

edition

Fachzeitschrift für Terminologie

1|22



Automatisierung

Von Terminologien
zu Ontologien

Seite 5

Normung

Standardized
Data Categories

Seite 11

Lernplattform

E-Learning in der
Terminologearbeit

Seite 19

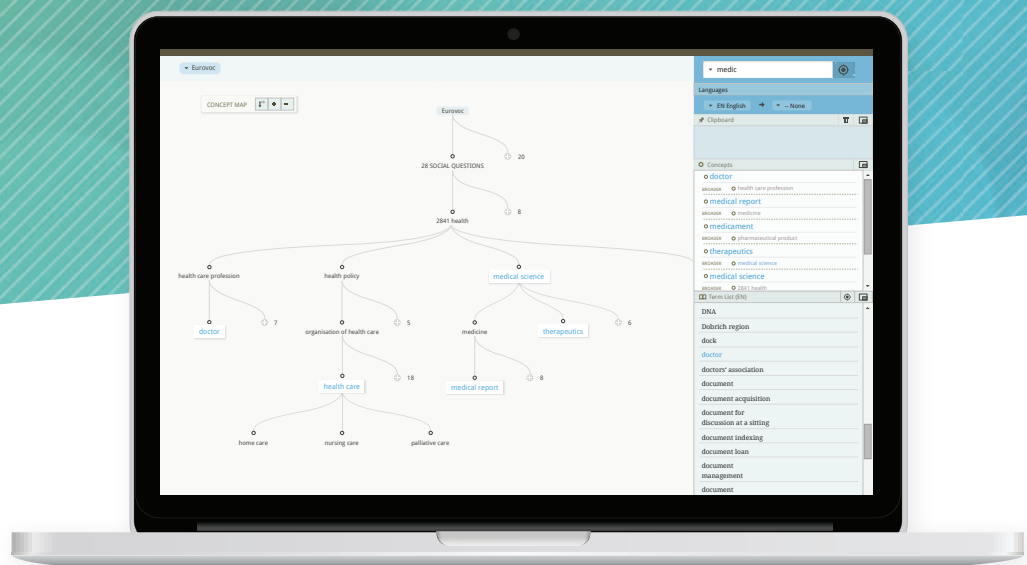
NMÜ-Engines

Beeinflussung der
Terminologietreue

Seite 25

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**



DTT-Jubiläum und Wachwechsel bei der edition

Mit diesem Heft wollen wir feiern: Der Deutsche Terminologie-Tag wird 35 Jahre alt. Sie haben sich sicher schon öfter mal gefragt: Woher kommt eigentlich dieser – für einen Verein – doch recht eigen-sinnige Name? Klaus-Dirk Schmitz blickt in einem launigen Beitrag auf die Gründungsjahre und die Entwicklung des DTT zurück. In seinen 35 Jahren hat der DTT wesentlich dazu beigetragen, dass Terminologiearbeit immer mehr in den Fokus rückt. Schließlich wurden mit den Best Practices für Terminologiearbeit Standards gesetzt, die weit über die deutschsprachige Terminologie hinausgehen.

Wir gehen auch fremd in diesem Heft: Im Interview mit dem Architekten Dennis Ewert schauen wir auf die Sprache von Mauern und die Zeichensysteme von Städten. Es gibt auch Terminologie jenseits unserer bisherigen Vorstellungskraft...

Weitere Beiträge befassen sich mit dem automatisierten Aufbau von Ontologien, der Standardisierung von terminologischen Datenkategorien,

dem Einsatz von E-Learning für die Terminologiearbeit und der Terminologietreue bei domänenspezifischen NMÜ-Engines. Unter Tools & Ressourcen „erschnüffeln“ wir schließlich ein Tool zur Terminologierecherche.

Und zum Schluss noch eine Neuigkeit in eigener Sache: Mit diesem Heft übernimmt Nicole Keller, die unseren Leserinnen und Lesern aus der Rubrik Tools & Ressourcen bestens bekannt ist, die Redaktionsaufgaben von Annette Weilandt. Annette zieht sich zwar aus der Redaktionsleitung zurück, wird dem edition-Team aber weiterhin beratend zur Seite stehen. Alle DTT-Vorstandsmitglieder, das DIT und das edition-Redaktionsteam bedanken sich sehr herzlich bei Annette Weilandt für vier Jahre sehr engagierte Redaktionsarbeit. Gleichzeitig begrüßen wir Nicole Keller ganz herzlich in der Redaktionsleitung und im Team.

Nun bleibt uns nur noch, Ihnen viel Spaß bei der Lektüre zu wünschen – und wie immer freuen wir uns über Ihre Rückmeldungen.



Dr. Nicole Keller
Redaktionsleitung
redaktion@dttev.org



Angelika Ottmann
Redaktionsleitung
redaktion@dttev.org

PRÜFEN SIE IHRE TERMINOLOGIE DOCH, WO SIE WOLLEN.

termXact.
Für jedes Gelände.

- ✓ prüft in Word, Excel, PowerPoint, Adobe Framemaker und InDesign
- ✓ prüft browserbasiert in Chrome, Edge und Firefox
- ✓ prüft in Sharepoint, Confluence, WordPress, Typo3
- ✓ prüft in MadCap Flare und oXygen
- ✓ prüft in Microsoft Translate, Google Translate & anderen maschinellen Übersetzungssystemen

Ab 17 EUR/mtl.

**Erfahren Sie mehr auf
termxact.de**

termXact erlaubt die direkte Übernahme terminologischer Daten aus gängigen Terminologieverwaltungssystemen – und mit tbxConnect auch aus anderen Beständen.

Editorial

- 3 DTT-Jubiläum und Wachwechsel bei der edition
Nicole Keller und Angelika Ottmann

Themen

- 5 Die Terminologie von heute ist die Ontologie von morgen
Thiemo von Gillhaußen, Maryline Hernandez und Axel Theofilidis
- 11 Standardized terminological data categories
Detlef Reineke
- 19 Das „Wundermittel“ E-Learning und seine Eignung für die Terminologearbeit
Lily Arnold
- 25 Je mehr Daten desto schlechter? Terminologietreue bei domänenspezifischen NMÜ-Engines
Frauke Hellerich

Tools & Ressourcen

- 32 Trufflepiggy: Terminologie komfortabel erschnüffeln
Armin Mutscheller

Interview

- 43 Ein Haus für die deutsche Sprache
Interview mit Architekt Dennis Ewert

Wissenswertes

- 10 DTT-Grundlagenseminar
- 36 Am Anfang war der Terminologie-Tag... 35 Jahre Vereinsgeschichte des DTT e.V.
Klaus-Dirk Schmitz
- 42 Termine
- 47 DTT-Vertiefungsseminar

Die Terminologie von heute ist die Ontologie von morgen

Thiemo von Gillhausen, Maryline Hernandez
und Axel Theofilidis

In this article, we show how it is possible to build a knowledge graph starting from a termbase. A natural language processing analyzer uses terms & definitions to identify relations existing between various concepts. Language patterns and phrases hint towards the existing relations which are best visualized using a knowledge graph.

Keywords: Terminology, ontology, knowledge graph, natural language processing, relations between concepts

Ausgangslage

Der Vorteil von Terminologiearbeit ist hinreichend bekannt. Erst durch die Festlegung einer firmeninternen Terminologie können Autoren und Übersetzer die Unternehmenssprache bei ihrer Arbeit verwenden.

Die ersten Terminologieverwaltungssysteme wurden als Nebenkomponente von Übersetzungssoftware und Translation Memorys entwickelt, sozusagen als digitale Karteikartenbox, in der den Benennungen einer Quellsprache die zielsprachlichen Äquivalente zugeordnet sind. Schnell wurde jedoch klar, dass es bei der Erfassung und Verwaltung von Terminologie um weitaus mehr geht, als den Übersetzungsprozess zu unterstützen. Es geht insbesondere auch darum, einen einheitlichen Sprachgebrauch zu etablieren und so eine klare und verständliche Unternehmenskommunikation sowohl nach innen wie nach außen sicherzustellen.

Terminologiebestände werden somit sukzessive nach begrifflichen Gesichtspunkten organisiert (begriffsorientierte Terminologie). Den Begriffen werden Definitionen, Fotos oder Abbildungen zugeordnet und für die Bezeichnung der unterschiedlichen Begriffe werden jeweils eindeutige Vorzugsbenennungen festgelegt. Terminologiebestände wachsen so zu umfangreichen Datenbasen heran, die das komplexe Kernwissen eines Unternehmens enthalten. Typischerweise liegt dieses Wissen aber in unstrukturierter und heterogener Form vor, z. B. als frei formulierte Definitionen, die über viele Jahre hinweg in unterschiedlichen Arbeitskontexten von unterschiedlichen Autorinnen und Autoren angelegt wurden. Damit ist das terminologische

Wissen nicht nur für maschinelle Anwendungen, sondern auch für Nicht-Fachexperten oft nur eingeschränkt nutzbar.

Im Folgenden wird gezeigt, dass der wertvolle Wissensschatz, der sich in einer Terminologie manifestiert, mit Hilfe von Sprachverarbeitungsverfahren effektiv gehoben und damit vielfältigen anderen Anwendungsszenarien zugänglich gemacht werden kann, insbesondere auch weiterführenden maschinellen Verarbeitungsmethoden. Der Ansatz besteht darin, aus den terminologischen Daten – Definitionen und Benennungen – auf der Grundlage linguistischer Analyseverfahren automatisiert semantische Netze abzuleiten.

Begriffssysteme sind eine bekannte Methode zur systematischen Aufbereitung und Darstellung von Wissen und werden schon in vielen Unternehmen eingesetzt. Der Aufbau erfolgt meist intellektuell und von Hand. Die automatisierte Extraktion von Begriffsbeziehung aus der Terminologie kann diesen manuellen Erstellungsprozess unterstützen.

Semantische Netze

Bei einem semantischen Netz handelt es sich um ein Netzwerk (gerichteter Graph) aus Begriffen, welches das Wissen über deren Bedeutung darstellt. Formal ist ein Graph als eine Menge von Knoten und Kanten definiert. Folgende Abbildung stellt den kleinsten **Wissensgraphen** dar, den wir erstellen können – er wird auch als Tripel bezeichnet:

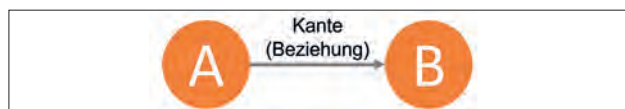


Abb. 1: Wissensgraph (Tripel)

Knoten A und B in Abb. 1 sind hier zwei verschiedene Einheiten (Begriffe). Diese Knoten sind durch eine Kante verbunden, welche die Beziehung zwischen den beiden definiert. Ein konkretes Beispiel ist „Getriebeöl ist ein Schmierstoff“, wobei „Getriebeöl“ und „Schmierstoff“ die zwei Knoten darstellen, welche durch die Beziehung „ist ein“ miteinander verbunden werden.

Um solche semantischen Netze aus Benennungen und Definitionen abzuleiten, müssen sie mit Hilfe von Verfahren der maschinellen Sprachverarbeitung (**natural language processing** oder kurz **NLP**) analysiert werden.

Linguistische Analyse

Die maschinelle Sprachverarbeitung erfolgt durch eine regelbasierte **Linguistic Engine** und wird typischerweise in mehreren aufeinanderfolgenden Analyseschritten durchgeführt.

Aus den Ausgangsdaten wird zunächst der sprachlich relevante Teil extrahiert (Rohtext), das heißt, Auszeichnungen wie XML-Tags oder Formatierungsmarkierungen werden abgetrennt. Der zu verarbeitende Text wird anschließend in einzelne Sätze zerlegt. Dieser Schritt ist komplexer als allgemein angenommen. So darf z. B. nicht pauschal nach jedem Punkt ein neuer Satzansatz vermutet werden. Der Punkt wird als Abkürzungszeichen in „z. B.“ oder bei Ordinalzahlen „1., 2., 3.“ mitten im Satz verwendet. Ähnlich verhält es sich mit der anschließenden Zerlegung der Sätze in Wörter.

Der nächste Schritt besteht in einer morphologischen Analyse der einzelnen Wörter, also einer Analyse der Wortformen. Dabei wird jedes Wort in Teileinheiten getrennt, die in einem Wörterbuch aus möglichen Wortbestandteilen nachgeschlagen werden. Das Wörterbuch für Deutsch umfasst mehrere Tausend einzelne Wortbestandteile. Für ein gegebenes Wort ergeben sich oft unterschiedliche Analysen und damit auch Bedeutungen. Für „Montage“ wird beispielsweise ermittelt, dass es sich entweder um „die Montage“ (von montieren) oder um die Pluralform von „der Montag“ handelt.

Diese (meist mehrdeutigen) Ergebnisse bilden die Eingabe für den nächsten Schritt. Ein linguistisches Regelwerk übernimmt die grammatische Analyse der Sätze. Dabei werden zusammengehörige Bestandteile erkannt und markiert. Im Satz „Die Montage der Mischbatterie erfordert Spezialwerkzeuge.“ wird „erfordert“ als Prädikat und „die Montage der Mischbatterie“ als Nominalphrase und Subjekt erkannt. Dadurch wird die für „Montage“ morphologisch mögliche Interpretation als Pluralform von „Montag“ verworfen. Der Satz wird dadurch vereindeutigt.

Nach dem gleichen Prinzip erkennen nachgelagerte Regelwerke unterschiedlichste Eigenschaften und Muster in den Sätzen eines Texts und vergeben entsprechende Markierungen. Je nach Anwendung können solche Markierungen auf orthografische, grammatische oder stilistische Fehler hinweisen, sie können korrekt oder inkorrekt verwendete Terminologie anzeigen und sie können auch begriffliche Beziehungen aufzeigen, die zwischen Satzbestandteilen bestehen.

Auf dieser Grundlage wurde ein Terminologiebestand von ca. 10.000 Begriffen analysiert. Aus den terminologischen Einträgen wurden jeweils die Vorzugsbenennung und die Definition untersucht. Dafür wurde ein spezielles Regelwerk erarbeitet, das unter Zuhilfenahme der linguistischen Analyse in mehreren Stufen aus den Benennungen und den Definitionen unterschiedliche Begriffsbeziehungen extrahiert. Die aus diesem Projekt gewonnenen Erkenntnisse werden im Folgenden vorgestellt.

Erkennung von Abstraktionsbeziehungen

Ein Teil des Regelwerks zur linguistischen Analyse betrachtet den Aufbau der Definitionen. Gängig sind Definitionen mit Nennung der Gattung (Oberbegriff) und des spezifischen Unterschieds, was in der philosophischen Tradition „Genus proximum et differentia specifica“ genannt wird (siehe auch [4]).

Das erarbeitete Regelwerk untersucht unterschiedliche Typen von Definitionen, die entweder als vollständiger Satz mit finitem Verb oder als Phrasen mit unterschiedlichen Attributen realisiert sind, und identifiziert so die Benennung des Oberbegriffs zur Ausgangsbenennung. Zwischen den beiden Begriffen besteht die Beziehung „ist ein“. Es entsteht ein **hierarchisches Begriffssystem** (vgl. [1] und [2]).

Ein Beispiel für eine Definition, welche die Beziehung zwischen der Ausgangsbenennung und dem Gattungsbegriff explizit durch das Verb „sein“ ausdrückt, ist: „Ein Kettenrad **ist** ein Maschinenelement, das zur Kraftübertragung mit Ketten dient“. Das eingesetzte Regelwerk ermittelt für diese Definition die Beziehung: „Kettenrad ist ein Maschinenelement“.

Das weitaus gebräuchlichere Definitionsmuster in Terminologiebeständen nennt dagegen den Gattungsbegriff direkt in Form einer Phrase, in der begleitende Adjektive, Partizipien oder Relativsätze die für die Ausgangsbenennung spezifischen Eigenschaften beschreiben, z. B.: „Ventil: Bauteil bestehend aus Schaft, Teller und Dichtfläche“ oder „Motoröl: Betriebsstoff zur Schmierung der Motorteile“. Aus diesen Definitionen leitet das Regelwerk die Begriffsbeziehungen „Ventil ist ein Bauteil“ bzw. „Motoröl ist ein Betriebsstoff“ ab.

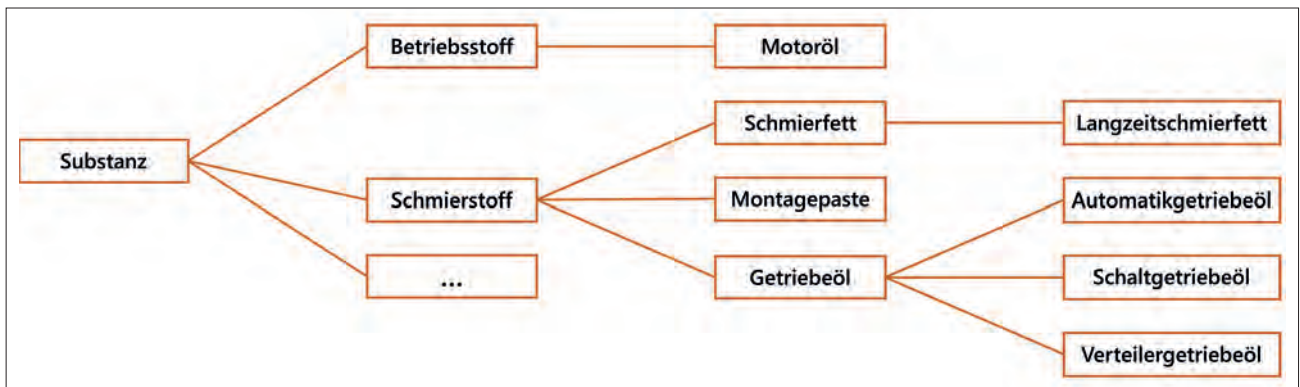


Abb. 2: Categoriesystem mit Ober- und Unterkategorien

Wenn dieses Verfahren mehrmals und über einen gesamten Terminologiebestand einschließlich der Definitionen zur Anwendung kommt, entstehen mehrere Hierarchieebenen und dadurch **Begriffsleitern**. Jede Benennung der Leiter stellt eine Abstraktionsebene dar.

Dem Beispiel in Abb. 2 kann entnommen werden, dass es sich bei „Schaltgetriebeöl“ um eine Unterkategorie von „Getriebeöl“ handelt. Weiter ist definiert, dass es sich bei „Getriebeöl“ um einen „Schmierstoff“ handelt, der wiederum zur Oberkategorie der „Substanzen“ gehört.

Hinweis: In der Terminologielehre spricht man von **Begriffssystemen mit Ober- und Unterbegriffen** (vgl. [1]), im Bereich der Ontologie spricht man dagegen von **Kategorie-systemen mit Ober- und Unterkategorien** (vgl. [7]).

Erkennung weiterer Begriffsbeziehungen

Durch die genaue Betrachtung der Definitionen kann nicht nur die Beziehung von Ober- und Unterkategorien abgeleitet werden. Aufgrund von weiteren sprachlichen Mustern können auch zusätzliche Begriffsbeziehungen identifiziert werden. Typischerweise drücken Präpositionen solche Begriffsbeziehungen aus.

Eine Präpositionalphrase mit „zu“ wie „zum Schmierren“ gibt beispielsweise die Beziehung der **Funktion** an, in welcher zwei Begriffe zueinanderstehen. Mit „an“ oder „in“ wird oft ein **Ort** angegeben („am Zylinderkopf“, „im Benzineinspritzsystem“) und mit „aus“ ein **Material** („aus Schmieröl und Eindicker“). All diese Beziehungen

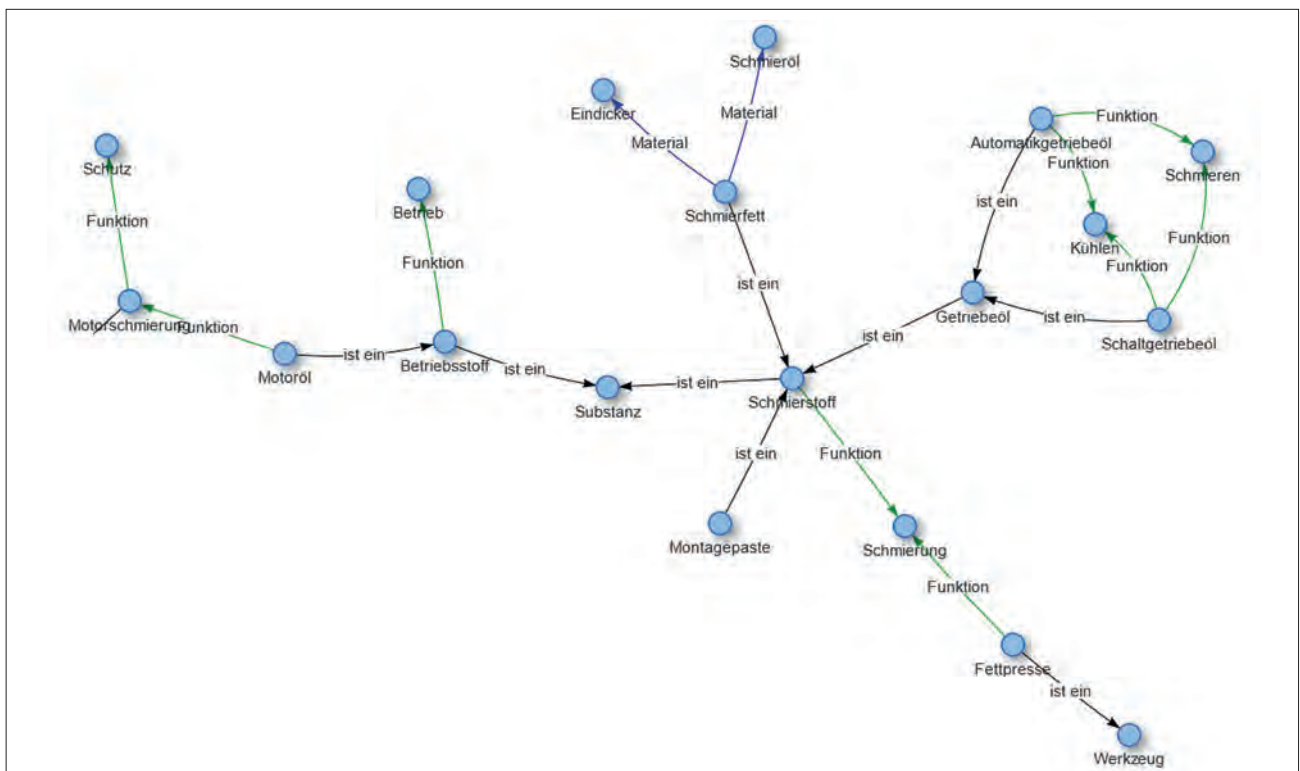


Abb. 3: Darstellung unterschiedlicher Begriffsbeziehungen: Gemischtes Begriffssystem



Abb. 4: Bestandsbeziehungen nach DIN 2331

charakterisieren weitere Eigenschaften der ermittelten Kategorien und verbinden diese untereinander (vgl. Abb. 3).

Auch die **Bestandsbeziehung** „ist Teil von“ wird durch spezielle Präpositionen und sprachliche Muster identifiziert, z. B. durch eine Präpositionalphrase mit „mit“, wie in „Augenschraube: Schraube mit einem augenförmigen Befestigungsring am Ende“, oder durch eine Genitivkonstruktion, wie in „Einlassventil: Bauteil des Verbrennungsmotors“.

Neben den schon genannten Beziehungen erkennt das linguistische Verfahren weitere Beziehungen wie **direktes Objekt** der ermittelten Funktion („zur Schmierung **der Motorteile**“) oder **Gegensatz**. Somit werden die Begriffe eines Terminologiebestands auf vielfältige Weise miteinander verknüpft. Das so entstehende Beziehungsnetz kann grafisch dargestellt werden. Concept Maps, wie in Abb. 3 gezeigt, sind dafür besonders geeignet, da eine eindeutige Beschriftung den Typ der Begriffsbeziehung kennzeichnet. Damit kann aus einem nicht-hierarchischen Textbestand an Benennungen und Definitionen erkannt werden, welche Begriffe in welcher Beziehung (oder welchen Beziehungen) zueinanderstehen. Dies kann als Grundlage für eine

Graphenvisualisierung dienen, um einen Terminologiebestand idealerweise vollständig als semantisches Netz darzustellen.

Erkennung von Abstraktionsbeziehungen aus Benennungen

Auch ohne Definitionen können Abstraktionsbeziehungen mit linguistischen Verfahren automatisch ermittelt werden: durch die morphologische Analyse der Benennungen. Terminologisch relevante Benennungen sind sehr oft zusammengesetzte Wörter wie „Expansionsventil“ oder „Getriebeöl“. Diese werden im Deutschen so gebildet, dass ein Grundwort („-öl“) durch Voranstellung von mindestens einem Bestimmungswort („Getriebe-“) näher definiert wird. Auf dieser Grundlage kann die linguistische Analyse aus Wortzusammensetzungen Oberkategorien ableiten und dabei zugleich prüfen, ob die Benennungen der ermittelten Oberkategorien ihrerseits im Terminologiebestand vorkommen.

Bei komplexeren Benennungen entstehen so mehrstufige Begriffsleitern. Wenn Benennungen in mehreren

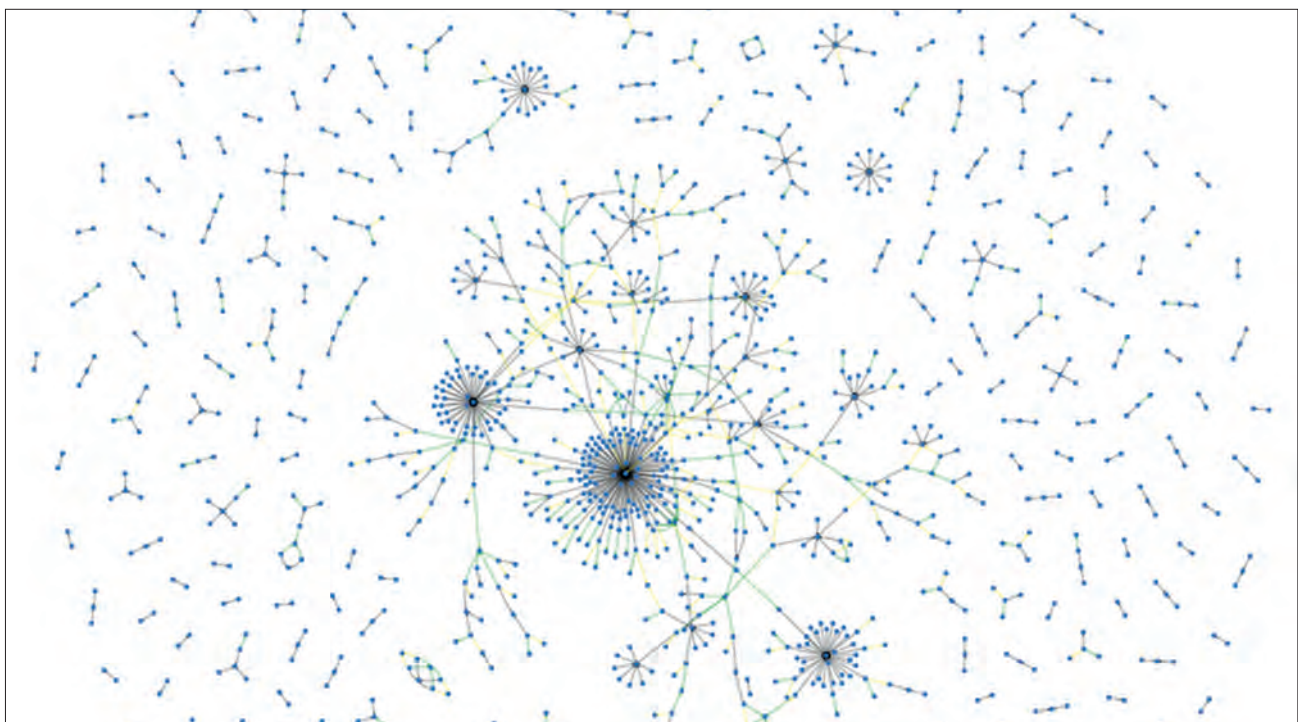


Abb. 5: Darstellung eines kompletten semantischen Netzes

Begriffsleitern vorkommen, wird auf diese Weise das Begriffssystem ergänzt. Die Terminologiearbeit kann dadurch systematisch angegangen werden, indem beispielsweise alle Benennungen betrachtet werden, die „Ventil“ als Grundwort haben.

In manchen Fällen kommt die automatische Verarbeitung an ihre Grenzen, nämlich bei Mehrdeutigkeiten. Wenn identische Benennungen mit unterschiedlichen Definitionen als mehrere terminologische Einträge existieren, würde derzeit die automatische Verarbeitung annehmen, dass es sich um ein und denselben Knoten im semantischen Netz handelt. Eine solche Mehrdeutigkeit fällt z. B. auf, wenn man in eine andere Sprache übersetzt, so etwa bei „Sender“ mit seinen zwei Definitionen:

1. elektrisches Bauteil, das Sendeimpulse erzeugt.
Auf Englisch: „transmitter“
2. Rundfunk- oder Fernsehveranstalter.
Auf Englisch: „sender“

Im Beispiel ergäbe sich ein einzelner Knoten für „Sender“ mit zwei Beziehungen des Typs „ist ein“ („ist ein Bauteil“, „ist ein Rundfunkveranstalter“) und der Beziehung **Funktion** („erzeugt Sendeimpulse“). Daraus ergibt sich die Unstimmigkeit, dass die zugewiesene Funktionsbeziehung für den „Rundfunkveranstalter“ nicht zutreffend ist.

Nutzen

Die linguistisch aufbereitete Terminologie bildet de facto eine **Ontologie**, da sie das Wissen einer Domäne modelliert und Informationen in Beziehung setzt. Die Methode, Terminologie als Begriffssystem abzubilden, ist ein Grundsatz der Terminologielehre. Das vorgestellte Verfahren ist in der Lage, diesen Prozess automatisiert zu unterstützen und damit einen wertvollen Beitrag zur Terminologiearbeit zu leisten.

Die grafische Darstellung der Terminologie erlaubt es Nicht-Fachexperten, die einzelnen Begriffe und deren Beziehungen als Teil eines großen Ganzen zu verstehen, und erleichtert so den Zugang zu dem enthaltenen Wissen.

Neben der visuellen Aufbereitung der Ergebnisse kann die Ontologie auch für andere Prozesse oder Anwendungen nutzbar gemacht werden. Aus dem Umfeld des Semantic Web sind hierfür standardisierte Technologien bekannt. Zwei der bekanntesten Vertreter sind wohl das RDF (Resource Description Framework) und die Ontologiesprache OWL (Web Ontology Language). Beides sind Austauschformate, welche die Ontologie als Tripel in Form von z. B. XML speichern. Es gibt auch die Möglichkeit, das extrahierte Wissen in einer Graphendatenbank zu speichern. Die auf diese Weise bereitgestellte Ontologie kann in viele

Prozesse eingebunden werden, die meistens außerhalb des Verantwortungsbereichs der Terminologie liegen. Beispielsweise können Suchanfragen auf der Webseite oder im Webshop optimiert werden. Die Suche nach einer Benennung kann um die Suche nach dem Oberbegriff ergänzt werden, um die Information im Kontext zu sehen.

Neben dem praktischen Nutzen für die Terminologiearbeit und andere Bereiche im Unternehmen bietet der **automatische Ontologieaufbau** auch einen deutlichen Vorteil an Kosten und Zeit. So muss niemand händisch die Ontologie aufbauen und dafür sämtliche Begriffe dahingehend intellektuell untersuchen.

Fazit

Durch das vorgestellte Verfahren konnte das in der Terminologie enthaltene Wissen schnell und einfach extrahiert werden. Auf dieser Basis war es möglich, die Arbeit der Terminologen erstmals grafisch darzustellen und so einem breiten Anwenderkreis zugänglich zu machen. Im Vergleich zu einer manuellen Ermittlung der Begriffsbeziehungen konnte mit dem automatischen Verfahren ein rechnerisches Einsparpotenzial von 70 % erreicht werden. In Zukunft wird beim Projektpartner die Aufbereitung der Terminologie als Concept Map ein fester Bestandteil der Terminologiearbeit sein.

Um das Verfahren für den automatischen Ontologieaufbau optimal anzuwenden, sind gute und richtige Definitionen von großem Vorteil. Wer diesen ersten Schritt noch nicht gegangen ist, erhält konkrete Vorgehensweisen bei Pulitano [5] und Schmitz [6]. Im Fall von nicht vorhandenen Definitionen beschränkt sich die Ontologie auf ein System aus Ober- und Unterkategorien.

Die Ergebnisse des automatischen, linguistisch fundierten Ontologieaufbaus konnten überzeugen. Tests mit anderen Terminologiebeständen haben die Verfahren bestätigt, gleichzeitig aber auch Ansätze für Verbesserungen bzw. Weiterentwicklungen des bestehenden Regelwerks aufgezeigt. Zugleich könnte es ein lohnender Ansatz sein, das Verfahren auf andere Textsorten bzw. auf ganze Texte auszuweiten.

Literatur

- [1] DIN 2331 (2019): Begriffssysteme und ihre Darstellung. Berlin: Beuth.
- [2] Drewer, Petra (2019): Begriffssysteme und ihre Darstellung – Zur Neuausgabe von DIN 2331 sowie zur grundsätzlichen Bedeutung von Begriffssystemen für die Terminologiearbeit. In edition – Fachzeitschrift für Terminologie, Ausgabe 2/2019, 15. Jahrgang, S. 5-12.
- [3] DTT – Deutscher Terminologie-Tag (2014) (Hrsg.): Terminologiearbeit – Best Practices 2.0. 2. überarbeitete und ergänzte Auflage. Koordination und Redaktion: Drewer, Petra / Pulitano, Donatella / Schmitz, Klaus-Dirk. Köln: Deutscher Terminologie-Tag e.V.

DTT-Fortbildung - online -



© pixelio.de

DTT-Grundlagenseminar

„Terminologiearbeit – Grundlagen, Werkzeuge, Prozesse“

14. Oktober 2022
9:00 - 17:30 Uhr
Online

Referenten

- Prof. Dr. Petra Drewer
- Prof. Dr. Rachel Herwartz
- Prof. Dr. Klaus-Dirk Schmitz

Programm

- Terminologiearbeit – Grundlagen
- Werkzeuge Teil I: Recherche und Extraktion
- Methoden, Prozesse, Beteiligte, Qualifikation
- Werkzeuge Teil II: Management und Kontrolle

Diese Veranstaltung zählt zum
DTT-Terminologiezertifikat.

Weitere Informationen und
Anmeldung unter:
dttev.org/fortbildung

Stand: 1. Mai 2022 | Änderungen vorbehalten.

Thema | Die Terminologie von heute ist die Ontologie von morgen

- [4] Fleischmann, Klaus; Theofilidis, Axel (2021): Ontologieaufbau muss kein Mammutprojekt sein. Vortrag tekomp-Jahrestagung 2021.
- [5] Pulitano, Donatella (2018): „Definitionen – Haute Couture oder Prêt-à-porter? In: Drewer, Petra / Mayer, Felix / Schmitz, Klaus-Dirk (Hrsg.) (2018): Terminologie und Text(e). Akten des DTT-Symposiums 2018. Köln: Deutscher Terminologie-Tag e.V., S. 53-66.
- [6] Schmitz, Klaus-Dirk (2015): „Definitionen in der Terminologiearbeit – die Definition als wichtiges Element der Terminologiearbeit und in Termbanken“. In edition – Fachzeitschrift für Terminologie, Ausgabe 1/2015, 11. Jahrgang, S. 10-16.
- [7] Stuckenschmidt, Heiner (2011): Ontologien – Konzepte, Technologien und Anwendung (2. Aufl.). Heidelberg: Springer.



Thimo von Gillhausen verantwortet bei der Congree Language Technologies GmbH den Unternehmensbereich Congree Search und damit die Themen rund um Information Extraction und Content Classification bis hin zum Ontologieaufbau. Er beschäftigt sich seit langem mit modularer Informationserstellung, Analyse von Massendaten und Strukturierung von Inhalten im Allgemeinen.

Kontaktadresse
tvgillhausen@congree.com
www.congree.com



Maryline Hernandez hat Fachübersetzen (Französisch, Deutsch, Englisch) und Computerlinguistik studiert. Sie hat an der Weiterentwicklung der Komponente Terminologieprüfung von CLAT (heute Congree) maßgeblich mitgewirkt. Als Linguistic Engineer bei der Congree Language Technologies GmbH liegt ihr Fokus auf Themen rund um Terminologie.

Kontaktadresse
mhernandez@congree.com
www.congree.com



Axel Theofilidis ist seit 1990 als Computerlinguist in Saarbrücken tätig – zunächst beim Institut für angewandte Informationsforschung (IAI) und der IAI Linguistic Content AG, seit 2020 bei der Congree Language Technologies GmbH. Im Rahmen sprachtechnologischer Entwicklungsprojekte bilden die sprachliche Qualitätssicherung sowie der Aufbau von Terminologie und Ontologie seine Schwerpunktthemen.

Kontaktadresse
atheofilidis@congree.com
www.congree.com

Standardized terminological data categories

Detlef Reineke

A fundamental aspect in terminology management is the appropriate, use case-specific design of terminology databases. This refers to the overall database structure as well as to the careful selection of terminological data categories. In this sense, the community of practice has made numerous efforts to provide standardized terminological data category resources and therefore support uniform data (re-)use and interoperability.

Keywords: terminology database design, terminological data categories, terminological data exchange, ISO 12620, ISOcat, DatCatInfo, IEC CDD

Every well-trained terminologist is aware that the foundation for success in interoperable terminology management is a thoroughly designed terminology database. This implies a sound use case analysis prior to any implementation of the database in a specific terminology management tool. The analysis should lead to a data model that describes the class of required data fields, i.e., the data categories, types, permissible values, cardinalities, and relations. In a hypothetical data model, the data category /definition/, for example:

- shall be an open data category, means it instantiates free text (but we could imagine a further restriction here concerning the maximum number of characters, for example);
- shall only occur once at the concept level;
- and shall always be accompanied by a source.

The data category /source/ itself could be a closed data category, i.e. a pick list with predefined values, and should be repeatable at any level of the concept entry structure to describe the source of other types of terminological data such as figures, contexts, explanations, or notes.

Now, given the case you do not (and never will) share terminology with others and permanently store your terminology database in the same terminology management tool, you probably would not have to think so much about a collaborate environment-specific data model as long as it fits your needs and you get along with it. However, if you plan to import, export, or publish your terminology for further re-use or processing in other tools or applications, it is advisable to adhere to the terminological metamodel defined in ISO 16642:2017

“Computer applications in terminology – Terminological markup framework” [1] and to make use of common standardized terminological data categories. Advantage: your terminology database is more intuitive and easier to use whilst raising options for a transparent and lossless data exchange and hence reducing potential costs due to tedious data analysis, data mapping, and data transformation.

ISO 12620:1999: the rise of standardized terminological data categories

The terminology community of practice has long been aware of the benefits of using standardized terminological data categories. In 1999, ISO published the first international standard on terminological data categories, ISO 12620:1999 “Computer applications in terminology – Data categories” [2]. It included 277 terminological data categories and respective data category specifications as well as 35 bibliographical data categories taken from ISO 12083:1994 “Information and documentation – Electronic manuscript preparation and markup” [3]. The terminological data categories were divided into three major groups – term and term-related data categories, concept-related data categories, and administrative data categories – and subdivided into a total of ten subgroups. Table 1 on the following page illustrates the specification for the data category /grammatical gender/.

The aforementioned data category specification does not represent the complete format. Additional fields for description include “Admitted name”, “Full form”, “Related name”, and “Nonadmitted name”.

| | |
|-----------------------|--|
| Notation number | A.2.2.2 |
| Preferred name | grammatical gender |
| Description | A grammatical category that indicates grammatical relationships between words in sentences. |
| Note | The concept of gender varies from language to language and is not a universal feature of all languages. |
| Example | In French, <i>vie</i> (life) is feminine and is used with feminine articles such as <i>la</i> , the feminine pronoun <i>elle</i> , and feminine adjective endings, e.g., <i>une vie longue</i> . |
| Permissible instances | Types of grammatical gender commonly documented in terminology databases include: a) masculine b) feminine c) neuter d) other |

Table 1: Data category specification in ISO 12620:1999

TBX data categories

The terminological data categories defined in ISO 12620:1999 were kind of agnostic, i.e. application-independent. However, a considerable number of these data categories were later included in the TermBase eXchange (TBX) standard ISO 30042:2008 “Systems to manage terminology, knowledge and content – TermBase eXchange (TBX)” [4] and the TBX standard published by the Localization Industry Standards Association (LISA) [5] which is identical to ISO 30042:2008 in title and content. In addition, some new data categories were defined such as /subordinate concept generic/, /coordinate concept generic/, /superordinate concept generic/ (as well as their partitive counterparts), /term location/ or /term qualifier/, for example [6]. All these TBX data categories had normative character and were declared default, thus coining the name TBX-Default as one among multiple potential TBX dialects. Nonetheless, in the new ISO TBX version ISO 30042:2019 “Management of terminology resources – TermBase eXchange (TBX)” [7] the number of commonly used terminological data categories was reduced to 21 and their status downgraded to non-normative.

In 2009, LISA launched TBX-Basic [8], a drastically scaled-down TBX version with only two dozen terminological data categories. After LISA became insolvent in 2011, TBX-Basic was picked up by TerminOrgs, an international terminology association for large organizations. Furthermore, the Language, Terminology/Translation and Acquisition Consortium (LTAC) developed an even smaller and less industry-relevant TBX dialect called TBX-Min with about a dozen terminological data categories [9].

In essence, it needs to be taken into account that, with the exception of ISO 30042:2008, TBX-Basic, and TBX-Min, TBX data categories are currently not a clear-cut, stable set of terminological data categories rather than terminological data categories that happen to be part of any kind of TBX document instance.

ISOcat, a wiki-based metadata registry

ISO 12620:1999 and even ISO 30042:2008 were still rooted in a PDF-based development and distribution environment. However, in the early 2000s part of the database management community experienced a paradigm shift towards a more wiki-based approach that also seemed to better fit the needs for a continuously updatable data category registry instead of cumbersome and time-consuming hardcopy publication procedures. Thus, in 2008, ISO launched the metadata registry ISOcat, the first publicly accessible electronic ISO database standard based on ISO/IEC 11179 series “Information technology – Metadata registries (MDR)” [10]. ISOcat was developed, hosted, and maintained by the Max Planck Institute for Psycholinguistics in Nijmegen, which also acted as registration authority.

At that time, ISO published ISO 12620:2009 “Terminology and other language and content resources – Specification of data categories and management of a Data Category Registry for language resources” [11]. This standard no longer listed terminological data categories. It provided a framework for the specification of data categories and the implementation of data category registries, not only for terminological data categories, but also for data categories used in other communities of practice such as lexicography, morphology, machine translation, or corpus linguistics.

Most of the terminological data categories from ISO 12620:1999 found their way into the new ISOcat registry. The underlying data category description format was heavily enriched to be able to cope with the challenges that arose from a continuously operational, authoritative, and multilingual online database registry. In ISOcat, a data category entry was divided into the following main sections further separated into subsections (see Table 2):

- a header including the data category’s key and permanent identifier (PID) as well as information about the type, owner, and scope of the data category;

- an administrative section including the data category identifier, status, justification, source, and versioning information;
- a description section providing information on the data category's profile affiliation as well as language-specific subsections containing the natural language name, the definition, examples, notes, and the corresponding sources;
- a conceptual domain section that gave information on data type and profile affiliation, and also the list of permissible values for closed data categories that were linked to their respective entries as these values themselves constitute data categories;
- and linguistic sections that allowed to describe language-specific constraints on permissible values.

| Data Category: grammatical gender | |
|--|---|
| Key | 245 |
| PID | http://www.isocat.org/datcat-DC-245 |
| Type | complex, closed |
| Owner | (Name) |
| Scope | public |
| 1. Administration Information Section | |
| 1.1 Administration Record | |
| Identifier | grammaticalGender |
| Version | 1:0 |
| Registration Status | private |
| Administration Status | private |
| Justification | Grammatical gender of nouns, pronouns, and adjectives is a standard feature of many Indo-European languages. [sic.] |
| Origin | ISO 12620:1999 |
| Explanatory Comment | ISO 12620A-020202 |
| Effective Date | 2001-09-11 |
| 1.1.1 Creation | |
| Creation Date | 2000-10-24 |
| Change Description | no change description found |
| 1.1.2 Last Change | |
| Last Change Date | 2012-01-01 |
| Change Description | added sentence on relation to natural gender |
| 2. Description Section | |
| Profile | Terminology |
| 2.1 Data Element Name Section | |
| Data Element Name | grammatical gender |
| Source | ISO 12620:1999; ISO 30042:2008 |
| 2.2 English Language Section | |
| Language | English (en) |
| 2.2.1 Name Section | |
| Name | grammatical gender |
| Name Status | standardized name |
| 2.2.2 Definition Section | |
| Definition | a grammatical category that indicates grammatical relationship between words in sentences |
| Source | ISO 12620:1999 |

Table 2 (Part 1 of 2): Data category specification in ISOcat

| | |
|------------------------------------|--|
| 2.2.3 Example Section | |
| Example | The association of a concept with a gender can be based on word form or other factors and is not necessarily always related to natural gender. In German, the noun "Brücke" (bridge) is feminine, but the French noun "pont" is masculine. |
| Source | Document recent studies |
| 2.2.4 Explanation Section | |
| Explanation | The concept of gender varies from language to language and is not a universal feature of all languages. |
| Source | ISO 12620:1999; SALT |
| 2.3 French Language Section | |
| Language | French (fr) |
| 2.2.1 Name Section | |
| Name | genre grammatical |
| ... | ... |
| 2.4 German Language Section | |
| Language | German (de) |
| 2.2.1 Name Section | |
| Name | Genus |
| ... | ... |
| 3. Conceptual Domain | |
| Data Type | string |
| Profile | Terminology |
| Value | /feminine/ |
| Value | /masculine/ |
| Value | /neuter/ |
| Value | /otherGender/ |
| 4. Linguistic Section | |
| Language | French (fr) |
| 4.1 Conceptual Domain | |
| Data Type | string |
| Value | /feminine/ |
| Value | /masculine/ |
| 5. Linguistic Section | |
| Language | German (de) |
| 5.1 Conceptual Domain | |
| Data Type | string |
| Value | /feminine/ |
| Value | /masculine/ |
| Value | /neuter/ |

Table 2 (Part 2 of 2): Data category specification in ISOCat

When migrating the terminological data categories from ISO 12620:1999 and ISO 30042:2008 to ISOCat, ISO experts subjected the data category specifications to a careful review:

- renaming some data categories ([/classification number/](#) became [/classification code/](#), for example, or [/terminology management transaction/](#) was changed to [/transaction/](#));
- changing definitions (see [/characteristic/](#) or [/context/](#));
- modifying permissible values of existing data categories (see [/grammatical number/](#) or [/register/](#));
- adding completely new data categories (such as [/causally related concept/](#));

- and removing others (/sequential relation/ and their values /temporal relation/ and /spatial relation/, /grammar/ or /nondescriptor/, for example).

One of ISOcat's most attractive new core features consisted of allowing registered users to select data categories according to their needs and download these data category selections in various formats (HTML, RDF, RNG, XSD, etc.) for further use. Interesting too that new data categories could be proposed by any authorized member and not only by ISO experts through traditional ISO standardization procedures. But, despite ISOcat's promising approach it was abandoned in 2014 due to both an overly complex proposal, validation, release, and maintenance procedure and the lack of industry demand. As a consequence, ISO relinquished all rights and ownership of its data category registry to LTAC/TerminOrgs who migrated approximately half of ISOcat's data category specifications (nearly 3,000) to a new repository called DatCatInfo [12] hosted on TermWeb, an online terminology management tool from Swedish tool provider Intervetrum.

The DatCatInfo repository

Lessons learned from ISOcat's deficiencies, the DatCatInfo data model was reduced and the complex registration procedure dropped with the consequence that DatCatInfo was not a registry anymore rather than a repository. However, most of the features are still available in DatCatInfo such as creation, editing, search, filter, and export of data category specifications. A new interesting feature is the possibility to visualize relations between closed data categories and its permissible values through a rudimentary graph. Table 3 on the following page depicts the data category specification for /grammatical gender/ in DatCatInfo.

Similar to ISOcat, the entry structure for specifying data categories is borrowed from the concept-orientated terminological metamodel defined in ISO 16642:2017. The entry structure complies with the data model recommended in ISO 12620:2019 "Management of terminology resources – Data category specifications" [13], that means the data describing data categories are deposited at the concept, the language, or the data category name level. Sections for data category specifications in other languages can be added as needed.

A critical point in both the ISO 12620:2019 and DatCatInfo data model is the fact that the definition for a data category can be allocated to the data category name level. This is not such a problem as long as there is only one data category name per language. But if, as stated in the standard, more than one data category name can be documented per language (likewise term autonomy in terminology

databases) and each data category name section is given all the data category name section fields to completely describe a data category name, different definitions for different data category names in one language are possible. And this is something that, in principle, abets bad database practice.

Another major concern is DatCatInfo's data quality. A quick look at some data category specifications reveals a significant number of inconsistencies, intransparencies, duplicates, gaps, and even errors. Some evidence (as of 15 May 2022):

- When including a standard source for a data category justification, definition, example or explanation, for example, the year of publication is sometimes added (ISO 12620:1999), sometimes it is missing (ISO 12620).
- Some data category specifications include more than one standard source such as /grammatical number/ (ISO 12620:1999 and ISO 30042:2008) although the definition and the permissible values are identical in both standards.
- Some reference to extinct ISOcat is still maintained (see /administrative status/).
- It is stated that the data category /responsibility/ can include: originator, inputter, updater, checker, approver, user, subset owner, withdrawer, exporter, importer. Some of the values are included in DatCatInfo such as /updater/, /exporter/ and /importer/ whilst others such as /originator/, /inputter/, /checker/, or /approver/ are missing.
- The data category /grammatical number/ within the profile "Terminology" contains data fields for the corresponding permissible values, but these data fields are empty even though these values are described in separate entries.
- The data categories /figure/ and /audio/ are implemented as string (or free text) whereas the data category /video/ is a pick list value. Moreover, the data category specification for /video/ does not include a link to the corresponding data category that allows for instantiating the permissible value /video/.
- The data category /figure/ has duplicate entries and overlaps with the data category /xGraphic/.
- The data category /reliability code 1/ is specified using the wrong name and identifier (/reliability code/ instead of /reliability code 1/ and reliabilityCode instead of reliabilityCode1).
- /broader concept generic/ and /broader concept partitive/ have no counterpart /narrower concept generic/ and /narrower concept partitive/.
- The search does not always work properly. For example, there is no result when searching for the data category /element working status/. However, when

| | |
|---------------------------------|---|
| Concept level | |
| Relations | Generic: feminine Generic: masculine Generic: neuter Generic: other gender |
| PID | http://datcatinfo.termweb.eu/datcat/DC-245 |
| Implemented as | pick list |
| DC ID | grammaticalGender |
| Justification | Grammatical gender of nouns, pronouns, and adjectives is a standard feature of many Indo-European languages |
| Origin | ISO 12620:1999 |
| Profile | Terminology |
| Standard | ISO 10241, TBX-Basic |
| Language level | |
| English | |
| Data category name level | |
| Name | grammatical gender |
| Status | standardized |
| Definition | A grammatical category that indicates grammatical relationships between words in sentences. |
| Definition source | ISO 12620 [sic.] |
| Example | The association of a concept with a gender can be based on word form or other factors and is not necessarily related to natural gender. In German, the noun "Brücke" (bridge) is feminine, but the French noun "pont" is masculine. |
| Example source | Document recent studies |
| Explanation | The concept of gender varies from language to language and is not a universal feature of all languages. |
| Explanation source | ISO 12620:1999; SALT |
| Reviewer comment | (Author) – (comment). |

Table 3: Data category specification in DatCatInfo

| | |
|-------------------------|---|
| Code | 0112/2///12620#ZZZ003 |
| Version | 001 |
| Revision | 01 |
| IRDI | 0112/2///12620#ZZZ003#001 |
| Preferred name | grammatical gender |
| Definition | A grammatical category that indicates grammatical relationships between words in sentences. |
| Note | |
| Definition source | ISO 12620:1999 |
| Proposal date | 2020-11-23 |
| Version initiation date | 2020-11-23 |
| Version release date | 2020-11-23 |
| Revision release date | 2020-11-23 |
| Responsible committee | ISO/TC 37/SC 3 |
| Version history | 001-01 (2020-11-23 07:19:36 by BATCH 00000001) standard |

Table 4: Potential description of the data category /grammatical gender/ in IEC CDD

searching for one of its permissible values such as /starter element/ a link to the full specification of /element working status/ appears.

- Etc.

ISO 12620:2019, a new attempt

As stated, in 2011, three years before ISOcat's shutdown, LISA, co-owner of the exchange standard TBX, became insolvent and was disbanded. These two incidents made it necessary to review ISO 12620:2009. The current version ISO 12620:2019 is now a much more slimline framework than its predecessor and all references to ISOcat and LISA have been removed. However, the standard was tailored around and very much focused on private, non-ISO data category repository DatCatInfo, thus conferring a quasi-normative character to a repository that is not at its best either. As a matter of fact, shortly after ISO 12620:2019 was published, ISO launched a new standard project series in 2020 in order to overcome ISO 12620:2019's flaws. The future standard will be subdivided in two parts: ISO 12620-1 "Management of terminology resources – Data categories – Part 1: Specifications" [14] and ISO 12620-2 "Management of terminology resources – Data categories – Part 2: Repositories" [15] where DatCatInfo will be referred to as just one of many potential ISO-compliant implementations of the standard.

Common Data Dictionary IEC CDD

Back in the early 2000s, and in parallel to the aforementioned ISO efforts, the International Electrotechnical Committee (IEC) started to develop a standardized metadata registry called Common Data Dictionary (CDD) aiming at providing industry with product data classification and product data descriptions to be used in smart manufacturing and Industry 4.0. Initially, IEC CDD addressed content developers in electrical and electronic domains. In fact, IEC CDD was published as an international database standard for electric components (IEC 61360-4 DB "Standard data element types with associated classification scheme for electric components – Part 4: IEC reference collection of standard data element types and component classes" [16]) and its methodology and information model are further described in IEC 61360-1 "Standard data element types with associated classification scheme – Part 1: Definitions – Principles and methods" [17] and IEC 61360-2 "Standard data element types with associated classification scheme for electric components – Part 2: EXPRESS dictionary schema" [18]. However, it soon became clear that the data dictionary approach was attractive to all kinds of industry sectors, even beyond electrotechnical product data.

Consequently, ISO jumped on the bandwagon and the scope of IEC CDD was extended converting it to a reference data registry for common concepts of all IEC and ISO domains. In this context, reference data refers to collections of items managed in a database such as graphical symbols, sets of definitions, sets of dimensions, dictionaries of data element types, etc., a description that also applies to terminological data categories.

Given the potential usefulness of IEC CDD as a registry for standardized terminological data categories, a test pilot was set up. Table 4 on the previous page illustrates the description of the data category /grammatical gender/ in an IEC CDD entry.

These entries can be linked to form small classifications, ontologies, or concept systems as depicted in Fig. 1. IEC CDD entries are multilingual. Apart from the English language section reference data can also be described in French, German, Japanese, and Chinese language sections.

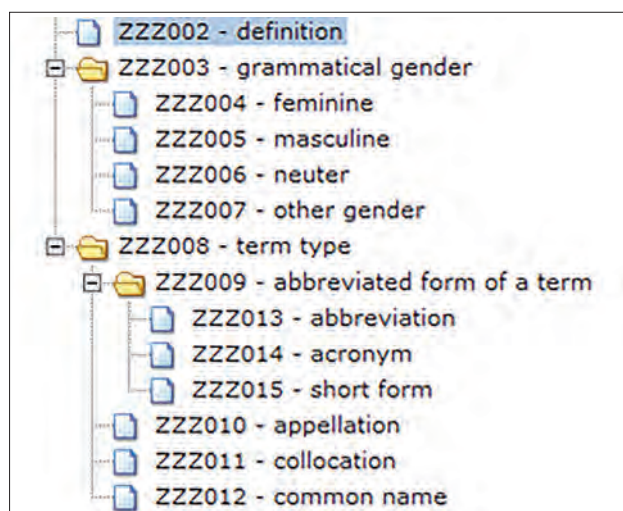


Fig. 1: Terminological data category classification in IEC CDD

As stated, IEC CDD started out as a database standard to support digital interoperability in the electrotechnical product data domain and segued into a domain-independent ISO/IEC reference data registry. Furthermore, in the context of society's and industry's digital transformation, ISO and IEC seek to keep pace with fundamental requirements regarding the development and use of standards shifting from today's PDF-based standards model to machine-interpretable (smart) content. The main objectives concerning the development of smart standards include semantic modelling, self-learning analysis and validation, interruption-free data flow within the value chain, and automatic question answering and predictive content supply. Most of the ISO and IEC standards include valuable micro-content and reference data ready to be reused in an automated, interactive way in industry applications. And terminological data categories would also fit the purpose. Possible scenarios

range from designing terminology databases, knowledge databases, thesauri, taxonomies, and classifications to terminology extraction, data annotation, data mining, data validation, data transformation, ultimately all kinds of smart and AI-driven environments that use smart terminology layers in their applications.

Outlook

At present, ISO and IEC are still examining the requirements for a common future smart standards development architecture. In this architecture, a centralized reference data registry is only one out of many cornerstones on the long way to operability that need to be taken into account such as the legal framework, governance and procedure solutions, commercial policy, third party providers, and user's adoption. Possible technical implementations for a centralized reference data registry range from out-of-the-box options for certain use cases relying on open-source processes and standard tooling to common, proprietary solutions for all ISO/IEC standardization fields whilst ensuring a viable business model. For a potential terminological data category ISO database standard two business models would be imaginable: a robust, rigorous (but not academically overloaded), fully operational, continuously updated, resource-sustainable, authoritative high-quality ISO database standard free of charge including all necessary distribution, user and application features and an optimized version with extra features distributed through an external provider such as ECLASS, for example.

It is quite unlikely that standardized terminological data categories will reappear in form of an ISO database standard in the short-term also due to unfortunate discussions within ISO/TC 37/SC 3 trying to confer on non-ISO DatCatInfo a worldwide SSoT status whilst hampering efforts to move towards an industry-orientated ISO database standard on terminological data categories. Nonetheless, in the mid- and long-term almost all ISO and IEC committees will provide standardized reference data through a single centralized reference data registry for industry use. So, the question is not whether it makes sense or not, it is just a question of time and the terminology community would be well-advised to take the lead in this transformation process as it has been and will always be a cutting-edge technology.

Literature

- [1] ISO 16642:2017 "Computer applications in terminology – Terminological markup framework". Geneva: ISO.
- [2] ISO 12620:1999 "Computer applications in terminology – Data categories". Geneva: ISO.
- [3] ISO 12083:1994 "Information and documentation – Electronic manuscript preparation and markup". Geneva: ISO.
- [4] ISO 30042:2008 "Systems to manage terminology, knowledge and content – TermBase eXchange (TBX)". Geneva: ISO.
- [5] LISA (2008): Systems to manage terminology, knowledge, and content – TermBase eXchange (TBX). https://www.gala-global.org/sites/default/files/migrated-pages/docs/tbx_oscar_0.pdf.
- [6] Reineke, Detlef / Romary, Laurent (2018): SKOS and TBX vocabularies. <https://hal.inria.fr/hal-01883377/file/2019-10-15-skos-tbx-references-v3.pdf>.
- [7] ISO 30042:2019 "Management of terminology resources – TermBase eXchange (TBX)". Geneva: ISO.
- [8] TerminOrgs (2014): TBX-Basic. <https://www.terminorgs.net/TBX-Basic.html>.
- [9] LTAC Global (2022): TBX-Min – Information. <https://www.tbxinfo.net/tbx-min-resources-and-tools>.
- [10] ISO/IEC 11179 series "Information technology – Metadata registries (MDR)". Geneva: ISO.
- [11] ISO 12620:2009 "Terminology and other language and content resources – Specification of data categories and management of a Data Category Registry for language resources". Geneva: ISO.
- [12] LTAC/TerminOrgs (2022): DatCatInfo. <https://datcatinfo.net>.
- [13] ISO 12620:2019 "Management of terminology resources – Data category specifications". Geneva: ISO.
- [14] ISO 12620-1 "Management of terminology resources – Data categories – Part 1: Specifications". Geneva: ISO.
- [15] ISO 12620-2 "Management of terminology resources – Data categories – Part 2: Repositories". Geneva: ISO.
- [16] IEC 61360-4 DB "Standard data element types with associated classification scheme for electric components – Part 4: IEC reference collection of standard data element types and component classes". Geneva: IEC.
- [17] IEC 61360-1 "Standard data element types with associated classification scheme – Part 1: Definitions – Principles and methods". Geneva: IEC.
- [18] IEC 61360-2 "Standard data element types with associated classification scheme for electric components – Part 2: EXPRESS dictionary schema". Geneva: IEC.



Dr. Detlef Reineke, philosopher, translator, terminologist, and programmer, is professor at the University of Las Palmas de Gran Canaria. He has been involved in national and international standardization for over 10 years, the last three years as chair of ISO/TC 37/SC 3 "Management of terminology resources". Detlef Reineke participated in the development of several fundamental terminology standards such as ISO 26162, DIN ISO 26162, ISO 30042, ISO 16642, ISO 12620, and ISO 5078, and is also active as data engineer and consultant in the language industry.

Kontaktadresse
data-engineering@posteo.de

Das „Wundermittel“ E-Learning und seine Eignung für die Terminologearbeit

Lily Arnold

E-learning is a key element of modern education and training, which has already become established in numerous companies in various industries. The question now under discussion is whether and to what extent this learning medium is also suitable for terminology work. How can terminology work be facilitated and perhaps even improved by e-learning?

Keywords: terminology training material, terminology e-learning, terminology education, e-learning benefits

E-Learning ist als Mittel der Aus- und Weiterbildung von Mitarbeiterinnen und Mitarbeitern nicht mehr aus den meisten Unternehmen wegzudenken. Laut Umfragen bedient sich der Großteil deutscher Unternehmen technologisch unterstützter Lernplattformen [2; S. 7]. Gerade in seinen Anfangszeiten um die Jahrtausendwende wurde E-Learning häufig als „Allzweckwaffe“ oder „Wundermittel“ betrachtet, mit dem sich sämtliche firmeninternen Probleme in Wohlgefallen auflösen. Geld- und Zeitersparnis, aber auch die Kompetenzsteigerung der Mitarbeitenden galten als quasi garantiert.

Unternehmen, die E-Learning hingegen nicht in ihre Weiterbildungsprozesse integriert haben und dies auch in Zukunft nicht vorhaben, nennen in Umfragen verschiedene Gründe: Mangelnde Erfahrung im Umgang mit E-Learning, nicht lohnende Produktionskosten, Datenschutzprobleme und Missbrauch durch die Nutzerinnen und Nutzer führen die Liste an [2; S. 12].

Ziel dieses Artikels ist, einen objektiven Blick auf die Eignung von E-Learning zu werfen, und zwar aus der Perspektive der Terminologearbeit. Bevor geklärt werden kann, ob und inwieweit Prozesse innerhalb der Terminologearbeit durch E-Learning unterstützt werden können, muss zunächst ein einheitliches Verständnis des Begriffs „E-Learning“ geschaffen werden. Kein leichtes Unterfangen, denn die Bezeichnung wird im Feld der Informations- und Kommunikationstechnologien (IKT) sehr unterschiedlich und zum Teil generisch verwendet [4; S. 35].

E-Learning steht für electronic learning, auf die direkte Übersetzung beschränkt bedeutet das also: Lernen mit elektronischer Unterstützung. Diese stark vereinfachte

Betrachtung lässt jedoch die technologischen Möglichkeiten „echter“ E-Learning-Anwendungen außer Betracht und unterschlägt die Tatsache, dass E-Learning-Anwendungen auch nach einer speziellen didaktischen Aufbereitung verlangen. Einen Überblick über die Terminologie im Bereich Informationstechnologien bietet die ISO/IEC 2382-36:2019, sie liefert zudem Definitionen zu den verschiedenen Begriffen rund um technologisch unterstütztes Lernen [5]. Hier wird E-Learning als durch IKT erleichtertes Lernen definiert.

Bevor sich ein Unternehmen für die Einführung von E-Learning-Technologien entscheidet, sollte sichergestellt werden, dass innerhalb des Betroffenenkreises dasselbe Begriffsverständnis vorliegt. Dabei können bestimmte Begriffsinhalte ausgeschlossen oder besonders hervorgehoben werden, solange diesbezüglich eine einheitliche und transparente Entscheidung getroffen wird. In diesem Artikel wird die Bezeichnung „E-Learning“ verwendet, die für das Lernen unter Verwendung speziell didaktisch aufbereiteter Lernmedien steht.

Grenzen und Möglichkeiten

Wie viele Technologien, die zur Vereinfachung verschiedener Arbeitsprozesse eingesetzt werden können, bietet E-Learning durchaus Chancen und Möglichkeiten. Gerade in den vergangenen zwei Jahren musste gezwungenermaßen verstärkt auf digitale Schulungsangebote zurückgegriffen werden, da die Corona-Pandemie starken Einfluss auf die Arbeitswelt genommen hat. Seit dem ersten Lockdown 2020 ist die Zahl der im Homeoffice arbeitenden Personen von 12 auf 25 % gestiegen [6]. Diese Entwicklung bestärkt den Einsatz von E-Learning-Angeboten.

E-Learning hat jedoch auch seine Grenzen. Der Einsatz sollte also gut geplant und reflektiert sein und sich nicht nur auf die bekannten Vorteile wie Zeitersparnis, die Unabhängigkeit von Lehrpersonen und die Unabhängigkeit von Ort und Zeit der Durchführung stützen. E-Learning ist nur so lange ein Gewinn, wie sich die technischen Möglichkeiten auch als tatsächlicher Mehrwert für den jeweiligen Anwendungsfall erweisen.

Eignung für bestimmte Anwendungsbereiche

Zunächst muss für jedes geplante E-Learning-Modul der Anwendungsbereich definiert und im Anschluss die grundsätzliche Eignung des Mediums „E-Learning“ für den jeweiligen Anwendungsbereich überprüft werden. Trotz der umfassenden Möglichkeiten bietet es sich in manchen Wissensbereichen nicht an, auf E-Learning-Lernmedien zurückzugreifen. Gerade in geisteswissenschaftlichen Bereichen, die sich in der Wissensvermittlung stark auf diverse Formen von Diskurs und Kommunikation stützen, sind E-Learning-Anwendungen nicht unbedingt geeignet [9; S. 4]. Sie können zwar ein gewisses Maß an Flexibilität erreichen und auf die individuellen Aktionen der Lernenden reagieren, dabei stehen sie jedoch nicht als Kommunikationspartner zur Verfügung.

Eine Umfrage zu den erwarteten Lernzielen von E-Learning-Anwendungen ergab, dass sowohl Anwenderinnen und Anwender als auch Personalverantwortliche das Aneignen und Auffrischen von Fachwissen als Hauptziele von E-Learning betrachten und auch die Steigerung von Selbstmanagementfähigkeiten und Organisationskompetenz von einer solchen Anwendung erwarten. Das Aneignen von Führungswissen und die Steigerung von Teamfähigkeit werden hingegen nur von höchstens einem Drittel der Befragten als Ziel angegeben [2; S. 10, nach Littig 2002[7]]. Die Erwartungshaltung ist hier also kongruent zu den häufig bemängelten Grenzen des E-Learnings, und

diese Einschätzung ist nicht unbegründet. Aus psychologischem Blickwinkel werden viele Soft Skills, die für Teamfähigkeit und Führungsqualifikationen unerlässlich sind, als angeborene Fähigkeiten klassifiziert. Sie stehen damit im Gegensatz zu Hard Skills, also den erlernten Fertigkeiten. E-Learning-Programme können Fertigkeiten vermitteln, Fähigkeiten jedoch allenfalls verstärken [1; /2; S. 140].

Auch die Überprüfung des Lernfortschritts in Form von Wissensabfragen und zugehörigem Feedback kann über E-Learning nur in begrenztem Maß erfolgen. E-Learning-Module bedienen sich zu diesem Zweck häufig sogenannter Assessments, die z. B. in Form von Quizzeinheiten umgesetzt werden [4; S. 43 /5]. Diese Assessments werden jedoch auch häufig kritisiert, denn die als individuell und abwechslungsreich angepriesenen Wissensabfragen bieten oft nur Multiple-Choice-Fragen oder andere, simple Auswahlfrageformen an. Damit bleiben sie weit hinter den Möglichkeiten zurück, die der multimediale Charakter von E-Learning eigentlich bietet bzw. bieten sollte (so eine Studie der Stiftung Warentest [2; S. 13]). Damit nicht selbst denkende E-Learning-Module überhaupt Feedback geben können, müssen die Antworten auf vorprogrammierte Fragen in einem ebenfalls vorgegebenen Erwartungsbereich liegen. Die Kreativität der Nutzerinnen und Nutzer ist jedoch nicht planbar. Dadurch können individuelle, stark variierende Antworten auf offene Fragen nur schwer überprüft und Transferleistungen schlecht abgefragt und verifiziert werden [2; S. 21].

Um E-Learning tatsächlich als geeigneten Ersatz für eine firmeninterne oder auch -externe Lehrperson einsetzen zu können, muss das E-Learning-Modul mehr als nur eine Sammlung von Inhalten sein. Der sogenannte Content will zwar gut durchdacht sein, doch dessen didaktische Aufbereitung (und Nachbereitung in Form von Feedback) ist mindestens genauso wichtig. Eine reine Aneinanderreihung

Anzeige

edition

Fachzeitschrift für Terminologie

Alle Ausgaben der edition finden Sie als PDF in unserem kostenlosen Online-Archiv. Perfekt zum Stöbern, Nachschlagen und Weitergeben.

www.dttev.org/edition



von Informationen unterschlägt nicht nur die mannigfaltigen Möglichkeiten von E-Learning, sondern verfehlt auch den Zweck, E-Learning als erfolgreiche Form der Weiterbildung weitgreifend einsetzen zu können.

Ausbildung verschiedener Kompetenzebenen

In jedem Lernprozess spielt der Wissensbegriff eine tragende Rolle, wobei Wissen – ebenso wie der Begriff des E-Learnings – unterschiedlich definiert werden kann. Einerseits kann hier der reine Content gemeint sein, also das sich anzeigende Faktenwissen, andererseits lässt diese Definition einmal mehr die pädagogische Sichtweise außen vor. Aus diesem Grund hat sich in diesem Zusammenhang der Begriff der *Kompetenz* etabliert. Kompetenz beinhaltet nicht nur das reine Faktenwissen, sondern auch die Fähigkeit, dieses Wissen aufzunehmen und anzuwenden. Man unterscheidet dabei mindestens drei verschiedene Kompetenzebenen: Kennen (Inhalte wiedergeben), Verstehen (Inhalte selbst erklären) und Handeln (Inhalte anwenden und Verantwortung übernehmen) [9; S. 3].

Das Erreichen höherer Kompetenzebenen, die über die simple Wiedergabe von erlerntem Faktenwissen hinausgehen, erfordert mehr Kontext sowie kommunikative Unterstützung [9; S. 4]. Durch das Ausbleiben von individuellem, differenziertem Feedback und das Fehlen einer Plattform, auf der Sachverhalte in eigenen Worten erklärt werden können, entfallen jedoch die wichtigsten Faktoren zum Ausbilden von tiefgehendem Verständnis.

Handlungskompetenzen können zwar über E-Learning-Anwendungen ausgebildet werden, dies ist allerdings nur begrenzt bzw. unter einigem Aufwand möglich. E-Learning-Anwendungen nutzen zu diesem Zweck Simulationen; diese sollen die Durchführung verschiedener Handlungssequenzen schulen. Bekannte Beispiele sind Simulationen chirurgischer Eingriffe für angehende Ärztinnen und Ärzte oder IT-Schulungen, bei denen die Lernenden in die Nutzung diverser digitaler Anwendungen eingeführt werden. Der Erstellungsaufwand für solche Simulationen ist allerdings wesentlich höher als der für theoretische Wissensvermittlung und Assessments in Form von Multiple-Choice-Fragen oder ähnlichen Frageformen.

E-Learning im Terminologiebereich

Schließlich stellt sich die Frage, wie sich die vorangegangenen Betrachtungen auf den Einsatzbereich der Terminologie anwenden lassen. Dazu muss zunächst differenziert werden, um welchen Anwendungsbereich innerhalb der Terminologiewissenschaft oder -arbeit es sich im Speziellen handelt (beispielsweise die Erarbeitung erster Datenbestände, deren Erweiterung und Kontrolle oder auch das Schaffen von Akzeptanz innerhalb des Unternehmens). Da

die Terminologiearbeit sehr viele Aufgabenfelder umfasst, ergeben sich auch verschiedene denkbare Anwendungsfälle, für die der Einsatz von E-Learning als Lernmedium jeweils evaluiert werden muss.

Vermittlung von Grundlagen

Wie bereits festgestellt, sind E-Learning-Anwendungen für die Ausbildung der niedrigen Kompetenzebenen sehr gut geeignet und bieten sich für die Vermittlung von reinem Faktenwissen grundsätzlich an. Gerade für die unterste Kompetenzebene des Kennens ist häufige Repetition eine sinnvolle Lernmethode. Repetitive Übungsformen lassen sich über E-Learning gut realisieren, da Lernende so viele Stoffwiederholungen durchführen können, wie sie für nötig erachten, ohne dabei auf die Zuwendung einer Lehrperson angewiesen zu sein. Das Feedback für die Lernenden in Form von Assessments kann auf dieser Ebene noch von E-Learning-Anwendungen gewährleistet werden und ist somit ebenfalls unabhängig von Lehrpersonen [9; S. 4]. Von dieser Zeit- und damit einhergehenden Kosteneinsparung profitiert natürlich auch die Ausbildung im Terminologiebereich. Auch die zweite Kompetenzebene (Verständnis) kann durch E-Learning erlangt werden, sofern die Module qualitativ hochwertig sind und der Content systematisch aufbereitet wird. Auf diese Weise werden die Lernenden intellektuell angesprochen und entwickeln im besten Fall den Ehrgeiz, den Stoff nicht nur aufzunehmen, sondern auch begreifen zu wollen [9; S. 3].

Bei der Einarbeitung neuer Mitarbeiterinnen und Mitarbeiter ist E-Learning ein valides Mittel, diese in das Thema Terminologie einzuführen, Interesse zu wecken oder einen gewissen Kenntnisstand der Grundlagen zu garantieren. Gerade für Quereinsteigende in diesem Bereich oder Mitarbeitende eines Unternehmens, in dem Terminologie neu eingeführt werden soll, ist das Schaffen von Grundlagenwissen unerlässlich.

Beschränkt sich der zu vermittelnde Content also auf terminologische Grundlagen wie die Differenzierung von Begriff, Bezeichnung und Benennung, den Aufbau und Vergleich qualitativ hochwertiger Definitionen oder den grundlegenden Ablauf von Terminologiearbeit, ist dies über E-Learning durchaus sinnvoll umsetzbar. Die einzelnen Themengebiete bedingen einander zudem, so wird die Bedeutung einzelner Schritte in der praktischen Terminologiearbeit erst dann ersichtlich, wenn die theoretischen Grundlagen bekannt sind. Mitarbeitende mit profundem Grundlagenwissen wissen daher um die Wichtigkeit jedes einzelnen Arbeitsschrittes und überspringen keine vermeintlich vernachlässigbare Aufgabe.

Aber auch außerhalb des direkten Anwendungsbereichs kann E-Learning der Terminologiearbeit zuträglich sein.

Schulungen zur Relevanz von kontrollierter Terminologie können auch in fachfremden Abteilungen durchgeführt werden, deren einzige Verbindung zur Terminologearbeit ist, dass sie die dort getroffenen Vorgaben einhalten müssen. Dadurch kann die Akzeptanz kontrollierter Sprache gesteigert werden, was wiederum die Arbeit des Terminologenteams enorm erleichtert, da es dann weniger persönliche Überzeugungsarbeit leisten muss. Auch nachfolgende Arbeitsschritte, beispielsweise im Content Management oder im Übersetzungsmanagement, profitieren von Terminologiemanagement. Das Bewusstsein um die weitreichenden Vorteile und der entsprechende Rückhalt kann gerade in konservativen Unternehmen bei der Neueinführung von Terminologiemanagement einen entscheidenden Unterschied bedeuten und sogar schon zu Beginn des Prozesses darüber entscheiden, ob ein Terminologieprojekt tatsächlich aufgesetzt werden kann [3; S. 35].

IT-Anwenderschulung

Ein weiterer sinnvoller Anwendungsbereich im Terminologiebereich ist die IT-Anwenderschulung. Gerade die Einarbeitung im Umgang mit Terminologieverwaltungstools lässt sich über E-Learning sehr gut realisieren, da der Anwendungsbereich und die Lernumgebung ähnlich sind. Die Lernenden setzen sich also schon bei der Bearbeitung der E-Learning-Anwendung mit einer vergleichbaren Toolumgebung auseinander. Einige E-Learning-Erstellungstools erlauben beispielsweise die Einbettung von bewegten Bildschirmaufnahmen und Mauspfaden, wodurch Mausbewegungen nachvollzogen und im Anschluss selbstständig durchgeführt werden können [11; S. 112 ff]. Simulationen dieser Art erlauben neuen Mitarbeitenden, Terminologieverwaltungssysteme kennenzulernen, ohne dass auf Testsysteme in den Tools selbst zurückgegriffen werden muss oder unnötig viele Benutzerzugänge angelegt werden müssen. Sensible Datenbestände sind damit sicher vor unzureichend ausgebildeten Nutzerinnen und Nutzern, während diese dennoch eine möglichst realitätsnahe Ausbildung erhalten können. Zudem sollte erwähnt werden, dass Terminologieverwaltung je nach Projekt und Unternehmen stark variieren kann (je nach Anbieter, Datenstruktur, Datenkategorien, etc.) [3; S. 100] und somit auch erfahrenere Terminologinnen und Terminologen erst in ein neues Terminologieverwaltungssystem eingearbeitet werden müssen.

Ausbildung oder Einarbeitung von Führungspersönlichkeiten

Wie zuvor bereits ausgeführt, ist E-Learning nur begrenzt zur Vermittlung von Soft Skills und zur Ausbildung persönlicher Kompetenzen geeignet. Laut dem Rat für Deutschsprachige Terminologie (RaDT) sind Terminologinnen und Terminologen aber auf genau solche Fähigkeiten an-

gewiesen. Neben der fachlichen Kompetenz müssen sie auch über ergänzende Fähigkeiten verfügen, die sie erst zu wirklich guten und erfolgreichen Vertreterinnen und Vertretern ihres Fachs machen: Kommunikative und soziale Kompetenz, Durchsetzungsvermögen und Organisationskompetenz seien hier nur als Beispiele genannt [8; S. 7]. Empathische Eigenschaften und das Verhalten gegenüber bzw. mit anderen kann man am besten im direkten Umgang mit Menschen entwickeln und nicht für sich allein bei der Bearbeitung eines elektronischen Moduls erlernen. Auch den für die Durchsetzungskraft nötigen Respekt erarbeiten sich Terminologinnen und Terminologen nur in der Zusammenarbeit. In diesem Anwendungsfall kann E-Learning also nicht als vollwertiger Ersatz für klassische Schulungen herhalten, die auf die Interaktion von Teilnehmenden und Lehrpersonen bzw. von Teilnehmenden untereinander angewiesen sind.

Terminologiebestände

Auch für die Vermittlung der Terminologien selbst eignet sich E-Learning nur bedingt. Zum einen sind Terminologien ständiger Änderung unterzogen, woraus sich ein hoher Aktualisierungsaufwand für die entsprechende E-Learning-Anwendung ergibt. Da viele Unternehmen in Umfragen genau diesen Aktualisierungsaufwand als Grund nennen, E-Learning gar nicht erst einzusetzen, ist vergleichsweise langlebiges Faktenwissen grundsätzlich besser als E-Learning-Content geeignet, als es kurzlebige Datenbestände sind.

Zum anderen müssen Terminologiebestände im Normalfall nicht erlernt werden, um deren konsistenten Gebrauch in einem Text zu gewährleisten. Die Überprüfung der ordnungsgemäßen Verwendung von Terminologie ist in der Regel technisch unterstützt, Autorinnen und Autoren müssen also nicht alle vorgegebenen Bezeichnungen im Kopf haben. Allerdings können reine Terminologiebestände durchaus auch als sinnvolle Lerngrundlage dienen, um mit der Fachsprache des entsprechenden Fachbereichs vertraut zu werden, da sie im Idealfall nicht nur die Bezeichnungen, sondern auch die Definitionen und begrifflichen Zusammenhänge enthalten.

Anwendungsbeispiel

Es existieren bereits verschiedene Online-Lernangebote für den Bereich Terminologearbeit, meist in Form von Trainings mit Trainern und Webinaren, also nicht ganz unabhängig von Lehrperson und Lerngruppe (z. B. CTM terminology courses [10]). Bei solchen Anwendungsfällen handelt es sich aber eher um eine Mischung aus E-Learning, wie es in diesem Artikel definiert wurde, und webbasiertem Training ähnlich einer Online-Schulung.

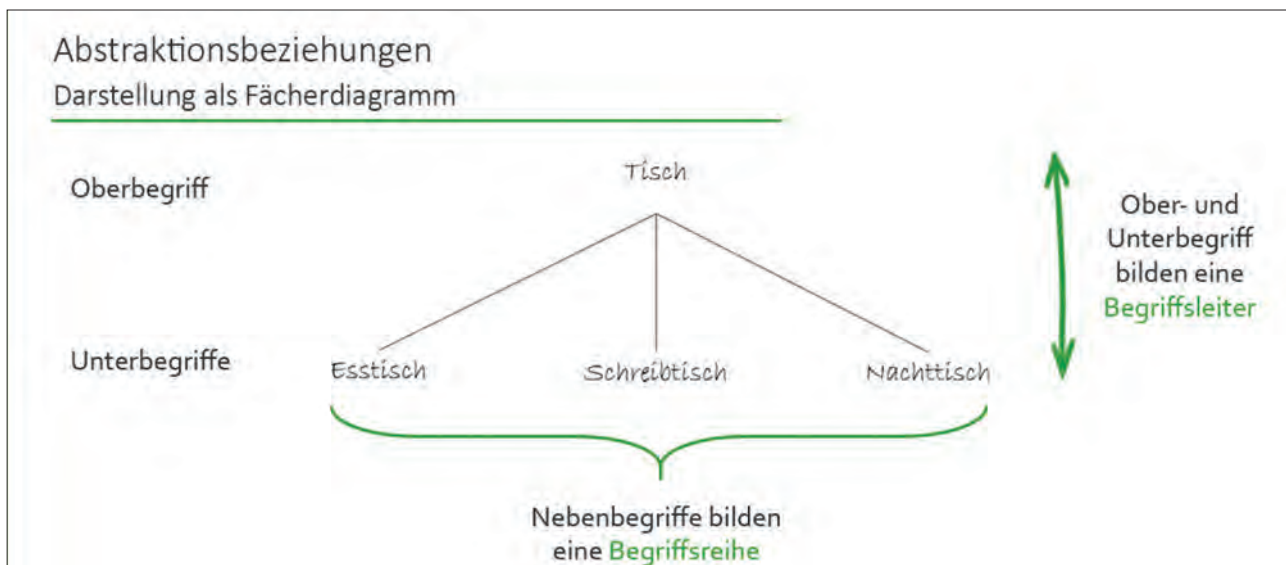


Abb. 1: Grundlagen Begriffssysteme (Auszug aus dem u.g. E-Learning <http://www.technischeredaktion.com/Thesis-Arnold/>)

Lückentext

Darstellung hierarchischer Begriffsbeziehungen

Abstraktionsbeziehungen werden als Fächerdiagramm* dargestellt, Bestandsbeziehungen als Klammerdiagramm*. Zur besseren Auffindbarkeit einzelner Begriffe können Notationen in jede Art von Begriffssystem aufgenommen werden. Einteilungskriterien werden hingegen nur in Abstraktionssysteme* dargestellt.

Abb. 2: Assessment (Auszug aus dem u.g. E-Learning <http://www.technischeredaktion.com/Thesis-Arnold/>)

Ein unabhängiges E-Learning-Modul zu den hier erörterten Themen und in der wie oben definierten Form wurde im Rahmen der 2021 unter dem Titel „Konzeption und Umsetzung eines vorlesungsbegleitenden E-Learning-Moduls zum Themengebiet ‚Terminologiewissenschaft und -arbeit‘ zur Unterstützung des selbstbestimmten Lernens“ entstandenen Bachelorarbeit erstellt. Das E-Learning-Modul vermittelt die Grundlagen der Terminologiewissenschaft und -arbeit und richtet sich vor allem an Studierende dieses Fachs, die erstmals mit diesen Grundlagen in Berührung kommen. Das integrierte Assessment dient der selbstständigen Wissensüberprüfung, das Modul kann außerdem beliebig oft durchgeführt werden. Auf eine Lehrperson sind die Lernenden nicht angewiesen.

Die didaktische Aufbereitung der Inhalte ist im Rahmen des Moduls, das für die Bachelorthesis erstellt wurde, speziell an die Zielgruppe und das Lernumfeld Hochschule angepasst. Der Anwendungsbereich wurde also auch hier vor der Erstellung genau definiert, um eine passende Wissensaufbereitung zu ermöglichen.

Auch wenn in den vorausgegangenen Ausführungen vom Arbeitsumfeld und nicht vom Studium die Rede ist, lässt

sich das Beispiel trotzdem zu einem gewissen Grad übertragen. Gerade in den Abteilungen, die indirekt an der Terminologiearbeit beteiligt sind, ist häufig vergleichbar wenig Grundlagenwissen vorhanden, das durch ein solches E-Learning-Modul aufgebaut werden kann.

Das E-Learning-Modul ist öffentlich zugänglich und kann über folgenden Link erreicht werden:

<http://www.technischeredaktion.com/Thesis-Arnold/>

Fazit

Um schließlich noch einmal auf die Ausgangsfrage zurückzukommen: E-Learning in der Terminologiearbeit? Bei guter Vorbereitung und basierend auf einem sinnvollen didaktischen Konzept können E-Learning-Anwendungen definitiv einen Mehrwert für die Terminologiearbeit darstellen. Für die Vermittlung von Grundlagen können sie als immer wieder verwendbares Schulungstool eingesetzt werden, bei der Akzeptanzbildung innerhalb des Unternehmens entlasten sie die Terminologinnen und Terminologen und erleichtern damit die Verbreitung von Terminologiebeständen und Schreibregeln. Und auch die Einpflegung neuer Datenbestände in Terminologiedatenbanken kann in

Anwenderschulungen trainiert werden. Von einem qualitativ hochwertigem E-Learning profitiert dadurch schlussendlich nicht nur die Terminologearbeit, sondern das gesamte Unternehmen.

Literatur

- [1] Bundesministerium des Inneren (o.J.): „Chancen und Grenzen von E-Learning und Blended Learning in der dienstlichen Fortbildung“. <elearning_Innenministerium.pdf > [Stand: k.A. Zugriff: 07.03.2022]
- [2] DGFP (Hrsg.) (2004): E-Learning in Unternehmen: Konzepte – Einsatzmöglichkeiten – Qualitätskriterien. Düsseldorf: DGFP
- [3] Drewer, Petra / Schmitz, Klaus-Dirk (2017): Terminologiemanagement: Grundlagen – Methoden – Werkzeuge. Berlin: Springer Vieweg
- [4] Handke, Jürgen / Schäfer, Anna Maria (2012): E-Learning, E-Teaching und E-Assessment in der Hochschullehre – Eine Anleitung. München: Oldenbourg Wissenschaftsverlag
- [5] ISO/IEC 2382-36 (Juni 2019): Information technology – Vocabulary – Part 36: Learning, education and trainings
- [6] Landeszentrale für Politische Bildung Baden-Württemberg (2021): „Wie verändert Corona die Arbeitswelt?“. <Folgen von Corona | Digitalisiert Corona die Arbeitswelt? (lpb-bw.de)> [Stand: Juni 2021. Zugriff: 05.04.2022]
- [7] Littig, Peter (2002): Klug durch E-Learning? Eine Marktstudie der

DEKRA Akademie. Bielefeld

- [8] RaDT (2020): „Berufsbild 2.0“. <RaDT-Berufsprofil 2-0.pdf> [Stand: April 2020. Zugriff: 07.03.2022]
- [9] Sesnik, Werner (2003): „Grenzen des E-Learning“. <eLearning_Grenzen.pdf> [Stand: Oktober 2003. Zugriff: 07.03.2022]
- [10] Termnet (2017): „This is for you: Online terminology courses“. <Online Courses * (termnet.eu)> [Stand: 2017. Zugriff: 31.03.2022]
- [11] Uhrig, Martin (2018): Adobe Captivate 2019 – Erfolgreich(e) Screen-casts und E-Learning-Anwendungen erstellen. North Charleston: Create Space



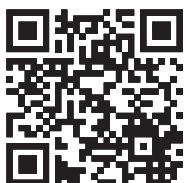
Lily Arnold hat im Sommer 2021 ihren Bachelor in Kommunikation und Medienmanagement an der Hochschule Karlsruhe (HKA) abgeschlossen. In ihrer Thesis hat sie sich eingehend mit den Grundlagen von Terminologiewissenschaft und -arbeit und deren Umsetzung in einem E-Learning-Modul beschäftigt. Neben dem Studium ist sie als Technische Redakteurin für ein kleines Software-Unternehmen tätig.

Kontaktadresse
lily.arnold99@gmail.com

Anzeige

Die gds-eigene neuronale Translation Engine PLURAVOX spricht „Maschinenbau“.

www.gds.eu/de/fachuebersetzungen



Redaktionslösungen
Content Delivery
Technische Dokumentation
Fachübersetzungen
CE-Support
Consulting | Projekte | Customizing

sprachenwelt

Je mehr Daten desto schlechter?

Terminologietreue bei domänenspezifischen NMÜ-Engines

Frauke Hellerich

Achieving terminology adherence in domain-adapted NMT engines remains an issue despite the generally high output quality. This paper shows how the type, amount, quality, and preparation of training data can influence the correct translation of terminology in NMT output of two domain-adapted and customized NMT engines for technical content.

Keywords: domain-adapted NMT, NMT customization, terminology consistency in NMT, clean training data, NMT output assessment

In der Recherche zu dieser Untersuchung las ich in einem Paper des spanischen MÜ-Forschers Mikel Forcada den Satz „Involvement can only occur through understanding“ [1]. Diese Aussage brachte mich zum Nachdenken: Als Übersetzerin und Terminologin wollte ich herausfinden, wie ich Einfluss auf die Terminologietreue einer fachspezifischen MÜ-Engine nehmen könnte, und lernen, ihren Output richtig zu evaluieren.

Auch bei neuronalen maschinellen Übersetzungssystemen (NMÜ), mit denen sich Engines für ein bestimmtes Fachgebiet trainieren (Domänenadaption) und darüber hinaus auf unternehmens- und projektspezifische Stil- und Terminologievorgaben anpassen lassen (Customization), kann die Terminologietreue im Output je nach Qualität der Trainingsdaten variieren und die Sicherung terminologischer Konsistenz ein anspruchsvolles Unterfangen darstellen. Der CEO der Globalese GmbH, Gabor Bessenyei, brachte es im Kick-off-Meeting für diese Untersuchung so auf den Punkt: „Viele Anbieter, aber auch Kunden bzw. Benutzer haben die Erwartung, dass sie ihre Terminologieliste an ihre Engine übergeben, und dann weiß die Engine plötzlich, dass sie statt ‚Hund‘ ‚Katze‘ übersetzen soll. Das ist natürlich ein schöner Traum [...].“

Bei Globalese wird Terminologie nicht über den Suchen-Ersetzen-Ansatz anhand einer von Nutzer*innen bereitgestellten Terminologieliste im Output integriert. Für domänenspezifische Engines wird vielmehr eine automatische Domänenadaption durchgeführt, bei der ein „Tuning“ auf

die Terminologie anhand von domänenspezifischen Trainingsdaten, die Nutzer*innen selbst beisteuern, stattfindet, um die Terminologietreue im NMÜ-Output zu verbessern [2]. Voraussetzung für dieses Vorgehen ist natürlich terminologisch einheitliches und sauberes Trainingsmaterial. Aufgrund der automatischen Domänenadaption können beim Training einer domänenspezifischen Engine generische, also gemeinsprachliche, Trainingsdaten eingesetzt werden, ohne dass sie ein Rauschen (*noise*) im Trainingsprozess verursachen. Diese Trainingsdaten entsprechen dann zwar nicht zwangsläufig den angesetzten Terminologievorgaben, dienen aber bei einer kleinen Menge domänenspezifischer Trainingsdaten der Vergrößerung des Trainingskorpus. Ziel ist es vielmehr, die sprachliche Wohlgeformtheit (*fluency*) des Engine-Outputs im Allgemeinen zu verbessern. Doch was tun, wenn den bereitgestellten Trainingsdaten auch die nötige terminologische Konsistenz fehlt?

Im Rahmen der Untersuchung für meine Masterarbeit an der Hochschule Würzburg-Schweinfurt in Zusammenarbeit mit der beo Gesellschaft für Sprachen und Technologie mbH und der NMÜ-Anbieterin Globalese GmbH habe ich zwei domänenadaptierte Engines für das Sprachpaar Deutsch-Englisch trainiert und die Terminologietreue im Output der Engines untersucht und verglichen. Die beiden Test-Engines wurden mit dem Globalese-NMÜ-System trainiert, unter Verwendung von domänen- und unternehmensspezifischen Trainingsdaten. Zur Verfügung gestellt wurden diese Daten von beo. Sie bestanden aus

Werkstattliteratur und After-Sales-Texten aus der Automobilbranche. Dabei wurden die Translation-Memory- und Terminologiedaten für die beiden Test-Engines jeweils unterschiedlich zusammengestellt, um herauszufinden, ob sich ein reduziertes, aber dafür terminologisch einheitliches Trainingskorpus positiv auf die Terminologietreue im NMÜ-Output auswirkt oder ob doch die Maxime „je mehr Daten desto besser“ gilt.

Trainingsdaten

Das Translation Memory (TM), das den Hauptanteil der domänenspezifischen Trainingsdaten in der Untersuchung ausmachte, wurde durch ein individuelles Skript für den Einsatz als Trainingsmaterial aufbereitet. Aussortiert wurden:

- Formatierungstags aus den Segmenten
- Segmente, bei denen der Ausgangssatz mehr als doppelt so lang war wie der Zielsatz
- Segmente, die lediglich Satzfragmente enthielten
- Segmente, die ausschließlich aus Abkürzungen bestanden
- Dubletten-Segmente

Bei Varianten, also zwei Segmenten mit identischem Ausgangstext aber unterschiedlichen Übersetzungen, wurde nur das Segment mit dem aktuelleren Änderungsdatum beibehalten. Das trainingsrelevante TM-Korpus umfasste nach der Aufbereitung noch 714.643 Segmente.

Die eingesetzten Terminologiedaten stammten aus vier Terminologiedatenbanken (Marketing, Technik, Recht und Verwaltung). Diese präskriptive Terminologie wird im Unternehmen selbst vornehmlich für die Redaktions- und Übersetzungsarbeit eingesetzt und Begriffseinträge enthalten demzufolge verwendungsbezogene und grammatikalische Metadaten. Darum wurden die Termbank-Daten mithilfe eines Skripts für die Verarbeitung im NMÜ-System aufbereitet,

indem Metainformationen entfernt und lediglich die eindeutig bevorzugten deutschen und englischen Äquivalente eines Termbankeintrags in eine bilinguale tsv-Liste (tab-separated value) extrahiert wurden. Dieser Export umfasste schlussendlich 11.320 Benennungspaare, die entweder aus Ein- und Mehrwortbenennungen, Phrasen oder Abkürzungen bestanden. Die Liste machte den zweiten Teil der domänen- und unternehmensspezifischen Trainingsdaten aus.

Terminologische Filterung

Die Sichtung des aufbereiteten TMs zeigte für viele Begriffe terminologische Unstimmigkeiten, z. B. durch Verwendung veralteter, verbotener oder gar nicht erfasster Benennungen für einen Begriff. Durch diese fehlerhaften Trainingsdaten würde eine NMÜ-Engine falsche Aktivierungsmuster ausbilden, durch die sie lernt, in bestimmten Kontexten ggf. eine verbotene Benennung zu priorisieren. Das Korpus wies für die Ausgangsbenennung „Ladezustand“ beispielsweise in nur 95 Segmenten die bevorzugte Zielbenennung „state of charge“ auf; in 32 Fällen wurde die zugelassene Abkürzung „SOC“ verwendet. In 451 Segmenten fand sich jedoch die verbotene Benennung „charge level“, der eine Engine im Verlauf des Trainings also öfter „begegnen“ würde, wodurch ihre Verwendung in vielen Kontexten wahrscheinlicher würde.

Infolge dieser Beobachtungen wurde das TM-Trainingskorpus mithilfe eines Skripts einer terminologischen Filterung unterzogen. Jedes TM-Segment wurde nach den folgenden Bedingungen geprüft: Ist eine der deutschen Ausgangsbenennungen (AB) aus der extrahierten Terminologieliste (T-Liste) enthalten? Wenn nicht, wurde das Segment beibehalten. War dies jedoch der Fall, wurde geprüft, ob die Zielbenennung (ZB) im Zielsegment entsprechend der Terminologieliste vorhanden war. Wenn dem nicht so war, wurde das Segment verworfen. Andernfalls wurde es beibehalten und alle weiteren Segmente mit derselben deutschen

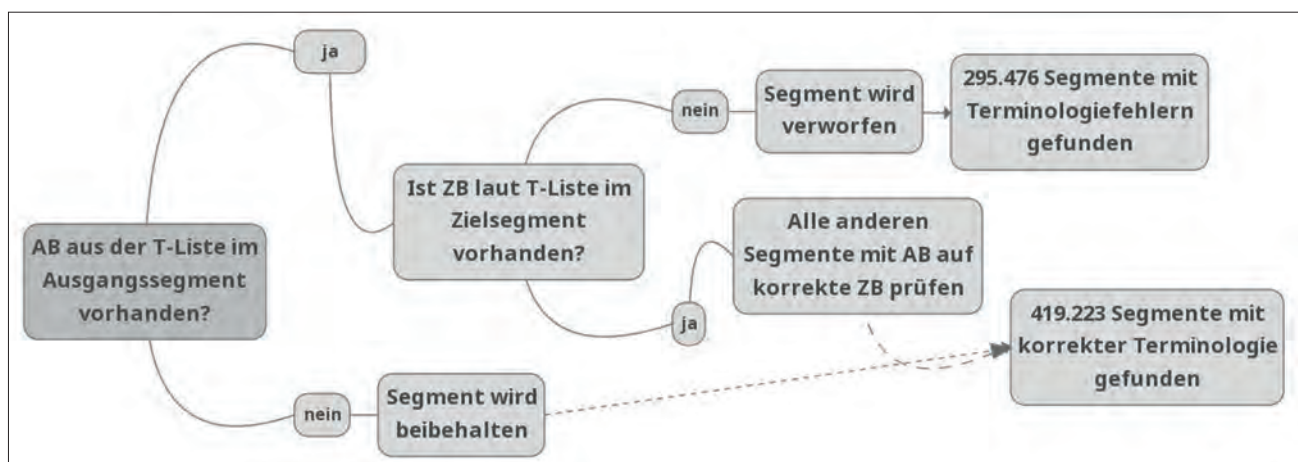


Abb. 1: Entscheidungsbaum zur terminologischen Filterung (eigene Darstellung)

| <input type="checkbox"/> | NAME | SOURCE | TARGET | SEGMENT PAIRS |
|--------------------------|----------------------|--------|---------|------------------|
| <input type="checkbox"/> | term_DEU_ENU.tsv | German | English | 11,320 |
| <input type="checkbox"/> | tm_ENU.tsv | German | English | 714,635 |
| <input type="checkbox"/> | tm_ENU_not_ok_cs.tsv | German | English | 295,498 |
| <input type="checkbox"/> | tm_ENU_ok_cs.tsv | German | English | 419,172 |
| Total | | | | 1,440,625 |

Abb. 2: Trainingskorpora-Anzeige auf dem Globalese-Dashboard (eigener Screenshot)

Ausgangsbennungen wurden auf die Verwendung der entsprechenden englischen Zielbennungen geprüft (siehe Abb. 1).

Nach der terminologischen Filterung standen neben dem vollständigen TM-Trainingskorpus und der Terminologieliste nun, wie in Abb. 1 zu sehen, zusätzlich das gefilterte, terminologisch konsistente Trainingskorpus mit 419.223 Segmentpaaren und das terminologisch inkonsistente „Rest“-Korpus mit 295.476 Segmentpaaren zum Einsatz im Engine-Training zur Verfügung. Nach dem Hochladen auf die Globalese-Plattform sortierte das System noch aus beiden Korpora wenige ungültige Segmente aus (siehe Abb. 2).

Engine-Typen bei Globalese

Je nach Datenlage stehen bei Globalese verschiedene Engine-Typen zur Auswahl: Die sogenannte Stock-Engine steht allen Nutzer*innen ohne Zugabe eigener Daten zur Verfügung. Gabor Bessenyei nennt sie das „Breitspektrum-Antibiotikum“, da sie mit gemeinsprachlichen Daten bzw. Daten aus verschiedenen Fachgebieten aus öffentlich zugänglichen Korpora (Stockdaten) trainiert ist. Eine domänenspezifische Engine mit selbst beigesteuerten Daten ist im Gegensatz dazu das Schmalspektrum-Antibiotikum für einen festgelegten Use Case.

Bei individuellen Trainingsdaten unterscheidet Globalese zwischen Masterdaten (*master corpora*) und Hilfsdaten (*auxiliary corpora*) [3]. Masterdaten machen den domänen- und unternehmensspezifischen Anteil des Trainingsmaterials aus und entsprechen genau den Stil- und Terminologievorgaben

der Inhalte, die mit der Engine übersetzt werden sollen. Da in den meisten Fällen eine zu geringe Menge an hochwertigen Masterdaten zur Verfügung steht, empfiehlt Globalese, diese durch sogenannte Hilfsdaten zu ergänzen. Hilfsdaten sind ebenso individuelle, domänenspezifische Daten, die jedoch nicht zwangsläufig den projektspezifischen Vorgaben entsprechen. Sie können ggf. bestehende Lücken in den Masterdaten füllen und im Engine-Output für eine verbesserte Fluency sorgen. In Bezug auf Terminologie können die Hilfsdaten beispielsweise nicht abgebildete Flexionsformen von Fachbennungen beisteuern. Die Master- und Hilfsdaten können bei zu geringem Volumen zudem durch die gemeinsprachlichen Stockdaten angereichert werden, die Globalese für viele Sprachpaare bereitstellt [3].

Mit der sogenannten Automated In-Domain Adaptation (AIDA) findet im Trainingsprozess für eine domänenspezifische Engine neben dem Tuning auf die Masterdaten dann eine Ergänzung durch die eingesetzten Hilfsdaten bzw. eine Auswahl von und Ergänzung durch die generischen Stockdaten statt. In diesem automatisierten Prozess entsteht dann unter Einsatz von Machine-Learning-Technologien eine Trainingsdatenmenge, die groß genug ist, um eine domänenspezifische Engine zu trainieren, aber klein genug, dass kein Rauschen für die Engine entsteht oder die Priorität der Masterdaten nicht überschrieben wird [2].

Ein typisches Szenario wäre beispielsweise ein TM mit 100.000 Segmenten aus der technischen Dokumentation, das als Masterkorpus genutzt werden soll. Ergänzend wird

ein weiteres TM mit 50.000 Segmenten aus dem Marketing als Hilfskorpus eingesetzt. Um eine domänenspezifische Engine zu trainieren, mit der Inhalte aus der technischen Dokumentation übersetzt werden sollen, empfiehlt Globalese bei dieser Datenlage zusätzlich die Einbindung generischer Stockdaten, um Datenlücken zu schließen und die Fluency im Output zu verbessern. Die Mindestmenge an Mastertrainingsdaten für eine zuverlässige domänenspezifische Globalese-Engine, die zusätzlich generische Stockdaten verwendet, wird bei 100.000 Segmenten angesetzt. Erst ab 300.000 Segmenten an hochwertigen Masterdaten empfiehlt es sich, eine domänenspezifische Engine ohne die anreichernden Globalese-Stockdaten zu trainieren. Das erforderliche Volumen an Masterdaten wird dabei auch immer von der Datenqualität bestimmt. Auch bei Globalese gilt es also, die Datenlage abzuwägen und auszuprobieren, welche Konfiguration für die eigenen Projektanforderungen am besten funktioniert [vgl. 2].

Test-Engines

Das große Volumen an Trainingsdaten in dieser Untersuchung machte den Einsatz von Stockdaten nicht zwangsläufig erforderlich. Um jedoch die Parameter zur Zusammensetzung der Trainingsdaten des Globalese-Systems zu testen, wurde das Trainingsmaterial wie folgt konfiguriert:

Test-Engine 1 wurde nach dem Vorbild der bei beo eingesetzten Engine zusammengestellt und verwendete als Masterdaten das ungefilterte TM und die Terminologieliste, was einen Gesamtumfang von rund 726.000 Segmenten ergab (siehe Abb. 3). Zusätzlich wurden die Trainingsdaten mit Globalese-Stockdaten angereichert, die in einer von Globalese entwickelten Ähnlichkeitsanalyse auf Basis der domänenspezifischen Daten ausgewählt wurden. Das Training der Engine dauerte insgesamt knapp 28 Stunden.

Um einen Vergleichswert zu schaffen, wurde für die Test-Engine 2 eine Konfiguration gewählt, die andere Optionen des NMÜ-Systems ausschöpfte: Die Mastertrainingsdaten waren das terminologisch gefilterte TM-Korpus und die Terminologieliste. Zusätzlich wurden die in der terminologischen Filterung herausgefilterten Segmente, also die terminologisch ungültigen und veralteten Daten, hier anreichernd als unternehmensspezifische Hilfsdaten eingesetzt (siehe Abb. 3). Das Ziel hierbei war die Schließung möglicher Lücken in den Masterdaten, ohne dabei die korrekte Terminologie aus den Masterdaten zu unterdrücken. Auf diese Weise konnten auch flektierte Formen von Benennungen, die in der Filterung eventuell durch das Raster fielen, wieder mit ins Training gegeben werden. Das Globalese-System stellt im Trainingsprozess sicher, dass die korrekte Terminologie der Mastertrainingsdaten der konkurrierenden

| NAME | STATUS | HEALTH | SOURCE | TARGET | MASTER CORPORA | AUXILIARY CORPORA | STOCK DATA |
|---|---------|---------|--------|---------|----------------|-------------------|------------|
| Engine 1 (unbereinigtes TM + TB + Stock) | TRAINED | Healthy | German | English | 725,955 | 0 | CORPORA |
| Engine 2 (bereinigtes TM + TB + Hilfsdaten) | TRAINED | Healthy | German | English | 430,492 | 295,498 | |

Displaying 1-2 Filtered 2 Total 18 25 per page
 ●●● Healthy
 ●●● Retraining recommended
 ●●● Engine version deprecated

Abb. 3: Engine-Anzeige auf dem Globalese-Dashboard (eigener Screenshot)

veralteten oder falschen Terminologie in den Hilfsdaten gegenüber priorisiert wird. Stockdaten wurden für die Engine 2 nicht eingesetzt, um festzustellen, inwieweit ihr Einsatz ein Rauschen im Engine-1-Output erzeugen würde. Das Training der Test-Engine 2 dauerte insgesamt knapp 10 Stunden. Es war somit deutlich schneller abgeschlossen, weil keine Stockdaten analysiert werden mussten.

Testdaten und Engine-Evaluation

Zur Evaluierung einer NMÜ-Engine bedarf es eines festgelegten Testdatensatzes mit bilingualen Segmentpaaren, der im Abgleich mit dem von den Test-Engines generierten Output als Goldstandard dient [4]. Dieser darf nicht Teil des Trainingskorpus sein und muss neben dem zu übersetzenden Ausgangstext einen validierten Zieltext umfassen, der die vorgeschriebene Terminologie verwendet. Der in der vorliegenden Untersuchung eingesetzte Testdatensatz wurde aus zwei Übersetzungsprojekten aus dem Automobilbereich zusammengestellt, damit die Testdaten die zu prüfende unternehmensspezifische Terminologie enthalten. Er bestand neben dem deutschen Ausgangstext aus einer englischen post-editierten NMÜ-Referenzübersetzung. Die knapp 1.500 Testsegmente enthielten 13.891 Wörter. Als untersuchungsrelevante Terminologie wurden 3.438 Ein- und Mehrwortbenennungen sowie fachsprachliche Wendungen identifiziert. Eigennamen wurden dabei nicht berücksichtigt.

Auch beim Einsatz von Testdaten aus der Praxis kann eine mangelnde terminologische Qualität und Konsistenz ein Problem darstellen. Idealerweise wird hier zunächst eine Terminologieprüfung und ggf. -bereinigung der Testdaten durchgeführt. Vor allem automatische Bewertungsmethoden für maschinellen Output, wie z. B. der BLEU-Score, verzerren ansonsten die Evaluationsergebnisse. Mit dem BLEU-Score wird der Grad der Abweichung der maschinellen Übersetzung von der Referenzübersetzung gemessen [5]. Hier ein Beispiel: Wenn die maschinelle Übersetzung die korrekte Benennung, die Referenzübersetzung aber eine nicht zugelassene Benennung verwendet, würde die Einhaltung der vorgegebenen Terminologie bestraft. Um dieses Problem zu umgehen, wurde in dieser Untersuchung eine manuelle Fehler-Annotation durchgeführt, auch wenn dies die ressourcenintensivere Wahl war. Auf diese Weise wurden Terminologiefehler im Testdatensatz direkt erkannt und, statt eine vollumfängliche Evaluation durchführen zu müssen, konnte der Fokus auf die im Output aufgetretenen Terminologiefehler gelegt werden. Die Annotation bot zudem detailliert Aufschluss darüber, welche Arten von Terminologiefehlern auftraten. Die Fehler konnten

systembedingt oder datenbedingt sein. Zu systembedingten Fehlern zählen beispielsweise Auslassungen im Output oder auch falsche oder nicht übertragene Interpunktion. Diese Fehler erlernt eine Engine nicht durch das Auslesen der Trainingsdaten, sondern sie werden vom NMÜ-System verursacht. Datenbedingte Fehler können entstehen, indem eine Engine aus Fehlern oder Mängeln in den Trainingsdaten ungewollte Muster ableitet, die dann unter anderem zur Verwendung von nicht zugelassenen Synonymen oder zu fehlübersetzten Benennungen führen können.

Die untersuchten Terminologiefehler wurden elf Typen aus dem MQM-Framework zugeordnet und nach Kritikalität in drei Schweregrade unterteilt [4, 6]. Die Wahl fiel auf das MQM-Framework, weil es ein flexibles Evaluationsystem mit einer Fehlertypologie zur menschlichen Annotation und Bewertung von Übersetzungsqualität bietet. Die Schweregrade, denen die Terminologiefehler zugeordnet wurden, werden als geringfügig, grob und schwerwiegend bezeichnet. Geringfügige Fehler sind z. B. Rechtschreib- oder Flexionsfehler und schränken Verständnis sowie Lesefluss nicht weiter ein. Grobe Fehler entstehen durch Verstöße gegen Terminologievorgaben und schränken die Verständlichkeit bereits stärker ein. Ein Fehlertyp ist hier z. B. die fehlerhafte Ergänzung einer Benennung („Auftrag“ wird statt als „order“ als „repair order“ übersetzt.) oder eine gedoppelte Benennung. Ein weiterer grober Fehlertyp der Test-Engines ist das Generieren einer veralteten oder abzulehnenden Benennung, die auf einem falschen TM-Match basiert, das die Engine aus den Trainingsdaten vollständig oder in großen Teilen übernimmt (siehe Beispiel in Zeile 1 in nachfolgender Tab. 1). Schwerwiegende Fehler machen ein Output-Segment für den vorhergesehenen Zweck unbrauchbar, weil es Fehlinformationen enthält oder dadurch Schaden verursacht werden könnte. Schwerwiegende Terminologiefehler sind z. B. die Auslassung von Benennungen oder wenn eine homonyme Benennung im Ausgangstext im Sinne eines falschen Begriffs übersetzt wird. Auch Fehlübersetzungen können Terminologiefehler sein, z. B., wenn im Output eine falsche, teils willkürliche Benennung generiert wird, aufgrund einer der Engine unbekannteren Benennung, die somit eine terminologische Lücke im Trainingsmaterial ist (siehe Zeile 2 in Tab. 1).

Den drei Schweregraden wurden Multiplikatoren zugeordnet, um aus den Fehlerzahlen der beiden Engine-Outputs einen prozentualen Qualitätsscore pro Engine zu errechnen. Diese Scores können für zukünftige Projekte, in denen eine andere Engine-Konfiguration oder eine Engine, die mit einem anderen System trainiert wurde, als Vergleichswerte erneut herangezogen werden.

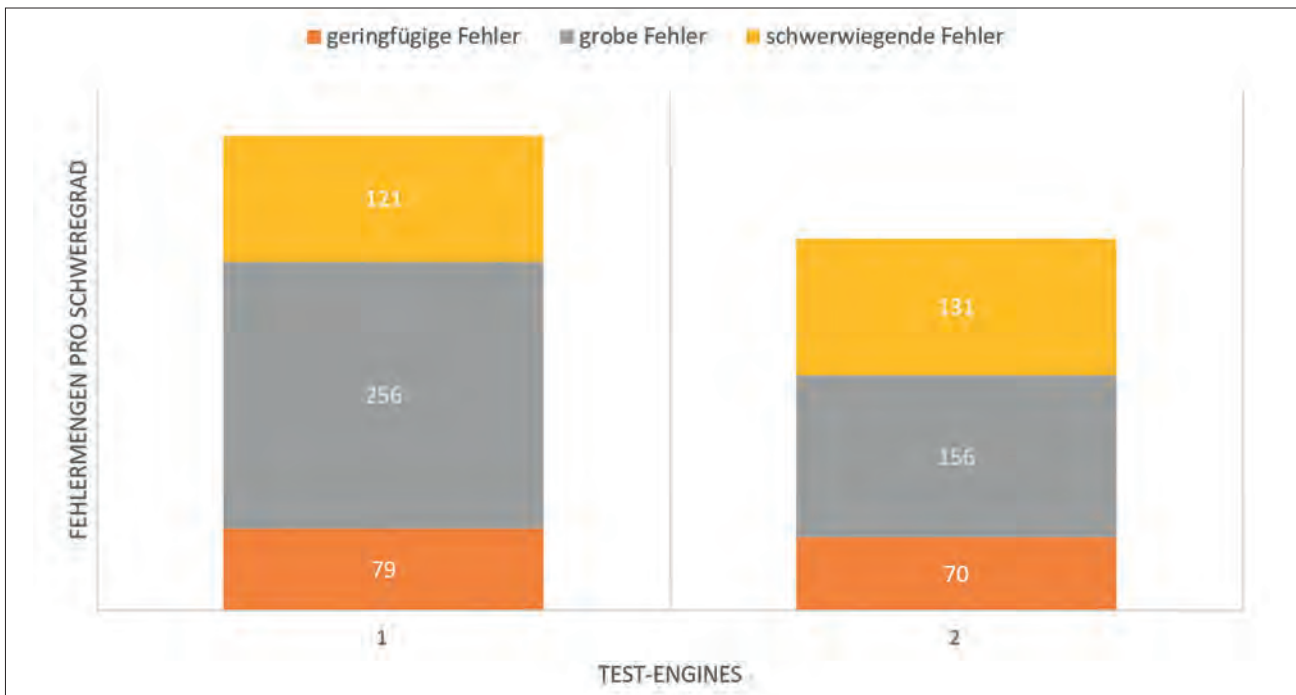


Abb. 4: Fehlermengen im Output der Test-Engines (eigene Darstellung)

Ergebnisse

Von Engine 1 wurden 87 % der untersuchten Terminologie korrekt generiert. 13 % hingegen waren nicht korrekt oder fehlten. Im Output von Engine 2 waren 90 % der Terminologie korrekt und 10 % fehlerhaft oder abwesend. In absoluten Zahlen ergaben sich folgende Werte: Von den 3.438 geprüften Benennungen und Wendungen wurden in 456 Fällen im Output von Engine 1 Fehler festgestellt; bei 357 Fällen im Engine-2-Output. Die errechneten Qualitätsscores lagen somit bei 92,6 % bzw. 94 %, womit eine verbesserte Terminologietreue in Engine 2 mit reduziertem, terminologischen Trainingsmaterial bestätigt werden kann.

Auffällig war die deutlich höhere Anzahl grober Fehler im Engine-1-Output. Diese ist auf den Fehlertyp „falsches TM-Match“ zurückzuführen (siehe Zeile 1 in Tab. 1), der 25 % der gesamten Fehlermenge der Engine ausmachte. Sein Auftreten konnte in Engine 2 durch die Filterung des TMs vermindert werden. Schwerwiegende Fehler wie z. B. Fehlübersetzungen (siehe Zeile 2 in Tab. 1) waren vor allem auf terminologische Lücken im Trainingsmaterial

zurückzuführen und sind deshalb im Engine-2-Output dominanter, da hier mit dem reduzierten Korpus und ohne Globalese-Stockdaten trainiert wurde. Die manuelle Evaluation konnte diese Lücken für viele Begriffe jedoch gezielt aufdecken und bietet einen Ansatz für Ergänzungen im Trainingsmaterial. Terminologiefehler aufgrund fachlich unpassender Terminologie (*Inconsistent with domain*) war in keinem der Outputs von Relevanz, was darauf schließen lässt, dass der Einsatz der Stockdaten für Engine 1 die Domänenadaptation nicht gestört hat.

Fazit

Die angebrachte Maxime „je mehr Daten desto besser“ kann zur Sicherung der Terminologietreue im Output einer domänenspezifischen NMÜ-Engine nicht immer standhalten [vgl. 3, 7]. Vor allem dann nicht, wenn terminologische Inkonsistenzen in einem großen Trainingskorpus ein Rauschen erzeugen. Ein zu kleines Korpus wiederum birgt das Risiko, dass die Engine zu viele terminologische Lücken hat und aufgrund dessen falsche oder keine Output-Benen-

| Ausgangstext | Referenzübersetzung | Engine-Output |
|--|--|---|
| Die elektrohydraulische Steuereinheit besteht im Wesentlichen aus folgenden Bauteilen: | The electrohydraulic controller unit basically consists of the following components: | The electrohydraulic controller unit basically consists of the following <u>component parts</u> : |
| Planetenaußenrad (15) zusammen mit Träger (14) abbauen. | Remove <u>planetary hub gear</u> (15) together with carrier (14). | Remove <u>hose assembly</u> (15) together with carrier (14) |

Tab. 1: Analysebeispiele für „falsches TM-Match“ und „Fehlübersetzung“

nungen generiert. Um es mit Riley Newmans Worten auszudrücken: „More data beats better models. Better data beats more data“ [8].

Unter der Voraussetzung, dass ein ausreichend großes Volumen an Trainingsdaten vorhanden ist, ist eine terminologische Filterung ein guter Ansatz, um Trainingsdaten für eine domänenspezifische NMÜ-Engine ohne größeren Bereinigungsaufwand vorzubereiten. Zudem finden die herausgefilterten Daten im Globalese-System weiterhin einen Nutzen als Hilfsdaten. In einem nächsten Schritt sollte natürlich eine terminologische Bereinigung dieser Daten durchgeführt werden, um das Mastertrainingskorpus entsprechend vergrößern zu können. Wichtig für den hier vorgestellten Ansatz ist außerdem, dass eine noch differenziertere Terminologieliste zur Filterung und auch zum Training erarbeitet wird. In der Filterung können die Daten auf diese Weise gründlicher geprüft werden und im Training können damit mehr grobe Datenlücken geschlossen werden. Grundvoraussetzung für eine gesicherte Terminologietreue in einer domänenspezifischen NMÜ-Engine ist also zunächst eine sauber gepflegte Terminologie.

Literatur

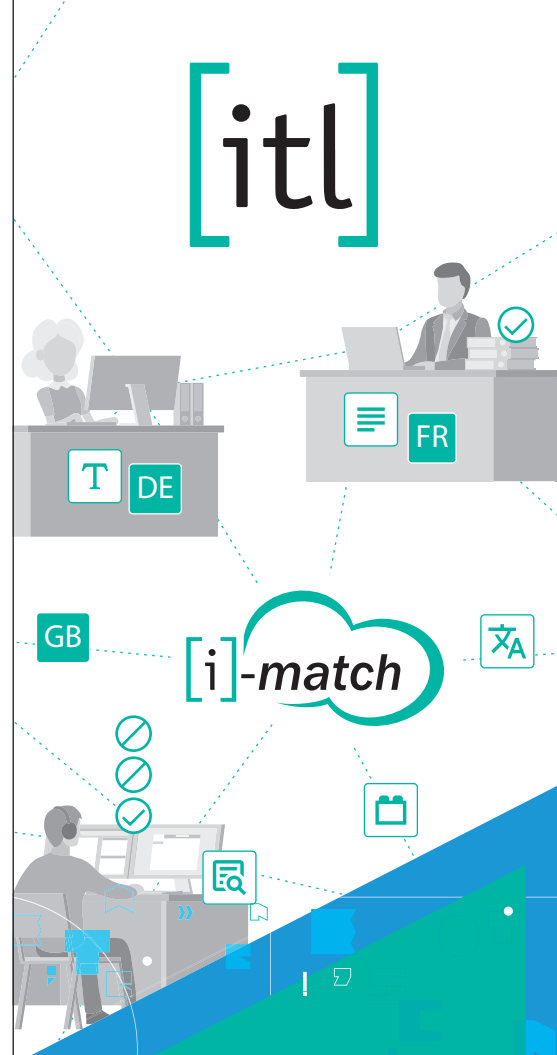
- [1] Forcada, Mikel L. (2017): Making sense of neural machine translation. In: Kenny, Dorothy (Hrsg.): Translation Spaces, Vol. 6-2. Amsterdam: John Benjamins Publishing Company, S. 291-309.
- [2] Bessenyei, Gabor (2019): <https://www.globalese-mt.com/2019/02/19/breaking-through-the-terminology-barrier-in-neural-machine-translation/> [Zugriff am 20.03.2022]
- [3] Horvath, Greg (2018): <https://www.globalese-mt.com/2018/12/06/using-stock-corpora-for-training-engines/> [Zugriff am 23.03.2022]
- [4] Winter, Tom; Zielinski, Daniel (2020): Terminologie in der neuronalen maschinellen Übersetzung. In: Porsiel, Jörg (Hrsg.): Maschinelle Übersetzung für Übersetzungsprofis. Berlin: BDÜ Weiterbildungs- und Fachverlagsgesellschaft mbH, S. 210-233.
- [5] Brownlee, Jason (2019): <https://machinelearningmastery.com/calculate-bleu-score-for-text-python/#:~:text=The%20Bilingual%20Evaluation%20Understudy%20Score,in%20a%20score%20of%200.0> [Zugriff am 19.03.2022]
- [6] Soloviev, Kirill (2017): <https://blog.taus.net/measuring-content-quality-with-error-typology-step-by-step-guide> [Zugriff am 14.03.2022]
- [7] Koehn, Phillip (2020): Neural Machine Translation. Cambridge, New York, Port Melbourne, New Delhi, Singapore: Cambridge University Press.
- [8] Sharma, Shikha (2020): <https://blog.taus.net/what-is-data-cleaning> [Zugriff am 19.03.2022]



Frauke Hellerich ist freiberufliche technische Übersetzerin. Sie hat Internationale Fachkommunikation (B. A.) und Fach- und Medienübersetzen (M. A.) studiert. Als Lehrbeauftragte an der Hochschule für angewandte Wissenschaften Würzburg-Schweinfurt unterrichtet sie Technisches Englisch, Business-Englisch und Niederländisch. Außerdem ist sie als Terminologin und Projektleiterin bei Büro b3 Terminologiemanagement tätig.

Kontaktadresse

frauke.hellerich@lehrbeauftragte.fhws.de



[i]-match Language Management in Ihrer Cloud

- Unternehmensübergreifende Nutzung von Terminologie
- Terminologie- und Webportal
- Autorenunterstützung
- Integration per REST-API

Terminologiemanagement testen in der [i]-match Test-Community



David Bodensohn
info.i-match.cloud / 089 892623-610

[itl] boost your content to success
www.itl.eu / itl.ch / itl.at

Trufflepiggy: Terminologie komfortabel erschnüffeln

Armin Mutscheller

Systematische Terminologierecherche gilt vielen als beschwerlich und zeitraubend. Als Übersetzer kann sich ein Lied davon singen: Das Internet ständig nach ein- und mehrsprachiger Terminologie oder nach Fachwissen durchsuchen zu müssen, kann frustrierend sein, ist jedoch für ordentliche Übersetzungsergebnisse unerlässlich.

Das von dem Programmierer Clemens Mautner Markhof in Wien entwickelte Browser-Add-on Trufflepiggy erweist sich dabei als sehr hilfreich. Der gelernte Spiele-Entwickler hat das pfiffige Tool in seinen Grundzügen bereits im Jahr 2010 entworfen, um sich die Recherche für seine Master-Arbeit zu erleichtern. Nach und nach spendierte er seinem Trüffelschwein immer mehr Funktionen und stellte es schließlich im Jahr 2015 der Öffentlichkeit zur Verfügung. Trufflepiggy gibt es für alle gängigen Webbrowser, und das sogar in zwei Versionen: Das mit allen Funktionen ausgestattete Quick Search und das im Funktionsumfang etwas reduzierte, dafür jedoch in puncto Browser-Berechtigungen besonders genügsame Context Search. Beide Versionen sind kostenlos. Wer sie benutzen möchte, muss lediglich ein Nutzerprofil anlegen, um Einstellungen browserübergreifend speichern zu können. Dazu genügt die Angabe eines Benutzernamens, einer E-Mail-Adresse und eines Kennworts. Weitere personenbezogene Daten werden nicht erhoben. Auch beim Setzen von Cookies ist das in Deutschland gehostete Tool sehr zurückhaltend. Die Datenschutzerklärung ist ebenso wie die Benutzeroberfläche und Dokumentation in englischer Sprache gehalten.

Flexibel und ergonomisch

Die Verwendung von Trufflepiggy ist schnell erklärt: Man markiert auf einer Webseite das Wort, nach dem man suchen möchte, und öffnet über eine Tastenkombination oder durch Klicken auf eine neben dem Text eingeblendete (und auf Wunsch deaktivierbare) rosa Schweinsnase den Suchdialog (siehe Abb. 1). Dieser kann entweder als eigenständige Seite (also in einem eigenen Tab) oder als sogenanntes Overlay angezeigt werden. Letzteres legt sich halbtransparent über die aktuelle Webseite, was aus meiner Erfahrung die bessere Option ist. Anschließend klickt man in einer Liste auf eine der mitgelieferten oder selbst definierten Suchquellen, und schon erschnüffelt Trufflepiggy die gesuchte Terminologie und zeigt eine Trefferliste an. Wer möchte, kann die Suche auch über das Kontextmenü des Browsers anstoßen.

Besonders praktisch ist die Möglichkeit, aus bereits angelegten (privaten oder öffentlichen) Suchquellen beliebig viele Suchgruppen zu erstellen. Möchte man beispielsweise auf duden.de gleichzeitig das Wörterbuch, die Synonym-sammlung und das Sprachwissen durchsuchen, fasst man die entsprechenden Suchseiten einfach zu einer Gruppe zusammen, um diese in einem Aufwasch zu durchforsten (siehe Abb. 2). Terminologen können also für ihre Arbeit spezielle Gruppen mit enzyklopädischen, grafischen und korpusorientierten Quellen anlegen. Für Übersetzungsprojekte lassen sich auf die gleiche Weise fach-, sprachen- oder geospezifische Gruppen erstellen.



Abb. 1: Wer bevorzugt mit der Maus hantiert, klickt auf den rosa Rüssel, um das Trüffelschwein auf die gesuchte Terminologie anzusetzen. Über die Tastatur bringt man es mit STRG+SHIFT+F dazu, die Fährte aufzunehmen.

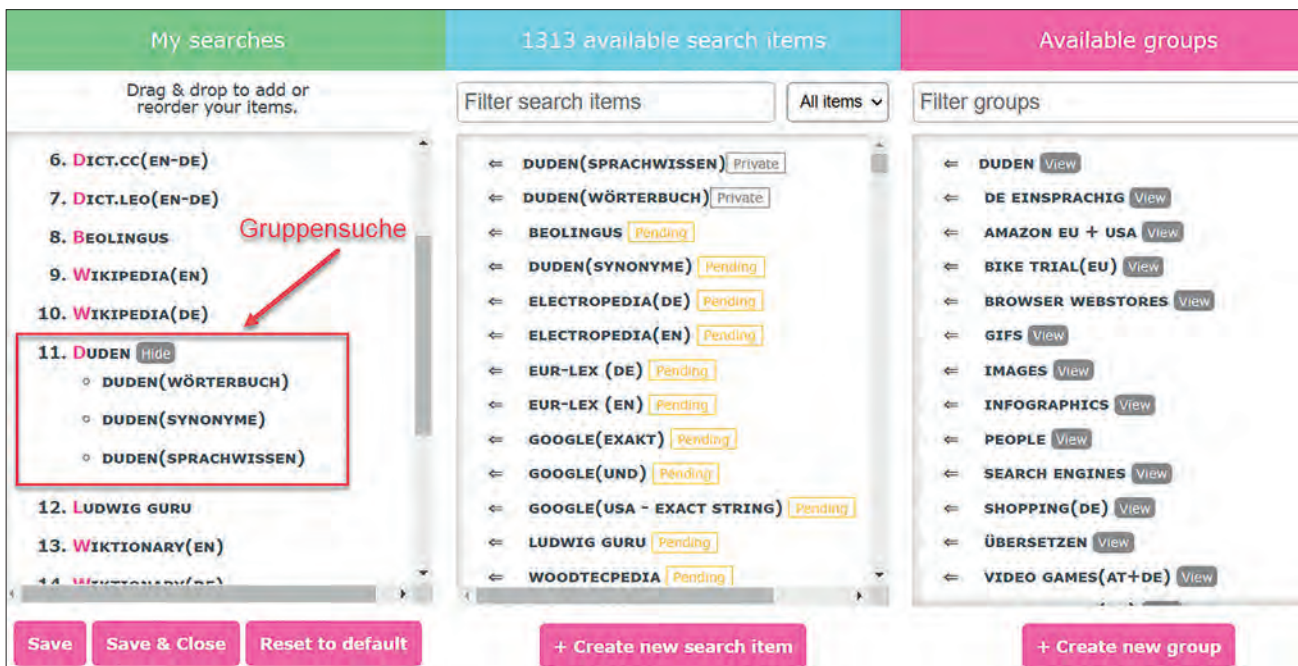


Abb. 2: Mit einer Gruppensuche lassen sich mehrere Suchquellen gleichzeitig durchsuchen. Neu angelegte Suchquellen sind in der Voreinstellung öffentlich verfügbar, können aber auch als „privat“ markiert werden.

Das Add-on kennt viele nützliche Optionen und Einstellmöglichkeiten, die auf einem Server zentral gespeichert werden. Zu den Möglichkeiten, die sich nicht auf den ersten Blick erschließen, gehört die gezielte Nutzung der Omnibox, also die Adressleiste des Browsers. Hat man den Cursor dort platziert (in Google Chrome zum Beispiel mit dem Hotkey STRG+L), genügt es, den Buchstaben „t“, ein Leerzeichen und eine Buchstabenfolge einzugeben, um die gewünschte Suchquelle zunächst einzukreisen. Hat man beispielsweise Wikipedia, Wiktionary und HandWiki als

Suchquellen angelegt, zaubert Trufflepiggy bei der Eingabe „wi“ alle drei Wikis in eine unterhalb der Omnibox eingeblendete Auswahlliste. Dann markiert man über die Pfeiltasten die gewünschte Suchquelle gezielt, gibt noch das gesuchte Wort ein und stößt über die Eingabetaste die Suche an. Mit dieser sehr praktischen Funktion gelangt man schnell zur benötigten Suchquelle. Dies macht das Add-on zu einem sehr ergonomischen Hilfsmittel, da man auf den ständigen Griff zur Maus verzichten und die Suchvorgänge weitestgehend per Tastatur steuern kann.

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

16./17.09.2022

Berlin (Kosmos)

Bundesverband der
Dolmetscher und Übersetzer



7. FACHKONFERENZ SPRACHE UND RECHT

Rechtsvergleichung in der Praxis

Am 16./17. September 2022 veranstaltet der Bundesverband der Dolmetscher und Übersetzer e. V. (BDÜ) die siebte Auflage seiner Fachkonferenz zum Thema „Sprache und Recht“, diesmal unter dem Motto: „Rechtsvergleichung in der Praxis“. Die Konferenz richtet sich an Dolmetscher und Übersetzer, die mit Recht und Rechtssprache arbeiten, sowie an interessierte angrenzende Berufsgruppen wie Rechtsanwälte, Richter, Notare, Vertreter der Polizei und Justiz und Vertreter von Behörden. Ihnen allen bietet die Konferenz eine Diskussions-, Weiterbildungs- und Netzwerkveranstaltung rund um die Themengebiete Recht in verschiedenen Sprachen und Systemen, Rechtssprache, Rechtsvergleichung und praktische Arbeit der Rechtsdolmetscher/-übersetzer. Erwartet werden rund 500 Teilnehmer aus Deutschland und dem Rest der Welt sowie Delegationen von Berufsverbänden der Dolmetscher und Übersetzer aus zahlreichen Mitgliedstaaten der Europäischen Union.

- **Anmeldeschluss (Teilnehmer):** 31. Juli 2022
- **Anmeldeschluss (Aussteller):** 15. August 2022

Weitere Informationen und Anmeldung unter:

→ www.fsr2022.de

Hint Type ! + your search-item to quickly load any search, e.g. *!amz* loads all Amazon searches. [Watch a demo.](#)

1 G GOOGLE(USA - EXACT STRING) 2 G GOOGLE(UND) 3 G GOOGLE MAPS

4 Sp STARTPAGE(COM) 5 DICT.LINGUEE(EN-DE) 6 DICT.CC(EN-DE)

7 DICT.LEO(EN-DE) 8 BEOLINGUS 9 W WIKIPEDIA(EN)

10 W WIKIPEDIA(DE) d DUDEN l LUDWIG GURU

W [W] WIKTIONARY(EN) W WIKTIONARY(DE) a AF ACRONYMFINDER(COM)

e IEC ELECTROPEDIA(EN) IEC ELECTROPEDIA(DE) EUR-LEX (DE)

EUR-LEX (EN) WOODTECPEDIA y YOUTUBE

Extras

TRY TO OPEN AS LINK

SEARCH ON DTTEV.ORG

Suche in der Domain der aktuell geöffneten Website

Trufflepiggy Quick Search

edit searches | options | profile | help

Abb. 3: Auch Websites ohne eigene Suchfunktion lassen sich mit Trufflepiggy problemlos durchsuchen.

Noch tiefer schnüffeln

Ein besonderer Leckerbissen, der die tägliche Übersetzungs- und Textarbeit enorm erleichtert, ist die Möglichkeit, die gerade im Browser geöffnete Domain zu durchsuchen (siehe Abb. 3). Trufflepiggy führt dazu eine Google-Suche unter Verwendung des Google-Suchparameters `site:` aus. Hilfreich ist dies vor allem dann, wenn die betreffende Website keine eigene Suchfunktion hat. Und erstaunlicherweise liefert Trufflepiggy selbst dort, wo die Website bereits ein Suchfeld besitzt, nicht selten mehr und bessere Ergebnisse als die dort implementierte Suchfunktion.

Kleine Wunschliste

Das Add-on lässt kaum Wünsche offen. Verbessern ließe sich jedoch meiner Ansicht nach das Layout des Suchdialogs. Dort ist man auf eine starre Drei-Spalten-Anordnung angewiesen, die sich nicht ändern lässt. Die einzelnen Einträge lassen sich zwar nach oben oder unten verschieben, doch könnten eine benutzerdefinierte Spaltenanzahl, eine farbliche Kennzeichnung thematisch verwandter Suchquellen oder frei platzierbare Trennlinien für mehr Übersicht sorgen. Auch eine flexiblere Nummerierung der Einträge bzw. das freie Zuweisen von Buchstaben für eine Schnellsuche per Tastendruck könnten dem Tool eine noch bessere Usability verleihen.

Als Sahnehäubchen auf meiner Wunschliste stelle ich mir eine Chronik der bisher durchgeführten Recherchen vor, um Suchvorgänge und Fundstellen besser dokumentieren zu können, als es der Browserverlauf ermöglicht. Doch auch dafür gibt es recht brauchbare Add-ons, wie zum Beispiel Better History für Google Chrome. Und wer keine Berührungängste mit Tools wie AutoHotKey hat, kann das Trüffelschwein mit wenigen Skriptzeilen sogar aus Textverarbeitungsprogrammen oder CAT-Tools heraus fröhlich schnüffeln lassen, zumindest unter Windows. Oink!



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 Heidelberger Institut für Übersetzen und Dolmetschen Lehrveranstaltungen ab.

Kontaktadresse
info@mutscheller.de
www.mutscheller.de

Am Anfang war der Terminologie-Tag...

Ein Rückblick auf 35 Jahre Vereinsgeschichte des DTT e.V.

Klaus-Dirk Schmitz



Abb. 1: 1987 – 2022 Deutscher Terminologie-Tag e.V. (Illustration: Tamara Arndt)

Erste terminologische Aktivitäten beim BDÜ

Im Bundesverband der Dolmetscher und Übersetzer e.V. (BDÜ) fanden sich bereits Ende der 1970er bzw. Anfang der 1980er Jahre Fachleute zusammen, die die Bedeutung von Terminologiearbeit im Bereich des Übersetzens erkannt hatten. Neben einzelnen Fachübersetzern gesellten sich auch Leiter von Sprachendiensten in Behörden und Unternehmen dazu. Auf Initiative dieser Expertengruppe veranstaltete der BDÜ am 11. Mai 1985 einen Tag zum Thema Terminologie in Köln, auf dem u. a. Ingo Hohnhold (Siemens AG München) ein Grundsatzreferat mit dem Titel „Was ist, was soll, was kann Terminologiearbeit – nicht nur für Übersetzer?“ hielt. Außerdem wurden in diesem Rahmen sowohl eine ständige Arbeitsgruppe „Terminologiearbeit“ im BDÜ gegründet als auch eine Rahmenresolution mit Zielsetzungen für die übersetzungsorientierte Terminologiearbeit formuliert, die den Fokus auf die folgenden drei Themenbereiche legte:

- Terminologiearbeit und -sicherung
- Rechnergestützte Verarbeitung von Terminologiebeständen
- Darstellung von terminologischen Einträgen und terminologische Zusammenarbeit

Die neu gegründete Arbeitsgruppe bereitete bereits für das Folgejahr eine größere Veranstaltung zum Thema Terminologiearbeit vor, die am 27. und 28. November 1986 ebenfalls in Köln als „1. Deutscher Terminologie-Tag“ stattfand. Es gab Kurzreferate aus dem universitären Bereich (Germersheim, Heidelberg, Hildesheim, Saarbrücken, Genf und Süd-Dänemark) und Vorträge von DIN und Infoterm. Anschließend teilten sich die etwa 200 Teilnehmer unter Leitung von Ingo Hohnhold, Yvan Mossmann, Michael Schaar und Helmut Hartmann in drei Gruppen auf, um zu den oben genannten Bereichen gemeinsam fachliche Inhalte und Empfehlungen zu erarbeiten, die daraufhin im Mitteilungsblatt des BDÜ publiziert wurden.

Gründung des Deutschen Terminologie-Tags

Um den Fokus der Aktivitäten nicht nur auf die übersetzungsbezogene Terminologiearbeit zu beschränken und unabhängiger vom BDÜ zu werden, versammelten sich die vier Organisatoren des 1. Deutschen Terminologie-Tags und weitere interessierte Fachleute aus der Industrie am 3. Mai 1987 in Köln zur Gründungsversammlung des „Deutschen Terminologie-Tag e.V.“, der dann offiziell am 15. Dezember 1987 in Köln ins Leben gerufen wurde.

Vorstand und Mitglieder

Nach der Gründung musste der neue Verein zunächst seine Position definieren und seine Ziele sowie Aktivitäten festlegen. Die folgenden Vorsitzenden prägten im Laufe der Jahre zusammen mit verschiedenen Vorstands- und Fachbeiratsmitgliedern maßgeblich die Verbandsarbeit:

- Ursula Reisen: 1989 – 1998
- John D. Graham: 1998 – 2001
- Jutta Zeumer: 2001 – 2011
- Mark D. Childress: 2011 – 2017
- Tom Winter: 2017 – heute

Die Zahl der aktiven und passiven Mitglieder war am Anfang noch recht überschaubar, da das Thema Terminologie damals noch als ausgesprochenes Nischenthema galt. Sie wuchs aber mit den Jahren kontinuierlich und knackte schließlich im Jahr 1994 die 100er-Grenze. Inzwischen

taz vom 21.2.1990, S. 11

Turmbau zu Brüssel

Grenzen mögen fallen, Sprachbarrieren bleiben / Englisch als EG-Sprache hat kaum größere Chancen ■ Mit dem SPRACHWIRRWARR auf du und du

Köln/Duisburg (dpa) – Als sich die hungrigen Nachkriegsdeutschen hauptsächlich von pappigem Maismehlbrot ernähren mußten, hatte dies eine kleine, aber folgenschwere Ursache: Ein Dolmetscher hatte im Hilfesuch an die Amerikaner das deutsche Wort „Korn“ allzu unbedacht mit dem ähnlich klingenden „corn“ übersetzt, was schlicht „Mais“ bedeutet. Und wenn, wie auf einer Karikatur im Blatt des Vereins Deutscher Ingenieure unlängst zu sehen, der „Deutsche Michel“ den Begriff „Schraube“ benutzt, versteht der Brite schnell „Mutter“, der Franzose „Schraubhaken“, der Italiener „Karabinerhaken“ und der schwarzgelockte Karikatur-Spanier gar „Angelhaken“.

Auch wenn 1993 die Grenzschranken Europas fallen, wird ein Hindernis bestehen bleiben: Sprachbarrieren, so Wirtschaftsfachleute, können gerade in der exportorientierten Bundesrepublik wie eine Bremse wirken und der Wirtschaft Milliardenverluste beschern. Betroffen sind besonders kleine und mittlere Unternehmen, die andererseits große Hoffnungen auf den Binnenmarkt setzen.

Findige Zeitgenossen haben dies längst bemerkt und bieten ihre Dienste als „Terminologie-Broker“ an. Sie weisen den Weg zu der richtigen Terminologie, zu den Fachbegriffen, wenn etwa ein deutscher Produzent Küchengeräte auf den griechischen Markt bringen will, dänischer Teppichboden in Portugal verlegt werden soll.

Systematisch geht der vor zwei Jahren gegründete „Deutsche Terminologie-Tag e.V.“ (DTT), der aus dem Bundesverband der Dolmetscher und Übersetzer (BDÜ) hervorgegangen ist, das Problem an: Es müssen nicht immer fremde Sprachen sein, wenn es zu einem regelrechten Wirrwarr im Gespräch der Fachleute kommt. Selbst innerhalb der Bundesrepublik gibt es etwa bei Technikern regionale Fachsprachen-Unterschiede, die auch auf die verschiedenen Ausbildungsorte zurückgehen. Auch von Firma zu Firma differieren die Fachvokabeln. Also, so unterstreicht Ursula Reisen (Köln), Vorsitzende des DTT und Leiterin des Sprachendienstes bei Rheinbraun, kommt es zunächst darauf an, die Terminologie innerhalb der eigenen Sprache festzuschreiben, diese Definition dann mit vergleichbaren Ausdrücken in der fremden Sprache zu vergleichen. Immer wichtiger wird natürlich auch die deutsch-deutsche Fachsprachen-Verständigung, die trotz politischer Entspannung noch die „Erblast“ jahrzehntelanger Isolation und ungleicher Entwicklung beispielsweise im

Bereich von Technik oder Wirtschaft trägt.

Englisch als allgemeine EG-Umgangssprache hat nach Einschätzung von Ursula Reisen kaum eine größere Chance, da einerseits jedes Mitgliedsland eifersüchtig über die Verwendung der eigenen Sprache wacht, andererseits im Zusammenhang mit dem neuen Produkthaftungsrecht der EG jedes Produkt in der jeweiligen Sprache des Abnehmerlandes dokumentiert werden muß: „Eine ungeheure terminologische Aufgabe!“ Überhaupt bestehe jedes Produkt „zu einem wesentlichen Teil aus Papier“, wie Betriebsanleitung, Montageanweisung, Wartungshandbuch – und Textqualität entspricht hier leicht Produktqualität. Ursula Reisen: „Eine Tatsache, der bisher viel zu wenig Bedeutung beigemessen wird.“ Sachgerechte Terminologiearbeit sei durchaus kostenaufwendig, „aber noch teurer ist es, keine Terminologie zu erarbeiten“.

Eindämmen kann natürlich auch der Terminologie-Tag die Flut der tagtäglich in vielen Bereichen entstehenden Fachvokabeln nicht, aber zumindest das Problem aufzeigen, etwa die EG-Datenbank der Fachbegriffe „Eurodicautom“ in Luxemburg und Fachpublikationen als „verborgene Terminologiequellen“ bestmöglich im Auge behalten.

■

Abb. 2: Der DTT und die damalige Vorstandsvorsitzende Ursula Reisen werden in der taz vom 21.2.1990 auf S. 11 erwähnt. (Quelle: <https://taz.de/Turmbau-zu-Bruessel/11779390/> Nachdruck mit freundlicher Genehmigung)



Abb. 3: Jutta Zeumer eröffnete das DTT-Symposium 2006 in Weimar, das in Zusammenarbeit mit der tekom stattfand (Foto: Gregor Schäfer)

gibt es neben Einzelmitgliedern auch zahlreiche Firmenmitgliedschaften und die Gesamtzahl beläuft sich auf etwas mehr als 300 Mitglieder. Gleichzeitig wuchs auch der Vorstand und zählt heute neun ehrenamtliche Mitglieder.

DTT-Publikationen

Unter der Federführung von Reiner Arntz wurde 1991 die bekannte gelbe Broschüre „Terminologie – Wozu?“ veröffentlicht, um die Terminologearbeit und den Verein zu bewerben. Heute ist sie – überarbeitet und „modernisiert“ – kostenfrei in drei Sprachen erhältlich (siehe Anzeige auf S. 33).

Seit 1992 wurde regelmäßig ein Rundbrief mit Pressepiegel an die Mitglieder des DTT verschickt, um sie über aktuelle Informationen aus dem Bereich der Terminologie, aber auch über vereinsrelevante Angelegenheiten zu informieren. 2005 wurde dieser Rundbrief von der Terminologiefachzeitschrift *edition* abgelöst, die terminologisch relevante Themen aus Praxis, Forschung sowie Lehre aufgreift und in zwei Ausgaben im Jahr auf professionellem Niveau Mitglieder und andere Interessierte über Terminologiethemata informiert. 2015 wurde die *edition* im Zuge der weiteren Professionalisierung und Internationalisierung in den European Reference Index for the Humanities and the Social Sciences aufgenommen. Sie entspricht damit den internationalen Anforderungen an eine wissenschaftliche Fachzeitschrift mit Redaktionsbeirat und doppeltem Blind-Review. Einen wesentlichen Anteil am Erfolg der *edition* haben Jutta Zeumer und Detlef Reineke. Alle DTT-Mitglieder erhalten ein Printexemplar der *edition*, die ansonsten auf der DTT-Website im elektronischen Format für alle Interessierten frei verfügbar ist (siehe Anzeige auf S. 20).

Neben den Tagungsbänden, die begleitend zu jedem Symposium erscheinen und Veröffentlichungen der jeweiligen Vorträge enthalten, ist der Best-Practices-Leitfaden sicherlich die bekannteste und am meisten verkaufte Publikation des DTT. Die praxisnahen Inhalte wurden von ehrenamtlichen Gruppen in sechs thematischen Modulen erarbeitet und „der Ordner“ erschien 2010 in seiner ersten Version. Die überarbeitete Version 2.0 wurde um zwei weitere Themen ergänzt und 2014 in aktualisierter Form auf den Markt gebracht. Die englischsprachige Fassung der Version 2.0 wurde schließlich 2020 veröffentlicht (siehe Anzeige auf S. 41).



Abb. 4: DTT-Stand auf der tekom Jahrestagung 2018. Im Bild v. l.: Geschäftsführerin Dr. Annette Weilandt, 1. Vorsitzender Tom Winter und Olga Buchstaller-Vodopyanova, Ansprechpartnerin in der DTT-Geschäftsstelle. (Foto: Petra Drewer)

Veranstaltungen des DTT

Seit 1989 veranstaltet der DTT im Abstand von zwei Jahren Symposien zu relevanten und aktuellen Themenbereichen aus dem Gebiet der theoretischen und praktischen Terminologiearbeit, die mit den Jahren immer größeres Interesse erfuhren (siehe Tabelle 1).

Im Bereich Fortbildung bietet der DTT ein Grundlagenseminar (Terminologiearbeit – Grundlagen, Werkzeuge, Prozesse) und ein Vertiefungsseminar (Terminologiearbeit – Projekte, Prozesse, Datenaustausch) an. Ergänzt werden diese Veranstaltungen durch mehrmals im Jahr stattfindende Webinare, in denen aktuelle Themen und Entwicklungen aufgegriffen oder Werkzeuge zur Terminologiearbeit praktisch vorgestellt werden.

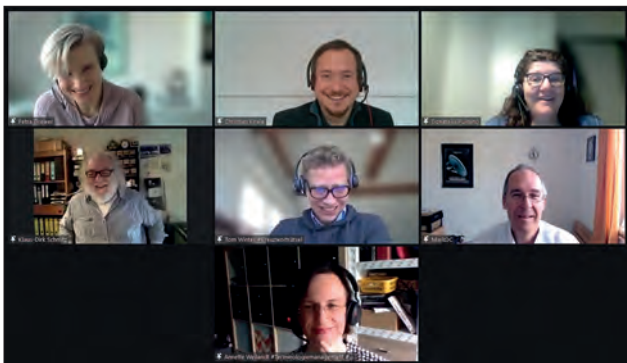


Abb. 5: 2021 fand das DTT-Symposion aufgrund der Corona-Krise zum ersten Mal virtuell statt. Im Bild v. l.: Prof. Dr. Petra Drewer, Christian Kriele, Donatella Pulitano, Prof. Dr. Klaus-Dirk Schmitz, Tom Winter, Mark Childress und Dr. Annette Weilandt (Screenshot: Tamara Arndt)

Weitere Aktivitäten des DTT

Bereits 2010 hatte der DTT-Vorstand die Idee, herausragende anwendungsorientierte oder wissenschaftliche Arbeiten aus dem Bereich der Terminologiewissenschaft und Terminologiearbeit durch die Vergabe eines Förderpreises auszuzeichnen. Nachdem Ausschreibungs- und Bewertungsmodalitäten geregelt und die erste Ausschreibung erfolgt waren, wurde 2014 auf dem DTT-Symposion der erste DTT-Förderpreis verliehen. Die bisherigen Preisträger sowie die Bewerbungsmodalitäten sind auf der DTT-Website zu finden.

Nach wie vor sind die Ausbildungsmöglichkeiten für Terminologen noch sehr begrenzt. Aus diesem Grund hat der DTT 2015 das DTT-Terminologiezertifikat ins Leben gerufen, um Interessierten die Möglichkeit zu bieten, zusätzliche terminologische Kenntnisse zu erwerben und diese auch nachweisen zu können. Die erforderlichen Veranstaltungen können problemlos berufsbegleitend und in (fast) beliebiger Reihenfolge absolviert werden. Außerdem stellen vor allem die Präsenzveranstaltungen eine gute Möglichkeit dar, sich mit anderen Fachleuten im Bereich Terminologie auszutauschen.

Mitfinanziert durch das NRW-Förderprogramm TRAFO und in Abstimmung mit dem DTT wurde an der TH-Köln das Deutsche Terminologie-Portal realisiert. Im Rahmen des Folgeprojekts „TIPPS“ (Terminology Information Policy, Portal and Service) wurden die Inhalte neu aufbereitet und bis 2017 ständig aktualisiert (www.termportal.de).

| Jahr | Ort | Thema des Symposions |
|---------|------------|--|
| 1989 | Duisburg | Terminologie für 1992 |
| 1991 | Köln | Terminologie als Qualitätsfaktor |
| 1992 | Köln | Geistiges Eigentum an Terminologien |
| 1994 | Köln | Terminologie als Produktivitätsfaktor |
| 1996 | Köln | Terminologie für ein vielsprachiges Europa |
| 1998 | Köln | Deutsche Terminologie im internationalen Wettbewerb |
| 2000 | Köln | Terminologie und Recht |
| 2002 | Köln | eTerminology: Professionelle Terminologiearbeit im Zeitalter des Internet |
| 2004 | Köln | Terminologie und Wissensmanagement |
| 2006 | Weimar | Terminologie von Anfang an (zusammen mit tekomp) |
| 2008 | Mannheim | Terminologie und Fachkommunikation |
| 2010 | Heidelberg | Best Practices in der Terminologiearbeit |
| 2012 | Heidelberg | Terminologieprozesse und Terminologiewerkzeuge |
| 2014 | Mannheim | Rechte, Rendite, Ressourcen: Wirtschaftliche Aspekte des Terminologiemanagements |
| 2016 | Mannheim | Terminologie und Kultur |
| 2018 | Mannheim | Terminologie und Text(e) |
| 2020/21 | virtuell | Terminologie: Industrie, Information, Intelligenz |
| 2023 | Mannheim | [Symposion in Vorbereitung] |

Tabelle 1: Übersicht über die Themen der DTT-Symposien seit Gründung des Vereins



Abb. 6: Treffen des DIT mit Gästen aus dem DTT-Vorstand 2013. Im Bild v. l.: Prof. Dr. Petra Drewer, Dr. Detlef Reineke, Dr. Nicole Keller, Mark Childress, Jutta Zeumer, Prof. Dr. Felix Mayer, Donatella Pulitano, Prof. Dr. Klaus-Dirk Schmitz, Ayten Turhan und Dr. François Massion (Foto: anonym).

Neben einführenden Informationen zum Thema Terminologie gibt es hier u. a. Hinweise zu Software, Fortbildungsangeboten und weiterführender Literatur.

Deutsches Institut für Terminologie e.V. (DIT)

Bereits in der Anfangsphase des DTT wurde erkannt, dass ein Mitgliederverband mit dem Status der Gemeinnützigkeit nicht Beratungs- und Dienstleistungsaufgaben für Unternehmen und Behörden anbieten und durchführen kann. Deshalb entstand die Idee, dafür einen eigenen, eng mit dem DTT verbundenen Verein zu gründen. Aus einer Expertengruppe des DTT entstand 1993 die Idee des Deutschen Instituts für Terminologie e.V. (DIT). Die Ausschreibung der Europäischen Kommission für das POINTER-Projekt (Proposals for an Operational Infrastructure for Terminology in Europe), bei der ein deutscher Projektpartner gesucht wurde, beschleunigte die Gründung des DIT, die am 30. September 1994 bestätigt wurde.

Das DIT war in den Jahren 1995/1996 als deutscher Partner maßgeblich an dem durch die Kommission kofinanzierten Projekt POINTER beteiligt. Die Ergebnisse dieses Projekts lieferten einen guten Überblick über terminologie-relevante Aktivitäten, terminologische Ressourcen sowie

vorhandene Strukturen im Europa der neunziger Jahre und boten konkrete Empfehlungen für die Umgestaltung und Verbesserung der Terminologielandschaft in Europa.

Zu Beginn der 2000er gab es eine Neuorientierung des DIT, das sich seitdem als Expertengruppe und „Think Tank“ des DTT präsentiert. Hierzu wurde das DIT als formaler Rahmen für die vom DTT benannten Fachbeiräte definiert. Die Fachbeiräte des DIT engagieren sich maßgeblich bei der Vergabe des Förderpreises, der inhaltlichen Gestaltung der Symposien sowie bei der Diskussion neuerer Themen wie etwa „KI und Terminologie“. Der Vorstand des DIT (aktuell: Petra Drewer, Donatella Pulitano, Felix Mayer) ist quasi der Sprecher der DTT-Fachbeiräte, koordiniert deren Aufgaben und hält die enge Verbindung zum DTT-Vorstand.

Prof. Dr. Klaus-Dirk Schmitz ist seit 1991 DTT-Mitglied und seit 1999 ununterbrochen im DTT-Vorstand.

Kontaktadresse
schmitz@dttev.org
www.dttev.org

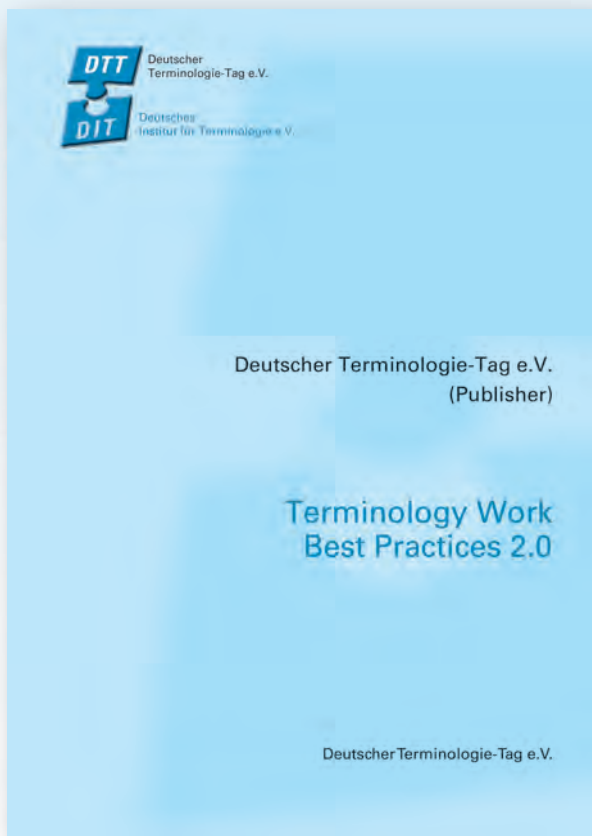
DTT-Symposion 2023: Termin jetzt vormerken!

Das nächste DTT-Symposion findet vom
2. bis 4. März 2023 in Mannheim statt.

Terminology Work Best Practices 2.0

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

kompakt | praxisnah | auf den Punkt



Geballtes Terminologiewissen 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**

Termine

Terminologie³ –

Aus der Praxis, für die Praxis

22. bis 23. Juni 2022, online und kostenlos

Weitere Informationen und Anmeldung unter terminologiehochdrei.com

Terminology Summer School 2022

1. bis 4. Juli 2022 in Athen und online

Weitere Informationen und Anmeldung unter shop.termnet.eu

DTT-Grundlagenseminar 2022

14. Oktober 2022, online

DTT-Vertiefungsseminar 2022

17./18., 24./25. November
und 2./3. Dezember 2022, online

Weitere Informationen und Anmeldung unter dttev.org/fortbildung

Einsendeschluss für die nächste Runde des DTT-Förderpreises

30. Juni 2022

Weitere Informationen unter dttev.org/dtt-foerderpreis

Concept Systems and Frames in Terminology (CSFT)

28. bis 29. September 2022 in Innsbruck

Weitere Informationen und Anmeldung unter uibk.ac.at/congress/csft-2022

Continental Connections – International Indexing Conference 2022

17. bis 18. Oktober 2022 in Berlin

Weitere Informationen und Anmeldung unter dgi-info.de



tekom
Jahrestagung **2022**

S T U T T G A R T

8. – 10. NOVEMBER 2022

tekom
MESSE22

jahrestagung.tekom.de



Ein Haus für die deutsche Sprache

Interview mit Architekt Dennis Ewert

In Ausgabe 2/2020 der edition wurde bereits über die Pläne berichtet, ein „Forum Deutsche Sprache“ in Mannheim zu errichten.

Mit seiner Masterarbeit hat der Architekt Dennis Ewert 2019 einen Entwurf für dieses „Haus für die deutsche Sprache“ vorgelegt. Dafür musste er sich sehr intensiv damit auseinandersetzen, wie sprachliche Mittel die architektonische Gestaltung prägen können. Dennis Ewert hat ein Raumprogramm für dieses Museum entworfen, in dem die aus der Recherche erarbeiteten Räume in den drei vorgegebenen Themenbereichen „Sprechen und Hören“, „Schrift und Literatur“ sowie „Mimik, Gestik und Situation“ inhaltlich umrissen werden.

Dennis Ewert hat an der Staatlichen Akademie der Bildenden Künste Stuttgart Architektur studiert und arbeitet heute als freier Architekt in Mannheim.

Herr Ewert, Sie schreiben in Ihrer Arbeit zu Sprache und Architektur: „Das wichtigste Merkmal natürlicher Sprachen und ihrer Komponenten, also ihrer sprach-

lichen Zeichen, ist ihre Repräsentationalität, also die Tatsache, dass sie semantische, propositionale Gehalte aufweisen. Sprachen lassen sich innerhalb der Bereiche Syntax, Semantik und Pragmatik untersuchen. Diese lassen sich auf der Oberfläche des Museums in ihrem wohl nahestehendsten architektonischen Element ablesen, der Mauer. In der Anordnung der Steine, ihrer Materialität und der Oberflächenbeschaffenheit repräsentieren sie diese.“

Wie sind Sie darauf gekommen, Sprache mit Architektur in Verbindung zu setzen?

Den inhaltlichen Weg für meine Masterarbeit bereitete meine wissenschaftliche Arbeit mit dem Titel „Regionale und überregionale Verwendung deutscher Natursteine in der Architektur“. Durch meine Untersuchung konnte ich feststellen, dass Steine allein durch die Art und Weise ihres Einsatzes auf ganz unterschiedlichen Bedeutungsebenen wirken können. Beispielsweise kann durch die vermehrte regionale Verwendung eines bestimmten Steins eine ortstypische und vielleicht sogar unverwechselbare Tonalität geprägt werden, es können Kontinuitäten geschaffen und

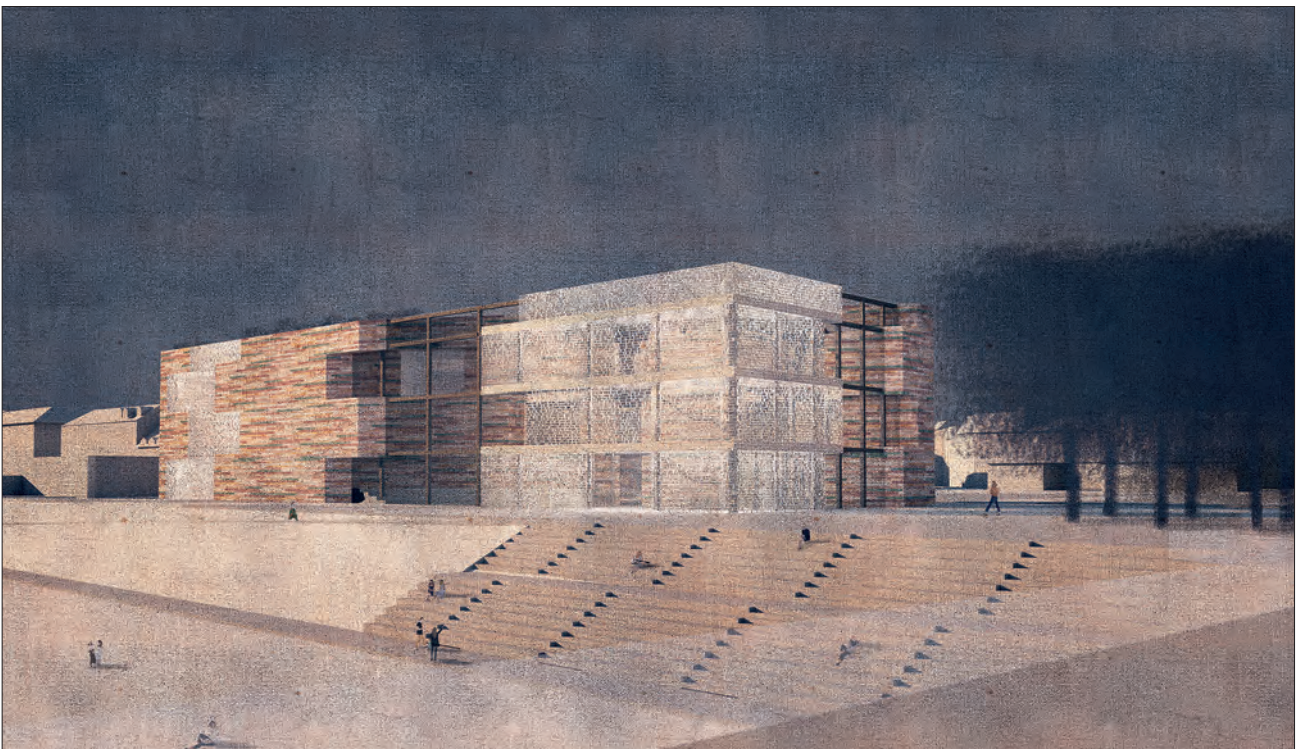


Abb. 1: Entwurf für das Forum Deutsche Sprache: Gesamtansicht



Abb. 2: Entwurf für das Forum Deutsche Sprache: Foyer

Identifikationsmerkmale gebildet werden. In der Region um Soest in Nordrhein-Westfalen beispielsweise wird seit Jahrhunderten der grüne Anröchter Sandstein verbaut. In der Altstadt von Soest finden sich daher zahlreiche grüne historische wie zeitgenössische Bauwerke. Für das unvoreingenommene Auge stellt sich der Eindruck von moosbewachsenen Mauern ein. Doch es handelt sich hier um einen ganz besonderen regionalen Code.

Gleichzeitig sind Gesteinsarten aber auch in der Lage, überregional qualitative Effekte zu erzeugen und über die vielleicht sogar profane Nutzung in der Abbauregion hinauszuwachsen. So ist der Villmarer Kalkstein, auch Lahnmarmor genannt, im Foyer des Empire State Building in New York, der Eremitage in Sankt Petersburg und sogar im Kreml zu finden. Ob geplant oder unbeabsichtigt – wir befinden uns hier schon im Bereich der Repräsentation.

Wie repräsentiert die Anordnung der Steine, ihre Materialität und ihre Oberflächenbeschaffenheit eine Mauer? Was hat eine Mauer mit Syntax, Semantik und Pragmatik zu tun?

Die Textur eines verwendeten Steins allein beschreibt die Ausdrucksfähigkeit noch nicht erschöpfend. Auch die Form

und die Anwendung im Kontext sind Träger von Bedeutungen, die man in klassischen sprachwissenschaftlichen Kategorien untersuchen kann. Mit meinem Entwurf für ein Haus für die deutsche Sprache habe ich diesen Aspekt visualisieren und für die zukünftigen Besucher:innen ablesbar machen wollen. Und zwar durch das unseren Sprachen wohl am besten entsprechende architektonische Element: die Mauer. Die Tradierung des Schichtens von Steinen in der Architektur bildet die syntaktische Struktur des Systems. Es handelt sich hier um die rein formale Beziehung zwischen den Zeichen, um Regeln der Verknüpfung und um Ordnungs- und Anordnungsfunktionen. Als kleinstes Element der Mauer trägt der einzelne Stein einen architektonisch-semantischen Gehalt. Es geht dabei um die Zuordnung von Bedeutungen zu den Zeichen. Durch seine besondere Oberflächenbehandlung wird ein Bezug zu einer bestimmten handwerklichen Bearbeitung hergestellt. Die individuelle Oberfläche, die einzigartige Erscheinung, die Art und Weise, wie Licht und Schatten auf dem Stein wirken, löst ganz verschiedene Assoziationen beim Betrachter aus. Jede Reihe, die aus den unterschiedlich behauenen Steinen gebildet wird, ist in einem anderen Naturwerkstein ausgeführt. Durch die regionalen Materialitäten entstehen unterschiedliche Verbindungen zwischen dem Betrachter

und der Wand. Diese pragmatische Dimension umfasst die Wirkung von Zeichen und Bedeutungen auf bestimmte Empfänger:innen. Als mit der Materialität Vertrauter erkennt man lokale Texturen der Heimat und steht bereits in einer bestimmten Beziehung zu diesen. Aber auch Außenstehende haben durch individuelle Erfahrungen – zum Beispiel durch Reisen, Besuche, die überregionale Verwendung des Baustoffs – eine ganz eigene Beziehung zu den Steinen. Die Zeichen auf der pragmatischen Ebene sind somit vieldeutig. Sinn der Vieldeutigkeit der Zeichen ist es, durch die Vielfalt der Codes möglichst viele Gruppen unserer pluralistischen Gesellschaft zu erreichen, um eindeutig für viele zu sein. Nicht Vieldeutigkeit ist hier das Thema, sondern Komplexität.

Sie haben Ihre Raumpäne von einem Steinmetz in Stein hauen lassen. Jedes Geschoss ist aus einem anderen Gestein. Was wollten Sie dadurch erreichen?

Bei meinem Entwurf spielt die sinnliche, ästhetische Erfahrung eine entscheidende Rolle. Zweidimensionale Pläne auf Papier können nur einen Teil dieser Erfahrung transportieren. Hinzu kommt, dass grafische Darstellungen und Visualisierungen oft den Ansprüchen, die an sie gestellt werden, in der Realität nicht gerecht werden können und vielleicht sogar nicht wollen. Die Steinplatten konnten mit allen Sinnen erfahren werden. Man konnte die Pläne anfassen und die Unterschiedlichkeit der Steine erforschen. Grundsätzlich finde ich es überlegenswert, in der Architekturkommunikation – insbesondere den Architekturwettbewerben –

weg von gewohnten Regeln wie genormten Papiergrößen, einheitlichen Darstellungsformaten und Visualisierungsanforderungen zu gehen. Mein Beispiel zeigt, dass es auch andere und präzisere Darstellungsformen geben kann.

Sie sprechen in Ihrer Arbeit von „Urban Coding“ und dass Städte Zeichensysteme sind. Was bedeutet Urban Coding? Und welche kulturellen Aspekte verbinden sich damit für Sie?

Städte können als sprachliche Systeme verstanden werden. Ihre Zeichen können decodiert und gelesen werden. Zusammen bilden die Zeichen die Identität einer Stadt. Die Wahl eines Materials in Kombination mit einer Farbe ist eine Ebene und Möglichkeit der Codierung einer Stadt. Das funktioniert sogar umgekehrt. Wir alle kennen den Farbton Siena, der von der Tonalität der italienischen Stadt abgeleitet ist und sich aus der Farbigekeit der die Stadt umgebenden Erde ergibt. Turin beispielsweise nutzt die Möglichkeiten des Urban Coding, um das Image der grauen Industriestadt aufzuwerten. Die Bedeutungen der komplexen städtischen Zeichensysteme können gelesen werden, sofern der Betrachter mit den kulturspezifischen Besonderheiten des Ortes vertraut ist.

Farbcodes funktionieren auch wie eine externe Festplatte. Erinnerungen, die mit entsprechenden Farbprofilen abgespeichert sind, können einem Ort zugewiesen werden. So können auch spezifische Zeitspannen unserer eigenen Biografie lokal eingeordnet werden. Daran lässt sich erkennen, dass Begriffe wie Heimat und Ferne stark atmosphärisch

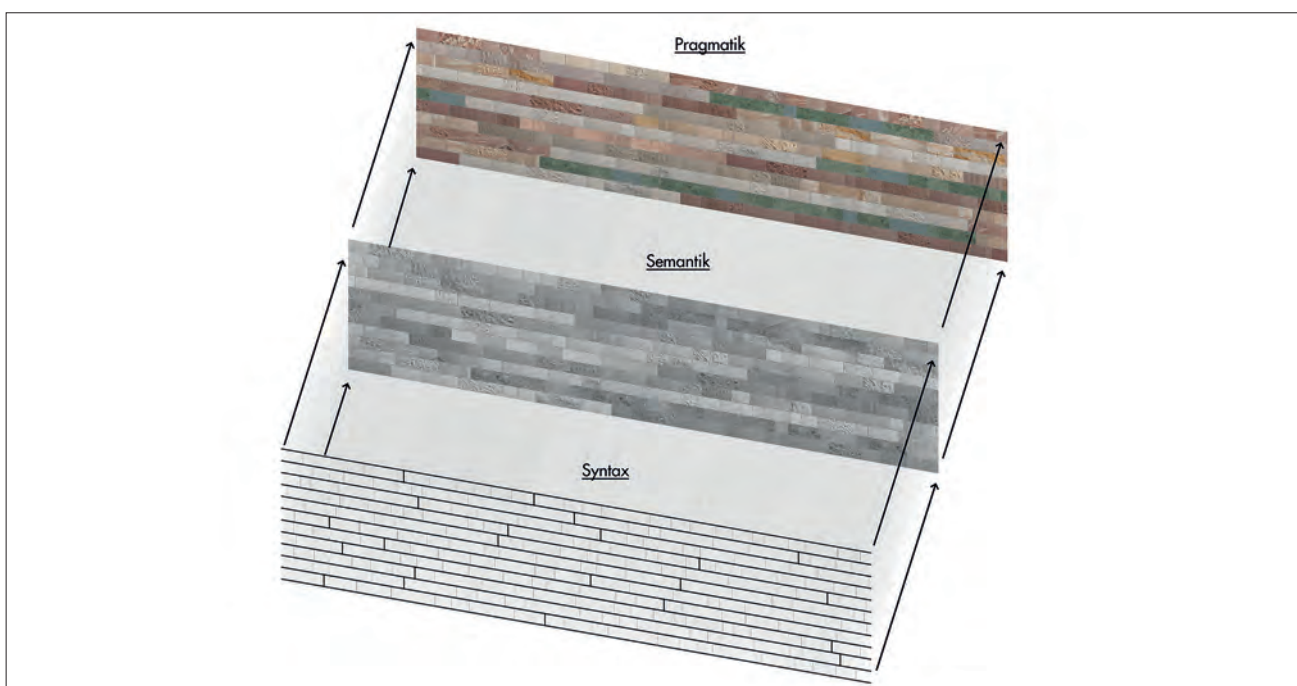


Abb. 3: Syntax, Semantik und Pragmatik einer Mauer

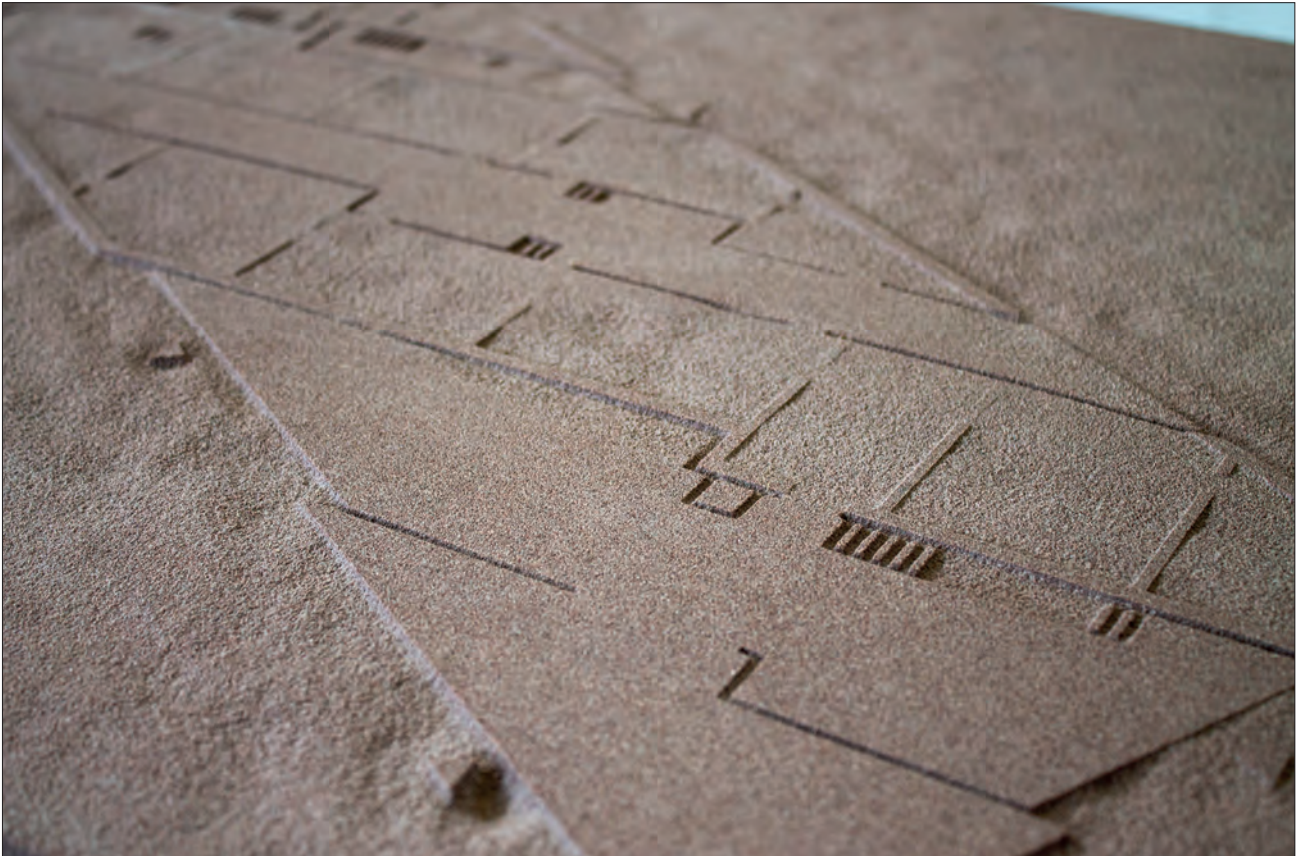


Abb. 4: In Stein gehauener Architekturplan

geprägt sind. Die Wahrnehmungspsychologie attestiert der einheitlichen Farbigkeit positive Effekte. Sie vereinfacht die Lesbarkeit. Ein aus der Systematik herausfallendes Gebäude dagegen stört diese Lesbarkeit wie ein verdrehtes Wort. In unserer pluralistischen Gesellschaft scheint sich ein Zielkonflikt zwischen der Objektästhetik und der Raumästhetik einzustellen. Während es der ersten um das individuelle Geltungsbedürfnis geht, entspricht letztere dem kollektiven Wunsch nach Rücksichtnahme, Einigkeit und städtebaulichen Qualitäten. Eine gemeinsame Tonalität eines Ortes kann diesen Widerspruch auflösen. In London wird beispielsweise seit dem großen Brand im Jahr 1666 vielfach der weiß-graue Portland Stone verwendet. Das Material verbindet seitdem unterschiedlichste Gebäude aus verschiedenen Epochen. Eine erkennbare Gemeinsamkeit kann also auch plurale Positionen zusammenhalten.

Können Sie uns konkrete Beispiele für Urban Coding nennen, die kulturelle Besonderheiten widerspiegeln?

Auch aus dem Kontext gerissene oder universelle Materialien können in der architektonischen Verwendung sinnvoll sein. Glas und Stahl verkörpern beispielsweise Internationalität. Passend für eine Stadt wie New York. Ein besonderes Beispiel für die Verwendung entwurzelter regionaler

Codes ist Venedig. Über die Jahrhunderte sammelten sich im Stadtbild Materialien und Ornamente aus verschiedensten Ländern. Der Markusdom weist zahlreiche byzantinische Elemente auf und die verschiedenfarbigen Säulen stammen aus unterschiedlichen Ländern. Zwischen die gotischen Formen der Altstadt mischen sich außerdem orientalische Elemente. Venedig war eine Stadt der Seefahrer, die im gesamten Mittelmeerraum tätig waren. Die Altstadt zeugt von dieser Tätigkeit und kann als Collage verschiedenster Materialitäten und Stile gesehen werden.

Um nun den Bogen zur Terminologie zu spannen: Wo findet in Ihrem Modell die Terminologie ihren Platz? Wo können wir als Terminologen Parallelen für unsere Modelle und unsere Arbeit ableiten?

Wir Architekten bewegen uns, wie Sie sehen, tagtäglich in ästhetischen Sprachwelten und gestalten passende Botschaften. Und auch wir gehen, wenn nötig, auf Terminologierecherche in Datenbanken. Sei es das passende Moodbild für das Materialkonzept oder die Beratung in einer der zahlreichen professionell geführten Materialbibliotheken. Dabei handelt es sich um Einrichtungen, in denen Produkte mit Informationen zu Materialität und Technologie zugänglich sind. Teilweise handelt es sich sogar um unbekannte,

neue und innovative, aber auch vergessene Baustoffe. Wenn Sie so wollen, können diese Bibliotheken auch als eine Art architektonische Terminologiedatenbank verstanden werden. Wobei die Ästhetik natürlich nur ein Aspekt der Recherche ist. Weitere Kategorien sind beispielsweise statische oder bauphysikalische Eigenschaften. Unser Exkurs zum Urban Coding zeigt, dass sich Planende und Auftraggebende der zahlreichen Zeichen bewusst bedienen, um Botschaften zu formulieren. Aber auch die pragmatische Verwendung verfügbarer oder lokaler Baumaterialien ist Träger von Bedeutungen, wie meine Analysen in der Arbeit „Regionale und überregionale Verwendung deutscher Natursteine in der Architektur“ verdeutlicht. „Man kann nicht nicht kommunizieren“ sagte Paul Watzlawick. Ich glaube, dieser Grundsatz gilt auch für die Architektur. Eine Fassade kann uns immer etwas erzählen, wenn wir nur zuhören.

Der Vergleich von Mauern mit sprachlichen Merkmalen und die Vorstellung des Urban Coding als ein Zeichensystem, das zur Verständigung über die Benutzung von Wörtern bzw. mündlicher oder schriftlicher Sprache weit hinausgeht, ist überaus spannend und wirft einen ganz anderen Blick auf Sprache und Terminologie. Herr Ewert, vielen Dank für dieses Gespräch.

Literaturangaben:

- [1] Regionale und überregionale Verwendung deutscher Natursteine in der Architektur - Dennis Ewert - Staatliche Akademie der Bildenden Künste Stuttgart 2019
- [2] STADTFARBEN Strategische und zukunftsfähige Planung von Stadtraum und Atmosphäre durch Farbmasterplanung - Markus Schlegel & Gerhard Fuchs - Societätverlag 2013. <https://lebendige-stadt.de/pdf/Stadtfarben.pdf>
- [3] Hans Bernoulli „Die Einheit des Materials im Aufbau der Städte“, Zeitschrift „Architektonische Rundschau“, 25. Jg., 1909, H. 9
- [4] John W. Root „The City House in the West“, Zeitschrift „Scribner's Magazine“, Bd. 8, Juli-Dezember, 1890

Das Interview führte:
Angelika Ottmann
redaktion@dttev.org



Dennis Ewert ist freier Architekt. Er hat Architektur an der Staatlichen Akademie der Bildenden Künste in Stuttgart studiert und ist Co-Vorsitzender und Gründungsmitglied des Vereins M O F A Mannheims Ort für Architektur e.V.

Kontaktadresse
info@dennis-ewert.com
www.dennis-ewert.com

Anzeige



DTT-Fortbildung – online

DTT-Vertiefungsseminar

„Terminologiewerk – Projekte, Prozesse, Datenaustausch“

17./18. und 24./25. November sowie 2./3. Dezember 2022
online

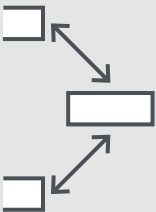
Weitere Informationen und Anmeldung unter: dttev.org/fortbildung

Stand: 1. Juni 2022 | Änderungen vorbehalten.

ENTERPRISE TERMINOLOGY MANAGEMENT



Unternehmensweiter
Terminologiezugriff:
Intuitiv, extrem anpassbar,
informativ



Wissensmanagement mittels
Concept Maps und Taxonomien



Kollaborative Workflows in allen
Sprachen



Single Source of Truth -
Die Termbank als Basis für alle an-
deren Anwendungen wie CAT-Tools,
MT, ERP, usw.



Terminologieprüfung bereits in der
Ausgangssprache mit Plug-ins für
MS Word, Adobe InDesign,
XMetaL, <oXygen>, Visual Studio, Vi-
sual Studio Code & Trados Studio



Mehr Information unter:
[KALEIDOSCOPE.AT/ED22](https://www.kaleidoscope.at/ED22)