



Working paper

2017

Open Access

This version of the publication is provided by the author(s) and made available in accordance with the copyright holder(s).

Was haben Wissensmodellierung, Wissensstrukturierung, künstliche Intelligenz und Terminologie miteinander zu tun?

Prof. Dr. Drewer, Petra; Dr. Massion, François; Pulitano, Donatella

How to cite

PROF. DR. DREWER, Petra, DR. MASSION, François, PULITANO, Donatella. Was haben Wissensmodellierung, Wissensstrukturierung, künstliche Intelligenz und Terminologie miteinander zu tun? 2017

This publication URL: <https://archive-ouverte.unige.ch/unige:100073>



**Deutscher
Terminologie-Tag e.V.**

**Deutsches
Institut für Terminologie e.V.**

Petra Drewer, François Massion, Donatella Pulitano

Was haben Wissensmodellierung, Wissensstrukturierung, künstliche Intelligenz und Terminologie miteinander zu tun?

Inhalt

1	Herausforderungen und Entwicklungen	3
1.1	Die digitale Gesellschaft.....	3
1.2	Big Data und Informationsexplosion.....	3
1.3	Industrie 4.0 und Internet der Dinge	4
1.4	Künstliche Intelligenz	6
1.5	Intelligente Assistenten, Digital Assistants, Bots	8
1.5.1	Assistentensysteme	8
1.5.2	Metaassistenten	8
1.5.3	„Smarte“ Maschinelle Übersetzung	10
1.6	Konsequenzen.....	10
2	Wissensmodellierung mit Hilfe von Ontologien	11
2.1	Ontologien	11
2.2	Verknüpfte Ontologien und Linked Open Data	12
3	Terminologie.....	14
3.1	Die Terminologiewissenschaft als Basis der Terminologiearbeit.....	14
3.2	Aufgabenspektrum in der Terminologiearbeit	14
3.3	Erstellen von Begriffssystemen.....	18
3.3.1	Grundlagen	18
3.3.2	Sprachlich-kulturelle Unterschiede	20
3.3.3	Begriffssysteme in der Praxis.....	21
4	Wechselwirkungen und Austausch zwischen innovativen Technologien und Terminologie	22
4.1	Terminologie und Ontologien	22
4.1.1	Ausgangssituation.....	22
4.1.2	Bessere Qualität von Ontologien auf sprachlicher Ebene	23
4.1.3	Erarbeitung von ontologischen Datenbeständen.....	23
4.2	Annäherungen von Terminologiedatenbanken und Wissensdatenbanken	24
4.3	Veränderungen im Berufsprofil von Terminologen	26
5	Ausblick	27
	Literaturverzeichnis.....	28
	Abbildungsverzeichnis.....	28

1 Herausforderungen und Entwicklungen

1.1 Die digitale Gesellschaft

Der Beginn des 21. Jahrhunderts ist durch einige große Trends und Bemühungen gekennzeichnet, die in den nächsten Jahrzehnten große Auswirkungen auf unseren Alltag, aber auch auf unser Berufsleben haben werden. Die wichtigsten im Überblick:

- Weiterentwicklung der Informationsgesellschaft: Wissen als entscheidende Ressource; rasante Zunahme von Daten und Informationen aus unterschiedlichen Quellen, in unterschiedlichen Medien und Formaten und für unterschiedliche Zielgruppen (Big Data)
- Internet der Dinge
- Wunsch nach digitaler Unterstützung in allen Bereichen des Alltagslebens und der Industrie (digitaler Wandel)

Um diese Anforderungen zu erfüllen bzw. die zunehmenden Datenmengen beherrschbar und nutzbar zu machen, werden Arbeiten in folgenden Bereichen unternommen:

- verstärkte Bemühungen und Fortschritte im Bereich selbstlernende Maschinen und Systeme
- Künstliche Intelligenz
- Entwicklung von intelligenten Assistenten

Die Basis für die Umsetzung von intelligenten Maschinen und Assistenten ist der Zugriff und das „Verstehen“ einer Unmenge von Daten. Die Hauptherausforderungen sind also:

- Aufbereitung von Daten und Informationen in maschinenlesbarer Form
- Modellierung und Strukturierung von Wissen, insbes. in Form von computerlesbaren Ontologien, semantischen Netzen etc.

1.2 Big Data und Informationsexplosion

Die Wissensexplosion und die Verteilung von Informationen aus menschlichen Aktivitäten oder aus Maschinen sowie Software nehmen kaum vorstellbare Dimensionen an.

Unter „Big Data“ versteht man Datenmengen, die aufgrund ihrer Größe, ihrer Komplexität, der Geschwindigkeit ihrer Entstehung sowie ihrer Übertragung, der Vielfalt ihrer Datentypen und -quellen sowie v. a. aufgrund ihrer schwachen oder fehlenden Strukturierung mit herkömmlichen Datenverarbeitungsmethoden nicht mehr auswertbar sind.

„Hintergrund der Diskussionen um Big Data ist der starke Anstieg des weltweiten Datenvolumens. Eine Vielzahl unterschiedlicher Quellen ist dafür verantwortlich: Sensordaten, Maschinendaten, Log-Daten, das WorldWideWeb oder RFID-Chips. Im Jahre 2011 knackte das weltweite Datenvolumen die Zettabyte-Barriere (1 mit 21

Nullen) und ein Ende des Wachstums ist nicht in Sicht. 2020 sollen es bereits 35 Zettabyte sein.” (<http://www.softselect.de/wissenspool/big-data>, Zugriff 13.02.2017)

Typische Quellen und Arten von Big Data sind:

- Sprachdaten (v. a. auch mehrsprachig)
- Daten, die von Behörden oder Unternehmen gesammelt und produziert werden
- Informationen und Interaktionen aus sozialen Netzwerken
- Nutzungsdaten (z. B. Reise-/Bewegungsdaten, Gesundheitsdaten-Tracking) von Geräten oder Systemen wie Navigationssysteme, Smartphones, Wearables
- vernetzte Technik im Alltag (z. B. in Häusern und Kraftfahrzeugen) und in der Produktion (Smart Factory, vgl. Abschnitt 1.3)

Die Aufgabe von Wissensspezialisten ist es, aus diesen vielen Quellen und Beständen Wissen zu generieren, das zu unterschiedlichen Zwecken verwendet werden kann.

Letztlich wird das Big-Data-Phänomen durch die Mehrsprachigkeit noch um ein Vielfaches verstärkt. Die ohnehin schon unbeherrschbare Menge an Daten wird durch die Multiplikation in viele Sprachen schier endlos. Gerade hier sind also weitere Verfahren notwendig, um die Informationen aus diesen mehrsprachigen Daten zugänglich zu machen und in computerlesbares Wissen umzuwandeln.

Während viel über Big Data gesprochen wird, darf jedoch nicht übersehen werden, wie wichtig es ist, auch brauchbare Ansätze zur Verarbeitung kleinerer Datenbestände (Small Data) zu entwickeln. Auch kleinere Datenpakete (z. B. Konversation in Echtzeit) müssen so aufbereitet werden, dass sie für Menschen und/oder Maschinen verständlich und verarbeitbar sind.

1.3 Industrie 4.0 und Internet der Dinge

„Industrie 4.0‘ [...] steht für ein ‚Zukunftsprojekt‘ [...] der deutschen Bundesregierung. Die sog. vierte industrielle Revolution, auf welche die Nummer verweist, zeichnet sich durch Individualisierung (selbst in der Serienfertigung) bzw. Hybridisierung der Produkte (Kopplung von Produktion und Dienstleistung) und die Integration von Kunden und Geschäftspartnern in Geschäfts- und Wertschöpfungsprozesse aus. Wesentliche Bestandteile sind eingebettete Systeme sowie (teil-)autonome Maschinen, die sich ohne menschliche Steuerung in und durch Umgebungen bewegen und selbstständig Entscheidungen treffen [...]. Die Vernetzung der Technologien und mit Chips versehenen Gegenstände resultiert in hochkomplexen Strukturen und cyber-physischen Systemen (CPS) bzw. im Internet der Dinge.“ (Gabler, <http://wirtschaftslexikon.gabler.de/Definition/industrie-4-0.html>, Zugriff: 14.02.2017)

Im Unternehmensumfeld geht es im Wesentlichen darum, dass Geräte, sog. eingebettete Systeme („Embedded Devices“) oder komplette Anlagen dazu fähig sind, selbstständig Entscheidungen zu treffen, auf teils unerwartete Ereignisse zu reagieren und miteinander zu arbeiten. Im Rahmen von Industrie 4.0 spricht man hier von der sog. „Smart Factory“ bzw. „intelligenten Fabrik“.

An einem Beispiel: Die ThyssenKrupp Elevator AG hat zusammen mit Microsoft und dem IT-Dienstleister CGI ein intelligentes vernetztes Monitoring-System entwickelt. Über die Cloud vernetzt das Unternehmen seine Aufzüge und überwacht so sämtliche Funktionen – von der Kabinengeschwindigkeit über die Zuladung bis hin zu den Türmechanismen. Auf diese Weise ist ein innovatives Wartungssystem entstanden, das nicht mehr auf Störungen reagiert, sondern es ermöglicht, auf Basis von Echtzeitdaten bereits vor dem Ausfall eines Aufzugs Maßnahmen zu ergreifen. Aus Big Data werden so Smart Data. Weitere Details unter <http://www.thyssenkrupp-aufzuege.de/unternehmen/neuigkeiten/detail/news/cebit-thyssenkrupp-elevator-praesentiert-innovative-serviceloesung/> (Zugriff: 21.02.2017).

Die Einsatzmöglichkeiten des Internets der Dinge sind also außerordentlich vielfältig. Sie reichen von der oben beschriebenen Smart Factory bis hin zu smarten Informationsagenten bzw. -assistenten, die eigenständig Daten sammeln, sie in Informationen umwandeln und für verschiedene Zwecke bereitstellen. Auch der Einsatz von „Wearables“ (tragbare Computersysteme) fällt in diesen Bereich. Anders als klassische Anwendungen wie Hörgeräte und Herzschrittmacher bauen die neuen Produkte auf Vernetzung, Kontextsensitivität und Ansätze der Künstlichen Intelligenz, die den Menschen in Alltag und Beruf nahezu unmerklich unterstützen. Ein weiteres Beispiel für eine mögliche Nutzung des Internets der Dinge ist eine „Smart Documentation“, die mit dem Anwender kommuniziert und Handlungen wie beispielsweise das Setzen von Parametern in einer Konfigurationsmaske übernimmt. Bei all diesen Anwendungen werden gleichzeitig wieder große Mengen an Daten gesammelt, die für andere Nutzungen und Auswertungen zur Verfügung stehen.

Die drei Hauptfunktionen eines Geräts oder einer Applikation, die im Internet der Dinge funktioniert, sind:

- (1) Fähigkeit, das Umfeld zu erkennen
- (2) Fähigkeit, das Umfeld zu verstehen und mit ihm zu kommunizieren
- (3) Fähigkeit, mit dem Umfeld zu interagieren und selbstständig Aktionen auszulösen

Terminologen und Linguisten sind speziell gefragt, wenn es um die zweite Hauptfunktion geht, denn sie verfügen über das Wissen und die Expertise, um Kommunikation zwischen Mensch und Maschine (Software) in natürlicher Sprache zu optimieren.

- Die Korpuslinguistik stellt Methoden bereit, um Sprachdaten (geschrieben und/oder gesprochen) digital so aufzubereiten, dass sie im Anschluss bei der automatischen Analyse von großen Datenmengen genutzt werden können.
 - Die Computerlinguistik verbindet Sprachwissenschaft und Informatik und widmet sich der gesamten Algorithmisierung von natürlicher Sprache. Das Spektrum der Computerlinguistik umfasst die Spracherkennung sowie morphologische, syntaktische und semantische Analysen;
-

es findet Anwendung z. B. in der Maschinellen Übersetzung, der automatischen Rechtschreib- und Grammatikkorrektur, bei der Suchmaschinenoptimierung (SEO), in Diktiergeräten (Spracherkennung), in Vorlesegeräten (Sprachsynthese und -ausgabe) etc.

- Die Terminologiewissenschaft rückt neben der Sprache das Wissen in den Vordergrund, das mit/in der Sprache transportiert wird, fragt somit nicht nur nach Wörtern, sondern v. a. nach deren Bedeutungen. Terminologen sprechen hier von Benennungen (Wörtern) und Begriffen (Bedeutungen). Die Terminologiewissenschaft liefert also Methoden, mit denen Begriffe und Begriffsverknüpfungen dargestellt und Benennungen mehrsprachig zur Verfügung gestellt werden können (zu weiteren Verbindungen siehe Kapitel 3 und 4).

1.4 Künstliche Intelligenz

Der Begriff der „artificial intelligence“ (dt. „Künstliche Intelligenz“, KI) entstand im Jahr 1956, auch wenn bereits früher in dieser Richtung geforscht wurde. Seine Definition ist schwierig und nicht abgeschlossen, zumal der grundlegende Begriff der Intelligenz nicht eindeutig geklärt ist. Typischerweise versteht man unter KI den Versuch, eine menschenähnliche Intelligenz nachzubilden, also Maschinen (Programme) zu entwickeln, die selbstständig Situationen erkennen und Probleme lösen können. Der Bereich des maschinellen Lernens ist hier von besonders großer Bedeutung.

Um diese Ziele zu erreichen, sind komplexe Algorithmen gefragt, wie z. B. beim gerade aktuellen Projekt von Google DeepMind, bei dem Google Software entwickelt, die in der Lage ist, selbstständig zu lernen, wie komplexe Fragen behandelt werden müssen, wodurch beispielsweise die Leistung von maschinellen Übersetzungssystemen deutlich verbessert wird.

Neben der maschinellen Intelligenz und ihren Lernprozessen spielt v. a. die Frage nach der Kommunikation zwischen Mensch und Maschine eine wichtige Rolle (siehe dazu auch Abbildung 1).

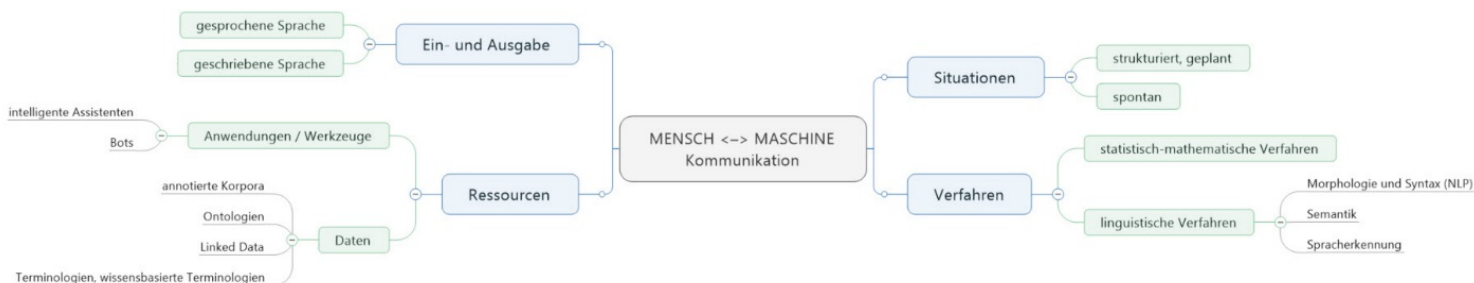


Abbildung 1: Mensch-Maschine-Kommunikation

Typische **Anwendungen bzw. Werkzeuge**, bei denen KI eine Rolle spielt, sind intelligente Assistenten oder Bots, die auf der Basis verschiedener Daten (z. B. annotierte Korpora, Ontologien, Linked Data) mit Menschen kommunizieren. Die **Situationen**, in denen diese Maschinen mit Menschen kommunizieren,

sind teilweise strukturiert und geplant, teilweise jedoch auch spontan und damit schwer steuerbar. Die **Ein- und Ausgabe der Daten** erfolgt im Regelfall über natürliche Sprache (geschrieben oder gesprochen).¹

Um die eingegebenen Daten, aber auch „Big Data“ zu verstehen, kommen verschiedene **Analyse- und Verarbeitungsverfahren** zum Einsatz, die sich gegenseitig nicht ausschließen, sondern eher ergänzen können:

- statistische und mathematische Verfahren
- morphologische und syntaktische Verfahren (Natural Language Processing, NLP)
- semantische Verfahren

Während statistische und mathematische Verfahren bei der Analyse großer Datenmengen sehr erfolgreich sind, sind sie bei der Analyse von Daten, die weniger häufig vorkommen, weniger tauglich. In der Statistik spricht man bei diesem relativ langen flachen Teil einer Häufigkeitsverteilungskurve vom „Long Tail“.

Der Ansatz, künstliche neuronale Netze einzusetzen, um Maschinen zu helfen, selbstständig aus Situationen zu lernen, ist vielversprechend, wie das aktuelle Beispiel „AlphaGo“ zeigt, bei dem das von Google DeepMind entwickelte Programm in der Lage ist, einen Go-Spiel-Meister zu schlagen.

Die rein statistischen Methoden stoßen an ihre Grenzen und versagen insbesondere in Situationen, bei denen die Kommunikation spontan und frei erfolgt (Kontextbezug, Mehrdeutigkeit) oder bei denen die Datenmengen verhältnismäßig gering sind (Small Data). Hier werden linguistische und v. a. semantische Verfahren den Ausschlag geben.

Linguisten und Terminologen interessieren sich besonders für alle Aspekte von KI, die natürliche Sprache involvieren. Es geht dabei um die Extraktion von Informationen aus geschriebenen und/oder gesprochenen Texten (im Regelfall nicht nur ein-, sondern mehrsprachig) sowie um das Modellieren von Wissen mit Hilfe von Begriffen und Benennungen, um die Kommunikation zwischen Mensch und Programmen bzw. zwischen Programmen untereinander auf Basis natürlicher Sprache zu ermöglichen.

Innerhalb der linguistischen Herangehensweise berücksichtigen die morphologischen Verfahren auf der Basis der vorab stattfindenden Tokenisierung der Texte insbesondere Flexionsendungen, um eine Lemmatisierung vorzunehmen bzw. grammatische Informationen auszulesen. Bezieht man die Syntax mit ein, so werden formal-strukturelle Beziehungen innerhalb der Sätze analysiert, bis letztlich die Semantik die Berücksichtigung der Bedeutungen einbringt. Gerade hier zeigen sich oft Probleme, da Bedeutungen und damit menschliches Wissen schwer fassbar sind. Gleichzeitig aber bieten sich gute Ansätze, um terminologische Kenntnisse und Arbeitsweisen einzubinden, da hier mehrsprachige Terminologien bereits taxonomisch oder ontologisch aufbereitet werden und sich so in wissensbasierte Terminologien verwandeln könnten (vgl. Kapitel 3).

¹ Bei der schriftlichen Texteingabe entfällt der zusätzliche Analyseschritt der Spracherkennung.

1.5 Intelligente Assistenten, Digital Assistants, Bots

1.5.1 Assistentensysteme

Intelligente Assistenten (auch: Intelligente Persönliche Assistenten (IPA) oder Sprachassistenten) sind kleinere Dienstprogramme, die mit Menschen – zunehmend in natürlicher Sprache – kommunizieren und bestimmte Aufgaben ausführen. Dazu suchen sie selbstständig Informationen in verschiedenen Quellen und tauschen sich mitunter mit anderen Assistenten aus. Ein typisches Beispiel ist ein Assistent, der Informationen für eine Flugreise (Konditionen, Datum, Verfügbarkeit) selbstständig sucht und zusammenstellt. Ein anderes, weniger anspruchsvolles Beispiel sind Preisvergleichsportale.

Spracherkennungssysteme sind in der Lage, natürliche gesprochene Sprache zu erkennen und zu verarbeiten. Beispiele für diesen Bereich sind Siri, Cortana oder Google Now. Auch Amazon Echo (Alexa) gehört zu den IPAs, obwohl sich Basis und Zielsetzung leicht unterscheiden. Mittlerweile gibt es intelligentere (wissensbasierte) Systeme wie Watson (IBM), die für umfangreichere semantische Suchen, inkl. Hypothesenbildung und Informationsableitung, gedacht sind.

1.5.2 Metaassistenten

Im Hinblick auf die Assistenten zeichnet sich eine interessante und voraussichtlich folgenreiche Entwicklung ab. Einige Firmen arbeiten an einer Art Metaassistent. Anstatt individuelle Assistenten für jede einzelne Aufgabe zu bemühen (Reise, bestimmte Einkäufe, Wissenssuche), soll eine einzige persönliche Anwendung diese Aufgaben übernehmen und sich mit den entsprechenden spezialisierten Assistenten in Verbindung setzen, fortlaufend lernen und dadurch effizienter werden. Diese zentrale Anwendung würde alle erforderlichen Informationen über den Anwender sammeln (Big Brother!), um somit die Präferenzen des Benutzers optimal zu berücksichtigen. Viv, die Weiterentwicklung von Siri, wird in dieser Richtung entwickelt (<http://viv.ai>).

Abbildung 2 (Quelle: <https://www.wired.com/2014/08/viv/>) illustriert eine komplexe Anfrage an Viv, zu deren Bearbeitung vom Assistenten verschiedene Datenquellen konsultiert und verknüpft werden müssen. Eine professionell aufbereitete Terminologie und Wissensmodellierung (verteilte Wissensbestände) sind hier entscheidend.

In diesem Zusammenhang entstehen auch neue Programmieransätze, eine Art „Programmieren on-the-fly“ auf der Basis einer hinterlegten Programmlogik. Dieses dynamischere Vorgehen wird dazu führen, dass deutlich mehr Applikationen entwickelt und zur Verfügung stehen werden, mit dem Ergebnis, dass der Bedarf nach semantischen Lösungen sehr schnell und sehr stark wachsen wird.

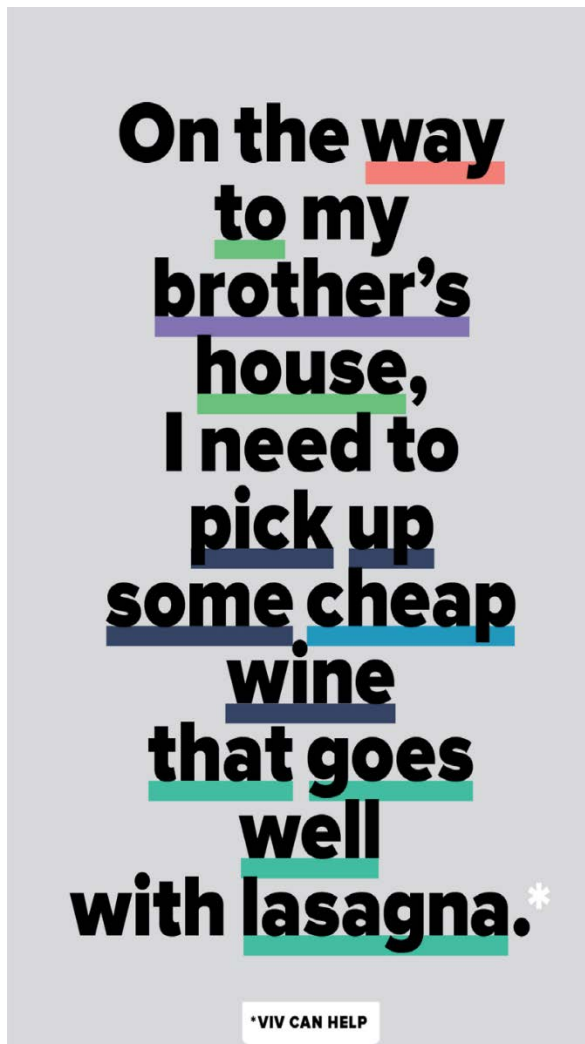
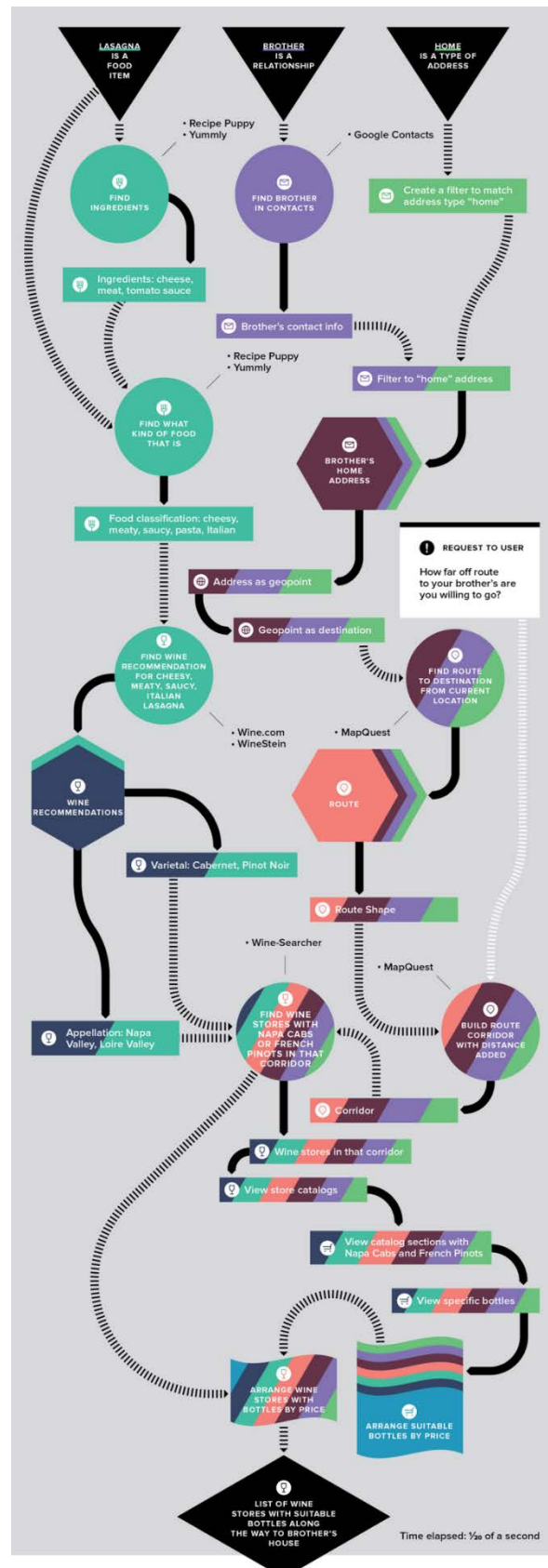


Abbildung 2: Komplexe Anfrage und Bearbeitungswege in Viv
(Quelle: <https://www.wired.com/2014/08/viv/>)



1.5.3 „Smarte“ Maschinelle Übersetzung

Auch im Bereich der Maschinellen Übersetzung (MÜ) sind deutliche Ansätze einer „Smartifizierung“ spürbar.

Zunächst setzte man in der MÜ auf regelbasierte Verfahren, die mit Hilfe von syntaktischen Mustern und Wörterbüchern die Konvertierung von Aussagen aus einer Sprache in die nächste umsetzten. Später und bis heute sind statistische Verfahren hinzugekommen (auch für die Auswertung von Big Data). Seit etwa zwei Jahren verhelfen künstliche neuronale Netze MÜ-Systemen zu besseren Ergebnissen (en: „NMT = Neural Machine Translation“). In November 2016 schaltete Google seine neuen, auf dieser Technologie basierenden MÜ-Engines frei und konnte damit die Ergebnisse bei manchen Sprachkombinationen außerordentlich verbessern.

1.6 Konsequenzen

KI entwickelt sich zunehmend zu einem inter- und pluridisziplinären Forschungsgebiet. Neben Programmierern, Mathematikern und vielen weiteren Spezialisten haben auch Sprachwissenschaftler und Terminologen einen entscheidenden Beitrag zu leisten.

Zentrale Herausforderungen sind hier v. a.

- das „Verstehen“ natürlicher Sprache
- das semantisch korrekte Interpretieren der Aussagen auf Basis von Wissensmodellierungen und -strukturierungen

Im Mittelpunkt steht hier das Zusammenspiel zwischen wissenshaltigen begriffsorientierten Terminologien und Ontologien (vgl. Abschnitt 3.3 und 4.1).

2 Wissensmodellierung mit Hilfe von Ontologien

2.1 Ontologien

Die „Ontologie“ ist ursprünglich eine philosophische Disziplin, die sich mit der Existenz im Allgemeinen sowie v. a. mit der Strukturierung der Realität befasst. Hier spielen die Einteilung der Realität in Entitäten mit spezifischen Eigenschaften sowie ihre Beziehungen zueinander eine wichtige Rolle.

In den letzten Jahrzehnten wird der Begriff mehr und mehr von der Informatik und der KI-Forschung besetzt. Sie verstehen unter Ontologien computerlesbare Wissensmodellierungen und arbeiten mit Klassen und Instanzen, die über Eigenschaften verfügen und miteinander über Relationen verbunden sind. Die in Ontologien genutzten Relationen verfügen ebenfalls über Eigenschaften. Sie sind vielfältig und umfassen sowohl hierarchische als auch nicht-hierarchische Ordnungen, z. T. sogar sehr freie, assoziative Beziehungen.

Inhalte und Aufbau von Ontologien hängen sehr stark vom Einsatzgebiet und vom Einsatzzweck ab. So können bestimmte Informationen für das eine Nutzungsszenario sehr bedeutsam sein, während sie für ein anderes Szenario völlig irrelevant sind.

Abbildung 3 zeigt das Funktionsprinzip einer Ontologie. Im oberen Teil befindet sich die eigentliche Ontologie aus Klassen/Begriffen und Relationen. Die Klassen/Begriffe werden durch Ellipsen dargestellt, die Relationen durch Pfeile. Vererbungen erfolgen v. a. durch hierarchisch aufeinander bezogene Begriffe, hier also z. B. zwischen Künstler (Oberbegriff) und Maler bzw. Bildhauer als Unterbegriffe. Die Vererbung wird durch dicke Pfeile angezeigt. Die Relationen *malt* und *gemaltVon* stehen in einer inversen Beziehung zueinander, wodurch weitere Logik in die Ontologie integriert wird, die es ermöglicht, dass von einem Maler auf seine Kunstwerke und umgekehrt, von einem Bild zum Maler, geschlossen werden kann.

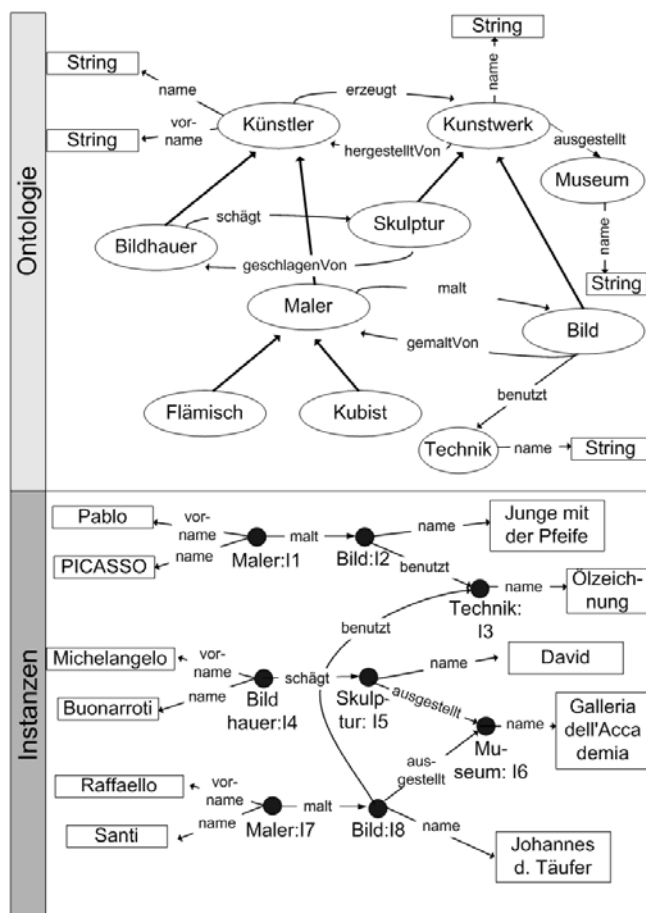


Abbildung 3: Beispiel für Ontologie
 (Quelle: [https://de.wikipedia.org/wiki/Ontologie_\(Informatik\)#/media/File:Ontschichten.gif](https://de.wikipedia.org/wiki/Ontologie_(Informatik)#/media/File:Ontschichten.gif))²

Der untere Teil der Abbildung zeigt Instanzen der Ontologie, jeweils repräsentiert durch schwarze Punkte. Das Kürzel "I1" steht dabei für den eindeutigen Ressourcennamen der Instanz (vgl. [https://de.wikipedia.org/wiki/Ontologie_\(Informatik\)](https://de.wikipedia.org/wiki/Ontologie_(Informatik))).

2.2 Verknüpfte Ontologien und Linked Open Data

Ontologien sind nach Sachgebieten (en: „domains“) organisiert. Ein Beispiel bilden Ontologien für die Produktionsplanung in Unternehmen. Da für bestimmte Aufgaben Wissen aus unterschiedlichen Bereichen benötigt wird, werden Ontologien miteinander verknüpft.

² Von Häger (de:Benutzer:Häger) - Eigenes Werk, Gemeinfrei, <https://commons.wikimedia.org/w/index.php?curid=4910952>

Eine spezielle Form der Ontologienverknüpfung liefert *Linked Open Data*. Linked Open Data (LOD) setzt den Ontologiegedanken in großem Rahmen um und ist damit Teil des Semantic Web. Unter Linked Open Data versteht man frei verfügbare Daten, die zu einem Netzwerk von Datenbeständen aus Begriffen mit eindeutiger Identifikationsnummer (sog. „URI“) verbunden werden (<http://lod-cloud.net/>). Auf dieses Netzwerk (sog. „Linked Open Data Cloud“ oder „Giant Global Graph“) können semantische Applikationen zugreifen³.

Das Basisformat für Ontologiedatenbestände ist das XML-basierte RDF, das „Resource Description Framework“. RDF arbeitet mit Tripeln, d. h. Verknüpfungen von Subjekt-Prädikat-Objekt (oder im Klartext: Ausgangsbegriff-Relation-Objekt dieser Relation).

Eine Weiterentwicklung von RDF ist OWL (Web Ontology Language). Es handelt sich um eine Spezifikation des World Wide Web Consortiums (W3C), um Ontologien anhand einer formalen Beschreibungssprache erstellen, publizieren und verteilen zu können. Zusätzlich zu RDF und RDF-Schema werden weitere Sprachkonstrukte eingeführt, die es erlauben, Ausdrücke ähnlich der Prädikatenlogik so zu formulieren, dass auch Software (z. B. Agenten) die Bedeutung verarbeiten („verstehen“) kann (vgl. https://de.wikipedia.org/wiki/Web_Ontology_Language oder auch <https://www.w3.org/2001/sw/wiki/OWL>)

Einer der bekanntesten OWL-basierten Ontologieeditoren ist Protégé, ein Open-Source-Tool der Stanford University in Kalifornien. Eine bekannte Abfragesprache für RDF und OWL ist SPARQL. Sie ermöglicht intelligente Abfragen („Reasoning“), die das hinterlegte Wissen (z. B. über Eigenschaften oder das Vererben von Eigenschaften zwischen Klassen und Subklassen) auswerten.

Eine besondere Herausforderung bei der Abfrage bzw. der Verknüpfung von Ontologien sind die Struktur und die Benennung der Klassen und Instanzen. Nicht nur die formelle Benennung von Klassen und Instanzen kann unterschiedlich sein, sondern auch die semantischen Inhalte und die Strukturierung der Relationen zwischen Klassen und Instanzen. Ferner berücksichtigt das modellierte Wissen meistens die semantischen Unterschiede nicht, die es zwischen Sprachen, Ländern und Kulturen gibt. Diese Aspekte verhindern einen reibungslosen Austausch von Wissen. Nur selten werden Synonyme, Mehrsprachigkeit und Ambiguitäten professionell genug implementiert, erarbeitet und verwaltet. Daher werden Zusammenhänge zum Teil nicht korrekt interpretiert, und auch das Zusammenführen von verschiedenen Ontologien stößt schnell an Grenzen. Hier können Terminologiewissenschaft und -arbeit verschiedene Lösungen anbieten (siehe Abschnitte 3.3 und 4.1).

³ In diesem Zusammenhang sei auf DBpedia hingewiesen, ein Gemeinschaftsprojekt der Universität Leipzig, der Universität Mannheim, des Hasso-Plattner-Instituts und OpenLink Software, mit dem strukturierte Informationen aus Wikipedia extrahiert und Web-Anwendungen zugänglich gemacht werden können. Dabei werden nicht nur der Fließtext der Wikipedia-Artikel, sondern auch die darin enthaltenen strukturierten Informationen (Infoboxen, Tabellen, Kategorien, geografische Koordinaten, Weblinks usw.) extrahiert und als Datenbasis für fortgeschrittene Fragen verwendet. (Wikipedia, 15.2.2017)

3 Terminologie

3.1 Die Terminologiewissenschaft als Basis der Terminologiearbeit

Das Fundament der praktischen Terminologiearbeit wird durch die Terminologiewissenschaft bereitgestellt, die sich grundlegend mit Begriffen und ihren Bezeichnungen in der Fachkommunikation befasst. Ihr Ziel ist es, terminologische Eigenschaften und Besonderheiten der Fachkommunikation und des Wissenstransfers zu beschreiben und Lösungsansätze für Probleme zu entwickeln. Sie bietet Theorien und Methoden zur Systematisierung und Ordnung von Wissen und zur Verwendung von sprachlichen und nichtsprachlichen Repräsentationen von Begriffen (vgl. RaDT 2017, www.radt.org).

3.2 Aufgabenspektrum in der Terminologiearbeit

Terminologiearbeit besteht darin, Fachwortschätze nach einschlägigen Regeln zusammenzustellen, zugänglich zu machen und bei Bedarf zu vereinheitlichen. Somit stehen zwar die Fachwörter (auch: Termini oder Benennungen) im Mittelpunkt des Interesses, zumal ja auch der Zugang zu den Daten über sie erfolgt, aber es ist keine Terminologiearbeit denkbar, die nicht die Begriffe hinter den Fachwörtern berücksichtigt.⁴

Terminologen befassen sich also

- mit Begriffen als Denk- oder Wissenseinheiten
- mit den Beziehungen zwischen den Begriffen
- mit den Gegenständen, die durch die Begriffe erfasst werden, wobei die Gegenstände materiell oder nichtmateriell, real oder imaginär sein können
- mit den Bezeichnungen und Beschreibungen der Begriffe

Typischerweise arbeitet ein Terminologe in einer Organisation (Unternehmen oder Institutionen), die das Ziel hat, die Fachsprache und Terminologie der eigenen Branche bzw. des eigenen Fachgebiets professionell einzusetzen. Gleichzeitig sollen durch die Terminologiearbeit i. d. R. Kosten gespart und die Verständlichkeit von Dokumenten erhöht werden. Der Terminologe führt also z. B. folgende Arbeitsschritte aus:

Er sammelt die Terminologie des gesamten Fachgebiets sowie seiner Organisation. Diese Sammlung kann einsprachig erfolgen, um beispielsweise die Texte/Dokumente der Organisation in Bezug auf ihre Verständlichkeit und Eindeutigkeit zu erhöhen. Meist jedoch erfolgt die Terminologiearbeit für mehrere Sprachen, da es im Zeitalter der Globalisierung kaum noch Unternehmen und Institutionen gibt, die nur mit einer Sprache arbeiten. Durch mehrsprachige Terminologiearbeit können unnötige Recherchen verhindert und Übersetzungskosten gespart werden.

⁴ Semantiker würden sagen: „Wörter haben Bedeutungen“, während Terminologen diesen Sachverhalt mit der Aussage „Benennungen stehen für Begriffe“ beschreiben würden.

Bei mehrsprachigen Projekten muss geprüft werden, ob die Begriffe in verschiedenen Sprachen übereinstimmen. Darüber hinaus wird innerhalb der Sprachen untersucht und erfasst, ob Synonyme vorliegen, also ob für ein und denselben Begriff verschiedenen Benennungen verwendet werden.

Bei einem Unternehmensprojekt besteht der nächste Schritt in der Regel darin, im Fall von Synonymen nur noch eine Benennung zuzulassen, damit der Sprachgebrauch vereinheitlicht und Übersetzungskosten reduziert werden können.

Ein Eintrag in einer zweisprachigen Terminologiedatenbank könnte so aussehen wie in Abbildung 4 dargestellt.

Ein Eintrag in einer Terminologiedatenbank gibt im Normalfall genau einen Begriff wieder: der Begriff im Beispiel weist allerdings viele Benennungen auf: In Abbildung 4 sind im Englischen 9 Synonyme erfasst, im Deutschen immerhin 5. All diese Benennungen stehen jedoch für denselben Begriff.

Wie zu erkennen ist, ist pro Sprache in dieser Datenbank eine Benennung als „preferred“ markiert, soll also im Unternehmen bevorzugt (oder ausschließlich) verwendet werden. Die anderen Benennungen haben den Status „deprecated“, sollen also nicht (mehr) verwendet werden. Sie bleiben jedoch als Synonyme in der Datenbank erhalten und werden kontinuierlich ergänzt, da sie für Suchen und Vergleiche zur Verfügung stehen müssen.

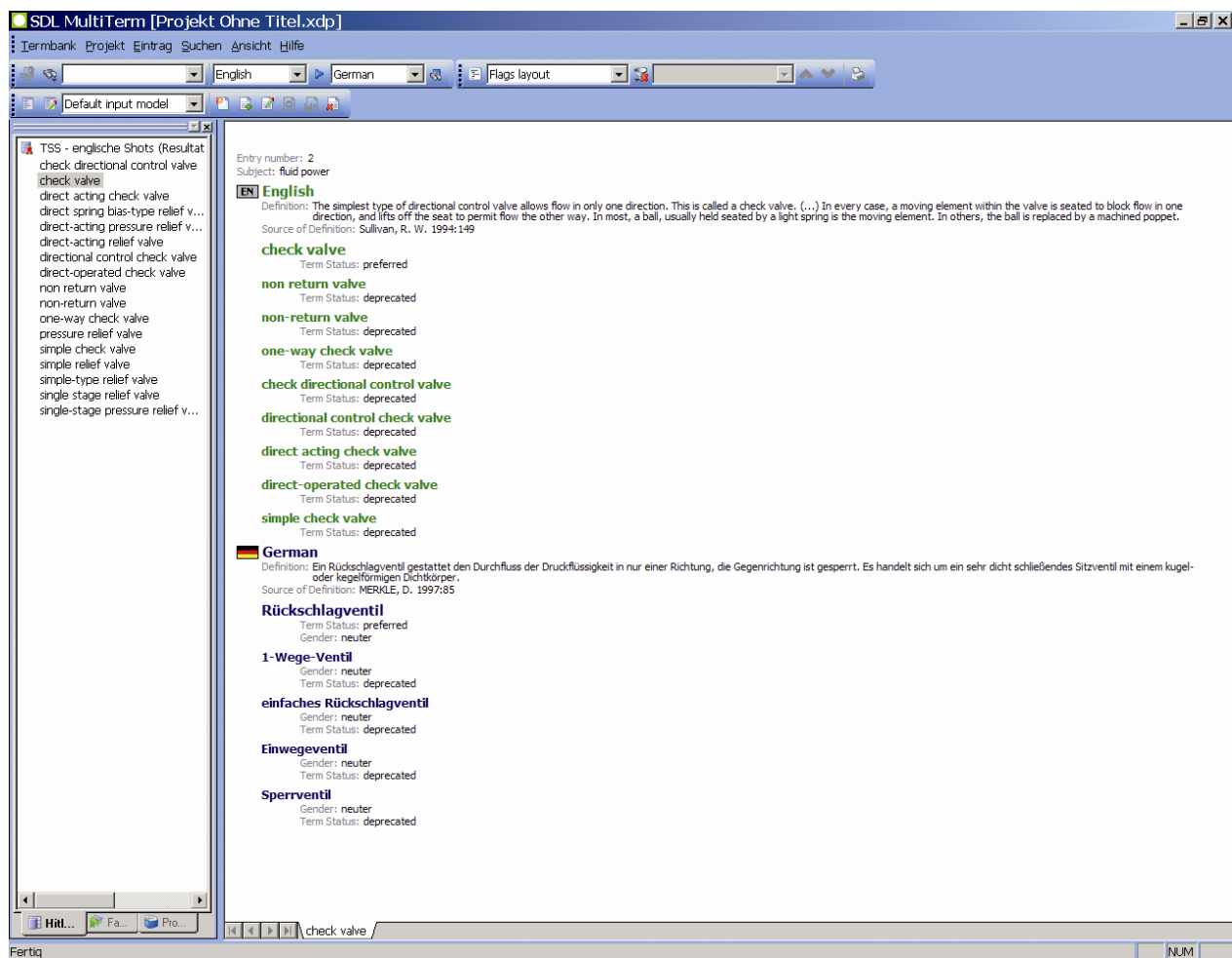


Abbildung 4: Beispiel eines zweisprachigen Eintrags in einer Terminologiedatenbank

Im dargestellten Beispiel ist es so, dass es auf begrifflicher Ebene im Deutschen und im Englischen keine Unterschiede gibt. Unter „Rückschlagventil“ stellt sich ein deutscher Muttersprachler dasselbe vor wie ein englischer Muttersprachler unter „check valve“. Diese Übereinstimmung ist auch an den Definitionen sichtbar. Doch dieser „einfache Fall“ der völligen Übereinstimmung („Volläquivalenz“) zwischen Sprachen ist nicht immer gegeben (zur näheren Erläuterung siehe Abschnitt 3.3).

Je mehr Sprachen ein Terminologieeintrag umfasst, desto schwieriger wird die Entscheidung, ob die erfassten bzw. vorgeschlagenen Benennungen tatsächlich immer denselben Begriff repräsentieren. Die Abbildungen 5 und 6 zeigen einen Fall von „Teiläquivalenz“, bei dem der Begriff in einer Sprache (Deutsch) weiter gefasst ist als in den zwei anderen Sprachen (Englisch und Französisch). Die Definition spielt in diesem Fall eine besonders wichtige Rolle, weil sie die begrifflichen Unterschiede (hier: die Größe der Fläche) versprachlicht. Als Mehrwert für den Leser unterstreicht zusätzlich die Anmerkung diesen Sachverhalt („a wood is of smaller extent“ bzw. „un bois couvre une surface de moindre étendue“).

Datensatz 10			
Deutsch			
<u>Definition</u>	Eine mit Forstpflanzen bestockte Grundfläche.		
<u>Quelle der Definition</u>	nach Bundeswaldgesetz §2		
Wald			
<u>Quelle der Benennung</u>	nach Bundeswaldgesetz §2		
Erstellungsdatum	Erstellt von	Änderungsdatum	
17.03.2017	admin	17.03.2017	
Französisch			
<u>Definition</u>	Vaste étendue de terrain peuplée d'arbres.		
<u>Quelle der Definition</u>	BT, Glossaire des génériques en usage dans les noms géographiques du Canada, 1987		
<u>Anmerkung</u>	D'une façon générale un bois couvre une surface de moindre étendue qu'une forêt.		
<u>Quelle der Anmerkung</u>	Métro, Terminologie forestière, 1975		
forêt			
<u>Quelle der Benennung</u>	BT, Glossaire des génériques en usage dans les noms géographiques du Canada, 1987		
Erstellungsdatum	Erstellt von	Änderungsdatum	
17.03.2017	admin	17.03.2017	
Englisch			
<u>Definition</u>	A community of trees and associated organisms covering a considerable area.		
<u>Quelle der Definition</u>	Sharpe et al., Introduction to forestry, 1976		
forest			
<u>Quelle der Benennung</u>	Sharpe et al., Introduction to forestry, 1976		
Erstellungsdatum	Erstellt von	Änderungsdatum	
17.03.2017	admin	17.03.2017	

Datensatz 11			
Deutsch			
<u>Definition</u>	Eine mit Forstpflanzen bestockte Grundfläche.		
<u>Quelle der Definition</u>	nach Bundeswaldgesetz §2		
Wald			
<u>Quelle der Benennung</u>	nach Bundeswaldgesetz §2		
Erstellungsdatum	Erstellt von	Änderungsdatum	
17.03.2017	admin	17.03.2017	
Französisch			
<u>Definition</u>	Ensemble d'arbres poussant plus ou moins près les uns des autres et couvrant un certain espace.		
<u>Quelle der Definition</u>	Métro, Terminologie forestière, 1975		
<u>Anmerkung</u>	D'une façon générale un bois couvre une surface de moindre étendue qu'une forêt.		
<u>Quelle der Anmerkung</u>	Métro, Terminologie forestière, 1975		
bois			
<u>Quelle der Benennung</u>	Métro, Terminologie forestière, 1975		
Erstellungsdatum	Erstellt von	Änderungsdatum	
17.03.2017	admin	17.03.2017	
Englisch			
<u>Definition</u>	A community of trees growing more or less closely together.		
<u>Quelle der Definition</u>	Ford-Robertson, Terminology of forest science, technology, practice, and products, 1971		
<u>Anmerkung</u>	A wood is of smaller extent than a forest and generally larger than a coppice.		
<u>Quelle der Anmerkung</u>	Ford-Robertson, Terminology of forest science, technology, practice, and products, 1971		
wood			
<u>Quelle der Benennung</u>	Ford-Robertson, Terminology of forest science, technology, practice, and products, 1971		
Erstellungsdatum	Erstellt von	Änderungsdatum	
17.03.2017	admin	17.03.2017	

Abbildungen 5 und 6: Beispiele für dreisprachige Einträge in einer Terminologiedatenbank

Ähnlich verhält es sich beispielsweise mit „télécharger“. Auf Französisch umfasst der Begriff sowohl das Herunter- als auch das Hochladen von Daten. Die Benennungen „runway“ (EN) bzw. „piste“ (FR) geben das wieder, was im Deutschen mit 2 Benennungen (Landebahn und Startbahn) ausgedrückt wird. Auch bei Fahrzeugklassen kommt es zu Abweichungen, mit denen viele Anwender von Terminologie nicht rechnen, so entspricht z. B. „subcompact car“ sowohl dem „Kleinstwagen“ als auch dem „Kleinwagen“ (DE).

Durch das begriffsorientierte Arbeiten, das an den Beispieleinträgen deutlich wird (1 Begriff = 1 Eintrag in der Datenbank), werden auch Mehrdeutigkeiten problemlos aufgelöst und sinnvoll abgelegt, denn Ambiguitäten werden in verschiedenen Einträgen verwaltet. In all diesen Fällen wird für jeden Begriffsinhalt ein eigener Eintrag angelegt, wobei die Informationen in einer Sprache wiederholt werden müssen. Beispiel: Maus (sowohl in der Zoologie als auch in der IT).

Da die Benennung „Maus“ unterschiedliche Begrifflichkeiten abbildet, müssen mehrere Datensätze angelegt werden, die jeweils einen Begriff umfassen. Die ambige Benennung kommt dann in beiden Einträgen vor. So kann man auch unterschiedliche Synonyme in den Einzelsprachen und unterschiedliche Äquivalente in verschiedenen Sprachen sauber zuordnen.

Dies ist ein zentraler Unterschied zwischen einer Terminologiedatenbank und einem Wörterbuch, denn letzteres würde nur einen Eintrag zum Schlagwort „Maus“ verzeichnen und innerhalb dieses Eintragsblocks beide Begriffe behandeln.

3.3 Erstellen von Begriffssystemen

3.3.1 Grundlagen

Begriffe werden in der professionellen Terminologearbeit nicht isoliert betrachtet, sondern zueinander in Beziehung gesetzt. Die Erarbeitung eines Terminologiebestands beginnt also häufig mit dem Aufbau eines (kompletten) Systems von Begriffsbeziehungen für ein bestimmtes (Teil-)Fachgebiet. Diese Begriffssysteme geben das Wissen eines Fachgebiets in strukturierter Form wieder.

Die Beziehungen zwischen den Begriffen können hierarchischer oder nicht-hierarchischer Natur sein, wobei die hierarchischen Begriffsbeziehungen für die Terminologearbeit bedeutsamer sind als die nicht-hierarchischen. Bei den hierarchischen Begriffsbeziehungen unterscheidet man Bestandsbeziehungen und Abstraktionsbeziehungen. Bei beiden hierarchischen Beziehungsarten werden Über- und Unterordnungen zwischen Begriffen her- bzw. dargestellt.

In nicht-hierarchischen Systemen werden z. B. sequentielle Beziehungen (wie chronologische Abfolgen oder Ursache-Wirkung-Beziehungen) oder Gegensatzbeziehungen dargestellt. Nur durch eine Kombination der verschiedenen Beziehungsarten können Sachgebiete vollständig und sinnvoll strukturiert werden.

Die folgenden Abbildungen zeigen zwei beispielhafte Visualisierungen. Während Abbildung 7 und Abbildung 9 auf Abstraktionsbeziehungen basieren, also Unterbegriffe zeigen, die durch Spezifizierung

entstehen, zeigt Abbildung 8 ein System, das auf Bestandsbeziehungen (Teil-Ganzes-Beziehungen) beruht.

Abstraktionsbeziehungen zeichnen sich dadurch aus, dass ein Unterbegriff alle Merkmale seines Oberbegriffs enthält, gleichzeitig aber über mindestens ein weiteres Merkmal verfügt. Grafisch wird diese Beziehungsart üblicherweise in Winkeldiagrammen dargestellt.

Bestandsbeziehungen sind dadurch charakterisiert, dass sich der Verbandsbegriff, d. h. der übergeordnete Begriff, in seine einzelnen Teilbegriffe zerlegen lässt. Grafisch wird diese Beziehungsart üblicherweise in Klammerdiagrammen dargestellt.

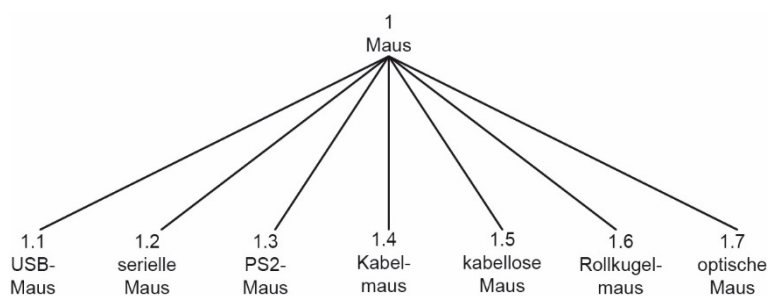


Abbildung 7: Beispiel für ein Begriffssystem (Abstraktionssystem mit polydimensionaler Reihe)
 Quelle: Drewer/Schmitz 2017:11

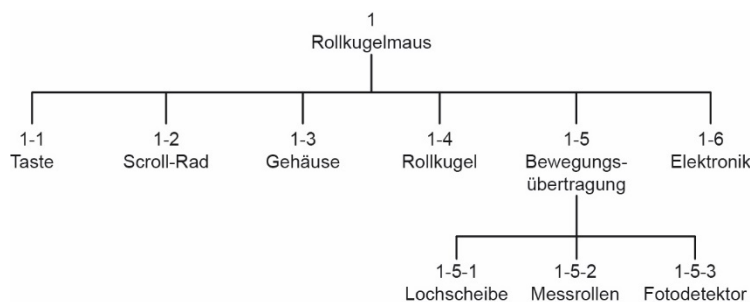


Abbildung 8: Beispiel für ein Begriffssystem (Bestandsbeziehungen)
 Quelle: Drewer/Schmitz 2017:10

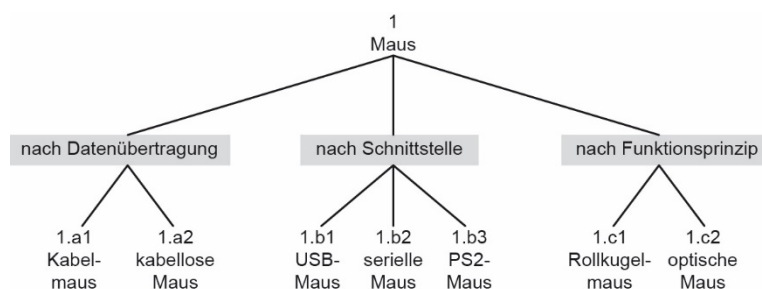


Abbildung 9: Beispiel für ein Begriffssystem (Abstraktionssystem mit monodimensionalen Reihen)
 Quelle: Drewer/Schmitz 2017:12

Die Reihen in Abstraktionssystemen entstehen auf unterschiedliche Art und Weise:

- **Monodimensionale Reihen** entstehen, wenn pro Unterteilungsschritt nur ein Kriterium verwendet wird, z. B. Einteilung des Oberbegriffs Maus nur nach der Art der Datenübertragung in die Unterbegriffe Kabelmaus und kabellose Maus.
- **Polydimensionale Reihen** entstehen, wenn in einem Unterteilungsschritt mehrere Kriterien verwendet werden (im Beispiel: „nach Schnittstelle“, „nach Datenübertragung“ und „nach Funktionsprinzip“). Werden die verschiedenen Einteilungskriterien nicht genannt, sondern alle Unterbegriffe gleichwertig aufgelistet, so ergibt sich eine ungeordnete und unübersichtliche Reihe (siehe Abbildung 7).
Diese unübersichtliche polydimensionale Reihe kann jedoch in mehrere monodimensionale Reihen unterteilt werden, indem die Einteilungskriterien explizit genannt werden (siehe Abbildung 9:
Beispiel für ein Begriffssystem (Abstraktionssystem mit monodimensionalen Reihen)
Quelle: Drewer/Schmitz 2017:12
- Dies erhöht die Lesbarkeit und Verständlichkeit des Abstraktionssystems und macht sichtbar, welche Überlegungen der Einteilung zugrunde liegen.

(vgl. Drewer/Schmitz 2017:11)

Eine weitere Unterteilungsmöglichkeit richtet sich nach der im System realisierten Hierarchie. Wenn jeder Begriff im System nur einen einzigen Oberbegriff hat, so spricht man von **Monohierarchie**.

Polyhierarchie hingegen liegt vor, wenn ein Begriff mehrere Oberbegriffe hat, er also durch Unterteilung „auf verschiedenen Wegen“ erreicht werden kann. Durch überkreuzte Linien und die komplexe Darstellung sind polyhierarchische Systeme oft unübersichtlich und schwer nachvollziehbar. Allerdings geben sie die fachlichen Begrifflichkeiten und Zusammenhänge oft treffender und angemessener wieder als monohierarchische Systeme. (vgl. Drewer/Schmitz 2017:12, vgl. DIN 2331 1980:13f).

Besonders wichtig für die Terminologearbeit ist die Verknüpfung der Begriffssysteme mit den terminologischen Kerndaten in der Terminologiedatenbank. Jeder Knoten in einem Begriffssystem (hier also jedes Kästchen, das einen Begriff repräsentiert) steht für eine Vielzahl an sprachlichen Informationen. Man könnte sich also vorstellen, dass hinter jedem Knoten ein terminologischer Eintrag steht, wie er in Abbildung 4 dargestellt wurde. Der Zugang zur begrifflichen Struktur könnte also über jedes beliebige Synonym und in jeder beliebigen Sprache erfolgen.

3.3.2 Sprachlich-kulturelle Unterschiede

Der Mensch sieht die Realität und versteht Begriffe durch verschiedene Filter: seine persönliche Erfahrung, seine Ausbildung, seine Kultur und seine Sprache. Das betrifft ganz allgemeine Begriffe aus dem Alltagsleben (z. B. Definition von *Frühstück*) aber auch Fachbegriffe aus Technik, Medizin oder Wirtschaft, um einige Beispiele zu nennen.

So bildet die deutsche Sprache im Gegensatz zur französischen Sprache keinen allgemeinen Begriff für „Herausbewegen“ („sortie“ auf Französisch). Dafür bildet die englische Sprache einen Oberbegriff für das Herausbewegen aus einem Ort („exit“), während die deutsche Sprache mehrere Unterbegriffe bildet, je

nachdem, mit welchem Fortbewegungsmittel diese Bewegung stattfindet: „Ausgang“ oder „Ausfahrt“ (vgl. Abbildung 10).

In der Übersetzerpraxis führen diese sprachabhängigen begrifflichen Unterschiede oft zu Problemen, weil der Kontext nicht immer verfügbar ist bzw. nicht ausreicht, um die nötigen Entscheidungen zu treffen, etwa im unteren Beispiel bei einer Übersetzung, die im Ausgangstext ein englisches Wort wie „exit“ enthält, bei der der Übersetzer sich im Deutschen entscheiden muss.

Die Aufgabe der Terminologen ist es also, sprachneutrale Begriffe zu definieren, um begriffliche Systeme zwischen Sprachen und Kulturen zu überbrücken. Bei fehlenden Benennungen für einen Begriff in einer bestimmten Sprache wird oft die Benennung eines Oberbegriffs bzw. eines Unterbegriffs vererbt.

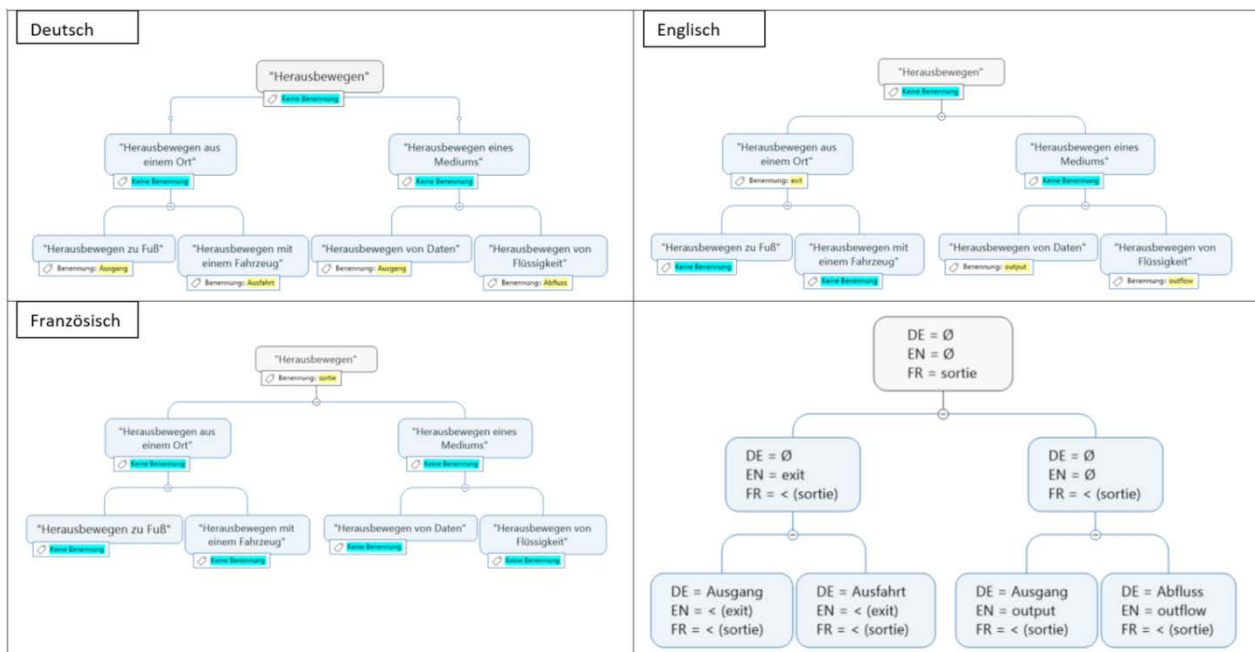


Abbildung 10: Begriffliche Unterschiede im Sprachvergleich

3.3.3 Begriffssysteme in der Praxis

In Unternehmen und Institutionen wird leider häufig auf die systematische Erarbeitung von Begriffssystemen verzichtet. Die Gründe: Zum einen handelt es sich um eine aufwendige und anspruchsvolle Aufgabe, die einigen zu kostspielig erscheint, zum anderen fehlen oft die entsprechenden Tools für die Aufbereitung und grafische Visualisierung. Vereinzelt ist es möglich, eine grafische Darstellung zu erarbeiten, die aber häufig nicht aus der Terminologiedatenbank dynamisch generiert wird, sondern statisch zu erstellen ist.

Hier ist jedoch inzwischen Bewegung auf dem Markt der Terminologieverwaltungssysteme entstanden. Das Arbeiten mit Begriffssystemen in direkter Interaktion mit den terminologischen Einträgen wird einfacher und komfortabler (siehe dazu Abschnitt 4.1.3).

4 Wechselwirkungen und Austausch zwischen innovativen Technologien und Terminologie

4.1 Terminologie und Ontologien

4.1.1 Ausgangssituation

Wie wird die Welt in einigen Jahren aussehen? Bis zum Jahr 2020 werden je nach Schätzung 20 bis 50 Mrd. Geräte miteinander vernetzt sein und kommunizieren. Ein Teil dieser Kommunikation wird in natürlicher Sprache erfolgen und unstrukturierte Daten verstehen müssen (im Gegensatz zu vordefinierten Szenarios und mit Metadaten angereicherten Inhalten).

Für diese Entwicklung sind semantische Verfahren erforderlich, die sehr schnell und kontextbezogen auf individuelle Situationen/Informationen reagieren und Aussagen formulieren können. Der genaue Umfang ist noch nicht abzusehen, aber der Bedarf an derartigen Lösungen wird in den kommenden Jahren sehr stark ansteigen. Um ihn zu decken, müssen in einem bisher unbekanntem Umfang mehrsprachige Ontologien bzw. ontologieartige Datenbestände zur Verfügung stehen und erreichbar sein. Das wird mit den bisherigen Mitteln (auch an Humanressourcen, d. h. an Wissensfachleuten) nicht zu bewältigen sein. Es bedarf neuer Verfahren, um ad hoc relativ schnell ontologieartig verwertbare Wissensbestände aufzubauen. Das kann Wissen für bestimmte Communities betreffen oder Fachwissen eines Unternehmens, einer Institution oder einer Branche.

In der Terminologiewissenschaft werden schon seit Jahrzehnten Begriffssysteme erarbeitet und visualisiert, die den Ontologien in vielen Aspekten sehr ähnlich sind (siehe Abschnitt 3.3). In Begriffssystemen wird strukturiertes Wissen dargestellt, und zwar in Form von Begriffen (Klassen) und ihren Beziehungen zueinander. Der Hauptunterschied zu den Ontologien besteht darin, dass Begriffssysteme dazu gedacht sind, die Begriffe und ihre Beziehungen für Menschen nachvollziehbar und verstehbar zu machen und sie ggf. in übersichtlicher Form graphisch darzustellen. Ontologien hingegen werden von Maschinen „gelesen“ und müssen daher einige andere Anforderungen erfüllen. Die Inhalte beider Wissensstrukturierungen (Begriffssysteme und Ontologien) sind jedoch außerordentlich ähnlich. Die Terminologiewissenschaft verfügt über ein ausgereiftes Methodenrepertoire zur Erarbeitung und Darstellung von Begrifflichkeiten und Relationen, v. a. unter Berücksichtigung sprachlicher Aspekte wie Synonymie, Ambiguität und Mehrsprachigkeit.

Terminologiewissenschaft und -arbeit werden davon profitieren, wenn sie sich für den Einsatz von Semantiken in Form von Ontologien und Frames (Frame Semantics nach Charles Fillmore) öffnen. Gleichzeitig bieten sich vielversprechende Anknüpfungspunkte für die Ersteller von Ontologien, wenn sie die Erkenntnisse der Terminologiewissenschaft nutzen und Ergebnisse der Terminologiewissenschaft in ihre Modelle sowie in konkrete Ontologien einarbeiten. Momentan sind insbesondere zwei Schwächen feststellbar, die die Erarbeitung von Ontologien behindern bzw. ihre Qualität einschränken:

1. Die Ontologien sind sprachlich meist nicht gut aufbereitet. Dies betrifft v. a. die Mehrsprachigkeit und auch die Erfassung und den Umgang mit Synonymen und

Quasisynonymen.

2. Die Ontologien werden von einer kleinen Zahl hochspezialisierter Fachleuten entwickelt, die den steigenden Bedarf nicht ansatzweise decken können.

Die folgenden Abschnitte greifen diese Problemfelder auf und erläutern, inwiefern terminologische Arbeitsmethoden und -ergebnisse helfen können.

4.1.2 Bessere Qualität von Ontologien auf sprachlicher Ebene

Eine der Schwächen von Ontologien ist, dass sie ihre Inhalte im sprachlichen Bereich oft schlecht aufbereiten. Es ist zwar technisch möglich, Synonyme und Fremdsprachen über spezielle Relationen wie *SameAs* einzubeziehen, aber in der Praxis ist dies selten der Fall. So entstehen unter anderem Schwierigkeiten bei der korrekten Interpretation von Fragen, die in natürlicher Sprache formuliert sind, denn der Zugang muss ja weiterhin über sprachliche Bezeichnungen erfolgen, die die Klassen repräsentieren.

Zudem ist es bislang nur schwer möglich, bestehende Ontologien miteinander zu verknüpfen, da sie verschiedene Bezeichnungen für dieselben Klassen verwenden (Synonymie) und zudem strukturelle Unterschiede aufweisen, v. a. in der Granularität der Strukturierung. Noch komplexer werden die Probleme, wenn man verschiedensprachige Ontologien verbinden möchte.

Die Terminologiearbeit kann wesentlich dazu beitragen, mehr Ontologien in besser verwertbarer Qualität aufzubauen.

- Mehrsprachigkeit, v. a. Umgang mit Teiläquivalenzen
- Professioneller Umgang mit Synonymen und Teilsynonymen (in allen beteiligten Sprachen!!!)

Wie die Terminologiearbeit bisher mit diesen Phänomenen umgeht, wurde in den Abschnitten 3.2 und 3.3 angedeutet. Die dort abgebildeten Möglichkeiten zeigen jedoch nur beispielhafte (relativ schlanke) Lösungen.

4.1.3 Erarbeitung von ontologischen Datenbeständen

Wie oben erläutert, kann die Terminologiewissenschaft geeignete Methoden und Vorgehensweisen zur Verfügung stellen, um qualitativ hochwertige und damit besser/breiter einsetzbare Ontologien zu erstellen. Zugleich können auch konkrete Arbeitsergebnisse aus dem terminologischen Bereich in Ontologien eingebunden werden.

Über die Terminologiearbeit kann die Wissensarbeit teilweise „demokratisiert“ werden, indem deutlich mehr Personen auf Grundlage bereits hinterlegter Relationsdefinitionen daran arbeiten können. Das ist z. B. entscheidend, wenn es darum geht, schnell neue Wissensgebiete zu modellieren. Man denke zum Beispiel an die vielen Spezialisten, die es in verschiedenen Unternehmen gibt: Redakteure, Entwickler, Supportmitarbeiter usw. Und zum anderen kann sie zur Interoperabilität von Wissen auch über die Sprachgrenzen hinweg beigetragen, indem mehrsprachige begriffsorientierte Modelle die oft irreführenden Klassenbezeichnungen ergänzen.

Dass der Terminologiebereich bereits in dieser Richtung arbeitet, erkennt man u. a. daran, dass es heute am Markt für Terminologieverwaltungssysteme schon einige Tools gibt, die Relationen zwischen Begriffen modellieren.

- Coreon (Coreon GmbH)
- i-Term (DANTERM Technologies)
- LookUp (D.O.G. Dokumentation ohne Grenzen GmbH)
- Quickterm (Kaleidoscope Communications Solutions GmbH)
- Termweb (Interverbum Technology)

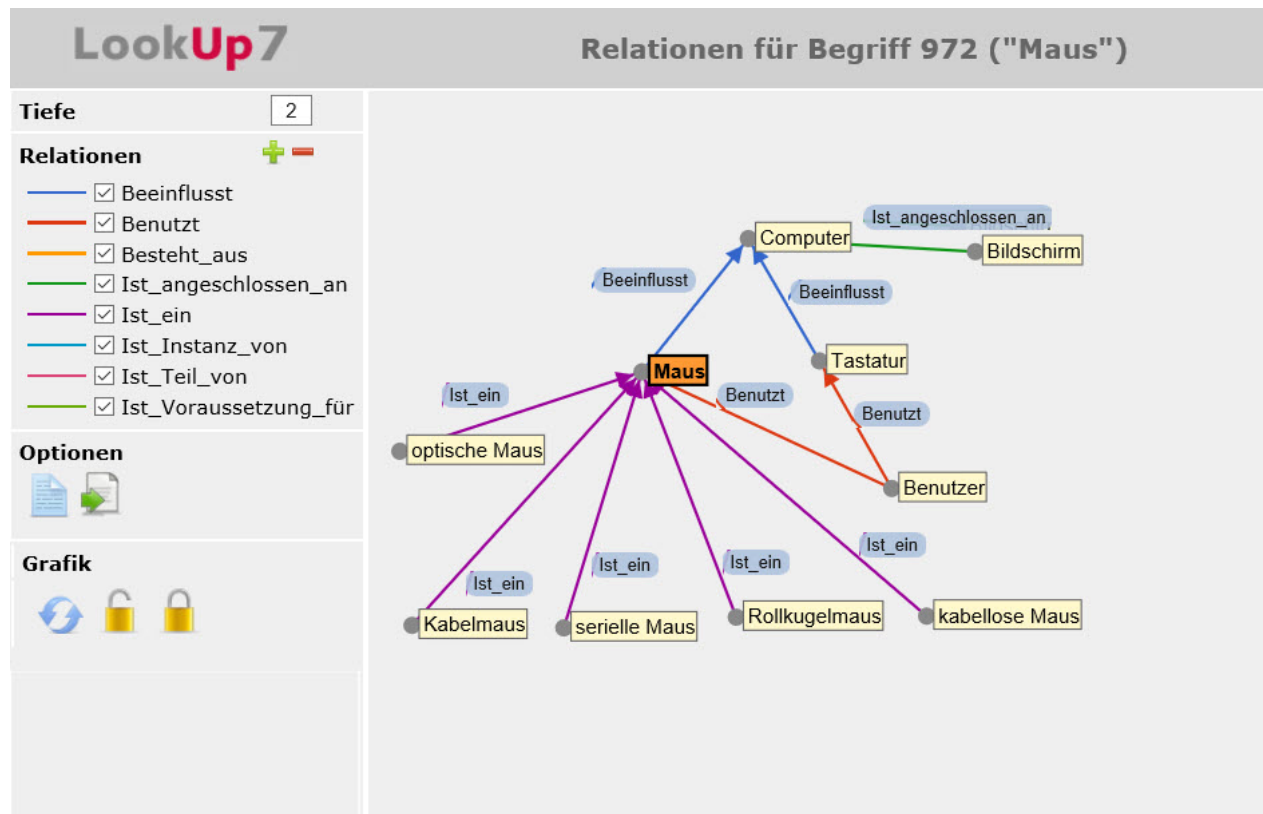


Abbildung 11: Beispiel für mit LookUp visualisiertes Relationssystem

4.2 Annäherungen von Terminologiedatenbanken und Wissensdatenbanken

Wie sieht ein typischer Eintrag in einer Terminologiedatenbank aus? Welche Datenkategorien (auch: Datenfelder) enthält er? Wie sind die Datenkategorien angeordnet und aufeinander bezogen?

Diese Fragen sind kaum zu beantworten, denn

„[f]ür die Auswahl von terminologischen Datenkategorien bei der Implementierung eines Terminologieverwaltungssystems ist entscheidend, welche Zielsetzung das Terminologieprojekt verfolgt, welche Nutzergruppen anvisiert werden und in welchem organisatorischen Umfeld das System eingesetzt werden soll. Davon hängt ab, welche terminologischen und administrativen Informationen in der Datenbank verwaltet werden und welche Datenkategorien man dafür braucht.“ (Drewer/Schmitz 2017:103)

Verschiedene Ersteller, verschiedene Nutzergruppen, verschiedene Anforderungen an die Terminologiedatenbank beeinflussen also deren Aufbau und Umfang, so dass sehr unterschiedliche, z.B: sehr individuelle Lösungen entstehen. Dennoch sollen an dieser Stelle einige Datenkategorien genannt und kurz erläutert werden, die in der Regel in Terminologiedatenbanken enthalten sind:

- **Benennung:** Selbstverständlich enthält jede Terminologiedatenbank ein Datenfeld zur Aufnahme von Benennungen, die den jeweiligen Begriff repräsentieren.
- **Definition:** Da Definitionen Begriffe beschreiben und von anderen Begriffen abgrenzen, ist diese Datenkategorie eine der wichtigsten für das Terminologiemanagement.
- **Abbildung:** Statische oder bewegte Bilder können Begriffe erläutern und somit ganz entscheidend zum Verständnis beitragen. Sie werden oft als Ergänzung, manchmal sogar als Ersatz von Definitionen in Terminologiedatenbanken aufgenommen.
- **Fachgebiet:** Diese Datenkategorie enthält Informationen darüber, zu welchem Fachgebiet ein Begriff gehört. Mit Hilfe dieser Angabe lassen sich Begriffe besser zuordnen und auch bei der Suche bzw. beim Filtern kann das Fachgebiet eine sehr gute Unterstützung sein.

In vielen Datenbanken findet man zudem Datenkategorien, die nähere Informationen zu einzelnen Benennungen enthalten, z. B.

- ob die Verwendung einer Benennung auf bestimmte Länder oder Regionen beschränkt ist (Datenkategorie „Region“, „Regionale Verwendung“ oder „Geografischer Gebrauch“)
- ob die Verwendung einer bestimmten Benennung verboten oder vorgeschrieben ist (Datenkategorie „Status“ oder „Gültigkeit“), eine Information, die in allen präskriptiven und normativen Projekten erforderlich ist
- ob eine Benennung einer bestimmten Stilebene angehört (Datenkategorie „Stil“, „Sprachregister“)
- ob die Verwendung einer Benennung auf bestimmte Projekte, Produkte, Abteilungen oder Kunden eingeschränkt ist

Zusätzlich gibt es häufig eine Datenkategorie „Anmerkung“, in der Kommentare, weiterführende Hinweise oder sonstige, für die Benutzer potentiell interessante Informationen abgelegt werden.

In der Datenkategorie „Kontext“ (oder manchmal „Phraseologie“) werden anhand von Beispielsätzen spezifische sprachliche Verwendungsweisen aufgezeigt.

Die hinterlegten Informationen sollten mit einer Quellenangabe versehen sein, um die Informationen

nachvollziehbar zu machen und um weitere Recherchen durchführen zu können, so dass auch die Datenkategorie „Quelle“ zum Standardumfang einer Terminologiedatenbank gehört.

Somit hat jede Organisation ihre eigene Terminologieeintragsstruktur und auch eigene Regeln für die Handhabung und die „Befüllung“ der Datenkategorien. Dadurch wird es rasch komplex und kompliziert, wenn es gilt, (Teil-)Bestände von einer Terminologiedatenbank in eine andere zu transferieren, Terminologiebestände auszutauschen oder für andere Systeme nutzbar zu machen. In den letzten Jahren haben die Bemühungen, ein standardisiertes Austauschformat zu erarbeiten, gefruchtet: TBX (ISO 30042) basiert auf XML und ist mit den meisten Terminologieverwaltungssystemen kompatibel.

Wissensdatenbanken, die im KI-Umfeld eingesetzt werden, sehen mit Sicherheit in vielen Aspekten anders aus, enthalten andere Daten und andere Verknüpfungen. Dennoch zeigen sich viele Verbindungen.

Die nächsten Schritte in der Weiterentwicklung (sowohl von Terminologie- als auch von Wissensdatenbanken) müssen im Rahmen einer pluri- und interdisziplinären Zusammenarbeit gemacht werden. Weder Terminologen noch Wissensmanager/Ontologen können allein das gesamte Spektrum der Möglichkeiten überblicken. Nur durch enge Zusammenarbeit und Austausch können neue Tendenzen erkannt und Synergien genutzt werden.

4.3 Veränderungen im Berufsprofil von Terminologen

Wie oben erläutert, ergeben sich für Terminologen völlig neue Aufgabengebiete. Auf der Basis ihres terminologischen Know-hows sind sie in der Lage, Daten zu erarbeiten, die in Ontologien und in der Folge von Smart Assistants und weiteren Tools genutzt werden können.

Innerhalb der Terminologiewissenschaft ist schon lange klar, dass Terminologen Wissensmanager sind. Nicht ohne Grund werden diese Berufsbezeichnungen oft synonym verwendet. Auch Bezeichnungen wie „Wissensmodellierer“ und „Informationsstrukturierer“ sind durchaus treffend bzw. zutreffend, um die Tätigkeiten von Terminologen zu beschreiben.

Terminologen leisten seit Jahren und Jahrzehnten vieles von dem, was nun im Rahmen der Ontologierstellung neues Gewicht und Ansehen bekommt. Sie ermitteln Begriffe, analysieren ihre Relevanz für bestimmte Kontexte und Anwendungsgebiete, definieren ihre Merkmale und setzen sie in Beziehung zueinander, indem sie Begriffssysteme erstellen. Diese Tätigkeiten sind nun gefragt, um Ontologien in verschiedensten Bereichen aufzubauen.

Gleichzeitig ergeben sich für Terminologen durch den Aufschwung der Ontologien neue Arbeitsmöglichkeiten in ihrem terminologischen Umfeld. Das bisher oft vernachlässigte Erstellen von Begriffssystemen, die wir oben bereits teilweise mit Ontologien gleichsetzen konnten, dürfte durch neue technische Möglichkeiten einfacher werden und neue Anwendungsgebiete erschließen.

5 Ausblick

Aus der Vogelperspektive betrachtet, erleben wir in bestimmten Abständen größere Revolutionen. Viele haben die Terminologiearbeit und die Terminologiewissenschaft beeinflusst. Zuerst haben das Entstehen moderner Wissenschaften und die industrielle Revolution die Grundlage für das Wirken von Wissenschaftlern wie Eugen Wüster geliefert. Der nächste große Schub für die Terminologie kam mit der Professionalisierung der Dokumentationserstellung und der Übersetzungsarbeit mithilfe von Content-Management- und Translation-Memory-Systemen. Dahinter steht der wachsende Bedarf an Informationen, die in immer kürzeren Abständen mehrsprachig zur Verfügung stehen müssen.

Nun stehen wir an der Schwelle zum nächsten Quantensprung und müssen aus der Perspektive der Terminologie den Trend richtig einschätzen, Lösungen anbieten und ggf. unser eigenes Tun anpassen.

Mit anderen Worten:

- Was können Terminologiewissenschaft und -arbeit liefern, um die neuen Entwicklungen im Umfeld von Big Data, Ontologien und Künstlicher Intelligenz zu unterstützen?
- Welche Rolle spielen anders herum Big Data, Ontologien und Künstliche Intelligenz für die Terminologiewissenschaft und -arbeit? Inwiefern beeinflussen sie terminologische Arbeiten und Methoden?

Die oben beschriebenen Entwicklungen beeinflussen sich gegenseitig sehr stark. Zur Entwicklung von Lösungen tragen Spezialisten aus verschiedenen Fachrichtungen bei: Mathematiker, Informatiker, Informationswissenschaftler, Computerlinguisten, unterschiedliche Fachleute (Medizin, Finanzen usw.), Neurologen, Kognitionspsychologen und viele mehr.

Die Terminologiewissenschaft ist gefragt, wenn es um die Verbindung von Wissen und natürlicher Sprache geht. Hier liegen ihre Spezialisierung und ihre Erfahrung. Während jedoch bisher die Aufbereitung von Wissen für menschliche Nutzer im Mittelpunkt stand, verschiebt sich der Fokus nun auf die Aufbereitung von Wissen für Maschinen. Es geht um mehrsprachige Kommunikation zwischen Menschen, zwischen Mensch und Maschine oder zwischen Maschinen, d. h. zwischen Softwareprodukten oder Softwareagenten.

Literaturverzeichnis

Drewer, Petra / Schmitz, Klaus-Dirk (2017): Terminologiemanagement : Grundlagen - Methoden -
Werkzeuge. Heidelberg: Springer Vieweg (Kommunikation und Medienmanagement 1)

Abbildungsverzeichnis

Alle Abbildungen können in hoher Auflösung online betrachtet werden.

- Abbildung 1: [Mensch-Maschine-Kommunikation](#)
- Abbildung 2a: [Komplexe Anfrage und Bearbeitungswege in Viv](#)
- Abbildung 2b: [Komplexe Anfrage und Bearbeitungswege in Viv](#)
- Abbildung 3: [Beispiel für Ontologie](#)
- Abbildung 4: [Beispiel eines zweisprachigen Eintrags in einer Terminologiedatenbank](#)
- Abbildung 5: [Beispiele für dreisprachige Einträge in einer Terminologiedatenbank](#)
- Abbildung 6: [Beispiele für dreisprachige Einträge in einer Terminologiedatenbank](#)
- Abbildung 7: [Beispiel für ein Begriffssystem \(Abstraktionssystem mit polydimensionaler Reihe\)](#)
- Abbildung 8: [Beispiel für ein Begriffssystem \(Bestandsbeziehungen\)](#)
- Abbildung 9: [Beispiel für ein Begriffssystem \(Abstraktionssystem mit monodimensionalen Reihen\)](#)
- Abbildung 10: [Begriffliche Unterschiede im Sprachvergleich](#)
- Abbildung 11: [Beispiel für mit LookUp visualisiertes Relationssystem](#)