55,04
	3	39,57	53,85	45,62	43,04	75,0	54,69	31,3	41,14	35,55
PL	4	41,67	53,8	46,96	29,41	44,12	35,29	20,1	26,8	22,97
	5	43,9	44,72	44,31	35,98	40,97	38,31	21,34	22,73	22,01
	6	35,0	51,47	41,67	41,0	53,59	46,46	20,5	28,47	23,84

Abbildung 20 veranschaulicht die erzielten Werte von Text2TCS und Goldstandard.

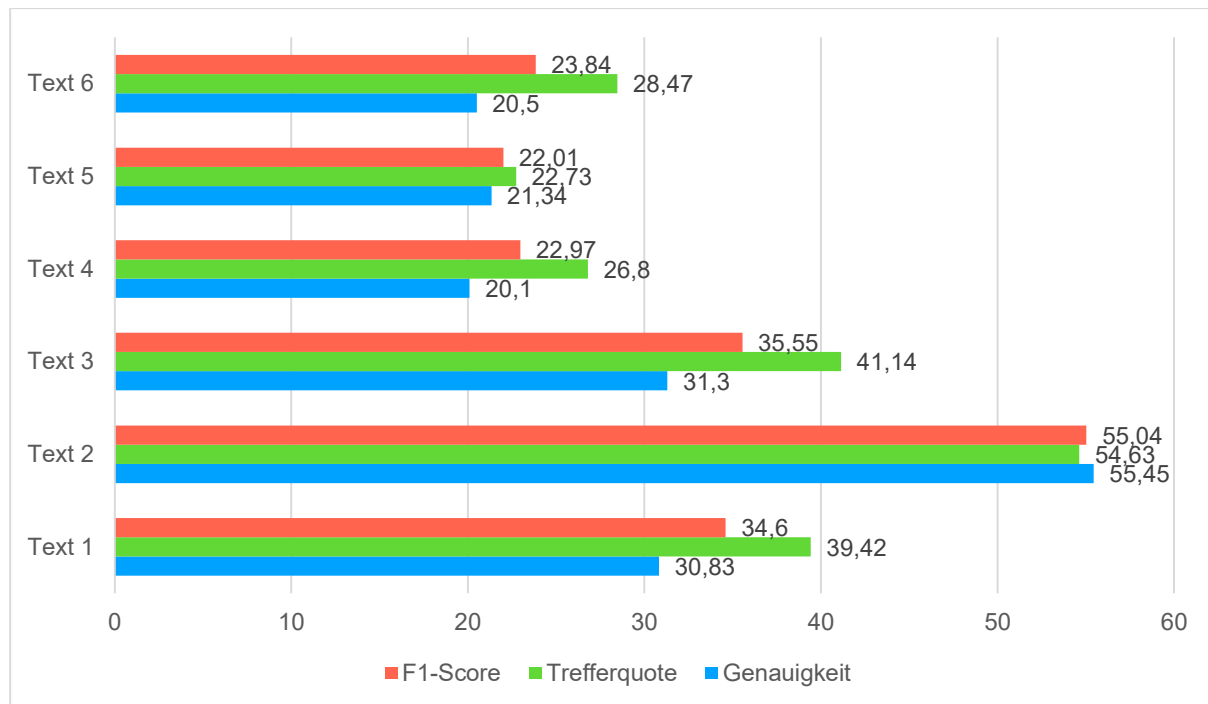


Abbildung 20: Ergebnisse von Text2TCS und Goldstandard (lemmatisiert)

Text2TCS erzielte den höchsten F1-Wert von 55,04 erneut bei Text 2, bei dem alle drei Werte sehr ausbalanciert waren. Für alle anderen Texte generierte das Tool höhere Trefferquoten als Genauigkeitswerte und erzielte bessere Ergebnisse bei der Extraktion österreichischer als polnischer Rechtsterminologie. Die Lemmatisierung verbesserte den F1-Score von Text2TCS bei allen Texten.

Tabellen 15-19 zeigen die Genauigkeits- und Trefferquoten sowie F1-Werte der Leistung von ChatGPT nach der Lemmatisierung. Das Balkendiagramm in Abbildung 21 stellt den Durchschnittswert von allen fünf Prompts pro Text dar. Die nach der Entfernung der Duplikate der Analyse unterzogenen Termkandidaten, die von ChatGPT generiert wurden, betrugen für Text 1: 58, 55, 46, 49 und 57 Termkandidaten. Für Text 2: 74, 66, 66, 62 und 53 Termkandidaten. Für Text 3: 51, 52, 55, 66 und 62 Termkandidaten. Für Text 4: 49, 62, 51, 57 und 70 Termkandidaten. Für Text 5: 43, 46, 37, 37 und 62 Termkandidaten. Für Text 6: 46, 57, 56, 49 und 40 Termkandidaten.

Tabelle 15: Ergebnisse von ChatGPT Prompt 1 (lemmatisiert)

Prompt	Text	Erstannotator			Zweitannotatorin			Goldstandard		
		G	T	F1	G	T	F1	G	T	F1
1	1	34,48	20,83	25,97	32,76	40,43	36,19	34,48	19,23	24,69
	2	58,11	21,83	31,74	51,35	30,65	38,39	45,95	16,59	24,38
	3	43,14	13,02	20,0	43,14	16,67	24,05	33,33	9,71	15,04
	4	42,86	13,29	20,29	38,78	13,97	20,54	18,37	5,88	8,91
	5	44,19	11,8	18,63	53,49	15,97	24,6	20,93	5,84	9,13
	6	26,09	8,82	13,18	45,65	13,73	21,11	13,04	4,17	6,32

Tabelle 16: Ergebnisse von ChatGPT Prompt 2 (lemmatisiert)

Prompt	Text	Erstannotator			Zweitannotatorin			Goldstandard		
		G	T	F1	G	T	F1	G	T	F1
2	1	34,55	19,79	25,17	32,73	38,3	35,3	29,09	15,38	20,12
	2	68,18	22,84	34,22	62,12	33,06	43,15	54,55	17,56	26,57
	3	63,46	19,53	29,87	53,85	21,21	30,43	44,23	13,14	20,26
	4	45,16	17,72	25,45	38,71	17,65	24,25	19,35	7,84	11,16
	5	32,61	9,32	14,5	47,83	15,28	23,16	17,39	5,19	7,99
	6	28,57	11,76	16,66	50,0	18,3	26,79	12,5	4,86	7,0

Tabelle 17: Ergebnisse von ChatGPT Prompt 3 (lemmatisiert)

Prompt	Text	Erstannotator			Zweitannotatorin			Goldstandard		
		G	T	F1	G	T	F1	G	T	F1
3	1	41,3	19,79	26,76	41,3	40,43	40,86	39,13	17,31	24,0
	2	63,08	20,81	31,3	60,0	31,45	41,27	53,85	17,07	25,92
	3	61,82	20,12	30,36	63,64	26,52	37,44	47,27	14,86	22,61
	4	33,33	10,76	16,27	29,41	11,03	16,04	15,69	5,23	7,84
	5	29,73	6,83	11,11	45,95	11,81	18,79	10,81	2,6	4,19
	6	31,48	12,5	17,89	50,0	17,65	26,09	9,26	3,47	5,05

Tabelle 18: Ergebnisse von ChatGPT Prompt 4 (lemmatisiert)

Prompt	Text	Erstannotator			Zweitannotatorin			Goldstandard		
		G	T	F1	G	T	F1	G	T	F1
4	1	36,73	18,75	24,83	34,69	36,17	35,41	34,69	16,35	22,22
	2	70,97	22,34	33,98	70,97	35,48	47,31	61,29	18,54	28,47
	3	56,06	21,89	31,49	51,52	25,76	34,35	36,36	13,71	19,91
	4	40,35	14,56	21,4	38,6	16,18	22,8	19,3	7,19	10,48
	5	27,03	6,21	10,1	40,54	10,42	16,58	13,51	3,25	5,24
	6	26,53	9,56	14,06	48,98	15,69	23,77	12,24	4,17	6,22

Tabelle 19: Ergebnisse von ChatGPT Prompt 5 (lemmatisiert)

Prompt	Text	Erstannotator			Zweitannotatorin			Goldstandard		
		G	T	F1	G	T	F1	G	T	F1
5	1	35,09	20,83	26,14	31,58	38,3	34,62	31,58	17,31	22,36
	2	66,04	17,77	28,0	62,26	26,61	37,28	54,72	14,15	22,49
	3	64,52	23,67	34,63	58,06	27,27	37,11	48,39	17,14	25,31
	4	25,71	11,39	15,79	22,86	11,76	15,53	10,0	4,58	6,28
	5	22,58	8,7	12,56	33,87	14,58	20,38	4,84	1,95	2,78
	6	22,5	6,62	10,23	32,5	8,5	13,48	7,5	2,08	3,26

Abbildung 21 zeigt die erzielten Durchschnittswerte von ChatGPT und Goldstandard.

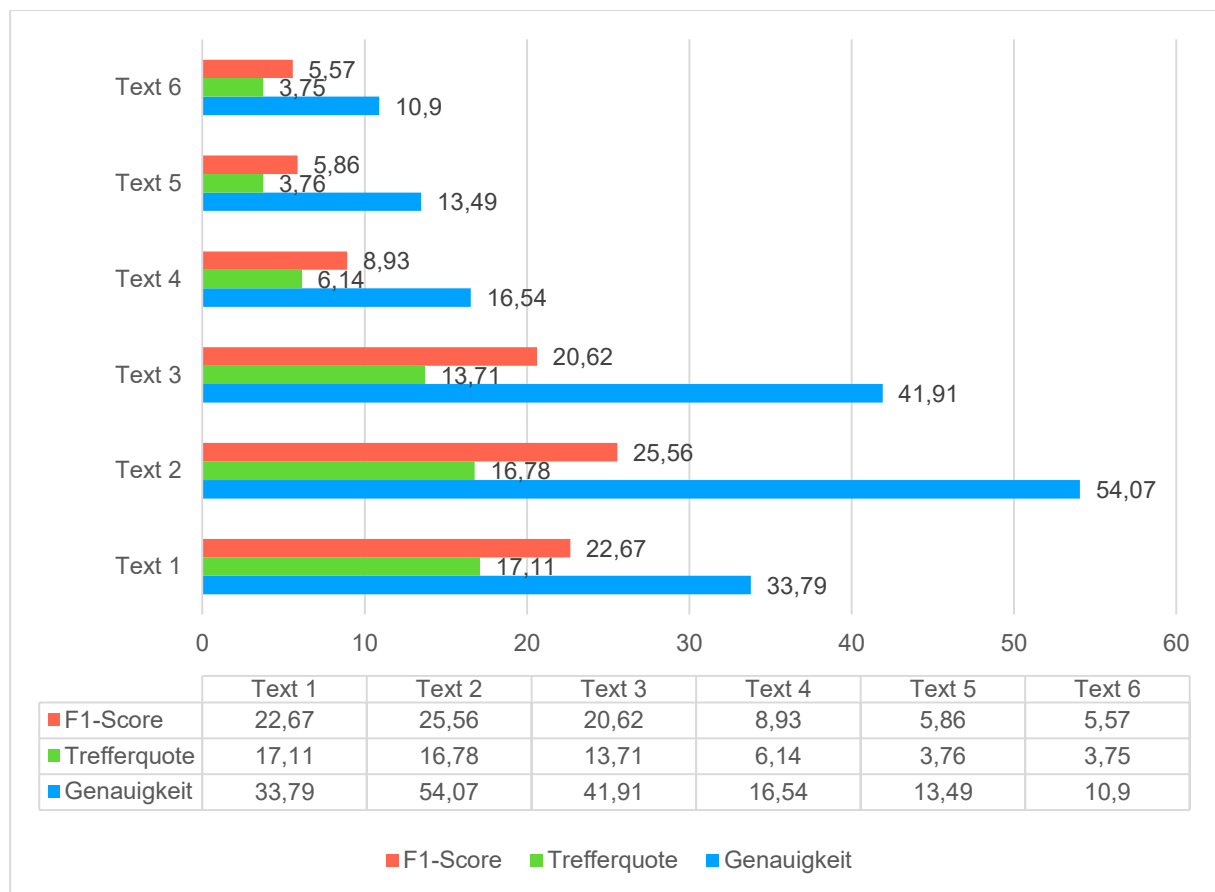


Abbildung 21: Ergebnisse von ChatGPT und Goldstandard (Durchschnittswert, lemmatisiert)

ChatGPT erzielte den höchsten F1-Wert von 25,56 im Durchschnitt bei Text 2. Bei einzelnen Prompts erzielte Prompt 4 bei Text 2 die Höchstleistung von 28,47 F1-Score. Für alle Texte generierte das Tool höhere Genauigkeits- als Trefferquotenwerte und erzielte bessere Ergebnisse bei der Extraktion österreichischer als polnischer Rechtsterminologie. Die Prompts haben bei gleichen Texten unterschiedliche Resultate erbracht. Bezugnehmend auf alle sechs Texte hat Prompt 2 das beste Ergebnis von 15,51 F1-Score erzielt. In Bezug auf die drei österreichischen Texte hat Prompt 3 den höchsten F1-Wert von 24,17. Für die polnischen Texte war es Prompt 2 mit 8,71 F1-Score. Der Unterschied im F1-Score zwischen den Prompts, die für den jeweiligen Text am besten und am schlechtesten abschnitten, betrug 4,57 für Text 1, 5,98 für Text 2, 10,27 für Text 3, 4,88 für Text 4, 6,35 für Text 5 und 3,74 für Text 6.

Im letzten Schritt wurde die Leistung beider Extraktionstools, wie in Tabelle 20 dargestellt, anhand eines größeren Korpus nach Lemmatisierung untersucht, das aus jeweils drei der o.g. österreichischen und drei der polnischen Texte bestand. Der Goldstandard betrug unverändert für Deutsch 409 Termkandidaten und für Polnisch 398 Termkandidaten. Der Erstannotator verzeichnete 400 und 391 Termkandidaten. Die Zweitannotatorinnen verzeichneten 264 und 404 Termkandidaten. Für ChatGPT betrug die Anzahl der Termkandidaten für Deutsch pro

Prompt 171, 162, 154, 163 und 159 Termkandidaten. Für Polnisch waren es 134, 161, 139, 141 und 168. Text2TCS verzeichnete 510 Termkandidaten für Deutsch und 511 für Polnisch.

Tabelle 20: Ergebnisse vom Gesamtkorpus für Text2TCS und ChatGPT (lemmatisiert)

		Erstannotator			Zweitannotatorin			Goldstandard		
		G	T	F1	G	T	F1	G	T	F1
AT	Text2TCS	44,31	56,5	49,67	40,0	77,27	52,71	38,63	48,17	42,88
	ChatGPT 1	44,44	19,0	26,62	42,11	27,27	33,1	36,84	15,4	21,72
	ChatGPT 2	54,32	22,0	31,32	49,38	30,3	37,56	41,98	16,63	23,82
	ChatGPT 3	54,55	21,0	30,33	53,9	31,44	39,71	45,45	17,11	24,86
	ChatGPT 4	52,76	21,5	30,55	50,92	31,44	38,88	41,72	16,63	23,78
	ChatGPT 5	54,72	21,75	31,13	49,69	29,92	37,35	42,77	16,63	23,95
	ChatGPT Durchschnitt	52,15	21,05	29,99	49,2	30,07	37,32	41,75	16,48	23,62
PL	Text2TCS	38,75	50,64	43,9	36,01	45,54	40,22	18,98	24,37	21,34
	ChatGPT 1	38,06	13,04	19,42	47,76	15,84	23,79	17,16	5,78	8,65
	ChatGPT 2	36,65	15,09	21,38	45,96	18,32	26,2	16,77	6,78	9,66
	ChatGPT 3	31,65	11,25	16,6	42,45	14,6	21,73	12,95	4,52	6,7
	ChatGPT 4	32,62	11,76	17,29	43,26	15,1	22,39	15,6	5,53	8,17
	ChatGPT 5	23,21	9,97	13,95	29,17	12,13	17,13	8,33	3,52	4,95
	ChatGPT Durchschnitt	32,43	12,22	17,72	41,72	15,19	22,24	14,16	5,22	7,62

Auch nach der Lemmatisierung blieb die Tendenz bezüglich der Leistung beider Extraktions-tools gleich. Text2TCS schnitt in beiden Sprachen besser beim F1-Score ab und erzielte bei verschiedenen Prompts unterschiedliche Werte, wenngleich der Unterschied zwischen dem höchsten und niedrigsten F1-Wert bei 3,14 für Deutsch und 4,71 für Polnisch lag. Auffällig ist die Steigerung der Werte von Text2TCS um die Hälfte für Polnisch (+10,95). Auch ChatGPT erzielte im Durchschnitt höhere Werte nach der Lemmatisierung für Polnisch (+3,67). Interessanterweise sank der Durchschnittswert von ChatGPT für Deutsch nach der Lemmatisierung

um 4,54. Der Unterschied im F1-Score zwischen den Prompts, die am besten und am schlechtesten abschnitten, betraf 3,14 für Deutsch und 4,71 für Polnisch.

In Tabelle 21 wurden die Unterschiede in der Leistung von Text2TCS und ChatGPT nach der Lemmatisierung für einzelne Texte aufgeschlüsselt.

Tabelle 21: Unterschiede nach der Lemmatisierung für Text2TCS und ChatGPT

			Text2TCS	GPT1	GPT2	GPT3	GPT4	GPT5	GPT Durchschnitt
Goldstandard	Text 1	G	+7,59	-6,9	-10,91	-6,52	-4,09	-12,28	-8,14
		T	+7,69	-3,85	-5,77	-2,88	-1,92	-6,73	-4,23
		F1	+7,77	-4,94	-7,55	-4	-2,62	-8,7	-5,57
	Text 2	G	+9,62	-8,1	-4,54	-3,73	-1,61	-1,88	-3,97
		T	+6,34	-2,92	-1,46	-1,47	-0,48	-0,48	-1,36
		F1	+8,01	-4,29	-2,21	-2,13	-0,74	-0,76	-2,03
	Text 3	G	+4,69	-3,92	-19,23	-9,09	-15,16	-12,9	-12,06
		T	+3,43	-1,15	-5,72	-2,85	-5,72	-4,57	-4
		F1	+4,35	-1,78	-8,82	-4,34	-8,31	-6,75	-6
	Text 4	G	+13,7	+18,37	+19,35	+15,69	+19,3	+10	+16,54
		T	+16,34	+5,88	+7,84	+5,23	+7,19	+4,58	+6,14
		F1	+15,03	+8,91	+11,16	+7,84	+10,48	+6,28	+8,93
	Text 5	G	+8,58	+20,93	+15,22	+10,81	+13,51	+4,84	+13,06
		T	+6,5	+5,84	+4,54	+2,6	+3,25	+1,95	+3,63
		F1	+7,72	+9,13	+6,99	+4,19	+5,24	+2,78	+5,66
	Text 6	G	+15,03	+6,52	-20,83	-26,45	+10,2	+7,5	-4,62
		T	+18,75	+2,09	-8,33	-10,42	+3,48	+2,08	-2,22
		F1	+16,84	+3,17	-11,9	-14,95	+5,19	+3,26	-3,04

Text2TCS verzeichnete die größte Steigerung bei F1-Score von Text 6 (+16,84). ChatGPT wies die größte Steigerung bei Prompt 2 in Text 4 (+11,16) und die größte Senkung bei Prompt 3 in Text 6 (-14,95) auf.

Zusammenfassend ergab die vorliegende Untersuchung, welche Genauigkeits- und Trefferquoten sowie F1-Werte Text2TCS und ChatGPT bei der Gewinnung von österreichischer und polnischer Rechtsterminologie aus dem Fachbereich des Urheberrechts anhand drei Entscheidungstexte pro Sprache erreichten. Text2TCS erzielte vor der Lemmatisierung bei allen Messungen höhere Trefferquoten als Genauigkeitswerte sowohl bei einzelnen Texten als auch beim Gesamtkorpus. Nach der Lemmatisierung lag nur ein Fall vor, bei dem der Trefferquotenwert niedriger als der Genauigkeitswert war und zwar bei Text 2. Sonst lagen ebenfalls die Trefferquotenwerte höher. Der höchste erzielte F1-Score lag vor der Lemmatisierung bei 47,03 und nach der Lemmatisierung bei 55,04 und betraf Text 2. Die Höchstwerte für die polnischen Texte beliefen sich auf 14,29 vor der Lemmatisierung für Text 5 und auf 23,84 für Text 6 danach. Anhand des größeren Korpus erzielte Text2TC 36,42 F1-Score für Deutsch und 10,39 für Polnisch vor der Lemmatisierung und 42,88 sowie 21,34 danach. ChatGPT generierte bei jeder Messung, die Ergebnisse brachte, höhere Genauigkeits- als Trefferquotenwerte. Im Durchschnitt erzielte ChatGPT vor der Lemmatisierung den höchsten F1-Wert von 28,24 für Text 1 und für die polnischen Texte 8,61 für Text 6. Nach der Lemmatisierung lag der höchste F1-Score im Durchschnitt bei 25,56 für Text 2 und für die polnischen Texte 8,93 für Text 4. Anhand des Gesamtkorpus erzielte ChatGPT im Durchschnitt vor der Lemmatisierung 28,16 F1-Score für Deutsch und 3,95 für Polnisch und nach der Lemmatisierung 23,62 für Deutsch und 7,62 für Polnisch. Da die Leistung der Prompts im Rahmen der verwendeten Zero-Shot-Strategie bei einzelnen Texten stark variierte, und keiner der Prompts konstant bei jedem Text die besten Resultate erzielte, lässt sich nicht eindeutig konstatieren, welcher Prompt am besten abgeschnitten hat. Bei den einzelnen Prompts und einzelnen Texten waren es Prompt 5 für Text 3 mit F1-Score 32,06 bei nicht lemmatisierten Texten und Prompt 4 für Text 2 mit F1-Score 28,47 nach der Lemmatisierung. Bei den durchschnittlichen Werten lassen sich anhand aller 6 Texte, 3 Texte pro Sprache und der Gesamtkorpora folgende Ergebnisse feststellen, die Tabelle 22 veranschaulicht.

Tabelle 22: Die besten Ergebnisse unter Prompts anhand der Durchschnittswerte

	Vor der Lemmatisierung	Nach der Lemmatisierung
Texte 1-6	Prompt 2 F1 17,57	Prompt 2 F1 15,51
Texte 1-3 AT	Prompt 3 F1 28,79	Prompt 3 F1 24,17
Texte 4-6 PL	Prompt 3 F1 6,6	Prompt 2 F1 8,71
Gesamtkorpus AT	Prompt 2 F1 29,77	Prompt 3 F1 24,17

Gesamtkorpus PL	Prompt 3 F1 7,4	Prompt 2 F1 9,66
Gesamtkorpora AT + PL	Prompt 2 F1 18,44	Prompt 2 F1 16,74

Anhand der Durchschnittswerte erzielte Prompt 2 die besten Resultate gefolgt von Prompt 3. Während Prompt 2 sehr direkt formuliert wurde, wurde ChatGPT im Prompt 3 gesiezt und höflich um die Termkandidatenliste gebeten. Da Prompt 1 nach dem gleichen Prinzip wie Prompt 2 und Prompt 4 ähnlich wie Prompt 3 formuliert wurden und diese Prompts in den Durchschnittswerten schlechter abgeschnitten haben, lässt sich der Einfluss der Höflichkeit auf die erbrachte Leistung anhand der erhobenen Daten nicht erkennen. Dies gilt ebenfalls für die Knappheit respektive Ausführlichkeit der Formulierung von Prompts. Prompt 2 wurde knapp formuliert und erzielte die besten Resultate, aber gleichzeitig wurde Prompt 3 ausführlicher formuliert und seine Leistung übertraf die Ergebnisse von dem ebenfalls knapp formulierten Prompt 1. Prompt 5, der über die längste Formulierung verfügt, hat generell schlechter abgeschnitten, aber gleichzeitig erreichte er einzeln bei Text 3 die Höchstleistung unter allen Messungen vor der Lemmatisierung. Was Prompt 2 und Prompt 3 verbindet, ist die Tatsache, dass ChatGPT-4o aufgefordert wurde, Fachtermini, im Gegensatz zu Fachbegriffen wie im Prompt 1 und Fachwörtern wie im Prompt 4, zu extrahieren. Dies gilt allerdings auch für Prompt 5, der wiederum schlechter abgeschnitten hat. Im Polnischen lässt sich dieser Zusammenhang nicht feststellen, da im Prompt 2 ChatGPT terminy specjalistyczne und im Prompt 3 terminy fachowe zu extrahieren hatte. Die Analyse der Prompts, die am schlechtesten abgeschnitten haben, lässt in Bezug auf die o.g. Punkte ebenfalls keine eindeutigen Schlüsse ziehen, denn das schlechteste Resultat vor der Lemmatisierung erzielte in Bezug auf alle Texte der lang und höflich formulierte Prompt 5 mit F1 13,74 und nach der Lemmatisierung der knapp und direkt formulierte Prompt 1 mit F1 13,04. Anhand der Ergebnisse der Gesamtkorpora lässt sich außerdem eine weitere Tendenz beobachten. Die Leistungsschwankung zwischen dem Prompt, der für das gesamte österreichische respektive polnische Korpus die beste Leistung und jenem, der die schlechteste Leistung erzielte ist sowohl vor als auch nach der Lemmatisierung für Polnisch höher als für Deutsch und beträgt für das österreichische Korpus jeweils 4,6 / 3,14 und für das polnische Korpus 7,12 / 4,71 in Bezug auf F1-Score.

9.3. Qualitative Analyse der automatisch generierten Termkandidaten

Die manuelle Detailanalyse bezieht sich z.T. auf die im theoretischen Teil gewonnenen Erkenntnisse und teilweise auf die wiederkehrenden Muster und Auffälligkeiten der extrahierten Terminologie.

Wie im Kapitel zu den lexikalischen Besonderheiten der Rechtssprache erwähnt (5.2.2.1.), stellen die für gewöhnlich als gemeinsprachlich bezeichneten Ausdrücke, denen allerdings in Rechtstexten eine fachliche Bedeutung zugewiesen wird, eine Herausforderung sowohl für die humane als auch die automatische Termextraktion dar. In den untersuchten Texten kommen mehrere solche Ausdrücke vor, die in den manuell erstellten Goldstandard aufgenommen wurden, z. B.:

- Die Vorinstanzen gaben der Klage statt.
- Die Beklagte bekämpft die rechtliche Beurteilung des Berufungsgerichts [...]
- Im allseitigen Interesse sollen aber lange Schwebezustände durch den Rückruf vermieden werden [...]
- [...] und die ordentliche Revision erklärte es für nicht zulässig.

Text2TCS erkannte zum Teil diese Ausdrücke und sie wurden der Termkandidatenliste hinzugefügt. Für Deutsch waren es u.a. *Erklärung, Rückruf, Nutzung, erklären* und für Polnisch *odpowiedzialność, przedmiot, materiał, strona*. Gleichzeitig wurden viele solcher Ausdrücke von Text2TCS nicht erkannt wie etwa *stattgeben, überholen, betonen, Gegenstand, aussprechen, bekämpfen, abweisen, wniesć, podnosić, opierać się, wyprowadzać*. ChatGPT generierte bei den von Text2TCS erkannten Einwortbenennungen unterschiedliche, zum größten Teil negative Ergebnisse wie in Tabelle 23 dargestellt.

Tabelle 23: ChatGPT: Erkennung gemeinsprachlicher Ausdrücke mit fachlicher Bedeutung

	Prompt 1	Prompt 2	Prompt 3	Prompt 4	Prompt 5
Erklärung	ja	nein	nein	nein	nein
Rückruf	ja	ja	ja	nein	ja
Nutzung	nein	nein	nein	nein	nein
Erklären	nein	nein	nein	nein	nein
Odpowiedzialność	nein	nein	nein	nein	nein
Przedmiot	nein	nein	nein	nein	nein
Materiał	nein	nein	nein	nein	nein
Strona	nein	nein	nein	nein	nein

Es ist allerdings anzumerken, dass einige dieser Einwortbenennungen als Bestandteile von Mehrwortbenennungen extrahiert wurden, was aber im Goldstandard einen separaten Eintrag hat, falls es sich um einen Fachausdruck handelt. Alle von Text2TCS nicht erkannten Termini wurden von ChatGPT ebenfalls nicht erkannt.

Aus dem theoretischen Teil der vorliegenden Arbeit (2.3.2.1.) geht außerdem hervor, dass die automatische Gewinnung von Mehrwortbenennungen problematisch ist, denn die Unithood einer Mehrwortbenennung nicht immer richtig erkannt wird. Infolgedessen werden solche Fachausdrücke von dem Extraktionstool zerlegt und nur Teile davon werden ggf. in die Termkandidatenliste aufgenommen. In den untersuchten Rechtstexten liegen Mehrwortbenennungen vor, die in den Goldstandard aufgenommen wurden. Darunter befinden sich jene, die gänzlich von Text2TCS extrahiert wurden, wie *abstrakte Rechtsfrage*, *Oberster Gerichtshof*, *Co Verlagsverhältnis*, *geistige Schöpfung*; jene, die zerlegt und als Einwortbenennungen gänzlich in die Termkandidatenliste aufgenommen wurden, wie *außerordentliche Revision*, *ordentliche Revision*; Mehrwortbenennungen, deren Teile als Einwortbenennungen oder kürzere Mehrwortbenennungen extrahiert wurden und Teile nicht, z. B. *erhebliche Rechtsfrage*, *Folge geben*, *Richtlinie 2001/29/EG des Europäischen Parlaments und des Rates vom 22. Mai 2001 zur Harmonisierung bestimmter Aspekte des Urheberrechts und der verwandten Schutzrechte in der Informationsgesellschaft* und jene, die gänzlich nicht erkannt wurden, wie *Pariser Fassung*, *in eventu*. Für Polnisch sind es jeweils *protokół stanu faktycznego*, *stawka minimalna*, *pozwolenie na użytkowanie*, *świadczenie pieniężne*, *umowa licencyjna*; *obowiązek zwrotu*, *moc dowodu*; *ulec przedawnieniu*, *zgodnie z prawem*, *opiewać na kwotę*, *dyrektywa 2004/48/WE Parlamentu Europejskiego i Rady z dnia 29 kwietnia 2004 r. w sprawie egzekwowania praw własności intelektualnej*; *ex tunc*, *instytucja pożytku publicznego*. ChatGPT erbrachte unterschiedliche Resultate mit verschiedenen Prompts. Beispielsweise wurde die Mehrwortbenennung *außerordentliche Revision*, deren Unithood von Text2TCS weder im Text 1 noch im Text 2 erkannt wurde, von ChatGPT im Text 1 mit allen Prompts generiert und im Text 2 mit Prompt 1 und 2. *Abstrakte Rechtsfrage*, die Text2TCS richtig erkannte, wurde von ChatGPT nur mit Prompt 1 und 4 gewonnen. *Ex tunc*, die von Text2TCS nicht generiert wurde, wurde mit Prompt 2 und 4 gewonnen.

Ein weiteres Muster lässt sich außerdem bei der Extraktion unterschiedlicher Wortarten feststellen. In die Goldstandards wurden unter Einwortbenennungen neben den Substantiven, die die größte Gruppe der Termini bilden, auch Verben und Adjektive/Partizipien aufgenommen. ChatGPT extrahierte in allen Texten beider Sprachen ausschließlich Substantive als Einwortbenennungen. Adjektive/Partizipien kamen lediglich dann vor, wenn sie Teile von Mehrwortbenennungen waren. Text2TCS hat sowohl Verben, wie *vervielfältigen*, *untersagen*, *erklären*, *widerrufen*, *zurückweisen*, *zasądzić*, als auch Adjektive/Partizipien, wie *unentgeltlich*, *beurteilend*, *gewerblich*, *ausschließlich*, *vertraglich*, *rechtskräftig*, *uiszczony*, *nienależyty*,

uprawniony, wiarygodny, udowodniony, wszechstronny als Einwortbenennungen extrahiert, wenngleich ihre Anzahl im Vergleich mit dem Goldstandard trotzdem sehr limitiert ist.

Weiters enthalten Goldstandards Kurzwörter und Eigennamen, die ebenfalls als Fachausdrücke gelten. Tabelle 24 erteilt anhand einer überschaubaren Auswahl Auskunft darüber, wie beide Extraktionstools bei Kurzwörtern und Eigennamen abschnitten. Es handelt sich um Ein- und Mehrwortbenennungen, in der Form, in der sie im Goldstandard zu finden sind.

Tabelle 24: Extraktion von Kurzwörtern durch Text2TCS und ChatGPT

	Text2TCS	ChatGPT Prompt 1	ChatGPT Prompt 2	ChatGPT Prompt 3	ChatGPT Prompt 4	ChatGPT Prompt 5
BGH	ja	ja	ja	ja	ja	ja
EuGH	ja	ja	ja	ja	ja	nein
iSd	nein	nein	nein	nein	nein	nein
Pariser Fassung	nein	nein	ja	ja	ja	ja
InfoRL	ja	nein	nein	nein	ja	ja
§ 42c UrhG	nein	ja	ja	ja	nein	ja
UWG	ja	ja	ja	ja	ja	ja
4 Ob 113/09k	nein	nein	nein	nein	nein	nein
RS0076187	nein	nein	nein	nein	nein	nein
Główny Urząd Nadzoru Budowlanego	nein	ja	ja	ja	ja	ja
k.p.c	nein	ja	ja	nein	ja	ja
LEX	ja	nein	ja	nein	nein	nein
ww.	nein	nein	nein	nein	nein	nein
OSNC 2006/5/92	nein	nein	nein	nein	nein	nein

10. Diskussion

Das folgende Kapitel lässt sich in drei Teile untergliedern. Zunächst werden die Ergebnisse der durchgeführten Untersuchung im Hinblick auf die Forschungsfragen interpretiert. Weiters werden Herausforderungen und Beschränkungen der Untersuchung beschrieben. Das Kapitel endet mit den Empfehlungen für die weitere Forschung.

10.1. Interpretation der Ergebnisse

Text2TCS erzielte in der vorliegenden Untersuchung bei der Termextraktion zwar höhere Werte als ChatGPT, aber gleichzeitig niedrigere Werte als in den anderen Studien. Dies ist auf verschiedene Faktoren zurückzuführen, die in dieser Untersuchung eine Rolle spielten. Zum einen wies Text2TCS grundsätzlich relativ hohe Trefferquotenwerte bei gleichzeitig niedrigeren Genauigkeitswerten auf, was in der Forschung als Noise-Effekt bekannt ist. Wie im vorherigen Kapitel angeführten Beispiele zeigen, extrahierte Text2TCS Ausdrücke, die eindeutig der Gemeinsprache zuzuordnen sind und keinen terminologischen Wert aufweisen. Weiters wurden Fachausdrücke gewonnen, die Bestandteile von Mehrwortbenennungen waren und nicht einzeln als Einwortbenennungen im Goldstandard enthalten waren. Dies ist auf die mangelnde Erkennung von der Unithood der jeweiligen Mehrwortbenennung zurückzuführen. Gleichzeitig lässt sich ebenfalls der sog. Silence-Effekt feststellen. Text2TCS ignorierte zum großen Teil bei der Extraktion Einwortbenennungen, die nicht zu der Wortart der Substantive gehörten, in dem konkreten Fall waren es Verben und Adjektive/Partizipien. Der Auseinandersetzung mit der Fachliteratur ist zu entnehmen, dass die meisten Termini Substantive sind und daraus lässt sich herleiten, dass die Extraktionstools höchstwahrscheinlich deswegen andere Wortarten weniger im Fokus haben. Die Rechtssprache stellt diesbezüglich allerdings eine zusätzliche Herausforderung dar, da sie einige allgemeine Ausdrücke terminologisiert, wie im theoretischen Teil beschrieben und anhand konkreter Beispiele im vorherigen Teil veranschaulicht wurde. Es handelt sich dabei trotzdem um Fachausdrücke, die zu gewinnen sind, denn sie bilden dennoch die rechtliche Terminologie und sind als alle anderen Termini zu behandeln. Bei der manuellen Terminologierecherche lassen sich solche Ausdrücke mithilfe des Kontexts feststellen und von Terminolog:innen mit ausreichendem Wissen erkennen. Die mangelnde bzw. instabile Trefferquote bei der Gewinnung von Namen und Kurzwörtern ist wahrscheinlich darauf zurückzuführen, wie oft sie in den Trainingsdaten vorgekommen sind. Wie die Beispiele zeigen, schnitt Text2TCS bei den gängigen Abkürzungen besser ab. Die verfügbaren Trainingsdaten, insbesondere ihre Menge, könnte den Unterschied zwischen den Ergebnissen der beiden Sprachen

erklären, insbesondere da die in Text2TCS verwendeten Modelle zwar auf Polnisch trainiert wurden, aber nicht explizit für die Termextraktion auf Polnisch trainiert wurden. Die durchgeführte Untersuchung belegte außerdem den positiven Einfluss der Lemmatisierung auf die erzielten Werte. Insbesondere stiegen die Genauigkeits- und Trefferquoten sowie F1-Werte der polnischen Texte, was möglicherweise auf den flexionsreichen Charakter der polnischen Sprache zurückzuführen ist.

Die ermittelten Werte der Termextraktion von ChatGPT lagen in der vorliegenden Untersuchung ebenfalls unter den Forschungsergebnissen anderer Forschenden. Dies liegt daran, dass auch ChatGPT mit dem Noise- und Silence-Effekt konfrontiert wurde. In dem Fall waren allerdings trotz der relativ hohen Genauigkeitswerte die sehr niedrigen Trefferquotenwerte für den niedrigeren F1-Score ausschlaggebend. Trotz mehrerer Extraktionsversuche mit unterschiedlich formulierten Prompts, extrahierte ChatGPT eine sehr konstante, im Vergleich mit Text2TCS und dem jeweiligen Goldstandard niedrige Anzahl von Termkandidaten. Viele der bei der Interpretation der Leistung von Text2TCS genannten Faktoren treffen auch auf ChatGPT zu, darunter die Verfügbarkeit von Trainingsdaten, mangelnde Erkennung der Unithood von Mehrwortbenennungen, die ausschließliche Extraktion von Substantiven im Falle von Einwortbenennungen sowie fehlende Erkennung der terminologisierten Ausdrücke. Einen Unterschied bilden die Kurzwörter und Eigennamen, bei deren Extraktion ChatGPT anhand der genannten Beispiele etwas bessere Ergebnisse als Text2TCS erzielte. Die Lemmatisierung wirkte sich positiv auf die Extraktion polnischsprachiger Terminologie aus, was der o.g. Interpretation entspricht, aber die Resultate der österreichischen Terminologie wiederum sinken ließ. Die niedrige Anzahl der extrahierten Termkandidaten könnte daran liegen, dass ChatGPT nicht explizit auf diese Aufgabe spezialisiert ist. Dies lässt sich daraus herleiten, dass Termextraktion in der Forschungsliteratur nicht unter den primären Einsatzbereichen von ChatGPT zu finden ist. Eine weitere logische Interpretation betrifft die Formulierung von Prompts. Die Analyse der Leistung der untersuchten Prompts ergab zwar, dass ChatGPT-4o bei verschiedenen Anfragen unterschiedliche Termkandidatenlisten generierte, aber die Inkonsistenz ihrer Leistung lässt aus den erhobenen Daten keine eindeutigen Schlüsse darüber ziehen, welche Formulierung bei der verwendeten Zero-Shot-Strategie am erfolgreichsten ist. Eine auch im Rahmen der vorliegenden Arbeit statistisch wenig signifikante Tendenz, die auf Durchschnittswerten basiert, lässt annehmen, dass Anfragen, die kurz und deutlich formuliert sind und den Ausdruck Fachterminus bzw. Fachtermini beinhalten, die besten Resultate bei der automatischen Termextraktion mit ChatGPT-4o erzielen.

Die niedrigere Erfolgsquote beider Extraktionstools kann schließlich auf den Faktor Mensch zurückgeführt werden. Die Ergebnisse der Evaluierung hingen nämlich stark mit dem manuell erstellten Goldstandard zusammen. Bei der humanen Annotation sind stets mehrere Faktoren zu berücksichtigen wie die zeitlichen Ressourcen, der gesundheitliche Zustand, das Interesse daran, die Lebenserfahrung u.a. Da die Annotator:innen weder Terminologie- noch Rechtsexpert:innen waren, konnte lediglich nach dem Prinzip des besten Wissens und Gewissens vorgegangen werden. Bei der Einigung auf den Goldstandard wurde zwar auf externe Quellen wie Lexika, Internet u.a. zugegriffen, aber auch dann herrschte unter den Annotator:innen gewisse Unsicherheit. Ein weiterer Aspekt, der einen Einfluss auf die erbrachte Leistung beider Annotator:innen haben konnte, betrifft die Art und Weise, wie Rechtstexte auf Deutsch und Polnisch formuliert werden. Insbesondere im Polnischen sind längere, verschachtelte Sätze für solche Texte üblich, die trotz der ersprachlichen Kompetenz eine Herausforderung beim Verstehen darstellen. Dies konnte auch die automatische Termextraktion in den polnischen Texten beeinflusst haben. Diese Unsicherheit spiegelte sich außerdem in den teilweise größeren Unterschieden zwischen den jeweiligen Annotator:innen bei der unabhängig voneinander durchgeführten Annotation wider.

Die aufgrund der Untersuchung gesammelten Daten überstiegen teilweise das Vorhaben der vorliegenden Arbeit. Dazu zählen z. B. die Evaluierungsergebnisse zwischen dem jeweiligen Extraktionstool und den einzelnen Annotator:innen. Diese Daten wurden zwar erfasst, aber im Rahmen der vorliegenden Arbeit nicht weiterverfolgt und bilden einen möglichen Forschungspunkt, der im weiteren Unterkapitel (10.2) beschrieben wird.

10.2. Herausforderungen und Beschränkungen der Untersuchung

Die im vorherigen Unterkapitel genannten Schwächen der Extraktionstools bei der Erkennung der Termhood und Unithood der Benennungen bildeten gleichzeitig die größte Herausforderung für die durchgeführte Untersuchung. Die ermittelten Werte sind trotzdem aussagekräftig und ermöglichen, eine Tendenz in Hinblick auf die gestellten Forschungsfragen aufzuzeigen, denn die Gütekriterien der quantitativen Forschung Validität und Objektivität wurden eingehalten, da insgesamt drei Annotator:innen an der Annotation arbeiteten und einander dabei nicht beeinflussten. Die Reliabilität ist aufgrund der Tatsache, dass Systeme immer wieder neu trainiert werden problematisch und dieser Aspekt wird ebenfalls im nächsten Unterkapitel (10.3) ausführlicher behandelt. Eine der Lösungen, um ggf. auf genauere Daten zu kommen, wäre Rechtsexpert:innen zur Bewertung des Goldstandards zuzuziehen. Da dies im Falle der Rechtsterminologie allerdings gleichzeitig die Auseinandersetzung mit allen Texten bedeutet, damit

der Kontext ermittelt werden kann, und aufgrund der Tatsache, dass es sich um zwei verschiedene Sprachen und Rechtssysteme handelt, war es aus monetären Gründen zu aufwändig.

Weiters bildet das Prompt Engineering einen eigenen Forschungsbereich. Möglicherweise könnte eine ausführlichere Auseinandersetzung mit dem Thema bzw. die Unterstützung eines:einer Experten:in und der Einsatz anderer Strategien bei der Formulierung von Prompts die von ChatGPT erbrachte Leistung steigern. Dazu zählen Beispieldaten in den Anfragen oder Folgeanfragen, die ChatGPT dazu bringen sollten, mehr Termkandidaten zu extrahieren bzw. bessere Resultate zu erbringen. In Bezug auf den Umfang und das tatsächliche Thema der Arbeit wurden diese Ansätze allerdings nicht verfolgt.

Der bereits genannte Umfang der Arbeit und die dafür vorgesehene Zeit wirkten ebenfalls beschränkend auf die Untersuchung. Weitere Prompts und/oder das umfangreichere Korpus waren anhand der verfügbaren Ressourcen nicht möglich.

Der finanzielle Aufwand war auch einschränkend bezüglich der Auswahl des Extraktionstools, mit dem die Leistung von ChatGPT-4o verglichen wurde. Wie im theoretischen Teil der Arbeit geschildert, sind mehrere Tools am Markt verfügbar, die auf die Termextraktion spezialisiert sind. Um sie bzw. ihre vollständige Version nutzen zu können, muss allerdings eine finanzielle Investition getätigt werden, die alleine zwecks der Untersuchung nicht vertretbar war.

10.3. Weiterführende Forschung

Die durchgeführte Untersuchung ist nicht nur in der Terminologieforschung, sondern auch in der Erforschung der LLMs eingebettet. Aufgrund der gewonnenen Daten lässt sich z. B. die Leistung beider Extraktionstools im Vergleich mit dem Goldstandard mit deren Leistung im Vergleich mit den Referenzlisten jeweiliger Annotator:innen erforschen. Die Unsicherheiten zwischen den Annotator:innen bilden die Grundlage für die Ermittlung der Kappa-Werte und die weitere qualitative Untersuchung, um die Gründe für die Unterschiede zu ermitteln. Inhaltlich lassen sich die automatisch extrahierten Mehrwortbenennungen ausführlicher analysieren, um genauer zu ermitteln, warum sie eine Herausforderung für die Extraktionstools darstellten und z. B. inwieweit sich die Anzahl der Bestandteile einer Mehrwortbenennung auf den Extraktionserfolg auswirkt.

Im Rahmen der Untersuchung wurden außerdem Daten darüber erhoben, welche Resultate die Prompts bei den jeweiligen Texten erzielten. Im Fokus der Untersuchung standen allerdings nur die gewonnenen Werte und nicht der Einfluss der einzelnen Textfragmente auf die Leistung der Extraktionstools. Die Daten könnten dementsprechend genutzt werden, um

eine qualitative Inhaltsanalyse durchzuführen, die erforscht, wie sich z. B. die Länge, der Grad der Komplexität, der ausgesuchte Textausschnitt und andere Faktoren auf die Leistung der Tools bzw. im Falle von ChatGPT einzelner Prompts auswirken.

Eine weitere Forschungsmöglichkeit bezieht sich auf die Reliabilität der gewonnenen Daten. Es könnte die gleiche Untersuchung z. B. in einem Jahr durchgeführt werden, um zu erforschen, wie sich die Liste mit den extrahierten Termkandidaten beim Einsatz derselben Prompts geändert hat. Die für die vorliegende Untersuchung verwendete Version von ChatGPT-4o kann außerdem in absehbarer Zukunft durch eine neuere Version ersetzt werden, mit der die Untersuchung ebenfalls wiederholt werden kann. Außerdem ist die momentan entgeltliche Version von ChatGPT-4.0 verfügbar, die ebenfalls dazu verwendet werden könnte.

Schließlich hilft die Untersuchung der Leistung von ChatGPT anhand anderer Sprachen als Englisch, die Forschungslücke zu schließen und dessen Leistung in den jeweiligen Sprachen festzustellen. Konkret lassen sich die Ergebnisse der Untersuchung mit jenen vergleichen, die andere Rechtsbereiche oder andere fachsprachliche Texte in den untersuchten Sprachen betreffen. Insbesondere für Polnisch, die im Vergleich mit Englisch und Deutsch eine unterrepräsentierte Sprache in Bezug auf die Menge der verfügbaren Trainingsdaten ist. Die meisten im Zuge der Recherche des aktuellen Stands der Forschung gefundenen Studien betreffen in erster Linie die Leistung von ChatGPT bei der Termextraktion in englischer Sprache.

11. Fazit

In der vorliegenden Arbeit wurde in erster Linie die Leistung von LLMs anhand von ChatGPT am Gebiet der automatischen Termextraktion erforscht. Die generierten Termkandidaten wurden mittels der Qualitätsmetriken Genauigkeit, Trefferquote und F1-Score anhand der Ergebnisse der humanen, manuellen Terminologierecherche evaluiert und mit den Ergebnissen eines spezialisierten Extraktionstools verglichen. Um die Leistung von ChatGPT zu relativieren und dadurch genauere Daten zu ermitteln, wurden die Termkandidaten mit fünf verschiedenen Prompts gewonnen. Weiters war das Ziel dieser Masterarbeit, zur Forschungslücke beizutragen, indem die Leistung von ChatGPT in diesem Gebiet für andere Sprachen als Englisch ermittelt wird, weswegen im Falle der vorliegenden Arbeit österreichische und polnische Texte herangezogen wurden. Die Forschungslücke betrifft Polnisch noch mehr als Deutsch, da vergleichsmäßig weniger Trainingsdaten in der polnischen Sprache zur Verfügung stehen. Mit den Ergebnissen der Untersuchung wird mindestens zum gewissen Grad u.a. auf die Studien von De Paiva et al. (2023); Mouratidis et al. (2022); Premasiri et al. (2023); Rokas et al. (2020) angeknüpft, weshalb sie sich in den aktuellen Forschungsstand eingliedern lassen.

Die für die Untersuchung ausgewählte Methodik führte erfolgreich zur Erhebung der im Hinblick auf die Forschungsfragen relevanten Daten. Es wurden drei österreichische und drei polnische Entscheidungstexte in vergleichbarer Länge ausgesucht, die den Prinzipien der Erstsprachlichkeit, Fachlichkeit und Aktualität folgten. Um die Vergleichbarkeit der Ergebnisse zu gewährleisten, kamen alle Texte aus dem für die neuesten Methoden der künstlichen Intelligenz relevanten Rechtsbereich des Urheberrechts. Die Systemauswahl, insbesondere die Entscheidung für das dem Vergleich dienende Extraktionstool Text2TCS, ermöglichte die Überprüfung der in der Einleitung gestellten Annahmen. Das Tool wurde aufgrund seiner Relevanz für das Vorhaben, wegen der in den anderen Studien erzielten hohen Ergebnisse und aus Verfügbarkeits- sowie Kostengründen ausgewählt. Die zwecks der Evaluierung verwendeten Qualitätsmetriken Genauigkeit, Trefferquote und F1-Score sowie die manuelle Erstellung vom Goldstandard entsprechen der in der ähnlichen Forschung üblichen Vorgehensweise und ermöglichen somit die Vergleichbarkeit der gewonnenen Ergebnisse und gesammelten Erkenntnisse mit anderen Studien. Die Metriken ließen sich außerdem dank der Hilfe einer Expertin mit dem speziell dafür in Python erstellten Code automatisch berechnen, was angesichts des Umfangs der Untersuchung mit über 250 F1-Messungen aus Zeitgründen unerlässlich war. Die Validität und Objektivität der erhobenen Daten wurden gesichert, indem insgesamt drei Annotator:innen unabhängig voneinander manuelle Annotationen durchgeführt haben.

Durch die im Zuge der durchgeführten Untersuchung gewonnenen Daten und gesammelten Erkenntnisse zeichnet sich die Tendenz in Bezug auf die erste Forschungsfrage und die Annahme ab, dass das neuronale generative Sprachmodell ChatGPT-4o, dessen primärer Nutzungsbereich die Termextraktion nicht einschließt, schlechter an diesem Gebiet als das Extraktionstool Text2TCS, das speziell zwecks der automatischen Termextraktion entwickelt wurde, abschneidet und seine Leistung zum großen Teil nicht den Ergebnissen der manuellen Terminologierecherche entspricht. Weiters erteilte die Datenerhebung Auskunft darüber, bei welchen der beiden analysierten Sprachen ChatGPT besser auf dem Gebiet der Termextraktion abschneidet. Aufgrund der Untersuchung zeichnet sich die Tendenz in Bezug auf die zweite Forschungsfrage und Annahme ab, dass ChatGPT bessere Resultate bei der Gewinnung der Rechtsterminologie aus den österreichischen als aus den polnischen Texten erzielte. Der andere Teil der zweiten Forschungsfrage und die Annahme, dass ChatGPT mit unterschiedlich formulierten Prompts verschiedene Termkandidaten generiert, zeichnet sich mit den gewonnenen Daten ebenfalls ab, allerdings lässt sich nicht eindeutig feststellen, welche Prompts die besten Resultate erzielten und warum, was die Notwendigkeit weiterer Forschung an diesem konkreten Gebiet untermauert.

Der zeitliche Aufwand, der geplante Umfang der Masterarbeit und die finanziellen Ressourcen wirkten sich beschränkend auf die durchgeführte Untersuchung aus. Dies äußerte sich in erster Linie in der überschaubaren Länge der Texte, aus denen die Terminologie extrahiert wurde. Weiters wurden im Zuge der Untersuchung Daten erhoben, insbesondere bezüglich der Leistung einzelner Annotator:innen, die, im Gegensatz zum Goldstandard, nicht direkt die Untersuchung betrafen und aus den o.g. Gründen nur aufgelistet und nicht weiterverfolgt wurden. Weiters sind Annotator:innen, die keine Rechtsexpert:innen waren, zum gewissen Grad trotz intensiver Recherche wegen der Fachlichkeit der Rechtstexte an ihre Grenzen bei der Erstellung vom Goldstandard gestoßen.

In weiterführender Forschung, die an die erhobenen Daten anknüpft, kann es daher durchaus sinnvoll sein, Fachexpert:innen zuzuziehen, um die Termhood und Unithood der im Goldstandard enthaltenen Ein- und Mehrwortbenennungen zu gewährleisten bzw. überprüfen. Weiters kann der Korpusumfang erweitert werden, indem entweder längere Textfragmente und/oder mehrere Voraussetzungen erfüllende Texte herangezogen werden. Überlegenswert ist die (weitere) Vorverarbeitung der Texte vor der Extraktion, allerdings würde dies der üblichen Praxis bei der Nutzung von ChatGPT widersprechen. Weiters könnten Forschende andere Strategien bei der Formulierung der Prompts einsetzen und die generierten Termkandidaten mit den

in dieser Untersuchung erhobenen Daten vergleichen. Besonders interessant sind die Ergebnisse der vorliegenden Arbeit für Forschende, die die Leistung von ChatGPT für Deutsch und Polnisch untersuchen. Außerdem könnten die gewonnenen Erkenntnisse dem Vergleich der Leistung von verschiedenen Extraktionstools bei der automatischen Termextraktion dienen oder die Ergebnisse der Untersuchung könnten aus einer ganz anderen Perspektive betrachtet werden und für den Vergleich der Leistung von ChatGPT innerhalb und zwischen den slawischen Sprachen von Nutzen sein. Schließlich kann es sinnvoll sein, die erhobenen Daten für den Vergleich der Leistung von verschiedenen Versionen von ChatGPT zu verwenden, um dessen Entwicklung zu erforschen.

Die vorliegende Arbeit und die erhobenen Daten sind aus mehreren Gründen relevant für die aktuelle Forschung. Dies ist an erster Stelle auf die rasche Entwicklung der Methoden der künstlichen Intelligenz zurückzuführen. LLMs sind immer mehr für die Allgemeinheit verfügbar und werden sowohl im privaten als auch im beruflichen Kontext eingesetzt. Zu dem letzteren zählt immer mehr die Arbeit mit Terminologie, insbesondere in solchen Bereichen wie Translation, Branding und Marketing, IT, Recht, Medizin, Technologie u.v.m. Die automatische Termextraktion kann u.a. beim Verfassen von Texten oder bei der Vorbereitung auf Verhandlungen oder Präsentationen von Nutzen sein und ist, solange sie verlässliche Resultate generiert, aus zeitlichen und finanziellen Gründen der manuellen Terminologierecherche überlegen. Daraus lässt sich schließen, dass es in absehbarer Zukunft zu weiteren Entwicklungen auf diesem Gebiet kommen wird. Da dies allerdings bei neuronalen Ansätzen immer größere Menge an Daten erfordert, wird man zukünftig häufiger mit der Frage des Urheberrechts konfrontiert, was ebenfalls zu weiteren Forschungsideen führen kann.

12. Literaturverzeichnis

- Adams, Jonathan (2024). *KI-Grundlagen des Prompt Engineering*. Wrocław: Jon Adams.
- Adamzik, Kirsten (2018). *Fachsprachen: die Konstruktion von Welten*. Tübingen: A. Francke Verlag.
- Ahmad, Khurshid; Davies, Andrea; Fulford, Heather & Rogers, Margaret (1994). What is a term? The semi-automatic extraction of terms from text. In: Snell-Hornby, Mary; Pöschhacker, Franz & Kaindl, Klaus (Hrsg.) *Translation Studies: An Interdiscipline. Selected papers from the Translation Studies Congress, Vienna, 1992*. Bd. 2. Amsterdam: John Benjamins Publishing Co.
- Ahmad, Khurshid & Rogers, Margaret (2001). Corpus Linguistics and Terminology Extraction. In: Budin, Gerhard & Wright, Sue Ellen (Hrsg.) *Handbook of Terminology Management: Volume 2: Application-Oriented Terminology Management*. Amsterdam: John Benjamins, 725–760.
- Ala, Hema & Sharma, Dipti Misra (2020). Graph Based Automatic Domain Term Extraction. *Proceedings of the 17th International Conference on Natural Language Processing (ICON): TermTraction 2020 Shared Task*, 1–4.
- Alegria, Iñaki; Urizar, Ruben; Gurrutxaga, Antton; Lizaso, Pili; Saralegi, Xabier & Ugartetxea, Sahats (2004). A XML-Based Term Extraction Tool for Basque. *LREC*, 1–4.
- Alpaydin, Ethem (2022). *Maschinelles Lernen*. 3. Aufl. München/Wien: De Gruyter Oldenbourg.
- Ammon, Ulrich (1977). *Probleme der Soziolinguistik*. Berlin/Boston: Max Niemeyer Verlag.
- Arcan, Mihael; O'Halloran, Rory; Robin, Cécile & Buitelaar, Paul (2022). Towards bootstrapping a chatbot on industrial heritage through term and relation extraction. *Proceedings of the 2nd International Workshop on Natural Language Processing for Digital Humanities*, 108–122.
- Arntz, Reiner (2001). Juristisches Übersetzen im Spannungsfeld von Recht und Sprache. In: Mayer, Felix (Hrsg.) *Dolmetschen und Übersetzen: der Beruf im Europa des 21. Jahrhunderts: Akten des Kongresses des BDÜ-Landesverbandes Bayern, 23. - 25. November 2001, München*. Freiburg: Freigang, Mauro + Reinke, 99–108.
- Arntz, Reiner; Picht, Heribert & Schmitz, Klaus-Dirk (2014). *Einführung in die Terminologiearbeit*. 7. Aufl. Hildesheim [u.a.]: Olms.
- Arntz, Reiner & Sandrini, Peter (2007). Präzision versus Vagheit: das Dilemma der Rechtssprache im Lichte von Rechtsvergleich und Sprachvergleich. In: Bassey, Antia (Hrsg.)

- Indeterminacy in LSP and Terminology: Studies in honour of Heribert Picht*. The Netherlands: John Benjamins Publishing Company, 135–153.
- Arppe, Antti (1995). Term Extraction from unrestricted Text. *The Nordic Conference on Computational Linguistics (NoDaLiDa)*.
- Astrakhansev, Nikita (2018). ATR4S: toolkit with state-of-the-art automatic terms recognition methods in Scala. *Language Resources and Evaluation* 52 (3), 853–872.
- Baeza-Yates, Ricardo & Ribeiro, Berthier de Araújo Neto (2011). *Modern information retrieval: the concepts and technology behind search*. 2. Aufl. Harlow [u.a.]: Pearson Education.
- Baker, Pam (2024). *ChatGPT für Dummies*. Weinheim: Wiley-VCH GmbH.
- Baker, Paul (2006). *Glossary of corpus linguistics*. Edinburgh: Edinburgh University Press.
- Baldinger, Kurt (1952). *Die Gestaltung des wissenschaftlichen Wörterbuchs: historische Betrachtungen zum neuen Begriffssystem als Grundlage für die Lexikographie von Halig und Wartburg*. Hamburg: De Gruyter.
- Bang, Yejin; Cahyawijaya, Samuel; Lee, Nayeon; Dai, Wenliang; Su, Dan; Wilie, Bryan; Lovenia, Holy; Ji, Ziwei; Yu, Tiezheng & Chung, Willy (2023). A multitask, multilingual, multimodal evaluation of chatgpt on reasoning, hallucination, and interactivity. *Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics* 1, 675–718.
- Bańko, Mirosław (2011). Szyk przymiotników. In: *Poradnia Językowa PWN*. Warszawa: Wydawnictwo Naukowe PWN. <https://sjp.pwn.pl/poradnia/haslo/szyk-przymiotnikow;12630.html> (Stand: 10.07.2024).
- Barrón, Alberto; Sierra, Gerardo & Villaseñor, Elio (2006). C-value aplicado a la extracción de términos multipalabra en documentos técnicos y científicos en español. *7th Mexican International Conference on Computer Science (ENC 2006)*, 1–6.
- Barrón-Cedeño, Alberto; Sierra, Gerardo; Drouin, Patrick & Ananiadou, Sophia (2009). An improved automatic term recognition method for Spanish. *International Conference on Intelligent Text Processing and Computational Linguistics*, 125–136.
- Barta, Janusz & Markiewicz, Ryszard (2010). *Prawo autorskie*. 2. Aufl. Warszawa: Wolters Kluwer Polska.
- Baumann, Klaus-Dieter (1992). *Integrative Fachtextlinguistik*. Tübingen: Narr.

- Baumann, Klaus-Dieter (2004). Emotionen in der Fachkommunikation: ein kommunikativ-kognitiver Untersuchungsansatz. In: Baumann, Klaus-Dieter & Kalverkämper, Hartwig (Hrsg.) *Pluralität in der Fachsprachenforschung*. Tübingen: Narr, 83–119.
- Baumann, Klaus-Dieter (2010). Specialist Thinking Strategies in LSP Communication of the Natural and Technical Sciences. *Reconceptualizing LSP: Online Proceedings of the XVII European LSP Symposium 2009*, 1–13.
- Bausch, Karl-Heinz; Schewe, Wolfgang H. U. & Spiegel, Heinz-Rudi (1976). Eindeutige Verständigung - ein Element technisch-wissenschaftlicher Verständigung. In: Bausch, Karl-Heinz, Schewe, Wolfgang H. U. & Spiegel, Heinz-Rudi (Hrsg.) *Fachsprachen: Terminologie, Struktur, Normung*. Berlin/Köln: Beuth, 11–18.
- Beltagy, Iz; Lo, Kyle & Cohan, Arman (2019). SciBERT: A Pretrained Language Model for Scientific Text. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Association for Computational Linguistics*, 3615–3620.
- Bendel, Oliver (2023). ChatGPT. In: *Gabler Wirtschaftslexikon*. Wiesbaden: Springer Gabler Verlag. <https://wirtschaftslexikon.gabler.de/definition/chatgpt-124904/version-389664> (Stand: 13.10.2024).
- Bender, Emily M.; Gebru, Timnit; McMillan-Major, Angelina & Shmitchell, Shmargaret (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*, 610–623.
- Bielawski, Paweł (2023). Zur (Un-)Verständlichkeit der Rechtssprache am Beispiel der deutschen Rechtssprache. *Lebende Sprachen* 68, 40–57.
- Biemann, Chris; Heyer, Gerhard & Quasthoff, Uwe (2022). *Wissensrohstoff Text: eine Einführung in das Text Mining*. 2. Aufl. Wiesbaden: Springer Vieweg.
- Birsak, Agnieszka; Dissauer, Nikolaus; Frass, Ursula; Müller, Magdalena; Schellenberg-Jaworski, Sylwia; Schellnegger, Iwona; Stadler, Ewa; Wolborska-Lauter, Joanna & Ziemska, Joanna (2022). *Die österreichische Rechtsterminologie. Öffentlich-rechtlicher Teil: Deutsch - Polnisch. Austriacka terminologia prawnicza. Część prawa publicznego: Niemiecki - Polski*. Wien: Österreichischer Verband der allgemein beeideten und gerichtlich zertifizierten Dolmetscher (ÖVGd).
- Birsak, Agnieszka; Dissauer, Nikolaus; Furmahska, Agnieszka; Frass, Ursula; Jaworski, Adam; Müller, Magdalena; Schellenberg-Jaworski, Sylwia; Stadler, Ewa; Wolborska-Lauter, Joanna & Ziemska, Joanna (2021). *Die österreichische Rechtsterminologie. Judizieller*

- Teil. Część sądownicza. Deutsch - Polnisch, Niemiecki - Polski.* Wien: Österreichischer Verband der allgemein beeideten und gerichtlich zertifizierten Dolmetscher (ÖVGD).
- Błęszyński, Jan (2015). Polskie prawo autorskie po dwudziestu latach obowiązywania ustawy z 4 lutego 1994 r. *Ruch Prawniczy, Ekonomiczny i Socjologiczny* 77 (2), 5–25.
- Bolten, Jürgen (1992). „Fachsprache“ oder „Sprachbereich“: Empirischpragmatische Grundlagen zur Beschreibung der deutschen Wirtschafts-, Medizin- und Rechtssprache. In: Bungarten, Theo (Hrsg.) *Beiträge zur Fachsprachenforschung: Sprache in Wissenschaft und Technik, Wirtschaft und Rechtswesen*. Tostedt: Attikon-Verl., 57–72.
- Bordea, Georgeta; Buitelaar, Paul & Coughlan, Barry (2014). Hot topics and schisms in NLP: Community and trend analysis with saffron on ACL and LREC proceedings. *9th Edition of Language Resources and Evaluation Conference (LREC2014)*, 1–6.
- Böttcher, Eike (2021). *Kirchner – Abkürzungsverzeichnis der Rechtssprache*. 10. Aufl. Berlin/Boston: De Gruyter.
- Bourigault, Didier (1992). Surface Grammatical Analysis for the Extraction of Terminological Noun Phrases. *Proceedings of Coling-92*, 977–981.
- Bowker, Lynne (2008). *Computer-aided translation technology: a practical introduction*. Ottawa: Univ. of Ottawa Press.
- Brown, Tom; Mann, Benjamin; Ryder, Nick; Subbiah, Melanie; Kaplan, Jared D.; Dhariwal, Prafulla; Neelakantan, Arvind; Shyam, Pranav; Sastry, Girish & Askell, Amanda (2020). Language models are few-shot learners. *Advances in neural information processing systems* 33, 1877–1901.
- Büchele, Manfred (2023). *Urheberrecht*. 3. Aufl. Wien: MANZ'sche Verlags- und Universitätsbuchhandlung.
- Budin, Gerhard (1999). *Theorie und Praxis der übersetzungsbezogenen Terminologiearbeit*. Wien: WUV-Univ.-Verl.
- Bundeskanzleramt der Republik Österreich (2023a). *Rechtsinformationssystem des Bundes (RIS)*. <https://www.ris.bka.gv.at/Dokument.wxe?ResultFunctionToken=37aa55a2-f0e7-4a13-8963-19500da35aba&Abfrage=Gesamtabfrage&SearchInAsylGH=&SearchInAvn=&SearchInAvsv=&SearchInBegut=&SearchInBgblAlt=&SearchInBgblAuth=&SearchInBgblPdf=&SearchInBks=&SearchInBundesnormen=&SearchInBvb=&SearchInBvwg=&SearchInDok=&SearchInDsk=&SearchInErlaesse=&SearchInGbK=&SearchInGemeinderecht=&SearchInGemeinderechtAuth=&SearchInJustiz=&SearchInKmGer=&SearchInLandesnormen=&SearchInLvwg=&SearchIn->

Lgbl=&SearchInLgblNO=&SearchInLgblAuth=&SearchInMrp=&SearchInNormenliste=&SearchInPruefGewO=&SearchInPvak=&SearchInRegV=&SearchInSpg=&SearchInUbas=&SearchInUmse=&SearchInUpts=&SearchInUvs=&SearchInVbl=&SearchInVerg=&SearchInVfgh=&SearchInVwgh=&ImRisSeitVonDatum=&ImRisSeitBisDatum=&ImRisSeit=&ResultPageSize=100&Suchworte=4Ob59%2f23i&Dokumentnummer=JJT_20231017_OGH0002_0040OB00059_23I0000_000 (Stand: 07.11.2024).

Bundeskanzleramt der Republik Österreich (2023b). *Rechtsinformationssystem des Bundes (RIS)*. https://www.ris.bka.gv.at/Dokument.wxe?ResultFunctionToken=816f7fa6-68f8-46aa-ab83-129e3c251a0d&Position=1&SkipToDocumentPage=True&Abfrage=Justiz&Fachgebiet=Urheberrecht&Gericht=&Rechtssatznummer=&Rechtssatz=&Fundstelle=&Spruch=&Rechtsgebiet=Zivilrecht&AenderungenSeit=Undefined&JustizEntscheidungsart=&Norm=&SucheNachRechtssatz=False&SucheNachText=True&GZ=&VonDatum=01.07.2022&BisDatum=26.03.2024&ImRisSeitVonDatum=&ImRisSeitBisDatum=&ImRisSeit=Undefined&ImRisSeitChangeSet=Undefined&ImRisSeitForRemotion=Undefined&ResultPageSize=100&Suchworte=&Dokumentnummer=JJT_20230131_OGH0002_0040OB00139_22B0000_000 (Stand: 07.11.2024).

Bundeskanzleramt der Republik Österreich (2024a). *Rechtsinformationssystem des Bundes (RIS)*. <https://www.ris.bka.gv.at> (Stand: 26.10.2024).

Bundeskanzleramt der Republik Österreich (2024b) *Rechtsinformationssystem des Bundes (RIS)*. https://www.ris.bka.gv.at/Dokument.wxe?ResultFunctionToken=7f2ba226-5ca0-4e36-911cf6bfd23f1b60&Position=1&SkipToDocumentPage=True&Abfrage=Justiz&Fachgebiet=Urheberrecht&Gericht=&Rechtssatznummer=&Rechtssatz=&Fundstelle=&Spruch=&Rechtsgebiet=Zivilrecht&AenderungenSeit=Undefined&JustizEntscheidungsart=&Norm=&SucheNachRechtssatz=False&SucheNachText=True&GZ=&VonDatum=01.07.2022&BisDatum=16.03.2024&ImRisSeitVonDatum=&ImRisSeitBisDatum=&ImRisSeit=Undefined&ImRisSeitChangeSet=Undefined&ImRisSeitForRemotion=Undefined&ResultPageSize=100&Suchworte=&Dokumentnummer=JJT_20240125_OGH0002_0040OB00120_23K0000_000 (Stand: 07.11.2024).

- Cabré Castellví, Maria Teresa; Estopà Bagot, Rosa & Vivaldi Palatresi, Jordi (2001). Automatic term detection. A review of current systems. In: Bourigault, Didier, Jacquemin, Christian, & L’Homme, Marie-Claude (Hrsg.) *Recent Advances in Computational Terminology*. Bd. 2. Amsterdam/Philadelphia: Benjamins, 53–87.
- Cabré, Maria Teresa (1999). *Terminology: theory, methods and applications*. Amsterdam [u.a.]: Benjamins.
- Campeato, Oswald (2020). *Artificial intelligence, machine learning, and deep learning*. Dulles, Virginia: Mercury Learning and Information.
- Carstensen, Kai-Uwe; Ebert, Christian & Ebert, Cornelia (Hrsg.) (2010). *Computerlinguistik und Sprachtechnologie: eine Einführung*. 3. Aufl. Heidelberg: Spektrum Akademischer Verlag.
- Chataut, Sandeep; Do, Tuyen; Gurung, Bichar Dip Shrestha; Aryal, Shiva; Khanal, Anup; Lushbough, Carol & Gnimpieba, Etienne (2024). Comparative Study of Domain Driven Terms Extraction Using Large Language Models. *arXiv preprint arXiv:2404.02330*, 1–10.
- Chatzitheodorou, Konstantinos & Kappatos, Vassilios (2021). Hybrid extraction of multi-word terms: an application on vibration-based condition monitoring technique. *Mathematical Models in Engineering* 7 (1), 1–9.
- Ciresa, Meinhard (2022). *Praxishandbuch Urheberrecht*. 3. Aufl. Wien: Linde.
- Clark, Kevin; Luong, Minh-Thang; Le, Quoc & Manning, Christopher (2020). ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators. *ICLR 2020*, 1–18.
- Conneau, Alexis; Khandelwal, Kartikay; Goyal, Naman; Chaudhary, Vishrav; Wenzek, Guillaume; Guzmán, Francisco; Grave, Edouard; Ott, Myle; Zettlemoyer, Luke & Stoyanov, Veselin (2020). Unsupervised Cross-lingual Representation Learning at Scale. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, 8440–8451.
- Costa, Hernani; Zaretskaya, Anna; Corpas Pastor, Gloria & Seghiri Domínguez, Míriam (2016). Nine terminology extraction Tools: Are they useful for translators? *MultiLingual* 27 (3), 14–20.
- Cover, Thomas M. & Thomas, Joy A. (2006). *Elements of information theory*. Hoboken, New Jersey: Wiley-Interscience.
- Daille, Béatrice (2017). *Term variation in specialised corpora: characterisation, automatic discovery and applications*. Amsterdam: John Benjamins Publishing Company.

- Daille, Béatrice; Gaussier, Éric & Langé, Jean-Marc (1994). Towards automatic extraction of monolingual and bilingual terminology. *COLING 1994 Volume 1: The 15th International Conference on Computational Linguistics*, 515–521.
- De la Durantaye, Katharina & Raue, Benjamin (2020). Urheberrecht und Zugang in einer digitalen Welt - Urheberrechtliche Fragestellungen des Zugangs für Gedächtnisinstitutionen und die Digital Humanities. *RuZ-Recht und Zugang*, 83–94.
- De Paiva, Valeria; Gao, Qiyue; Kovalev, Pavel & Moss, Lawrence S. (2023). Extracting Mathematical Concepts with Large Language Models. *CEUR Workshop Proceedings*, 1–13.
- Devlin, Jacob; Chang, Ming-Wei; Lee, Kenton & Toutanova, Kristina (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. *North American Chapter of the Association for Computational Linguistics*, 4171–4186.
- Di Nunzio, Giorgio Maria; Marchesin, Stefano & Silvello, Gianmaria (2023). A systematic review of Automatic Term Extraction: What happened in 2022? *Digital Scholarship in the Humanities* 38, 41–47.
- Drewer, Petra & Horend, Sybille (2007). The Janus Head Article - How Much Terminology Theory Can Practical Terminology Management Use? *Language at work* 2, 1–22.
- Drewer, Petra & Schmitz, Klaus-Dirk (2017). *Terminologiemanagement: Grundlagen - Methoden - Werkzeuge*. Berlin: Springer Vieweg.
- Drouin, Patrick (2003). Term extraction using non-technical corpora as a point of leverage. *Terminology* 9, 99–115.
- Durango, María C.; Torres-Silva, Ever A. & Orozco-Duque, Andrés (2023). Named entity recognition in electronic health records: A methodological review. *Healthcare Informatics Research* 29 (4), 286–300.
- Eckardt, Birgit (2000). *Fachsprache als Kommunikationsbarriere? Verständigungsprobleme zwischen Juristen und Laien*. Wiesbaden: Dt. Univ.-Verl.
- Eckstein, Karina (2009). Toolgestützte Terminologieextraktion. In: Mayer, Felix & Seewald-Heeg, Uta (Hrsg.) *Terminologiemanagement: von der Theorie zur Praxis*. Berlin: Bundesverb. der Dolmetscher und Übers., 108–120.
- Eder, Maciej; Rybicki, Jan & Kestemont, Mike (2016). Stylometry with R: A package for computational text analysis. *The R journal* 8, 107–121.
- Edo Marzá, Nuria (2009). *The specialised lexicographical approach: a step further in dictionary-making*. Bern/Wien [u.a.]: Lang.

- Efing, Christian (2023). Berufssprache. In: Efing, Christian & Kalkavan-Aydın, Zeynep (Hrsg.) *Berufs- und Fachsprache Deutsch in Wissenschaft und Praxis*. Bd. 3. Berlin/Boston: Walter de Gruyter GmbH, 655–658.
- Elsen, Hilke (2007). Wortgruppenlexeme - Beispiele aus Enzyklopädie, Zeitung, Baurecht und Wasserbau. *Internationale Zeitschrift für Fachsprachenforschung, -didaktik und Terminologie* 29, 44–55.
- Elsen, Hilke (2017). Wortgruppenlexeme zwischen Wortbildung und Phraseologie. *Yearbook of phraseology* 8, 145–160.
- Engberg, Jan (1992). Gläser, Rosemarie: Fachtextsorten im Englischen, Tübingen: Gunter Narr Verlag 1990. *Hermes* 8, 93–99.
- Engberg, Jan (2017). Fachkommunikation und fachexterne Kommunikation. In: Felder, Ekkehard & Vogel, Friedemann (Hrsg.) *Handbuch Sprache im Recht*. Berlin/Boston: De Gruyter Mouton, 118–137.
- Ermert, Karl (1983). Für eine bürgerfreundliche Justiz. In: Wassermann, Rudolf & Peterson, Jürgen (Hrsg.) *Recht und Sprache: Beiträge zu einer bürgerfreundlichen Justiz*. Heidelberg: Müller Juristischer Verlag, 11–17.
- Europäische Kommission (2023). *Technische Dokumentation und EU-Konformitätserklärung*. https://europa.eu/youreurope/business/product-requirements/compliance/technical-documentation-conformity/index_de.htm (Stand: 26.10.2024).
- Europäisches Parlament (2024a). *Gesetz über künstliche Intelligenz: Parlament verabschiedet wegweisende Regeln*. <https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law> (Stand: 07.10.2024).
- Europäisches Parlament (2024b). *KI-Gesetz: erste Regulierung der künstlichen Intelligenz*. <https://www.europarl.europa.eu/topics/de/article/20230601STO93804/ki-gesetz-erste-regulierung-der-kunstlichen-intelligenz> (Stand: 07.10.2024).
- Evans, David A. & Zhai, Chengxiang (1996). Noun-phrase analysis in unrestricted text for information retrieval. *Proceedings of ACL, Santa Cruz, University of California*, 17–24.
- ExplosionAI GmbH (2024). *spaCy*. <https://spacy.io> (Stand: 08.11.2024).
- Farm Sanctuary SA (2022). *Pigcasso*. <https://pigcasso.org> (Stand: 06.10.2024).
- Fedorenko, Denis; Astrakhantsev, Nikita & Turdakov, Denis (2014). Automatic recognition of domain-specific terms: an experimental evaluation. *Proceedings of the Institute for System Programming* 26 (4), 55–72.
- Felber, Helmut & Budin, Gerhard (1989). *Terminologie in Theorie und Praxis*. Tübingen: Narr.

- Fiołka, Janusz (2015). Kilka uwag krytycznych o ustawie o prawie autorskim i prawach pokrewnych z 1994 r. *Ruch Prawniczy, Ekonomiczny i Socjologiczny* 77 (2), 263–276.
- Firlit-Fesnak, Grażyna; Jaroszevska, Emilia & Oberloskamp, Helga (2012). *Integracja społeczna jako wyzwanie dla polityki społecznej i pracy socjalnej = Soziale Integration als Herausforderung für die Sozialpolitik und Sozialarbeit*. Warszawa: Instytut Polityki Społecznej Uniwersytetu Warszawskiego: Oficyna Wydawnicza Aspra-JR.
- Fluck, Hans-Rüdiger (1996). *Fachsprachen: Einführung und Bibliographie*. 5. Aufl. Tübingen [u.a.]: Francke.
- Fluck, Hans-Rüdiger (2018). *Fachsprachen – Fachkommunikation – Fachsprachenvermittlung: Beiträge aus 50 Jahren Forschung*. Tübingen: Stauffenburg Verlag.
- Frantzi, Katerina & Ananiadou, Sophia (1995). Statistical Measures for Terminological Extraction. *Proceedings of the Third International Conference on Statistical Analysis of Textual Data*, 297–308.
- Frantzi, Katerina; Ananiadou, Sophia & Mima, Hideki (2000). Automatic recognition of multi-word terms: The C-value/NC-value method. *International journal on digital libraries* 3, 115–130.
- Georgantopoulos, Byron & Piperidis, Stelios (2000). A hybrid technique for automatic term extraction. *Proceedings of the ACIDCA 2000 Conference*, 1–5.
- Gerbert, Manfred (1995). Review of *Fachtextsorten im Englischen*, von Rosemarie Gläser. *Multilingua* 14, 436–438.
- Ghojogh, Benyamin & Ghodsi, Ali (2020). Attention Mechanism, Transformers, BERT, and GPT: Tutorial and Survey, 1–14.
- Gius, Evelyn (2016). Goldstandard. In: *forText*. Darmstadt: Technische Universität Darmstadt. <https://fortext.net/ueber-fortext/glossar/goldstandard> (Stand: 05.08.2024).
- Glapiak, Agnieszka & Stalmach, Sławomir (2023). *Kultura języka w pismach urzędowych: Urzędowy nie taki straszny. Podręcznik: Jak pisać pisma urzędowe*. Warszawa: Wojskowe Centrum Edukacji Obywatelskiej.
- Gläser, Rosemarie (1990). *Fachtextsorten im Englischen*. Tübingen: Narr.
- Gnutzmann, Claus (1980). Fachsprachen und Jargon. In: Gnutzmann, Claus & Turner, John (Hrsg.) *Fachsprachen und ihre Anwendung*. Tübingen: Narr, 49–59.
- Goldberg, Yoav (2017). *Neural network methods for natural language processing*. Morgan & Claypool.
- Goodfellow, Ian; Bengio, Yoshua & Courville, Aaron (2016). *Deep learning*. Cambridge/London: MIT press.

- Göpferich, Susanne (2007). Kürze als Prinzip fachsprachlicher Kommunikation. In: Steinhauer, Anja, Roelcke, Thorsten & Bär, Jochen A. (Hrsg.) *Sprachliche Kurze: konzeptuelle, strukturelle und pragmatische Aspekte*. Bd. 27. Berlin/New York: De Gruyter, 412–433.
- Grabb, Declan (2023). The impact of prompt engineering in large language model performance: a psychiatric example. *Journal of Medical Artificial Intelligence* 6, 1–5.
- Grätz, Axel (2021). *Künstliche Intelligenz Im Urheberrecht: Eine Analyse der Zurechnungskriterien und der Prinzipien der Verwandten Schutzrechte Vor Dem Hintergrund Artificialer Erzeugnisse*. Wiesbaden: Springer Fachmedien Wiesbaden GmbH.
- Gromann, Dagmar (2021a) *Dokumentation zur Text2TCS-Applikation*. <https://text2tcs.univie.ac.at/text2tcs-dokumentation> (Stand: 13.10.2024).
- Gromann, Dagmar (2021b) *Text to Terminological Concept System. Text2TCS. Version: 1.1.1*. <https://live.european-language-grid.eu/catalogue/tool-service/7375> (Stand: 13.10.2024).
- Gromann, Dagmar; Heinisch, Barbara & Lang, Christian (2021). *KG Annotation Tool*. <http://kgannotationtool.trans.univie.ac.at/AnnotationTool.html> (Stand: 08.11.2024).
- Gromann, Dagmar; Wachowiak, Lennart; Lang, Christian & Heinisch, Barbara (2023). Extracting Terminological Concept Systems from Natural Language Text. In: Rehm, Georg (Hrsg.) *European Language Grid: A Language Technology Platform for Multilingual Europe*. Cham: Springer International Publishing, 289–294.
- Grucza, Sambor (2013). *Od lingwistyki tekstu do lingwistyki tekstu specjalistycznego*. 3. Aufl. Warszawa: Wydawnictwo Naukowe Instytutu Kulturologii i Lingwistyki Antropocentrycznej. Uniwersytet Warszawski.
- Grynbaum, Michael M. & Mac, Ryan (2024). The New York Times sues OpenAI and Microsoft over AI use of copyrighted work. *Editor & Publisher Magazine – Newspaper & News Publishing Industry News*. <https://www.editorandpublisher.com/stories/the-new-york-times-sues-openai-and-microsoft-over-ai-use-of-copyrighted-work,247366> (Stand: 06.10.2024).
- Grzelak, Joanna (2010). *Polski język prawa - w perspektywie glottodydaktycznej*. Dissertation, Uniwersytet im. Adama Mickiewicza, Poznań.
- Guadamuz, Andres (2021). Do Androids Dream of Electric Copyright? In: Lee, Jyh-An, Hilty, Reto M. & Liu, Kung-Chung (Hrsg.) *Artificial Intelligence and Intellectual Property*. Oxford: Oxford University Press, 147–176.

- Gulbins, Jürgen & Kahrmann, Christine (2000). *Mut zur Typographie: ein Kurs für Desktop-publishing: mit 40 Tabellen*. 2. Aufl. Berlin [u.a.]: Springer.
- Hałas, Bożena (1995). *Terminologia języka prawnego*. Zielona Góra: Wydawnictwo Wyższej Szkoły Pedagogicznej im. Tadeusza Kotarbińskiego w Zielonej Górze.
- Hartley, Tony (2009). Technology and translation. In: Munday, Jeremy (Hrsg.) *The Routledge companion to translation studies. Revised edition*. London/New York: Routledge, 106–127.
- Hartmann, Dietrich (1980). Über den Einfluß von Fachsprachen auf die Gemeinsprache. Semantische und variationstheoretische Überlegungen zu einem wenig erforschten Zusammenhang. In: Gnutzmann, Claus & Turner, John (Hrsg.) *Fachsprachen und ihre Anwendung*. Tübingen: Narr, 27–48.
- Hegenbarth, Rainer (1982). *Juristische Hermeneutik und linguistische Pragmatik: dargestellt am Beispiel der Lehre vom Wortlaut als Grenze der Auslegung*. Königstein: Athenäum.
- Heid, Ulrich (2001). Verfahren zur Extraktion von Termkandidaten aus Texten: Ein Überblick. In: Mayer, Felix (Hrsg.) *Dolmetschen und Übersetzen: der Beruf im Europa des 21. Jahrhunderts: Akten des Kongresses des BDÜ-Landesverbandes Bayern, 23. - 25. November 2001, München*. Freiburg: Freigang, Mauro + Reinke, 186–204.
- Heller, Klaus (1970). Der Wortschatz unter dem Aspekt des Fachwortes. Versuch einer Systematik. *Wissenschaftliche Zeitschrift der Karl-Marx-Universität Leipzig* 19, 531–544.
- Herzog, Gottfried & Mühlbauer, Holger (2007). *Normen für Übersetzer und technische Autoren*. 2. Aufl. Berlin/Wien [u.a.]: Beuth.
- Heyer, Gerhard; Quasthoff, Uwe & Wittig, Thomas (2008). *Text Mining: Wissensrohstoff Text: Konzepte, Algorithmen, Ergebnisse*. Herdecke [u.a.]: W3L-Verl.
- Heylen, Kris & De Hertog, Dirk (2015) Automatic term extraction. In: Kockaert, Hendrik & Steurs, Frieda (Hrsg.) *Handbook of terminology*. Bd. 1. Amsterdam: John Benjamins, 203–221.
- Hoffmann, Lothar (1976). *Kommunikationsmittel Fachsprache: eine Einführung*. Berlin: Akad.-Verl.
- Hoffmann, Lothar (1985). *Kommunikationsmittel Fachsprache: eine Einführung*. 2. Aufl. Tübingen: Narr.
- Hoffmann, Lothar (1987). Fachsprachen - Instrument und Objekt (An Stelle eines Vorworts). In: Hoffmann, Lothar (Hrsg.) *Fachsprachen: Instrument und Objekt*. Leipzig: VEB Verl. Enzyklopädie, 7–9.

- Hoffmann, Lothar (1988). *Vom Fachwort zum Fachtext: Beiträge zur angewandten Linguistik*. Tübingen: Narr.
- Hoffmann, Lothar (1993). Fachwissen und Fachkommunikation. Zur Dialektik von Systematik und Linearität in den Fachsprachen. In: Bungarten, Theo (Hrsg.) *Fachsprachentheorie*. Bd. 2. Tostedt: Attikon-Verl, 595–617.
- Hoffmann, Lothar (2022). *Kommunikationsmittel Fachsprache*. 2. Aufl. Berlin/Boston: De Gruyter.
- Hoffmann, Ludger (1989). *Rechtsdiskurse: Untersuchungen zur Kommunikation in Gerichtsverfahren*. Tübingen: Narr.
- Hofmann, Norbert (2016). *Redundanz und Äquivalenz in der literarischen Übersetzung: Dargestellt an fünf deutschen Übersetzungen des „Hamlet“*. Tübingen: Max Niemeyer Verlag.
- Hofmarcher, Dominik (2023). Urheberrecht. In: Kucsko, Guido (Hrsg.) *Geistiges Eigentum*. 2. Aufl. Wien: MANZ'sche Verlags- und Universitätsbuchhandlung, 126–170.
- Höhm, Doris (2010). Zum Vorkommen von Stütz- bzw. Funktionsverbgefügen in deutschen und italienischen Rechtstexten. Oder: Vom Umgang mit komplex belassenen Sprachdaten in der empirischen Forschung. In: Hahn, Walther von & Vertan, Cristina (Hrsg.) *Fachsprachen in der weltweiten Kommunikation: Akten des XVI. Europäischen Fachsprachensymposiums, Hamburg 2007*. Frankfurt am Main: Lang.
- Hohnhold, Ingo (1990). *Übersetzungsorientierte Terminologearbeit: eine Grundlegung für Praktiker*. Stuttgart: InTra 1. Fachübersetzer-genossenschaft.
- Homar, Philipp (2024). *Künstliche Intelligenz: Eine Herausforderung für das Urheberrecht*. <https://www.wu.ac.at/forschung/forschungsportal/news/details-news/detail/kuenstliche-intelligenz-eine-herausforderung-fuer-das-urheberrecht> (Stand: 06.10.2024).
- Hong, Munpyo; Fissaha, Sisay & Haller, Johann (2001). Hybrid filtering for extraction of term candidates from German technical texts. *TIA-2001*, 1–10.
- Humphrey, A.; Kuberski, W.; Bialek, J.; Perrakis, N.; Cools, W.; Nuytens, N.; Elakhrass, H. & Cunha, P. A. C. (2022). Machine-learning classification of astronomical sources: estimating F1-score in the absence of ground truth. *Monthly Notices of the Royal Astronomical Society: Letters* 517 (1), 116–120.
- IATE (2024). *Über IATE*. <https://iate.europa.eu/about> (Stand: 26.10.2024).
- Icelandic Language Institute (2001). *Icelandic: at once ancient and modern*. Bd. 13. Reykjavik: Ministry of Education, Science and Culture. [https://www.stjornarradid.is/gogn/rit-og-](https://www.stjornarradid.is/gogn/rit-og)

- Iluk, Jan (1993). Zur Struktur polnischer und deutscher Personenbezeichnungen in juristischen Texten. In: Eichinger, Ludwig M. & Raith, Joachim (Hrsg.) *Sprachkontakte. Konstanten und Variablen*. Bd. 20. Bochum: Universitätsverlag Dr. N. Brockmeyer.
- Ilyas, Ihab F. (2018). Top-k Queries. In: Liu, Ling & Özsu, M. Tamer (Hrsg.) *Encyclopedia of Database Systems*. New York: Springer New York, 4168–4169.
- Janke, Regina (2013). Anforderungen an die Terminologieextraktion: eine vergleichende Untersuchung der Bedürfnisse von Terminologen, technischen Fachübersetzern und technischen Redakteuren. In: Hennig, Jörg & Tjarks-Sobhani, Marita (Hrsg.) *tekomp HOCHSCHULSCHRIFTEN*. Bd. 20. Stuttgart: tcworld.
- Jasińska, Katarzyna (2023). Trenowanie sztucznej inteligencji a naruszenie praw autorskich. Aspekty dowodowe. *PRAWO i WIEŻ* 4 (47), 419–434.
- Ji, Luning; Sum, Mantai; Lu, Qin; Li, Wenjie & Chen, Yirong (2007). Chinese terminology extraction using window-based contextual information. In: Gelbukh, Alexander (Hrsg.) *Computational linguistics and intelligent text processing: 8th International Conference, CICLing 2007, Mexico City, Mexico, February 18-24, 2007. Proceedings*. Berlin/Heidelberg: Springer, 62–74.
- Jurafsky, Dan & Martin, James H. (2024). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models*. 3. Aufl. Upper Saddle River, NJ: Prentice Hall Prentice-Hall International.
- Justeson, John S. & Katz, Slava M. (1995). Technical terminology: some linguistic properties and an algorithm for identification in text. *Natural language engineering* 1, 9–27.
- Kageura, Kyo & Umino, Bin (1996). Methods of automatic term recognition: A review. *International Journal of Theoretical and Applied Issues in Specialized Communication* 3 (2), 259–289.
- Kalverkämper, Hartwig (1998). Fachsprache und Fachsprachenforschung. In: Hoffmann, Lothar, Wiegand, Herbert Ernst, Galinski, Christian, Hüllen, Werner & Kalverkämper, Hartwig (Hrsg.) *Fachsprachen / Languages for Special Purposes. Halbband 1. (Handbücher zur Sprach- und Kommunikationswissenschaft 14.1.)*. Berlin/New York: Walter de Gruyter, 48–59.

- Katz, Slava M. (1987). Estimation of probabilities from sparse data for the language model component of a speech recognizer. *IEEE transactions on acoustics, speech, and signal processing* 35, 400–401.
- Kavila, Selvani Deepthi; Rajesh, B.; Vyshnavi, N. & Deep, K. Moni Sushma (2017). Automatic Key Term Extraction from Research Article using Hybrid Approach. *International Journal of Computer Applications* 166 (6), 17–21.
- Kęsicka, Karolina (2020). *Die staatliche Prüfung für beeidigte ÜbersetzerInnen und DolmetscherInnen. Das deutsche und polnische Strafrecht. Theorie und Übungsmaterial.* Warszawa: C.H.Beck.
- Kilian, Alina & Kilian, Agnieszka (2021a) *Słownik języka prawniczego i ekonomicznego: Niemiecko-polski.* 4. Aufl. Warszawa: Wolters Kluwer Polska.
- Kilian, Alina & Kilian, Agnieszka (2021b) *Słownik języka prawniczego i ekonomicznego: Polsko-niemiecki.* 4. Aufl. Warszawa: Wolters Kluwer Polska.
- Kirchhof, Paul (1977). Rechtsänderungen durch geplanten Sprachgebrauch? In: Wilke, Dieter & Weber, Harald (Hrsg.) *Gedächtnisschrift für Fridrich Klein.* München: Franz Vahlen, 227–244.
- Knauer, Babette (2007). *Gesellschaftsvertrag Deutsch-Englisch: die Besonderheiten der englischen und deutschen Rechtssprache und Probleme der juristischen Übersetzung am Beispiel einer Übersetzung eines Gesellschaftsvertrages (Diplomarbeit).* Berlin: Bundesverband der Dolmetscher und Übersetzer e.V. (BDÜ).
- Kocoń, Jan; Cichecki, Igor; Kaszyca, Oliwier; Kochanek, Mateusz; Szydło, Dominika; Baran, Joanna; Bielaniewicz, Julita; Gruza, Marcin; Janz, Arkadiusz; Kanclerz, Kamil; Kocoń, Anna; Koptyra, Bartłomiej; Mieszczenko-Kowszewicz, Wiktoria; Miłkowski, Piotr; Oleksy, Marcin; Piasecki, Maciej; Radliński, Łukasz; Wojtasik, Konrad; Woźniak, Stanisław & Kazienko, Przemysław (2023). ChatGPT: Jack of all trades, master of none. *Information Fusion* 99, 1–37.
- Konertz, Roman (2023). Urheberrechtliche Fragen der Textgenerierung durch Künstliche Intelligenz: Insbesondere Schöpfungen und Rechtsverletzungen durch GPT und ChatGPT. *Wettbewerb in Recht und Praxis* 69 (7), 796–804.
- Korzynski, Pawel; Mazurek, Grzegorz; Krzypkowska, Pamela & Kurasinski, Artur (2023). Artificial intelligence prompt engineering as a new digital competence: Analysis of generative AI technologies such as ChatGPT. *Entrepreneurial Business and Economics Review* 11 (3), 25–37.

- Kreutz, Peter (2020). Autonomes Fahren: Produkt und Produzentenhaftung. In: Oppermann, Bernd H. & Stender-Vorwachs, Jutta (Hrsg.) *Autonomes Fahren: Rechtsfolgen, Rechtsprobleme, technische Grundlagen*. 2. Aufl. München: C.H. Beck, 195–199.
- Kreutzer, Ralf T. (2023). *Künstliche Intelligenz verstehen: Grundlagen – Use-Cases – unternehmenseigene KI-Journey*. 2. Aufl. Wiesbaden: Springer Gabler.
- Krüger, Ralph (2021). Die Transformer-Architektur für Systeme zur neuronalen maschinellen Übersetzung -eine popularisierende Darstellung. *Trans-kom* 14, 278–324.
- KÜDES: Konferenz der Übersetzungsdienste europäischer Staaten (2018). *Empfehlungen für die Terminologearbeit*. 3. Aufl. Bern: Schweizerische Bundeskanzlei.
- Kuś, Iwona & Senda, Zofia (2004). *Prawo autorskie i prawa pokrewne. Poradnik przedsiębiorcy. Zmiany prawa w związku z dostosowaniem przepisów do prawa Unii Europejskiej*. Radom: Polska Agencja Rozwoju Przedsiębiorczości.
- Kyrychenko, Roman & Grishina, Anastasia (2024). *GPT-3 vs. BERT: Ending The Controversy*. <https://softteco.com/blog/bert-vs-chatgpt?WPACRandom=1722121622048> (Stand: 28.07.2024).
- Le, Tan & Sadat, Fatiha (2021). Multilingual Automatic Term Extraction in Low-Resource Domains. *The International FLAIRS Conference Proceedings* 34, 1–5.
- L’Homme, Marie-Claude; Benali, Loubna; Bertrand, Claudine & Laudique, Patricia (1996). Definition of an evaluation grid for term-extraction software. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 3 (2), 291–312.
- Li, Xue; Xiao, Ziyang; Tang, Yiyi; Zhang, Qingpeng; Chan, Esther W. & Wong, Ian CK (2024). Large Language Model in Medical Information Extraction from Titles and Abstracts with Prompt Engineering Strategies: A Comparative Study of GPT-3.5 and GPT-4. *medRxiv*, 1–27.
- Libor, Christine (2022). Kein Copyright für KI-Kunst. *Zeitschrift für das gesamte Medienrecht* 2, 124.
- Liu, Yinhan; Ott, Myle; Goyal, Naman; Du, Jingfei; Joshi, Mandar; Chen, Danqi; Levy, Omer; Lewis, Mike; Zettlemoyer, Luke & Stoyanov, Veselin (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. *CoRR*, 1–13.
- López-Rodríguez, Clara Inés & Sánchez-Cárdenas, Beatriz (2021). *Theory and digital resources for the English-Spanish medical translation industry*. Berlin: Peter Lang.
- Lucchi, Nicola (2023). ChatGPT: a case study on copyright challenges for generative artificial intelligence systems. *European Journal of Risk Regulation*, 1–23.

- Łukasik, Marek (2023). Corpus linguistics and generative AI tools in term extraction: a case of Kashubian—a low-resource language. *Applied Linguistics Papers* 27 (4), 34–45.
- Łuksza, Serafina (2006). Zarys specyfiki zapożyczeń we współczesnym rosyjskim i polskim języku prawnym. *Acta Polono-Ruthenica* (11), 373–378.
- Lyding, Verena & Ties, Isabella (2008). Computerlinguistische Anwendungen zur Nutzung normierter terminologischer Daten. In: Chiocchetti, Elena & Voltmer, Leonhard (Hrsg.) *Normazione, armonizzazione e pianificazione linguistica*. Bozen: EURAC, 75–95.
- Maamar, Niklas (2021). *Computer als Schöpfer: der Schutz von Werken und Erfindungen künstlicher Intelligenz*. Tübingen: Mohr Siebeck.
- Malinowski, Andrzej (2006). *Polski język prawny. Wybrane zagadnienia*. Warszawa: LexisNexis.
- Mamakas, Dimitris; Tsotsi, Petros; Androutsopoulos, Ion & Chalkidis, Ilias (2022). Processing long legal documents with pre-trained transformers: Modding legalbert and longformer. *Association for Computational Linguistics*, 130–142.
- Manning, Christopher D. & Schütze, Hinrich (1999). *Foundations of statistical natural language processing*. Cambridge, Mass. [u.a.]: MIT Press.
- Marciniak, Malgorzata; Mykowiecka, Agnieszka & Rychlik, Piotr (2016). Termopl-a flexible tool for terminology extraction. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 2278–2284.
- Marciniak, Malgorzata; Rychlik, Piotr & Mykowiecka, Agnieszka (2023). TermoUD-a language-independent terminology extraction tool. *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, 178–186.
- Mayer, Felix (2001). Terminologie in der Fachübersetzung. In: Mayer, Felix (Hrsg.) *Dolmetschen und Übersetzen: der Beruf im Europa des 21. Jahrhunderts: Akten des Kongresses des BDÜ-Landesverbandes Bayern, 23. - 25. November 2001, München*. Freiburg: Freigang, Mauro + Reinke, 164–168.
- Mayer, Felix (2009). Terminologielehre und Terminologiemanagement. In: Mayer, Felix & Seewald-Heeg, Uta (Hrsg.) *Terminologiemanagement: von der Theorie zur Praxis*. Berlin: Bundesverb. der Dolmetscher und Übers., 12–26.
- Maynard, Diana & Ananiadou, Sophia (2001). Term extraction using a similarity-based approach. In: Jacquemin, Christian., L'Homme, Marie-Claude. & Bourigault, Didier (Hrsg.) *Recent Advances in Computational Terminology*. Philadelphia: John Benjamins Publishing Company.

- Mazzi, Francesca (2024). Authorship in artificial intelligence-generated works: Exploring originality in text prompts and artificial intelligence outputs through philosophical foundations of copyright and collage protection. *The Journal of World Intellectual Property*, 1–18.
- McCrae, John Philip & Doyle, Adrian (2019). Adapting Term Recognition to an Under-Resourced Language: The Case of Irish. *Proceedings of the Celtic Language Technology Workshop*, 48–57.
- McMullen, Matthew (2023). *The emergence of prompt engineers: The next in-demand role in AI*. <https://www.datasciencecentral.com/the-emergence-of-prompt-engineers-the-next-in-demand-role-in-ai/> (Stand: 27.10.2024).
- Mima, Hideki & Ananiadou, Sophia (2000). An application and evaluation of the C/NC-value approach for the automatic term recognition of multi-word units in Japanese. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 6 (2), 175–194.
- Ministerstwo Sprawiedliwości (2022). *Portal Orzeczeń Sądów Powszechnych*. [https://orzeczenia.ms.gov.pl/content/\\$N/154500000003527_VII_AGa_000851_2021_Uz_2022-04-22_003](https://orzeczenia.ms.gov.pl/content/$N/154500000003527_VII_AGa_000851_2021_Uz_2022-04-22_003) (Stand: 07.11.2024).
- Ministerstwo Sprawiedliwości (2023a). *Portal Orzeczeń Sądów Powszechnych*. [https://orzeczenia.ms.gov.pl/content/\\$N/151015000008527_XVII_GW_000121_2021_Uz_2023-10-18_001](https://orzeczenia.ms.gov.pl/content/$N/151015000008527_XVII_GW_000121_2021_Uz_2023-10-18_001) (Stand: 07.11.2024).
- Ministerstwo Sprawiedliwości (2023b). *Portal Orzeczeń Sądów Powszechnych*. [https://orzeczenia.ms.gov.pl/content/\\$N/151520000009627_XXIV_GWo_000498_2023_Uz_2023-10-25_002](https://orzeczenia.ms.gov.pl/content/$N/151520000009627_XXIV_GWo_000498_2023_Uz_2023-10-25_002) (Stand: 07.11.2024).
- Ministerstwo Sprawiedliwości (2024). *Portal Orzeczeń Sądów Powszechnych*. <https://orzeczenia.ms.gov.pl> (Stand: 26.10.2024).
- Mishra, Siba & Sharma, Arpit (2020). Automatic Word Embeddings-Based Glossary Term Extraction from Large-Sized Software Requirements. In: Madhavji, Nazim, Pasquale, Lili-ana, Ferrari, Alessio & Gnesi, Stefania (Hrsg.) *Requirements Engineering: Foundation for Software Quality. REFSQ 2020. Lecture Notes in Computer Science*. Cham: Springer International Publishing, 203–218.
- Möhn, Dieter & Pelka, Roland (1984). *Fachsprachen: eine Einführung*. Tübingen: Niemeyer.
- Moorkens, Joss & Lewis, Dave (2019). Copyright and the re-use of translation as data. In: O'Hagan, Minako (Hrsg.) *The Routledge Handbook of Translation and Technology*. London: Routledge, 469–481.

- Moroz, Andrzej (2020). Terminologia prawna w aspekcie normatywnym. *Białostockie Archiwum Językowe* (20), 161–176.
- Mouratidis, Despoina; Mathe, Eirini; Voutos, Yorghos; Stamou, Klio; Kermanidis, Katia Lida; Mylonas, Phivos & Kanavos, Andreas (2022). Domain-specific term extraction: a case study on Greek Maritime legal texts. *Proceedings of the 12th Hellenic Conference on Artificial Intelligence*, 1–6.
- Mügge, Uwe (2007). Automatische Terminologieextraktion. In: Schmitt, Peter A. & Jüngst, Heike E. (Hrsg.) *Translationsqualität. Leipziger Studien zur angewandten Linguistik und Translatologie*. Frankfurt: Peter Lang, 712–726.
- Müller, Elke (2010). *Sprache - Recht - Übersetzen: Betrachtungen zur juristischen Fachkommunikation: mit einer Darstellung am Beispiel von deutschen und spanischen Strafurteilen*. Hamburg: Kovač.
- Nazarenko, Adeline & Zargayouna, Haifa (2009). Evaluating term extraction. *International Conference Recent Advances in Natural Language Processing (RANLP'09)*, 299–304.
- Olejniczak, Adam W. (2018). Język prawny w prawie polskim. Ustawodawca dla profesjonalistów czy obywateli? In: Jakuszewicz, Adam (Hrsg.) *Język i prawo*. Bydgoszcz: Wydawnictwo Uniwersytetu Kazimierza Wielkiego, 27–46.
- Oliver, Antoni (2017). A system for terminology extraction and translation equivalent detection in real time: Efficient use of statistical machine translation phrase tables. *Machine Translation* 31 (3), 147–161.
- OpenAI (2024). *ChatGPT*. <https://chatgpt.com> (Stand: 13.10.2024).
- OPENAI CHATGPT (2024a). *ChatGPT-Antwort an M. S. auf „Bist du der Urheber / die Urheberin der von dir erzeugten Inhalte?“ (persönliche Mitteilung, 07.10.2024)*.
- OPENAI CHATGPT (2024b). *ChatGPT-Antwort an M. S. auf „Könnten Sie bitte alle Fachtermini aus dem folgenden Text extrahieren? Das wäre sehr nett von Ihnen“ (persönliche Mitteilung, 18.11.2024)*.
- Padilla, Rafael; Netto, Sergio L. & Da Silva, Eduardo AB (2020). A survey on performance metrics for object-detection algorithms. *2020 international conference on systems, signals and image processing (IWSSIP)*, 237–242.
- Pardey, Charlotte (2021). *Sekundenschnell erfunden*. <https://www.forschung-und-lehre.de/recht/sekundenschnell-erfunden-4015> (Stand: 10.10.2024).
- Pazienza, Maria Teresa (1998). A domain-specific terminology-extraction system. *Terminology* 5, 183–201.

- Periñán-Pascual, Carlos & Mestre-Mestre, Eva M. (2015). DEXTER: automatic extraction of domain-specific glossaries for language teaching. *Procedia-Social and Behavioral Sciences* 198, 377–385.
- Pikor, Gabriela Anna (2023). Rechtssprache im deutsch-polnischen Kontrast. *SŁOWO. Studia językoznawcze* 14, 173–189.
- Premasiri, Damith; Haddad, Amal Haddad; Ranasinghe, Tharindu & Mitkov, Ruslan (2023). Deep Learning Methods for Identification of Multiword Flower and Plant Names. *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing*, 879–887.
- Prill, Jonathan (2024). *Künstliche Intelligenz und Versicherung: Eine zivil-, datenschutz- und versicherungsrechtliche Analyse des bestehenden und geplanten Rechts mit Blick auf Chancen und Hürden von KI für die Versicherungswirtschaft*. Karlsruhe: VVW Imprint: VVW.
- Radford, Alec; Narasimhan, Karthik; Salimans, Tim & Sutskever, Ilya (2018). Improving language understanding with unsupervised learning. *OpenAI blog*, 1–12
- Radford, Alec; Wu, Jeffrey; Child, Rewon; Luan, David; Amodei, Dario & Sutskever, Ilya (2019). Language models are unsupervised multitask learners. *OpenAI blog* 1 (8), 1–12.
- Ralli, Natascia & Andreatta, Norbert (2018). Bistro – ein Tool für mehrsprachige Rechtsterminologie. *Trans-kom - Journal of Translation and Technical Communication Research* 11, 7–44.
- Rat für Deutschsprachige Terminologie (RaDT) (2015). *Terminologie – Grundlage für Fachkommunikation, Fachinformation, Fachwissen*. 3. Aufl. Bern: Staatskanzlei des Kantons Bern.
- Reuther, Ursula (2008). Terminologiegewinnung und kontrollierte Anwendung von Terminologie. In: Hennig, Jörg & Tjarks-Sobhani, Marita (Hrsg.) *Terminologiewerk für Technische Dokumentation*. Lübeck: Schmidt-Römhild.
- Rigouts Terryn, Ayla; Hoste, Veronique; Drouin, Patrick & Lefever, Els (2020). Termeval 2020: Shared task on automatic term extraction using the annotated corpora for term extraction research (acter) dataset. *6th International Workshop on Computational Terminology (COMPUTERM 2020)*, 85–94.
- Rigouts Terryn, Ayla; Hoste, Véronique & Lefever, Els (2021). HAMLET: hybrid adaptable machine learning approach to extract terminology. *Terminology* 27 (2), 290–328.
- Rijsbergen, Cornelis Joost van (1975). *Information retrieval*. London [u.a.]: Butterworths.

- Ring, Gerhard; Kiefel, Sebastian & Möller-Klapperich, Julia (2021). *Urheberrecht*. Baden-Baden: Nomos Verlagsgesellschaft mbH & Co. KG.
- Roelcke, Thorsten (2022). Mehr- und vielsprachige berufliche Kommunikation – ein Modell. *Informationen Deutsch als Fremdsprache* 49, 594–604.
- Roelcke, Thorsten D. (2020). *Fachsprachen*. 4. Aufl. Berlin: Erich Schmidt Verlag.
- Röhner, Jessica & Schütz, Astrid (2020). *Psychologie der Kommunikation*. 3. Aufl. Berlin/Heidelberg: Springer Berlin Heidelberg.
- Rokas, Aivaras; Rackevičienė, Sigita & Utkas, Andrius (2020). Automatic extraction of lithuanian cybersecurity terms using deep learning approaches. In: Utkas, Andrius, Vaičenonienė, Jurgita, Kovalevskaitė, Jolanta & Kalinauskaitė, Danguolė (Hrsg.) *Human Language Technologies. The Baltic Perspective: Proceedings of the Ninth International Conference Baltic HLT 2020*. Amsterdam/Washington: IOS Press, 39–46.
- Rothman, Denis & Gulli, Antonio (2022). *Transformers for Natural Language Processing: Build, Train, and Fine-Tune Deep Neural Network Architectures for NLP with Python, Hugging Face, and OpenAI's GPT-3, ChatGPT, and GPT-4*. 2. Aufl. Birmingham: Packt Publishing.
- Rzecznik Praw i Wolności Akademickich. Uniwersytet im. Adama Mickiewicza w Poznaniu (2024). *Wytyczne językowe*. <https://bezdyskryminacji.amu.edu.pl/plan-rownosci-plci/wytyczne-jezykowe> (Stand: 22.09.2024).
- Šajatović, Antonio; Buljan, Maja; Šnajder, Jan & Bašić, Bojana Dalbelo (2019). Evaluating automatic term extraction methods on individual documents. *Proceedings of the Joint Workshop on Multiword Expressions and WordNet (MWE-WN 2019)*, 149–154.
- Salton, Gerard M.; Wong, Andrew & Yang, Chungshu (1975). A vector space model for automatic indexing. *Communications of the ACM* 18 (11), 613–620.
- Sander, Gerald G. (2004). *Deutsche Rechtssprache: ein Arbeitsbuch*. Tübingen/Basel: Francke.
- Sandrini, Peter (1996). *Terminologiarbeit im Recht: deskriptiver begriffsorientierter Ansatz vom Standpunkt des Übersetzers*. Wien: TermNet - Internat. Network for Terminology.
- Sandrini, Peter (1999). *Übersetzen von Rechtstexten: Fachkommunikation im Spannungsfeld zwischen Rechtsordnung und Sprache*. Tübingen: Narr.
- Sandrini, Peter (2014). Open Translation Data: Neue Herausforderung oder Ersatz für Sprachkompetenz? In: Ortner, Heike, Pfurtscheller, Daniel & Rizzolli, Michaela (Hrsg.) *Datenflut und Informationskanäle*. Innsbruck: Innsbruck University Press, 185–194.
- Sauron, Véronique Anne (2002). Tearing out the terms: evaluating terms extractors. *Proceedings of Translating and the Computer* 24, 1–18.

- Schmidt, Jürgen (1972). Einige Bemerkungen zur Präzision der Rechtssprache. In: Albert, Hans, Luhmann, Niklas, Maihofer, Werner & Weinberger, Ota (Hrsg.) *Rechtstheorie als Grundlagewissenschaft der Rechtswissenschaft*. Bd. 2. Düsseldorf: Bertelsmann Verlag, 390–438.
- Schmidt, Wilhelm (1969). Charakter und gesellschaftliche Bedeutung der Fachsprachen. *Sprachpflege* 18 (1969), 10–21.
- Schubert, Klaus (2007). *Wissen, Sprache, Medium, Arbeit: ein integratives Modell der ein- und mehrsprachigen Fachkommunikation*. Tübingen: Narr.
- Schubert, Klaus (2009). Kommunikationsoptimierung: Vorüberlegungen zu einem fachkommunikativen Forschungsfeld. *Trans-kom. Zeitschrift für Translationswissenschaft und Fachkommunikation* 2 (1), 109–150.
- Seemann, Michael (2023). Künstliche Intelligenz, Large Language Models, ChatGPT und die Arbeitswelt der Zukunft. In: *Working Paper Forschungsförderung*. Bd. 304. Düsseldorf: Hans-Böckler-Stiftung, 1–100.
- Seewald, Uta (1993). Traitement morphosémantique des noms composés de structure NN à partir de l'analyse morphosémantique des mots simples. *TAL. Traitement automatique des langues* 34 (2), 75–88.
- Simon, Heike (2022). *Deutsche Rechtssprache: ein Studien- und Arbeitsbuch mit Einführung in das deutsche Recht*. 7. Aufl. München: C.H. Beck.
- Slancar, Helene (2024). „Verknallt in einen Talahon“ und Ben Gaya: Wie KI-Musik in die Charts drängt. *STANDARD Verlagsgesellschaft m.b.H.*, 2024. <https://www.derstandard.at/story/3000000232997/trendige-ki-musik-von-ben-gaya-und-butterbros-verknallt-in-einen-talahon> (Stand: 08.10.2024).
- Spasic, Irena; Owen, David; Knight, Dawn & Artemiou, Andreas (2019). Unsupervised multiword term recognition in Welsh. *Machine Translation Summit XVII. Proceedings of the Celtic Language Technology Workshop 2019*, 1–6.
- Staudacher, Klaus & Nida-Rümelin, Julian (2020). *Philosophische Überlegungen zur Verantwortung von KI. Eine Ablehnung des Konzepts der E-Person*. München: bidt – Bayerisches Forschungsinstitut für Digitale Transformation.
- Strehlow, Richard A. (2001). The role of terminology in retrieving information. In: Wright, Sue Ellen. & Budin, Gerhard. (Hrsg.) *Handbook of Terminology Management: Volume 2: Application-Oriented Terminology Management*. Amsterdam: John Benjamins Publishing Company, 426–441.

- Studienservice und Lehrwesen der Universität Wien (2024). *Zulassungsfrist*. https://studieren.univie.ac.at/begriffs-abc/verwandte-eintraege/?tx_irfaq_pi1%5BshowUid%5D=31354&cHash=aa5f3a1949060958bf97df534af462a6 (Stand: 22.09.2024).
- Teubert, Wolfgang (2006). Language as an economic factor: the importance of terminology. In: Barnbrook, Geoff, Danielsson, Pernilla & Mahlberg, Michaela (Hrsg.) *Meaningful Texts: The Extraction of Semantic Information from Monolingual and Multilingual Corpora*. London: Bloomsbury Publishing Plc, 96–106.
- Thakur, Kutub; Barker, Helen G. & Pathan, Al-Sakib Khan (2024). *Artificial intelligence and large language models: an introduction to the technological future*. Boca Raton, FL: CRC Press.
- The Gorilla Foundation / Koko.org (2024). *Art by Gorillas*. <https://www.koko.org/product-category/gorilla-art/> (Stand: 06.10.2024).
- Thormann, Isabelle (2016). *Rechtssprache: klar und verständlich für Dolmetscher, Übersetzer, Germanisten und andere Nichtjuristen*. Berlin: BDÜ Fachverlag.
- Thurmair, Gregor (2003). Making term extraction tools usable. *EAMT Workshop: Improving MT through other language technology tools: resources and tools for building MT*, 1–10.
- Tran, Hanh Thi Hong; Martinc, Matej; Caporusso, Jaya; Doucet, Antoine & Pollak, Senja (2023) The Recent Advances in Automatic Term Extraction: A survey. *ACM Computing Surveys Journal*, 1–25.
- Tran, Hanh Thi Hong; Martinc, Matej; Repar, Andraz; Ljubešić, Nikola; Doucet, Antoine & Pollak, Senja (2024). Can cross-domain term extraction benefit from cross-lingual transfer and nested term labeling? *Machine Learning* 113 (7), 4285–4314.
- Trosborg, Anna (1992). The performance of legal discourse. *HERMES-Journal of Language and Communication in Business* (9), 9–18.
- Tunstall, Lewis; Von Werra, Leandro & Wolf, Thomas (2022). *Natural language processing with transformers. Building Language Applications with Hugging Face*. Sebastopol, CA: O'Reilly Media, Inc.
- Ungerer, Friedrich (1992). Gläser, Rosemarie, „Fachtextsorten im Englischen“ (Book Review). *Anglia* 110, 464–468.

- Utkā, Andrius; Mockienė, Liudmila; Laurinaitis, Marius; Rackevičienė, Sigita; Rokas, Aivaras & Bielinskienė, Agnė (2022). Building of parallel and comparable cybersecurity corpora for bilingual terminology extraction. *Selected Papers from the CLARIN Annual Conference 2021 Virtual Event, 2021, 27–29 September*, 126–138.
- Valaski, Joselaine; Reinehr, Sheila & Malucelli, Andreia (2015). Approaches and Strategies to Extract Relevant Terms: How are they being applied? *Int'l Conf. Artificial Intelligence (ICAI'15)*, 478–484.
- Vaswani, Ashish; Shazeer, Noam; Parmar, Niki; Uszkoreit, Jakob; Jones, Llion; Gomez, Aidan N.; Kaiser, Łukasz & Polosukhin, Illia (2017). Attention is all you need. *Advances in neural information processing systems* 30, 1–11.
- Vintar, Spela (2010). Bilingual term recognition revisited: The bag-of-equivalents term alignment approach and its evaluation. *Terminology* 16, 141–158.
- Vivaldi, Jorge & Rodriguez, Horacio (2007). Evaluation of terms and term extraction systems: A practical approach. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 13 (2), 225–248.
- Voltmer, Leonhard; Streiter, Oliver; Zielinski, Daniel & Ties, Isabella (2003). *Termextraktion durch Beispielterme. Ansätze und Versuchsergebnisse für eine Minderheitensprache*. Bozen: Eurac Research, 1–25.
- Voorhees, Ellen M. (1999). Natural Language Processing and Information Retrieval. In: Pazi-enza, Maria Teresa (Hrsg.) *Information extraction: towards scalable, adaptable systems*. Berlin [u.a.]: Springer.
- Wachowiak, Lennart; Lang, Christian; Heinisch, Barbara & Gromann, Dagmar (2021). Towards learning terminological concept systems from multilingual natural language text. *3rd Conference on Language, Data and Knowledge (LDK 2021)* 93, 22:1-22:18.
- Wandtke, Artur-Axel & Ostendorff, Saskia (2023). *Urheberrecht*. Berlin/Boston: De Gruyter.
- Wank, Rolf (2020). *Juristische Methodenlehre: eine Anleitung für Wissenschaft und Praxis*. München: Verlag Franz Vahlen.
- Wassermann, Rudolf (1979). Sprachliche Mittel in der Kommunikation zwischen Fachleuten und Laien im Bereich des Rechtswesens. In: Mentrup, Wolfgang (Hrsg.) *Fachsprachen und Gemeinsprache*. Düsseldorf: Schwann, 114–124.
- Wei, Jason; Wang, Xuezhi; Schuurmans, Dale; Bosma, Maarten; Xia, Fei; Chi, Ed; Le, Quoc V. & Zhou, Denny (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems* 35, 24824–24837.

- Weller, Marion; Gojun, Anita; Heid, Ulrich; Daille, Béatrice & Harastani, Rima (2011). Simple methods for dealing with term variation and term alignment. In: Kageura, Kyo & Zweigenbaum, Pierre (Hrsg.) *Proceedings of the 9th International Conference on Terminology and Artificial Intelligence (TIA 2011)*, 87–93.
- Wiebe, Andreas; Appl, Clemens; Heidinger, Roman; Homar, Philipp; Sedef, Arzu & Winner, Martin (2022). *Wettbewerbs- und Immaterialgüterrecht: Patentrecht, Urheberrecht, Markenrecht, Musterschutzrecht, UWG, Kartellrecht*. 5. Aufl. Wien: facultas.
- Wiesmann, Eva (2007). Qualität der Rechtsübersetzung - der Beitrag von JUSLEX. In: Schmitt, Peter A. & Jüngst, Heike E. (Hrsg.) *Translationsqualität*. Bd. 5. Frankfurt am Main: Lang, 637–647.
- Wissler, Lars; Almashraee, Mohammed; Díaz, Dagmar Monett & Paschke, Adrian (2014). The Gold Standard in Corpus Annotation. *IEEE GSC* 21, 1–4.
- Witschel, Friedrich (2004). *Text, Wörter, Morpheme — Möglichkeiten einer automatischen Terminologie-Extraktion*. Diplomarbeit, Universität Leipzig.
- Witschel, Hans Friedrich (2005). Text, Wörter, Morpheme - Möglichkeiten einer automatischen Terminologie-Extraktion. In: Fisseni, Bernhard (Hrsg.) *Sprachtechnologie, mobile Kommunikation und linguistische Ressourcen: Beiträge zur GLDV-Tagung 2005 in Bonn*. Frankfurt am Main/Wien [u.a.]: Lang, 659–672.
- Wüster, Eugen (1974). Die Allgemeine Terminologielehre -ein Grenzgebiet Zwischen Sprachwissenschaft, Logik, Ontologie, Informatik und den Sachwissenschaften. *Linguistics* 12, 106.
- Wüster, Eugen (1991). *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie*. 3. Aufl. Bonn: Romanist. Verl.
- Yacouby, Reda & Axman, Dustin (2020). Probabilistic extension of precision, recall, and f1 score for more thorough evaluation of classification models. *Proceedings of the first workshop on evaluation and comparison of NLP systems*, 79–91.
- Yang, Zhilin; Dai, Zihang; Yang, Yiming; Carbonell, Jaime; Salakhutdinov, Ruslan & Le, Quoc V. (2019). XLNet: generalized autoregressive pretraining for language understanding. *Advances in neural Information processing Systems* 32, 1-18.
- Yenduri, Gokul; Ramalingam, M.; Selvi, G. Chemmalar; Supriya, Y.; Srivastava, Gautam; Maddikunta, Praveen Kumar Reddy; Raj, G. Deepti; Jhaveri, Rutvij H.; Prabadevi, B. & Wang, Weizheng (2024). Gpt (generative pre-trained transformer)—a comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. *IEEE Access*, 1–40.

- Zamfirescu-Pereira, JD; Wong, Richmond Y.; Hartmann, Bjoern & Yang, Qian (2023). Why Johnny can't prompt: how non-AI experts try (and fail) to design LLM prompts. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*, 1–21.
- Zeck, Michael (2021). KI im Unternehmen: Haftet der Arbeitgeber oder die KI als e-Person? In: Knappertsbusch, Inka & Gondlach, Kai (Hrsg.) *Arbeitswelt und KI 2030: Herausforderungen und Strategien Für Die Arbeit Von Morgen*. Wiesbaden: Springer Fachmedien Wiesbaden GmbH, 93–102.
- Zerfaß, Angelika (2006). Terminologieextraktion. *eDITion – Terminologiemagazin* 2, 21–25.
- Zhang, Jianzhang; Chen, Sisi; Hua, Jinping; Niu, Nan & Liu, Chuang (2022). Automatic terminology extraction and ranking for feature modeling. *2022 IEEE 30th International Requirements Engineering Conference (RE)*, 51–63.
- Zhang, Ziqi; Gao, Jie & Ciravegna, Fabio (2018). Semre-rank: Improving automatic term extraction by incorporating semantic relatedness with personalised pagerank. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 12 (5), 1–41.
- Zhang, Ziqi; Iria, José; Brewster, Christopher & Ciravegna, Fabio (2008). A comparative evaluation of term recognition algorithms. *LREC* 5, 2108–2113.
- Zhang, Ziqi; Petrak, Johann & Maynard, Diana (2018). Adapted textrank for term extraction: A generic method of improving automatic term extraction algorithms. *Procedia Computer Science* 137, 102–108.
- Zhao, Zihao; Wallace, Eric; Feng, Shi; Klein, Dan & Singh, Sameer (2021). Calibrate before use: Improving few-shot performance of language models. *International conference on machine learning*, 12697–12706.
- Zhong, Maosheng; Kuang, Jiangling; Liu, GanLin; Hua, Xin & Wu, Jiahua (2023). The Dynamic Vectors-Based Attention Model for Chinese Mathematical Term Extraction. *IEEE Intelligent Systems* 38 (3), 12–20.
- Zhu, Wen; Zeng, Nancy & Wang, Ning (2010). Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. *NESUG proceedings: health care and life sciences, Baltimore, Maryland* 19, 1–9.
- Zia-Ul-Haq, Muhammad (2023). ChatGPT & Copyright: Key Legal and Moral Implications. *Bulletin of the Transilvania University of Braşov, Series VII: Social Sciences and Law* 16, 151–158.

- Zielinski, Daniel & Ramírez-Safar, Yamile (2005). *Research meets practice: t-survey 2005: An online survey on terminology extraction and terminology management*. In: *Proceedings of the ASLIB International Conference Translating & the Computer 27*, 1–27.
- Zipf, George Kingsley (1949). *Human behavior and the principle of least effort: an introduction to human ecology*. Cambridge, Mass.: Addison-Wesley.
- Zoll, Fryderyk (1926). *Polska ustawa o prawie autorskiem i Konwencja Berneńska*. Warszawa: Gebethner i Wolff.

Anhangsverzeichnis

Anhang 1	173
Anhang 2	174
Anhang 3	176
Anhang 4	178
Anhang 5	179
Anhang 6	181
Anhang 7	183
Anhang 8	186
Anhang 9	189
Anhang 10	195
Anhang 11	201
Anhang 12	203
Anhang 13	205
Anhang 14	207
Anhang 15	210
Anhang 16	213

Anhang 1

[1] Der Kläger beobachtete zufällig, wie bei einem Stromausfall mehrere Personen aus diversen Fahrgeschäften eines Vergnügungsparks befreit wurden. Er machte davon einige Fotos mit seinem Handy und veröffentlichte sie auf seinem Twitter Account.

[2] Ein Mitarbeiter des beklagten Rundfunks ersuchte den Kläger um Erlaubnis, die Bilder für einen Bericht über den Stromausfall in einer Nachrichtensendung verwenden zu dürfen. Der Kläger erteilte die Erlaubnis nicht, weil die Nutzung unentgeltlich erfolgen sollte. Der Beklagte sendete die Bilder dennoch und stellte seine Nachrichtensendung auch zum Abruf auf seiner Website bereit.

[3] Der Kläger begehrte Unterlassung, Urteilsveröffentlichung und 4.000 EUR als angemessenes Entgelt nach § 86 UrhG.

[4] Die Vorinstanzen gaben der Klage statt. Der Beklagte könne sich nicht auf ein freies Werknutzungsrecht nach § 42c UrhG berufen, weil die Fotos des Klägers nicht bei der Berichterstattung über ein Tagesereignis wahrnehmbar geworden seien. Außerdem fehle es am Informationszweck, weil ein Bericht über den Stromausfall im Bezirk auch ohne die Lichtbilder des Klägers möglich gewesen wäre.

Rechtliche Beurteilung

[5] Mit seiner außerordentlichen Revision will der Beklagte die Klagsabweisung erreichen. Sie zeigt keine erhebliche Rechtsfrage auf.

[6] 1. Gemäß § 42c UrhG dürfen zur Berichterstattung über Tagesereignisse Werke, die bei Vorgängen, über die berichtet wird, öffentlich wahrnehmbar werden, in einem durch den Informationszweck gerechtfertigten Umfang vervielfältigt, verbreitet, durch Rundfunk gesendet, der Öffentlichkeit zur Verfügung gestellt und zu öffentlichen Vorträgen, Aufführungen und Vorführungen benutzt werden.

[7] Ob die Voraussetzungen für diese freie Werknutzung vorliegen, richtet sich grundsätzlich nach den Umständen des Einzelfalls und bildet im Allgemeinen keine Rechtsfrage von erheblicher Bedeutung iSd § 502 ZPO (RS0108466; vgl. zuletzt 4 Ob 58/23t Rz 5).

[8] 2. Der Beklagte argumentiert, die Richtlinie 2001/29/EG des Europäischen Parlaments und des Rates vom 22. Mai 2001 zur Harmonisierung bestimmter Aspekte des Urheberrechts und der verwandten Schutzrechte in der Informationsgesellschaft (InfoRL) gestatte die Nutzung von Werken oder sonstigen Schutzgegenständen bereits in Verbindung mit der Berichterstattung über Tagesereignisse. Eine Wahrnehmbarkeit bei der Berichterstattung wie nach § 42c UrhG sei nach dem Unionsrecht kein Erfordernis, die österreichische Judikatur sei deshalb durch die neuere Rechtsprechung des EuGH überholt. Der BGH habe seine Linie zum vergleichbaren § 50 dUrhG bereits geändert.

[9] 2.1. Tatsächlich können die Mitgliedstaaten gemäß Art 5 Abs 3 lit c zweiter Fall der InfoRL die Nutzung von Werken „in Verbindung mit der Berichterstattung über Tagesereignisse“ vorsehen, soweit es der Informationszweck rechtfertigt und sofern nach Möglichkeit der Umstände die Quelle, einschließlich des Namens des Urhebers angegeben wird.

[10] Diese Bestimmung der Richtlinie führt aber eindeutig keine Vollharmonisierung durch (so auch EuGH C 469/17 Rn 54 und C 516/17 Rn 27, 39). Daher spricht ihr Wortlaut auch nicht gegen die vom österreichischen Gesetzgeber geforderte Art der Verbindung, die überdies in Einklang mit Art 10bis Abs 2 der Berner Übereinkunft (Pariser Fassung) zum Schutz von Werken der Literatur und Kunst steht. Dessen Text lautet: „Ebenso bleibt der Gesetzgebung der Verbandsländer vorbehalten zu bestimmen, unter welchen Voraussetzungen anlässlich der Berichterstattung über Tagesereignisse durch Photographie oder Film oder im Weg der Rundfunksendung oder Übertragung mittels Draht an die Öffentlichkeit Werke der Literatur oder Kunst, die im Verlauf des Ereignisses sichtbar oder hörbar werden, in dem durch den Informationszweck gerechtfertigten Umfang vervielfältigt und der Öffentlichkeit zugänglich gemacht werden dürfen.“

[11] 2.2. In den vom Revisionswerber genannten Fällen [gemeint offenbar] C 516/17 Spiegel Online (= BGH I ZR 139/15, „Reformistischer Aufbruch“) sowie C 469/17 Funke Mediengruppe (= BGH I ZR 139/15, „Afghanistan Papiere“) hatte der EuGH Vorabentscheidungsersuchen des BGH zu § 50 dUrhG zu beantworten. Diese Norm lässt zur Berichterstattung über Tagesereignisse die Vervielfältigung, Verbreitung und öffentliche Wiedergabe von Werken, die im Verlauf dieser Ereignisse wahrnehmbar werden, in einem durch den Zweck gebotenen Umfang zu.

[12] Die nationale Regelung in Deutschland ist daher deutlich weiter als in Österreich, weil das freie Werknutzungsrecht nach § 42c UrhG nur für Werke gilt, die bei Vorgängen, über die berichtet wird, öffentlich wahrnehmbar werden.

[13] Entsprechend lässt sich nach gefestigter Rechtsprechung des Senats aus dieser Norm keine allgemeine Rechtfertigung für die Verwertung von Lichtbildern ableiten, die (selbst) Tagesereignisse zeigen oder damit in Zusammenhang stehen (4 Ob 53/19a [Pkt 1.1] mwN).

[14] 2.3. Der Vollständigkeit halber sei darauf hingewiesen, dass nicht nur die nationalen Normen sich unterscheiden, sondern auch die Sachverhalte der Ausgangsverfahren beim BGH keine Ähnlichkeit mit der hier vorliegenden

Sachverhaltskonstellation haben. In diesen beiden Fällen war das Tagesereignis nämlich die Existenz und Verbreitung bzw die konkrete Formulierung eines Textes, den die Medien im Zuge dieser Berichterstattung jeweils auch in voller Länge veröffentlichten. Anders als Lichtbilder des Klägers, die nur das Tagesereignis Stromausfall abbilden, waren also beim BGH die urheberrechtlich geschützten Werke selbst das Thema der Berichterstattung und damit das Tagesereignis iSd § 50 dUrhG.

[15] 3. Schließlich betont der Beklagte, dass es laut EuGH keine außerhalb der Art 5 Abs 2 und 3 InfoRL liegenden, verfassungsunmittelbaren freien Nutzungen geben könne (unter Verweis auf C 469/17 Rn 64 und C 516/17 Rn 49).

[16] Dieses Argument spielte aber jedenfalls für die klagsstattgebende Entscheidung des Berufungsgerichts keine Rolle. Auch die Revision stellt keinen erkennbaren Zusammenhang mit dem zu beurteilenden Sachverhalt her. Die Beantwortung abstrakter Rechtsfragen ist aber nicht Aufgabe des Obersten Gerichtshofs (RS0111271 [T2]; RS0002495).

[17] 4. Ob die weiteren Voraussetzungen für eine freie Werknutzung vorliegen, insbesondere ob der Informationszweck im vorliegenden Fall die freie Werknutzung rechtfertigt, kann daher dahinstehen.

Anhang 2

[9] Der Kläger begehrte restliches Entgelt (in bar und in Form eines Bildes) und erhob auf UWG und UrhG gestützte Feststellungsansprüche und Unterlassungsansprüche wegen Eingriffen in seine Verwertungsrechte am Film. Gegenstand des Revisionsverfahrens sind nur zwei Eventualbegehren gegenüber der Erstbeklagten, nämlich zum Einen das Begehren 4 a) auf Feststellung, dass der Erstbeklagten kein Werknutzungsrecht am Film zukomme, zum Anderen das im Verhältnis zu diesem hilfsweise erhobene Begehren 5), mit dem der Kläger beiden Beklagten untersagen lassen will, Dritte ohne seine Zustimmung zur Zahlung von Schadenersatz und/oder Nutzungsentgelt für den Film aufzufordern und/oder diesen Film gewerblich zu nutzen. Dem Kläger stünden als Urheber und Filmhersteller nämlich die ausschließlichen Verwertungsrechte am Film zu, er habe den Beklagten nur eine Werknutzungsbewilligung für die Dauer der Ausstellung erteilt. Er habe den Film mit weiteren Filmschaffenden an fünf Drehtagen verwirklicht, sich aber gegenüber der Auftraggeberin volle künstlerische Freiheit und kreative Freiheit als Drehbuchautor, Regisseur, Filmer und Cutter ausbedungen. Während des erstinstanzlichen Verfahrens erklärte der Kläger in einem Schriftsatz den Rückruf aller ausschließlichen Nutzungsrechte für den Fall, dass das Gericht von einem Werknutzungsrecht der Beklagten ausgehen sollte.

[10] Die Erstbeklagte wendete ein, dass ihr die Verwertungsrechte zustünden, weil sie die Filmherstellerin iSd § 38 UrhG sei. Außerdem kämen ihr auch laut dem Vertrag mit dem Kläger alle Nutzungsrechte, ausgenommen jener für eine kommerzielle Verwertung, zu. Der Film sei technisch und inhaltlich so mangelhaft, dass er in der derzeitigen Form nicht für die Promotion des Künstlers verwertet werden könne. Deshalb stehe dem Kläger weder Entgelt noch ein Rückrufrecht zu.

[11] Das Erstgericht wies alle hier relevanten Klagebegehren ab. Filmherstellerin sei die Erstbeklagte, sodass ihr die Verwertungsrechte gemäß § 38 Abs 1 UrhG zukämen. Außerdem habe der Kläger ihr vertraglich Werknutzungsrechte ausgenommen für TV und Kino eingeräumt.

[12] Das Berufungsgericht entschied rechtskräftig über alle Ansprüche gegen die Zweitbeklagte. Hinsichtlich der Zahlungsbegehren und Herausgabebegehren gegen die Erstbeklagte verwies es die Rechtssache zur Verfahrensergänzung ans Erstgericht zurück. Anders als das Erstgericht gab es aber dem Eventualbegehren 4 a) mit Teilurteil statt und stellte fest, dass der Erstbeklagten kein Werknutzungsrecht am Film zukomme. Über das gegenüber Begehren 4 a) nur hilfsweise gestellte Unterlassungsbegehren 5) entschied das Berufungsgericht deshalb nicht mehr. Rechtlich vertrat es die Ansicht, dass der Kläger und nicht die Erstbeklagte Hersteller des gewerbsmäßig hergestellten Films sei. Ihm kämen daher nach § 38 Abs 1 UrhG die originären Verwertungsrechte daran zu. Der Kläger habe der Erstbeklagten im Filmproduktionsvertrag zwar ein umfassendes Werknutzungsrecht mit Ausnahme der kommerziellen Verwertung wie Kino und TV eingeräumt, dieses jedoch später gemäß § 29 UrhG widerrufen. Der Widerruf könne nach der Rechtsprechung auch in einem Prozessschriftsatz erfolgen. Da die Erstbeklagte den Widerruf erst in der nächsten Verhandlung – und damit außerhalb der 14 Tages Frist – bestritten habe, könne das Gericht nicht mehr prüfen, ob die Voraussetzungen für den Widerruf vorgelegen hätten. Auf die Brauchbarkeit oder Mangelhaftigkeit des Filmwerks komme es daher nicht an.

[13] Die außerordentliche Revision der Erstbeklagten zielt auf eine Wiederherstellung der Klagsabweisung laut Spruchpunkten 4 a) und 5) im Ersturteil ab.

[14] Der Kläger beantragt in der ihm freigestellten Revisionsbeantwortung, die Revision zurückzuweisen, hilfsweise ihr nicht Folge zu geben.

[15] Die Revision ist zur Klarstellung der Rechtslage zulässig, aber nicht berechtigt.

[16] 1. Das zentrale Argument in der Revision der Erstbeklagten lautet, dass ein Widerruf des Verwertungsrechts bei gewerbsmäßig hergestellten Filmen ausgeschlossen sei.

[17] 1.1. Wird von einem Werknutzungsrecht ein dem Zwecke seiner Bestellung entsprechender Gebrauch überhaupt nicht oder nur in so unzureichendem Maße gemacht, dass wichtige Interessen des Urhebers beeinträchtigt werden, so kann dieser, wenn ihn kein Verschulden daran trifft, das Vertragsverhältnis, soweit es das Werknutzungsrecht betrifft, gemäß § 29 Abs 1 UrhG vorzeitig lösen. Die Bestimmung gewährt dem Urheber also eine Möglichkeit zur vorzeitigen Vertragsauflösung, das sogenannte Rückrufrecht bei Nichtausübung bzw wegen Nichtgebrauchs (Ciresa in Ciresa, Österreichisches Urheberrecht [23. Lfg 2023] § 40 UrhG Rz 8; Walter, Österreichisches Urheberrecht I [2008] Rz 1819).

[18] 1.2. Wie die Revision richtig aufzeigt, ist § 29 Abs 1 UrhG aber gemäß § 40 Abs 3 UrhG für Werknutzungsrechte an gewerbsmäßig hergestellten Filmwerken nicht anwendbar. Die Bestimmung nimmt diese Werkart also vom gesetzlichen Auflösungsrecht aus.

[19] Die Wirksamkeit einer vom Urheber abgegebenen Erklärung, das Vertragsverhältnis aufzulösen, kann nach § 29 Abs 4 UrhG jedoch nicht mehr bestritten werden, wenn der Werknutzungsberechtigte diese Erklärung nicht binnen 14 Tagen nach ihrem Empfang zurückweist.

[20] Damit stellt sich die Frage, wie das Spannungsverhältnis zwischen diesen beiden Normen aufzulösen ist.

[21] 1.3. Der historische Gesetzgeber beabsichtigte mit diesen Regelungen, den Urheber vor einer Unterdrückung seines Werks zu schützen (EB zur RV des UrhG 1936, abgedruckt in proLIBRIS, Urheberrechtsgesetz – Texte Materialien Judikatur [2011] 138). Dieser Gedanke ist heute unvermindert gültig (vgl Art 22 UrheberrechtsRL 2019/790).

[22] Im allseitigen Interesse sollen aber lange Schwebezustände durch den Rückruf vermieden werden, sodass ehestens volle Sicherheit über seine Wirksamkeit besteht. Entwürfe mit längerer Nachfrist oder Bestreitungsfrist lehnte der historische Gesetzgeber deshalb ausdrücklich ab, zumal der Werknutzungsberechtigte typischerweise von der Erklärung des Urhebers nicht überrascht werde und damit genügend Zeit zur Überlegung habe (EB zur RV des UrhG 1936, abgedruckt in proLIBRIS, Urheberrechtsgesetz – Texte Materialien Judikatur [2011] 138).

[23] Hat der Werknutzungsberechtigte es aber unterlassen, die Auflösungserklärung rechtzeitig zurückzuweisen, so darf er mit der Behauptung, sie sei aus irgendwelchen Gründen unwirksam, nicht mehr gehört werden (EB zur RV des UrhG 1936, abgedruckt in proLIBRIS, Urheberrechtsgesetz – Texte Materialien Judikatur [2011] 139).

[24] 1.4. Die ständige Rechtsprechung versteht die 14 tägige Frist des § 29 Abs 4 UrhG deshalb als Fallfrist, mit deren Versäumung der Verlust des Bestreitungsrechts kraft Gesetzes eintritt. Damit ist es dem Werknutzungsberechtigten auch nicht mehr möglich, die Wirksamkeit der Auflösungserklärung betreffende Fragen aufzurollen (4 Ob 158/09b).

[25] Im Hinblick auf die umfassende Formulierung in den Materialien („aus irgendwelchen Gründen“) ließ der Senat bisher nach Fristende Einwände aller Art nicht mehr zu. Konkret wurde die Verschweigung des Einwands angenommen, dass die nach § 29 Abs 2 UrhG erforderliche Nachfristsetzung unterblieben war (RS0077758), ebenso des Einwands, dass materielle Wirksamkeitsvoraussetzungen nicht gegeben seien, etwa der Auflösungsgrund oder die Passivlegitimation aufgrund eines Co Verlagsverhältnisses und sich den daraus ergebenden Beschränkungen der Lösungsmöglichkeit (4 Ob 113/09k).

[26] 1.5. Die mit den Ausnahmetatbeständen verknüpften Fragen, ob es sich bei einem Filmwerk um ein gewerblich hergestelltes handelt, oder ob für ein Auftragswerk iSd § 28 Abs 2 UrhG eine Verwertungspflicht bestand, betreffen ebenfalls die Wirksamkeit des Rückrufs wegen Nichtgebrauchs. Sie können – wie auch der vorliegende Fall zeigt – sowohl auf faktischer, als auch auf rechtlicher Ebene strittig sein und einer langwierigen Klärung bedürfen.

[27] Aus diesem Grund scheint es sachgerecht, dass der Werknutzungsberechtigte den Rückruf auch dann binnen der 14 tägigen Frist bestreiten muss, wenn er meint, dass der Rechterückruf aufgrund eines gesetzlichen Ausnahmetatbestands wie § 40 Abs 3 UrhG (gewerbliches Filmwerk) oder etwa § 30 Abs 1 UrhG (Auftragswerk ohne Verwertungspflicht) unzulässig sei. Insbesondere Hersteller und Verwerter gewerblich hergestellter Filme werden vielfach nämlich über ausreichende Rechtskenntnisse und gut organisierte Unternehmen verfügen und sind daher zumindest nicht schutzwürdiger als andere Werknutzungsberechtigte.

[28] 2. Die Erstbeklagte kritisiert die vom Berufungsgericht zitierte Entscheidung (4 Ob 113/09k), wonach der Rückruf auch in einem Schriftsatz erfolgen kann und verweist dazu auf eine kritische Glosse von Walter. Die Verfahrenspartei und ihr Vertreter hätten nämlich keinen Grund gehabt, sich sofort nach Erhalt des Schriftsatzes mit dem darin enthaltenen Vorbringen zu beschäftigen.

[29] 2.1. Der Senat hat bereits ausgesprochen, dass durch die Verbindung der Auflösungserklärung mit einer Prozesshandlung keine Verlängerung der gesetzlichen Zurückweisungsfrist als materiell rechtlicher Fallfrist eintritt. Da die Bestreitung des Rückrufs auch keiner Prozessklärung bedarf, ist eine (allenfalls längere) Frist für die nächste Prozesshandlung für eine Verschweigung nicht maßgeblich (vgl 4 Ob 113/09k).

[30] 2.2. Tatsächlich kritisierte Walter das Ergebnis im damaligen Fall (MR 2010, 93). Er begründete dies jedoch vor allem damit, dass in auf das UrhG gestützten Verfahren Anwaltszwang herrsche und der Beklagte vier Wochen Zeit habe, um einen Rechtsanwalt mit einer Klagebeantwortung zu betrauen. Unter diesen Umständen sei mit den Vorinstanzen anzunehmen, dass der Kläger einem Widerspruch gegen den Rechterückruf auch innerhalb der Frist zur Erstattung der Klagebeantwortung zugestimmt habe.

[31] Diese oder eine vergleichbare Fallkonstellation liegt hier jedoch nicht vor. Vielmehr erfolgte der Rückruf erst im Laufe des Verfahrens, sodass die Erstbeklagte bereits anwaltlich vertreten und damit gegenüber einem möglicherweise rechtsunkundigen Erklärungsempfänger außerhalb eines Zivilverfahrens sogar im Vorteil war.

[32] Nur der Vollständigkeit halber sei darauf verwiesen, dass die Entscheidung 4 Ob 113/09k im Schrifttum sonst offenbar ohne Kritik – allenfalls unter Hinweis auf die „Gefährlichkeit“ des Rückrufs für den Werknutzungsberechtigten – referiert wird.

Anhang 3

[6] Der Geschäftsführer der Klägerin habe die Illustrationen von ihrem etwas „dämmlichen“ Aussehen (Glubschäugen, heraushängenden Zungen und Micky Maus artigen Gliedmaßen) befreit und ihr Erscheinungsbild dem Zeitgeist angepasst. Er habe ihnen insbesondere ein cooles, freundlicheres, beschwingteres, unbeschwerteres sowie aufgeschlosseneres Aussehen mit schlanken (und daher eleganteren) Gliedmaßen verliehen und sie in eine zeitgemäße, humorvolle, sympathische Illustration verwandelt. Die Schokoladetafel habe er in 4 x 6 Schokostücke (statt bisher 3 x 4) zerlegt und das Gesicht in der Tafel oben (statt unten) und damit dort, wo man es erwarte, gezeichnet. Die neuen Illustrationen seien Werke im Sinn einer eigentümlichen geistigen Schöpfung im Rahmen einer Bearbeitung (§ 5 UrhG), an welchen der Urheber, nämlich der Geschäftsführer der Klägerin, dieser die Nutzungsrechte eingeräumt habe. Die Beklagte nutze die Illustrationen unberechtigt und verstoße damit gegen das der Klägerin eingeräumte Nutzungsrecht nach dem UrhG. Überdies habe die Beklagte die Illustrationen teilweise ohne Genehmigung der Klägerin abgeändert oder abändern lassen. Sollte der Illustration der urheberrechtliche Schutz versagt werden, stünde der Klägerin als Verwendungsanspruch gemäß § 1041 ABGB ein angemessenes Entgelt von 15.030 EUR zu.

[7] Die Beklagte wendete ein, die Klägerin sei nicht aktivlegitimiert; es stehe nicht fest, dass der Geschäftsführer der Klägerin Urheber der Illustrationen sei und ihr Nutzungsrechte daran eingeräumt habe. Den Illustrationen fehle die urheberrechtliche Werkqualität. Es handle sich dabei nur um technisch bedingte Vereinfachungen der bereits vorhandenen Illustrationen mit geringen Veränderungen ohne Schaffung von Neuem oder Originellem. So seien anstelle von offenen Augen Striche und anstelle einer komplexen Zunge ein Punkt gezeichnet und die Nase entfernt worden. Die Finger der bestehenden Maskottchen wären nur schwer zu animieren gewesen und seien deshalb durch Punkte oder tropfenartige Gebilde ersetzt worden.

[8] Das Erstgericht wies die Klage wegen mangelnder Aktivlegitimation der Klägerin ab. Es sei nämlich unklar, ob die Klägerin oder deren Geschäftsführer als Privatperson mit der Erstellung der Illustrationen beauftragt worden sei und ob der Geschäftsführer die Illustrationen im Rahmen seiner dienstlichen Verpflichtung geschaffen habe.

[9] Das Berufungsgericht gab der Klage mit Ausnahme eines Teils des Urteilsveröffentlichungsbegehrens mittels Teilurteils statt. Den Wert des Entscheidungsgegenstands bemaß es mit 30.000 EUR übersteigend und die ordentliche Revision erklärte es für nicht zulässig. Die Aktivlegitimation der Klägerin bejahte das Berufungsgericht schon aufgrund des auch im Namen des Urhebers und Geschäftsführers der Klägerin gezeichneten Mahnschreibens an die Beklagte, wonach er sowohl mit der Geltendmachung der urheberrechtlichen Ansprüche als auch mit der Einforderung eines Entgelts unter anderem für die Einräumung ausschließlicher und übertragbarer Nutzungsrechte an den gegenständlichen Illustrationen durch die Klägerin einverstanden sei, was seine Zustimmung zur Einräumung von Nutzungsrechten an die Klägerin durch ihn voraussetze. Das ergebe sich auch aus dem anwaltlichen Mahnschreiben. Der Klägerin stehe daher das Werknutzungsrecht an den in Rede stehenden Illustrationen zu. Den vom Geschäftsführer der Klägerin gezeichneten bzw. bearbeiteten Illustrationen käme auch urheberrechtlicher Schutz zu, weil sie individuelle Eigenart aufwiesen. Die wesentlichen Unterschiede zwischen den originalen Illustrationen und den vom Genannten bearbeiteten bestünden darin, dass er die Körper der Figuren selbst, auch wenn er ihre äußere Form beibehalten habe, durch die Reduktion von Schattierungen weniger plastisch, sondern flacher und damit schlanker gestaltet habe, wobei dieser Effekt bei der Schokoladetafel durch die Aufteilung in 4 x 6 statt 3 x 4 Stücke verstärkt worden sei. Außerdem habe er die Micky Maus-artigen Gliedmaßen sowie die großen runden Augen samt Pupillen vereinfacht, sodass die Figuren ihr comic artiges Aussehen und joviales Aussehen eingebüßt hätten. Aufgrund des in den Bearbeitungen weiterhin auffällig gestalteten großen lachenden Mundes wirkten die Figuren vergnügt und froh und strahlten Zufriedenheit aus. Damit seien eine Reihe von individuellen Gestaltungselementen nicht nur geringfügig verändert worden, sodass nicht mehr von einer bloßen Vervielfältigung, rein mechanischen Veränderung oder im Wesentlichen unveränderten Wiedergabe der Originale gesprochen werden könne, sondern es dem Urheber, auch wenn es darum gegangen sein sollte, die Figuren zur leichteren Animation bzw. technisch bedingt zu vereinfachen, gelungen sei, etwas Neues, Individuelles und Originelles und damit eine urheberrechtlich geschützte Bearbeitung iSd § 5 UrhG zu schaffen. Die Beklagte habe keine Werknutzungsrechte an der Bearbeitung erworben. Die Klägerin habe daher infolge Einräumung der Werknutzungsrechte an den gegenständlichen Bearbeitungen Anspruch auf Unterlassung (§ 81 UrhG), Beseitigung (§ 82 f UrhG), Rechnungslegung (§ 87a UrhG) und Urteilsveröffentlichung (§ 85 UrhG). Substanzierte Einwendungen

gegen die Berechtigung des Beseitigungs-, Rechnungslegungs- und Veröffentlichungsanspruchs habe die Beklagte nicht erhoben. Zu veröffentlichen sei aber nur der über das Unterlassungs-, Beseitigungs- und Veröffentlichungsbegehren ergangene Spruch, nicht aber der Ausspruch über das Rechnungslegungsbegehren. Das Berufungsgericht erließ daher ein Teilurteil, mit dem es dem Unterlassungs-, Beseitigungs- und Rechnungslegungsbegehren sowie dem Urteilsveröffentlichungsbegehren im genannten Umfang stattgab.

[10] Die Beklagte macht mit ihrer – auf gänzliche Klageabweisung, in eventu Aufhebung, gerichteten – außerordentlichen Revision geltend, es fehle an ausreichendem Tatsachenvorbringen und -substrat für den Rechteerwerb der Klägerin, der Beklagten seien aufgrund des Auftragszwecks sämtliche Nutzungsrechte eingeräumt worden und die von der Klägerin vorgenommenen Änderungen an den Figuren seien nur marginal und hätten nur dazu gedient, eine leichtere Übertragung in die Animation zu ermöglichen; schließlich sei die Veröffentlichungsermächtigung ohne Begründung und Interessenabwägung ausgesprochen worden.

[...]

[11] Die Revision ist in Ermangelung von erheblichen Rechtsfragen nicht zulässig.

[12] 1.1. Die Beklagte bekämpft die rechtliche Beurteilung des Berufungsgerichts, wonach von einer konkludenten Rechteeinräumung des Geschäftsführers der Klägerin an diese auszugehen sei. Dazu fehle es nämlich an einem konkreten Tatsachenvorbringen der Klägerin.

[13] 1.2. Dem ist entgegen zu halten, dass die Klägerin bereits in ihrer Klage vorgebracht hat, der Geschäftsführer habe alle seine Rechte an der gegenständlichen Bearbeitung an die Klägerin abgetreten und dies werde immer so gehandhabt, weil das Anfertigen von Illustrationen für Dritte das Geschäft der Klägerin sei.

[14] 1.3. Die Beurteilung dieses Vorbringens als ausreichend konkret ist im Hinblick darauf, dass derartige Rechteeinräumungen bei vergleichbaren Sachverhalten gang und gäbe sind, vertretbar. So hat der Senat in einer vergleichbaren Konstellation die konkludente Einräumung eines ausschließlichen Werknutzungsrechts an Lichtbildern durch den alleinigen und geschäftsführenden GmbH-Gesellschafter an diese im Rahmen seiner Tätigkeit für die Gesellschaft ebenfalls für vertretbar gehalten, zumal die GmbH sämtliche Nutzungsentgelte an Dritte verrechnete (4 Ob 23/15h). Auf die konkreten Umstände der Rechteeinräumung des Geschäftsführers an die Gesellschaft kam es dort und kommt es auch hier nicht an.

[15] 1.4. Die Klägerin forderte die Beklagte mittels eines Mahnschreibens auf, für die Nutzungsrechte ihrer Illustrationen den geforderten Geldbetrag zu leisten. Später kontaktierte sie die Beklagte mittels anwaltlichen Mahnschreibens erneut und forderte sie zur Unterzeichnung einer Unterlassungs- und Verpflichtungserklärung auf. Das Berufungsgericht hat bereits aus dem unter anderem mit dem Namen des Geschäftsführers (= Urhebers der in Rede stehenden Illustrationen) gezeichneten Mahnschreiben darauf geschlossen, dass dieser mit dem Inhalt und damit auch mit der Geltendmachung urheberrechtlicher Ansprüche durch die Klägerin einverstanden war, was seine Zustimmung zur Einräumung von Nutzungsrechten an diese voraussetzt. Darüber hinaus haben beide Gesellschafter, also auch der Urheber bzw Geschäftsführer die Klagevertreter mit der Verfolgung der Urheberrechtsansprüche der Klägerin beauftragt.

[16] 1.5. Soweit die Beklagte moniert, dass für die konkludente Rechteeinräumung eine Gegenleistung der Klägerin fehle, verkennt sie, dass ein Werknutzungsentgelt den Wert der Gesellschaftsanteile und damit indirekt auch das Vermögen des Gesellschafters erhöht. Daher ist es unerheblich, ob der Gesellschafter-Geschäftsführer für die Rechteeinräumung ein gesondertes Entgelt erhält (vgl 4 Ob 100/20i).

[17] 2.1. Die Beklagte erachtet die Wertung des Berufungsgerichts, wonach die vom Geschäftsführer der Klägerin vorgenommenen Bearbeitungen der gegenständlichen Illustrationen aufgrund ihrer individuellen Ausgestaltung urheberrechtlich geschützte Werke iSd § 5 Abs 1 UrhG seien, als unrichtig und gegen die ständige Rechtsprechung des Obersten Gerichtshofs verstoßend.

[18] 2.2. Werke iSd § 1 Abs 1 UrhG sind eigentümliche geistige Schöpfungen auf den Gebieten der Literatur, der Tonkunst, der bildenden Künste und der Filmkunst. Zu den Werken der bildenden Künste gehören auch solche des Kunstgewerbes (RS0076423), weil es auf den Zweck des Werks nicht ankommt. Erfasst werden auch Werke der Gebrauchsgrafik (RS0076187; 4 Ob 53/92, Lindwurm; 4 Ob 159/99g, Zimmermann FITNESS), wenn sie eine schöpferische Eigenart aufweisen. Rein handwerkliche, routinemäßige Leistungen sind aber auch bei Gebrauchsgrafiken nicht schutzfähig (RS0115332). Ebenfalls schutzunfähig ist der künstlerische Stil als solcher (RS0076695; RS0076734). Eine „Bearbeitung“ im Rechtssinn (§ 5 UrhG) ist die Umgestaltung äußerer Merkmale bei gleichzeitiger Identität des Werks, also eine – nicht rein mechanische, sondern aus eigener schöpferischer Gestaltungskraft entwickelte – Änderung der äußeren Form unter Beibehaltung des Kerns des Werks, nicht aber eine geringfügige Änderung oder Umgestaltung des Originals (RS0076389; vgl auch RS0076406). Sie lässt das Werk in seinem Wesen unberührt, muss ihm aber wenigstens in der äußeren Form eine neue Gestalt geben, die als eigentümliche geistige Schöpfung des Bearbeiters zu werten ist (RS0076413). Eine „Bearbeitung“ setzt damit ein Werk iSd § 1 Abs 1 UrhG voraus (vgl RS0076443). Als Bearbeitung darf das Produkt nur mit Einwilligung des Urhebers des bearbeiteten Werks oder dessen Werknutzungsberechtigten verwertet werden (vgl RS0125380).

Anhang 4

Pozwem z 1 kwietnia 2021 r. wniesionym przeciwko A. P. powód J. J., po ostatecznej modyfikacji roszczeń (oświadczenia pełnomocników powoda – k. 561 – 562), domagał się: zasądzenia kwoty 139.292,00 zł wraz z ustawowymi odsetkami za opóźnienie od 20 października 2020 r. do dnia zapłaty, zasądzenia od pozwanej na rzecz instytucji pożytku publicznego, tj. Fundacji (...)z siedzibą w G., kwoty 20.000 zł tytułem naruszenia przez pozwaną przysługujących mu osobistych praw autorskich, nakazania pozwanej zaniechania działań polegających na rozpowszechnianiu w sieci należących do powoda utworów w postaci wykonania wizualizacji fotorealistycznych inwestycji pod nazwą (...) w G. przy ul. (...) na działkach nr (...) (dalej jako Inwestycja) na portalu społecznościowym F. pozwanej, a także projektów aranżacji wnętrz lokali inwestycji na stronie internetowej pozwanej w zakładce zobacz/pobierz (...) nakazania pozwanej usunięcia skutków naruszenia praw autorskich powoda poprzez zamieszczenie w (...) i ogólnopolskim wydaniu Gazety (...), a także na stronie pozwanej (...) i portalu F. pozwanej (...) w ciągu 14 dni od daty wykonalności wyroku oświadczenia w formie ogłoszenia o wielkości nie mniejszej niż 9x12 cm następującej treści: „A. P. (...)przeprasza Pana J. J., prowadzącego działalność gospodarczą pod firmą: (...)za naruszenie jego praw autorskich do projektu architektonicznego zjazdu do garażu podziemnego wraz z przebudową układu drogowego z drogi publicznej – ul. (...) w G. do inwestycji A. P. na działkach nr (...), koncepcji zabudowy oraz projektu aranżacji dwunastu wnętrz lokali Inwestycji pn. (...)w G. przy ul. (...)”, zobowiązania pozwanej do zwrotu do rąk powoda koncepcji, koncepcji zabudowy, kart opisowych i graficznych projektów aranżacji wnętrz, projektu architektonicznego zjazdu do garażu podziemnego wraz z przebudową układu drogowego z drogi publicznej – ul. (...) w G. do Inwestycji, decyzji o lokalizacji zjazdu na czas nieokreślony i decyzji o warunkach zabudowy, zasądzenia od pozwanej na swoją rzecz kosztów procesu, w tym kosztów zastępstwa procesowego według norm przepisanych oraz opłaty skarbowej od pełnomocnictwa, w przypadku oddalenia wniosku powoda o zwolnienie od kosztów sądowych w całości.

Uzasadniając swe roszczenia powód wskazał, że w ramach prowadzonej pod firmą (...) działalności gospodarczej zawarł z pozwaną A. P. prowadzącą działalność gospodarczą pod firmą (...) w R. umowę (...) oraz ugodę z 8 grudnia 2015 r., przedmiotem których było m.in. wykonanie przez powoda dokumentacji projektowej Inwestycji. Powód wykonał na rzecz pozwanej m.in. materiały marketingowe, koncepcje zabudowy Inwestycji, projekt zjazdu do garażu podziemnego wraz z przebudową układu drogowego, prace skutkujące wydaniem decyzji o lokalizacji zjazdu i decyzje o warunkach zabudowy, 16 sztuk kart projektowych (opisowych i graficznych), w tym projekty aranżacji wnętrz i fotorealistyczne wizualizacje.

Dalej powód twierdził, że pozwana odstąpiła od umowy w całości pismem z 4 marca 2016 r., w związku z czym wezwała powoda do zwrotu uiszczanego wynagrodzenia w kwocie 92.500 zł. Powód wskazał, że w związku z odstąpieniem od umowy po stronie pozwanej powstał obowiązek zwrotu wszystkich wykonanych w jej ramach utworów, z którego pozwania nie wywiązała się, mimo skierowania przez powoda wezwania z 1 października 2020 r.

Powód wskazał, że do 22 września 2020 r., kiedy dokonał oględzin Inwestycji, pozostawał w przekonaniu, że przy jej realizacji pozwania nie wykorzystuje efektów jego pracy udostępnionych na podstawie umowy.

Według powoda, między stronami nie została zawarta umowa licencyjna ani nie zostały przeniesione majątkowe prawa autorskie do wskazanych utworów. W związku z powyższym, w ocenie powoda, pozwana dopuściła się naruszenia praw autorskich przysługujących mu do utworów będących przedmiotem umowy, tj. prac, materiałów, koncepcji oraz dokumentacji projektowej, z których, mimo odstąpienia od umowy łączącej strony, nie przestała korzystać.

W przedmiocie roszczenia o zapłatę kwoty w wysokości 139.292,00 zł powód twierdził, że stanowi ona roszczenie o naprawienie szkody poprzez zapłatę sumy pieniężnej w wysokości odpowiadającej dwukrotności stosownego wynagrodzenia tytułem udzielenia przez uprawnionego zgody na korzystanie z utworu. Wykazując wysokość roszczenia powód przedstawił wyliczenie oparte o zasady wyceny prac projektowych Stowarzyszenia (...) (k. 11), zgodnie z którym jednostkowe wynagrodzenie opiewało na kwotę 69.646,00 zł. Podniósł również, iż dochodzenie przedmiotowego roszczenia nie jest zależne od wykazania winy oraz wysokości szkody, a odpowiedzialność wynika jedynie z bezprawności działań podmiotu korzystającego z utworu.

Powód wskazał, że pozwana dopuściła się również naruszenia jego osobistych praw autorskich poprzez brak informacji o autorstwie powoda 12 projektów aranżacji wnętrz lokali Inwestycji znajdujących się na stronie internetowej pozwanej wskazanej w punkcie 3. petitum pozwu oraz fotorealistycznych wizualizacji Inwestycji opublikowanych przez pozwaną na portalu F. (...) oraz (...)

Odnosząc się do twierdzeń przedstawionych przez pozwaną w odpowiedzi na pozew powód zaprzeczył, by roszczenia dochodzone pozwem uległy przedawnieniu. Zakwestionował również stanowisko pozwanej, w myśl którego stron nie łączyła umowa o treści wynikającej z załączonego do pozwu dokumentu „Umowa nr (...)”, a jedynie porozumienie. W treści pisma z 14 grudnia 2021 r. powód szczegółowo odniósł się do okoliczności

związanych z odstąpieniem przez pozwaną od łączącej strony umowy, zaprzeczając, by wynikało to z nienależytego wykonania umowy, za które ponosił odpowiedzialność.

Powołując się na załączony do pisma z 14 grudnia 2021 r. protokół stanu faktycznego powód zarzucił, że pozwana nie usunęła ze swojej strony internetowej 12 projektów aranżacji wnętrz lokali w ramach Inwestycji i znajdują się one pod adresem wskazanym we wniosku skierowanym do komornika sądowego.

Pozwana A. P. wniosła o oddalenie powództwa w całości oraz zasądzenie od powoda na swoją rzecz kosztów procesu, w tym kosztów zastępstwa procesowego od każdego z roszczeń objętych pozwem według norm przepisanych z uwzględnieniem dwukrotności stawki minimalnej lub według spisu kosztów wraz z odsetkami w wysokości odsetek ustawowych za opóźnienie w spełnieniu świadczenia pieniężnego, za czas od dnia uprawomocnienia się orzeczenia, którym je zasądzono, do dnia zapłaty oraz opłatą od pełnomocnictwa.

Pozwana przyznała, że w roku 2014 zawarła z powodem umowę, zmienioną porozumieniem z 8 grudnia 2015 r., której celem było uzyskanie przez nią ostatecznej decyzji o pozwoleniu na budowę dwóch budynków wielorodzinnych w ramach Inwestycji, prowadzenie przez powoda nadzoru autorskiego, wykonanie dokumentacji powykonawczej, obsługa w zakresie odbiorów, pozwolenia na użytkowanie i zaświadczeń o samodzielności lokali.

W ocenie pozwanej powództwo nie jest zasadne. Strony nie zawarły umowy w dniu 8 grudnia 2015 r., na odstąpienie od której powołuje się powód, oraz z której wywodzi skutki stanowiące podstawę jego roszczeń przeciwko pozwanej. W tym kontekście wskazała również, że dokument załączony do pozwu nazwany „Umowa nr (...)” opatrzony jest datą 28 września 2015 r. oraz nie został podpisany przez strony.

Pozwana podniosła, że roszczenia powoda objęte pozwem uległy przedawnieniu. Przy zachowaniu należytej staranności powód mógł dowiedzieć się o szkodzie oraz osobie obowiązanej do jej naprawienia z dniem otrzymania pisma zawierającego oświadczenie pozwanej o odstąpieniu od umowy. Do roszczeń powoda, jako związanych z działalnością gospodarczą, zastosowanie miał trzyletni okres przedawnienia. W zakresie weryfikacji bezprawnego wykorzystania dokumentacji projektowej zjazdu do hali garażowej w ramach Inwestycji – przy zachowaniu należytej staranności powód mógł jej dokonać już w grudniu 2016 r., kiedy pozwana złożyła wniosek o pozwolenie na budowę, a informacja o tym fakcie podlegała udostępnieniu na stronie Głównego Urzędu Nadzoru Budowlanego oraz stronie Biuletynu Informacji Publicznej Urzędu Miasta G..

Pozwana wskazała, że jej działania były zgodne z prawem, a przedstawiona przez powoda koncepcja „wstecznej bezprawności”, która miałaby być konsekwencją odstąpienia przez nią od umowy, nie znajduje oparcia w przepisach prawa. Zaprzeczyła również, by powód poniósł szkodę w związku z jej działaniami.

Pozwana oświadczyła, że konieczność odstąpienia od umowy łączącej strony była następstwem nienależytego wykonania zobowiązania przez powoda. W jej ocenie okoliczność ta nie może stanowić podstawy dla roszczenia powoda o zapłatę dwukrotności wynagrodzenia wywiedzionego z art. 79 Ustawy z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (j.t. Dz. U. z 2022 r. poz. 2509; dalej jako „pr. aut.”), ponieważ mimo tego, że oświadczenie o odstąpieniu od umowy wywiera skutek *ex tunc* nie doszło do naruszenia praw autorskich powoda, a jego twierdzenie o wstecznej utracie praw do utworów jest nieuzasadnione. W okresie, którego dotyczy roszczenie powoda pozwana korzystała ze wskazanych w pozwie utworów za jego zezwoleniem, czego nie mogło zmienić ze skutkiem wstecznym odstąpienie od umowy łączącej strony. Przepisy prawa przywołane przez powoda nie znajdują, w ocenie pozwanej, zastosowania do utworów oraz ich wykorzystania.

Pozwana wskazała, że, wbrew twierdzeniom powoda, nie naruszyła majątkowych praw autorskich do koncepcji zabudowy, projektu zjazdu do garażu podziemnego wraz z przebudową układu drogowego, uzgodnienia umowy drogowej, uzyskanej przez powoda działającego jako pełnomocnik pozwanej decyzji o warunkach zabudowy, kart projektów aranżacji wnętrz udostępnionych na jej stronie internetowej oraz wizualizacji zamieszczonych na jej profilu w portalu F. (k. 381 – 389). Zaprzeczyła również, by dopuściła się naruszenia osobistych praw autorskich powoda przez niezamieszczenie informacji o jego autorstwie wskazanych w treści pozwu dwunastu projektów aranżacji oraz dwóch wizualizacji na portalu F..

Odnosząc się do roszczenia o wydanie utworów i dokumentów pozwana podniosła, że nie otrzymała od powoda w ramach realizacji umowy łączącej strony niczego, co podlegałoby zwrotowi. Wskazała, że niemożliwym jest dokonanie zwrotu praw niematerialnych, a żaden przepis nie statuuje obowiązku zwrotu pełnomocnikowi decyzji administracyjnej wydanej w postępowaniu, w którym ustanowiony był pełnomocnikiem.

Anhang 5

Wnioskodawca J. K. jest tłumaczem. Wykonuje on tłumaczenia literatury anglojęzycznej na język polski. Dnia 26 sierpnia 1999 r. powód zawarł z Wydawnictwo (...) spółką z ograniczoną odpowiedzialnością z siedzibą w W. (zwana dalej Wydawnictwo (...)) umowę wydawniczą, na mocy której zobowiązał się do wykonania tłumaczenia na język polski książki autorstwa T. H. pt. (...). W § 3 powyższej umowy strony postanowiły, że tłumacz przenosi na wydawcę wyłączone prawo wydawania utworu bez ograniczeń czasowych, ilościowych i terytorialnych na

wszystkich polach eksploatacji związanych ze zwielokrotnianiem utworu i jego rozpowszechnianiem na każdym znanym w dacie umowy nośniku informacji, jak w szczególności w postaci książkowej, prasowej, dysków magnetycznych optycznych, wprowadzania do pamięci komputera i korzystania zeń w systemie on line (umowa z dnia 26 sierpnia 1999 r. – k. 12-15).

Wnioskodawca J. K. dnia 2 listopada 1999 r. zawarł z Wydawnictwem (...) umowę wydawniczą, na mocy której zobowiązał się on do sporządzenia tłumaczenia na język (...) pt. (...). W § 3 powyższej umowy strony postanowiły, że tłumacz przenosi na wydawcę wyłączne prawo wydawania utworu bez ograniczeń czasowych, ilościowych i terytorialnych na wszystkich polach eksploatacji związanych ze zwielokrotnianiem utworu i jego rozpowszechnianiem na każdym znanym w dacie umowy nośniku informacji, jak w szczególności w postaci książkowej, prasowej, dysków magnetycznych optycznych, wprowadzania do pamięci komputera i korzystania zeń w systemie on line (umowa z dnia 2 listopada 1999 r. – k. 16-19).

Dnia 1 lipca 2016 r. Wydawnictwo (...) udzieliło obowiązanej (...) spółce z ograniczoną odpowiedzialnością z siedzibą w K. licencji wyłącznej na wykorzystanie tłumaczenia książki (...) na polach eksploatacji:

a) prawo do wydania przekładu w formie książkowej, w oprawie broszurowej i twardej, w formie e-book i audiobook;

b) prawo wprowadzenia wydania przekładu do obrotu i wykorzystania handlowego, w tym w sprzedaży wysyłkowej.

Licencji udzielono na okres pięcioletni począwszy od dnia 1 lipca 2016 r. (umowa licencyjna z dnia 1 lipca 2016 r. – k. 20-21).

Dnia 20 grudnia 2018 r. Wydawnictwo (...) udzieliło pozwanej licencji wyłącznej na wykorzystanie tłumaczenia książek (...) i (...) na polach eksploatacji: wydania w formie książkowej, w oprawie broszurowej i twardej, w formie e-book i audiobook oraz wprowadzania wydania przekładu do obrotu i wykorzystanie handlowego, w tym w sprzedaży wysyłkowej (umowa licencyjna z dnia 20 grudnia 2018 r. – k. 22-23).

Pozwana wprowadziła do obrotu i oferowała do sprzedaży tłumaczenia książek (...) i (...) w formie e-booków na portalach: publio, upolujeboka.pl, Księgarnia medyczna (...), O.. (okoliczność, o której informacje są powszechnie dostępne w rozumieniu art. 228 § 2 k.p.c.).

Uprawniony był skłonny zawrzeć z pozwaną umowę licencyjną w zakresie „e-booka” na okres 2-letni za kwotę 700 zł brutto. Do zawarcia takiej umowy nie doszło (korespondencja mailowa pomiędzy stronami – k. 24-26).

Ustalając powyższy stan faktyczny, Sąd oparł się na dokumentach szczegółowo wskazanych przy każdej z okoliczności. Dokumentom tym należało przypisać walor wiarygodności, ponieważ na podstawie wszechstronnego rozważenia zebranego materiału Sąd nie uznał, aby ujawniły się jakiejkolwiek okoliczności mogące z urzędu budzić co do nich wątpliwości.

Ponadto, zauważyć trzeba, że sąd nie jest zobligowany do wskazania w uzasadnieniu orzeczenia przyczyn niewzięcia pod uwagę dowodów, które nie dotyczyły faktów istotnych dla rozstrzygnięcia i z tego względu nie stanowiły i nie mogły stanowić podstawy dla ustalenia stanu faktycznego sprawy (por. postanowienie Sądu Najwyższego z dnia 28 września 2000 r., sygn. akt IV CKN 109/00, LEX nr 532102).

Należało mieć też na uwadze, że Sąd, dokonując oceny zgromadzonego materiału po myśli art. 233 § 1 k.p.c., ocenia wiarygodność i moc dowodów według własnego przekonania na podstawie wszechstronnego rozważenia zebranego materiału, w tym wyjaśnień informacyjnych stron, oświadczeń, zarzutów przez nie zgłaszanych, jak i ich zachowania się w trakcie procesu (tak Sąd Najwyższy w wyroku z dnia 24 marca 1999 r., sygn. akt I PKN 632/98, OSNP z 2000 r., nr 10, poz. 382).

Sąd nie podziela zarówno wyrażonego w doktrynie poglądu, zgodnie z którym wiarygodne wykazanie okoliczności to nic innego jak jego uprawdopodobnienie, a wiarygodność nie oznacza pewności zaistnienia faktu, pewność ta wynika dopiero z udowodnienia faktu, jeżeli więc w art. 479 113 § 1 k.p.c. jest mowa o wykazaniu w wiarygodny sposób okoliczności wskazujących na naruszenie prawa, to znaczy, że ustawa zadowala się uprawdopodobnieniem tychże okoliczności w miejsce ich udowodnienia (por. A. Jakubecki, Wezwanie do udzielenia informacji w postępowaniu odrębnym w sprawach własności intelektualnej, PPC 2020, nr 3, s. 490), jak i poglądu, zgodnie z którym zawarty w art. 479 113 § 1 k.p.c. wymóg wykazania w sposób wiarygodny okoliczności wskazujących na naruszenie prawa własności intelektualnej, wymaga ich udowodnienia, nie zaś ich uprawdopodobnienia (tak m.in. E. Stefańska [w:] Kodeks postępowania cywilnego. Komentarz. Tom II. Art. 478-1217, wyd. IV, red. M. Manowska, Warszawa 2021, art. 479116, LEX).

Prawidłowej interpretacji pojęcia wykazania w sposób wiarygodny okoliczności wskazujących na naruszenie prawa własności intelektualnej należy dokonywać, odwołując się do brzmienia dyrektywy 2004/48/WE Parlamentu Europejskiego i Rady z dnia 29 kwietnia 2004 r. w sprawie egzekwowania praw własności intelektualnej (Dz. U. UE. L. z 2004 r. Nr 157, str. 45, zwanej dalej „dyrektywa”), która w art. 8 ust. 1 stanowi, iż państwa członkowskie zapewniają, że w kontekście postępowania sądowego dotyczącego naruszenia prawa własności intelektualnej oraz w odpowiedzi na uzasadnione żądanie powoda właściwe organy sądowe mogą nakazać przedstawienie informacji o pochodzeniu i sieciach dystrybucji towarów lub usług naruszających prawo własności intelektualnej przez naruszającego i/lub jakąkolwiek inną osobę. Zgodnie z tezą 14 preambuły

dyrektywy prawo do informacji ma być stosowane do naruszeń dokonywanych na skalę handlową, a więc działań podejmowanych w celu czerpania bezpośredniego lub pośredniego zysku gospodarczego.

Wskazać także należy, że ustawodawca w przypadkach, właściwych wyłącznie postępowaniu w sprawach własności intelektualnej, zabezpieczenia środka dowodowego (art. 479 98 § 1 k.p.c.) oraz wyjawienia lub wydania środka dowodowego (art. 479 106 k.p.c.) wymaga uprawdopodobnienia roszczenia i okoliczności leżących u jego podstawy. Przyjmując paradygmat racjonalnego ustawodawcy, nie można dojść do wniosku, że prawodawca zrównuje pojęcie wykazania z uprawdopodobnieniem, gdyby tak bowiem było, nie istniałaby potrzeba sięgania po pojęcie wykazania, na co wskazuje § 10 rozporządzenia Prezesa Rady Ministrów z dnia 20 czerwca 2002 r. w sprawie „Zasad techniki prawodawczej” (t.j. Dz. U. z 2016 r. poz. 283), który stwierdza, że do oznaczenia jednakowych pojęć używa się jednakowych określeń, a różnych pojęć nie oznacza się tymi samymi określeniami. Pojęcie wykazania pojawia się w k.p.c. m.in. na gruncie art. 126 § 1 pkt 5 k.p.c., zgodnie z którym każde pismo procesowe powinno zawierać w przypadku gdy jest to konieczne do rozstrzygnięcia co do wniosku lub oświadczenia – wskazanie faktów, na których strona opiera swój wniosek lub oświadczenie, oraz wskazanie dowodu na wykazanie każdego z tych faktów, art. 235 1 k.p.c. stanowiącego, że we wniosku o przeprowadzenie dowodu strona jest obowiązana oznaczyć dowód w sposób umożliwiający przeprowadzenie go oraz wyszczególnić fakty, które mają zostać wykazane tym dowodem, art. 235 2 § 1 pkt 2 i 3 k.p.c., według którego sąd może w szczególności pominąć dowód mający wykazać fakt bezsporny, nieistotny dla rozstrzygnięcia sprawy lub udowodniony zgodnie z twierdzeniem wnioskodawcy lub też nieprzydatny do wykazania danego faktu, art. 236 k.p.c. stwierdzającego, iż w postanowieniu o dopuszczeniu dowodu sąd oznaczy środek dowodowy i fakty, które mają nim zostać wykazane, a w miarę potrzeby i możliwości – także termin i miejsce przeprowadzenia dowodu, na gruncie art. 458 9 § 7 zd. 1 k.p.c., według którego fakty, które miałyby zostać wykazane dowodami wyłączonymi przez umowę dowodową, sąd może ustalić na podstawie twierdzeń stron, biorąc pod uwagę całokształt okoliczności sprawy, czy też art. 479 116 pkt 4 k.p.c., zgodnie z którym w przypadku wniosku o wezwanie do udzielenia informacji konieczne jest wykazanie, że żądane informacje są konieczne do określenia źródła lub zakresu naruszenia prawa.

Anhang 6

Powód dochodził w niniejszej sprawie zapłaty odszkodowania w wysokości 200 000 zł na podstawie art. 79 ust. 1 pkt 3 lit. b Pr. aut., wskazując na naruszenie przez pozwanych autorskich praw majątkowych do utworu w postaci projektu koncepcyjnego kładki mostowej dla pieszych otwieralnej nad kanałem w porcie w U.. Wierzytelność tę powód nabył od spółki (...) na podstawie umowy przelewu z dnia 2 marca 2016 r.

Prawo autorskie służy ochronie dobra niematerialnego w postaci utworu, którym - zgodnie z definicją zawartą w art. 1 ust. 1 Pr. aut. - jest każdy przejaw działalności twórczej o indywidualnym charakterze, ustalony w jakiegokolwiek postaci, niezależnie od wartości, przeznaczenia i sposobu wyrażenia. Powyższa definicja nakazuje oddzielenie utworu jako dobra o charakterze niematerialnym od przedmiotu materialnego, w którym utwór został utrwalony. Utworem w rozumieniu art. 1 ust. 1 Pr. aut. może być również dzieło techniczne, np. projekt techniczny przedstawiony w dokumentacji technicznej. O ile zatem w orzecznictwie wskazuje się, że przedmiotem ochrony prawnoautorskiej jest dokumentacja techniczna, to nie należy utożsamiać jej z konkretnym przedmiotem materialnym, ewentualnie odczytywać jako pewien skrót myślowy, bowiem utworem i przedmiotem ochrony jest utwór techniczny (projekt techniczny) opisany w tej dokumentacji. Odwołanie się do dokumentacji projektowej jest zaś o tyle usprawiedliwione, że zgodnie z treścią art. 1 ust. 3 Pr. aut., utwór jest przedmiotem prawa autorskiego - a tym samym i przedmiotem ochrony udzielanej przez ustawę - od chwili ustalenia, chociażby miał postać nieukończoną. Chwilą stworzenia utworu jest zaś moment, w którym akt twórczości przejawia się na zewnątrz, poza sferę przeżyć intelektualnych twórcy. Może to przybrać właśnie - choć oczywiście niewyłącznie - postać dokumentacji projektowej lub szerzej - technicznej.

Kwalifikacja dzieł stanowiących rezultat działalności technicznej jako utworów w rozumieniu ustawy o prawie autorskim i prawach pokrewnych rodzi szereg problemów, na co zwraca się uwagę w orzecznictwie i piśmiennictwie.

Po pierwsze, przejawem działalności twórczej jest będący wynikiem samodzielnego wysiłku autora rezultat różniący się od innych rezultatów, a indywidualny charakter dzieła wyraża się w tym, że odbiega ono od innych przejawów podobnego działania w sposób świadczący o jego swoistości i niepowtarzalności. Cechą procesu twórczego jest to, aby był on subiektywnie nowy dla samego twórcy, a rezultat działalności twórczej autora powinien być dla niego poprzednio nieznan. Nadto, przedmiotem ochrony prawa autorskiego jest tylko taki rezultat, który odróżnia się, choćby w minimalnym zakresie, od innych rezultatów takiego samego działania. Oznacza to, że warunkiem uznania określonego wytworu pracy umysłowej człowieka za utwór jest także to, aby posiadał on cechę indywidualności (por. wyroki Sądu Najwyższego z dnia 30 czerwca 2005 r., IV CK 763/04,

OSNC 2006/5/92 i z dnia 19 lutego 2014 r., V CSK 180/13 LEX nr 1455199 oraz wyrok Sądu Apelacyjnego w Warszawie z dnia 7 października 2015 r., VI ACa 1200/14, LEX nr 1936794).

Po drugie, wymaga podkreślenia, że na gruncie prawa autorskiego ochronie nie podlega techniczne rozwiązanie zastosowane w projekcie i przedstawione w dokumentacji technicznej, gdyż z reguły stoi temu na przeszkodzie brak cechy indywidualności, ale może ona objąć „dzieła techniczne”, jeżeli konkretna dokumentacja ma cechy wyróżniające i zawiera - uzyskane przy wykorzystaniu swobody twórczej - swoiste cechy osiągnięte dzięki indywidualnemu wkładowi pracy twórczej autora (por. wyroki Sądu Najwyższego z dnia 30 czerwca 2005 r., IV CK 763/04 i z dnia 19 lutego 2014 r., V CSK 180/13).

Po trzecie, pamiętać należy, że ochroną prawnoautorską objęty może być wyłącznie sposób wyrażenia; nie są natomiast objęte ochroną odkrycia, idee, procedury, metody i zasady działania oraz koncepcje matematyczne (art. 1 ust. 2 1 Pr. aut.).

Sąd Okręgowy, opierając się przede wszystkim na opiniach biegłych przeprowadzonych w toku postępowania pierwszoinstancyjnego, wskazał, że szkic koncepcyjny powoda nie zawierał elementów, które pozwalałyby uznać go za utwór podlegający ochronie na gruncie prawa autorskiego. Przede wszystkim koncepcja ta zawierała z. B. yt mało szczegółów, aby ją ocenić pod względem technicznym, architektonicznym „jako przejaw niepowtarzalności, nowości twórczej. Szkice i koncepcje podają idee i propozycje rozwiązań w sposób bardzo ogólny.” Sąd I instancji wskazał również, że kładka w U. jest dziełem technicznym, w którym zastosowano różnorodne elementy pochodzące z konstrukcji kładek i mostów już z. B. udowanych i modyfikowanych. Przyjęte rozwiązania, mimo że są rezultatem działalności kreatywnej powoda, to jednak nie powstał z tego utwór, który może być objęty ochroną prawem autorskim. Koncepcja powoda nie wykazuje dostatecznie doniosłych różnic w porównaniu z uprzednio wytworzonymi produktami tego typu. Nie powstała na skutek autorskiej twórczości powoda, ale wyraża techniczne uwarunkowania jej powstania.

Apelujący zarzucił Sądowi Okręgowemu wadliwą ocenę dowodów - przede wszystkim opinii sporządzonej przez biegłego P. W. (...). Wskazał, że właściwa ocena opinii biegłych - przy uwzględnieniu pozostałego zgromadzonego w sprawie materiału dowodowego - powinna prowadzić do zupełnie odmiennych wniosków niż wysnute przez Sąd I instancji, a mianowicie, że projekt koncepcyjny kładki mostowej dla pieszych w porcie w U. sporządzony przez spółkę (...) posiadał cechy wzoru przemysłowego lub patentu, a zatem powinien też podlegać ochronie prawnoautorskiej jako utwór (dzieło techniczne).

Z zarzutami strony powodowej można się było zgodzić tylko o tyle, że rzeczywiście nie wszystkie twierdzenia i wnioski zawarte w opiniach biegłych J. H. i P. W. (...) były uzasadnione i powinny być w całości przyjęte przez Sąd. Niemniej, w swoim podstawowym zakresie, w jakim odnosiły się do problemów wskazanych w postanowieniu dowodowym z dnia 1 września 2017 r., opinie mogły być wykorzystane przez Sąd do oceny tych kwestii, w przypadku których konieczne było odwołanie się do wiadomości specjalnych.

Argumentacja powoda przedstawiona w apelacji na uzasadnienie zarzutów procesowych zmierzających do podważenia oceny dowodów dokonanej przez Sąd Okręgowy oparta została na założeniach, które nie znajdowały potwierdzenia w zgromadzonym materiale dowodowym, nadto na nieuzasadnionych uproszczeniach.

W pierwszym rzędzie wskazać należało, że pozwani nigdy w toku sprawy nie przyznali, aby projekt kładki spółki (...) stanowił wzór przemysłowy. Przede wszystkim przyznaniu podlegają fakty, co wynika wprost z art. 229 k.p.c., który stanowi, że nie wymagają dowodu fakty przyznane w toku postępowania przez stronę przeciwną, jeżeli przyznanie nie budzi wątpliwości. Powód nie sformułował w niniejszej sprawie twierdzenia faktycznego, aby ww. projekt kładki stanowił wzór przemysłowy, nie mogło zatem ze strony pozwanych nastąpić przyznanie tej okoliczności. Nadto, to nie wola stron decyduje, czy dany wytwór stanowi wzór przemysłowy, ale zależy to od spełnienia określonych przesłanek wskazanych w przepisach prawa materialnego. Podlega to zatem ocenie sądu orzekającego w danej sprawie, a nie uznaniu stron. Definicję wzoru przemysłowego zawiera art. 102 ust. 1 ustawy z dnia 30 czerwca 2000 r. Prawo własności przemysłowej (t. j. Dz. U. z 2020 r., poz. 286 ze zm.; dalej jako p.w.p.), zgodnie z treścią którego, wzorem przemysłowym jest nowa i posiadająca indywidualny charakter postać wytworu lub jego części, nadana mu w szczególności przez cechy linii, konturów, kształtów, kolorystykę, fakturę lub materiał wytworu oraz przez jego ornamentację. Zgodnie z art. 1 ust. 2 pkt 5 Pr. aut., przedmiotem prawa autorskiego mogą być także utwory z zakresu wzornictwa przemysłowego, jednakże wówczas, gdy spełniają ogólne przesłanki utworu w wskazane w art. 1 ust. 1 Pr. aut. Nie ma tu zatem żadnego automatyzmu, zatem nawet spełnienie kryteriów rejestracji wzoru nie przekłada się na równoczesne objęcie ochroną prawnoautorską (por. wyrok Sądu Najwyższego z dnia 20 kwietnia 1984 r., IV CR 139/84, OSNC 1984/10/188).

W świetle powyższego stwierdzić należało, że stanowisko powoda, iż z faktu złożenia przez pozwanych do akt sprawy prywatnej opinii rzeczownika patentowego D. N., w której wskazano, że projekt kładki posiada walory wzoru przemysłowego, należało wyprowadzić wniosek, iż projekt kładki spółki (...) stanowił wzór przemysłowy, nie znajdowało oparcia ani w przepisach procesowych, ani w przepisach prawa materialnego.

Podobnie zakwestionować należało twierdzenie apelującego, iż projekt spornej kładki uznać należało za wynalazek, o czym zdaniem strony powodowej przesądzała opinia biegłego P. W. (...). Po pierwsze, wynalazek jest innym dobrem niematerialnym niż utwór. Jak już była o tym mowa powyżej, przedmiotem ochrony na gruncie prawa autorskiego objęty jest wyłącznie sposób wyrażenia, nie obejmuje ono natomiast ochroną odkryć, idei, czy

też metod i zasad działania. Zgodnie z treścią art. 24 p.w.p., patenty są udzielane - bez względu na dziedzinę techniki - na wynalazki, które są nowe, posiadają poziom wynalazczy i nadają się do przemysłowego stosowania. Jeżeli zatem kwestia ta podlegałaby rozważeniu, to wynalazkiem mogłoby być rozwiązanie techniczne przedstawione w danym projekcie technicznym, o ile spełniałoby przesłanki z art. 24 p.w.p. Po drugie, nie stanowiło przedmiotu opinii biegłego P. W. (...) ustalenie, czy projekt koncepcyjny kładki mostowej dla pieszych w porcie w U. spełnia przesłanki wynalazku w rozumieniu przepisów prawa własności przemysłowej. Na podstawie tej opinii nie można było zatem ustalić przesłanek nowości, poziomu wynalazczego i przemysłowego zastosowania ewentualnego wynalazku - w rozumieniu rozwiązania technicznego ujętego w projekcie koncepcyjnym spółki (...). Opinia biegłego P. W. (...) nie mogła tym samym stanowić dowodu pozwalającego na uznanie spornego projektu za wynalazek, bowiem stawiając taką tezę biegły ten nie tylko wyszedł poza przedmiot zleconej mu opinii, ale w istocie przedstawił wyłącznie swoje własne zdanie. Oczywiście, opinia ta nie mogła także stanowić dowodu na to, że ww. projekt stanowił wzór przemysłowy (niezarejestrowany).

Anhang 7

Goldstandard Text 1 (Anh. 1)	Goldstandard Text 2 (Anh. 2)	Goldstandard Text 3 (Anh. 3)
1.Ausgangsverfahren	1.Anspruch	1.Aktivlegitimation
2.Beklagte	2.Anwaltszwang	2.Anspruch
3.Berufungsgericht	3.Auflösungserklärung	3.Aufhebung
4.Bestimmung	4.Auflösungsgrund	4.Auftragszweck
5.Beurteilung	5.Auflösungsmöglichkeit	5.Ausgestaltung
6.BGH	6.Auflösungsrecht	6.Ausspruch
7.Entgelt	7.Auftraggeberin	7.Begründung
8.Entscheidung	8.Auftragswerk	8.Beklagte
9.Erlaubnis	9.Ausnahmetatbestand	9.Berechtigung
10.EuGH	10.Begehren	10.Berufungsgericht
11.Fall	11.Beklagte	11.Beseitigung
12.Gesetzgeber	12.Berufungsgericht	12.Beseitigungsanspruch
13.Gesetzgebung	13.Bestimmung	13.Beseitigungsbegehren
14.Harmonisierung	14.Bestreitung	14.Beurteilung
15.InfoRL	15.Bestreitungsfrist	15.Dritte
16.Informationsgesellschaft	16.Bestreitungsrecht	16.Eigenart
17.Informationszweck	17.Dritte	17.Einforderung
18.iSd	18.EB	18.Einräumung
19.Judikatur	19.Eingriff	19.Einwendung
20.Klage	20.Einwand	20.Einwilligung
21.Kläger	21.Entgelt	21.Entgelt
22.Klagsabweisung	22.Entscheidung	22.Entscheidungsgegenstand
23.Lichtbild	23.Erklärung	23.Ermangelung
24.Norm	24.Erklärungsempfänger	24.Erstgericht
25.Nutzung	25.Erstattung	25.Gegenleistung
26.Rechtfertigung	26.Erstbeklagte	26.Geltendmachung
27.Rechtsfrage	27.Erstgericht	27.Genehmigung
28.Rechtsprechung	28.Ersturteil	28.Geschäft
29.Revision	29.Eventualbegehren	29.Geschäftsführer
30.Revisionswerber	30.Fallfrist	30.Gesellschaft
31.Richtlinie	31.Fallkonstellation	31.Gesellschafter
32.RS0002495	32.Feststellung	32.Gesellschafter-Geschäftsführer
33.RS0108466	33.Feststellungsanspruch	33.Gesellschaftsanteil
34.Rundfunk	34.Filmproduktionsvertrag	34.GmbH
35.Sachverhalt	35.Filmwerk	35.GmbH-Gesellschafter
36.Sachverhaltskonstellation	36.Frist	36.Identitytät
37.Schutzgegenstand	37.Fristende	37.Interessenabwägung
38.Schutzrecht	38.Gegenstand	
39.Senat	39.Gericht	
40.Umfang	40.Gesetz	
41.Unionsrecht	41.Gesetzgeber	
42.Unterlassung	42.Herausgabebegehren	

43.Urheber	43.Interesse	38.iSd
44.Urheberrecht	44.iSd	39.Klage
45.Urteilsveröffentlichung	45.Judikatur	40.Klägerin
46.Verwertung	46.Klagebeantwortung	41.Klagevertreter
47.Vollharmonisierung	47.Klagebegehren	42.Kunstgewerbe
48.Vorabentscheidungsersuchen	48.Kläger	43.Lichtbild
49.Vorinstanz	49.Klagsabweisung	44.Mahnschreiben
50.Werk	50.Klarstellung	45.Nutzungsentgelt
51.Werknutzung	51.Klärung	46.Nutzungsrecht
52.Werknutzungsrecht	52.MR	47.Privatperson
53.Wortlaut	53.Nachfrist	48.Rechnungslegung
54.argumentieren	54.Nachfristsetzung	49.Rechnungslegungsanspruch
55.begehren	55.Nichtausübung	50.Rechnungslegungsbegehren
56.berufen	56.Nichtgebrauch	51.Recht
57.bestimmen	57.Norm	52.Rechteeinräumung
58.betonen	58.Nutzungsentgelt	53.Rechteerwerb
59.dahinstehen	59.Nutzungsrecht	54.Rechtsprechung
60.erfolgen	60.Passivlegitimation	55.Rechtssinn
61.ersuchen	61.Prozesserklärung	56.Revision
62.erteilen	62.Prozesshandlung	57.RS0076187
63.gestatten	63.Prozessschriftsatz	58.RS0076389
64.rechtfertigen	64.Rechterückruf	59.RS0076406
65.stattgeben	65.Rechtsanwalt	60.RS0076413
66.überholen	66.Rechtsslage	61.RS0076423
67.vervielfältigen	67.Rechtsprechung	62.RS0076443
68.vorliegen	68.Rechtssache	63.RS0076695
69.beklagt	69.Revision	64.RS0076734
70.beurteilend	70.Revisionsbeantwortung	65.RS0115332
71.gemäß	71.Revisionsverfahren	66.RS0125380
72.gerechtfertigt	72.RS0077758	67.Sachverhalten
73.klagsstattgebend	73.Rückruf	68.Schaffung
74.rechtlich	74.Rückrufrecht	69.Senat
75.unentgeltlich	75.RV	70.Spruch
76.urheberrechtlich	76.Schadenersatz	71.Tatsachensubstrat
77.verfassungsunmittelbar	77.Schriftsatz	72.Tatsachenvorbringen
78.§ 42c UrhG	78.Schrifttum	73.Teilurteil
79.§ 50 dUrhG	79.Schwebezustand	74.Umfang
80.§ 502 ZPO	80.Senat	75.Unterlassung
81.§ 86 UrhG	81.Spannungsverhältnis	76.Unterlassungsbegehren
82.4 Ob 53/19a [Pkt 1.1] mwN	82.Spruchpunkt	77.Unterlassungserklärung
83.4 Ob 58/23t Rz 5	83.Teilurteil	78.Unterzeichnung
84.abstrakte Rechtsfrage	84.Unterdrückung	79.Urheber
85.Art 10bis Abs 2 der Berner Übereinkunft	85.Unterlassungsanspruch	80.Urheberrechtsanspruch
86.Art 5 Abs 2 und 3 InfoRL	86.Unterlassungsbegehren	81.UrhG
87.Art 5 Abs 3 lit c zweiter Fall der InfoRL	87.Urheber	82.Urteilsveröffentlichung
88.außerordentliche Revision	88.Urheberrecht	83.Urteilsveröffentlichungsbegehren
89.BGH I ZR 139/15, „Afghanistan Papiere“	89.Urheberrechtsgesetz	84.Verfolgung
90.BGH I ZR 139/15, „Reformistischer Aufbruch“	90.UrhG	85.Veröffentlichungsanspruch
91.C 469/17 Funke Mediengruppe	91.UWG	86.Veröffentlichungsbegehren
92.C 469/17 Rn 64	92.Verfahren	87.Veröffentlichungsermächtigung
93.C 516/17 Rn 27, 39	93.Verfahrensergänzung	88.Verpflichtung
94.C 516/17 Rn 49	94.Verfahrenspartei	89.Verpflichtungserklärung
95.C 516/17 Spiegel Online	95.Verhandlung	90.Vervielfältigung
96.erhebliche Rechtsfrage	96.Versäumung	
97.EuGH C 469/17 Rn 54	97.Verschulden	
98.gefestigte Rechtsprechung	98.Verschweigung	
	99.Vertrag	
	100.Vertragsauflösung	
	101.Vertragsverhältnis	
	102.Vertreter	

99.in Einklang 100.Oberster Gerichtshof 101.Pariser Fassung 102."Richtlinie 2001/29/EG des Europäischen Parlaments und des Rates vom 22. Mai 2001 zur Harmonisierung bestimmter Aspekte des Urheberrechts und der verwandten Schutzrechte in der Informationsgesellschaft" 103.RS0111271 [T2]	103.Verwerter 104.Verwertung 105.Verwertungspflicht 106.Verwertungsrecht 107.Vorbringen 108.Vorinstanz 109.Werk 110.Werkart 111.Werknutzungsberechtigte 112.Werknutzungsbewilligung 113.Werknutzungsrecht 114.Widerruf 115.Widerspruch 116.Wiederherstellung 117.Wirksamkeit 118.Wirksamkeitsvoraussetzung 119.Zahlungsbegehren 120.Zivilverfahren 121.Zurückweisungsfrist 122.Zweck 123.Zweitbeklagte 124.abweisen 125.auffordern 126.auflösen 127.aufrollen 128.ausbedingen 129.aussprechen 130.beantragen 131.beeinträchtigen 132.begehren 133.berechtigen 134.bestreiten 135.betrauen 136.einräumen 137.eintreten 138.einwenden 139.entscheiden 140.erfolgen 141.erheben 142.erklären 143.erteilen 144.feststellen 145.freistellen 146.gewähren 147.herrschen 148.lösen 149.rechtsunkundigen 150.referieren 151.stattdgeben 152.unterbleiben 153.unterlassen 154.untersagen 155.vertreten 156.verweisen 157.verwerten 158.vorlegen 159.vorliegen 160.widerrufen 161.zukommen 162.zulassen	91.Verwendungsanspruch 92.Vorbringen 93.Werk 94.Werknutzungsberechtigter 95.Werknutzungsentgelt 96.Werknutzungsrecht 97.Werkqualität 98.Wertung 99.Wesen 100.Zustimmung 101.abtreten 102.abweisen 103.aufweisen 104.aussprechen 105.beauftragen 106.bejahen 107.bekämpfen 108.einräumen 109.einwenden 110.erachten 111.ergeben sich 112.ergehen 113.erheben 114.erklären 115.erlassen 116.fordern 117.monieren 118.stattdgeben 119.verkennen 120.versagen 121.verstoßen 122.verwerten 123.vorbringen 124.vorliegen 125.werten 126.zukommen 127.zulassen 128.zurückverweisen 129.zustehen 130.aktivlegitimiert 131.angemessen 132.anwaltlich 133.ausschließlich 134.dienstlich 135.eigentümlich 136.gegenständlich 137.geltend 138.geschäftsführend 139.gesondert 140.konkludent 141.rechtlich 142.schöpferisch 143.schutzzfähig 144.schutzunfähig 145.ständig
--	---	--

163.zurückverweisen 164.zurückweisen 165.zustehen 166.anwaltlich 167.anwendbar 168.ausschließlich 169.erstinstanzlich 170.gemäß 171.gesetzlich 172.gewerblich 173.gewerbsmäßig 174.hilfsweise 175.kommerziell 176.kraft 177.maßgeblich 178.materiell 179.originär 180.rechtlich 181.rechtskräftig 182.sachgerecht 183.schutzwürdig 184.ständig 185.strittig 186.umfassend 187.unwirksam 188.unzulässig 189.vertraglich 190.zulässig 191.§ 28 Abs 2 UrhG 192.§ 29 Abs 2 UrhG 193.§ 29 Abs 4 UrhG 194.§ 30 Abs 1 UrhG 195.§ 38 Abs 1 UrhG 196.§ 40 Abs 3 UrhG 197.4 Ob 113/09k 198.4 Ob 158/09b 199.Art 22 UrheberrechtsRL 2019/790 200.außerordentliche Revision 201.Co Verlagsverhältnis 202.Folge geben 203.kreative Freiheit 204.künstlerische Freiheit 205.Rz 1819	146.substanziert 147.übersteigend 148.übertragbar 149.unberechtigt 150.unerheblich 151.unrichtig 152.urheberrechtlich 153.verstoßend 154.vertretbar 155.zulässig 156.§ 1 Abs 1 UrhG 157.§ 1041 ABGB 158.§ 5 Abs 1 UrhG 159.§ 5 UrhG 160.§ 5 UrhG 161.§ 81 UrhG 162.§ 82 f UrhG 163.§ 85 UrhG 164.§ 87a UrhG 165.4 Ob 100/20i 166.4 Ob 159/99g 167.4 Ob 23/15h 168.4 Ob 53/92 169.außerordentliche Revision 170.entgegen halten 171.geistige Schöpfung 172.in eventu 173.Oberster Gerichtshof 174.ordentliche Revision 175.erhebliche Rechtsfrage 176.gänzliche Klageabweisung
--	---

Anhang 8

Goldstandard Text 4 (Anh. 4)	Goldstandard Text 5 (Anh. 5)	Goldstandard Text 6 (Anh. 6)
1.autorstwo 2.bezprawność 3.decyzja 4.dochodzenie 5.inwestycja 6.nakazanie 7.naruszenie 8.odpowiedzialność 9.odsetki 10.okoliczność	1.brutto 2.doktryna 3.dokument 4.dowód 5.dyrektywa 6.fakt 7.interpretacja 8.k.p.c 9.LEX 10.licencja	1.apelacja 2.apelujący 3.argumentacja 4.autor 5.biegły 6.dobro 7.dowód 8.dzieło 9.fakt 10.naruszenie

11.orzeczenie	11.materiał	11.niepowtarzalność
12.oświadczenie	12.naruszenie	12.odszkodowanie
13.pełnomocnik	13.okoliczność	13.odwołanie się
14.petitum	14.orzeczenie	14.okoliczność
15.pismo	15.oświadczenie	15.oparcie
16.podmiot	16.paradygmat	16.opinia
17.porozumienie	17.postanowienie	17.orzecznictwo
18.postępowanie	18.postępowanie	18.p.w.p.
19.powód	19.powód	19.patent
20.powództwo	20.pozwana	20.piśmiennictwo
21.pozew	21.prawo	21.podważenie
22.pozwana	22.prawodawca	22.postępowanie
23.przedmiot	23.preambuła	23.potwierdzenie
24.przepis	24.proces	24.powód
25.roszczenie	25.roszczenie	25.pozwany
26.samodzielność	26.rozpowszechnianie	26.procedura
27.stanowisko	27.rozstrzygnięcie	27.przedmiot
28.staranność	28.rozważenie	28.przepis
29.stowarzyszenie	29.sąd	29.przesłanka
30.strona	30.sprawa	30.przyznanie
31.szkoła	31.strona	31.sąd
32.tytułem	32.sygn.	32.spółka
33.udostępnienie	33.udowodnienie	33.sprawa
34.ugoda	34.umowa	34.strona
35.umowa	35.uprawdopodobnienie	35.swoistość
36.uprawomocnienie	36.ustalenie	36.twierdzenie
37.utwór	37.ustawa	37.twórca
38.weryfikacja	38.ustawodawca	38.ustalenie
39.wezwanie	39.utwór	39.ustawa
40.wniosek	40.uzasadnienie	40.utwór
41.wykonanie	41.wezwanie	41.uwzględnienie
42.wynagrodzenie	42.wiarygodność	42.uzasadnienie
43.zaniechanie	43.wniosek	43.uznanie
44.zasądzenie	44.wnioskodawca	44.wierzytelność
45.zastosowanie	45.wskazanie	45.wniosek
46.zaświadczenie	46.wydanie	46.wola
47.zezwolenie	47.wyjawienie	47.ww.
48.zgoda	48.wykazanie	48.wyrok
49.zobowiązanie	49.wymóg	49.wytwór
50.dochodzić	50.wyrok	50.założenie
51.domagać się	51.żądanie	51.zarzut
52.dopuszczać się	52.zaistnienie	52.dochodzić
53.oświadczyć	53.zarzut	53.nabyć
54.podlegać	54.zwielokrotnianie	54.nakazać
55.podnosić	55.dokonać	55.opierać się
56.powołać się	56.nakazać	56.podlegać
57.przywołać	57.odwołać się	57.przesądzić
58.przyznać	58.postanowić	58.przyznać
59.stanowić	59.stanowić	59.stanowić
60.statuować	60.stosować	60.stwierdzić
61.wezwać	61.ustalić	61.ustalić
62.wnieść	62.uznać	62.wskazać
63.wskazać	63.wskazać	63.wykazać
64.wykazać	64.wykazać	64.wymagać
65.wywodzić	65.wymagać	65.wyprowadzać
66.zakwestionować	66.zadowalić się	66.zakwestionować
67.zaprzeczyć	67.zawrzeć	67.zarzucać
68.zarzucić	68.zobowiązać się	68.nieuzasadniony
69.zasądzić	69.dokonywany	69.orzekający
70.zawrzeć	70.naruszający	70.pierwszoinstancyjny

71.administracyjny	71.obowiązany	71.prawnoautorski
72.bezprawny	72.stanowiący	72.procesowy
73.dochodzony	73.stwierdzający	73.przyznany
74.jednostkowy	74.udowodniony	74.subiektywny
75.należyty	75.uprawniony	75.świadczący
76.nienależyty	76.uzasadniony	76.swoisty
77.nieuzasadniony	77.wiarygodny	77.usprawiedliwiony
78.obowiązany	78.wskazany	78.uzasadniony
79.przedmiotowy	79.wskazujący	79.wadliwy
80.przysługujący	80.wszechstronny	80.wskazany
81.stosowny	81.wyłączny	81.wysnuty
82.uiszczony	82.żądany	82.akt sprawy
83.uprawniony	83.zobligowany	83.akt twórczości
84.ustanowiony	84.akt I PKN 632/98	84.art. 1 ust. 1 Pr. aut.
85.wydany	85."§ 10 rozporządzenia Prezesa Rady Ministrów z dnia 20 czerwca 2002 r. w sprawie „Zasad techniki prawodawczej”"	85.art. 1 ust. 2 1 Pr. aut.
86.wywieziony	86.§ 3	86.art. 1 ust. 2 pkt 5 Pr. aut.
87.zasadny	87.akt IV CKN 109/00	87.art. 1 ust. 3 Pr. aut.
88.zawarty	88.art. 126 § 1 pkt 5 k.p.c.	88.art. 102 ust. 1 ustawy z dnia 30 czerwca 2000 r.
89."art. 79 Ustawy z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych"	89.art. 228 § 2 k.p.c.	89.art. 229 k.p.c.
90.Biuletyn Informacji Publicznej Urzędu Miasta G	90.art. 233 § 1 k.p.c.	90.art. 24 p.w.p.
91.do rąk	91.art. 235 1 k.p.c.	91.art. 79 ust. 1 pkt 3 lit. b Pr. aut.
92.dokumentacja powykonawcza	92.art. 235 2 § 1 pkt 2 i 3 k.p.c.	92.autorska twórczość
93.działalność gospodarcza	93.art. 236 k.p.c.	93.autorskie prawo majątkowe
94.ex tunc	94.art. 458 9 § 7 zd. 1 k.p.c.	94.cecha indywidualności
95."Główny Urząd Nadzoru Budowlanego"	95.art. 479 106 k.p.c.	95.charakter niematerialny
96.instytucja pożytku publicznego	96.art. 479 113 § 1 k.p.c.	96.dobro niematerialne
97.j.t. Dz. U. z 2022 r. poz. 2509	97.art. 479 116 pkt 4 k.p.c.	97.Dz. U. z 2020 r., poz. 286 ze zm.
98.k. 381 – 389	98.art. 479 98 § 1 k.p.c.	98.działalność kreacyjna
99.k. 561 – 562	99.art. 479 116	99.działalność techniczna
100.komornik sądowy	100.art. 8 ust. 1	100.działalność twórcza
101.koszty procesu	101.dojść do wniosku	101.dzieło techniczne
102.koszty sądowe	102.dopuszczenie dowodu	102.IV CK 763/04
103.koszty zastępstwa procesowego	103."dyrektywa 2004/48/WE Parlamentu Europejskiego i Rady z dnia 29 kwietnia 2004 r. w sprawie egzekwowania praw własności intelektualnej"	103.IV CR 139/84
104.majątkowe prawa autorskie	104.Dz. U. UE. L. z 2004 r. Nr 157, str. 45	104.LEX nr 1936794
105.na rzecz	105.Dz. U. z 2016 r. poz. 283	105.materiał dowodowy
106.nadzór autorski	106.fakt bezsporny	106.na gruncie
107.normy przepisane	107.k. 12-15	107.na podstawie
108.obowiązek zwrotu	108.k. 16-19	108.ocena dowodów
109.odalenie wniosku	109.k. 20-21	109.ochrona dobra niematerialnego
110.odsetki ustawowe	110.k. 22-23	110.ochrona prawa autorskiego
111.odstąpić od umowy	111.k. 24-26	111.ochrona prawnoautorska
112.odstąpienie od umowy	112.Kodeks postępowania cywilnego	112.opinia biegłego
113.okres przedawnienia	113.LEX nr 532102	113.OSNC 1984/10/188
114.opatrzonej datą	114.moc dowodu	114.OSNC 2006/5/92
115.opiewać na kwotę	115.na gruncie	115.podlegający ochronie
116.opłata od pełnomocnictwa	116.na mocy	116.postanowienie dowodowe
117."opłata skarbową od pełnomocnictwa"	117.oprzec wniosek	117.praca twórcza
118.osobiste prawa autorskie	118.organ sądowy	118.prawo autorskie
119.po stronie	119.OSNP z 2000 r., nr 10, poz. 382	119.prawo materialne
120.podstawa dla roszczenia	120.państwo członkowskie	120.prawo własności przemysłowej
121.ponosić odpowiedzialność		121.proces twórczy
122.powstanie obowiązku		122.przedmiot materialny
123.pozwolenie na budowę		123.przedmiot ochrony
124.pozwolenie na użytkowanie		124.rzecznik patentowy
		125.Sąd Apelacyjny
		126.Sąd I instancji

125.prawa autorskie 126.prawa niematerialne 127.protokół stanu faktycznego 128.przedmiot roszczenia 129.przedmiot umowy 130.przenieść prawo 131.przepis prawa 132.roszczenie o zapłatę 133.skierowanie wniosku 134.skutek wsteczny 135.spis kosztów 136.stawka minimalna 137 świadczenie pieniężne 138.udzielenie zgody 139.ulec przedawnieniu 140.umowa drogowa 141.umowa licencyjna 142.w całości 143.w postaci 144.wniosek powoda 145.wsteczna utrata praw 146.wydanie decyzji 147.wykazanie winy 148.wykonalność wyroku 149.wysokość szkody 150.wywrzeć skutek 151.zgodne z prawem 152.znaleźć oparcie 153.zwolnienie od kosztów sądowych	121.pismo procesowe 122.po myśli 123.podzielać pogląd 124.pole eksploatacji 125.postępowanie odrębne 126.postępowanie sądowe 127."postępowanie w sprawach własności intelektualnej" 128.PPC 2020, nr 3, s. 490 129.prawo do informacji 130.prawo własności intelektualnej 131.przenosić prawo 132.przeprowadzenie dowodu 133.racjonalny ustawodawca 134.Sąd Najwyższy 135.spółka z ograniczoną odpowiedzialnością 136.środek dowodowy 137.stan faktyczny 138.teza 14 139.Tom II. Art. 478-1217 140.udzielić licencji 141.umowa dowodowa 142.umowa licencyjna 143.umowa wydawnicza 144.walor wiarygodności 145.własność intelektualna 146."wniosek o wezwanie do udzielenia informacji" 147.wprowadzenie do obrotu 148.wprowadzić do obrotu 149.wyłączne prawo 150.z urzędu 151.zabezpieczenia środka dowodowego 152.zawarcie umowy 153.zysk gospodarczy 154.ujawnić okoliczności	127.Sąd Najwyższy 128.Sąd Okręgowy 129.strona powodowa 130.strona przeciwna 131.swoboda twórcza 132.twierdzenie faktyczne 133.umowa przelewu 134.ustawa o prawie autorskim i prawach pokrewnych 135.V CSK 180/13 136.V CSK 180/13 LEX nr 1455199 137.VI ACa 1200/14 138.w rozumieniu 139.W świetle 140.w toku 141.wyprowadzić wniosek 142.wytwór pracy umysłowej człowieka 143.zarzut procesowy 144.złożyć opinię
---	---	--

Anhang 9

TEXT2TCS Text 1 (Anh. 1) Ursprüngliche Liste	TEXT2TCS Text 2 (Anh. 2) Ursprüngliche Liste	TEXT2TCS Text 3 (Anh. 3) Ursprüngliche Liste
1.§ 50 2.Abruf 3.abstrakter Rechtsfragen 4.Argument 5.Art 6.Aufführungen 7.Ausgangsverfahren 8.außerordentlichen 9.beantworten 10.Beantwortung 11.befreit 12.Beklagte 13.beklagten	1.abgegebenen 2.allseitigen 3.Ansicht 4.Ansprüche 5.anwaltlich vertreten 6.Anwaltszwang 7.Argument 8.Auflösungserklärung 9.Auflösungsgrund 10.Auflösungsmöglichkeit 11.Auftraggeberin 12.Auftragswerk 13.Ausnahmetatbeständen	1.abändern 2.ABGB 3.abgeändert 4.Aktivlegitimation 5.aktivlegitimiert 6.Änderungen 7.Anfertigen 8.Animation 9.animieren 10.Anspruch 11.anwaltlichen Mahnschreiben 12.anwaltlichen Mahnschreibens 13.artigen

14.Bericht	14.Ausnahmetatbestands	14.auffällig
15.Berichterstattung	15.ausschließlichen	15.aufgeschlosseneres
16.Berichterstattung über	16.außerordentliche	16.Aufhebung
17.Berichterstattung über Tageser-	17.Ausstellung	17.Aufteilung
eignisse	18.bar	18.Auftragszwecks
18.Berichterstattung über Tageser-	19.Beklagte	19.Augen
eignisse“	20.Berufungsgericht	20.ausgesprochen
19.berichtet; vervielfältigt; ver-	21.Berufungsgericht	21.Ausgestaltung
breitet	22.Beschränkungen	22.ausschließlicher und übertrag-
20.Berner Übereinkunft	23.Bestellung	barer
21.Berufungsgerichts	24.Bestimmung	23.Aussehen
22.Bestimmung	25.Bestreitung	24.äußere Form
23.beurteilenden	26.Bestreitungsfrist	25.äußeren Form
24.Bezirk	27.Bestreitungsrechts	26.äußerer Merkmale
25.BGH	28.Bildes	27.außerordentlichen
26.BGH I ZR	29.Brauchbarkeit	28.Ausspruch
27.Bilder	30.Co Verlagsverhältnisses	29.Bearbeiters
28.Deutschland	31.Dauer	30.bearbeiteten
29.Draht	32.Drehbuchautor; Regisseur; Fil-	31.bearbeiteten
30.dUrhG	mer; Cutter	32.Bearbeitung
31.Entgelt	33.Drehtagen	33.Bearbeitung
32.Ereignisse	34.Dritte	34.Bearbeitungen
33.Ereignisses	35.ehestens	35.Bearbeitungen
34.Erlaubnis	36.Eingriffen	36.befreit
35.EuGH	37.Einwands	37.Begründung
36.Existenz	38.Empfang	38.Beklagte
37.Fahrgeschäften	39.Entgelt	39.Beklagte; Beklagten
38.Fällen	40.Entscheidung	40.Berechtigung
39.Film	41.Entwürfe	41.Berufungsgericht
40.Formulierung	42.Ergebnis	42.Berufungsgericht
41.Fotos	43.Erhalt	43.Berufungsgerichts
42.freie	44.erklärt	44.beschwingteres
43.freie Werknutzung	45.Erklärung	45.Beseitigung
44.freie Werknutzungsrecht	46.Erklärungsempfänger	46.Beseitigungs-, Rechnungsle-
45.freien Nutzungen	47.Erstattung	gungs- und Veröffentlichungsan-
46.freies Werknutzungsrecht	48.Erstbeklagte	spruchs
47.Funke Mediengruppe	49.Erstbeklagte	47.bildenden Künste; Kunstgewer-
48.gefestigter	50.Erstbeklagten; Beklagten	bes
49.Gesetzgebung	51.Erstgericht	48.comic artiges
50.Handy	52.erstinstanzlichen Verfahrens	49.cooler
51.Harmonisierung	53.Ersturteil	50.dienstlichen Verpflichtung
52.hörbar	54.Eventualbegehren; Begehren	51.Dritte
53.InfoRL	55.faktischer	52.eigentümliche
54.Informationszweck	56.Fallfrist	53.eigentümliche geistige Schöp-
55.Klage	57.Fallkonstellation	fung
56.Kläger	58.Feststellung	54.eigentümlichen geistigen
57.Klägers	59.Feststellungsansprüche	Schöpfung
58.Klagsabweisung	60.Film	55.Einräumung
59.klagsstattgebende Entschei-	61.Filme	56.Einräumung
dung	62.Filmhersteller	57.Einwilligung
60.Kunst	63.Filmherstellerin	58.eleganteren
61.Lichtbilder	64.Filmproduktionsvertrag	59.entfernt
62.Lichtbildern	65.Filmschaffenden	60.Entgelt
63.Linie	66.Filmwerk	61.Entgelt
64.Literatur	67.Filmwerken	62.Entgelts
65.Medien	68.Filmwerks	63.Entscheidungsgegenstands

66.Mitarbeiter	69.Folge	64.Erscheinungsbild
67.Mitgliedstaaten	70.Formulierung	65.Erstellung
68.Nachrichtensendung	71.Fragen	66.Erstgericht
69.Namens	72.Frist	67.Figuren
70.nationale Regelung	73.Fristende Einwände	68.Filmkunst
71.nationalen Normen	74.Gebrauch	69.Finger
72.Norm	75.Gedanke	70.flacher
73.Nutzung	76.gehört	71.freundlicheres
74.Obersten Gerichtshofs	77.Gericht	72.froh
75.öffentlich wahrnehmbar	78.Gesetzes	73.Gebilde
76.öffentliche Wiedergabe	79.gesetzlichen	74.Gebrauchsgrafik
77.öffentlichen Vorträgen	80.gesetzlichen Auflösungsrecht	75.Gebrauchsgrafiken
78.Öffentlichkeit	81.gewerblich	76.Gegenleistung
79.Österreich	82.gewerblich hergestellter	77.gegenständlichen
80.österreichische Judikatur	83.gewerblich hergestelltes	78.gegenständlichen
81.österreichischen Gesetzgeber	84.gewerbsmäßig hergestellten	79.geistige Schöpfungen
82.Photographie	85.gewerbsmäßig hergestellten	80.Geldbetrag
83.Quelle	Filmen	81.Geltendmachung
84.Rechtliche Beurteilung	86.gewerbsmäßig hergestellten	82.Geltendmachung
85.Rechtsfrage	Films	83.Genehmigung
86.Rechtsprechung	87.Herausgabebegehren	84.geschaffen
87.Revision	88.Hersteller	85.Geschäft
88.Revisionswerber	89.Hersteller	86.geschäftsführenden
89.Richtlinie	90.hilfsweise	87.Geschäftsführer
90.Rundfunk gesendet	91.historische Gesetzgeber	88.Geschäftsführer
91.Rundfunks	92.Interessen	89.Geschäftsführers; Urhebers
92.Rundfunksendung	93.Kino	90.Gesellschaft
93.Sachverhalt	94.Kino und	91.Gesellschafter
94.Sachverhalte	95.Klagebeantwortung	92.Gesellschafter-Geschäftsführer
95.Sachverhaltskonstellation	96.Klagebegehren	93.Gesellschafter
96.Schutz von	97.Kläger	94.Gesellschaftsanteile
97.Schutzgegenständen	98.Kläger	95.Gesicht
98.Schutzrechte in der Informati- onsgesellschaft	99.Klagsabweisung	96.Gestaltungselementen
99.Senats	100.Klarstellung	97.Gestaltungskraft
100.sichtbar	101.Klärung	98.gezeichnet
101.Spiegel Online	102.kommerzielle Verwertung	99.gezeichneten
102.Stromausfall	103.kommerziellen Verwertung	100.gezeichneten
103.Tagesereignis	104.kreative Freiheit	101.Gliedmaßen
104.Tagesereignisse	105.Kritik	102.GmbH
105.Text	106.kritische Glosse	103.GmbH-Gesellschafter
106.Textes	107.kritisierte	104.handwerkliche
107.Twitter Account	108.künstlerische Freiheit	105.heraushängenden
108.Übertragung	109.Künstler	106.humorvolle
109.Umfang	110.Mangelhaftigkeit	107.Identity
110.Umstände	111.Materialien	108.Illustration
111.Umständen	112.materiell rechtlicher Fallfrist	109.Illustration der
112.unentgeltlich	113.materielle	110.Illustrationen
113.Unionsrecht	114.Nachfrist	111.Illustrationen
114.Unterlassung	115.Nachfristsetzung	112.individuelle Eigenart
115.urheberrechtlich geschützten Werke	116.Nichtausübung	113.individuellen
116.Urheberrechts	117.Nichtgebrauchs	114.individuellen
117.Urhebers	118.Nichtgebrauchs	115.Individuelles
118.UrhG	119.Normen	116.Inhalt
119.Urteilsveröffentlichung	120.Nutzungsentgelt	117.Interessenabwägung
	121.Nutzungsrechte	118.joviales
	122.originären Verwertungsrechte	119.Kerns

120. Verbandsländer	123. Österreichisches Urheberrecht; UrhG; Österreichisches Urheberrecht I	120. Klage
121. Verbreitung	124. Passivlegitimation	121. Klage
122. verfassungsunmittelbaren	125. proLIBRIS	122. Klageabweisung
123. Vergnügungsparks	126. proLIBRIS	123. Klägerin
124. veröffentlichten	127. Promotion	124. Klägerin
125. Vervielfältigung	128. Prozesserklärung	125. Klagevertreter
126. Verwertung	129. Prozesshandlung	126. komplexen
127. Vollharmonisierung	130. Prozessschriftsatz	127. konkludente
128. Vollständigkeit	131. prüfen	128. konkludenten
129. Vorabentscheidungsersuchen	132. Rechterückruf	129. konkret
130. Vorführungen	133. Rechtlich	130. Körper
131. Vorgängen	134. rechtlicher	131. künstlerische Stil
132. Vorinstanzen	135. Rechtsanwalt	132. lachenden
133. wahrnehmbar	136. Rechtskenntnisse	133. leisten
134. Wahrnehmbarkeit	137. rechtskräftig	134. Lichtbildern
135. Website	138. Rechtslage	135. Literatur
136. Werke	139. Rechtsprechung	136. Mahnschreiben
137. Werken	140. Rechtsprechung	137. Mahnschreibens
138. Werknutzung	141. Rechtssache	138. Mahnschreibens
139. Wortlaut	142. rechtsunkundigen	139. Maskottchen
140. ZPO	143. referiert	140. mechanische
141. zugänglich	144. Regelungen	141. mechanischen Veränderung
142. Zweck	145. Revision	142. Micky Maus
	146. Revisionsbeantwortung	143. Micky Maus-artigen
	147. Revisionsverfahrens	144. Mundes
	148. Rückruf	145. Namen
	149. Rückruf	146. Nase
	150. Rückrufrecht	147. Nutzungsentgelte
	151. Rückrufs	148. Nutzungsrecht
	152. RV	149. Nutzungsrechte
	153. Schadenersatz und/oder	150. Nutzungsrechte; Nutzungsrechten
	154. Schriftsatz	151. Nutzungsrechten
	155. Schriftsatz	152. Obersten Gerichtshofs
	156. Schriftsatzes	153. offenen
	157. Schrifttum	154. ordentliche
	158. schutzwürdiger	155. Originale
	159. Schwebezustände	156. Originals
	160. Senat	157. Originellem
	161. Sicherheit	158. Originelles
	162. Spannungsverhältnis	159. plastisch
	163. Spruchpunkten	160. Privatperson
	164. strittig	161. Produkt
	165. technisch und inhaltlich	162. Punkt
	166. Teilurteil	163. Punkte
	167. Texte Materialien Judikatur	164. Pupillen
	168. Texte Materialien Judikatur	165. Rechnungslegung
	169. TV	166. Rechnungslegungsbegehren
	170. Überlegung	167. Rechte
	171. Unterdrückung	168. Rechteeinräumung
	172. Unterlassungsansprüche	169. Rechteeinräumungen
	173. Unterlassungsbegehren	170. Rechteerwerb
	174. untersagen	171. rechtliche Beurteilung
	175. unwirksam	172. Rechtsfragen
	176. Urheber	173. Rechtsprechung
		174. Rechtssinn

177.Urheberrechtsgesetz	175.Reduktion
178.Urheberrechtsgesetz	176.Revision
179.UrheberrechtsRL	177.Revision
180.Urhebers	178.routinemäßige Leistungen
181.UrhG	179.runden
182.UWG	180.Sachverhalten
183.Verfahren	181.Schattierungen
184.Verfahrens	182.schlanken
185.Verfahrensergänzung	183.schlanker
186.Verfahrenspartei	184.Schokoladetafel
187.Verhandlung	185.Schokostücke
188.Verlängerung	186.schöpferische Eigenart
189.Verlust	187.schöpferischer
190.Versäumung	188.schutzfähig
191.Verschulden	189.schutzunfähig
192.Verschweigung	190.Senat
193.Vertrag	191.Spruch
194.vertraglich	192.ständige
195.Vertragsauflösung	193.Striche
196.Vertragsverhältnis	194.Substanzierte Einwendungen
197.Vetreter	195.sympathische
198.Verwerter	196.Tafel
199.Verwertungspflicht	197.Tätigkeit
200.Verwertungsrechte	198.Tatsachenvorbringen
201.Verwertungsrechts	199.Tatsachenvorbringen und -
202.Vollständigkeit	substrat
203.Voraussetzungen	200.technisch bedingt; vereinfachen
204.Vorbringen	201.technisch bedingte
205.Vorinstanzen	202.Teilurteil
206.Walter	203.Teilurteils
207.Werkart	204.Tonkunst
208.Werknutzungsberechtigte	205.tropfenartige
209.Werknutzungsberechtigte	206.Übertragung
210.Werknutzungsberechtigten	207.Umgestaltung; Änderung
211.Werknutzungsbewilligung	208.Umstände
212.Werknutzungsrecht	209.unbeschwerteres
213.Werknutzungsrechte	210.unrichtig
214.Werks	211.Unterlassung
215.Widerruf	212.Unterlassungs- und Verpflichtungserklärung
216.widerrufen	213.Unterlassungs-, Beseitigungs- und Rechnungslegungsbegehren
217.Widerspruch	214.Unterlassungs-, Beseitigungs- und Veröffentlichungsbegehren
218.Wiederherstellung	215.Unterzeichnung
219.Wirksamkeit	216.Urheber
220.Wirksamkeit	217.Urheber
221.Wirksamkeitsvoraussetzungen	218.urheberrechtlich geschützte
222.Zahlung	219.urheberrechtlich geschützte Werke
223.Zahlungsbegehren	220.urheberrechtliche Schutz
224.zitierte	221.urheberrechtliche Werkqualität
225.Zivilverfahrens	222.urheberrechtlichen Ansprüche
226.zurückweist	223.urheberrechtlicher Ansprüche
227.Zurückweisungsfrist	
228.zurückzuweisen	
229.zurückzuweisen	
230.Zweitbeklagte	

		224.urheberrechtlicher Schutz 225.Urheberrechtsansprüche 226.Urhebers und Geschäftsführers 227.UrhG 228.UrhG); UrhG 229.Urteilsveröffentlichung 230.Urteilsveröffentlichungsbegehren 231.Urteilsveröffentlichungsbegehrens 232.Veränderungen 233.Vereinfachungen 234.Verfolgung 235.Vermögen 236.veröffentlichen 237.Veröffentlichungsermächtigung 238.verrechnete 239.vertretbar 240.Vervielfältigung 241.Verwendungsanspruch 242.verwertet 243.Vorbringens 244.vorgebracht 245.Werk 246.Werke 247.Werke 248.Werken; Werks 249.Werknutzungsberechtigten 250.Werknutzungsentgelt 251.Werknutzungsrecht 252.Werknutzungsrechte 253.Werknutzungsrechts 254.Wert 255.Wert 256.Wertung 257.Wesen 258.Wiedergabe 259.Zeitgeist 260.zeitgemäße 261.zerlegt 262.Zimmermann; Lindwurm; FITNESS) 263.Zufriedenheit 264.Zunge 265.Zungen 266.Zustimmung 267.Zustimmung 268.Zweck
--	--	--

Anhang 10

TEXT2TCS Text 4 (Anh. 4) Ursprüngliche Liste	TEXT2TCS Text 5 (Anh. 5) Ursprüngliche Liste	TEXT2TCS Text 6 (Anh. 6) Ursprüngliche Liste
1.adresem 2.aranżacji 3.aranżacji wnętrz 4.architektonicznego zjazdu 5.autorstwie 6.autorstwie 7.bezprawnego 8.bezprawności 9.Biuletynu Informacji Publicznej 10.budowę 11.budynków wielorodzinnych 12.decyzje o 13.decyzji 14.decyzji 15.decyzji administracyjnej 16.decyzji o 17.dnia zapłaty 18.dokument 19.dokumentacji 20.dokumentacji 21.dokumentacji projektowej 22.dokumentacji projektowej 23.dokumentów 24.dokumentu 25.dowiedzieć 26.drogi publicznej 27.działalności gospodarczej; działalność gospodarczą 28.działalnością gospodarczą 29.działania 30.działaniami 31.działkach 32.efektów 33.fakcie 34.firmą 35.fotorealistyczne wizualizacje 36.fotorealistycznych wizualizacji 37.Fundacji 38.garażu podziemnego 39.garażu podziemnego 40.Gazety 41.graficznych) 42.hali garażowe 43.informacja 44.informacji 45.informacji 46.Inwestycji 47.Inwestycji 48.inwestycji; Inwestycja 49.jednostkowe wynagrodzenie	1.akt 2.anglojęzycznej 3.art 4.audiobook 5.bezsporny 6.broszurowej 7.brzmienia 8.doktrynie 9.dokumentach 10.Dokumentom 11.dopuszczeniu 12.dowód 13.dowodami 14.dowodem 15.dowodów 16.dowodu 17.dyrektywy 18.dysków magnetycznych optycznych 19.dystrybucji 20.e-book 21.e-booków 22.egzekwowania praw własności intelektualnej 23.eksploatacji 24.fakt 25.faktów 26.faktów 27.faktu 28.faktu 29.fakty 30.formie 31.gruncie art 32.handlową 33.informacje 34.informacji 35.informacji 36.informacyjnych stron 37.język 38.język polski 39.Kodeks postępowania cywilnego 40.komputera 41.korzystania 42.książek 43.książki 44.książkowej 45.książkowej, prasowej, 46.Księgarnia medyczna 47.LEX)	1.akt sprawy 2.akt twórczości 3.apelacji 4.apelującego 5.Apelujący 6.architektonicznym 7.Argumentacja 8.art 9.automatyzmu 10.autora 11.autorskich praw majątkowych 12.autorskiej twórczości 13.biegłego 14.biegłego; biegły 15.biegłych 16.cechę 17.cechy 18.cechy 19.części 20.definicja 21.Definicję 22.dobra 23.dobra niematerialnego 24.dobrem niematerialnym 25.dokumentacja 26.dokumentacja techniczna 27.dokumentacji 28.dokumentacji projektowej 29.dokumentacji technicznej 30.dowodów 31.dowodów 32.dowodu 33.działalności twórczej 34.działalności twórczej autora 35.działania 36.dziedzinę 37.dzieł 38.dzieła 39.dziełem 40.dzieło techniczne 41.elementy 42.fakturę 43.fakty 44.idee 45.idei 46.indywidualności 47.indywidualny charakter 48.indywidualny charakter 49.indywidualnym 50.intelektualnych twórcy

50.kart opisowych i graficznych projektów	48.Licencji	51.kanałem
51.kart projektów aranżacji wnętrz	49.licencji wyłącznej	52.kładek
52.kart projektowych	50.literatury	53.kładka
53.komornika sądowego	51.materiału	54.kładki
54.koncepcja	52.miejsce	55.kładki mostowej
55.koncepcje zabudowy	53.moc	56.kładki mostowej
56.koncepcji	54.naruszającego	57.kładki mostowej dla pieszych
57.koncepcji zabudowy	55.naruszeń	58.kładki spółki
58.koncepcji zabudowy	56.naruszenia prawa	59.kolorystykę
59.korzystanie	57.naruszenia prawa własności	60.koncepcja
60.kosztów	intelektualnej	61.Koncepcja
61.kosztów	58.naruszenie	62.koncepcje
62.kosztów sądowych	59.naruszenie	63.koncepcje matematyczne
63.kwoty	60.naruszenie prawa	64.konstrukcji
64.łączącej	61.nośniku	65.konturów
65.łączącej strony	62.O	66.kreacyjnej powoda
66.lokali	63.obrotu	67.kryteriów
67.lokali	64.oceny	68.kształtów
68.lokalizacji	65.okoliczności	69.Kwalifikacja
69.majątkowe prawa autorskie	66.okoliczności	70.kwestii
70.majątkowych praw autorskich	67.okoliczności sprawy	71.linii
71.materiałów	68.określeń	72.materiał
72.materiały marketingowe	69.określeniami	73.materiale dowodowym
73.modyfikacji	70.oprawie	74.materiału dowodowego
74.nadzoru autorskiego	71.organy sądowe	75.metod
75.nakazania	72.orzeczenia	76.metody
76.naprawienia	73.osobę	77.modyfikowanych
77.naprawienie	74.oświadczeń	78.moment
78.naruszenia	75.oświadczenia	79.mostów
79.naruszenia	76.oświadczenie	80.niematerialnym
80.naruszenia praw autorskich	77.oznaczenia	81.niepowtarzalności
81.naruszenie	78.oznaczyć	82.nieuzasadnionych
82.nienależytego	79.pamięci	83.nowości
83.norm przepisanych	80.państwa członkowskie	84.nowości twórczej
84.norm przepisanych	81.paradygmat	85.ocena
85.obowiązek	82.pewność	86.ocenę
86.obsługa	83.pewności	87.ocenie
87.odbiorów	84.pismo procesowe	88.oceny
88.oddalenia	85.pochodzeniu	89.oceny
89.oddalenie	86.podstawy	90.ochroną odkryć
90.odpowiedzialność	87.poglądu	91.ochroną odkrycia
91.odsetkami; odsetek usta-	88.pojęć	92.ochroną prawem autorskim
wowych	89.polach	93.ochroną prawnoautorską
92.odstąpienia	90.portalach	94.ochroną prawnoautorską
93.odstąpienia	91.postanowienie	95.ochronie
94.odstąpienie	92.postanowieniu	96.ochronie prawnoautorskiej
95.odstąpieniem	93.postępowania sądowego	97.ochrony
96.odstąpieniu od umowy	94.postępowaniu odrębnym	98.ochrony
97.odstąpiła	95.postępowaniu w sprawach	99.ochrony prawa autorskiego
98.ogłędzin	własności intelektualnej	100.ochrony prawnoautorskiej
99.ogłoszenia	96.powoda	101.odszkodowania
100.ogólnopolskim wydaniu	97.Pozwana	102.odwołanie
101.okoliczności	98.PPC	103.okoliczności
102.okres	99.prawa własności intelektualnej	104.opinia
103.opłatą	100.prawa własności intelektual-	105.Opinia
	nej	106.opiniach

104.opłaty skarbowej od pełnomocnictwa	101.Prawidłowej interpretacji	107.opinie
105.opóźnienie	102.prawo	108.opinii
106.opóźnienie	103.prawo do informacji	109.opinii
107.orzeczenia	104.prawo własności intelektualnej	110.ornamentację
108.osobie obowiązanej	105.prawo wprowadzenia	111.orzecznictwie
109.osobistych praw autorskich	106.prawodawca	112.OSNC
110.osobistych praw autorskich	107.preambuły	113.patentu
111.ostatecznej	108.Prezesa Rady Ministrów	114.patenty
112.oświadczenia	109.procesu	115.pierwszoinstancyjnego
113.oświadczenie	110.przedstawienie	116.piesznych
114.oświadczenie o	111.przekładu	117.piśmiennictwie
115.pełnomocnictwa	112.przekonania	118.podważenia
116.pełnomocnik	113.przeprowadzenia	119.porcie
117.pełnomocnikiem	114.przeprowadzenie	120.porcie
118.pełnomocników	115.przyczyn	121.postać
119.pełnomocnikowi	116.publio	122.postaci
120.petitum pozwu	117.racjonalnego ustawodawcy	123.postanowieniu dowodowym
121.pisma	118.roszczenia	124.postępowania
122.pisma	119.rozpowszechnianiem	125.postępowania
123.podmiotu	120.rozstrzygnięcia	126.potwierdzenia
124.podpisany	121.rozstrzygnięcia	127.powód
125.porozumienie	122.rozważenia	128.Powód
126.porozumieniem	123.Sąd	129.powoda
127.portalu	124.sąd	130.powoda
128.portalu	125.sąd	131.poziom
129.portalu F	126.Sąd Najwyższy	132.poziomu
130.portalu społecznościowym	127.Sądu Najwyższego	133.pozwani
131.postępowaniu	128.sieciach	134.pozwanych
132.powód	129.spółce z ograniczoną odpowiedzialnością	135.Pr
133.Powód	130.spółką z ograniczoną odpowiedzialnością	136.Pr
134.powód	131.sporządzenia	137.pracy twórczej
135.powoda	132.sprawy	138.prawa autorskiego
136.powoda	133.sprawy	139.prawa autorskiego
137.powodem	134.sprzedaży	140.Prawo autorskie
138.powództwa	135.sprzedaży wysyłkowej	141.Prawo własności przemysłowej; p.w.p.); wzorem przemysłowym
139.powództwo	136.środek dowodowy	142.problemów
140.powykonawczej	137.środka dowodowego	143.procedury
141.pozew	138.stan faktyczny	144.procesu twórczego
142.pozwana	139.stanu faktycznego	145.produktami
143.Pozwana	140.stron	146.projekcie
144.pozwana	141.strona	147.projekcie koncepcyjnym
145.pozwaną	142.stronami	148.projekcie technicznym
146.pozwanej	143.strony	149.projekt
147.pozwanej	144.systemie on line	150.projekt kładki
148.pozwania	145.termin	151.projekt koncepcyjny
149.Pozwem	146.tłumacz	152.projekt koncepcyjny
150.pozwem	147.towarów lub usług	153.projekt techniczny
151.pozwem	148.twardej	154.projektu koncepcyjnego
152.pozwie	149.twierdzeń	155.prywatnej
153.pozwolenia na użytkowanie	150.twierdzeniem	156.przedmiotem
154.pozwoleniu	151.udowodnienia	157.przedmiotem materialnym
155.pozwu	152.udowodniony	158.przedmiotu materialnego
156.pozwu	153.umowę	159.przedstawione
157.prac		160.przejaw działalności twórczej
158.prac projektowych		

159.prace	154.umowę dowodową	161.przemysłowego stosowania
160.pracy	155.umowę licencyjną	162.przemysłowego zastosowania
161.praw autorskich	156.umowy	163.przepisach prawa materialnego
162.praw autorskich; praw	157.upolujebooka.pl	164.przepisach procesowych
163.praw niematerialnych	158.uprawdopodobnienia	165.przepisów prawa własności
164.prawem	159.uprawdopodobnienia	przemysłowej
165.procesu	160.uprawdopodobnienie	166.przesłanek
166.procesu	161.uprawdopodobnieniem	167.przesłanki
167.profilu	162.uprawdopodobnieniem	168.przeznaczenia
168.projekt	163.Uprawniony	169.przeżyć
169.projektów	164.urzędu	170.przyznane
170.projektów	165.ustalenia	171.przyznanie
171.projektowej Inwestycji	166.ustawa	172.przyznaniu
172.projektu	167.ustawodawca	173.rejestracji
173.projektu	168.utworu	174.rezultat
174.projektu aranżacji	169.uzasadnieniu	175.rezultat działalności technicznej
175.projekty aranżacji wewnątrz	170.walor	176.rezultatów
176.protokół stanu faktycznego	171.wątpliwości	177.rozważeniu
177.przebudową	172.Wezwanie do udzielenia informacji	178.rozwiazania
178.przebudową	173.wiarygodne	179.rozwiazania technicznego
179.przebudową układu drogowego	174.wiarygodność	180.rozwiazanie techniczne
180.przedawnienia	175.wiarygodności	181.rzecznika patentowego
181.przedawnieniu	176.wiarygodny	182.Sąd
182.przedawnieniu	177.wiarygodny	183.Sąd I instancji
183.przedmiotem	178.wnioskodawcy	184.Sąd Okręgowy
184.przepis	179.wniosku o przeprowadzenie	185.Sąd Okręgowy
185.przepisach prawa	180.wniosku o wezwanie do udzielenia informacji	186.Sądowi Okręgowemu
186.Przepisy prawa	181.wniosku; wniosek lub	187.Sądu Apelacyjnego w Warszawie
187.roszczeń	182.wprowadzania	188.Sądu Najwyższego
188.roszczeń	183.wszechstronnego	189.Sądu Najwyższego
189.roszczenia	184.wydania	190.sądu orzekającego
190.roszczenia	185.wydania	191.samodzielnego
191.roszczenia o	186.wydawania	192.sferę
192.roszczenia o zapłatę	187.wydawcę	193.skrót myślowy
193.roszczenie	188.Wydawnictwem	194.spółkę
194.roszczenie o	189.Wydawnictwo	195.spółki
195.rozpowszechnianiu	190.wydawniczą	196.spółki
196.samodzielnosci	191.wyjaśnień	197.spornego projektu
197.sieci	192.wyjawienia lub	198.spornej kładki
198.skierowania	193.wykazane	199.sporządzonej
199.skutek	194.wykazania	200.sporządzony
200.skutki	195.wykazania	201.sposób wyrażenia
201.skutków	196.wykazanie	202.sposób wyrażenia
202.spelnieniu	197.wykazaniu	203.sposobu
203.spisu	198.wykonania	204.sprawie
204.stanowisko	199.wykorzystania handlowego	205.stanowisko
205.staranności	200.wykorzystanie; wykorzystanie handlowego	206.stronę przeciwną
206.stawki minimalnej	201.wymóg	207.strony powodowej
207.Stowarzyszenia	202.wyroku	208.strony powodowej
208.stronami	203.zabezpieczenia	209.stworzenia
209.stronie	204.zachowania	210.swobody twórczej
210.stronie	205.zarzutów	211.swoistości
211.stronie Głównego Urzędu Nadzoru Budowlanego	206.zawarcia	212.szkie koncepcyjny
212.stronie internetowej		

213.stronie internetowej	207.zgłaszanych	213.Szkice
214.Strony	208.źródła	214.techniczne rozwiązanie
215.strony	209.zwielokrotnianiem	215.techniczne uwarunkowania
216.strony internetowej	210.zysku gospodarczego	216.technicznej
217.sumy pieniężnej		217.technicznym
218.świadczenia pieniężnego		218.techniki
219.szkołę		219.tezę
220.szkoły		220.twierdzenia
221.szkoźcie		221.twierdzenia faktycznego
222.tręci		222.twierdzenie
223.twierdzeń		223.twórcy
224.twierdzeniom		224.umowy przelewu
225.udostępnieniu		225.ustalenia
226.udostępnionych		226.ustalenie
227.ugodę		227.ustawę
228.uiszczzonego		228.ustawy
229.układu drogowego		229.ustawy o prawie autorskim
230.układu drogowego		230.utrwalony
231.umowa licencyjna		231.utwór
232.umowa o treści		232.utwór techniczny
233.umowę		233.Utworem
234.umowę		234.utworem
235.umowy		235.utworów
236.umowy		236.utworu
237.uprawnionego		237.utworu; utwór
238.uprawomocnienia		238.utwory
239.ustawowymi odsetkami		239.uzasadnienie
240.Ustawy		240.uzasadnione
241.usunęła		241.uznaniu stron
242.usunięcia		242.walory
243.utworów		243.wartości
244.utworów		244.wątpliwości
245.utworu		245.Wierzytelność
246.uzgodnienia umowy drogowej		246.wkładowi
247.warunkach zabudowy		247.wnioski
248.warunkach zabudowy		248.wniosków
249.weryfikacji		249.wola stron
250.wezwania		250.wynalazczego
251.winy		251.wynalazczy
252.wizualizacji		252.wynalazek
253.wizualizacji fotorealisty-		253.wynalazki
cznych		254.wynalazku
254.wnętrz		255.wyrażenia
255.wniosek o pozwolenia na bu-		256.wyrok
dowę		257.wyrok
256.wniosku		258.wyroki
257.wydanej		259.wyróżniające
258.wydanie		260.wysnute
259.wydaniem decyzji		261.wytwór
260.wykazania		262.wytworu
261.wykonalności		263.wytworu pracy umysłowej
262.wykonania		człowieka
263.wykonania		264.wytworzonymi
264.wykonanie		265.wzornictwa przemysłowego
265.wykonanie		266.wzoru
266.wykonanych		267.wzoru przemysłowego

267.wykorzystania 268.wyliczenie 269.wynagrodzenia 270.wynagrodzenia 271.wyroku 272.wysokości 273.wywiedzionego 274.zabudowy 275.załączonego 276.zamieszczenie 277.zapłatę 278.zapłatę 279.zapłaty 280.zasady wyceny 281.zasądzenia 282.zasądzenie; zasądzono 283.zastępstwa procesowego 284.zastępstwa procesowego 285.zaświadczeń 286.zezwoleniem 287.zgody 288.zjazdu 289.zjazdu 290.zobowiązania 291.zobowiązania 292.zwolnienie 293.zwrotowi 294.zwrotu 295.zwrotu		268.wzoru przemysłowego; wzór przemysłowy 269.założeniach 270.zapłaty 271.zarzutami 272.zarzutów procesowych 273.zasad działania 274.zasady działania 275.z. B.udowanych 276.zdanie
---	--	---

Anhang 11

ChatGPT Text 1 (Anh. 1) Prompt 1	ChatGPT Text 1 (Anh. 1) Prompt 2	ChatGPT Text 1 (Anh. 1) Prompt 3	ChatGPT Text 1 (Anh. 1) Prompt 4	ChatGPT Text 1 (Anh. 1) Prompt 5
1.§ 42c UrhG 2.§ 502 ZPO 3.§ 86 UrhG 4.abstrakte Rechtsfragen 5.Afghanistan Papiere 6.Art 5 Abs 3 lit c InfoRL 7.Aufführungen 8.außerordentliche Revision 9.Berichterstattung 10.Berner Übereinkunft 11.Berufungsgericht 12.BGH 13.dUrhG 14.Entgelt 15.erhebliche Bedeutung 16.EuGH 17.Fahrgeschäfte 18.freie Werknutzung 19.freies Werknutzungsrecht 20.Funke Mediengruppe 21.gefestigte Rechtsprechung 22.Informationsgesellschaft 23.Informationszweck 24.Kläger 25.Klagsabweisung 26.Lichtbilder 27.Mitgliedstaaten 28.Nachrichtensendung 29.Name des Urhebers 30.nationale Regelung	1.§ 42c UrhG 2.§ 86 UrhG 3.angemessenes Entgelt 4.Art 5 Abs 3 lit c InfoRL 5.Aufführungen 6.außerordentliche Revision 7.Berichterstattung 8.Berner Übereinkunft 9.Berufungsgericht 10.BGH 11.BGH I ZR 139/15 12.BGH I ZR 139/15 „Reformistischer Aufbruch“ 13.C 469/17 Funke Medien- gruppe 14.C 516/17 Spiegel Online 15.EuGH 16.Europäisches Parlament 17.Fahrgeschäfte 18.freies Werknutzungsrecht 19.Informationsgesellschaft 20.Informationszweck 21.iSd § 502 ZPO 22.Kläger 23.Klagsabweisung 24.klagsstattgebende Entschei- dung 25.Lichtbilder 26.Nachrichtensendung	1.§ 42c UrhG 2.§ 502 ZPO 3.§ 86 UrhG 4.angemessenes Entgelt 5.Art 5 Abs 3 lit c zweiter Fall 6.außerordentliche Revision 7.Berichterstattung 8.Berner Übereinkunft 9.Berufungsgericht 10.BGH 11.Deutschland 12.EuGH 13.EuGH C 469/17 14.EuGH C 516/17 15.Fahrgeschäfte 16.freies Werknutzungsrecht 17.Funke Mediengruppe 18.Informationsgesellschaft 19.Judikatur 20.Kläger 21.Klagsabweisung 22.Lichtbilder 23.Nachrichtensendung 24.nationale Regelung 25.Oberster Gerichtshof (OGH) 26.Österreich 27.Pariser Fassung 28.Rechtsfrage 29.Rechtsprechung 30.Revisionswerber	1.abstrakte Rechtsfragen 2.Afghanistan Papiere 3.Art 5 Abs 3 lit c 4.außerordentliche Revision 5.Berichterstattung 6.Berner Übereinkunft 7.Berufungsgericht 8.Bezirk 9.BGH 10.Entgelt 11.EuGH 12.Fahrgeschäfte 13.Funke Mediengruppe 14.InfoRL 15.Informationszweck 16.Kläger 17.Kunst 18.Lichtbilder 19.Lichtbilder 20.Literatur 21.Medien 22.Mitgliedstaaten 23.Nachrichtensendung 24.Oberster Gerichtshof 25.Öffentlichkeit 26.Öffentlichkeitsrecht 27.Pariser Fassung 28.Quelle 29.Rechtliche Beurteilung 30.Rechtsfrage	1.§ 42c UrhG 2.§ 50 dUrhG 3.§ 502 ZPO 4.§ 86 UrhG 5.Afghanistan Papiere 6.Art 5 Abs 3 lit c InfoRL 7.außerordentliche Revision 8.Beklagter 9.Berichterstattung über ein Ta- gesereignis 10.Berner Übereinkunft 11.BGH 12.BGH I ZR 139/15 13.Entgelt 14.erhebliche Rechtsfrage 15.EuGH C 469/17 16.EuGH C 516/17 17.Europäisches Parlament 18.Existenz und Verbreitung 19.Formulierung eines Textes 20.freies Werknutzungsrecht 21.Funke Mediengruppe 22.gefestigte Rechtsprechung 23.InfoRL 24.Informationsgesellschaft 25.Informationszweck 26.Klage 27.Kläger 28.Klagsabweisung 29.Lichtbilder

31.Oberster Gerichtshof 32.Öffentlichkeit 33.österreichische Judikatur 34.Quelle 35.Rechtsfrage 36.Revisionswerber 37.Richtlinie 2001/29/EG 38.Rundfunk 39.Sachverhaltskonstellation 40.Senat 41.Spiegel Online 42.Stromausfall 43.Tagesereignis 44.Twitter Account 45.Umstände des Einzelfalls 46.Unionsrecht 47.Unterlassung 48.Urheberrecht 49.Urteilsveröffentlichung 50.Verbreitung 51.Vergnügungspark 52.Vervielfältigung 53.verwandte Schutzrechte 54.Vollharmonisierung 55.Vorfürhungen 56.Vorinstanzen 57.Vorträge 58.Werke der Literatur und Kunst	27.Name des Urhebers 28.nationale Regelung 29.Oberster Gerichtshof 30.öffentliche Vorträge 31.Öffentlichkeit zur Verfügung gestellt 32.österreichische Judikatur 33.Pariser Fassung 34.Photographie 35.Rat der Europäischen Union 36.Rechtsfrage 37.Richtlinie 2001/29/EG 38.Rundfunk 39.Rundfunksendung 40.Sachverhaltskonstellation 41.Stromausfall 42.Tagesereignis 43.Übertragung mittels Draht 44.Unionsrecht 45.Unterlassung 46.Urheberrecht 47.Urteilsveröffentlichung 48.Verbreitung 49.Vergnügungspark 50.Vervielfältigung 51.verwandte Schutzrechte 52.Vorabentscheidungsersuchen 53.Vorfürhungen 54.Vorinstanzen 55.Werken der Literatur und Kunst	31.Richtlinie 2001/29/EG 32.Rundfunk 33.Rundfunk 34.Spiegel Online 35.Stromausfall 36.Tagesereignis 37.Twitter Account 38.Unionsrecht 39.Unterlassung 40.Urheber 41.Urheberrecht 42.Urteilsveröffentlichung 43.verfassungsunmittelbaren freien Nutzungen 44.Vergnügungspark 45.verwandte Schutzrechte 46.Vorinstanzen 47.Werke der Literatur und Kunst	31.Rechtsfrage 32.Rundfunk 33.Rundfunk gesendet 34.Sachverhaltskonstellation 35.Spiegel Online 36.Stromausfall 37.Twitter Account 38.Unionsrecht 39.Unterlassung 40.Urheber 41.UrhG (§ 86, § 42c) 42.Urteilsveröffentlichung 43.Verbreitet 44.Vergnügungspark 45.Vervielfältigt 46.Vollharmonisierung 47.Vorabentscheidungsersuchen 48.Vorinstanzen 49.Wahrnehmbarkeit 50.Website 51.ZPO (§ 502)	30.Nachrichtensendung 31.nationalen Normen 32.Oberster Gerichtshof 33.Öffentlichkeit 34.österreichische Judikatur 35.Pariser Fassung 36.Photographie 37.Quelle 38.Rat der Europäischen Union 39.Rechtsprechung des EuGH 40.Reformistischer Aufbruch 41.Richtlinie 2001/29/EG 42.Rundfunk 43.Rundfunksendung 44.Stromausfall 45.Tagesereignisse 46.Übertragung mittels Draht 47.Unionsrecht 48.Unterlassung 49.Urheber 50.Urheberrecht 51.urheberrechtlich geschützte Werke 52.Urteilsveröffentlichung 53.verfassungsunmittelbaren freien Nutzungen 54.Verwertung 55.Vollharmonisierung 56.Vorinstanzen 57.Werke der Literatur und Kunst
---	---	--	--	--

Anhang 12

ChatGPT Text 2 (Anh. 2) Prompt 1	ChatGPT Text 2 (Anh. 2) Prompt 2	ChatGPT Text 2 (Anh. 2) Prompt 3	ChatGPT Text 2 (Anh. 2) Prompt 4	ChatGPT Text 2 (Anh. 2) Prompt 5
1.§ 29 UrhG 2.§ 40 Abs 3 UrhG 3.14-Tages-Frist 4.anwaltlich vertreten 5.Auftraggeberin 6.ausschließliche Nutzungsrechte 7.außerordentliche Revision 8.Ausstellung 9.Berufungsgericht 10.Bestreitungsrecht 11.Bestreitungsrecht kraft Gesetzes 12.Ciresa 13.Cutter 14.Drehbuchautor 15.Drehtage 16.Entgelt 17.Entscheidung 4 Ob 113/09k 18.Erklärung 19.Erstbeklagte 20.erstinstanzliches Verfahren 21.Eventualbegehren 22.Feststellungsansprüche 23.Filmer 24.Filmhersteller 25.Filmproduktionsvertrag 26.Filmschaffende 27.gesetzlicher Ausnahmetatbestand	1.§ 29 UrhG 2.§ 38 UrhG 3.Abs (Absatz) 4.Anwaltszwang 5.Auflösungserklärung 6.Auflösungsmöglichkeit 7.Auftragswerk 8.ausschließliche Nutzungsrechte 9.außerordentliche Revision 10.Berufungsgericht 11.Bestreitungsrechts 12.Bestreitungsrechts 13.Co Verlagsverhältnisses 14.Cutter 15.Drehbuchautor 16.Drehtage 17.Entgelt 18.Entgelt 19.Erklärungsempfänger 20.Erstbeklagte 21.Erstgericht 22.Eventualbegehren 23.Fallfrist 24.Feststellungsansprüche 25.Filmer 26.Filmherstellerin 27.Filmproduktionsvertrag 28.gesetzlichen Ausnahmetatbestands	1.Auflösungserklärung 2.Auftraggeberin 3.Auftragswerk 4.Ausschließliche Nutzungsrechte 5.Ausstellung 6.Berufungsgericht 7.Bestreitungsrecht 8.Co Verlagsverhältnis 9.Cutter 10.Drehbuchautor 11.Entgelt 12.Erstbeklagte 13.Erstinstanzliches Verfahren 14.Eventualbegehren 15.Fallfrist 16.Feststellungsansprüche 17.Filmer 18.Filmhersteller 19.Filmherstellerin 20.Filmproduktionsvertrag 21.Gesetzlicher Auflösungsrecht 22.Gesetzlicher Ausnahmetatbestand 23.Gewerbsmäßig hergestellte Filme 24.Glosse 25.Herausgabebegehren 26.Herstellungsverhältnis	1.Anwaltszwang 2.Auflösungserklärung 3.Auflösungsmöglichkeit 4.Auflösungsrecht 5.Auftraggeberin 6.Auftragswerk 7.Ausnahmebestimmungen 8.Berufungsgericht 9.Bestreitungsfrist 10.Bestreitungsrecht 11.Co Verlagsverhältnis 12.Cutter 13.Drehbuchautor 14.Drehtagen 15.Entgelt 16.Erklärungsempfänger 17.Erstbeklagte 18.Erstgericht 19.Eventualbegehren 20.Feststellungsansprüche 21.Film 22.Filmer 23.Filmhersteller 24.Filmherstellerin 25.Filmproduktionsvertrag 26.Filmschaffenden 27.Glosse 28.Herausgabebegehren 29.Klagebeantwortung 30.Klagebegehren	1.Abs (Absatz) 2.Anwaltszwang 3.Auflösungsrecht 4.Auftraggeberin 5.Ausnahmebestimmung 6.Berufungsgericht 7.Bestreitungsrecht 8.Cutter 9.Drehbuchautor 10.Erstinstanzliches Verfahren 11.Eventualbegehren 12.Feststellungsansprüche 13.Filmer 14.Filmhersteller 15.Filmherstellerin 16.Filmproduktionsvertrag 17.Gesetzgeber 18.Gesetzlicher Auflösungsrecht 19.Gewerbliche Nutzung 20.Herausgabebegehren 21.Klagebeantwortung 22.Kreative Freiheit 23.Künstlerische Freiheit 24.Materiell rechtliche Fallfrist 25.Nichtausübung 26.Nichtgebrauch 27.Nutzungsentgelt 28.Nutzungsrechte 29.Prozessklärung

28.gesetzliches Auflösungsrecht 29.gewerblich 30.gewerbsmäßig hergestellte Filme 31.gut organisierte Unternehmen 32.Herausgabebegehren 33.Kino 34.Klagebeantwortung 35.Kläger 36.Klagsabweisung 37.kreative Freiheit 38.künstlerische Freiheit 39.Nutzungsentgelt 40.originäre Verwertungsrechte 41.Prozesserklärung 42.Prozesshandlung 43.Prozessschriftsatz 44.Rechtskenntnisse 45.Rechtslage 46.Rechtsunkundiger Erklärungsempfänger 47.Regisseur 48.Revisionsbeantwortung 49.Revisionsverfahren 50.Rückruf 51.Rückrufrecht 52.Rückrufrecht bei Nichtausübung bzw. wegen Nichtgebrauchs 53.Schadenersatz 54.Schriftsatz 55.Schrifttum 56.Schrifttum 57.Teilurteil	29.gewerbsmäßig hergestellten Films 30.Herausgabebegehren 31.Hersteller 32.Klagebeantwortung 33.kommerzielle Verwertung 34.kreative Freiheit 35.kritische Glosse 36.künstlerische Freiheit 37.materiel rechtlicher Fallfrist 38.materiellen Wirksamkeitsvoraussetzungen 39.originären Verwertungsrechte 40.Passivlegitimation 41.Prozesserklärung 42.Prozesshandlung 43.Prozessschriftsatz 44.Rechtslage 45.Regisseur 46.Revisionsbeantwortung 47.Revisionsverfahren 48.Rückruf 49.Rückrufrecht 50.Rückrufrecht bei Nichtausübung 51.Schriftsatz 52.Schrifttum 53.Teilurteil 54.Unterlassungsansprüche 55.UrhG (Urheberrechtsgesetz) 56.UWG (Gesetz gegen den unlauteren Wettbewerb) 57.Verfahrenspartei 58.Vertragsverhältnis 59.Verwerter	27.Kläger 28.Kommerzielle Verwertung 29.Kreative Freiheit 30.Künstlerische Freiheit 31.Nichtausübung 32.Nichtgebrauch 33.Nutzungsentgelt 34.Nutzungsrechte 35.Passivlegitimation 36.Promotion 37.Prozesserklärung 38.Prozesshandlung 39.Prozessschriftsatz 40.Rechtslage 41.Rechtsunkundiger Erklärungsempfänger 42.Regisseur 43.Revisionsbeantwortung 44.Revisionsverfahren 45.Rückruf 46.Rückrufrecht 47.Schadenersatz 48.Schriftsatz 49.Schrifttum 50.Teilurteil 51.Umfassendes Werknutzungsrecht 52.Unterlassungsansprüche 53.UrhG (Urheberrechtsgesetz) 54.UWG (Gesetz gegen den unlauteren Wettbewerb) 55.Verfahrenspartei 56.Vertrag 57.Vertragsverhältnis 58.Verwertungspflicht 59.Verwertungsrecht	31.Kläger 32.Mangelhaftigkeit 33.Nachfrist 34.Nutzungsentgelt 35.Nutzungsrechte 36.Passivlegitimation 37.Promotion 38.Prozesserklärung 39.Prozesshandlung 40.Rechtskenntnisse 41.Regisseur 42.Revisionsverfahren 43.Rückrufrecht 44.Schriftsatz 45.Schrifttum 46.Teilurteil 47.Unterlassungsansprüche 48.Urheber 49.UrheberrechtsRL (Urheberrechtsrichtlinie) 50.UrhG (Urheberrechtsgesetz) 51.UWG (Gesetz gegen den unlauteren Wettbewerb) 52.Vertrag 53.Vertragsverhältnis 54.Verwerter 55.Verwertungspflicht 56.Verwertungsrechte 57.Werknutzungsbewilligung 58.Werknutzungsrecht 59.Widerrufen 60.Wirksamkeitsvoraussetzungen 61.Zahlungsbegehren 62.Zivilverfahren	30.Prozesshandlung 31.Rechterückruf 32.Regisseur 33.Restliches Entgelt 34.Revisionsverfahren 35.Rückruf 36.Rückrufrecht 37.Rückweisung 38.Schadenersatz 39.Schriftsatz 40.Schrifttum 41.Teilurteil 42.Unterlassungsansprüche 43.Urheber 44.UrheberrechtsRL (Urheberrechtsrichtlinie) 45.UrhG (Urheberrechtsgesetz) 46.UWG (Gesetz gegen den unlauteren Wettbewerb) 47.Verfahrenspartei 48.Verwertungsrechte 49.Werknutzungsbewilligung 50.Werknutzungsrecht 51.Widerspruch 52.Zahlungsbegehren 53.Zivilverfahren
---	---	---	--	---

58.TV 59.Unterlassungsansprüche 60.Urheber 61.UrhG (Urheberrechtsgesetz) 62.UWG (Gesetz gegen den unlauteren Wettbewerb) 63.Verfahrensergänzung 64.Verfahrenspartei 65.Vertrag 66.Vertragsverhältnis 67.Verwerter 68.Verwertungsrechte 69.Verwertungsrechte gemäß § 38 UrhG 70.Walter 71.Werknutzungsbewilligung 72.Werknutzungsrecht 73.Widerruf 74.Zahlungsbegehren 75.zivilrechtliches Verfahren	60.Verwertungspflicht 61.Verwertungsrechte 62.Vorinstanzen 63.Werknutzungsbewilligung 64.Werknutzungsrecht 65.Widerruf 66.Widerspruch 67.Zahlungsbegehren 68.Zivilverfahren	60.Verwertungsrechte 61.Vorzeitige Vertragsauflösung 62.Werknutzungsbewilligung 63.Werknutzungsrecht 64.Wirksamkeit 65.Zahlungsbegehren 66.Zivilverfahren		
--	---	---	--	--

Anhang 13

ChatGPT Text 3 (Anh. 3) Prompt 1	ChatGPT Text 3 (Anh. 3) Prompt 2	ChatGPT Text 3 (Anh. 3) Prompt 3	ChatGPT Text 3 (Anh. 3) Prompt 4	ChatGPT Text 3 (Anh. 3) Prompt 5
1.Aktivlegitimation 2.aktivlegitimiert 3.Auftrag 4.äußere Form 5.außerordentliche Revision 6.Bearbeitung (§ 5 UrhG) 7.Berufungsgericht 8.Beseitigung (§ 82 f UrhG) 9.bildende Künste	1.§ 1041 ABGB 2.§ 5 UrhG 3.§ 81 UrhG 4.§ 82 f UrhG 5.§ 85 UrhG 6.§ 87a UrhG 7.Aktivlegitimation 8.aktivlegitimiert	1.ABGB (Allgemeines Bürgerliches Gesetz. B.uch) 2.Abtretung 3.Aktivlegitimation 4.Auftragszweck 5.außerordentliche Revision 6.Bearbeitung 7.Beklagte 8.Berufungsgericht	1.§ 1 Abs 1 UrhG 2.§ 1041 ABGB 3.§ 5 UrhG 4.§ 81 UrhG 5.§ 82 f UrhG 6.§ 85 UrhG 7.§ 87a UrhG 8.Aktivlegitimation 9.aktivlegitimiert	1.§ 1041 ABGB 2.§ 5 UrhG 3.§ 81 UrhG 4.§ 82 f UrhG 5.§ 85 UrhG 6.§ 87a UrhG 7.Abtretung 8.Aktivlegitimation 9.Aufhebung

10.eigentümliche geistige Schöpfung	9.ausschließliche und übertragbare Nutzungsrechte	9.Beseitigung	10.äußere Merkmale	10.Auftragszweck
11.eigentümliche geistige Schöpfung (§ 1 Abs 1 UrhG)	10.äußere Form	10.eigentümliche geistige Schöpfung	11.außerordentliche Revision	11.äußere Merkmale
12.Einwilligung des Urhebers	11.außerordentliche Revision	11.eigentümliche geistige Schöpfung des Bearbeiters	12.Bearbeitung	12.außerordentliche Revision
13.Entgelt	12.Bearbeitung	12.Einträumung	13.Beklagte	13.Bearbeitung
14.Entscheidung	13.Beklagte	13.Einwilligung des Urhebers	14.Berufungsgericht	14.Berufungsgericht
15.Erstgericht	14.Berufungsgericht	14.Entgelt	15.Beseitigung	15.Beseitigung
16.Filmkunst	15.Beseitigung	15.Entscheidungsgegenstand	16.bildende Künste	16.Beseitigungsbegehren
17.Gebrauchsgrafik	16.eigentümliche geistige Schöpfung	16.Erstgericht	17.dienstlichen Verpflichtung	17.Einwilligung
18.Geschäftsführer	17.Einwilligung des Urhebers	17.Gebrauchsgrafik	18.eigentümliche geistige Schöpfung	18.Entgelt
19.Gesellschaft	18.Entscheidungsgegenstand	18.Gegenleistung	19.Einwilligung	19.erhebliche Rechtsfragen
20.Gesellschaftsanteile	19.erhebliche Rechtsfragen	19.Geschäftsführer	20.Entscheidungsgegenstand	20.Erstgericht
21.Identität des Werks	20.Erstgericht	20.GmbH (Gesellschaft mit beschränkter Haftung)	21.erhebliche Rechtsfragen	21.Gebrauchsgrafik
22.Illustrationen	21.Gebrauchsgrafik	21.handwerkliche, routinemäßige Leistungen	22.Erstgericht	22.Gegenleistung
23.Kunstgewerbe	22.Geschäftsführer	22.Identität des Werks	23.Filmkunst	23.Geltendmachung
24.künstlerischer Stil	23.Identität des Werks	23.Illustration	24.Gebrauchsgrafik	24.Geschäftsführer
25.Literatur	24.Illustrationen	24.Klägerin	25.geistige Schöpfungen	25.Gesellschaft
26.Mahnschreiben	25.Klägerin	25.Kunstgewerbe	26.Geschäftsführer	26.Gesellschaftsanteile
27.mechanische Veränderung	26.Kunstgewerbe	26.künstlerischer Stil	27.Gesellschaftsanteile	27.GmbH-Gesellschafter
28.mechanische Veränderung	27.künstlerischer Stil	27.Mahnschreiben	28.Glubschaugen	28.Identität des Werks
29.Nutzungsrechte	28.Mahnschreiben	28.Mahnschreiben	29.handwerkliche Leistungen	29.Illustrationen
30.ordentliche Revision	29.Nutzungsrechte	29.Maskottchen	30.Identität des Werks	30.Interessenabwägung
31.Rechnungslegung (§ 87a UrhG)	30.Oberster Gerichtshof	30.Nutzungsentgelte	31.Illustrationen	31.Klageabweisung
32.Rechteeinräumung	31.ordentliche Revision	31.Nutzungsrechte	32.individuelle Eigenart	32.konkludente Rechteeinräumung
33.routinemäßige Leistungen	32.Rechnungslegung	32.Nutzungsrechte	33.iSd § 5 UrhG	33.künstlerischer Stil
34.Schokostücke	33.Rechteabtretung	33.Oberster Gerichtshof	34.Klägerin	34.Lichtbilder
35.schöpferische Eigenart	34.Rechteeinräumung	34.ordentliche Revision	35.konkludente Rechteeinräumung	35.Mahnschreiben
36.Tatsachenvorbringen	35.Rechteerwerb	35.Rechnungslegung	36.Kunstgewerbe	36.Maskottchen
37.Teilurteil	36.routinemäßige Leistungen	36.Rechteeinräumung	37.künstlerischer Stil	37.Nutzungsentgelte
38.Tonkunst	37.schöpferische Eigenart	37.Schokoladetafel	38.Literatur	38.Nutzungsrechte
39.Unterlassung (§ 81 UrhG)	38.Teilurteil	38.schöpferische Eigenart	39.Mahnschreiben	39.Originell
40.Urheber	39.Umgestaltung	39.Tatsachenvorbringen	40.Micky Maus-artigen Gliedmaßen	40.Rechnungslegung
41.Urheberrechtlicher Schutz	40.Unterlassung	40.Teilurteil	41.Nutzungsrechte	41.Rechteerwerb
	42.urheberrechtlicher Schutz		42.ordentliche Revision	42.Routineleistungen
				43.Schaffung
				44.schöpferische Eigenart

42. Urteilsveröffentlichung (§ 85 UrhG) 43. Vermögen 44. Veröffentlichung 45. Vervielfältigung 46. Verwendungsanspruch (§ 104 I ABGB) 47. Werke 48. Werknutzungsentgelt 49. Werknutzungsrecht 50. Werknutzungsrechte 51. Werkqualität 52. Zeitgeist	43. Urteilsveröffentlichung 44. Veröffentlichungsermächtigung 45. Verwendungsanspruch 46. Werke 47. Werke der bildenden Künste 48. Werknutzungsberechtigter 49. Werknutzungsentgelt 50. Werknutzungsrecht 51. Werkqualität 52. Zeitgeist	41. Umgestaltung 42. Unterlassung 43. Urheber 44. UrhG (Urheberrechtsgesetz) 45. Urteilsveröffentlichung 46. Urteilsveröffentlichungsbegehren 47. Veröffentlichungsermächtigung 48. Verwendungsanspruch 49. Werke 50. Werke der bildenden Künste 51. Werknutzungsberechtigten 52. Werknutzungsentgelt 53. Werknutzungsrecht 54. Werkqualität 55. Wert der Gesellschaftsanteile 56. Wertung 57. Zeitgeist	43. Privatperson 44. Rechnungslegung 45. Rechteerwerb 46. routinemäßige Leistungen 47. Schokoladetafel 48. schöpferische Eigenart 49. Tatsachenvorbringen 50. Teilurteil 51. Tonkunst 52. Umgestaltung 53. Unterlassung 54. Urheber 55. urheberrechtlich geschützte Bearbeitung 56. urheberrechtliche Werkqualität 57. Urteilsveröffentlichung 58. Vermögen 59. Veröffentlichungsermächtigung 60. Vervielfältigung 61. Verwendungsanspruch 62. Werke 63. Werknutzungsberechtigten 64. Werknutzungsentgelt 65. Werknutzungsrecht 66. Zeitgeist	45. schöpferische Gestaltungskraft 46. Schöpfung 47. Substrat 48. Tatsachenvorbringen 49. Teilurteil 50. Unterlassung 51. Unterlassungsbegehren 52. Urheber 53. Urheberrecht 54. Urteilsveröffentlichung 55. Vermögen 56. Veröffentlichungsbegehren 57. Veröffentlichungsermächtigung 58. Verwendungsanspruch 59. Werk 60. Werknutzungsberechtigten 61. Werknutzungsrecht 62. Werkqualität
---	---	--	--	---

Anhang 14

ChatGPT Text 4 (Anh. 4) Prompt 1	ChatGPT Text 4 (Anh. 4) Prompt 2	ChatGPT Text 4 (Anh. 4) Prompt 3	ChatGPT Text 4 (Anh. 4) Prompt 4	ChatGPT Text 4 (Anh. 4) Prompt 5
-------------------------------------	-------------------------------------	-------------------------------------	-------------------------------------	-------------------------------------

1.Aranżacja wnętrz 2.Art. 79 Ustawy o prawie autorskim i prawach pokrewnych 3.Biuletyn Informacji Publicznej 4.Decyzja administracyjna 5.Decyzja o lokalizacji 6.Decyzja o warunkach zabudowy 7.Dokumentacja powykonawcza 8.Dokumentacja projektowa 9.Działalność gospodarcza 10.Główny Urząd Nadzoru Budowlanego 11.Inwestycja 12.Komornik sądowy 13.Koszty procesu 14.Koszty zastępstwa procesowego 16.Licencja 17.Majątkowe prawa autorskie 18.Majątkowe prawa autorskie 19.Materiały marketingowe 20.Nadzór autorski 21.Naprawienie szkody 22.Nienależyte wykonanie zobowiązania 23.Odsetki ustawowe za opóźnienie 24.Odsetki ustawowe za opóźnienie 25.Odstąpienie od umowy 26.Ogólnopolskie wydanie	1.Art. 79 Ustawy z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych 2.Bezpodstawne wz. 3.Biuletyn Informacji Publicznej 4.Decyzja o lokalizacji zjazdu 5.Decyzja o pozwoleniu na budowę 6.Decyzja o warunkach zabudowy 7.Dokumentacja powykonawcza 8.Dokumentacja projektowa 9.Działalność gospodarcza 10.Ex tunc 11.Fotorealistyczne wizualizacje 12.Gazeta 13.Główny Urząd Nadzoru Budowlanego 14.Instytucja pożytku publicznego 15.Inwestycja 16.Komornik sądowy 17.Koncepcja zabudowy 18.Koszty procesu 19.Koszty zastępstwa procesowego 20.Koszty zastępstwa procesowego 21.Majątkowe prawa autorskie 22.Materiał marketingowy 23.Materiały marketingowe 24.Nadzór autorski	1.Aranżacja wnętrz 2.Art. 79 Ustawy o prawie autorskim i prawach pokrewnych 3.Bezprawność działań 4.Biuletyn Informacji Publicznej 5.Decyzja administracyjna 6.Decyzja o lokalizacji zjazdu 7.Decyzja o warunkach zabudowy 8.Dokumentacja projektowa 9.Działalność gospodarcza 10.Faktura 11.Główny Urząd Nadzoru Budowlanego 12.Inwestycja 13.Karty projektowe (opisowe i graficzne) 14.Koncepcja zabudowy 15.Koszty procesu 16.Koszty zastępstwa procesowego 17.Licencja 18.Majątkowe prawa autorskie 19.Należyta staranność 20.Naprawienie szkody 21.Naruszenie praw autorskich 22.Odstąpienie od umowy 23.Odstąpienie od umowy ex tunc 24.Odwołanie od umowy 25.Oględziny 26.Opłata skarbową 27.Osobiste prawa autorskie 28.Oświadczenie	1.Art. 79 Ustawy o prawie autorskim i prawach pokrewnych 2.Bezprawność działań 3.Biuletyn Informacji Publicznej 4.Decyzja administracyjna 5.Decyzja o lokalizacji zjazdu 6.Decyzja o warunkach zabudowy 7.Dokumentacja powykonawcza 8.Dokumentacja projektowa 9.Działalność gospodarcza 10.Ex tunc 11.Gazeta ogólnopolska 12.Główny Urząd Nadzoru Budowlanego 13.Karty projektowe (opisowe i graficzne) 14.Komornik sądowy 15.Koncepcja „wstecznej bezprawności” 16.Koncepcja zabudowy 17.Koszty procesu 18.Koszty zastępstwa procesowego 19.Majątkowe prawa autorskie 20.Materiały marketingowe 21.Nadzór autorski 22.Naprawienie szkody 23.Nienależyte wykonanie umowy 24.Odbiory 25.Odsetki ustawowe	1.Autorskie prawa majątkowe 2.Bezpieczeństwo pracy 3.Bezsprawność działań podmiotu korzystającego z utworu 4.Biuletyn Informacji Publicznej 5.Decyzja o lokalizacji zjazdu 6.Decyzja o pozwoleniu na budowę 7.Decyzja o warunkach zabudowy 8.Dokumentacja powykonawcza 9.Dokumentacja projektowa 10.Działalność gospodarcza 11.Fotorealistyczne wizualizacje 12.Główny Urząd Nadzoru Budowlanego 13.Informacja o autorstwie 14.Inwestycja 15.Jednostkowe wynagrodzenie 16.Karty projektowe 17.Komornik sądowy 18.Koncepcje zabudowy 19.Koszty procesu 20.Koszty zastępstwa procesowego 21.Licencja 22.Majątkowe prawa autorskie 23.Materiały marketingowe 24.Nadzór autorski 25.Naruszenie osobistych praw autorskich 26.Naruszenie praw autorskich
---	--	---	---	--

27.Okres przedawnienia 28.Opłata skarbową od pełnomocnictwa 29.Osobiste prawa autorskie 30.Osobiste prawa autorskie 31.Oświadczenie 32.Pełnomocnik 33.Petition pozwu 34.Powód 35.Pozew 36.Pozwana 37.Pozwolenie na budowę 38.Pozwolenie na użytkowanie 39.Protokół stanu faktycznego 40.Przedawnienie roszczeń 41.Przysługujące prawa autorskie 42.Roszczenia 43.Rozpowszechnianie 44.Skutki naruszenia praw autorskich 45.Stawki minimalne 46.Ugoda 47.Umowa 48.Wezwanie 49.Wizualizacje fotorealistyczne 50.Wsteczna bezprawność 51.Wynagrodzenie 52.Wynagrodzenie stosowne 53.Zaświadczenie o samodzielności lokali	25.Nienależyte wykonanie umowy 26.Odsetki ustawowe za opóźnienie 27.Odstąpienie od umowy 28.Odszkodowanie 29.Opłata od pełnomocnictwa 30.Opłata skarbową od pełnomocnictwa 31.Oświadczenie 32.Oświadczenie o odstąpieniu od umowy 33.Oświadczenie pozwanej 34.Pełnomocnik 35.Porozumienie 36.Powód 37.Pozew 38.Pozew 39.Pozwana 40.Pozwolenie na użytkowanie 41.Projekt aranżacji wnętrz 42.Projekt zjazdu do garażu podziemnego 43.Protokół stanu faktycznego 44.Przedawnienie 45.Przeprosiny 46.Przysługujące prawa autorskie 47.Roszczenie 48.Roszczenie 49.Rozpowszechnianie 50.Stosowne wynagrodzenie 51.Stowarzyszenie 52.Ugoda 53.Układ drogowy 54.Umowa	29.Oświadczenie o odstąpieniu od umowy 30.Pełnomocnik 31.Pełnomocnik 32.Petition pozwu 33.Powód 34.Pozew 35.Pozwana 36.Prawa autorskie 37.Projekt architektoniczny 38.Protokół stanu faktycznego 39.Przebudowa układu drogowego 40.Przedawnienie 41.Roszczenia 42.Rozpowszechnianie 43.Skutek ex tunc 44.Stowarzyszenie 45.Ugodą 46.Umowa 47.Umowa o dzieło 48.Ustawowe odsetki za opóźnienie 49.Wizualizacje fotorealistyczne 50.Wycena prac projektowych 51.Wynagrodzenie 52.Zwolnienie od kosztów sądowych	26.Odsetki ustawowe za opóźnienie 27.Odstąpienie od umowy 28.Opłata skarbową 29.Osobiste prawa autorskie 30.Oświadczenie 31.Pełnomocnik 32.Porozumienie 33.Powód 34.Pozew 35.Pozwana 36.Pozwolenie na budowę 37.Pozwolenie na użytkowanie 38.Projekt aranżacji wnętrz 39.Projekt zjazdu do garażu podziemnego 40.Protokół stanu faktycznego 41.Przebudowa układu drogowego 42.Przedawnienie roszczeń 43.Przysługujące prawa autorskie 44.Roszczenie 45.Rozpowszechnianie 46.Stosowne wynagrodzenie 47.Stowarzyszenie 48.Ugoda 49.Umowa 50.Umowa licencyjna 51.Umowa nr 52.Wizualizacje fotorealistyczne 53.Wycena prac projektowych 54.Wynagrodzenie 55.Zaświadczenie o samodzielności lokali	27.Obługa w zakresie odbiorów 28.Odstąpienie od umowy 29.Odstąpienie od umowy ex tunc 30.Ogólnopolskie wydanie 31.Opisowe i graficzne projekty aranżacji wnętrz 32.Opłata skarbową 33.Osobiste prawa autorskie 34.Ostateczna modyfikacja roszczeń 35.Oświadczenia pełnomocników 36.Oświadczenie w formie ogłoszenia 37.Pełnomocnik 38.Portal społecznościowy 39.Powód 40.Pozew 41.Pozwana 42.Pozwolenie na użytkowanie 43.Prawa autorskie 44.Prawo autorskie i prawa pokrewne 45.Projekt architektoniczny 46.Projekt zjazdu 47.Projekty aranżacji wnętrz 48.Protokół stanu faktycznego 49.Przebudowa układu drogowego 50.Przedawnienie roszczeń 51.Przeprosiny za naruszenie praw autorskich 52.Roszczenie o naprawienie szkody
--	--	--	--	--

	55.Umowa drogowa 56.Umowa licencyjna 57.Umowa nr 58.Ustawowe odsetki za opóźnienie 59.Wezwanie 60.Wizualizacje fotorealistyczne 61.Wsteczna bezprawność 62.Wycena prac projektowych 63.Wynagrodzenie 64.Zastępstwo procesowe 65.Zaświadczenie o samodzielności lokali		56.Zezwolenie 57.Zezwolenie na korzystanie z utworu	53.Roszczenie o zapłatę dwukrotności wynagrodzenia 54.Rozpowszechnianie utworów 55.Skutek ex tunc 56.Staranność 57.Treść umowy 58.Ugoda 59.Umowa 60.Ustawowe odsetki za opóźnienie 61.Usunięcie skutków naruszenia 62.Wizualizacje fotorealistyczne 63.Wsteczna bezprawność 64.Wycena prac projektowych 65.Wysokość roszczenia 66.Zaświadczenia o samodzielności lokali 67.Zezwolenie na korzystanie z utworów 68.Zjazd do garażu podziemnego 69.Zwolnienie od kosztów sądowych 70.Zwrot uiszczonego wynagrodzenia
--	---	--	--	---

Anhang 15

ChatGPT Text 5 (Anh. 5) Prompt 1	ChatGPT Text 5 (Anh. 5) Prompt 2	ChatGPT Text 5 (Anh. 5) Prompt 3	ChatGPT Text 5 (Anh. 5) Prompt 4	ChatGPT Text 5 (Anh. 5) Prompt 5
---	---	---	---	---

1.Dowody 2.Dyrektywa 3.Fakt 4.Kodeks postępowania cywilnego (k.p.c.) 5.Licencja wyłączna 6.Naruszenie prawa 7.Nośnik informacji 8.Obowiązana 9.Okoliczności sprawy 10.Organ sądowy 11.Oświadczenie 12.Państwa członkowskie 13.Pismo procesowe 14.Pochodzenie i sieci dystrybucji 15.Pola eksploatacji 16.Postępowanie sądowe 17.Prawa własności intelektualnej 18.Prawo wydawania 19.Przekład 20.Roszczenie 21.Rozpowszechnianie 22.Sąd 23.Sprzedaż wysyłkowa 24.Środek dowodowy 25.Stan faktyczny 26.Sygn. akt 27.Tłumacz 28.Twierdzenia stron 29.Umowa licencyjna 30.Umowa wydawnicza 31.Uprawdopodobnienie 32.Ustalenie stanu faktycznego 33.Uzasadnienie orzeczenia	1.Art. 228 § 2 k.p.c. 2.Art. 233 § 1 k.p.c. 3.Audiobook 4.Doktryna 5.Dyrektywa 2004/48/WE 6.Dz. U. 7.E-book 8.Egzekwowanie praw własności intelektualnej 9.k.p.c. 10.LEX 11.Licencja wyłączna 12.Nośnik informacji 13.Obrót 14.OSNP 15.Paradygmat racjonalnego ustawodawcy 16.Pismo procesowe 17.Pola eksploatacji 18.Postanowienie 19.Postępowanie dowodowe 20.Postępowanie w sprawach własności intelektualnej 21.Prawo do wydania przekładu 22.Prawo własności intelektualnej 23.Preambuła 24.Rozporządzenie Prezesa Rady Ministrów 25.Rozpowszechnianie 26.Sprzedaż wysyłkowa 27.Środek dowodowy 28.Stan faktyczny 29.Sygn. akt 30.Tłumaczenia literatury anglojęzycznej	1.Dokumenty 2.Dyrektywa 3.Dyski magnetyczne 4.Dyski optyczne 5.Licencja wyłączna 6.Naruszenie prawa 7.Nośniki informacji 8.Okoliczności 9.Pamięć komputera 10.Pismo procesowe 11.Pola eksploatacji 12.Postanowienie o dopuszczeniu dowodu 13.Postępowanie sądowe 14.Przeniesienie praw 15.Rozpowszechnianie 16.Sieci dystrybucji 17.Skala handlowa 18.Sprzedaż wysyłkowa 19.Środek dowodowy 20.Stan faktyczny 21.Stan faktyczny sprawy 22.System on line 23.Twierdzenia stron 24.Udowodnienie faktu 25.Umowa dowodowa 26.Umowa wydawnicza 27.Walor wiarygodności 28.Wezwanie do udzielenia informacji 29.Wiarygodne wykazanie okoliczności 30.Własność intelektualna 31.Wniosek 32.Wnioskodawca 33.Wprowadzenie do obrotu	1.Audiobook 2.Dokumenty 3.Dowód 4.Dyrektywa 5.E-book 6.Egzekwowanie praw 7.Fakty 8.Licencja wyłączna 9.Naruszenie prawa 10.Nośnik informacji 11.Pismo procesowe 12.Pola eksploatacji 13.Postanowienie o dopuszczeniu dowodu 14.Postępowanie sądowe 15.Prawa wyłączne 16.Prawo wydawania utworu 17.Rozporządzenie Prezesa Rady Ministrów 18.Rozpowszechnianie 19.Sieci dystrybucji 20.Sprzedaż wysyłkowa 21.Stan faktyczny 22.System on line 23.Tłumacz 24.Tłumaczenia literatury 25.Umowa dowodowa 26.Umowa wydawnicza 27.Uprawdopodobnienie roszczenia 28.Ustawa 29.Wezwanie do udzielenia informacji 30.Wiarygodność 31.Własność intelektualna 32.Wprowadzenie do obrotu	1.Art. 126 § 1 pkt 5 k.p.c. 2.Art. 228 § 2 k.p.c. 3.Art. 233 § 1 k.p.c. 4.Audiobook 5.Dopuszczenie dowodu 6.Dowody 7.Dyrektywa 2004/48/WE Parlamentu Europejskiego i Rady 8.Dyski magnetyczne 9.Dyski optyczne 10.E-book 11.E-booki 12.Egzekwowanie praw własności intelektualnej 13.Korespondencja mailowa 14.Licencja wyłączna 15.Naruszenie prawa 16.Nośnik informacji 17.Obowiązana 18.Oświadczenia 19.Pamięć komputera 20.Pismo procesowe 21.Pojęcie wykazania 22.Pola eksploatacji 23.Pominięcie dowodu 24.Portale (publio, upolujeboka.pl, Księgarnia medyczna) 25.Postanowienie Sądu Najwyższego 26.Postępowanie sądowe 27.Pozwana 28.Prawo własności intelektualnej 29.Preambuła dyrektywy
---	---	---	---	---

34. Waler wiarygodności 35. Wiarygodność 36. Wniosek 37. Wnioskodawca 38. Wprowadzenie do obrotu 39. Wydanie przekładu 40. Wydawca 41. Wykorzystanie handlowe 42. Wyrok 43. Zwielokrotnianie utworu	31. Umowa dowodowa 32. Umowa wydawnicza 33. Uprawdopodobnienie 34. Uprawdopodobnienie roszczenia 35. Ustalenie stanu faktycznego 36. Ustawodawca 37. Wiarygodność 38. Wniosek o wezwanie do udzielenia informacji 39. Wnioskodawca 40. Wprowadzenie do obrotu 41. Wykorzystanie handlowe 42. Wyłączne prawo wydawania 43. Wyrok 44. Zabezpieczenie środka dowodowego 45. Zasady techniki prawodawczej 46. Zwielokrotnianie utworu	34. Wyjawienie środka dowodowego 35. Wykorzystanie handlowe 36. Zabezpieczenie środka dowodowego 37. Zwielokrotnianie utworu	33. Wydawnictwo 34. Wyjawienie środka dowodowego 35. Wykorzystanie handlowe 36. Zabezpieczenie środka dowodowego 37. Zwielokrotnianie utworu	30. Przeniesienie prawa 31. Przeprowadzenie dowodu 32. Rozporządzenie Prezesa Rady Ministrów 33. Rozpowszechnianie utworu 34. Sieci dystrybucji 35. Skala handlowa 36. Sprzedaż wysyłkowa 37. Stan faktyczny 38. Sygn. akt 39. System on-line 40. Tłumacz 41. Umowa dowodowa 42. Umowa licencyjna 43. Umowa wydawnicza 44. Uprawdopodobnienie 45. Uprawdopodobnienie roszczenia 46. Uzasadnienie orzeczenia 47. Wiarygodność dokumentów 48. Wnioskodawca 49. Wprowadzenie do obrotu 50. Wskazanie dowodu 51. Wszechstronne rozważenie materiału 52. Wydanie środka dowodowego 53. Wydawnictwo 54. Wyjaśnienia informacyjne 55. Wykorzystanie handlowe 56. Wyłączne prawo wydawania 57. Zabezpieczenie środka dowodowego 58. Zakres naruszenia prawa 59. Zarzuty
--	--	---	--	--

				60.Zasady techniki prawodawczej 61.Źródło naruszenia prawa 62.Zwielokrotnianie utworu
--	--	--	--	--

Anhang 16

ChatGPT Text 6 (Anh. 6) Prompt 1	ChatGPT Text 6 (Anh. 6) Prompt 2	ChatGPT Text 6 (Anh. 6) Prompt 3	ChatGPT Text 6 (Anh. 6) Prompt 4	ChatGPT Text 6 (Anh. 6) Prompt 5
1.Apelacja 2.Art. 102 ust. 1 ustawy Prawo własności przemysłowej 3.Art. 229 k.p.c. 4.Art. 24 p.w.p. 5.Art. 79 ust. 1 pkt 3 lit. b Pr. aut. 6.Autorskie prawa majątkowe 7.Biegły 8.Dobra niematerialne 9.Dokumentacja projektowa 10.Dokumentacja techniczna 11.Dzieło techniczne 12.Idee 13.Indywidualność 14.Indywidualny charakter 15.Kwalifikacja dzieł 16.Materiał dowodowy 17.Metody 18.Nowość 19.Ocena dowodów 20.Ochrona prawnoautorska 21.Odkrycia 22.Odszkodowanie	1.akt twórczości 2.art. 79 ust. 1 pkt 3 lit. b Pr. aut. 3.autorskie prawa majątkowe 4.cechy wyróżniające 5.dobra niematerialne 6.dobro niematerialne 7.dobro niematerialne 8.dokumentacja projektowa 9.dokumentacja techniczna 10.działalność twórcza 11.dzieło techniczne 12.idea 13.indywidualny charakter dzieła 14.indywidualny wkład pracy twórczej 15.k.p.c. (Kodeks postępowania cywilnego) 16.koncepcja matematyczna 17.metoda działania 18.ocena dowodów 19.ochrona prawna 20.ochrona prawna	1.akt twórczości 2.art. 79 ust. 1 pkt 3 lit. b Pr. aut. 3.autorskie prawa majątkowe 4.cecha indywidualności 5.cechy linii, konturów, kształtów, kolorystyka, faktura, materiał wytworu, ornamentacja 6.definicja wzoru przemysłowego 7.dobro niematerialne 8.dokumentacja projektowa 9.dokumentacja techniczna 10.dzieło techniczne 11.dzieło techniczne 12.indywidualny charakter 13.indywidualny charakter dzieła 14.kwalifikacja dzieł 15.materiał dowodowy 16.moment stworzenia utworu 17.nowość twórcza	1.Apelacja 2.Autorskie prawa majątkowe 3.Cechy indywidualności 4.Dobro niematerialne 5.Dokumentacja techniczna 6.Dowody 7.Działalność twórcza 8.Dzieło techniczne 9.Fakty 10.Idee 11.Indywidualny charakter 12.Kładka mostowa 13.Kodeks postępowania cywilnego (k.p.c.) 14.Koncepcje matematyczne 15.Kreacyjność 16.Kryteria rejestracji 17.Linie, kontury, kształty, kolorystyka, faktura, ornamentacja 18.Metody 19.Ochrona prawa autorskiego 20.Ochrona prawnoautorska 21.Odkrycia	1.Akt twórczości 2.Autorskie prawa majątkowe 3.Biegły sądowy 4.Cechy indywidualności 5.Definicja wzoru przemysłowego 6.Dobra niematerialne 7.Dokumentacja techniczna 8.Działalność twórcza 9.Dzieła techniczne 10.Indywidualny wkład pracy twórczej 11.Kodeks postępowania cywilnego (k.p.c.) 12.Konstrukcja kładek i mostów 13.Kwalifikacja dzieł 14.Materiał dowodowy 15.Nowość i poziom wynalazczy 16.Ocena dowodów 17.Ochrona prawnoautorska 18.Odkrycia, idee, procedury, metody i zasady działania

23.Opinia biegłego 24.Orzecznictwo 25.Patent 26.Powód 27.Poziom wynalazczy 28.Prawo autorskie 29.Prawo autorskie i prawa pokrewne 30.Procedury 31.Projekt koncepcyjny 32.Projekt techniczny 33.Przejaw działalności twórczej 34.Przemysłowe stosowanie 35.Rezultat działalności twórczej 36.Sąd Apelacyjny 37.Sąd Najwyższy 38.Sąd Okręgowy 39.Sposób wyrażenia 40.Subiektywnie nowy 41.Umowa przelewu 42.Utwór techniczny 43.Własność przemysłowa 44.Wynalazek 45.Wzór przemysłowy 46.Zasady działania	21.ochrona prawnoautorska 22.odszkodowanie 23.opinia biegłego 24.opinie biegłych 25.orzecznictwo 26.p.w.p. (Prawo własności przemysłowej) 27.patent 28.postać nieukończona 29.poziom wynalazczy 30.poziom wynalazczy 31.prawa pokrewne 32.prawo autorskie 33.procedura 34.projekt koncepcyjny 35.projekt techniczny 36.przedmiot prawa autorskiego 37.przemysłowe stosowanie 38.przesłanki 39.przesłanki nowości 40.przeżycia intelektualne 41.rejestracja wzoru 42.rozwiązanie techniczne 43.sposób wyrażenia 44.swoboda twórcza 45.szkie koncepcyjny 46.umowa przelewu 47.umowa przelewu 48.ustalenie 49.ustawa o prawie autorskim i prawach pokrewnych 50.utwór 51.wynalazek 52.wynik działalności twórczej 53.wyniki samodzielnego wysiłku autora	18.nowość, poziom wynalazczy, przemysłowe zastosowanie 19.ocena dowodów 20.ocena opinii biegłych 21.ochrona prawnoautorska 22.ochrona prawnoautorska 23.odkrycia, idee, procedury, metody i zasady działania, koncepcje matematyczne 24.odszkodowanie 25.opinia biegłego 26.opinie biegłych 27.orzecznictwo 28.patent 29.patenty 30.poziom wynalazczy 31.prawo autorskie 32.proces twórczy 33.projekt koncepcyjny kładki mostowej 34.projekt koncepcyjny kładki mostowej 35.projekt techniczny 36.przedmiot materialny 37.przedmiot ochrony prawnoautorskiej 38.przejaw działalności twórczej 39.przejaw niepowtarzalności 40.przemysłowe stosowanie 41.przesłanki rejestracji wzoru 42.przyznanie faktów 43.rezultat 44.rezultat działalności technicznej	22.Odszkodowanie 23.Opinia rzecznika patentowego 24.Opinie biegłych 25.Orzecznictwo 26.Osoby trzecie 27.Patent 28.Piśmiennictwo 29.Poziom wynalazczy 30.Prawo autorskie 31.Prawo własności przemysłowej (p.w.p.) 32.Procedury 33.Proces twórczy 34.Projekt koncepcyjny 35.Projekt techniczny 36.Przemysłowe stosowanie 37.Sąd Okręgowy 38.Sfera przeżyć intelektualnych twórcy 39.Sposób wyrażenia 40.Subiektywnie nowy 41.Swoboda twórcza 42.Szkic koncepcyjny 43.Techniczne rozwiązanie 44.Umowa przelewu 45.Ustawa o prawie autorskim i prawach pokrewnych 46.Utwór 47.Wynalazek 48.Wzór przemysłowy 49.Zasady działania	19.Opinia rzecznika patentowego 20.Orzecznictwo 21.Patent 22.Piśmiennictwo 23.Prawo autorskie 24.Prawo autorskie i prawa pokrewne 25.Prawo własności przemysłowej 26.Projekt koncepcyjny 27.Przedmiot ochrony prawnoautorskiej 28.Przemysłowe stosowanie 29.Przesłanki rejestracji 30.Przyznanie faktów (art. 229 k.p.c.) 31.Sposób wyrażenia 32.Świadczenie ochronne 33.Swoboda twórcza 34.Szkic koncepcyjny 35.Twórczość autorska 36.Umowa przelewu 37.Utwór 38.Utwór techniczny 39.Wynalazek 40.Wzór przemysłowy
--	---	--	--	--

	54.wyrażenie odkrycia 55.wyrażenie techniczne 56.wyrok Sądu Apelacyjnego 57.wyrok Sądu Najwyższego 58.wytwór pracy umysłowej 59.wzór przemysłowy 60.zarzut procesowy 61.zasada działania	45.rezultat działalności twórczej 46.rozwiązania kreacyjne 47.swoboda twórcza 48.szkie koncepcyjny 49.techniczne rozwiązanie 50.umowa przelewu 51.ustalenie 52.ustawa o prawie autorskim i prawach pokrewnych 53.ustawa z dnia 30 czerwca 2000 r. Prawo własności przemysłowej 54.utwór 55.wkład pracy twórczej 56.wynalazek 57.wzór przemysłowy 58.wzornictwo przemysłowe 59.zarzuty procesowe		
--	---	---	--	--

Abstract (Deutsch)

Die Rechtssprache ist aus vielerlei Hinsicht ein besonderer Fall, sei es aufgrund der Rechtsterminologie, die sowohl fachspezifische Termini als auch gemeinsprachliche Ausdrücke mit der fachlichen Bedeutung enthält oder hinsichtlich der Tatsache, dass sie, anders als viele andere Fachsprachen, alle Gesellschaftsschichten einschließt. Ausgerechnet dadurch weist Rechtsterminologie ein breites Nutzungsspektrum sowohl im privaten als auch im beruflichen Kontext auf. Infolgedessen steigt, insbesondere für gewisse Berufsgruppen, der Bedarf an der automatischen Gewinnung rechtlicher Termini aus Fachtexten und somit an einem verlässlichen Extraktionstool, das mit hoher Präzision korrekte Termini gewinnt, da Verwendung fehlerhafter Rechtsterminologie ggf. gravierende Konsequenzen mit sich ziehen kann.

Die vorliegende Arbeit geht diese Herausforderungen aus der der Sicht der praktischen Terminologearbeit an. Im Rahmen einer Untersuchung wird die Leistung von ChatGPT-4o anhand von fünf Prompts und 6 Textfragmenten mit der Leistung eines spezialisierten Extraktionstools Text2TCS und mit einem manuell erstellten Goldstandard anhand der Qualitätsmetriken Genauigkeit, Trefferquote und F1-Score ermittelt. Untersucht wird die Leistung beider Extraktionssysteme am Beispiel der österreichischen und polnischen Rechtsterminologie aus dem Gebiet des Urheberrechts, mit dem Ziel, ihre Eignung für diese Aufgabe gezielt für die genannten Sprachen zu bestimmen.

Anhand der erhobenen Daten lässt sich ein Verbesserungsbedarf bei der automatischen Gewinnung von österreichischer und polnischer Rechtsterminologie feststellen. Die Untersuchung zeigte, dass das spezialisierte Extraktionstool bessere Ergebnisse als ChatGPT verzeichnete. Außerdem generierte ChatGPT mit unterschiedlichen Prompts jeweils eine andere Liste mit Termkandidaten. Schließlich konnten bessere Ergebnisse für Deutsch als für Polnisch erzielt werden.

Die Erkenntnisse der vorliegenden Arbeit sollten in erster Linie einen Beitrag zu der laufenden Forschung am Gebiet der Termextraktion durch LLMs leisten. Die Untersuchung des aktuellen Leistungsstands von ChatGPT bei der automatischen Termextraktion in anderen Sprachen als Englisch hilft die Forschungslücke zu schließen und erteilt Auskunft darüber, wie verlässlich dabei aktuell das Tool ist, und dass zukünftig noch mehr an diesem Gebiet getan werden muss.

Abstract (English)

More often than not, legal language can pose a challenge on account of its extensive terminology, consisting not only of specialist terms but also of common language expressions with specialist legal meanings. Unlike many other specialist languages, legal language is not limited solely to a group of experts. Instead, it affects people from all strata of society. The widespread use of legal terminology, both in private and professional contexts, requires, especially for some occupational groups, a certain degree of automation in the field of term extraction from legal texts. Since the use of incorrect legal terms can result in serious consequences, there is a need for a reliable automatic term extraction tool to ensure the precision of the extracted legal terminology.

The following study addresses these challenges from the perspective of practical terminology work. It aims to determine the performance of ChatGPT-4o in the area of automatic term extraction by using five different prompts on 6 different text fragments. The results are then compared with the performance of a specialised extraction tool, Text2TCS, and with a manually created gold standard, using the evaluation metrics of precision, recall, and F1 score. The performance of both extraction tools are tested using Austrian and Polish legal texts from the field of copyright law in order to determine their applicability for the specific task of extracting legal Austrian and Polish terms.

Collected data indicates the need for improvement in the automatic extraction of Austrian and Polish legal terms. The study shows that the specialised extraction tool produces better results than ChatGPT. In addition, ChatGPT generates a different list of term candidates depending on the prompts. Finally, the study concludes that automatic term recognition obtains better results for German than for the Polish language.

The findings of the present work can primarily contribute to the ongoing research in the field of automatic term extraction by Large Language Models (LLMs). Investigating the current performance of ChatGPT in automatic term extraction in languages other than English, helps close the research gap and provides information on how reliable the tool currently is for this specific task, and that there is room for further improvement in this area in the future.