20. Jahrgang • 2005-2024

edition

Fachzeitschrift für Terminologie

Deutscher Terminologie-Tag e.V.
www.dttev.org

2|24

Worldbuilding: Benennungsbildung in der Fantasy-Literatur

Hochschulterminologie
Das Projekt Baden-Württerbuch
Seite 5

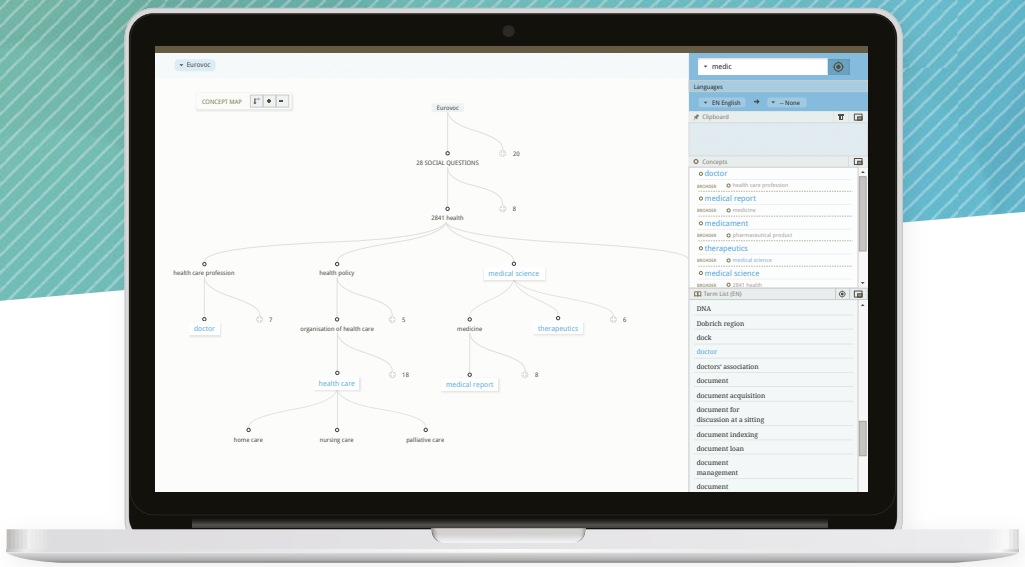
Vergleich
Software- vs. KI-gestützte Termextraktion
Seite 13

Normen
ISO 704 im Zeitalter der KI
Seite 39

DTT-Symposion 2025
Terminologie in der KI – KI in der Terminologie
27.-29. März in Worms

Your Multilingual Knowledge System

Extreme Visual Terminology



360-degree control through systematic approach: **multilingual Concept Maps**

Proven enterprise deployments, scalable with Single sign-on: appealing, **plug-in free browser solution**

Enable NMT workflows, tune enterprise search, annotate and classify texts: **versatile integrations**

coreon

Knowledge meets language.

coreon.com

Nibelungenhelden und Mondgötter

Der Science-Fiction-Roman „When the Moon Hatched“ zeichnet eine unvergessliche Welt, in der der Kampf um die Macht von mächtigen Gottheiten, uralter Magie und einem Himmel voller Drachen bestimmt wird. Literaturübersetzerin Christine Heinzius war Mitglied im Team, das die deutsche Übersetzung erstellt hat. Im **Interview** erzählt sie vom Worldbuilding, also davon, wie sie und ihre Kolleginnen und Kollegen solche Neuen Welten terminologisch erschließen.

Im **Hauptteil** stellen Vanessa Knies und Ellise Meyer das Projekt des Baden-Württerbuchs für das Hochschulwesen in Baden-Württemberg vor, Vanessa Jochum und Jasmin Nesbigall präsentieren eine Untersuchung, in der verschiedene Methoden der Terminologieextraktion auf ihre Effizienz hin durchleuchtet werden.

Die Rubrik **Tools & Ressourcen** hält viele Tipps bereit: eine nützliche Funktion, mit der man Websites in Apps umwandelt, ein Typing Assistant für schnelleres Tippen sowie die Funktionserweitungen von flashterm.

Schließlich stellt Barbara Karsch in der Rubrik **Normen** die Frage, wie ISO 704 Terminology Work im Zeitalter der KI von Nutzen sein kann – gefolgt von einer kurzen Vorstellung der neuesten Normen aus dem Terminologie- und Übersetzungsbereich.

Große Ereignisse werfen ihre Schatten voraus. Die Anmeldung zum Symposion 2025 ist jetzt freigeschaltet. Veranstaltungsort ist dieses Mal die sagenumwobene Nibelungenstadt Worms. Das Thema schaut weiter in die Zukunft: **Terminologie in der KI – KI in der Terminologie**. Nicht verpassen!

Wir wünschen eine interessante Lektüre und freuen uns auf zahlreiche Rückmeldungen.

Dr. Nicole Keller
Redaktionsleitung
redaktion@dttev.org

Angelika Ottmann
Redaktionsleitung
redaktion@dttev.org

TERM SOLUTIONS
termXact • termXplorer • myTerm

Haben Sie Ihr Terminologie-management im Griff?

Falls nicht, sollten Sie baldmöglichst dafür sorgen – vor allem wenn Sie Künstliche Intelligenz im Unternehmen nutzen wollen. Denn:

There is no AI without terminology.

Ob in ChatBots, beim Prompting oder in der Maschinellen Übersetzung: Das Corporate Wording muss auch bei KI-erstellten Texten berücksichtigt werden.

Unsere ausgezeichneten Terminologie-Tools **myTerm**, **termXplorer** und **termXact online** unterstützen Sie dabei:

- **Multilingual** durch 3-Ebenen-Modell
- **Flexibel** für maßgeschneiderte Abstimmungsprozesse
- **Planbar** dank firmenweiter Lizenz
- **Skalierbar** als Pay-per-Feature
- **User-zentriert** mit übersichtlichem UX-Design und Filter-Dashboard
- **Leicht integrierbar** über einfachen Import und Export sowie REST-API und Single-Sign-On

Greifen Sie zu.

Alle Infos auf **termsolutions.ai**

Inhalt

Editorial

3 Nibelungenhelden und Mondgötter
Nicole Keller und Angelika Ottmann

Themen

5 Das Baden-Württerbuch: Hochschulterminologie für Baden-Württemberg
Vanessa Knies und Ellise Meyer

13 Das Beste rausholen – Termextraktion mit KI und Verstand
Vanessa Jochum und Jasmin Nesbigall

Aus dem Verband

10 DTT-Symposion 2025

23 20 Jahre edition

Interview

24 Worldbuilding – Benennungsbildung einmal anders
Literaturübersetzerin Christine Heinzius

Tools & Ressourcen

28 flashterm: Aktuelle Funktionserweiterungen im Überblick
Frauke Eschenbeck und Lola Klan Reyes

31 Aus Web mach App – oder wie man Websites in Apps umwandelt
Maurice Mayer-Dewor

34 Typing Assistant: Terminologie als Treibstoff für Predictive Typing
Armin Mutscheller

Normen

39 ISO 704 Terminology work – Principles and methods: Immer noch wichtig oder immer wichtiger?
Barbara Karsch

43 Neues aus der Normenwelt

20. Jahrgang.
edition erscheint zweimal im Jahr (Juni/Dezember).
ISSN 1862-023X.
Herausgeber: Deutscher Terminologie-Tag e.V. (DTT) – www.dttev.org
Redaktionsleitung: Dr. Nicole Keller und Angelika Ottmann – redaktion@dttev.org
Titelseite, DTT-Anzeigen, Layout und Satz: Tamara Arndt – layout@dttev.org
Anzeigen: Olga Buchstaller-Vodopiyanova – anzeigen@dttev.org
Lektorat: Maurice Mayer-Dewor
Mediadaten und komplettes Impressum: www.dttev.org/edition/impressum

Das Baden-Württerbuch: Hochschulterminologie für Baden-Württemberg

Vanessa Knies und Ellise Meyer

The online dictionary for higher education terminology of the State Coordination Office for Translations Relating to Higher Education in Baden-Württemberg at the University of Mannheim currently includes around 1,750 entries. It helps staff members of 45 higher education institutions to translate texts into English, write English texts and use terminology consistently.

Keywords: higher education institution terminology, online dictionary, internationalization

Seit einigen Jahren hat die Internationalisierung verstärkt Einzug in deutsche Hochschulen[1] gefunden. Dies betrifft neben diversen Austauschprogrammen für Mitarbeiter*innen und Studierende auch die Übersetzung der Infrastruktur, meist ins Englische als Lingua franca. Für viele Programme und Studienangebote sind keine Deutschkenntnisse erforderlich, was dennoch Auswirkungen auf die Verwaltung hat, da trotz englischsprachiger Angebote auch der in der Regel deutschsprachige Verwaltungsweg in der Hochschule beschritten werden muss. Die Individualität des deutschen Hochschulsystems sowie die Komplexität inneruniversitärer Verwaltungsprozesse stellen für die Übertragung deutscher Texte ins Englische oft eine besondere Herausforderung dar [1]. Vor diesem Hintergrund wurde 2016 die Landeskoordinationsstelle für Übersetzungsangelegenheiten im Hochschulwesen Baden-Württemberg geschaffen. Sie ist eine Serviceeinrichtung für die Themen englische Übersetzung, Terminologie und sprachliche Internationalisierung, die an der Universität Mannheim angesiedelt ist und von 45 Landeshochschulen getragen wird. Eine zentrale Ressource ist das Baden-Württerbuch, ein zweisprachiges Online-Wörterbuch mit Hochschulterminologie. Es soll Verwaltungsmitarbeiter*innen aus allen Bereichen der Hochschulverwaltung, darunter auch Übersetzer*innen, sowie wissenschaftlichen Mitarbeiter*innen als Ressource dienen. Ziel des Wörterbuchs ist, ausgewählte hochschulspezifische Begriffe zweisprachig aufzubereiten und somit die Verwendung der Bezeichnungen sowie die einheitliche Übersetzung hochschulspezifischer Texte zu erleichtern. Eingebettet ist das Online-Wörterbuch in das Portal der Landeskoordinationsstelle [1], auf dem auch alle weiteren Ressourcen wie Übersetzungen bereitgestellt werden. Zum Portal haben alle Mitarbeiter*innen der teilnehmenden Hochschulen Zugriff, derzeit sind rund 900 Personen registriert.

Das Baden-Württerbuch ist für das Sprachenpaar Deutsch – Englisch ausgelegt. Andere Sprachen deckt das Wörterbuch nicht ab. Dies liegt in der Zielgruppe der englischen Texte und Übersetzungen, in denen die Termini des Wörterbuchs verwendet werden sollen, begründet: Die ausländischen Studierenden und Gastwissenschaftler*innen, die für ihren Studien- oder Forschungsaufenthalt nach Deutschland kommen, stammen aus den unterschiedlichsten Ländern der Welt. Wir haben es also mit einer sehr heterogenen Personengruppe zu tun, für die Englisch als Lingua franca die naheliegende Wahl ist.

Sowohl für das Aufbaukonzept des Baden-Württerbuchs als auch für die Termsuche ist es wichtig zu verstehen, dass das Wörterbuch aus der deutschen Perspektive konzipiert wurde. Demnach bietet es englischsprachige Übersetzungslösungen[2] für deutschsprachige Bezeichnungen von Konzepten aus dem deutschen Bildungs- und Hochschulsystem. Dies ist vor dem Hintergrund der Systemunterschiede zwischen den verschiedenen Ländern bedeutend, denn Hochschulsysteme sind immer auch stark an das jeweilige nationale Rechtssystem geknüpft [2]. Diese Unterschiede führen oft zu einer Unübersetzbarkeit bestimmter Benennungen aufgrund mangelnder Äquivalente im Zielsystem oder aber zu „false friends“ (so ist beispielsweise die deutsche Professorenlaufbahn nicht mit der US-amerikanischen Laufbahn vergleichbar). Das Wörterbuch hat also nicht den Anspruch, etwa die US-amerikanische oder britische Hochschulsprache abzubilden.

Grundlage für die Erstellung des Wörterbuchs im Jahr 2019 war zunächst ein gemeinsames Glossar von 13 Landeshochschulen, dessen Einträge systematisch ausgewählt

[1] „Hochschule“ bezeichnet nach dem Landeshochschulgesetz (LHG) alle Hochschularten in Baden-Württemberg: Universitäten, Pädagogische Hochschulen, Kunsthochschulen, Hochschulen für Angewandte Wissenschaften und die Duale Hochschule Baden-Württemberg.
[2] Wir sprechen hier bewusst nicht von Äquivalenten, da es diese für hochschulspezifische Benennungen nicht immer gibt. Stattdessen finden wir häufig auch Ersatzlösungen wie Lehnübersetzungen, Umschreibungen, Neologismen etc.

und aufbereitet wurden. Zum Zeitpunkt der Veröffentlichung 2022 umfasste das Wörterbuch 1300 Einträge. Es wurde seitdem stetig erweitert und zählt heute rund 1750 Einträge.

In diesem Artikel sollen der Aufbau und die Nutzung des Wörterbuchs genauer vorgestellt werden. Dies umfasst sowohl die Auswahl von Termini und die Eintragsstruktur als auch die Suchfunktionen und technischen Grundlagen. Außerdem werden mögliche Kritikpunkte aufgeführt und begründet.

Zielgruppe

In der Landeskoordinationsstelle agieren wir als Sprachmittler*innen für eine heterogene Zielgruppe. Die Hauptnutzer*innen des Wörterbuchs sind deutschsprachige Hochschulmitarbeiter*innen aus diversen Arbeitsbereichen, vor allem aus Verwaltungs- und Serviceeinrichtungen. Sie haben in der Regel gemein, dass sie zwar keine Englischexpert*innen sind, aber Spezialist*innen in ihrem Fachgebiet. Das Wörterbuch nutzen sie im Rahmen der Internationalisierung, um englischsprachige Texte zu erstellen und unbekannte englischsprachige Benennungen nachzuschlagen. Zu einem kleinen Teil wird das Wörterbuch auch von Übersetzer*innen und Wissenschaftler*innen genutzt.

Aufgrund dieser besonderen Zielgruppe hat die Landeskoordinationsstelle nicht den Anspruch, mit dem Wörterbuch eine Sprachnorm zu schaffen – auch vor dem Hintergrund der jeweils hochschulinternen Vorgaben zur Corporate Language –, sondern vielmehr die Nutzer*innen in ihrer individuellen Sprachverwendung zu unterstützen.

Kriterien für die Auswahl von Termini

Generell werden keine allgemeinsprachlichen Termini in das Wörterbuch aufgenommen. Sie müssen immer einen Hochschulbezug haben und einem der vordefinierten Fachgebiete zugeordnet werden können. Die Liste der Fachgebiete (s. Abschnitt Aufbau) ist dabei erweiterbar, sofern eine gewisse Anzahl an Termini unter ein Fachgebiet fällt. Wichtig ist die Abgrenzung von Benennungen, die zwar häufig im Hochschulkontext verwendet, aber einem anderen Fachgebiet zugeordnet werden, z. B. Benennungen wie „Nachhaltigkeit“ und „Kooperation“. Des Weiteren finden sich im Wörterbuch zwar Komposita, jedoch keine Mehrwort-Benennungen.

Aufbau

Zur Erarbeitung der Eintragsstruktur wurden neben der einschlägigen Fachliteratur [3],[4] vor allem andere frei zugängliche wie auch hochschulinterne Online-Wörterbücher als Vorlage konsultiert, konkret die Terminologie-Datenbanken der Universität Zürich [5], Universität Freiburg [6], Johannes Gutenberg-Universität Mainz [7], Friedrich-Alexander-Universität Nürnberg-Erlangen [8] und die Terminologiedatenbank der EU, IATE [9]. Da die Zielgruppe des Wörterbuchs hauptsächlich aus Nicht-Übersetzer*innen bzw. Nicht-Terminolog*innen besteht, sollten die Einträge neben deren inhaltlicher Korrektheit entsprechend einfach, eindeutig und übersichtlich gestaltet werden. Diese Zielgruppenorientierung führte dazu, dass von den klassischen Terminologieeintragsstrukturen aus der Terminologielehre abgewichen wurde (s. Abschnitt Kritik). Es werden zum Beispiel nicht alle standardmäßigen Datenkategorien, wie grammatische Angaben oder Definitionen, aufgeführt. Die Benennungen der verwendeten Datenkategorien sind außerdem gemeinsprachlich, weshalb z. B. der „Kontext“ als „Satzbeispiel“ und „Kollokationen“ als „Wortverbindungen“ bezeichnet wurden.

Ein weiterer Anspruch war, das Wörterbuch nicht über die von den Hochschulen festgelegte hochschuleigene Terminologie (Corporate Language) zu stellen. Daher wurde das Eintragsfeld „Vorzugsbenennung“ integriert, unter dem die die jeweilige Benennung verwendende Institution aufgeführt wird. Das Baden-Württerbuch beinhaltet zum Großteil Substantive. Verben und Adjektive werden vornehmlich in den Wortverbindungen erfasst und sind durch die Volltextsuche auffindbar (s. Abschnitt Suchoptionen). Neben der Fachterminologie finden sich auch Eigennamen wie Hochschulnamen, Positionsbezeichnungen und Gesetzestitel im Wörterbuch.

Ein Eintrag enthält sechs Datenkategorien, die nachfolgend aufgeführt und erläutert werden. Abbildung 1 auf Seite 7 zeigt die Eintragsstruktur anhand des Beispielterminus „Studiendekan/Studiendekanin“.

Wortverbindungen

Unter den Wortverbindungen finden sich gängige Kollokationen und Komposita (s. auch Abbildung 2 auf Seite 7). Die englischen Einträge spiegeln dabei jeweils die deutschen. Ausgewählt werden die Kollokationen bzw. Komposita auf Grundlage von Fachwörterbüchern [10],[11],[12], Kollokationswörterbüchern [13],[14] sowie Einträgen aus den Translation Memorys der Servicestelle für Übersetzung der Universität Mannheim. Die Auswahl der Kollokationen und Komposita erfolgt subjektiv und nicht aufgrund statistischer Auswertungen der tatsächlichen Häufigkeit im Sprachgebrauch.

Satzbeispiele

Hier handelt es sich nicht um reine Kontextbeispiele, sondern um Sätze, die im Arbeitsalltag hilfreich sein sollten. Auch hier entsprechen die deutschen den englischen Beispielen. Die Satzbeispiele sind allgemein gehalten, da sie für über 40 Hochschulen geeignet sein müssen. Sie stammen, neben den unter „Wortverbindungen“ genannten Quellen, von Webseiten etablierter Hochschulen aus dem englischsprachigen Raum, den Übersetzungen von Bundes- [15] und EU-Gesetzen [16] sowie dem Bundesinnenministerium [17].

Fachgebiet
Auch im Hochschulkontext gibt es Bezeichnungen, die je nach Kontext unterschiedliche Bedeutungen haben und somit einem bestimmten Fachgebiet zugeordnet werden können. Ein Eintrag kann in der Regel mehreren Fachgebieten angehören. Derzeit gibt es 27 Fachgebiete, darunter Ämter und Positionen, Bewerbung und Zulassung, Forschung, Kurse und Module, Organisationseinheiten, Personalwesen und Prüfungsangelegenheiten.

Hinweise
Hinweise beinhalten nützliche Zusatzinformationen für die Nutzer*innen, zum Beispiel Informationen zur Groß- und Kleinschreibung, der Verwendung der Bezeichnung im Englischen etc.

Vorzugsbenennung
Dieses Feld gibt an, an welcher Hochschule die jeweilige Benennung verwendet wird. Es gibt dementsprechend für einen Begriff teilweise mehrere Einträge bzw. Benennungen. Aktuell haben lediglich die 13 Hochschulen des Grundglossars ihre Vorzugsbenennungen mit der Landeskoordinationsstelle geteilt.

Wenn keine Vorzugsbenennung aufgeführt ist, bedeutet dies entweder, dass der Großteil der Hochschulen die gleiche Übersetzung verwendet, oder, dass es sich um allgemeine, hochschul-unspezifische Lösungen handelt oder die aufgeführte(n) Variante(n) die Empfehlung vonseiten der Landeskoordinationsstelle darstellt/darstellen.

Verwandte Einträge
Hier wird auf andere Einträge verlinkt, die inhaltlich mit dem Begriff verwandt sind und z. B. den gleichen Wortstamm haben.

	484	Studiendekan/Studiendekanin	dean of studies
Wortverbindungen		~ der Fakultät X	~ of the School/Faculty of X
Satzbeispiele		Prof. X ist der neue Studiendekan/die neue Studiendekanin der Fakultät X. Er war von X bis Y Studiendekan der Fakultät X. Der Studiendekan/die Studiendekanin bestimmt Zeit und Ort der Einsichtnahme. Vorschläge sind an den Studiendekan/die Studiendekanin zu richten. Die Entscheidung wurde von Prof. X, Studiendekanin der Fakultät X, getroffen.	Prof. X is the new dean of studies of the School/Faculty of X. From X to Y, he was dean of studies of the School/Faculty of X. The dean of studies decides on the time and place for accessing the examination records. Suggestions must be submitted to the dean of studies. The decision was made by Prof. X, dean of studies of the School/Faculty of X.
Fachgebiet		Ämter und Positionen	
Hinweise		Diese Übersetzung sollte sich nach der Übersetzung von "Dekan/Dekanin" richten. Im Englischen nur großgeschrieben bei alleinstehender Amtsbezeichnung eines bestimmten Studiendekans/einer bestimmten Studiendekanin, z.B. unter einer Unterschrift, in einer E-Mail-Signatur, einem Büroschild oder einer Kontaktadresse.	
Vorzugsbenennung		U Konstanz, HS Albstadt-Sigmaringen, HS Offenburg, U Heidelberg	
Verwandte Einträge		Dekan/Dekanin, Prodekan/Prodekanin	

Abb. 1: Eintragsstruktur des Baden-Württerbuchs (Quelle: eigener Screenshot)

	978	Studiengang	degree program (AE)/programme (BE)
Wortverbindungen		sich für einen ~ bewerben	to apply for a ~
		einen ~ aufnehmen	to take up a ~
		zu einem ~ zugelassen werden	to be admitted to a ~
		in einem ~ eingeschrieben sein	to be enrolled in a ~
		einen ~ abschließen	to complete a ~
		den ~ wechseln	to change one's ~
		Interesse an einem ~ haben	to be interested in a ~
		weiterführender ~	(post)graduate ~
		Studiengangswechsel	change of one's ~
		Studiengangskoordinator/in	~ coordinator
		Studiengangsleiter/in	~ director
		Studiengangssteckbrief	~ description
		Kooperationsstudiengang	cooperative ~

Abb. 2: Wortverbindungen (Quelle: eigener Screenshot)

Regeln für Eintragsinhalte

Für die Eintragsinhalte hat die Landeskoordinationsstelle diverse Regeln aufgestellt, die entweder sprachlich oder technisch bedingt sind.

- Positionsbezeichnungen werden sowohl in der weiblichen als auch in der männlichen Form angegeben und durch einen Schrägstrich getrennt. Die Verwendung von „*“ oder Ähnlichem zur nicht-binären Erfassung von Positionsbezeichnungen war technisch bislang nicht möglich, da Sucheinträge so nicht gefunden werden.
- Synonyme im Kollokationsfeld werden mit Schrägstrich voneinander getrennt (s. Abbildung 7 auf Seite 9).
- Wo relevant, werden in der englischen Benennung sowohl die amerikanische als auch die britische Schreibweise aufgeführt und mit (AE) oder (BE) gekennzeichnet.
- Offene Angaben, wie Daten, Namen oder Orte, werden als Platzhalter mit „X“ gekennzeichnet, da die Einträge hochschulunabhängig und allgemeingültig sein sollen.
- Bei Wortverbindungen wird die Bezeichnung durch eine Tilde (~) ersetzt. Falls ein Kompositum als Wortverbindung aufgeführt wird, muss die entsprechende Bezeichnung ausgeschrieben werden, z. B. „Abiturnote“ und nicht „~note“. Nur so kann der Terminus „Abiturnote“ in der Suche gefunden werden. Deklinationen bleiben von dieser Regel unberührt. Die Eintragsinhalte sind nicht für maschinenlesbare Zwecke gedacht, weshalb die Verwendung der Tilde keine Nutzungseinschränkung darstellt.
- Bei Eigennamen, beispielsweise Hochschulnamen, werden keine anderen Eintragsfelder befüllt.

Suchoptionen

Das Wörterbuch kann durch Filter gezielt durchsucht werden. Die Suchoptionen werden nachfolgend vorgestellt.

Allgemeine Suche

Die Suchfunktion ist sehr einfach aufgebaut: In das Suchfeld wird die gewünschte Benennung eingegeben (s. Abbildung 3). Über „Normale Suche“ oder die Enter-Taste wird die Suche ohne weiteren Filter direkt gestartet. Das Programm durchsucht in diesem Fall die gesamte Datenbank, unabhängig von Hochschule und Fachgebiet. Dies bietet sich zum Beispiel an, wenn alle Fachgebiete gleichzeitig durchsucht werden sollen oder der/dem Nutzer*in nicht klar ist, welchem Fachgebiet der Begriff zuzuordnen ist.

Suche nach Fachgebiet

Auch im Hochschulbereich gibt es bestimmte Bezeichnungen, die je nach Kontext unterschiedliche Bedeutungen haben können und somit im Englischen teilweise verschiedene Übersetzungen erfordern. Um schneller zur passenden

Abb. 3: Suchmaske mit Suchfiltern (Fachgebiet, Hochschule) und Suchoptionen (normale Suche, Volltextsuche) (Quelle: eigener Screenshot)

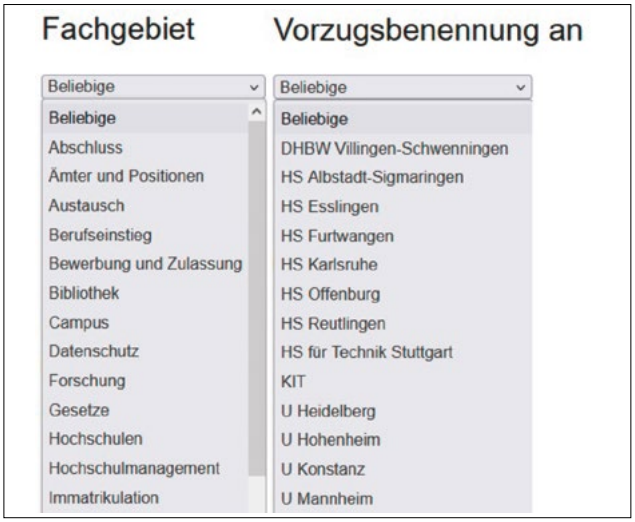


Abb. 4: Ausschnitte aus den Suchfiltern nach Fachgebiet und Hochschule (Quelle: eigener Screenshot)

Übersetzung für den jeweiligen Kontext zu gelangen, kann bei der Suche direkt das gewünschte Fachgebiet ausgewählt werden (s. Abbildung 4). Ergebnisse aus anderen Fachgebieten werden dann nur angezeigt, wenn es Überschneidungen gibt. So wird vermieden, dass Bezeichnungen ausgewählt werden, die eigentlich in einem anderen fachlichen Kontext verwendet werden. Ein Beispiel aus dem Hochschulkontext ist das „Referat“, das entweder die Organisationseinheit oder der Vortrag von Studierenden sein kann. Abbildungen 5 und 6 auf Seite 9 zeigen, wie der Begriffsunterschied im Baden-Württerbuch nicht nur durch das Fachgebiet direkt gekennzeichnet wird, sondern auch in den Wortverbindungen und Satzbeispielen entsprechend zum Ausdruck kommt.

Suche nach Vorzugsbenennung

Hier kann gezielt nach den an der jeweiligen Hochschule verwendeten Übersetzungen gesucht werden (s. Abbildung 4).

Volltextsuche

Einige gesuchte Termini haben keinen eigenen Eintrag, sondern sind als Wortverbindung unter einem allgemeineren

	1110	Referat	presentation
Wortverbindungen		ein ~ über X ein ~ halten ein ~ vorbereiten Thema eines ~s Note eines ~s	a ~ on X to hold/give a ~ to prepare a ~ topic of a ~ grade for a ~
Satzbeispiele		Teilnehmer*innen dieser Veranstaltung müssen ein Referat halten. Das Thema des Referats ist X. Das Referat sollte nicht länger als X Minuten dauern. Die Prüfungsleistung für diesen Kurs besteht aus einem Referat und einer Hausarbeit.	Participants of this course must hold a presentation. The topic of the presentation is X. The presentation should not exceed X minutes. The graded assessment for this course includes a presentation and a term paper.
Fachgebiet		Kurse und Module, Prüfungsangelegenheiten, Studium	
Hinweise		-	
Vorzugsbenennung		-	
Verwandte Einträge		Vortrag	

Abb. 5: Suchergebnis für „Referat“ mit Filter „Prüfungsangelegenheiten“ in der Fachgebietssuche [1] (Quelle: eigener Screenshot)

	1173	Referat	department
Wortverbindungen		~ für X ein ~ einrichten zuständiges ~ Aufgaben des ~s Referatsleiter/Referatsleiterin	~ of X to establish a ~ ~ responsible responsibilities of the ~ head of ~
Satzbeispiele		Das Referat für X unterstützt Sie bei Fragen zu X. Bei Rückfragen können Sie sich gerne an das Referat für X wenden. Für das Thema X hat das Rektorat ein neues Referat eingerichtet.	The Department of X is happy to assist you if you have questions about X. Please do not hesitate to contact the Department of X if you have any questions. The president's office/rectorate has established a new department dedicated to X.
Fachgebiet		Organisationseinheiten	
Hinweise		Im Englischen als Eigenname eines bestimmten Referats großgeschrieben.	
Vorzugsbenennung		-	
Verwandte Einträge		Abteilung, Stabsstelle, Referent/Referentin	

Abb. 6: Suchergebnis für „Referat“ mit Filter „Organisationseinheiten“ in der Fachgebietssuche [1] (Quelle: eigener Screenshot)

	865	Doktorgrad	doctorate
Wortverbindungen		einen ~ besitzen einen ~ anstreben/promovieren einen ~ erlangen jmd. einen ~ verleihen einen ~ entziehen ~ der X [Fachrichtung]	to hold a ~ to pursue a ~ to obtain a ~/to earn a ~ to confer a ~ on sb./to award sb. a ~ to withdraw a ~/to revoke a ~ ~ in X

Abb. 7: Ergebnisse der Volltextsuche werden gelb hinterlegt (Quelle: eigener Screenshot)

Eintrag aufgelistet. Dies trifft häufig auf Komposita zu sowie allgemein auf Verben und Adjektive. Zum Beispiel werden Termini wie „Drittmittelpersonal“ oder „promovieren“ nicht als eigene Einträge, sondern als Wortverbindungen unter „Drittmittel“ bzw. „Promotion“ aufgeführt.

In solchen Fällen kann die Volltextsuche verwendet werden (s. Abbildung 3 auf Seite 8). Es erscheinen dann alle Wörterbucheinträge, in denen die gesuchte Benennung entweder in der Hauptbenennung oder als Wortverbindung enthalten ist (s. Abbildung 7).

Deutscher
Terminologie-Tag e.V.

Deutsches
Institut für Terminologie e.V.

nibelungenstadt
worms

DTT-Symposion 2025

Terminologie in der KI – KI in der Terminologie

27.-29. März 2025, Wormser Tagungszentrum, Worms

Weitere Informationen und Anmeldung auf www.dttev.org

Technische Umsetzung

Die Terminologiedatenbank wird in MultiTerm gepflegt. Das Programm wurde ausgewählt, da es bereits an der Universität Mannheim im Einsatz ist und Datenkategorien frei bestimmt werden können. In der MultiTerm-Datenbank werden für jeden Eintrag die oben genannten Eintragsfelder angelegt und befüllt. In regelmäßigen Abständen wird das Glossar aus MultiTerm exportiert, mithilfe eines eigens geschriebenen Programms konvertiert und in die Online-Wörterbuch-Struktur im Content-Management-System (CMS) Joomla importiert.

Das Wörterbuch im Portal wurde eigens von einem IT-Mitarbeiter der Universität programmiert. Es bringt daher einige technische Besonderheiten mit sich. Werden beispielsweise bestimmte Eintragsfelder ausgelassen, müssen sie mit „-“ gekennzeichnet werden, damit der Import in das CMS reibungslos läuft. Außerdem dürfen Semikolons nicht verwendet werden, da sie im Programm als Trennzeichen festgelegt wurden und beim Import sonst zu Fehlern führen.

Die technischen Besonderheiten beeinträchtigen im Wesentlichen weder die Eintragserstellung noch die Eintragsqualität.

Kritik

Vor allem die Abweichung von etablierten Terminologieeintragsstrukturen kann aus Terminolog*innensicht kritisiert werden. Einige Kritikpunkte sollen hier aufgegriffen und erläutert werden.

Zunächst ist hervorzuheben, dass die meisten Entscheidungen gegen gängige Eintragsstrukturen im Sinne der Zielgruppenorientierung und Benutzerfreundlichkeit und aus einem pragmatischen Ansatz getroffen wurden. Die Landeskoordinationsstelle kann nur einen begrenzten Teil ihrer zeitlichen Ressourcen für die Terminologiearbeit aufwenden, weshalb Termini bewusst ausgewählt und ggf. Informationen kompakt in einem Eintrag gebündelt werden.

Daher weichen z. B. Einträge, bei denen sowohl die britische als auch die amerikanische Schreibweise aufgeführt werden, vom Grundsatz der Benennungsautonomie ab. Die Entscheidung wurde bewusst getroffen, um den Nutzer*innen die verschiedenen Schreibweisen vor Augen zu führen und sie zur reflektierten Verwendung der an ihrer Hochschule gängigen Schreibweise zu bewegen. Nicht-Linguist*innen fällt der Unterschied zwischen den Schreibweisen mitunter nicht auf und sie könnten bei zwei separaten Einträgen Gefahr laufen, die für ihre Hochschule untypische Variante zu verwenden.

Dass nicht alle standardmäßigen Datenfelder wie Definitionen und grammatikalische Angaben aufgeführt werden, ist damit zu begründen, dass die Zielgruppe diese Informationen für gewöhnlich nicht benötigt. In der Regel suchen die Nutzer*innen nur nach englischsprachigen Übersetzungsvorschlägen und konkreten Anwendungsbeispielen für Begriffe, deren Bedeutung sie durch ihre Arbeit kennen.

Die Darstellung der Einträge ist speziell und zielt vor allem auf Kompaktheit ab. Dadurch entstehen in Einzelfällen unklare Strukturen, wie z. B. bei Kollokationen, welche die Verwendung der Tilde als Platzhalter und des Schrägstrichs zur Synonymaufführung kombinieren. Als Beispiel sei hier die unter dem Eintrag „Doktorgrad“ aufgeführte Kollokation „einen ~ anstreben/promovieren“ genannt (s. Abbildung 7 auf Seite 9). Es ist auf den ersten Blick nicht eindeutig, dass „promovieren“ separat von der vorausgehenden Phrase als Synonym zu verstehen ist. Geschuldet ist dies auch der Tatsache, dass aus technischen Gründen keine Leerzeichen zwischen den Schrägstrichen stehen dürfen, da es ansonsten beim Import der Datenbank zu Fehlern käme. Hier ist also zu berücksichtigen, dass es sich beim Interface des Wörterbuchs um eine eigene Programmierung handelt. Eine derartige, ggf. missverständliche Verwendung von Sonderzeichen wird beim anstehenden Umzug der Datenbank (s. Abschnitt Weiterentwicklung) bedacht und nach Möglichkeit verbessert.

Die Datenfelder der Wörterbucheinträge enthalten entgegen der allgemeinen Praxis keine Quellenangaben. Dies kann kritisiert werden, da Nutzer*innen nicht nachvollziehen können, woher Kollokationen, Beispiele etc. stammen und ob die Quellen verlässlich sind. Die Entscheidung für dieses Vorgehen fiel wieder zugunsten eines einfachen und übersichtlichen Aufbaus des Wörterbuchs. Die Landeskoordinationsstelle dient in diesem Fall als Prüfinstanz der benutzten Inhalte, die, wie in den Unterabschnitten „Wortverbindungen“ und „Satzbeispiele“ beschrieben, aus qualitätsgesicherten Wörterbüchern, Enzyklopädien, Gesetzestexten und eigenen Übersetzungen stammen.

Weder die Nutzeroberfläche noch die Hinweise in den Einträgen sind zweisprachig, was einen Bruch mit den Grundsätzen der Datenmodellierung darstellt. In der Praxis würde dies nicht-deutschsprachigen Hochschulmitarbeiter*innen die Nutzung der Ressource erschweren. Jedoch besteht die primäre Zielgruppe des Wörterbuchs, wie eingangs beschrieben, aus deutschsprachigen Hochschulmitarbeiter*innen, die im Sinne der Internationalisierung Texte auf Englisch bereitstellen wollen.

Die Verwendung der Bezeichnung „Vorzugsbenennung“ für die Datenkategorie, unter der im Fall des Baden-Württerbuchs die Hochschulen angeführt werden, an denen die englische Übersetzungslösung verwendet wird, kann missverständlich sein. Nach der Terminologielehre erscheinen unter dieser Datenkategorie in der Regel die von der Organisation empfohlenen Bezeichnungen. Eine Umbenennung der Datenkategorie wäre daher angebracht.

Weiterentwicklung

Bisher nicht vollständig erfasst sind die Einträge zu Modulen und Kursnamen, spezifischen Fakultäten, Instituten und Abteilungen sowie hochschulspezifischen Bezeichnungen für Studiengänge. Geplant ist, diese Termini langfristig in das Wörterbuch aufzunehmen, da solche Bezeichnungen häufig verwendet werden, beispielsweise in Transcripts of Records. Aufgrund ihrer spezifischen Art hatten diese Termini bislang im Aufbau und der Erweiterung des Wörterbuchs keine Priorität. Des Weiteren werden ständig neue Termini auf Basis häufiger Suchanfragen in das Wörterbuch aufgenommen. Im Backend des CMS lässt sich einsehen, welche Wörter die Nutzer*innen in die Suchmaske eingeben. Die Anfragen werden analysiert, den Kriterien entsprechende Termini ausgewählt und in die Datenbank eingetragen. Mit Blick auf die Vorzugsbenennungen sind zwei weitere Hochschulen aktuell im Prozess, Glossare zu erstellen und ihre Vorzugsbenennungen ins Baden-Württerbuch überführen zu lassen.

Perspektivisch soll das Portal in ein anderes CMS überführt werden, was eine große technische Herausforderung für das Baden-Württerbuch darstellt. Durch den eigens programmierten Code wird es sich nicht eins zu eins in einer neuen Umgebung darstellen lassen. Die technische Umsetzung muss noch erarbeitet werden. Mögliche Lösungen sind die Umstrukturierung und Anpassung des Wörterbuchs an das neue CMS, die Einbindung von MultiTerm oder die Erhaltung der aktuellen Instanz allein für die Wörterbuchfunktion. Die erste und gleichzeitig wahrscheinlichste Lösung wird unumgänglich Neuerungen mit sich bringen. Daher bringt der Umzug auch eine Chance für Verbesserungen mit sich. Denkbar sind hier z. B. die Optimierung der Suchfunktion, eine verbesserte Darstellung der Einträge oder die unter dem Abschnitt „Kritik“ aufgeführten Aspekte.

Um eine adäquate Nutzung des Wörterbuchs durch die Zielgruppe zu gewährleisten, werden von der Koordinationsstelle in regelmäßigen Abständen Schulungen angeboten, um die Hochschulmitarbeiter*innen für die verschiedenen Suchoptionen, aber auch für die eingangs angesprochenen Besonderheiten der Verwaltungssprache an deutschen Hochschulen sowie die Systemunterschiede zwischen den Ländern zu sensibilisieren.

Literatur

[1] Online-Portal der Landeskoordinationsstelle für Übersetzungsangelegenheiten im Hochschulwesen Baden-Württemberg an der Universität Mannheim: https://hochschuluebersetzung-bw.uni-mannheim.de/index.php (letzter Zugriff 09.10.2024)

[2] Ballansat-Aebi, Suzanne (2014): Grundlegende Faktoren bei der juristischen Übersetzung. http://www.ballansat-translation.ch/wp-content/uploads/2013/09/Grundlegende-Faktoren-bei-der-juristischen-Uebersetzung1.pdf (letzter Zugriff 14.10.2024)

[3] Arntz, Reiner; Picht, Heribert; Schmitz, Klaus-Dirk (2014): Einführung in die Terminologiearbeit. 7., vollständig überarbeitete und aktualisierte Auflage. Hildesheim, Georg Olms Verlag.

[4] Drewer, Petra; Schmitz, Klaus-Dirk (2017): Terminologiemanagement. Grundlagen – Methoden – Werkzeuge. Berlin, Springer Vieweg (Kommunikation und Medienmanagement, 1).

[5] Terminologie-Datenbank Universität Zürich, https://www.uniterm.uzh.ch/search.php (letzter Zugriff 10.10.2024)

[6] Wörterbuch der Universität Freiburg, https://uni-freiburg.de/transfer/wissenstransfer-und-wissenschaftskommunikation/newsroom/service-fuer-universitaetsmitglieder/#woerterbuch-der-universitaet (intern; letzter Zugriff 10.10.2024)

[7] Online-Glossar der Johannes Gutenberg-Universität Mainz, https://uebersetzen.uni-mainz.de/online-glossar (intern; letzter Zugriff 13.10.2024)

[8] LookUp, Datenbank der Friedrich-Alexander-Universität Nürnberg-Erlangen, https://lookup.sz.uni-erlangen.de/Logon.aspx (intern; letzter Zugriff 10.10.2024)

[9] IATE, https://iate.europa.eu/home (letzter Zugriff 10.10.2024)

[10] Siepmann, Dirk (2015): Wörterbuch Hochschule. Forschung, Lehre und Management. Deutsch-Englisch/Englisch-Deutsch. Bonn, Deutscher Hochschulverband.

[11] Dietl, Clara Erika; Lorenz, Egon (2020): Fachwörterbuch Recht, Wirtschaft und Politik = Dictionary of Law, Commerce and Politics. Band 2. Englisch – Deutsch. 6. Auflage. München, C.H. Beck; Basel, Helbing & Lichtenhahn.

[12] Behmel, Albrecht; Neudorfer, Kelly (2016): The Foreigner's Guide to German Universities. Origin, Meaning, and Use of Terms and Expressions in Everyday University Life. Hannover, ibidem-Verlag.

[13] Online Oxford Collocation Dictionary of English: https://m.freecollocation.com/ (letzter Zugriff 16.07.2024)

[14] Kollokationenwörterbuch, https://kollokationenwoerterbuch.ch/web/projekt/ (letzter Zugriff 16.07.2024)

[15] Bundesministerium der Justiz, https://www.gesetze-im-internet.de/ (letzter Zugriff 10.10.2024)

[16] EurLex, https://eur-lex.europa.eu/homepage.html?lang=en&locale=de (letzter Zugriff 10.10.2024)

[17] Bundesministerium des Innern und für Heimat, https://www.bmi.bund.de/DE/themen/oeffentlicher-dienst/oeffentlicher-dienst-node.html (letzter Zugriff 10.10.2024)

Vanessa Knies ist Landeskoordinatorin für Übersetzungsangelegenheiten im Hochschulwesen Baden-Württemberg an der Universität Mannheim.

Kontaktadresse
vanessa.knies@uni-mannheim.de
www.uni-mannheim.de

Ellise Meyer ist ehemalige Landeskoordinatorin für Übersetzungsangelegenheiten im Hochschulwesen Baden-Württemberg an der Universität Mannheim.

Kontaktadresse
ellise.meyer@gmail.com

Das Beste rausholen – Termextraktion mit KI und Verstand

Vanessa Jochum und Jasmin Nesbigall

Large language models (LLMs) are good at summarising or extracting information from given texts. It therefore seems reasonable to use them for term extraction as well. However, it is not yet clear whether they can outperform humans and specialised software. Comparison tests between LLMs, extraction software and terminologists show the advantages and disadvantages of each competitor and potential use cases.

Keywords: terminology work, term extraction, large language models, extraction software

Terminologiearbeit benötigt Akribie, Expertise und Liebe zum Detail. Viele Schritte der Terminologiearbeit können daher sehr zeitintensiv sein. Auch die Termextraktion als erster Schritt zum Aufbau eines Terminologiebestands benötigt je nach Datenbasis viel Zeit und ist ab einem gewissen Umfang an Ausgangsdokumenten manuell kaum noch effizient durchzuführen. Überall dort, wo aufwendige manuelle Prozesse unterstützt und automatisiert werden sollen, erfolgt aktuell der Ruf nach KI.

Gerade die Termextraktion scheint auf den ersten Blick durchaus eine geeignete Aufgabe für Large Language Models zu sein, denn eine klare Stärke dieser Modelle besteht darin, gezielt Informationen aus Texten zu ziehen. Erste Ergebnisse aus der Terminologie-Community schienen vielversprechend, erfolgten aber meist ohne Vergleich zu einem Referenzergebnis. So entstand die Idee, drei Kontrahenten gegeneinander antreten zu lassen: KI und klassische Extraktionssoftware sowie den Menschen als Referenz, um den Output der KI quantitativ und qualitativ einschätzen zu können.

Eine simple Bewertung des Outputs als „gut“ und „schlecht“ kann sicherlich schnell erfolgen, sobald ein Text eingegeben und der Chatbot gebeten wird, Fachtermini daraus zu extrahieren. Erhalten wir eine Liste mit Termini aus dem Sachgebiet des entsprechenden Texts, werden wir das Ergebnis als gut einstufen. Erhalten wir hingegen eine Liste allgemeinsprachlicher Wörter oder einen Fließtext mit einer Zusammenfassung des Textes, wurde die Aufgabenstellung nicht erfüllt. Die tatsächliche Qualität des Outputs können wir aber erst im Vergleich zu einer Referenz feststellen, da nur diese Aufschluss über das zu erwartende Ergebnis liefert, wodurch Fehler im generierten Output erst identifiziert werden können. Erhalten wir vom LLM eine Liste mit Fachtermini, von denen einige gar nicht im Text vorkamen, kann das Ergebnis nicht als verlässlich betrachtet werden. Wurden im Vergleich bei einer manuellen Extraktion deutlich mehr Termini extrahiert, ist auch die Quantität des Ergebnisses infrage zu stellen.

Ring frei also für umfangreiche Tests zur manuellen Extraktion, dem Einsatz von Extraktionssoftware und der Anwendung künstlicher Intelligenz.

Material und Versuchsaufbau

Im ersten Schritt wurde der Versuchsaufbau konzipiert und passendes Ausgangsmaterial gesucht. Dabei wurden zwei Texte zum Thema Schwammstadt und klimaangepasster Städtebau ausgewählt, in denen der Anteil an Fachterminologie aus den Sachgebieten Umwelt, Stadtentwicklung und Klimaanpassung nach erster Sichtung hoch war. Um einen möglichen Einfluss der Textlänge auf das Ergebnis untersuchen zu können, wurden Texte ausgewählt, die sich hinsichtlich ihrer Länge deutlich unterschieden: Text 1 umfasste 1.658 Wörter, Text 2 umfasste 9.803 Wörter. Text 2 war also mehr als fünfmal so lang wie Text 1. Noch umfangreichere Texte wurden aufgrund der Anforderung, den Text auch manuell zu bearbeiten, ausgeschlossen.

Beide Texte wurden als Word-Dateien verarbeitet, da alle drei zu untersuchenden Methoden dieses Datenformat unterstützen. Für die softwaregestützte Extraktion wurde MultiTerm Extract verwendet und für die KI-gestützte Extraktion wurden unterschiedliche Modelle von ChatGPT (s.u.) eingesetzt. Die Wahl der Tools erfolgte nach dem Kriterium der Gängigkeit bzw. Verbreitung, um einen Versuchsaufbau zu bieten, der reale Arbeitsbedingungen abbildet und bei

Bedarf von Interessierten nachgestellt werden kann. Da das Ergebnis der manuellen Extraktion als Referenzergebnis dienen sollte, wurde diese als erste durchgeführt.

Manuelle Extraktion

Termextraktion kann sicherlich zum kleinen Einmaleins der Terminologiearbeit gezählt werden. Das Auffinden relevanter Termini aus Fachtexten und Unternehmensmaterial ist für viele Terminolog:innen, aber auch für Redakteur:innen und Übersetzer:innen Bestandteil des Arbeitsalltags. Mit zunehmendem Textumfang nimmt die Effizienz der manuellen Extraktion allerdings deutlich ab: Parallel zum zeitlichen Aufwand steigt auch die Fehleranfälligkeit, z. B. die Gefahr, dass relevante Termini übersehen werden [1]. Hinzu kommt, dass objektive Bewertungsgrundlagen für die Entscheidung darüber, ob es sich bei einem Termkandidaten um einen Fachterminus handelt oder nicht, fehlen und jede Termextraktion daher nach subjektivem Ermessen der Beteiligten erfolgt [1]. In der Praxis kann das Ergebnis einer Termextraktion also je nach Prozessbeteiligten und häufig auch bei nochmaliger Prüfung des Ergebnisses durch die gleiche Person unterschiedlich ausfallen. Dennoch kann die manuelle Termextraktion als umfangreiche und qualitativ hochwertige Extraktionsmethode angesehen werden, da mit wenigen False Negatives und False Positives zu rechnen ist [1]. Sie kann daher als „Goldstandard“ dienen, um einen Referenzwert für andere Methoden – in unserem Fall für den Vergleich der mithilfe von Extraktionssoftware und LLMs gewonnenen Ergebnisse – zu erhalten.

Um die angesprochene Subjektivität zu reduzieren, wurde die manuelle Extraktion bei beiden Texten von jeweils zwei Terminolog:innen durchgeführt. Dies hatte den weiteren Vorteil, dass auch der zeitliche Aufwand verglichen und gemittelt werden konnte. Die Termkandidaten wurden in Listenform erfasst und direkt bei der Eingabe in die Liste durch die Funktion „Doppelte Werte“ auf Dubletten geprüft. Gerade mit zunehmender Textlänge stellt die doppelte oder mehrfache Erfassung von Termini eine übliche Fehlerquelle bei der manuellen Termextraktion dar.

Die Ergebnisse der unabhängig erfolgten Extraktionen wurden anschließend miteinander verglichen. Bei

Anzeige

DTT Deutscher Terminologie-Tag e.V.
DIT Deutsches Institut für Terminologie e.V.

Terminologiearbeit Best Practices 2.0

Das bewährte Praxishandbuch für Terminologiearbeit und Terminologiemanagement mit dem Know-how zahlreicher Experten aus Industrie und Wissenschaft ist auch **in englischer Sprache** erhältlich.

kompakt | praxisnah | auf den Punkt

2. überarbeitete und ergänzte Auflage
Stand: März 2014

Koordination und Redaktion:
Petra Drewer
Donatella Pulitano
Klaus-Dirk Schmitz

Geballtes Terminologiewissen in acht Modulen für Unternehmen und Freiberufler

1. Argumentationshilfen
2. Grundsätze und Methoden
3. Benennungen
4. Werkzeuge und Technologien
5. Projekt- und Prozessmanagement
6. Berufsprofile, Anforderungen, Ausbildungsinhalte
7. Urheberrecht an Terminologie
8. Wirtschaftlichkeit

Erhältlich in Deutsch oder Englisch auf www.dttev.org

Differenzen zwischen den beiden Listen erfolgte die Entscheidung für oder gegen einen Termkandidaten aufgrund fehlender objektiver Verfahren auf Basis der subjektiven Einschätzung der beiden Terminolog:innen.

Ergebnisse

Die Subjektivität zeigte sich deutlich in den Ergebnissen der Terminolog:innen: Beim kürzeren Text 1 unterschieden sich die Listen um 20 Termini, beim längeren Text 2 um knapp 60 Termini. Als End- und Referenzergebnis für die weiteren Extraktionsmethoden wurden aus Text 1 nach dem oben genannten Abgleich der Ergebnisse insgesamt 113 Termini validiert, aus Text 2 insgesamt 299 Termini. Auch wenn Text 2 mehr als fünfmal so lang war wie Text 1, betrug die Anzahl der validierten Termini nur circa das Zweieinhalbfache. Bei einer insgesamt hohen Terminusdichte der Texte decken sich diese Zahlen mit der Praxiserfahrung, dass Terminologie in kurzen Fachtexten meist komprimiert vorkommt, während sie sich in längeren Texten häufiger wiederholt.

Die Ergebnisse der manuellen Extraktion dienten sowohl als quantitative als auch als qualitative Referenzwerte für die beiden weiteren untersuchten Methoden: Auf Basis der Anzahl extrahierter Termini ließen sich quantitative Unterschiede verdeutlichen. Qualitativ wurden wiederum bei den weiteren Methoden nur diejenigen Termini final validiert, die auch im Referenzergebnis der manuellen Extraktion vorkamen. Auf diese Weise wurde die Vergleichbarkeit der Ergebnisse so objektiv wie möglich gehalten.

Softwaregestützte Extraktion

Als erster Herausforderer ging die Extraktionssoftware MultiTerm Extract an den Start. Extraktionstools bieten üblicherweise Einstellmöglichkeiten zur Silence-Noise-Ratio (Datenrauschen), zur minimalen und maximalen Terminuslänge und zur maximalen Anzahl extrahierter Termini. Außerdem können systemseitige oder benutzerspezifische Stoppwortlisten hinterlegt werden, um z. B. allgemeinsprachliche Termini oder in vorherigen Extraktionen geprüfte (und nicht validierte) Termkandidaten von der Extraktion auszuschließen. Für die Tests wurden für beide Texte insgesamt drei Durchläufe mit unterschiedlichen Einstellungen durchgeführt. Dabei wurde vor allem der Parameter Datenrauschen, aber auch die maximale Terminuslänge verändert (siehe Tabelle 1). Eine Verschiebung der Silence-Noise-Ratio zugunsten von Noise führt üblicherweise zu mehr Termkandidaten und damit zu einem umfangreicheren Rohergebnis für die Validierung, wie ebenfalls aus Tabelle 1 hervorgeht. Als Stoppwortliste wurde die vom System standardmäßig hinterlegte Liste für die Ausgangssprache Deutsch genutzt. Die maximale Anzahl extrahierter Termini wurde bei keiner Einstellung beschränkt, da dies für den Versuchsaufbau nicht zielführend gewesen wäre. Jede Einstellung wurde jeweils für Text 1 und Text 2 angewandt.

Nach der automatisierten Termextraktion durch MultiTerm Extract erfolgte eine manuelle Validierung der Termkandidaten, um tatsächlich relevante Fachtermini zu identifizieren. Die Validierung wurde von einer der beiden Personen vorgenommen, die auch die manuelle Extraktion durchgeführt hatten. Anschließend wurde das Ergebnis mit der Referenzliste aus der manuellen Extraktion abgeglichen und es wurde bewertet, wie viele der Termini aus der manuellen Extraktion im Ergebnis der softwaregestützten Extraktion vorhanden waren.

Ergebnisse

Wie aus Tabelle 1 ersichtlich, hatten die unterschiedlichen Einstellungen großen Einfluss auf das Rohergebnis der Extraktion. Im kürzeren Text 1 kamen Termini meist nur einmal vor. Dies führte mit den Standardeinstellungen (Durchlauf 1) zu unzureichenden Ergebnissen. Das Rohergebnis lag bei nur 67 Termkandidaten, die Anzahl validierter Termini, die dem Referenzergebnis entsprachen, lag bei 11 Termini und damit bei 9,73 % des Referenzergebnisses. Beim längeren Text 2 führte die Standardeinstellung zu einem Rohergebnis von 538 Termkandidaten und einer Anzahl validierter Termini von 78 bzw. 26,09 % des Referenzergebnisses.

Text	Durchlauf	Silence-Noise-Ratio (1= max. Noise, 11 = max. Silence)	Maximale Terminuslänge	Rohergebnis	Validierte Termini
1	1	6 von 11	3	67	11
1	2	4,5 von 11	3	151	23
1	3	1 von 11	2	612	98
2	1	6 von 11	3	538	78
2	2	4,5 von 11	3	1046	99
2	3	1 von 11	2	2526	259

Tab. 1: Einstellungen und Ergebnisse pro Durchlauf für Texte 1 und 2 in MultiTerm Extract

Als zielführendste Einstellung stellte sich Durchlauf 3 mit einer maximalen Noise-Einstellung und einer herabgesetzten Terminuslänge von 2 heraus. Hierbei wurden bei Text 1 insgesamt 98 Termini und damit 86,73 % des Referenzergebnisses validiert. Das Rohergebnis betrug hier allerdings 612 Termkandidaten, sodass die Validierung entsprechend zeitintensiv war. Auch im zweiten Text führte Durchlauf 3 zu den besten Ergebnissen mit 259 validierten Termini, also 86,62 % des Referenzergebnisses. Das Rohergebnis war aufgrund des größeren Textumfangs nochmals deutlich höher und betrug 2526 Termkandidaten, die geprüft und validiert werden mussten. Zusätzlich zum Umfang des Rohergebnisses mussten die Termini teilweise auch sprachlich bereinigt werden. Beispielsweise wurden Plural- in Singularformen umgewandelt.

Interessant war bei der manuellen Validierung des Rohergebnisses für Text 2 die Beobachtung, dass bei der softwaregestützten Extraktion zusätzliche Termini erfasst worden waren, die nicht im Referenzergebnis enthalten waren. Insgesamt handelte es sich um 30 weitere Termini, also zusätzliche 10 % des Referenzergebnisses. Bei der Prüfung, warum diese bei der manuellen Extraktion „übersehen“ wurden, fielen zwei Aspekte auf: die Darstellungsform in der Software und die Position der Termini im Text. Da die Extraktionssoftware das Rohergebnis in Listenform liefert, werden Termini konzentrierter und isolierter als im Fließtext dargestellt. Dadurch fallen auch Termini auf, die im Dokument eventuell überlesen werden, weil sie in Bildunterschriften oder Fußnoten vorkommen. Die Listenform birgt gleichzeitig jedoch das Risiko, dass Termini erfasst und validiert werden, die nur Teil eines Organisationsnamens sind oder im Literaturverzeichnis vorkommen.

KI-gestützte Extraktion

Da auch im Terminologiebereich die Neugier auf die Einsatzmöglichkeiten für KI-Systeme wächst, wurde als letzte Extraktionsmethode ChatGPT in den Ring geschickt. Dabei wurden drei Versionen des Large Language Models verwendet: GPT-3.5, GPT-4 Turbo und die im Frühjahr 2024 erschienene Version GPT-4o. Bei der Nutzung generativer KI kommt dem Prompting eine besondere Bedeutung zu [2],[3]. Als Prompting wird die Formulierung einer Arbeitsanweisung an das Large Language Model bezeichnet, damit dieses eine bestimmte Aufgabe erfüllen und den gewünschten Output generieren kann. Es haben sich verschiedene Prompting-Strategien entwickelt, die sich u. a. im Translationskontext bewährt haben [4]. Auch in unseren Tests wurden verschiedene Strategien angewandt, die sich teilweise zusätzlich hinsichtlich der Detailliertheit der Arbeitsanweisung und der verlangten Zusatzaufgaben unterschieden.

Die Wahl fiel auf Domain-Specific-Prompting, Task-Specific-Prompting und Reverse-Prompting sowie naives Prompting als Ausgangspunkt. Der erste Durchlauf erfolgte mit GPT-3.5 und insgesamt zehn unterschiedlichen Prompts (siehe Tabelle 2 auf Seite 17). Die Prompts mit den besten Ergebnissen wurden zudem mehrmals, teilweise unmittelbar hintereinander, teilweise in einem gewissen zeitlichen Abstand zueinander, verwendet. Sie wurden außerdem anschließend in GPT-4 Turbo und GPT-4o getestet.

Startpunkt war Text 1 mit einem naiven Prompt (siehe Tabelle 2, P1) als Ausgangspunkt, um mit jedem weiteren Prompt einordnen zu können, ob sich Prompting-Strategien wie erwartet positiv auf das Ergebnis auswirken [3],[4]. Es folgte Domain-Specific-Prompting, bei dem Informationen zum Sachgebiet des eingegebenen Textes oder des erwarteten Outputs im Prompt mitgegeben werden. Da die Eingrenzung von Sachgebieten als grundlegend für die Termextraktion angesehen wird, wurden für diese Strategie mehrere Prompts formuliert, die entweder das Thema des Texts und das Sachgebiet der zu extrahierenden Termini enthielten (P2) oder nur letzteres (P3), um keine Interferenz zwischen Sachgebieten zu riskieren. Bei P3 fiel die Wahl auf das breitere Sachgebiet „Umwelt“ im Vergleich zum enger gefassten Thema „Schwammstadt“, sodass der Prompt als allgemeiner als P2 gelten kann. P3 wurde dann mit kleineren Umformulierungen nochmals als P4 getestet. Anschließend wurde ein Domain-Specific-Prompt um weitere Anweisungen zur Aufgabe ergänzt (P5 und P6), um den Einfluss detaillierterer Arbeitsanweisungen auf die Menge und Qualität des Outputs untersuchen zu können.

Als zweite Strategie wurde Task-Specific-Prompting getestet, bei welchem dem LLM in der Anweisung eine bestimmte Rolle zugewiesen wird, hier die eines Extraktionstools (P7). Auch dadurch erhält das Modell mehr Informationen zur gestellten Aufgabe, was besonders bei komplizierten Aufgabenstellungen die Qualität des Outputs steigern soll [4]. In einem weiteren Versuch (P8) wurde spezifiziert, welches Verfahren bei der Extraktion genutzt werden soll. In beiden Fällen wurde die Nennung des Sachgebiets beibehalten, da diese als wesentlich für die Aufgabenstellung betrachtet wurde.

Eine weitere getestete Strategie war Reverse-Prompting. Hierbei wird entweder ein gewünschtes Ergebnis oder eine klare Zielsetzung als Prompt eingegeben und das LLM aufgefordert, selbst einen Prompt zu formulieren, der zu diesem Ergebnis führt. Im ersten Versuch (P9.0) gab ChatGPT einen naiven und zu allgemeinen Prompt aus, der das Sachgebiet nicht explizit angab. Der Prompt wurde daher als nicht zielführend bewertet und das Reverse-Prompting nochmals überarbeitet. Der neue Prompt enthielt zusätzlich

die Frage, welche Informationen das LLM benötigt, um das gewünschte Ergebnis zu erreichen. Das Ergebnis dieses Reverse-Prompting war weiterhin überraschend simpel: ChatGPT schlug vor, nur den Text, aus dem Termini extrahiert werden soll und das Sachgebiet zu nennen (P9). Damit deckte sich die Empfehlung mit dem bereits getesteten Domain-Specific-Prompting (P3) und brachte im Vergleich zu den bis dahin verwendeten Prompts weder eine qualitative noch eine quantitative Verbesserung. Aus allen neun bis hierhin getesteten Prompts hatte der naive Prompt (P1), also der Ausgangspunkt, das beste Ergebnis geliefert.

Da die Ergebnisse quantitativ hinter den Ergebnissen der manuellen und der softwaregestützten Extraktion zurückblieben, wurde dem Modell schließlich ein Orientierungswert aus dem Referenzergebnis mitgegeben. Dazu wurde ein Domain-Specific-Prompt um das Ergebnis (als Zahl) der manuellen Extraktion ergänzt (P10). Obwohl das Ergebnis auch bei diesem Prompt nicht an den Referenzwert herankam, gehörte es trotzdem zu den besten der Testreihe. Abbildung 1 auf Seite 19 zeigt die Ergebnisse der zehn getesteten Prompts im Vergleich zueinander.

Die beiden besten Prompts (1 und 10) wurden anschließend für den längeren Text 2 mehrfach und mit den drei verschiedenen Modellen von ChatGPT getestet. Dabei zeigte sich der Output als sehr instabil, obwohl Vorgehensweise, Prompt und Ausgangsdatei dieselben waren. Abbildung 2

Nummer	Strategie	Prompt
P1	Naiv	Extrahiere aus dem folgenden Text alle Fachtermini. Liste die Fachtermini einzeln untereinander auf.
P2	Domain-Specific	Extrahiere aus dem folgenden Text über Schwammstädte alle Fachtermini aus dem Sachgebiet „Umwelt". Liste alle gefundenen Fachtermini einzeln untereinander auf.
P3	Domain-Specific	Extrahiere alle Fachtermini aus dem gegebenen Text zu dem Sachgebiet „Umwelt" und liste sie einzeln auf.
P4	Domain-Specific	Extrahiere aus dem folgenden Text alle Fachtermini aus dem Sachgebiet „Umwelt". Liste die Fachtermini einzeln untereinander auf.
P5	Domain-Specific plus Details zur Aufgabe	Extrahiere aus dem folgenden Text alle Termini aus dem Bereich „nachhaltiges Starkregenmanagement" und liste sie einzeln untereinander auf. Extrahiere auch die vorhandenen Mehrwortbenennungen.
P6	Domain-Specific plus Details zur Aufgabe	Extrahiere aus dem folgenden Text alle Termini aus dem Bereich „nachhaltiges Starkregenmanagement" und liste sie einzeln untereinander auf. Extrahiere auch die vorhandenen Mehrwortbenennungen. Reduziere die Termini auf ihre Grundform.
P7	Task-Specific + Domain-Specific	Du bist ein Extraktionstool und wirst für die Terminologieextraktion eingesetzt. Extrahiere alle Fachtermini des Sachgebiets „Umwelt" aus dem folgenden Text.
P8	Task-Specific + Domain-Specific	Extrahiere aus dem folgenden Text alle Fachtermini aus dem Sachgebiet „Umwelt". Verwende für die Extraktion das linguistische Verfahren. Liste die Fachtermini einzeln untereinander auf. Der Text lautet wie folgt:
P9.0	Reverse-Prompting	Extrahiere alle Fachtermini aus dem gegebenen Text zu einem bestimmten Sachgebiet und liste sie einzeln auf.
P9	Domain-Specific nach Reverse-Prompting	Extrahiere alle Fachtermini zum Thema „Umwelt" aus dem gegebenen Text.
P10	Domain-Specific plus Referenzwert	Extrahiere aus dem folgenden Text alle Fachtermini aus dem Sachgebiet „Umwelt". In Texten mit vergleichbarer Länge wurden knapp 115 Termini gefunden. Liste die Fachtermini einzeln untereinander auf. Der Text lautet wie folgt:

Tab. 2: Prompting-Strategien und ausformulierte Prompts

Jetzt Leseprobe herunterladen und online bestellen!

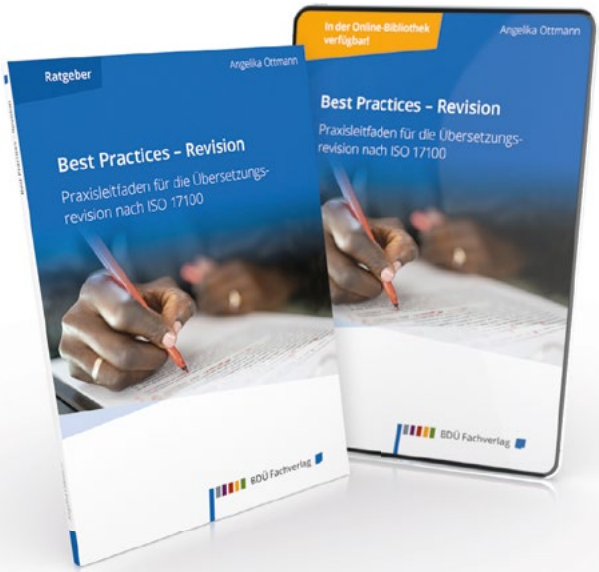
Bild: © Andrey Popov / stock.adobe.com

ANGELIKA OTTMANN

Best Practices – Revision

Praxisleitfaden für die Übersetzungsrevision nach ISO 17100

Die Revision ist ein wichtiger Bestandteil der Qualitätssicherung im Übersetzungsprozess und ist ein obligatorischer Schritt gemäß ISO 17100. In der Praxis herrscht jedoch viel Unsicherheit darüber, was genau Revision ist, welche Maßnahmen sie beinhaltet, wie sie durchzuführen ist, welche Kenntnisse man dafür benötigt und auch darüber, wann sie notwendig ist. Dieser Leitfaden ist als Nachschlagewerk und Anleitung für die Aufgabe Revision gedacht und richtet sich nicht nur an Übersetzerinnen und Übersetzer sowie Revisorinnen und Revisoren, sondern auch an Auftraggeberinnen und Auftraggeber. **Behandelt werden, neben Definition und Abgrenzung der Begriffe, die Grundsätze der Revision, die Erstellung von Revisionsaufträgen, die Techniken und Entscheidungsfaktoren bei der Revision, die Gestaltung des Revisionsprozesses, die Bewertung von Revisionsqualität, die für die Revision erforderlichen Kompetenzen sowie die Zeitplanung und Preisgestaltung für die Revision.** Der Leitfaden soll einen Beitrag für die Entwicklung von Branchenstandards für den Übersetzungsprozessschritt Revision leisten.



Bestellbar über:

→ **www.bdue-fachverlag.de**/detail_book/170

Umfang: 94 Seiten, ISBN: 978-3-946702-27-6,
Erscheinungsjahr: 2024, Preis: 44,00 € inkl. MwSt.

Weiterbildungs- und
Fachverlagsgesellschaft
Fachverlag

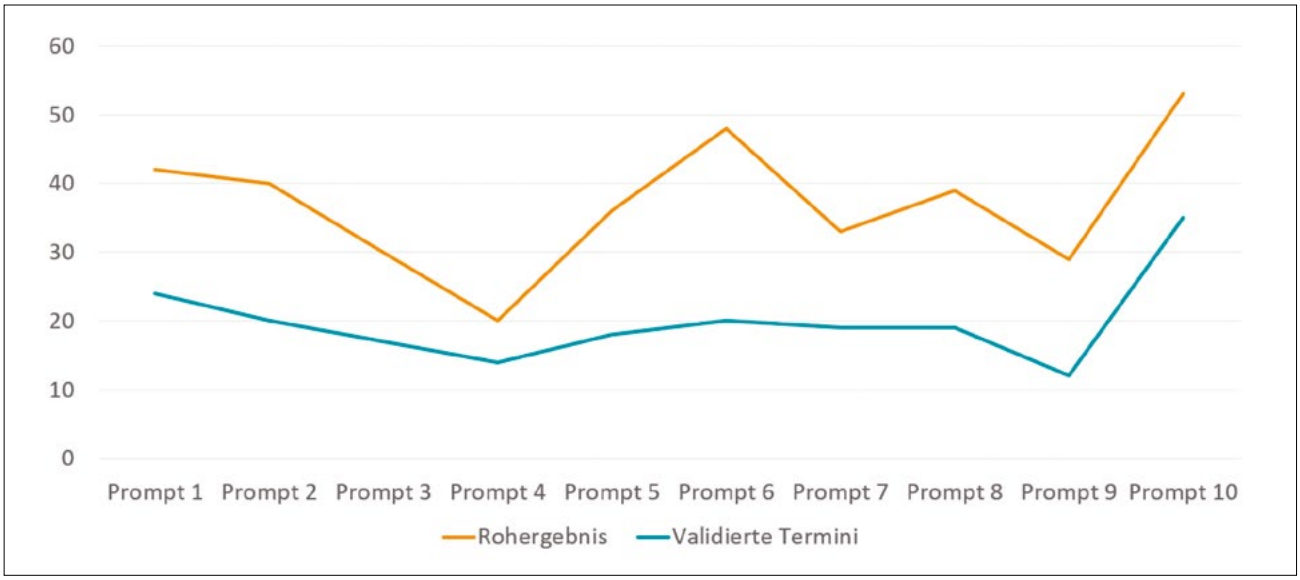


Abb. 1: Rohergebnisse und Anzahl validierter Termini für die zehn verwendeten Prompts im jeweils ersten Durchlauf, getestet an Text 1 (Quelle: eigene Darstellung)

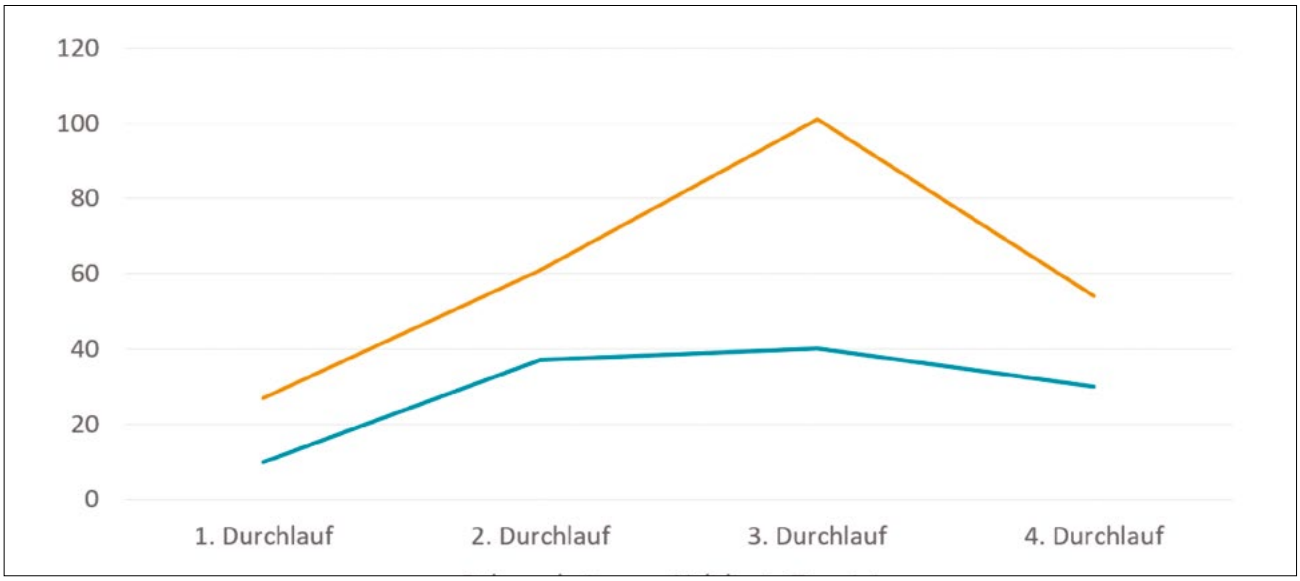


Abb. 2: Rohergebnisse und Anzahl validierter Termini aus vier Durchläufen für den identischen Prompt P1, getestet an Text 2 (Quelle: eigene Darstellung)

zeigt die unterschiedlichen Ergebnisse für Prompt P1 in vier Durchläufen.

Ergebnisse

Die insgesamt instabilen Ergebnisse – insbesondere bei identischen Prompts – deuten darauf hin, dass das Prompting, anders als anfangs angenommen, wenig Einfluss auf die Ergebnisse hatte. Weder das qualitative noch das quantitative Ergebnis ließen sich mit den angewandten Strategien zuverlässig und stabil steigern.

Das Rohergebnis aus ChatGPT betrug je nach Prompt zwischen 20 und 53 Termkandidaten für den kürzeren Text 1, von denen anschließend zwischen 12 und 35 Termini validiert wurden. Damit konnte selbst beim besten Ergebnis nur ein Anteil von 30,97 % des Referenzergebnisses erreicht werden. Beim längeren Text 2, bei dem nur noch die besten Prompts aus den ersten Tests verwendet wurden, lag das Rohergebnis zwischen 25 und 157 Termkandidaten, von denen nur zwischen 5 und 75 Termini dem Referenzergebnis entsprachen. Die Abdeckung im Vergleich zum manuellen Ergebnis lag also im besten Versuch bei nur 25,08 %. Auffällig war hierbei, dass der identische Prompt (P1) in zwei aufeinanderfolgenden Durchläufen einmal zu 40, einmal zu nur 10 validierten Termini führte.

Auch bei den Tests mit ChatGPT wurden zusätzliche Termini über das Referenzergebnis hinaus extrahiert – hierbei handelte es sich allerdings um halluzinierte Termini. Bei

mehreren Prompts enthielt das Ergebnis ein oder zwei Termini, die thematisch zwar passten, aber in dieser Form nicht im Text vorkamen. Sie konnten allerdings aus der vorhandenen Terminologie abgeleitet werden, wie zum Beispiel „Wassersensibilität“ von dem im Text vorkommenden Terminus „wassersensible Stadtentwicklung“ oder „Klimaresilienz“ von „klimaresilient“. „Wassersensibilität“ wurde sogar bei verschiedenen Prompts halluziniert. Auffallend war auch, dass vorhandene Komposita von ChatGPT aufgeteilt oder neu zusammengesetzt und dann als Termini ausgegeben wurden. So wurde etwa „Wärmehaushalt“ als Terminus vorgeschlagen, obwohl nur „Wärmehaushaltsfunktion“ im Text vorkam. Aus den im Text direkt aufeinanderfolgenden Termini „Dachfläche“ und „Dachbegrünung“ fügte das Modell den Vorschlag „Dachflächenbegrünung“ zusammen. Die Beobachtungen lassen vermuten, dass ChatGPT die Halluzinationen direkt aus dem Text zieht und nicht allgemein aus dem Sachgebiet.

Weitere Bewertungsfaktoren

Die Termextraktionen führten je nach eingesetzter Methode zu teilweise sehr unterschiedlichen Ergebnissen (siehe Abbildung 3). Der gesetzte Referenzwert aus der manuellen Extraktion kann jedoch nicht als alleinige Bewertungsgrundlage dienen. Denn im Arbeitsalltag von Terminolog:innen, Übersetzer:innen und Redakteur:innen spielen außer den Faktoren Quantität und Qualität bezüglich einer Extraktion auch immer die Faktoren Zeit und Kosten eine Rolle. Zusätzlich sollten das Fehlerpotenzial und daraus resultierende Prozessrisiken betrachtet werden.

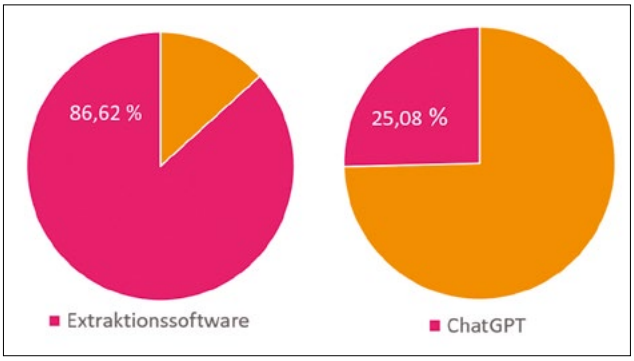


Abb. 3: Anteil extrahierter Termini bezogen auf das manuelle Referenzergebnis für Text 2 (Quelle: eigene Darstellung)

Um einen Vergleich des Faktors Zeit zu ermöglichen, wurde der Zeitaufwand pro Methode und Test dokumentiert. Obwohl bei der software- und KI-gestützten Extraktion einzelne Schritte automatisiert werden konnten, erfolgte die Validierung immer durch die beteiligten Terminolog:innen. Die Werte hängen daher immer auch davon ab, wie routiniert eine Aufgabe oder ein Tool um- bzw. eingesetzt wird. Bei ChatGPT wird die Bearbeitungszeit maßgeblich davon beeinflusst, ob verschiedene Prompts getestet werden müssen oder ob auf einen bereits vorformulierten Prompt zurückgegriffen wird.

Abbildung 4 zeigt, wie deutlich der Zeitunterschied beim längeren Text 2 ausfiel. Der Einsatz einer Extraktionssoftware sparte fast die Hälfte der Zeit einer manuellen Extraktion ein. Die Extraktion mit ChatGPT war allerdings mit nur einem Achtel der benötigten Zeit die ungeschlagen schnellste Methode. Das Formulieren und die Eingabe eines Prompts benötigten damit deutlich weniger Zeit als das Anlegen eines Extraktionsprojekts im Tool oder die manuelle Durchsicht mehrerer Seiten eines Dokuments, lieferten aber auch deutlich weniger validierte Termini. Die erhaltenen Ergebnisse waren durchweg sprachlich sauber und hätten zumindest in dieser Hinsicht ohne Nachbearbeitung weiterverwendet werden können. Sowohl bei der manuellen als auch bei der softwaregestützten Extraktion mussten Termini hingegen oft noch in die Grundform gebracht verwenden.

Der Einsatz von ChatGPT ermöglicht es außerdem, sich das Ergebnis für eine optimale Weiterverarbeitung in

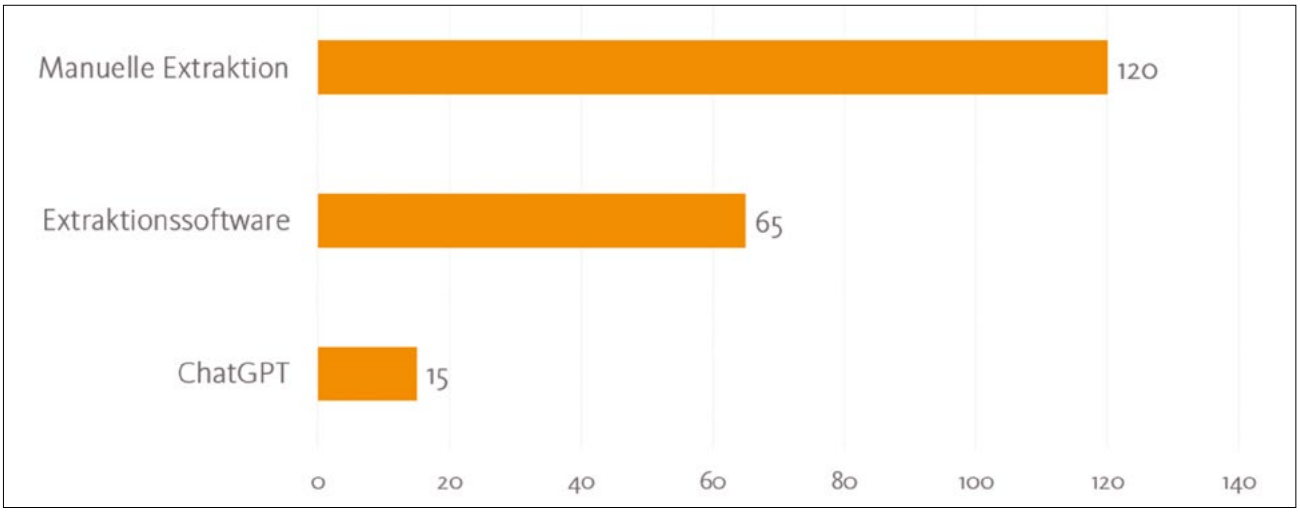


Abb. 4: Dauer der Extraktion und Validierung in Minuten für Text 2 (Quelle: eigene Darstellung)

unterschiedlicher Darstellungsweise (z. B. als Liste) oder in verschiedenen Dateiformaten ausgeben zu lassen. Damit sind Prozessautomatisierungen zum direkten Befüllen einer Datenbank möglich. Die Verarbeitung und Ausgabe der Ergebnisse sind dabei meist innerhalb weniger Sekunden abgeschlossen.

Obwohl die manuelle Extraktion in den Tests als Referenzwert entscheidend war und hier grundsätzlich von der höchsten Qualität ausgegangen werden kann, ist sie im Arbeitsalltag aufgrund des Zeitaufwands und der damit einhergehenden Kosten meist keine realistische Option.

Das Fehlerpotenzial ist je nach Methode unterschiedlich und wurde in den Abschnitten oben bereits erwähnt. Zusammenfassend lässt sich sagen, dass bei allen drei Methoden die Subjektivität bei der Validierung der Termkandidaten eine große Rolle spielt. Hinzu kommt bei der manuellen Extraktion die schwindende Konzentration bei großen Textmengen. Bei der softwaregestützten Extraktion muss pro Terminus die Fundstelle im Text berücksichtig werden, da extrahierte Termini zum Beispiel auch in Literaturangaben oder als Teil eines Unternehmensnamens vorkommen können. Bei aufgesplitteten Komposita (z. B. Stadt- und Dachbegrünung) wird der erste Teil (Stadtbegrünung) von den meisten Extraktionstools außerdem nicht berücksichtigt. Das Vorkommen im Text gilt auch als Fehlerpotenzial beim Einsatz von LLMs, bei denen zusätzlich noch Halluzinationen und die fehlende Reproduzierbarkeit der Ergebnisse hinzukommen.

Kritik

Während der Durchführung der Tests fielen einige Punkte auf, die kritisiert werden und als Gegenstand weiterer Untersuchungen dienen können.

Das bereits erwähnte Fehlen objektiver Kriterien zur Bewertung von Fachtermini bei der Extraktion führt zu subjektiven Entscheidungen. Zwei Terminolog:innen werden wahrscheinlich nie das exakt gleiche Ergebnis erhalten, da die Frage, was als Fachterminus gilt, nicht abschließend geklärt ist [1].

Zusätzlich kann die Vorgabe des Sachgebiets das Ergebnis maßgeblich beeinflussen: Je nachdem, wie weit ein Sachgebiet gefasst wird, werden mehr oder weniger Termini validiert, da mehr oder weniger Termini eingeschlossen werden. In unserem Beispiel gilt „Umwelt“ als übergeordnetes und „nachhaltiges Starkregenmanagement“ als spezifisches Sachgebiet.

Ebenso wird vermutet, dass das sachgebietsspezifische Vorwissen der Terminolog:innen die Extraktion beeinflusst. Termini, die zum eigenen Arbeitsalltag gehören, werden eventuell als weniger fachlich eingestuft und damit nicht validiert. Termini aus fremden Sachgebieten werden ggf. wiederum tendenziell als fachlicher eingestuft [5].

Bezüglich des Vergleichs des Faktors Zeit bei den drei Methoden ist anzumerken, dass aufgrund der Durchführung der Tests mit denselben Texten und Durchführenden die Fachtermini nach den ersten Durchläufen als bekannt gelten konnten. Im ersten Durchlauf (manuelle Extraktion) musste pro Terminus noch eine Grundentscheidung zur Validierung getroffen werden, die für alle weiteren Tests entfiel. Es wird daher vermutet, dass die Validierung bei jedem folgenden Test tendenziell schneller erfolgen konnte.

Bei der software- und KI-gestützten Extraktion erfolgte die Validierung der Termini im ersten Schritt unabhängig vom Referenzergebnis, denn nur dies entspricht den realen Bedingungen einer Termextraktion. Die validierten Termini wurden anschließend mit dem Referenzergebnis abgeglichen und nur die darin vorkommenden Termini wurden als Ergebnis gezählt. Aufgrund der angesprochenen Subjektivität ist es möglich, dass Termini aus dem Referenzergebnis im Rohergebnis der beiden untersuchten Methoden enthalten waren, aber nicht validiert wurden. In diesem Fall wären sie nicht im Endergebnis enthalten, obwohl sie eventuell zu den Termkandidaten gehörten. Dieses Fehlerpotenzial kann jedoch als moderat bis gering eingeschätzt werden, da die Termini nach mehreren Tests als bekannt gelten durften.

Im Rahmen der softwaregestützten Extraktion gibt es weitere Kritikpunkte, welche die Parametereinstellungen und die Auswahl der Extraktionssoftware betreffen. Im jeweils dritten Durchlauf beider Texte wurde zusätzlich zur Silence-Noise-Ratio auch die maximale Terminuslänge verringert und damit zwei Variablen gleichzeitig geändert. Diese Anpassung könnte weitere, hier nicht berücksichtigte Auswirkungen auf das Extraktionsergebnis gehabt haben. Zur Untersuchung möglicher Auswirkungen müssen Durchläufe erfolgen, bei denen jeweils nur einer der beiden Parameter angepasst wird.

Ebenfalls anzumerken ist die Auswahl des Extraktionstools, da MultiTerm Extract Termextraktionen auf Grundlage eines statistischen Verfahrens ausführt. Ergänzend wäre der Vergleich mit einem linguistisch oder hybrid arbeitenden Extraktionstool aufschlussreich.

Eine letzte kritische Betrachtung gilt den Prompting-Strategien. Reines Task-Specific-Prompting wurde für die Aufgabenstellung einer Termextraktion als nicht zielführend eingeschätzt, weshalb es nicht alleinstehend, sondern nur in Kombination mit Domain-Specific-Prompting getestet wurde. Diese Entscheidung liegt darin begründet, dass es sich bei einem Terminus um die „sprachliche Bezeichnung eines Allgemeinbegriffs aus einem Fachgebiet“ [6] handelt. Ob es sich bei einem Wort um einen Terminus

handelt oder nicht, hängt somit vom jeweiligen Sachgebiet ab, in dem das Wort vorkommt. Wird kein Sachgebiet vorgegeben (was bei reinem Task-Specific-Prompting der Fall wäre), lässt sich aus theoretischer Perspektive auch kein Terminus extrahieren. Es erfolgt daher in den Versuchen kein Vergleich zwischen den reinen Prompting-Strategien, sondern immer nur die Analyse der Beeinflussung zusätzlicher Arbeitsanweisungen.

Aufgrund des Einsatzes einer statistischen Extraktionssoftware wäre zusätzlich ein Prompt mit der Aufforderung der Anwendung eines rein statistischen Verfahrens bei der Termextraktion von Interesse (z. B. „Extrahiere die häufigsten Fachtermini aus dem vorhandenen Text.“).

Fazit

Letztendlich kann jede Methode für einen bestimmten Anwendungsfall nützlich sein. Manuelle Extraktion als Goldstandard liefert die umfangreichsten und verlässlichsten Ergebnisse, scheitert im Arbeitsalltag aber oftmals an der Umsetzbarkeit. KI liefert schnelle, aber teilweise auch falsche Ergebnisse und konnte rein quantitativ nicht mit den anderen beiden Methoden mithalten. Die Schnelligkeit und Automatisierbarkeit des Prozesses ermöglichen aber einen Terminologieaufbau „on the fly“, bei dem ohne großen Aufwand nach und nach ein Bestand aufgebaut werden kann.

Auf Termextraktion spezialisierte Software stellte sich im Test als der optimale Mittelweg heraus. In unseren Vergleichen erreichte die Software bei geringerem zeitlichen Aufwand gute Ergebnisse, die jedoch stark von den Einstellungen abhingen.

Es lässt sich also festhalten, dass viele Wege zur Termextraktion führen und kein klarer Sieger aus dem Ring steigt. Ausschlaggebend für die Entscheidung, welche Methode eingesetzt wird, ist daher die Frage nach der Zielsetzung und den vorhandenen Ressourcen. Und ganz unabhängig davon, in welchem Maße der Mensch sich durch Software und künstliche Intelligenz unterstützen lässt, bleibt er der zentrale Faktor für die Überprüfung und Validierung der Rohergebnisse.

Literatur

[1] Drewer, Petra; Schmitz, Klaus-Dirk (2017): Terminologiemanagement. Grundlagen, Methoden, Werkzeuge. Berlin: Springer Verlag

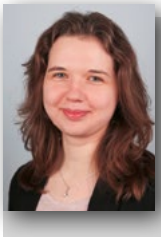
[2] Fagbohun, Oluwole; Harrison, Rachel M.; Dereventsov, Anton (2024): An Empirical Categorization of Prompting Techniques for Large Language Models: A Practitioner's Guide. Preprint. URL: https://arxiv.org/pdf/2402.14837 [letzter Zugriff: 20.08.2024]

[3] Zhou, Yongchao; Muresanu, Andrei Ioan; Han, Ziwen; Paster, Keiran; Pitis, Silviu; Chan, Harris; Ba, Jimmy (2023): Large Language Models are Human-Level Prompt Engineers. Version 2. arXiv. URL: https://arxiv.org/pdf/2211.01910 [letzter Zugriff: 20.08.2024]

[4] Peng, Keqin; Ding, Liang; Zhong, Qihuang; Shen, Li; Liu, Xuebo; Zhang, Min; Ouyang, Yuanxin; Tao, Dacheng (2023): Towards Making the Most of ChatGPT for Machine Translation. arXiv. URL: https://arxiv.org/pdf/2303.13780 [letzter Zugriff: 20.08.2024]

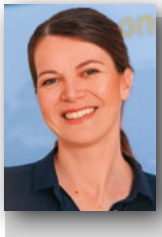
[5] Janke, Regina (2013): Anforderungen an die Terminologieextraktion: Eine vergleichende Untersuchung der Bedürfnisse von Terminologen, Technischen Fachübersetzern und Technischen Redakteuren. Stuttgart: tcworld (tekom Hochschulschriften 20)

[6] DIN 2342:2022-07: Terminologiewissenschaft und Terminologiearbeit – Begriffe. Berlin: DIN Media



Vanessa Jochum studiert derzeit im Masterstudiengang Terminologie und Sprachtechnologie an der Technischen Hochschule Köln und arbeitete ein Jahr lang als Werkstudentin im Terminologiemanagement bei der oneword GmbH. Ihr Interessensschwerpunkt liegt auf unterschiedlichen Einflussfaktoren auf die Termextraktion.

Kontaktadresse
vanessajochum@aol.com



Jasmin Nesbigall ist Diplom-Übersetzerin und Fachleitung für MTPE und Terminologiemanagement bei der oneword GmbH in Böblingen. Sie berät und betreut Unternehmen bei allen Schritten der Terminologiearbeit und beim Einsatz von KI und Automatisierungen in Terminologieprozessen.

Kontaktadresse
j.nesbigall@oneword.de
www.oneword.de

kostenloses Online-Archiv

20. Jahrgang • 2005-2024

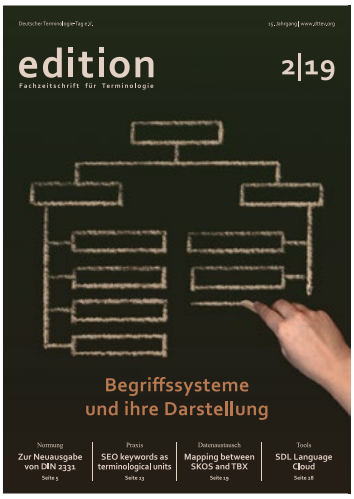
edition

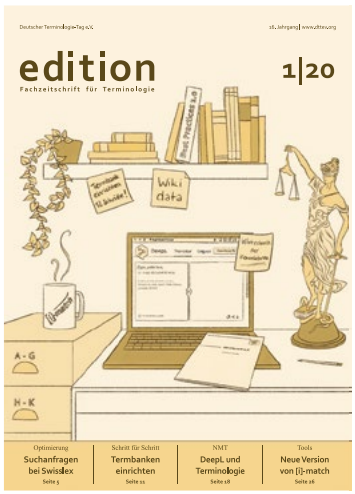
Fachzeitschrift für Terminologie

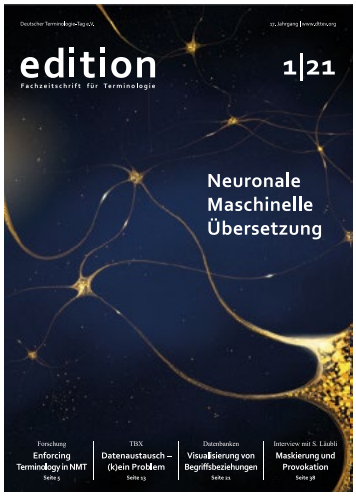
seit 2005

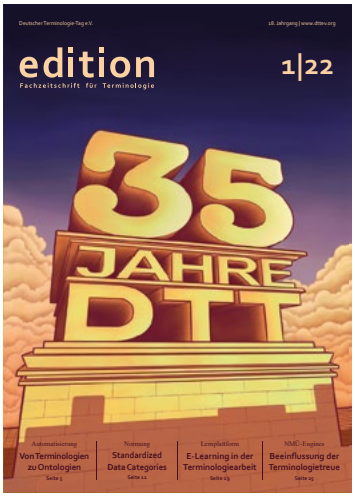
www.dttev.org/edition

Alle seit 2005 erschienenen Ausgaben der edition finden Sie als PDF in unserem kostenlosen Online-Archiv.
Perfekt zum Stöbern, Nachschlagen und Weitergeben. Wir wünschen viel Spaß beim Lesen!









Worldbuilding – Benennungsbildung einmal anders

Interview mit

Literaturübersetzerin Christine Heinzius

Christine Heinzius übersetzt Krimis, Handarbeitsbücher, Kochbücher – und Fantasy-Romane. Seit dem Abschluss ihrer Studiums der Anglistik, Romanistik und Skandinavistik in Saarbrücken, Paris, Leicester und Helsinki arbeitet sie als freiberufliche Übersetzerin mit Schwerpunkt Literatur. Bei der Übersetzung von Fantasy-Romanen gibt es teilweise ganz besondere terminologische Herausforderungen. Christine Heinzius hat bei der Übersetzung eines solchen Fantasy-Romans mitgewirkt: „When The Moon Hatched" von Sarah A. Parker.

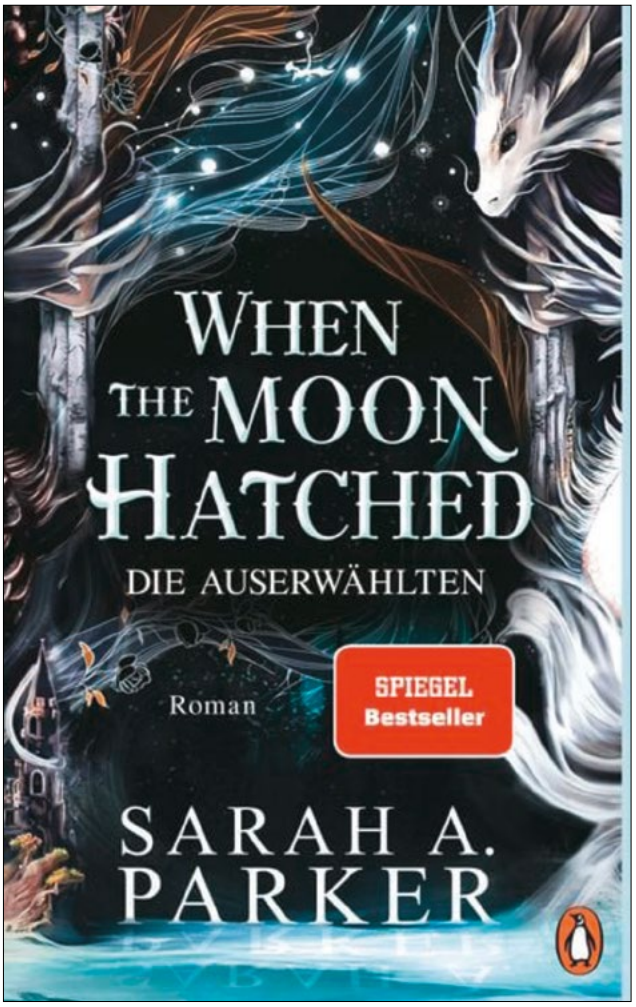


Abb. 1: Buchcover von „When The Moon Hatched" (Quelle: Penguin Verlag)

Redaktion: Frau Heinzius, Sie haben vor kurzem an der Übersetzung des Fantasy-Romans „When The Moon Hatched" mitgearbeitet. Wenn Sie im Team an einem Buch arbeiten, in dem eine ganz neue Welt der Begriffe erfunden werden muss – Worldbuilding genannt – wie lösen Sie das zusammen? Setzen Sie sich vorab zusammen und erarbeiten die neuen Benennungen für die Fantasy World? Oder bilden Sie Benennungen „on the fly" und teilen diese dann den anderen mit? In „When The Moon Hatched" gibt es ja am Ende sogar ein Glossar, das auch ins Deutsche übersetzt wurde. Haben Sie das zuerst übersetzt und als Vorlage für die weitere Übersetzung genommen?

Heinzius: Im Englischen ist das Glossar dem Roman vorangestellt, so dass der Kollege, der für den ersten Teil zuständig war, damit begonnen hat. Er hat die Übersetzung dann sofort als Google-Docs-Dokument eingestellt, so dass wir anderen Zugriff darauf hatten und uns daran orientieren konnten. Im Großen und Ganzen haben wir seine Vorschläge übernommen, bei einzelnen Wörtern gab es dann auch mal Gegenvorschläge, Nachfragen oder Diskussionen. Das alles blieb aber komplett unter uns Übersetzern; der Verlag hatte nichts damit zu tun.

Ganz konkret in diesem Fall haben wir einfach losgelegt, was natürlich auch daran lag, dass es ein „Schnellschuss" war, d. h., das Buch musste so schnell wie möglich auf Deutsch erscheinen. Deswegen haben wir auch im Team daran gearbeitet. Das Buch wurde auf fünf gleiche Teile aufgeteilt (ich hatte den zweitletzten) und dann hat jeder sofort mit der Übersetzung begonnen. Aber ganz ehrlich, angesichts der Seitenhonorare hatten wir gar nicht die Zeit, vorher lange zu diskutieren und abzuwägen. Wir hatten gegen Ende ein Zoom-Meeting von einer halben Stunde, um letzte Fragen abzuklären, alles andere lief schriftlich.

Redaktion: Wir haben uns die englische Leseprobe zu „When The Moon Hatched" angeschaut. Da sind ja einige sehr ungewöhnliche Wortschöpfungen enthalten, wie „sowmoth powder", „parchment lark" oder „spangle poop". Wie ge-

hen Sie bei der Bildung neuer deutscher Benennungen für solche Begriffe vor? Haben Sie dafür eine Strategie?

Heinzius: Ich schaue mir die Wörter und was sie bezeichnen erstmal an. Erinnert das Wort an gängige englische Wörter? Ist es vielleicht aus einem normalen Wort plus Fantasieergänzung aufgebaut? Spielt es mit alten, nicht mehr gebräuchlichen Wörtern oder mit eher ungebräuchlichen Bedeutungen eines Worts? Enthält es ein Wortspiel? Ist es vielleicht sogar Satire und bezieht sich auf etwas in der realen Welt oder auf Fantasy-Klischees oder Ähnliches? Ist es eher eine witzige Formulierung?

Dann lese ich mir die Wörter laut vor: Spielt die Autorin mit dem Klang? Gibt es Alliterationen oder phonetische Auffälligkeiten? Gibt es einen bestimmten Rhythmus? Ist es kurz und knapp oder lang und ausschweifend? Soll es einer konkreten Fantasiesprache entspringen? Falls ja, gibt es mehrere, unterschiedliche im Roman, mit unterschiedlichen Klangbildern? Und wenn ich das analysiert habe (was natürlich nicht so systematisch mit abzuhakender Liste geschieht), dann gilt es, ein entsprechendes deutsches Wort zu prägen, d. h. eins, das ebenfalls z. B. altmodisch klingt oder melodiös. Dazu kommt noch eine Grundsatzentscheidung: Was kann/muss man komplett im Original stehen lassen oder doch fast im Original, so dass man das Wort nur grammatikalisch ans Deutsche anpasst? Letzteres ist durch Social Media häufiger geworden, weil Leserinnen bereits auf BookTok oder Instagram von dem Buch gehört haben und dann dieselben oder doch sehr ähnliche Begriffe in der deutschen Ausgabe erwarten. Ich persönlich finde es schade, weil dadurch viele kreative, neue Wörter gar nicht erst entstehen.

Redaktion: Sind Sie bei solchen neuen Wörtern frei in Ihrer Übersetzungsentscheidung oder stimmen Sie alle neuen Benennungen mit dem Lektorat ab? Oder leistet der Verlag die Vorarbeit und gibt die neue Terminologie vor?

Heinzius: Nein, der Verlag hat gar nicht die Kapazitäten, um da irgendwas vorzuarbeiten, das liegt ganz allein bei den Übersetzerinnen und Übersetzern. In diesem Fall haben wir uns untereinander abgestimmt, d. h., wenn einem Teammitglied eine Übersetzung nicht gefiel, hat es das angesprochen. Oder wenn wir uns bei einer eigenen Übersetzung unsicher waren, haben wir bei den anderen nachgefragt, ob die den Vorschlag gut finden oder eine bessere Idee haben.

Wenn ich so einen Roman allein übersetze, wende ich mich bei solchen Zweifelsfällen an die Verlagslektorin, aber hier war das nicht nötig. Auch mit der freien Lektorin (in beiden Fällen sind es fast immer Frauen), könnte man sich absprechen, wenn das Bedürfnis besteht.

Redaktion: Selbst wenn Sie die meisten Wortschöpfungen vorab übersetzen und klären können, tauchen mit Sicherheit noch weitere neue Benennungen während der Übersetzung auf, die auch einheitlich übersetzt werden sollen. Wie gehen Sie da vor? Wie erfolgt die Abstimmung und wie werden die neuen Benennungen festgehalten und mit allen geteilt?

Heinzius: In diesem Projekt haben wir ein zentrales Google-Docs-Dokument angelegt. Jedes Teammitglied hat seine eigene Schriftfarbe bekommen, so dass sofort ersichtlich war, wer eine neue Übersetzung eingestellt, eine andere kritisiert oder irgendeine Frage gestellt hat. Das Glossar wurde als Erstes eingestellt und damit schon der Großteil der Neologismen. Wenn ein Teammitglied bei der weiteren Übersetzung auf neue Wörter stieß, wurden sie ebenfalls in das Glossar eingetragen, so dass wir alle immer auf dem Laufenden waren.

Redaktion: Sie haben auch viele historische Romane übersetzt. Unterscheidet sich die Vorgehensweise hier von modernen Romanen?

Heinzius: Ich habe viele historische Romane zu zweit übersetzt. Auch da haben wir das Buch einfach in der Mitte geteilt und jede Kollegin fing mit ihrem Teil direkt an. Ich kenne Kolleginnen, die die Kapitel abwechselnd übersetzen, um den Stilunterschied zu minimieren. Das habe ich selbst aber noch nie gemacht.

Bei meiner Arbeit unterscheide ich zwischen verschiedenen Arten, sich abzusprechen: Zum einen erstellen wir eine Liste mit Wörtern; darauf landen z. B. die Adelstitel oder Anreden wie „Eure Durchlaucht“. Bei modernen Romanen halten wir auch noch fest, wer wen siezt und wer wen – ab wann – duzt. Diese Liste ist normalerweise auch ein Google-Docs-Dokument, damit alle Beteiligten immer auf die aktuellste Version zugreifen können. Eine andere Möglichkeit bieten Word-Dokumente, in denen wir Links und/oder Erklärungen und teilweise auch Bilder aufnehmen. Schließlich haben wir uns auch ganz frei per E-Mail oder am Telefon ausgetauscht. Das gilt vor allem für romanimmanente Sachverhalte, damit auch da die Terminologie übereinstimmt und man nicht z. B. an einer Stelle vom „lila Morgenmantel“ und an einer anderen vom „violetten Bademantel“ schreibt. Das dient auch dazu abzustimmen, welche Wörter in der Ausgangssprache stehen bleiben, was ja für das Lokalkolorit durchaus wichtig sein kann.

Redaktion: Laut Ihrem Portfolio haben Sie auch verschiedene Handarbeitsbücher – vor allem Strickbücher – übersetzt. Würden Sie solche Übersetzungen aus der Hobby-Welt als Fachübersetzungen bezeichnen? Schließlich

müssen Sie auch in diesem Kontext die Fachbegriffe kennen und konsistent verwenden. Haben Sie sich für diese Art von Übersetzungen eine Terminologiedatenbank mit Strickterminologie angelegt oder gibt es vielleicht kommerzielle Datenbanken für Strickterminologie?

Heinzius: Ja, das sind für mich schon auch Fachübersetzungen. Als Strickerin merkt man sofort, wenn der Übersetzer oder die Übersetzerin gar nicht stricken oder häkeln kann. Außerdem denke ich beim Übersetzen mit und finde immer wieder Fehler im Original, die dann mit der Designerin abgeklärt werden müssen. Ich habe keine echte Terminologiedatenbank angelegt, obwohl ich natürlich im Grunde eine verwende, aber die befindet sich vor allem in meinem Kopf und in den schon übersetzten Büchern oder aber auf Papier. Denn von manchen Verlagen bekomme ich Wörterlisten; die habe ich dann handschriftlich ergänzt: entweder mit darauf fehlenden gängigen Übersetzungen oder mit von mir neu geprägten.

Selbst kreiert habe ich z. B. die Abkürzung „aaz“, die auf Englisch „ssk“ heißt, „slip slip knit“ bzw. „abheben, abheben, zusammenstricken“. Diese Technik ist relativ neu und stammt aus dem englischen Sprachraum, weswegen es keinen Begriff und keine Abkürzung dafür auf Deutsch gibt. Da sie aber sehr praktisch und beliebt ist, hat sie sich recht schnell weit verbreitet und brauchte eine deutsche Abkürzung.

Kommerzielle Datenbanken für Strickbegriffe kenne ich nicht, was oft hilft, ist diese Seite: https://www.garnstudio.com/glossary.php?langf=de&langt=en&lang=de. Dort können etliche Sprachkombinationen nachgeschlagen werden. Inzwischen kenne ich alle gängigen Abkürzungen. Taucht in einem Buch doch eine neue auf, dann ist die so speziell, dass sie auch auf dieser Seite nicht zu finden ist. Da recherchiere ich dann in Blogs oder auf YouTube und immer wieder kommt es vor, dass ich selbst eine neue Benennung prägen muss. Außerdem gibt es immer wieder Designerinnen, die ihre eigenen Abkürzungen erfinden, die übernehme ich allerdings nicht, sondern ersetze sie im Deutschen durch die gängigen.

Redaktion: Die Übersetzung von Kochbüchern gehört auch zu Ihrem Repertoire. Können Sie unseren Leserinnen

Wortliste

Damit möglichst viele Personen von unserer Garnstudio Homepage und den Gratisanleitungen profitieren können, veröffentlichen wir hier Wortlisten, Ihnen hoffentlich weiterhelfen.

Translate terms from Deutsch to English (UK)

Deutsch	English (UK)
1 cm = 0,39 inch	1 inch=2,54 cm
1 Krausrippe (2 R. re.)	garter stitch (2 rows)
1 M. abheben	slip 1 stitch, sl 1
1 M. re. abheben	slip one knitwise (sl1 kw)
1 re, 1 li	k1p1, knit one, purl one
2 M. li. zusammenstricken	p2tog - purl 2 together
2 M. re zusammenstricken	k2tog knit 2 together
abketten	cast off /fasten off
abketten	cast off / fasten off
abketten	bind off (BO)
abketten	casting off

Abb. 2: Online-Strickglossar (Quelle: https://www.garnstudio.com/glossary.php)

und Lesern irgendwelche „terminologischen Leckerbissen" servieren, zum Beispiel aus dem Kochbuch zum Thema Lakritz? Lakritz ist in Deutschland ja nicht so verbreitet wie in Skandinavien. Mussten Sie da auch neue Benennungen bilden? Vielleicht haben Sie daraus ein Beispiel, das Ihnen besondere Schwierigkeiten bereitet hat – oder vielleicht auch eines, das lustig war?

Heinzius: Das Nervigste bei skandinavischen Kochbüchern ist nicht die Terminologie, sondern die Maßeinheit. Sehr oft wird ein Hohlmaß verwendet und nicht das Gewicht wie bei uns. Also 1 Liter Erdbeeren. Im Internet finden sich viele Listen für die Umrechnung und ich habe auch selbst eine Umrechnungstabelle angelegt, aber manchmal muss man die Autorin fragen, z. B. bei Dingen wie 300 ml Tahini oder 250 ml Tacochips oder Perl-Couscous. Letzteren kannte ich überhaupt nicht und musste lange recherchieren.

Schwierig sind auch Zutaten, die es bei uns nicht gibt, wie „Filmjölk" in Schweden. Die wird hier als Schwedenmilch verkauft, aber nur sehr, sehr selten.

Das letzte Kochbuch, das ich übersetzt habe, hatte Nachhaltigkeit als Thema, da musste ich sehr viel an Deutschland anpassen. Nicht bei den Rezepten, sondern bei den Begleittexten. Da gab es z. B. Ökosiegel, die es nur in Schweden gibt, oder hier nicht bekannte Organisationen und Aktionen. Am Ende fiel – natürlich in Absprache mit dem Verlag – ein ganzes Kapitel einfach weg.

Bei dem Lakritzkochbuch musste ich zunächst einmal herausfinden, was es an Grundstoffen gibt: Lakritzgranulat und -pulver kannte ich z. B. vorher auch nicht. Witzig fand ich die Lakritzfestivals und die finnische Salmiakgala, bei der jährlich der Preis für das beste Salmiakprodukt verliehen wird.

Bei einem Kochbuch zu Insekten war die Herausforderung, die genauen Bezeichnungen der Krabbeltierchen herauszufinden und dann alles so zu formulieren, dass es nicht eklig klingt. Das Kochbuch war allerdings nur kurz auf dem Markt (im Gegensatz zum Longseller Lakritz), so dass die Strategie wohl nicht aufgegangen ist.

Redaktion: Naturbücher haben Sie auch übersetzt. Können Sie aus diesem Bereich noch etwas Interessantes für unsere Leserschaft berichten?

Heinzius: Ich übersetze seit Jahren einen Vogelkalender aus dem Schwedischen und habe auch schon ein Gartenbuch übersetzt. Da ist bei der Terminologie das Problem, dass Tier- und Pflanzennamen im Deutschen regional sehr unterschiedlich sind. Ist es Feldsalat oder Rapunzel oder Vogerlsalat oder Mausohr? Ich versuche dann, das auszuwählen, was mir am verbreitetsten erscheint. Aber natürlich kann es passieren, dass ein Norddeutscher das völlig anders einschätzen würde.

Der Vogelkalender, bei dem ich teilweise auch texte, wenn es „zu schwedisch" ist, wird von einem Ornithologen lektoriert. Der hat in einem Jahr aus meinem Dompfaff einen Gimpel gemacht. Beide Namen sind in Deutschland gebräuchlich, aber ich musste es zurückändern, weil es im Text darum ging, woher der Vogel seinen Namen hat, nämlich vom Domherren, dem er mit rotem Bauch und schwarzem Mützchen ähnelt. Mit Gimpel hat der ganze Text nicht mehr funktioniert.

Redaktion: Frau Heinzius, wir bedanken uns für das Gespräch und wünschen Ihnen weiterhin viel Kreativität und Einfallsreichtum bei Ihren Übersetzungsprojekten.

Das Interview führten
Dr. Nicole Keller und
Angelika Ottmann

Abb. 3: Literaturübersetzerin Christine Heinzius (Quelle: privat)

flashterm: Aktuelle Funktionserweiterungen im Überblick

Frauke Eschenbeck und Lola Klan Reyes

flashterm ist ein spezialisiertes System der Eisenrieth Dokumentations GmbH zur Organisation, Verwaltung und Kontrolle von Fachbegriffen und deren Übersetzungen. Seit seiner Einführung im Jahr 2008 hat sich flashterm (www.flashterm.eu) zu einem vielseitigen und leistungsstarken Werkzeug für das Terminologie- und Wissensmanagement entwickelt.

In jüngster Zeit wurde eine Reihe von Erweiterungen umgesetzt, die das Anwendungsspektrum deutlich erweitert haben. Dieser Artikel beleuchtet folgende zentrale Neuerungen: Wissensansicht, strukturierte Prompts, Aktualisierung des ClipInspectors, verbesserte Visualisierung von Begriffsbeziehungen sowie flashterm Stage.

Wissensansicht

Die Wissensansicht (InfoView) in flashterm ist eine erweiterte Ansicht, die neben der Begriffsdefinition auch weiterführende spezifische Informationen liefert, um ein tieferes Verständnis des jeweiligen Begriffs zu ermöglichen. Im Vergleich zur Standardansicht, der TermView, umfasst die InfoView neben der Darstellung eines Begriffs kontextbezogene Zusatzinformationen, die für die umfassende Erschließung von Fachbegriffen von Bedeutung sind. Die Befüllung dieser InfoView erfolgt über ein Freitextfeld, in das Informationen eingegeben werden können, die mit Markdown-Elementen formatiert und nach Bestätigung der Eingabe automatisch in eine HTML-Darstellung umgewandelt werden.

Die Wahl zwischen der InfoView und der TermView kann in den Benutzereinstellungen getroffen werden, sodass die bevorzugte Ansicht standardmäßig angezeigt wird. Die TermView konzentriert sich auf grundlegende Informationen wie Definition oder Kontextbeispiele, während die InfoView zusätzliche Inhalte bietet, die besonders hilfreich für Nutzer sind, die sich intensiver mit den dargestellten Fachbegriffen auseinandersetzen möchten.

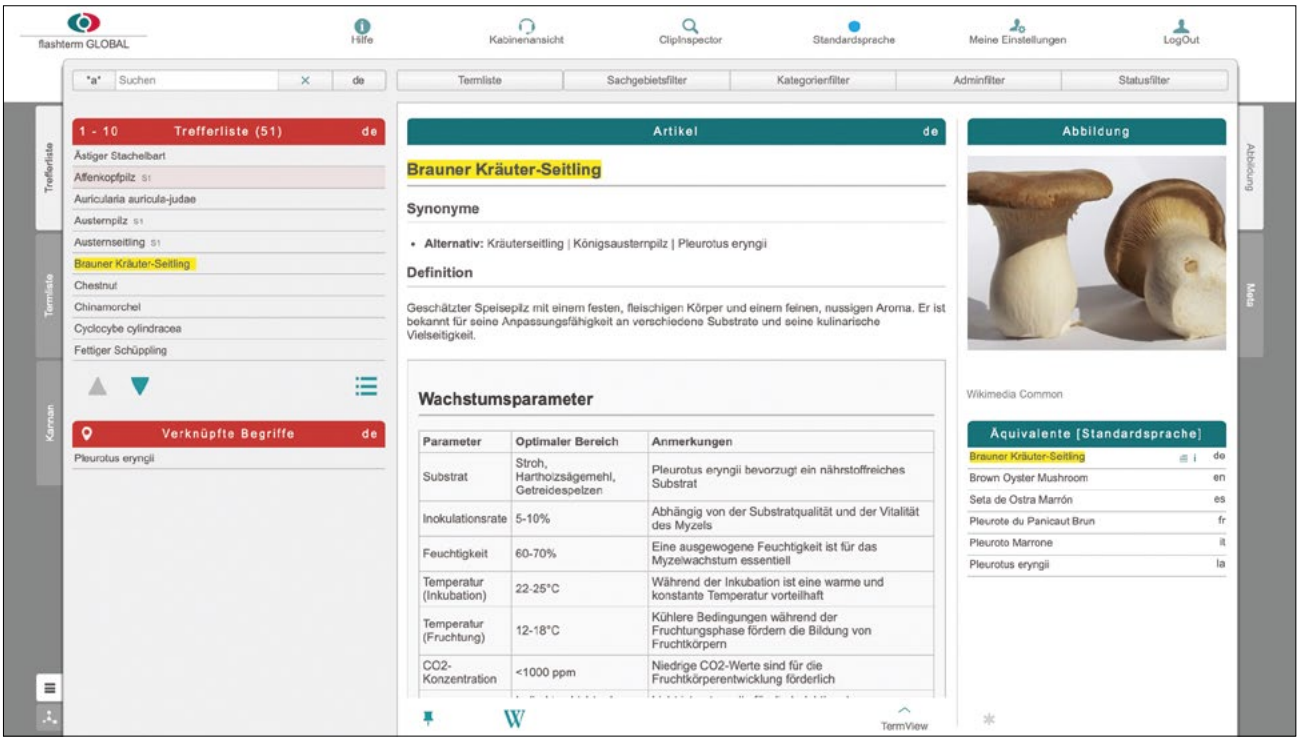


Abb. 1: InfoView (Quelle: eigener Screenshot)

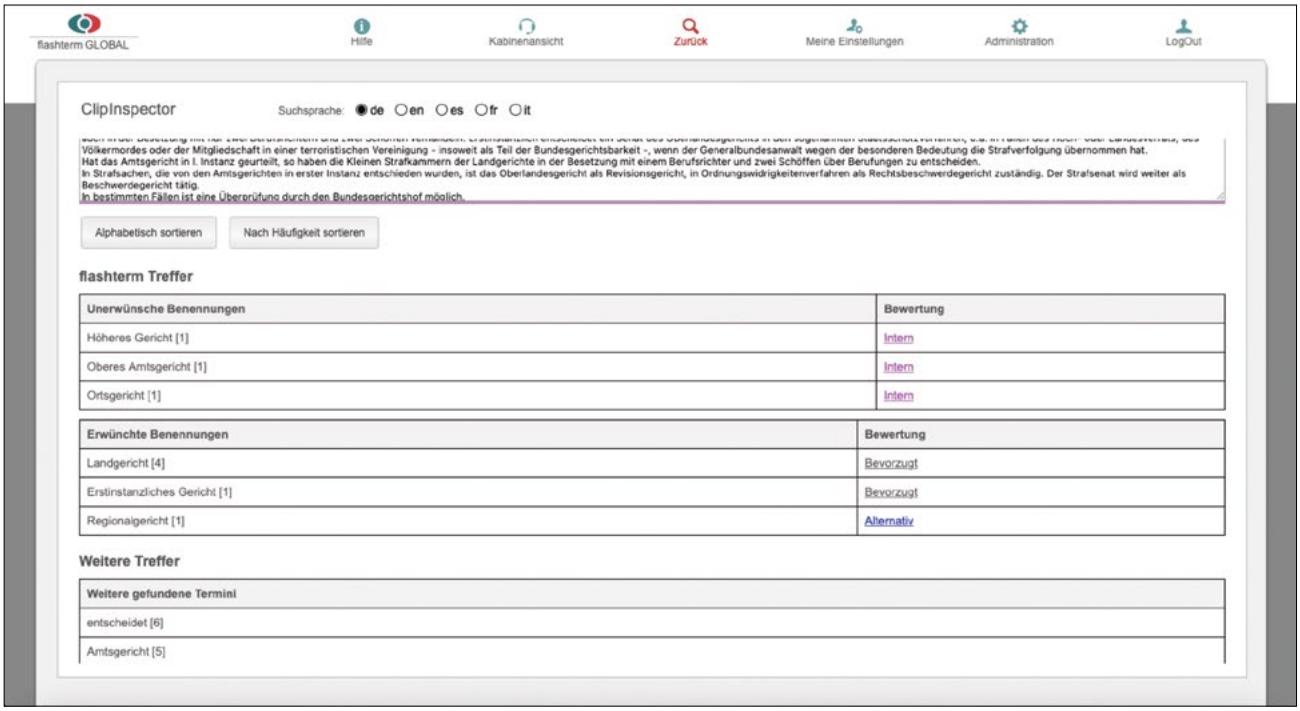


Abb. 2: ClipInspector (Quelle: eigener Screenshot)

Abbildung 1 auf S. 28 zeigt die InfoView zur Benennung „Brauner Kräuter-Seitling“. Neben der Begriffsdefinition und den Synonymen enthält sie weiterführende Informationen wie Wachstumsparameter, die in der TermView nicht sichtbar sind. Außerdem umfasst die InfoView in diesem Fall eine Abbildung des Pilzes sowie die äquivalenten Benennungen in weiteren Sprachen, die beide auch in der TermView zu sehen sind.

Die InfoView nimmt somit eine zentrale Rolle in der Benutzeroberfläche von flashterm ein, indem sie durch ihre weiterführenden Erklärungen und Zusatzinformationen eine wertvolle Ergänzung für die semantische Erschließung bietet. Sie trägt maßgeblich dazu bei, Fachwissen nicht nur darzustellen, sondern auch vertiefend zu vermitteln, was sie zu einem wichtigen Werkzeug für die Wissensintegration macht.

Strukturierte Prompts

Die Funktion „Strukturierte Prompts“ ermöglicht es den Nutzern, vordefinierte Prompts zu erstellen, die anschließend mittels Copy-and-Paste an einen externen KI-Chatbot wie beispielsweise ChatGPT übermittelt werden können. Die Prompts sind dabei flexibel gestaltbar und unterstützen verschiedene Variablen. Ein Beispiel für einen solchen Prompt lautet: „Verfasse einen Artikel über [Benennung] im Bereich [Sachgebiet]. Die Länge des Textes beträgt etwa 400 Zeichen.“ Die Auswahl der Variablen für Benennungen und Sachgebiete kann dabei aus der Datenbank erfolgen. Es steht eine Auswahl vordefinierter Prompts zur Verfügung, um Artikel zu verfassen, Zusatzinformationen zu liefern, Synonyme und Kontext zu bestimmten Benennungen vorzuschlagen und Äquivalente aufzulisten. Zudem soll die Funktion der strukturierten Prompts um die Möglichkeit erweitert werden, beliebig viele individuelle Prompts zu erstellen. Die Funktion bietet eine erhebliche Zeitersparnis bei der Terminologierecherche und verbessert die Qualität der Ergebnisse durch vordefinierte und flexibel einsetzbare Vorlagen.

Aktualisierung des ClipInspectors

Der ClipInspector ist eine Textanalyse-Funktion in flashterm, die eine gezielte Suche nach spezifischen Benennungen in Texten ermöglicht. Nach Auswahl der zu analysierenden Sprache wird der Text in das dafür vorgesehene Eingabefeld eingegeben. Im Anschluss führt der ClipInspector eine gründliche Analyse durch. Dabei werden vorhandene Benennungen aus der Datenbank identifiziert und neue, bisher in der Datenbank nicht erfasste Benennungen erkannt. Die Resultate der Analyse werden gemäß der in Abbildung 2 dargestellten Gliederung übersichtlich präsentiert. In der Kategorie „flashterm Treffer“ werden Benennungen angezeigt, die bereits in der Datenbank vorhanden sind. Die Kategorie „weitere Treffer“ hingegen erstellt eine Liste der neuen Benennungen, die noch nicht in die Datenbank aufgenommen wurden. Eine grafische Darstellung in Form einer Wortwolke veranschaulicht die Häufigkeit der im Text identifizierten Benennungen. Dabei werden häufigere Benennungen größer und auffälliger dargestellt. Des Weiteren bietet der ClipInspector eine detaillierte Übersicht

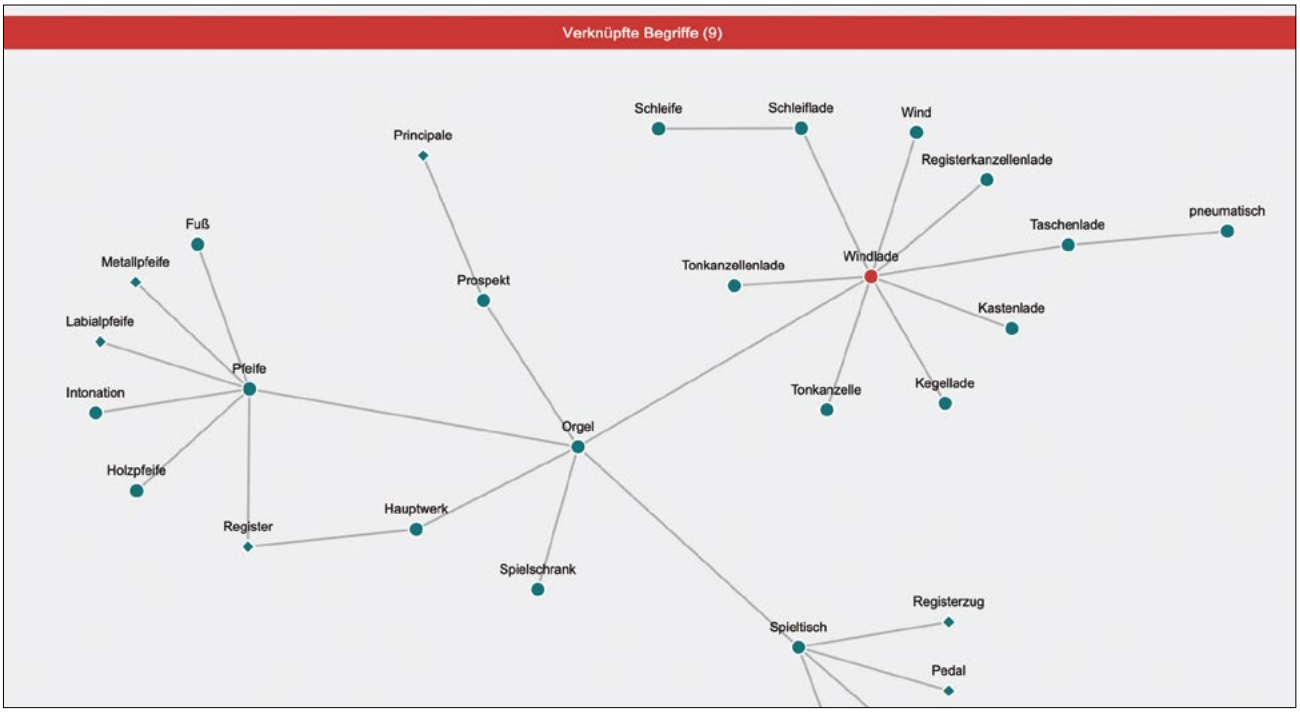


Abb. 3: Grafisches Begriffssystem (Quelle: eigener Screenshot)

über die im Text vorkommenden Kollokationen. Mit dem ClipInspector lässt sich systematisch überprüfen, ob alle relevanten Fachbegriffe eines bestimmten Themengebiets in der Datenbank erfasst sind. Das Einpflegen bislang nicht erfasster Begriffe ist derzeit noch zeitaufwendig. Allerdings ist eine neue Funktion zum Hinzufügen neuer Begriffe in die Datenbank geplant. Insgesamt zielt der ClipInspector darauf ab, die terminologische Analyse fachsprachlicher Texte zu vereinfachen und zu beschleunigen.

Visualisierung der Begriffsbeziehungen

Die Vernetzung thematisch verwandter Begriffe stellt ein neues Element der Terminologiearbeit in flashterm dar. Die Darstellung der Beziehungen erfolgt entweder in Form einer Liste oder als visuelles Begriffssystem. Die grafische Darstellung, wie in Abbildung 3 illustriert, bietet im Vergleich zur Listenform eine Visualisierung der Verknüpfungen zwischen den Begriffen. Der fokussierte Begriff wird dabei stets als roter Kreis hervorgehoben (vgl. „Windlade“ in Abbildung 3). Verknüpfte Begriffe erscheinen als grüne Kreise, während Begriffe, die weitere Verknüpfungen aufweisen, wegen des begrenzten Platzes durch Rauten dargestellt werden. Die Bezeichnungen der Begriffe sind direkt über den jeweiligen Symbolen platziert. Um die Übersichtlichkeit bei komplexen und verflochtenen Begriffssystemen zu verbessern, kann diese Darstellung auch im Vollbildmodus angezeigt werden. Ein Nachteil bei der Visualisierung von Begriffsbeziehungen bei flashterm ist, dass keine Auswahl verschiedener Relationen möglich ist. Es wird lediglich angezeigt, welche Begriffe miteinander verknüpft sind, ohne darzustellen, in welcher Beziehung sie zueinander stehen. Dem Nutzer bleibt somit selbst überlassen, ob er die Begriffe beispielsweise nach Ober- und Unterbegriffen gliedert oder eine andere Relation bevorzugt.

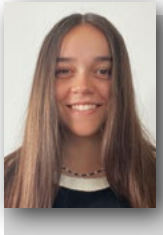
Weitere Neuentwicklungen

Die neuste Entwicklung ist die Unterscheidung zwischen flashterm Backstage und flashterm Stage. flashterm Backstage beinhaltet die bisherigen Funktionen von flashterm, um Terminologie zu erstellen und zu verwalten. flashterm Stage hingegen ist eine neue Funktion zur Terminologierecherche und Textanalyse. Der Zugang erfolgt über den Browser, ohne dass spezielle Zugriffsrechte oder komplexe Voreinstellungen notwendig sind, was die Nutzung für eine breite Zielgruppe, von Fachleuten über Unternehmen bis hin zu Laien, erleichtert. Die neue Funktion ermöglicht es, auf die firmeneigene Datenbank von flashterm zuzugreifen und Benennungen direkt über den Browser zu suchen, ohne einen zusätzlichen Zugang zur Datenbank zu benötigen. Nach Eingabe eines Textes werden die entsprechenden Info-Views aus der flashterm-Datenbank zu den gefundenen Begriffen angezeigt. Zudem haben die Nutzer die Möglichkeit, Texte zur Analyse einzugeben, um zum Beispiel Benennungen zu identifizieren, die in der Datenbank als abgelehnt markiert sind. Die Funktion unterstützt sowohl einsprachige als auch zweisprachige Analysen und Suchen. Besonders im Fokus stehen dabei die Usability und eine übersichtliche Programmstruktur. Für die Entwickler ist die

grafische Gestaltung ein zentraler Punkt, bei dem der Fokus auf der Benutzerfreundlichkeit liegt. Die Datenbank bleibt im Hintergrund und soll auf den ersten Blick nicht erkennbar sein, um den Arbeitsfluss nicht zu stören.

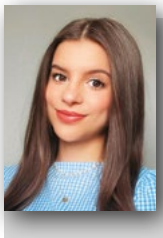
Fazit

flashterm bleibt dank kontinuierlicher Entwicklungen stets aktuell und überzeugt durch seine übersichtliche und intuitive Bedienung. Besonders mit der Einführung von flashterm Stage wird eine bisher nicht abgedeckte Zielgruppe angesprochen, die nicht nur Fachleute, sondern auch Nutzer ohne fundierte Kenntnisse im Bereich der Terminologieverwaltung umfasst. Mit diesen stetigen Fortschritten verspricht die Zukunft weitere interessante und vielversprechende Neuerungen.



Frauke Eschenbeck ist Studentin an der Universität Heidelberg im Masterstudiengang „Translation, Kommunikation und Sprachtechnologie".

Kontaktadresse
frauke.eschenbeck@stud.uni-heidelberg.de



Lola Klan Reyes ist Studentin an der Universität Heidelberg im Masterstudiengang „Translation, Kommunikation und Sprachtechnologie".

Kontaktadresse
lola.klan_reyes@stud.uni-heidelberg.de

Aus Web mach App – oder wie man Websites in Apps umwandelt

Maurice Mayer-Dewor

Das World Wide Web mit seinen unzähligen nützlichen Websites ist nicht mehr aus dem Berufsalltag von Terminologen und Übersetzern wegzudenken. Die Vorteile, die die Nutzung von Websites z. B. für die Terminologierecherche bietet, liegen auf der Hand. Bei der häufigen Nutzung von Websites kann die Effizienz aber auch ein wenig auf der Strecke bleiben – vor allem dann, wenn man die Websites wiederholt aufrufen muss. Sicher, man kann ein Lesezeichen für die Websites setzen oder aber man lässt die Websites einfach im Browser geöffnet. Für die konkrete Nutzung müssen diese aber im Wust der geöffneten Tabs erst einmal wiedergefunden werden. (Hierzu bieten die etablierten Webbrowser mittlerweile tatsächlich auch entsprechende Suchfunktionen an.)

Eine weitere Möglichkeit, häufig genutzte Websites rasch und einfach auffindbar zu machen bzw. etwas stärker in den Vordergrund zu rücken, besteht darin, diese in der Registerkarten-Leiste „anzuheften", wie die Funktion in Firefox heißt, bzw. „anzupinnen" (so in Chrome und Edge). Dadurch werden die entsprechenden Registerkarten stets angezeigt – auch wenn man eine Vielzahl an Registerkarten geöffnet hat, und zwar als erste Registerkarten auf der linken Seite in der Reihe der angezeigten Website-Registerkarten.

Für diejenigen, die bestimmte Websites sehr häufig nutzen, empfiehlt sich eine weitere Variante, nämlich die Website als App zu installieren. Beim Öffnen der App wird der entsprechende Browser im Hintergrund gestartet und die Website wird innerhalb der App angezeigt. Der Inhalt der App ist dann auf die jeweilige Website reduziert und es wird keine Adressleiste angezeigt (s. Abb. 1 auf Seite 32). Das fühlt sich bei der Verwendung praktisch wie eine native Desktop-App an. Die Website-Apps sind denkbar schlank und nehmen im Normalfall nur wenige hundert Kilobyte an Festplattenspeicher ein. Auch wird die Website-App zum Startmenü hinzugefügt und lässt sich problemlos an die Taskleiste anheften oder als Verknüpfung auf dem Desktop ablegen, so dass sie leicht gefunden bzw. gestartet werden kann. Die zuletzt gewählten Einstellungen innerhalb der Website-App

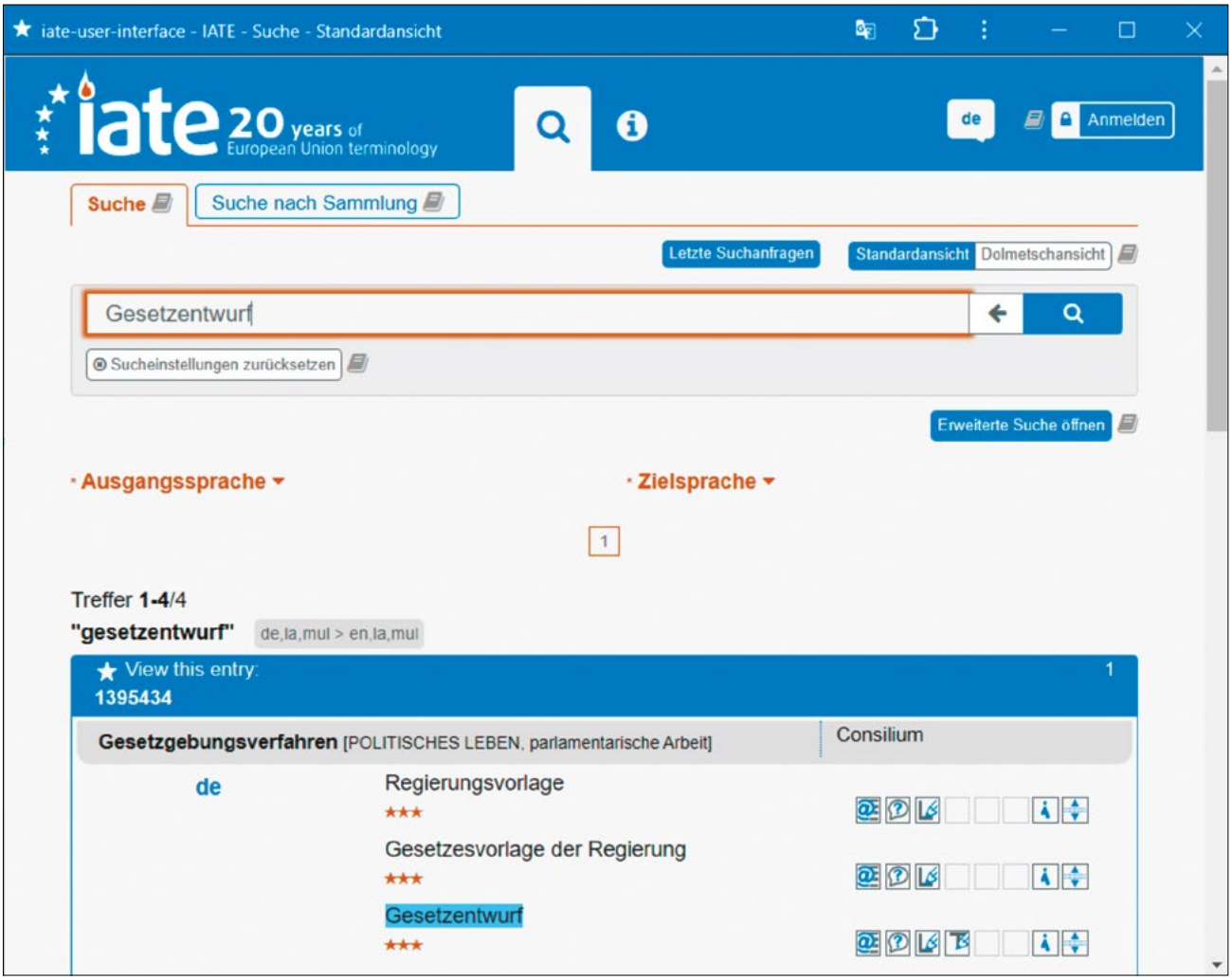


Abb. 1: Die IATE-Website als Website-App – auf das Wesentliche reduziert (Quelle: eigener Screenshot)

Anzeige

Kostenfreie Terminologiebroschüre

Der Deutsche Terminologie-Tag e. V. hat eine Informationsbroschüre mit grundlegenden Argumenten für Terminologiearbeit in Unternehmen und Organisationen herausgegeben. Die dreisprachige Broschüre im handlichen DIN-A6-Format kann kostenfrei und in begrenzter Menge bei der Geschäftsstelle des DTT bestellt werden.

Ihre Ansprechpartnerin:
Olga Buchstaller-Vodopiyanova
geschaeftsstelle@dttev.org

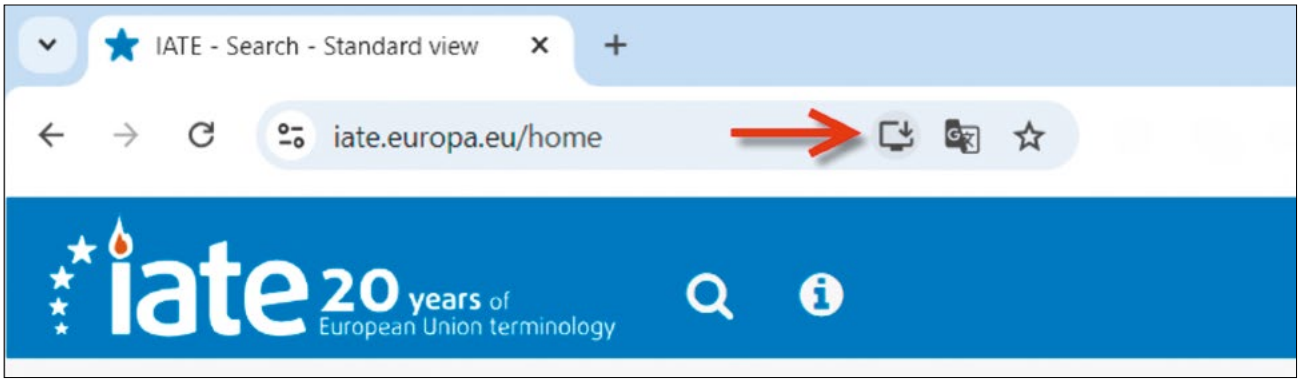


Abb. 2: Installation der Website-App von IATE am Beispiel von Google Chrome (Quelle: eigener Screenshot)

wie z. B. die Sucheinstellungen (Sprachrichtung etc.) sowie die Suchhistorie bleiben so lange gespeichert, bis die Cookies im zugrunde liegenden Browser gelöscht werden.

Die Umwandlung von Websites in Apps wird aktuell nicht von allen Browsern unterstützt. In Windows Edge ist es schon seit einigen Jahren möglich [1] – wie auch in Google Chrome [2]. Mozilla als Entwickler von Firefox hat die Unterstützung von Website-Apps bereits Ende 2020 eingestellt [3], aber auch Firefox-Nutzer sind nicht außen vor (s.u.).

Aber wie wird nun eine Website zu einer Website-App?

In manchen Fällen gestaltet sich die Umwandlung von Websites in eine App denkbar einfach: Denn teilweise wird die Installation der App direkt über ein entsprechendes Icon in der Adresszeile des Browsers angeboten. Dies ist beispielsweise bei IATE der Fall (s. Abb. 2). Es reicht aus, auf das Icon zu klicken, die Installation der App zu bestätigen – und schon ist die App installiert und bereit für die Nutzung. Aber auch wenn die Installation der Website-App nicht unmittelbar in der Adresszeile des Browsers angeboten wird, ist die Vorgehensweise für die Umwandlung denkbar einfach. Sie unterscheidet sich leicht je nach verwendetem Browser:

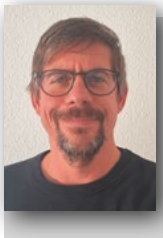
- Chrome: In Chrome öffnet man nach dem Aufruf der gewünschten Website zunächst die Einstellungen von Chrome (über das Icon mit den drei vertikal angeordneten Punkten) und wählt danach die Befehle „Streamen, speichern und teilen“ und abschließend „Seite als App installieren...“.
- Edge: In Edge öffnet man zunächst die Browser-Einstellungen (hier über das Icon mit den drei horizontal angeordneten Punkten) und wählt dann die Befehle „Apps“ und abschließend „Diese Site als eine App installieren“.
- Firefox: Wer Firefox als Standardbrowser nutzt, hat wie gesagt – auf den ersten Blick – leider das Nachsehen, da die Umwandlung von Websites in Website-Apps nicht mehr nativ unterstützt wird. In Firefox ist die Umwandlung einer Website in eine App über den Umweg einer Firefox-Erweiterung namens „Progressive Web Apps for Firefox“ sowie der zusätzlich benötigten Anwendung „PWAsForFirefox MSI installer“ möglich. Natürlich können die Firefox-Nutzer für die Umwandlung einer Website in eine App auch immer in einen anderen unterstützten Browser wechseln.

Nicht mehr benötigte Website-Apps können – wie jede herkömmliche Windows-Anwendung auch – einfach über die Systemsteuerung bzw. über die Einstellungen von Windows wieder deinstalliert werden.

Jetzt bleibt nur noch zu sagen: Viel Spaß beim Appen!

Quellen:

[1] https://en.wikipedia.org/wiki/Progressive_web_app#Initial_introduction [26.09.2024]
[2] https://www.chip.de/news/Neue-Google-Funktion-Wie-Sie-eine-Webseite-in-eine-App-verwandeln-koennen_185181438.html [26.09.2024]
[3] https://entwickler.de/javascript/firefox-beendet-pwa-support-fur-den-desktop [26.09.2024]
[4] https://addons.mozilla.org/de/firefox/addon/pwas-for-firefox/ [26.09.2024]



Maurice Mayer-Dewor ist Diplom-Übersetzer für die Sprachen Italienisch und Französisch und arbeitet als Dozent am Institut für Übersetzen und Dolmetschen (IÜD) der Universität Heidelberg. Dort lehrt er u. a. im Bereich der Übersetzungstechnologien. Darüber hinaus arbeitet er als freiberuflicher Übersetzer und Technischer Redakteur.

Kontaktadresse
maurice.mayer@iued.uni-heidelberg.de
www.iued.uni-heidelberg.de

Typing Assistant: Terminologie als Treibstoff für Predictive Typing

Armin Mutscheller

Texte schnell und fehlerfrei tippen: Das kann heutzutage kaum noch jemand aus dem Effeff. Übersetzer und andere Vielschreiber schätzen daher Arbeitshilfen wie Textbausteinverwaltungen oder Diktierprogramme à la Dragon Naturally Speaking. Doch nicht immer passen solche Tools zur Aufgabenstellung, zum Computersystem oder zur Eingabesprache. Manche sind auch aus ergonomischer Sicht nicht immer die beste Wahl. Zum Glück gibt es eine Alternative: Predictive Typing. Diese Technologie ist bei Smartphones schon lange gang und gäbe: Das System merkt sich Tastatureingaben, wertet ihre Häufigkeit statistisch aus und schlägt bei jeder neuen Eingabe ein oder mehrere Wörter vor, um den Satz zu vervollständigen. Auch die Anbieter von Translation-Memory-Systemen stellen solche Funktionen für ihre Produkte – und zwar nur für diese – zur Verfügung. Wer die Vorzüge einer Eingabeprognose auch in anderen Anwendungen nutzen möchte, ist auf zusätzliche Software angewiesen. Einen besonders bemerkenswerten Vertreter dieser Spezies möchte ich hier vorstellen. Er heißt „Typing Assistant“ und präsentiert sich als eine Art Schweizer Offiziersmesser für das Schreiben von Texten jeglicher Art.

Der Blick fürs Ganze

Auf iPhones war das „Recycling“ von Wörtern und Wortfolgen für schnellere Texteingaben von Anfang an eine hochgeschätzte Funktion vieler Apps. Daher fand Predictive Typing auch ziemlich schnell seinen Weg in die Android-Welt und ist seitdem von Smartphones nicht mehr wegzudenken. Auf Desktop-Computern war diese Schreibhilfe allerdings nie ein großes Thema, dabei wird der meiste Content nach wie vor am PC erstellt. Inzwischen haben Entwickler erkannt, wie wichtig schnelle und möglichst fehlerfreie Schreibergebnisse für Textschaffende sind, ob Blogger, Redakteure, Journalisten oder Übersetzer. In der Windows-Welt dürfte das hier vorgestellte Tool Typing Assistant inzwischen die Nase vorn haben. Und das liegt vor allem an der Flexibilität und Funktionsvielfalt, mit der es selbst hoffnungslose Zwei-Finger-Tipper wie mich zu rasanten Schnellschreibern macht. Man fragt sich, warum ein Tool mit diesen Fähigkeiten nicht schon längst zum Funktionsumfang von Windows gehört.

Systemübergreifend und sicher

Anders als viele Predictive-Typing-Apps für Smartphones speichert Typing Assistant seine Schreibvorschläge nicht in der Cloud des Herstellers, sondern auf dem lokalen System. Das ist unter dem Aspekt des Datenschutzes ein enormer Vorteil. Als Speicherort für das Programm und seine Daten kann auch ein USB-Stick oder ein Ordner in einer privaten Cloud verwendet werden. Für das letztgenannte Szenario empfiehlt sich die portable USB-Version der Software: Damit kann von verschiedenen Computern aus auf ein- und dieselben Wörterbuch-Datenbanken zugegriffen werden.

Typing Assistant ist zu Windows 10 und 11 kompatibel und funktioniert dort in nahezu allen Programmen. Nach Auskunft des Anbieters (Stand: Juni 2024) wird das Programm aktiv weiterentwickelt. Seine weitreichende Kompatibilität macht das Tool für Übersetzer besonders interessant, denn auf die nur eingeschränkt konfigurierbaren Predictive-Typing-Funktionen der verschiedenen Translation-Memory-Tools, die nur im jeweiligen CAT-Ökosystem funktionieren, ist man dann nicht mehr angewiesen und kann anwendungsübergreifend aus dem Vollen schöpfen. Die Entwickler haben daher Übersetzer als Zielgruppe fest im Blick und führen beispielsweise memoQ und Trados Studio ausdrücklich als kompatible Anwendungen auf. Für kleinere Probleme in diesen CAT-Umgebungen, die mit der intensiven Nutzung der Windows-Zwischenablage zu tun haben, empfehlen die Entwickler gut funktionierende Workarounds. Unterm Strich fühlt sich das Schreiben mit Typing Assistant sehr homogen an und man sprengt damit eine der Ketten, die die Hersteller von CAT-Tools ihren Kunden gerne anlegen, wenn es um die Einbindung von Zusatzanwendungen geht.

Typing Assistant bündelt auf Wunsch gleich mehrere Wortquellen zu einer großen Vorschlagsdatenbank. Wer alle diese Quellen aktiviert hat, kann zusehen, wie Typing Assistant bei der Texteingabe einen Turbo nach dem anderen zündet.

Turbo 1: Wörterbücher für über 50 Sprachen

Hauptwörterbücher gibt es derzeit für über 50 Sprachen bzw. Varietäten. Diese Wörterbücher enthalten auch sämtliche Flexionen, das ergibt für Deutsch rund 1,7 Millionen

Einträge. Wem das zu viel des Guten ist, dem stehen zumindest für Deutsch zwei reduzierte Wörterbücher mit deutlich begrenztem Wortschatz zur Verfügung. Welche Kriterien bei dieser Ausdünnung angewandt wurden, bleibt leider im Dunkeln. Hat man sich erst einmal für ein Sprachprofil und das dazugehörige Hauptwörterbuch entschieden, macht sich Typing Assistant sofort ans Werk. Kaum hat man die ersten paar Buchstaben eingegeben, prüft das Tool auch schon, ob eines der aktiven Wörterbücher passende Schreibvorschläge enthält, und blendet diese in einem Vorschlagsfenster ein. Typing Assistant verwendet dabei verschiedene Symbole, um die Herkunft der im Vorschlagsfenster angezeigten Vorschläge zu kennzeichnen (s. Abbildung 1).

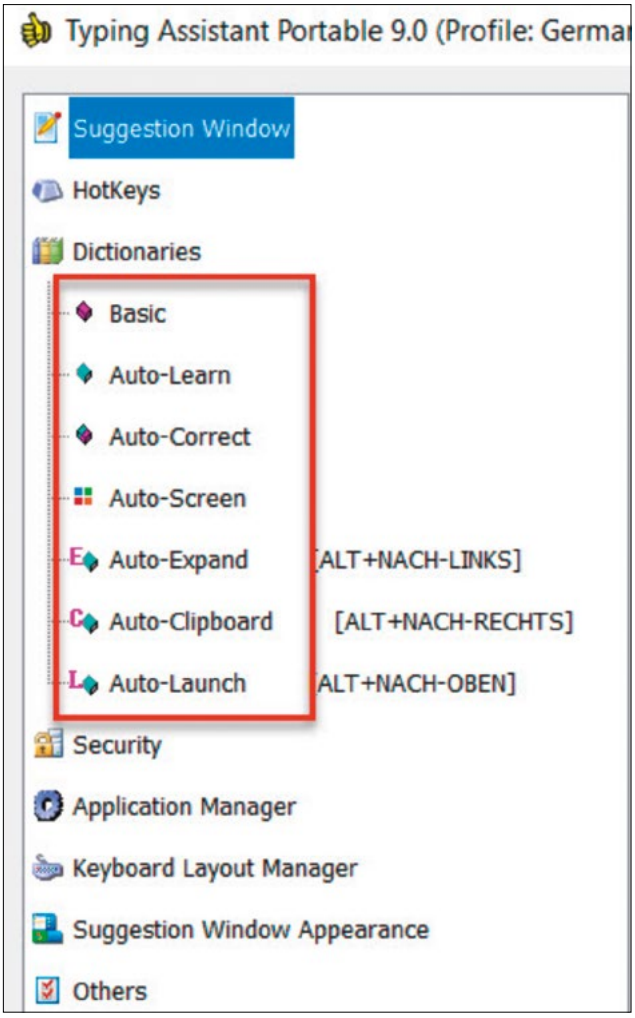


Abb. 1: Übersicht im Einstellungsmenü für Herkunft eines Vorschlags

Der gewünschte Eintrag kann durch Eintippen der vorangestellten Vorschlagsnummer in den Text übernommen werden. Die Taste oder Tastenkombination zum Auswählen eines Eintrags kann frei definiert werden, ebenso das Aussehen und die Position des Vorschlagsfensters. Auch die Zahl der Buchstaben, die man eingeben muss, damit die Auswahlliste erscheint, ist konfigurierbar. Dies gilt übrigens auch für die Dauer der Einblendung. Überhaupt sind die gesamte Benutzeroberfläche und das Verhalten des Programms bis ins kleinste Detail einstellbar.

Turbo 2: Eingabeverlauf und statistische Sortierung

Typing Assistant nutzt für Schreibvorschläge nicht nur das aktive Hauptwörterbuch, es lernt auch ständig dazu. So speichert es automatisch alle Benutzereingaben und bietet diese bei der nächsten Eingabe als Schreibvorschläge an. Mehr noch: Es führt im Hintergrund eine Statistik und stellt die am häufigsten verwendeten Wörter im Vorschlagsfenster an oberste Stelle. Wer häufig mit Urkundenübersetzungen oder schwierigen Fachtexten zu tun hat, wird das schnell als hilfreich empfinden. Repetitive Wortungetüme wie „Desoxyribonukleinsäure-Analyseverfahren“ oder lange Orts-/Eigennamen wie „Schmedeswurtherwesterdeich“ müssen nur ein einziges Mal fehlerfrei eingegeben werden. Fortan genügt es, sie durch Eingabe der Anfangsbuchstaben anzudeuten, um sie dann wiederum fehlerfrei in den Text einfügen zu können.

Turbo 3: Vom Anwender aktiv hinzugefügte Wörter

Wem die sehr umfangreichen Hauptwörterbücher und die oben beschriebene, passive Wortschatzerweiterung nicht genügen, hat einige Möglichkeiten, um noch mehr Schreibvorschläge im System zu hinterlegen. Eine davon ist besonders simpel und effektiv: Man markiert im Dokument ein Wort oder eine Wortfolge, drückt einen frei definierbaren Hotkey und schon ist der Textschnipsel in einem separaten, benutzerdefinierten Wörterbuch gespeichert. Man ahnt es bereits: Typing Assistant lässt sich auf diese Weise auch als Textbausteinverwaltung nutzen. Grußformeln, Anreden, Abwesenheitsnotizen, chemische Substanznamen, pharmazeutische Wirkstoffe, Länder-, Orts- und Familiennamen, Beglaubigungsvermerke und so weiter und so fort: All das muss man nur noch ein einziges Mal eintippen, danach übernimmt Typing Assistant die Schreibarbeit.

Wer fleißig Terminologie pflegt, kann daraus einen doppelten Nutzen ziehen: Korrekte Fachterminologie sorgt in Dokumenten und Übersetzungen bekanntlich für hohe Konsistenz und Qualität. Und die Texterstellung geht schneller von der Hand, wenn validierte Benennungen als Schreibvorschläge genutzt werden können. Denn neben einer Adhoc-Befüllung des aktiven Wörterbuchs kennt Typing Assistant noch eine weitere, sehr effektive Methode, um zum Beispiel Fach- oder Produktterminologie einzubinden: den Import von Wortlisten in bestehende oder neue Wörterbücher. Dazu eignen sich Exporte aus Terminologie-Datenbanken ganz hervorragend (siehe Abschnitt „Terminologie einbinden“ auf Seite 37). Man kann Typing Assistant auch mitteilen, ab welcher Wortlänge es Benennungen aufnehmen

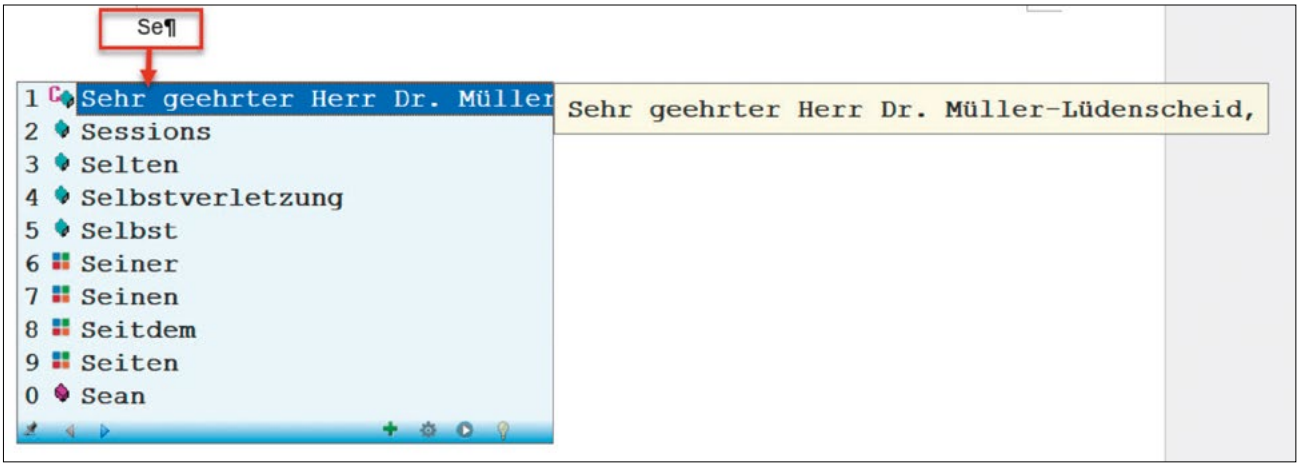


Abb. 2: Vorschlagsfenster für die Zeichenkette „Se“

soll. Das dürfen gerne auch ganze Fügungen sein, die später beim Schreiben ähnlich wie Textbausteine automatisch vorgeschlagen werden. In Abbildung 2 wurden zwei Zeichen („Se“) eingegeben, um eine häufig verwendete Anredeformel aus dem Clipboard-Wörterbuch abzurufen. Schnellere Tipper sollten eine größere Zeichenanzahl zum Triggern des Vorschlagsfensters einstellen, z. B. 5 oder 6 Zeichen, um nur längere Textvorschläge zu erhalten.

Turbo 4: Texteingabe durch Kürzel

Seine Qualitäten als verkappte Textbausteinverwaltung zeigt das Tool auch mit der Funktion „Auto-Expand“: Hierbei definiert man eine Abkürzung für eine häufig benötigte Floskel oder Formulierung. Mit der Abkürzung mfg lässt sich beispielsweise die Schlussformel „Mit freundlichen Grüßen“ blitzschnell in eine E-Mail einfügen. Das ähnelt der bekannten „Autokorrektur“-Funktion von Word & Co., hat jedoch mit dem Korrigieren von Tippfehlern erstmal nichts zu tun. Zudem ist Typing Assistant hier deutlich flexibler, denn ein und dasselbe Kürzel kann auch für mehrere Langformen verwendet werden. So kann man beispielsweise die am häufigsten verwendeten E-Mail-Adressen in das Auto-Expand-Wörterbuch eintragen und mit dem Tastenkürzel @ verknüpfen. Sobald man das @ eingibt, wird die Vorschlagsliste angezeigt, und schon kann man die gewünschte Adresse in das Empfängerfeld übernehmen. Über Kürzel lassen sich übrigens auch Anwendungen starten oder Dokumente in ihrer jeweiligen Standardanwendung öffnen – dazu verwendet Typing Assistant ein spezielles „Wörterbuch“ namens Auto-Launch.

Turbo 5: Tippfehler elegant korrigieren

Die oben erwähnte, „klassische“ Autokorrektur-Funktion gibt es in Typing Assistant ebenfalls. Für die Sprachen Polnisch, Französisch und Englisch bietet das Tool spezielle Wörterbücher mit häufigen Tippfehlern an, die bei der Texteingabe fortan automatisch korrigiert werden. Solche Wörterbücher lassen sich auch für andere Sprachen leicht anlegen und pflegen. Wer sich also partout nicht merken kann, ob es „Pallette“ oder „Palette“, „vorrübergehend“ oder „vorübergehend“ heißt, kann aufatmen – ein gewisses Maß an Fleiß für die Wörterbuchpflege vorausgesetzt.

Turbo 6: Vorschläge aus der Zwischenablage

Die Möglichkeiten von Typing Assistant sind damit aber noch lange nicht ausgeschöpft. Wer möchte, kann sich alle Textschnipsel, die über den Hotkey STRG+C (oder einen anderen, frei wählbaren Hotkey) jemals in die Windows-Zwischenablage kopiert wurden, als Schreibvorschläge anzeigen lassen. Dabei respektiert das Tool auch Absatzmarken. Und das heißt konkret: Man kann auch ganze Tabellenspalten oder Zellenblöcke als Schreibvorschläge definieren und später wiederverwenden.

Turbo 7: Optische Zeichenerkennung im aktiven Bildschirm

Und jetzt kommen wir zu der wohl schärfsten Klinge, die das Tool zu bieten hat: Typing Assistant kann nämlich auf Wunsch eine optische Zeichenerkennung (*optical character recognition*, OCR) auf den gesamten aktiven Bildschirm oder auf einen definierbaren Bildausschnitt anwenden. Und plötzlich stehen auch Fehlermeldungen, Menübefehle oder der gesamte sichtbare Kontext eines Translation-Memory-Editors in einem gesonderten OCR-Wörterbuch als Schreibvorschläge zur Verfügung. Und zwar unabhängig davon, ob der Text im aktiven Segment, in einem bereits übersetzten Segment oder irgendwo in einer Trefferliste steht oder stand. Die Erkennungsgenauigkeit ist dabei außerordentlich hoch, und die OCR-Funktion funktioniert praktisch verzögerungsfrei. Diese Funktion ist sehr hilfreich, um zum Beispiel bei einem PC-Problem den Text und/oder Codezeilen aus einer Fehlermeldung an den Support senden zu können, wenn dies per Screenshot nicht möglich ist.

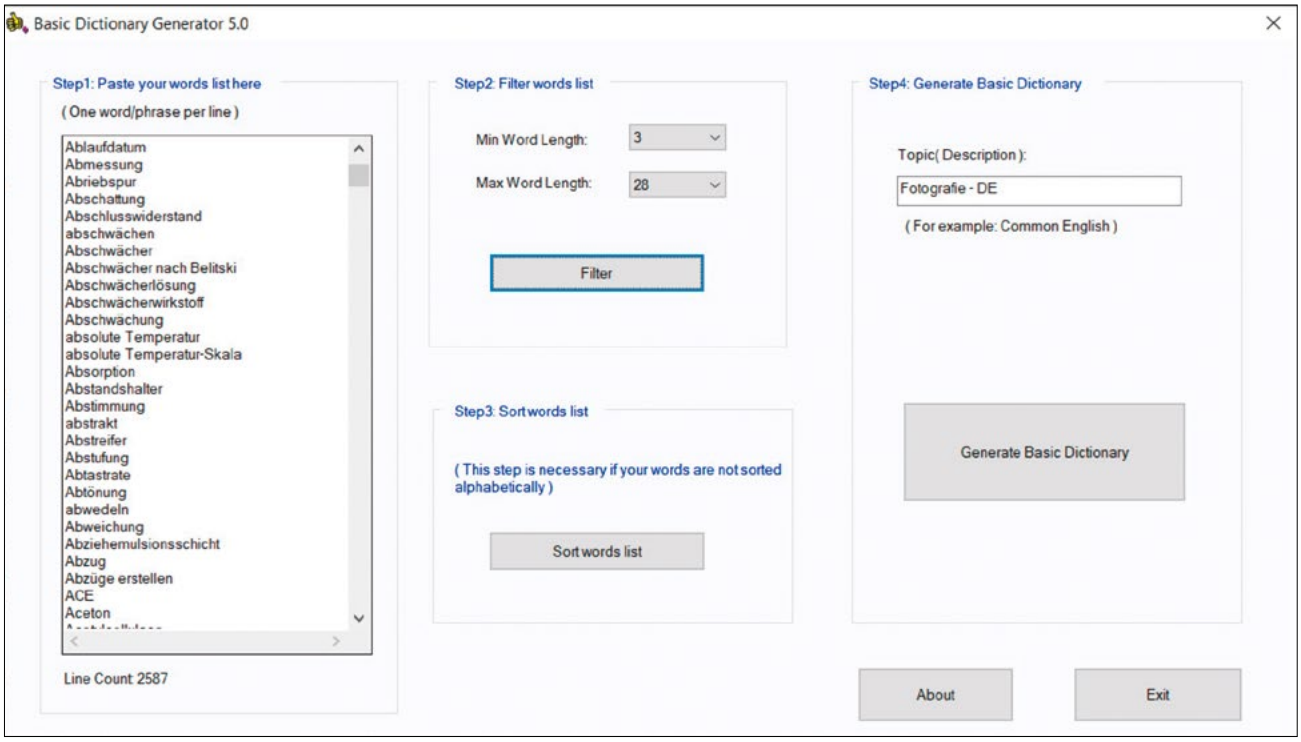


Abb. 3: Einfach und gut: Die Benutzeroberfläche des Wörterbuch-Generators ist selbsterklärend.

Terminologie einbinden

Typing Assistant kennt zwei Wege, um aus Terminologie-Datenbanken exportierte Wortlisten für die Tipparbeit nutzbar zu machen: Durch einfaches Einfügen einer Wortliste mit bis zu 10.000 Einträgen in ein *AutoLearn*-Wörterbuch (per Copy and Paste) oder durch Importieren eines selbst generierten Wörterbuchs im Format BSC. Dieses BSC-Wörterbuch kann dann als Basis-Wörterbuch (*Basic Dictionary*) verwendet werden. Diese Methode hat den Vorteil, dass weit über eine Million Einträge importiert werden können. Für beide Methoden empfiehlt es sich, ein neues Profil in Typing Assistant anzulegen. Ein kostenloses Konvertierungstool, das aus einsprachigen Wortlisten BSC-Dateien erzeugt, findet sich auf der Homepage von Typing Assistant.

Im ersten Schritt exportiert man die benötigte Wortliste aus der Terminologie-Datenbank. Als Ergebnis benötigt man eine simple, einsprachige Liste. Die Möglichkeit, eine solche Liste zu exportieren, bieten eigentlich alle gängigen Terminologie-Management-Systeme. Im Idealfall wird die

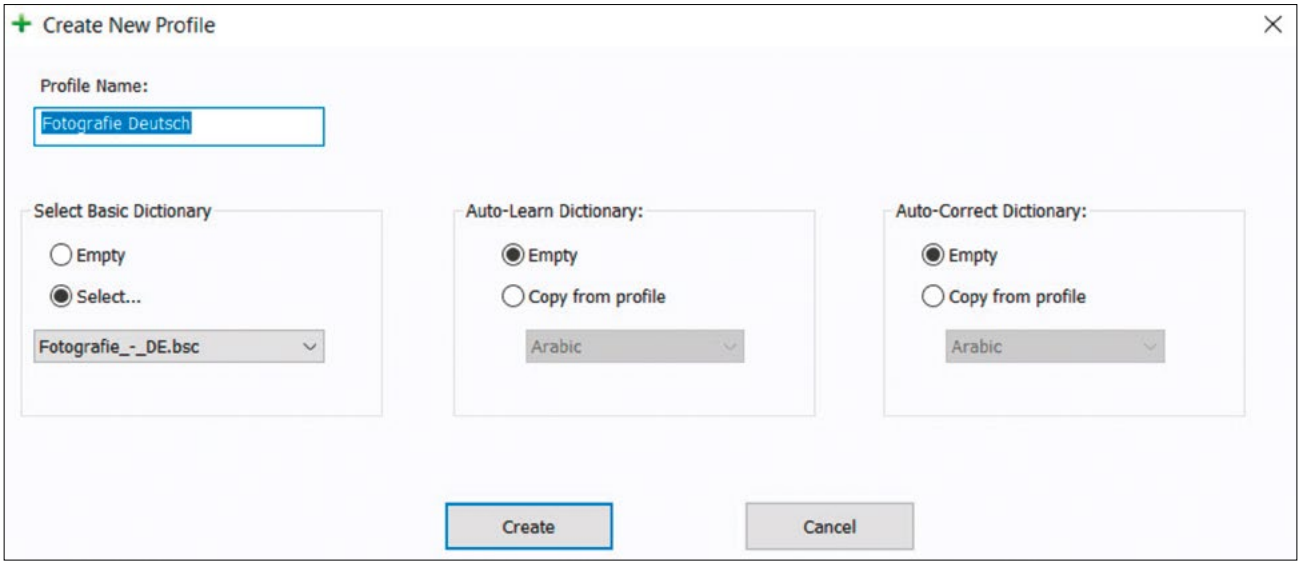


Abb. 4: Beim Anlegen eines neuen Profils (hier: „Fotografie Deutsch“) bindet man die Wortliste als Basis-Wörterbuch ein.

Wortliste in das Format XLSX (Microsoft Excel) konvertiert, denn dann kann man die entsprechende Spalte anhand der Filter- und Sortierfunktionen von Excel vor dem Import bearbeiten und beispielsweise sehr kurze oder duplizierte Einträge löschen. Formatierungsmerkmale wie Fett- oder Kursivschrift gehen beim späteren Import in Typing Assistant verloren.

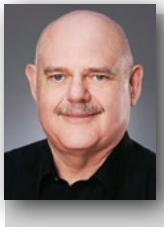
Hat man die Wortliste erstellt, überträgt man sie einfach per Copy and Paste in das Konvertierungstool von Typing Assistant und erzeugt daraus die BSC-Datei. Diese kopiert man im Installationsordner von Typing Assistant in das Verzeichnis „Basic“. Danach lässt es sich bei der Erstellung eines neuen Profils als Wörterbuch einbinden und liefert fortan Schreibvorschläge.

Wo liegen die Tücken?

Man muss ziemlich tief bohren, um kleinere Nachteile zu finden. Bisweilen kann es vorkommen, dass das Tool mit anderen Programmen kollidiert, die sehr dynamisch und intensiv auf die Windows-Zwischenablage zugreifen. Ein Beispiel hierfür ist das Automatisierungstool AutoHotkey (AHK). Wer sich damit das eine oder andere Helferlein für die tägliche Arbeit programmiert hat, wird in Verbindung mit Typing Assistant schon bald auf typische Probleme dieser (eher seltenen) Konstellation stoßen. Zum Glück lässt sich jedoch Typing Assistant per Hotkey temporär deaktivieren, was manchmal sinnvoll ist, wenn gleichzeitig AHK-Skripte ihren Dienst verrichten. Zudem ist es möglich, problematische Anwendungen aus Typing Assistant auszuschließen oder nur diejenigen Anwendungen freizugeben, die zu 100 % kompatibel sind. Problemchen dieser Art dürften aber nur Bastler und Coder betreffen, denen das Lösen und Umgehen von Softwareproblemen ohnehin im Blut liegt.

Was kostet das Tool?

Im Online-Shop des Anbieters wird eine Typing-Assistant-Lizenz (gültig für bis zu drei Computer) für 149,00 USD angeboten. Das klingt erstmal nach einem Batzen Geld, doch nach meiner Erfahrung ist jeder einzelne Cent davon gut angelegt. Wer zunächst nur ein wenig experimentieren möchte, kann das 45 Tage lang kostenlos tun. Man wird auch nicht – wie bei vielen anderen Tools – mit schöner Regelmäßigkeit zu kostenpflichtigen Lizenzverlängerungen oder Upgrades gedrängt, sondern kann das investierte Geld jahrelang für sich arbeiten lassen. Zudem ist das Tool sehr gut dokumentiert, wenn auch in teilweise etwas holprigem Englisch. Ich selbst genieße es bei meiner täglichen Textarbeit immer wieder aufs Neue.



Armin Mutscheller ist Diplom-Übersetzer und arbeitet als freiberuflicher Übersetzer, Lektor und Terminologe, hauptsächlich im Sprachpaar Englisch/Deutsch und in den Fachgebieten Medizin, IT und Industrietechnik. Er berät Unternehmen zu methodischen und linguistischen Aspekten der Produktkommunikation und hält am Institut für Übersetzen und Dolmetschen der Universität Heidelberg Lehrveranstaltungen ab.

Kontaktadresse
info@mutscheller.de
www.mutscheller.de

Anzeige

DTT-Terminologiezertifikat

Kennen Sie schon das **DTT-Terminologiezertifikat**?
Für den Erhalt des Zertifikats muss an einem DTT-Symposion, einem DTT-Grundlagenseminar, einem DTT-Vertiefungsseminar sowie an zwei DTT-Webinaren teilgenommen werden.

Ausführliche Informationen unter:
dttev.org/fortbildung/terminologiezertifikat

ISO 704 Terminology work – Principles and methods: Immer noch wichtig oder immer wichtiger?

Barbara Karsch

Jeder kommuniziert, jeder schreibt und jeder kann in Zukunft den KI-Knopf drücken und Inhalte automatisch generieren. Es ist davon auszugehen, dass Fehlkommunikation zunehmen wird. Das macht die Norm ISO 704 aktueller denn je. In diesem Artikel stellen wir diese Norm für diejenigen, die sie nicht ohnehin schon gut kennen, kurz vor. Dann konzentrieren wir uns auf den Unterschied zu den entsprechenden nationalen Normen in der jeweiligen Behandlung von Bezeichnungen. Und schließlich erläutern wir, warum die Inhalte dieser Norm mit zunehmendem Einfluss der künstlichen Intelligenz noch an Bedeutung gewinnen werden.

Überblick

ISO 704 Terminology work – Principles and methods [1] ist Pflichtlektüre für Terminologen, denen keine entsprechende nationale Norm des Dokuments zur Verfügung steht. Im deutschen Sprachraum bieten sich DIN 2330:2022 Terminologiearbeit – Grundsätze und Methoden [2] bzw. ÖNORM A 2704:2015 Terminologiearbeit – Grundsätze und Methoden [3] an, wobei anzumerken ist, dass die Inhalte trotz der Titel nicht 100 % deckungsgleich sind. Neben dem englischen Original von ISO 704 gibt es auch eine Version in französischer Sprache mit dem Titel ISO 704 Travail terminologique – Principes et méthodes. Die derzeitige Version ist 2022 erschienen und ersetzt die Vorgängerversion von 2009.

Ursprünglich war es Ziel des Dokuments, ein Fundament für die Terminologiearbeit in der Normenarbeit zu schaffen. Daher steht es allen, die an ISO-Normen arbeiten, kostenlos zur Verfügung. Heute dürfte der Anteil der Terminologen, die die Norm im terminologiewissenschaftlichen Unterricht bzw. in der Praxis in Unternehmen und Organisationen einsetzen, mindestens so groß sein wie der der Normungsexperten. Sie vermittelt uns die Grundlagen für die Erstellung eines korrekten terminologischen Eintrags in einer Termbank.

Die Norm umfasst insgesamt über 80 Seiten. Wie üblich in einer ISO-Norm werden nach den Abschnitten Anwendungsbereich („Scope“) und Normative Verweisungen („Normative References“) im Abschnitt 3 („Terms and Definitions“) die wichtigsten Begriffe definiert. Diese sind über die Online Browsing Platform [4] auch kostenlos abrufbar. Dort findet sich auch eine kostenfreie Version von ISO 1087:2019 Terminology work and terminology science, in der die für ISO 704 notwendige Terminologie festgelegt wurde.

Im Abschnitt Realität und Sprache („Reality and Language“) wird zunächst das Thema „Gegenstand“ (object) behandelt. Die für uns sichtbaren Gegenstände sind leicht vorstellbar. Komplizierter wird es, wenn es um einen (noch) nicht existierenden Gegenstand, also zum Beispiel um einen in der Entwicklung befindlichen Prototypen oder um einen abstrakten Gegenstand, z. B. eine Theorie, geht. Auch darüber muss Kommunikation möglich sein. Daher sind gedankliche Konstrukte ebenfalls Gegenstände im Sinne der Norm. Das Dokument erläutert den für die Kommunikation so wichtigen Zusammenhang zwischen Gegenständen in der wirklichen oder gedanklichen Welt, Begriffen und den sprachlichen Mitteln, die wir zur Kommunikation darüber verwenden. Diese Grundsätze gelten für die Terminologiearbeit in verschiedenen Arbeitsbereichen, z. B. in Wissenschaft, Technik, Industrie und Verwaltung. Erst durch das Verständnis, wie durch Abstraktion aus Gegenständen ein Begriff wird und wie diese Begriffe dann mit wohl gebildeten Bezeichnungen belegt werden, wird präzise Kommunikation zuverlässig möglich.

Es folgt ein umfassender Abschnitt zu Begriffen („Concepts“), in dem der Unterschied zwischen Allgemein- und Individualbegriff erläutert wird sowie Merkmale, Begriffsrelationen und -systeme beschrieben werden. Die Begriffsbildung wird durch ein Diagramm veranschaulicht.

Auch das Thema Definitionen wird auf 15 Seiten im Abschnitt „Definitions“ relativ ausführlich behandelt. Als Terminologen arbeiten wir uns häufig ausgehend von einer bereits bestehenden Bezeichnung (engl. designation) zurück zum Gegenstand mit seinen Eigenschaften hin zum Begriff mit seinem Oberbegriff und den Merkmalen. Erst nachdem diese terminologische Analyse – oft in Minutenschnelle, manchmal mit fachlicher Unterstützung – erfolgt

ist, können wir unser Verständnis in der Definition explizit machen. Wie wichtig das ist, dringt in öffentlichen Diskussionen mit Informationsarchitekten immer häufiger durch. Michael Iantosca von Avalara Inc. bezeichnet Terminologie beispielsweise im Rahmen seines Semantic Content Maturity Model als die „Quelle der Wahrheit" [5]. Ein hoher Prozentsatz der in einer Datenbank dokumentierten Begriffe wird erst durch eine Inhaltsdefinition eindeutig; Definitionen sind folglich ein wichtiger Schritt in Richtung Wahrheitsfindung.

Bezeichnungen – kein einfaches Thema

Zu einem angestrebten vollständigen und richtigen Eintrag fehlen noch einige terminologische Daten, zum Beispiel die Bezeichnungen, um die es im letzten großen Abschnitt („Designations") von ISO 704 geht. Hier hat sich in den letzten Jahrzehnten viel getan. Während sich unsere Vorgänger fast ausschließlich um Benennungen, also die Bezeichnungen von Allgemeinbegriffen, kümmerten, ist die Firmensprache heute gespickt mit Bezeichnungen wie Apple® iPhone®, VW T-Roc® oder Windows®. In der Praxis ergeben sich dann Fragen, wie

- Sollen diese Bezeichnungen dokumentiert werden?
- Wenn ja, sind sie Benennungen oder Namen, und wie erstelle ich dafür einen richtigen und vollständigen Eintrag?

Für diese Überlegungen lohnt es sich, die Abschnitte zu Bezeichnungen in der ISO 704 oder der ÖNORM A 2704 zu betrachten. In DIN 2330 fehlt ein entsprechender Teil, das Thema wird dafür kurz in DIN 2342:2022-07 behandelt.

Der Oberbegriff und die nächste Stufe im Begriffssystem sind in den genannten Dokumenten jeweils fast gleich und könnten zunächst wie in Abb. 1 dargestellt werden.

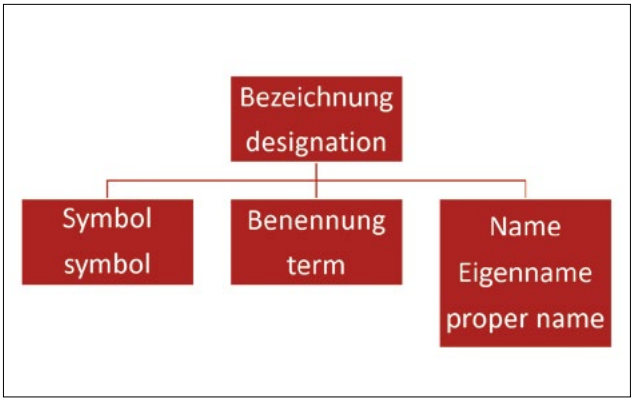


Abb. 1: Begriffssystem „Bezeichnung"[1] (Quelle: eigene Darstellung nach ISO 704, DIN 2342:2022-07 und ÖNORM A 2704)

Auf die Frage, ob beispielsweise Apple iPhone eine Benennung ist, da ja Millionen solcher Gegenstände existieren, oder ob es sich dabei eher um einen Namen handelt, gibt es mehrere Antworten.

Im Anhang C von ISO 704 bzw. im entsprechenden terminologischen Eintrag von ISO 1087 gibt es den Begriff der Benennung für „eine Gruppe von Gegenständen, deren relevante Eigenschaften identisch sind" [eigene Übersetzung nach [6]]. Apple iPhone wäre demnach also eine appellation. Das Begriffssystem in Abb. 1 wird wie in Abb. 2 erweitert.

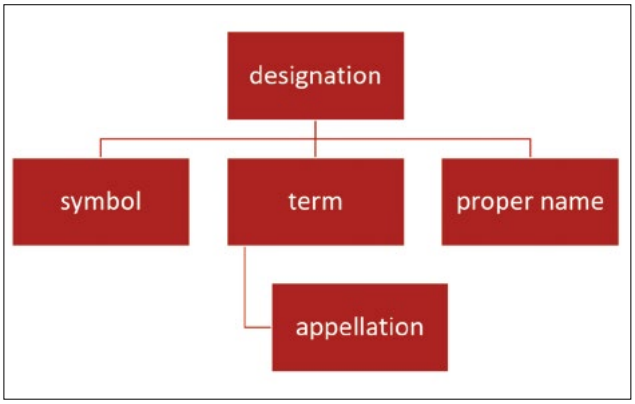


Abb. 2: Begriffssystem „Bezeichnung", in der ISO 704 erweitert um die sogenannte „appellation" (Quelle: eigene Darstellung)

In DIN 2342 wird für Benennung bzw. Eigenname nur jeweils einer der wichtigsten Unterbegriffe genannt. Eine Produktbezeichnung ist eine „Benennung, die von einer Organisation für ein spezielles Produkt festgelegt wird", und Markenname wird als „Eigenname, der von einer Organisation festgelegt und offiziell registriert oder zur Registrierung angemeldet wurde" definiert [7]. Apple iPhone ist gemäß DIN 2342 eine Produktbezeichnung. Grafisch sieht das wie in Abb. 3 aus.

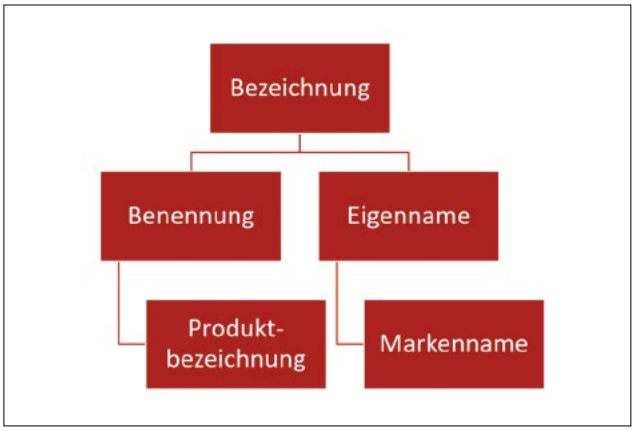


Abb. 3: Begriffssystem „Bezeichnung" nach DIN 2342 (Quelle: eigene Darstellung)

[1] „Name" deckt in der ÖNORM zwei Begriff ab; hier ist die „sprachliche Bezeichnung eines Individualbegriffs" gemeint.

In der ÖNORM wird die Thematik sehr viel umfassender behandelt. Der grundlegende Unterschied zu ISO 704 und DIN 2342 ist, dass der Begriff Name sowohl Bezeichnungen eines einzelnen Gegenstandes als auch Bezeichnungen mehrerer identischer Gegenstände umfasst [3]. Hier wäre Apple iPhone ein Sammelname, der sich auf mehrere Gegenstände, die in sämtlichen Eigenschaften übereinstimmen, bezieht [3].

Ein in englischer Sprache von den ÖNORM-Projektleitern verfasster Artikel [8] bietet ein umfangreiches Begriffssystem zu Namen.

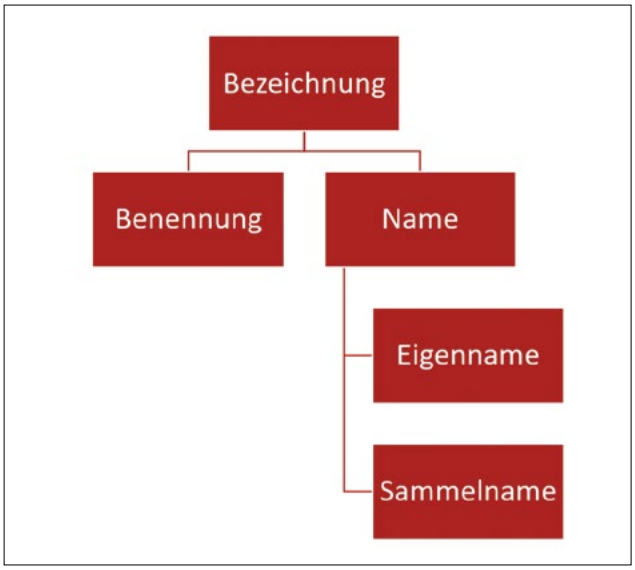


Abb. 4: Begriffssystem „Bezeichnung" nach ÖNORM A 2704 (Quelle: eigene Darstellung)

In der Praxis spielt dies insofern eine Rolle, dass zunächst eine Entscheidung getroffen werden muss: Ist der VW T-Roc nun ein Modell im Gegensatz zum Tiguan oder Golf und daher ein Markenname nach DIN, ein Eigenname nach der ÖNORM bzw. ein proper name nach ISO 704 und soll als solches dokumentiert werden? Oder geht es um tausende von VW T-Roc, die irgendwo vom Fließband laufen, womit es sich dann um eine Produktbezeichnung, einen Sammelnamen bzw. eine appellation handeln würde? In der übersetzungsorientierten Terminologiearbeit würde es meistens wenig Sinn ergeben, beide Begriffe zu dokumentieren. Wenn es jedoch Ziel ist, ein Begriffssystem explizit zu machen, könnten zwei Einträge notwendig sein. Wichtig wäre mindestens, dass eine Lösung systematisch auf ähnliche Bezeichnungen bzw. auf verwandte Begriffe angewandt wird. Dazu ist ein gutes Verständnis der in ISO 704 ausgeführten Grundlagen unerlässlich.

ISO 704: Content-Erstellung und KI

Stellen wir uns vor, dass in Zukunft immer mehr Content nicht mehr von Menschen mit diesem Wissen, sondern von Maschinen aus bereits bestehenden Texten erstellt wird. Dann wird schnell klar, dass ISO 704 enorm an Bedeutung gewinnen wird. Gehen wir zunächst jedoch zurück in die Vergangenheit und überlegen, wer früher Content erstellt hat.

Vor der Benennung steht die Erfindung. Grob gesagt gab es vor dem 19. Jahrhundert wenige Erfinder, die selbst über ihre Erfindungen schrieben. Ihre Texte waren eventuell nur für sie selbst verfasst, dürften aber in den meisten Fällen von Präzision geprägt gewesen sein, stammten sie doch direkt aus der Quelle. Mit zunehmender Innovationsgeschwindigkeit unter Beteiligung von mehr Menschen, die nun auch in Teams arbeiteten, entstand die Notwendigkeit, Produkte für Gewinn zu verkaufen. Daher konnte nicht mehr erwartet werden, dass Erfinder die Dokumentation übernahmen. Zudem sollten die Dokumente in vielen Sprachen und barrierefrei erstellt werden. Und schließlich stieg die Zahl sowohl der neuen Begriffe als auch der unterschiedlichen Leser derart an, dass sogar professionelle technische Redakteure und Übersetzer nicht mehr mithalten konnten. Sie wurden mehr und mehr von Maschinen unterstützt, die in Zukunft einen Teil der Arbeit übernehmen sollten.

Während dieser Evolution ging verständlicherweise Präzision verloren. Wenn Erfinder selbst in ihrem Fachgebiet über die von ihnen erfundenen und auf bestehenden Begriffen aufbauenden, neuen Begriffe schreiben, ist zu erwarten, dass sie präzise arbeiten und korrekte Fachterminologie verwenden. Wenn viele technische Redakteure und Übersetzer, die örtlich, zeitlich, aber auch kognitiv weiter von der Innovationsquelle entfernt sind, an Content für Konsumenten arbeiten, geht unweigerlich Präzision bezogen auf einen einzelnen Fachbegriff, einen Satz oder den gesamten Text verloren. Denn nicht jeder Aspekt, den Laienautoren möglicherweise sogar unbewusst nicht verstehen, wird hinterfragt.

Maschinen können prinzipiell systematischer arbeiten als der Mensch. Wir alle schätzen heute beispielweise, dass wir je nach Wortart und Sprache mehr oder weniger per Knopfdruck alle Vorkommen eines Wortes in einem Text ersetzen können. Wie sieht es jedoch aus, wenn der Maschine – oder sprechen wir doch von generativer künstlicher Intelligenz – nur Texte zur Verfügung stehen, in denen falsche oder keine Fachwörter verwendet wurden; in denen zu wenig Kontext vorhanden ist, um Homographen zu disambiguieren; in denen Begriffszusammenhänge nicht klar genug dargestellt wurden; die nach der Übertragung aus einer Fremdsprache in eine neue Kultur an Genauigkeit verloren haben?

Wenn die Grundsätze und Methoden gemäß ISO 704 bei der Erstellung von Ausgangstexten angewendet werden,

können Unklarheiten aufgrund von falsch dargestellten Begriffsrelationen, Homographen, inkonsistenter, falscher oder unpassender Terminologie vermieden werden. Je mehr Texte auf der Grundlage von ISO 704 erstellt werden, desto präziser dürften damit auch durch generative KI erstellte Texte sein.

Gleichzeitig können natürlich die auf Basis von ISO 704 beruhenden Termbank-Inhalte die Textgenerierung beispielsweise mithilfe von der Retrieval Augmented Generation (RAG) beeinflussen. Bei RAG werden die generierten Texte anhand von zuverlässigen Datenquellen (z. B. einer Termbank) verbessert. Und zu guter Letzt können die Inhalte dieser Norm auch die Terminologiearbeit mit KI-Unterstützung erleichtern. Die Integration von generativer KI in Termbanken ist heute bereits Wirklichkeit. Aber selbst, wenn kein direkter Zugriff darauf im Tool besteht, sondern beispielsweise ChatGPT oder der Microsoft Copilot zur Erstellung von Definitionen verwendet wird, lohnt es sich die in ISO 704 beschriebene Definitionsformel – Oberbegriff plus einschränkende Merkmale – in den Prompt zu integrieren.

Zusammenfassung

Für Einsteiger in unser Fachgebiet ist ISO 704 gewiss keine leichte Kost. Die Version von 2022 ist sicherlich leichter verdaulich als die Vorgängerversionen. Die Autorin hat mindestens über zehn Jahre hin immer wieder neue Erkenntnisse beim Nachschlagen gewonnen. Es lohnte sich also schon immer auch für alte Terminologiehasen, die Norm immer wieder aufzugreifen. In Zukunft wird das Verständnis dieser Grundlagen eine noch größere Rolle spielen, wenn erstellter Content von KI weiterverwertet wird. Die Verbreitung des in ISO 704 sowie in den entsprechenden nationalen Normen enthaltenen Wissens würde sich also lohnen. Wer nach ISO 704 arbeitet, kann mindestens im eigenen Fachgebiet kommunizieren. Dass dies eine hohe Anforderung ist, wissen wir Terminologen allerdings alle.

Literatur

[1] ISO TC 37 Language and Terminology (2022): ISO 704:2022 Terminology work – Principles and methods. ISO TC 37 SC1: Genf.

[2] Schmitz, Klaus-Dirk (Hrsg.) (2022): „DIN 2330:2022 Terminologiearbeit – Grundsätze und Methoden“. In: Normen für Terminologiearbeit, Technische Redaktion und Übersetzen. DIN Deutsches Institut für Normung e. V. Beuth Verlag GmbH: Berlin.

[3] Austrian Standards (2015): ÖNORM A 2074 Terminologiearbeit – Grundsätze und Methoden. Österreichisches Normungsinstitut/Austrian Standards Institute: Wien.[i]

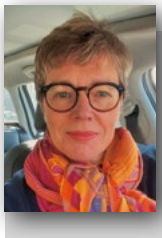
[4] ISO, Online Browsing Platform. https://www.iso.org/obp/ui/ [Letzter Zugriff am 15. September 2024]

[5] Iantosca, Michael (2022): From Information Wars to Knowledge Wars, Blog: ThinkingDocumentation. https://thinkingdocumentation.com/blog/f/from-information-wars-to-knowledge-wars [Letzter Zugriff am 15. September 2024]

[6] ISO TC37 Language and Terminology (2019): ISO 1087:2019 Terminology work – Vocabulary. ISO TC 37 SC1: Genf.

[7] Schmitz, Klaus-Dirk (Hrsg.) (2022): „DIN 2342:2022 Terminologiewissenschaft und Terminologiearbeit – Begriffe“. In: Normen für Terminologiearbeit, Technische Redaktion und Übersetzen, DIN Deutsches Institut für Normung e. V. Beuth Verlag GmbH: Berlin.

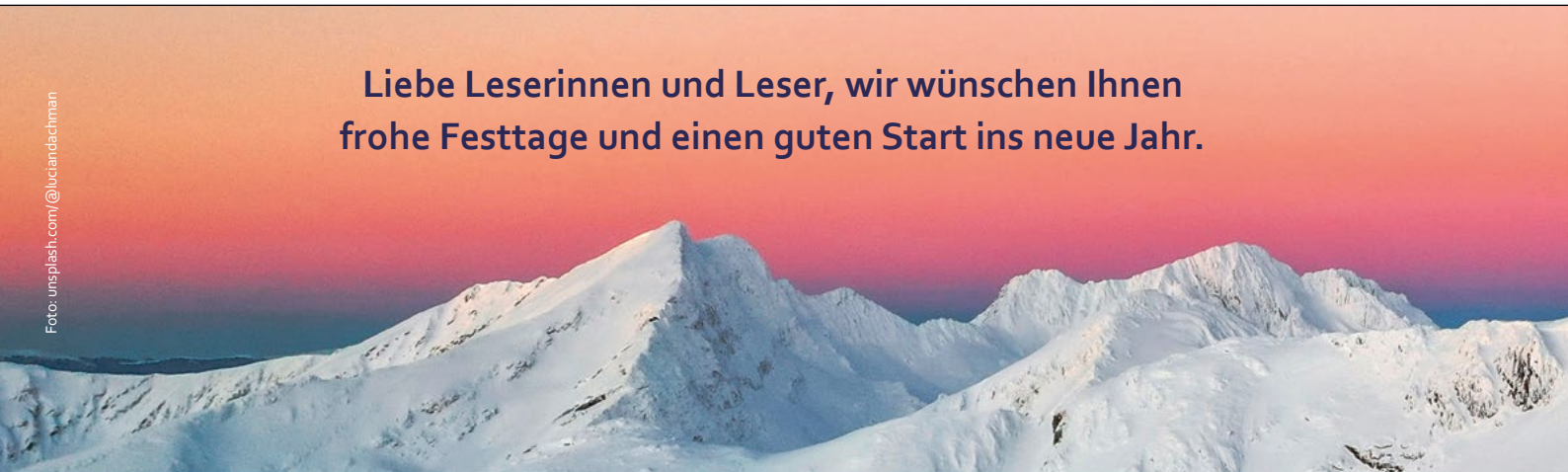
[8] Löckinger, Georg und Hebenstreit, Gernot (2016): „Terminology Work Revisited according to ÖNORM A 2704:2015”. In: Fachinformation 22. Austrian Standards Institute: Wien.



Barbara Karsch ist Terminologieberaterin und -trainerin bei BIK Terminology. Schwerpunkt ihrer Arbeit sind Terminologieschulungen, Implementierung von Terminologieverwaltungssystemen und Terminologiearbeit. Dabei greift sie auf ihre 14-jährige Erfahrung als Terminologin bei J.D. Edwards und Microsoft zurück. Seit 2009 ist sie Mitglied der amerikanischen ISO-Delegation und hat u. a. an der letzten Version von ISO 704 mitgearbeitet. Sie unterrichtet seit 2011 im Studiengang „Master's in Translation and Interpreting Online“ der New York University.

Kontaktadresse
bikterminology@gmail.com
www.bikterminology.com

[i] Herzlichen Dank an Austrian Standards dafür, dass die ÖNORM 2704 für diesen Artikel kostenlos zur Verfügung gestellt wurde.



Neues aus der Normenwelt

In dieser kurzen Übersicht stellen wir neu erschienene Normen, Normentwürfe, geplante Normen und Projekte aus der Welt der Terminologie und Übersetzung vor.

DIN ISO 5060:2024-06 Übersetzungsdienstleistungen – Evaluierung von Übersetzungen – Allgemeine Leitlinien (ISO 5060:2024)

Die Norm DIN ISO 5060 ist im Juni 2024 erschienen und enthält Leitlinien für die Evaluierung von Humanübersetzungen sowie für die Evaluierung von posteditierten maschinell erstellten Übersetzungen und nicht editierten maschinell erstellten Übersetzungen, außerdem Leitlinien im Hinblick auf Qualifikationen und Kompetenzen von Evaluierenden. Die Rolle der Stichprobenprüfung ist ebenso Thema. Der Fokus liegt auf der analytischen Evaluierungsmethode von Übersetzungen unter Verwendung von Fehlerkategorien und Fehlerpunkten.

DIN ISO 20539:2024-07 Übersetzen, Dolmetschen und zugehörige Technik – Begriffe (ISO 20539:2023)

Diese Norm legt Begriffe für Internationale Normen aus dem Bereich Übersetzen, Dolmetschen und zugehörige Technik fest.

DIN ISO 11669:2024-07 Übersetzungsvorhaben – Allgemeine Richtlinien (Entwurf)

DIN ISO 11669 stellt allgemeine Leitlinien für alle Phasen eines Übersetzungsprojekts auf. Hauptzweck ist die Sicherstellung von Effizienz und Qualität durch verbesserte Kommunikation aller Projektbeteiligten. Die Norm gibt einen Rahmen für die Entwicklung von Übersetzungsprojektspezifikationen als Grundlage für das Anfragen, Erstellen, Managen und Beurteilen von Übersetzungsprojekten vor und bietet Leitlinien für Bedarfsanalyse, Risikobeurteilung und Arbeitsabläufe.

DIN 8579 Übersetzungsgerechtes Schreiben – Texterstellung und Textbewertung (ISO/AWI 18968 Translation-oriented writing – Text production and text evaluation)

Im Jahr 2022 ist DIN 8579 „Übersetzungsgerechtes Schreiben – Texterstellung und Textbewertung“ erschienen und soll nun auch auf internationaler Ebene erarbeitet werden. Die Norm gibt Empfehlungen für die Erstellung und Bewertung von technischen Texten, die für die Übersetzung in andere Sprachen mittels menschlicher und/oder maschineller Übersetzung bestimmt sind.

DIN 19460 Mehrsprachige Terminologiearbeit – Grundsätze und Methoden (Projekt)

Die geplante Norm gibt Handlungsempfehlungen für konkrete Probleme in der mehrsprachigen Terminologiearbeit und richtet sich an alle, die in Unternehmen und Organisationen an mehrsprachigen Terminologieprozessen beteiligt sind. Sie vermittelt Grundlagen der mehrsprachigen Terminologiearbeit, indem sie unter anderem auf kulturelle Unterschiede und terminologische Lücken eingeht. Darauf aufbauend werden Probleme und Herausforderungen thematisiert, die sich in der mehrsprachigen Terminologiearbeit ergeben können. Anschließend werden Methoden und Lösungsmöglichkeiten für die Verwaltung der mehrsprachigen Terminologie und die Weiterverwendung der terminologischen Daten in anderen Systemen aufgezeigt. Die Entwurfsveröffentlichung ist für Juni 2025 geplant.

Alle Normen können bei DIN Media (vormals Beuth-Verlag) bestellt werden:
https://www.dinmedia.de/de

Wir freuen uns auf ein Wiedersehen 2025 in Worms.

Der DTT-Vorstand

RWS Trados | KALEIDOSCOPE

ENTERPRISE TERMINOLOGY MANAGEMENT

- Mühelose Terminologiebereitstellung
- Transparentes Terminologiemanagement
- Systemintegration mit CMS und TMS
- Globale Workflows
- Wissensmanagement
- Einfache Zusammenarbeit
- Rückfragenmanagement
- Terminologieprüfung

Trados GmbH | Waldburgstr. 21 | 70563 Stuttgart
www.rws.com/de | kontakt@rws.com

Kaleidoscope GmbH | Landstr. Hauptstr. 99 | 1030 Wien
www.kaleidoscope.at | info@kaleidoscope.at