

DIN ISO 26162



ICS 01.020; 35.240.30

Einsprüche bis 2016-05-18

Entwurf**Systeme zur Verwaltung von Terminologie, Wissen und Content –
Gestaltung, Einrichtung und Pflege von Terminologieverwaltungssystemen
(ISO 26162:2012);
Text Deutsch und Englisch**

Systems to manage terminology, knowledge and content –
Design, implementation and maintenance of terminology management systems
(ISO 26162:2012);
Text in German and English

Systèmes de gestion de la terminologie, de la connaissance et du contenu –
Conception, mise en oeuvre et maintenance des systèmes de gestion de la terminologie
(ISO 26162:2012);
Texte en allemand et anglais

Anwendungswarnvermerk

Dieser Norm-Entwurf mit Erscheinungsdatum 2016-03-18 wird der Öffentlichkeit zur Prüfung und Stellungnahme vorgelegt.

Weil die beabsichtigte Norm von der vorliegenden Fassung abweichen kann, ist die Anwendung dieses Entwurfes besonders zu vereinbaren.

Stellungnahmen werden erbeten

- vorzugsweise online im Norm-Entwurfs-Portal des DIN unter www.entwuerfe.din.de bzw. für Norm-Entwürfe der DKE auch im Norm-Entwurfs-Portal der DKE unter www.entwuerfe.normenbibliothek.de, sofern dort wiedergegeben;
- oder als Datei per E-Mail an nat@din.de möglichst in Form einer Tabelle. Die Vorlage dieser Tabelle kann im Internet unter www.din.de/stellungnahme oder für Stellungnahmen zu Norm-Entwürfen der DKE unter www.dke.de/stellungnahme abgerufen werden;
- oder in Papierform an den DIN-Normenausschuss Terminologie (NAT), 10772 Berlin (Hausanschrift: Burggrafenstr. 6, 10787 Berlin).

Die Empfänger dieses Norm-Entwurfs werden gebeten, mit ihren Kommentaren jegliche relevanten Patentrechte, die sie kennen, mitzuteilen und unterstützende Dokumentationen zur Verfügung zu stellen.

Gesamtumfang 167 Seiten

DIN-Normenausschuss Terminologie (NAT)

Inhalt

	Seite
Nationales Vorwort.....	5
Nationaler Anhang NA (informativ) Literaturhinweise	5
Einleitung	6
1 Anwendungsbereich.....	7
2 Normative Verweisungen	7
3 Begriffe	7
3.1 Ressourcen.....	7
3.2 Datenkategorien	9
3.3 Datenmodellierung.....	12
3.4 Anwendungsbereiche	13
4 Terminologieverwaltungssysteme (TVS)	14
4.1 Allgemeine Beschreibung.....	14
4.2 Kommerzielle TVS oder Eigenprogrammierungen.....	14
4.3 Vorgegebene oder frei definierbare Eintragsstruktur	14
4.4 Desktop-, Client-Server- oder webbasierte TVS.....	15
4.5 Eigenständige, integrierte oder hybride TVS	15
4.6 Einsprachige, zweisprachige oder mehrsprachige TVS.....	16
4.7 Datenbankbasierte oder strukturierte textbasierte TVS	16
4.8 TVS mit einer oder mehreren zugrunde liegenden Datenbanken	16
5 Projektphasen.....	16
5.1 Übersicht.....	16
5.2 Erstellung einer Vorstudie	17
5.3 Erstellung einer Machbarkeitsstudie.....	17
5.4 Durchführung einer Anwendungsfallanalyse	17
5.5 Festlegung der Systemanforderungen.....	18
5.6 Durchführung einer Kosten-Nutzen-Analyse	18
5.7 Systemplanung	18
5.8 Systementwicklung.....	19
5.9 Bereitstellung des Systems	19
5.10 Testbetrieb des Systems	19
5.11 Auffüllung, Nutzung und Pflege des TVS	19
6 Benutzerzentriertes Design.....	20
6.1 Grundlagen.....	20
6.2 Schritte zur Umsetzung des benutzerzentrierten Ansatzes	21
6.3 Ermittlung von Benutzergruppen und -bedürfnissen	21
6.4 Ermittlung der Endprodukte.....	25
6.5 Durchführung einer Aufgabenanalyse und Vorbereitung von Anwendungsfällen.....	25
6.6 Anforderungen identifizieren und priorisieren	26
6.7 Durchführung einer Wettbewerbsanalyse.....	28
6.8 Entwicklung und Bewertung eines Prototypen.....	28
6.9 Anpassung der Gestaltung an die Benutzerresonanz.....	28
6.10 Durchführung eines Beta-Tests.....	28
7 Terminologische Datenkategorien	29
7.1 Einführung in das Thema Datenkategorien.....	29

7.2	Grundsätze zur Auswahl und Verwendung von Datenkategorien	30
7.2.1	Begriffsorientierung.....	30
7.2.2	Kontextuelle und semantische Datenkategorien	31
7.2.3	Benennungsautonomie	31
7.2.4	Wiederholbarkeit.....	32
7.2.5	Datengranularität.....	32
7.2.6	Datenelementarität	33
7.3	Arten von Datenkategorien	33
7.3.1	Offene, geschlossene und einfache Datenkategorien.....	33
7.3.2	Obligatorische, fakultative, automatische und voreingestellte Datenkategorien	34
7.3.3	Datenkategorien mit Nur-Lesezugriff, Lese- und Schreibzugriff sowie versteckte Datenkategorien	34
7.3.4	Multimediadateien.....	35
7.3.5	Gemeinsame Ressourcen.....	35
7.3.6	Beziehungen zwischen Einträgen.....	35
7.4	Dateneintragsstrukturen	37
7.5	Auswahl der Datenkategorien.....	37
7.6	Übersetzungsspezifische Datenkategorien	38
7.7	Präskriptive Datenkategorien.....	38
7.8	Arbeitsorganisatorische Datenkategorien	38
7.9	Standardisierte Datenkategorienamen und Datenkategoriebegriffe.....	38
8	Datenmodellierung.....	39
8.1	Das terminologische Metamodell.....	39
8.2	Datenmodellierung zur Begriffsorientierung.....	40
8.3	Anwendungsorientierte Vorgehensweise	45
8.4	Beispiele für Datenmodellierungen	45
8.5	Bestandsdatenübernahme.....	52
9	Einrichtung des TVS.....	53
10	Bereitstellung des TVS.....	55
10.1	Bereitstellungsaktivitäten	55
10.2	Vorbereitung von Dokumentation, Onlinehilfen und Lehrmaterial	55
10.3	Support- und Wartungsangebot.....	56
10.4	Abhängigkeiten zwischen Projektbeteiligten klären.....	56
10.5	Ankündigung des TVS sowie Werbemaßnahmen	56
10.6	Auslieferung des TVS.....	56
10.7	Schulungsangebot	57
10.8	Ermittlung der Benutzerzufriedenheit	57
11	Benutzeroberflächen	57
11.1	Gestaltung der Benutzeroberfläche.....	57
11.2	Darstellung terminologischer Datenkategorien.....	58
11.3	Anzeige und Anordnung von terminologischen Einträgen.....	59
12	Eingeben und Bearbeiten von Daten.....	62
12.1	Manuelle Dateneingabe.....	62
12.2	Datenimport.....	70
12.3	Datenbearbeitung	71
12.4	Datenvalidierung.....	71
12.5	Automatische Generierung oder Modifizierung von Daten.....	71
12.6	Hinzufügen von Hyperlinks	72
12.7	Hinzufügen von Multimediadateien.....	72
13	Suchfunktionen	73
13.1	Datenbanksuche	73
13.2	Suche nach einer Benennung.....	75
13.3	Suche anhand von Begriffsnummern oder Eigenschaften	75

13.4	Komplexe Filter- und Suchfunktionen.....	76
13.5	Suche in Textfeldern.....	76
13.6	Durchblättern	76
14	Datenausgabe.....	77
14.1	Arten der Datenausgabe	77
14.2	Darstellung der Suchergebnisse	77
14.3	Sortierung.....	79
14.4	Ausdrucke	82
14.5	Datenexport in andere Dateiformate.....	84
14.6	Datenexport in andere Anwendungen.....	84
15	Organisation und Verwaltung des TVS	85
15.1	Erstellung eines Verwaltungsplans	85
15.2	Bedeutung des Datenflussmanagements.....	85
15.3	Änderungen des Datenmodells	85
15.4	Datensicherheit bieten	86
15.5	Zugriffskontrolle.....	87
15.6	Unterstützung von Austauschformaten	87
15.7	Personal für das TVS.....	87
15.8	Controlling und Ressourcenmanagement.....	88
Anhang A (informativ) Fallbeispiele: Datenkategorien und Datenmodellierung		89
Literaturhinweise.....		92

Nationales Vorwort

Dieses Dokument ist die deutsche Übersetzung der ISO 26162:2012. Fachlich zuständig für diesen deutschen Norm-Entwurf ist der Arbeitsausschuss NA 105-00-05 AA „Systeme für die Verwaltung von Terminologie, Wissen und Content“ im DIN-Normenausschuss Terminologie (NAT).

Das dieser Norm zugrunde liegende Dokument (ISO 26162:2012) wurde von der Arbeitsgruppe WG 4 „Database management“ vom Technischen Komitee ISO/TC 37 „Terminology and other language and content resources“, Subkomitee SC 3 „Systems to manage terminology, knowledge and content“, erarbeitet.

Um Zweifelsfälle in der Übersetzung auszuschließen, ist die englische Originalfassung der ISO 26162 entsprechend der diesbezüglich durch ISO erteilten Erlaubnis beigelegt. Die Nutzungsbedingungen für den deutschen Text des Norm-Entwurfes gelten gleichermaßen auch für den englischen ISO-Text.

Wir möchten auf folgende Fehler in der ISO-Norm von 2012 hinweisen:

Bild 12 – In der ISO-Ausgabe wurde versehentlich Bild 7 doppelt eingesetzt.

Bild 14 – In der ISO-Ausgabe wurde versehentlich Bild 8 doppelt eingesetzt.

Bild 22 – In der ISO-Ausgabe wurde versehentlich Bild 10 doppelt eingesetzt.

Bilder 23, 25, 26, 27, 28, 29, 30 bzw. 31 – In der ISO-Ausgabe wurden versehentlich die Bilder 11, 13, 14, 15, 16, 17, 18 bzw. 19 doppelt eingesetzt.

Eine Korrektur der Bilder wurde bei ISO bereits veranlasst.

Für die in diesem Dokument zitierten Internationalen Normen wird im Folgenden auf die entsprechenden Deutschen Normen hingewiesen:

ISO 704 siehe DIN 2330

ISO 1087-1 siehe DIN 2342

Nationaler Anhang NA (informativ)

Literaturhinweise

DIN 2330, *Begriffe und Benennungen — Allgemeine Grundsätze*

DIN 2342, *Begriffe der Terminologielehre*

Systeme zur Verwaltung von Terminologie, Wissen und Content — Gestaltung, Einrichtung und Pflege von Terminologieverwaltungssystemen

Einleitung

Zur Sammlung, Verwaltung und Speicherung von terminologischen Daten steht eine große Auswahl an Terminologieverwaltungssystemen (TVS) zur Verfügung. Diese verwenden verschiedene Arten von Datenbankmanagementsystemen – von Computeranwendungen für einzelne Nutzer bis hin zu Client-Server-Systemen bzw. webbasierter Software für große Unternehmen und Behörden. Terminologische Datensammlungen basieren auf verschiedenen Datenmodellen und bestehen aus unterschiedlichen Zusammenstellungen von Datenkategorien (Datenkategorieauswahl). Um die Interoperabilität der Systeme zu gewährleisten und Mehraufwand zu vermeiden, ist es wichtig, Normen und Richtlinien auszuarbeiten, die die Entwicklung und Verwendung von terminologischen Datensammlungen sowie die gemeinsame Nutzung und den Austausch von Daten regeln.

Um den Austausch terminologischer Daten zu erleichtern und einen integrierten Ansatz zur Analyse bestehender terminologischer Datensammlungen sowie zur Entwicklung neuer Systeme zu schaffen, veröffentlichte ISO/TC 37 bisher die folgenden Normen: ISO 704, ISO 12620, ISO 16642.

Eine gewisse Kenntnis der genannten Dokumente ist für die Lektüre dieser Internationalen Norm von Vorteil.

Die vorliegende Internationale Norm enthält Leitlinien für die Auswahl relevanter Datenkategorien, die Entwicklung und Einführung eines Datenmodells sowie einer TVS-Benutzeroberfläche und berücksichtigt dabei die geplanten Benutzergruppen. Die in diesem Dokument beschriebenen Richtlinien sind für die erfolgreiche Entwicklung eines TVS unverzichtbar und helfen, kostenintensive Fehler zu vermeiden, sowie das geeignete TVS für einen konkreten Verwendungszweck auszuwählen.

1 Anwendungsbereich

Diese Internationale Norm legt Kriterien für die Gestaltung, Einrichtung und Pflege von Terminologieverwaltungssystemen (TVS) fest.

Diese Internationale Norm nennt die Gründe für den Einsatz eines TVS und informiert über Benutzergruppen und deren Bedürfnisse, notwendige Schritte bei der Gestaltung und Einrichtung eines TVS sowie Organisation und Verwaltung eines terminologischen Datenbestandes. Sie enthält allgemeine Grundsätze zur Auswahl und Nutzung von Datenkategorien zur Terminologieverwaltung in verschiedenen Nutzungsumgebungen.

Diese Internationale Norm richtet sich an Terminologen, Softwareentwickler und andere Benutzergruppen, die an der Entwicklung und Beschaffung eines TVS beteiligt sind.

2 Normative Verweisungen

Die folgenden zitierten Dokumente sind für die Anwendung dieses Dokuments erforderlich. Bei datierten Verweisungen gilt nur die in Bezug genommene Ausgabe. Bei undatierten Verweisungen gilt die letzte Ausgabe des in Bezug genommenen Dokuments (einschließlich aller Änderungen).

ISO 704, *Terminology work — Principles and methods*

ISO 12620, *Terminology and other language and content resources — Specification of data categories and management of a Data Category Registry for language resources*

ISO 16642, *Computer applications in terminology — Terminology markup framework*

ISO 30042, *Systems to manage terminology, knowledge and content — TermBase eXchange (TBX)*

3 Begriffe

Für die Anwendung dieses Dokuments gelten die folgenden Begriffe.

3.1 Ressourcen

3.1.1

terminologischer Datenbestand

terminologische Datensammlung

Text- oder Datenbestand, der aus **terminologischen Einträgen** (3.1.4) aufgebaut ist

ANMERKUNG Nach ISO 24613:2008.

3.1.2

Terminologieverwaltungssystem

Terminologieverwaltungsprogramm

Terminologiemanagementsystem

TVS

Softwaretool, das speziell der Sammlung und Pflege terminologischer Daten sowie dem Zugriff auf diese Daten dient

3.1.3

Terminologiedatenbank

TDB

terminologische Datenbank

Termbank

Datenbank, die einen **terminologischen Datenbestand** (3.1.1) enthält

3.1.4

terminologischer Eintrag

Teil eines **terminologischen Datenbestandes** (3.1.1), der terminologische Daten zu einem Begriff enthält

ANMERKUNG Nach ISO 1087-2, 2.22.^{N1)}

3.1.5

Begriffsorientierung

Ansatz zur Verwaltung von Terminologie, wonach ein **terminologischer Eintrag** (3.1.4) genau einen Begriff bzw. zwei oder mehr **quasi-äquivalente Begriffe** (3.1.7) beschreibt

ANMERKUNG Ein begriffsorientierter Eintrag kann mehrere Benennungen enthalten, welche daher als semantisch äquivalent gelten.

3.1.6

äquivalenter Begriff

Begriff in einer Sprache, der dieselben Merkmale aufweist wie ein entsprechender Begriff in einer anderen Sprache

3.1.7

quasi-äquivalenter Begriff

Begriff in einer Sprache, der zwar nicht genau dieselben Merkmale aufweist wie ein entsprechender Begriff in einer anderen Sprache, aber in manchen Kontexten als Äquivalent verwendet werden kann

3.1.8

Querverweis

Benennung, die in einem Textfeld – beispielsweise */definition/* oder */context/* – verwendet wird und einen Begriff bezeichnet, der in einem anderen **terminologischen Eintrag** (3.1.4) im selben **terminologischen Datenbestand** (3.1.1) definiert wird

3.1.9

Dublette

Doppeleintrag

terminologischer Eintrag (3.1.4), der denselben Begriff beschreibt wie ein anderer Eintrag

ANMERKUNG normalerweise werden Dubletten vom TVS erkannt, indem zwei Benennungen derselben Form identifiziert werden. Dubletten werden allerdings aufgrund identischer Begriffe, nicht aufgrund identischer Benennungen bestimmt. Dubletten sind nicht mit Homografen zu verwechseln.

3.1.10

Begriffssystem

Menge von Begriffen, die entsprechend ihren Beziehungen untereinander angeordnet sind

[ISO 1087-1:2000, 3.2.11]

^{N1)} Nationale Fußnote: Im Ausgangstext der ISO 26162 folgt eine Anmerkung, die sich auf eine ausgangssprachliche Benennung bezieht. Diese Anmerkung wurde nicht übersetzt.

3.1.11**Begriffsplan**

grafische Darstellung eines **Begriffssystems** (3.1.10)

[ISO 1087-1:2000, 3.2.12]

3.1.12**Bestandsdaten**

Altdaten

terminologische Daten, die in einer bereits bestehenden Datei oder Datenbank enthalten sind und für den Import in ein **TVS** (3.1.2) in Betracht kommen

ANMERKUNG Bestandsdaten können in Form von bisher verwendeten Datenbanken, Textverarbeitungsdateien, CSV-, SGML-, HTML- bzw. XML-Dateien oder dergleichen vorliegen. Die Umwandlung solcher Dateien in ein Format, das mit einem neuen TVS kompatibel ist, kann eine große Herausforderung darstellen.

3.1.13**Benennung**

Terminus

Fachausdruck

ein Wort oder mehrere Wörter zur Bezeichnung eines Begriffs

BEISPIELE „Weltmeisterschaft“ und „Olympische Spiele“ sind zwei Benennungen aus dem Fachgebiet Sport.

ANMERKUNG 1 Kann das Wort bzw. können die Wörter mehr als einen Begriff bezeichnen, stellt jedes Paar aus Wort und Begriff eine eigene Benennung dar. So handelt es sich beispielsweise bei „Maus“ (Eingabegerät eines Computers) und „Maus“ (Nagetier) um zwei verschiedene Benennungen.

ANMERKUNG 2 In der Terminologiewissenschaft bezeichnen Benennungen Begriffe aus speziellen Fachgebieten; allgemeinsprachliche Wörter gelten nicht als Benennungen. In Einträgen terminologischer Datensammlungen werden dennoch manchmal allgemeinsprachliche Wörter erfasst und dort als Benennung betrachtet.

3.2 Datenkategorien**3.2.1****Datenkategorie**

Ergebnis der Spezifikation eines Datenfeldes

[ISO 1087-2:2000, 6.14]

3.2.2**Datenelement**

Dateneinheit, die in einem bestimmten Zusammenhang als unteilbar gilt

[ISO 1087-2:2000, 6.11]

3.2.3**Datengranularität**

Grad der Datengenauigkeit

ANMERKUNG Die Gesamtheit der einzelnen **Datenkategorien** (3.2.1) */part of speech/, /grammatical gender/* und */grammatical number/* beispielsweise verfügt gemeinsam über eine höhere Datengranularität als die alleinige Datenkategorie */grammar/*.

3.2.4**Datenelementarität**

Prinzip, wonach ein einzelnes Datenfeld nur eine einzige Information enthalten darf

ANMERKUNG Werden beispielsweise sowohl Vollform als auch Abkürzung im selben Datenfeld aufgeführt, so stellt dies einen Verstoß gegen das Prinzip der Datenelementarität dar.

3.2.5

Benennungsautonomie

Prinzip, wonach alle Benennungen in einem **terminologischen Eintrag** (3.1.4) mit der gleichen Menge von **Datenkategorien** (3.2.1) beschrieben werden können

3.2.6

Verzeichnis der Datenkategorien

Data Category Registry

DCR

Zusammenstellung standardisierter **Datenkategorien** (3.2.1), die als Referenz für die Definition linguistischer Annotationsschemata oder anderer Formate im Bereich der Sprachressourcen dienen

[ISO 12620:2009, 3.2.1]

ANMERKUNG Das DCR des ISO/TC 37 enthält **Datenkategoriespezifikationen** (3.2.7), die historische, deskriptive und administrative Informationen sowie weitere Metadaten enthalten.

3.2.7

Datenkategoriespezifikation

Reihe von Attributen, die zur vollständigen Beschreibung einer bestimmten **Datenkategorie** (3.2.1) dienen

[ISO 12620:2009, 3.2.2]^{N2)}

3.2.8

Auswahl der Datenkategorien

Data Category Selection

DCS

Zusammenstellung von **Datenkategorien** (3.2.1), die aus der **Datenkategorieregistratur** (3.2.6) ausgewählt werden

[ISO 12620:2009, 3.2.3]

3.2.9

komplexe Datenkategorie

Datenkategorie (3.2.1) mit **konzeptuellem Bereich** (3.2.11)

[ISO 12620:2009, 3.1.7]

3.2.10

offene Datenkategorie

komplexe Datenkategorie (3.2.9), deren **konzeptueller Bereich** (3.2.11) nicht auf eine festgelegte Wertemenge begrenzt ist

[ISO 12620:2009, 3.1.8]

3.2.11

konzeptueller Bereich

Menge zulässiger **Wertebedeutungen** (3.2.14)

ANMERKUNG 1 Auf Grundlage von ISO/IEC 11179-1:2004, 3.3.6.

ANMERKUNG 2 Die Wertebedeutungen in einer konzeptuellen Domäne können einzeln spezifiziert, durch zusätzliche Einschränkungen präzisiert oder durch eine Beschreibung ausgedrückt werden. Die **Datenkategorie** (3.2.1) /term/ beispielsweise ist durch ihre Definition festgelegt und darf daher z. B. keine kontextuellen oder grammatikalischen Informationen enthalten. Dennoch wäre es unmöglich, sämtliche Werte zu spezifizieren, die dieser Datenkategorie zugeordnet werden können.

^{N2)} Nationale Fußnote: Im Ausgangstext der ISO 26162 folgt eine Anmerkung, die sich auf eine ausgangssprachliche Benennung bezieht. Diese Anmerkung wurde nicht übersetzt.

3.2.12**geschlossene Datenkategorie**

komplexe Datenkategorie (3.2.9), deren **konzeptueller Bereich** (3.2.11) auf eine Reihe festgelegter **einfacher Datenkategorien** (3.2.13) beschränkt ist

[ISO 12620:2009, 3.1.13]

3.2.13**einfache Datenkategorie**

Datenkategorie (3.2.1) ohne **konzeptuellen Bereich** (3.2.11)

[ISO 12620:2009, 3.1.12]

3.2.14**Wertebedeutung**

Bedeutung oder semantischer Inhalt eines Wertes

[ISO/IEC 11179-1:2004, 3.3.39]

ANMERKUNG ISO/TC 37 betrachtet Aufzählungswerte als einfache – also als eigenständige – Datenkategorien. Die Wertebedeutung wird stets in Verbindung mit dem gesamten Wertebereich und der zugehörigen geschlossenen Datenkategorie betrachtet und ist nicht nur eine Eigenschaft des Wertes selbst.

3.2.15**Wertebereich**

Menge **zulässiger Werte** (3.2.16)

[ISO/IEC 11179-1:2004, 3.3.38]

3.2.16**zulässiger Wert**

Ausdruck einer **Wertebedeutung** (3.2.14), die in einem bestimmten **Wertebereich** (3.2.15) zugelassen ist

[ISO/IEC 11179-1:2004, 3.3.28]

3.2.17**thematischer Bereich**

Klasse von Anwendungen, die durch die Ähnlichkeit der zu manipulierenden Datenstrukturen identifiziert wird

[ISO 12620:2009, 3.4.3]

BEISPIELE Terminologie, Lexikographie, morphosyntaktische Annotation.

3.2.18**Themenbereichsprofil**

Profil

Darstellungsweise innerhalb einer **Datenkategoriespezifikation** (3.2.7) des **thematischen Bereichs** (3.2.17), der eine **Datenkategorie** (3.2.1) zugeordnet wird

[ISO 12620:2009, 3.4.4]

ANMERKUNG Eine Datenkategorie kann über mehrere Themenbereichsprofile verfügen, die auf eine Nutzung durch mehrere thematische Domänen hinweisen.

3.3 Datenmodellierung

3.3.1

Datenmodell

grafische und/oder lexikalische Darstellung von Daten, die deren Eigenschaften, Struktur sowie wechselseitige Beziehungen präzisiert

[ISO/IEC 11179-1:2004, 3.2.7]

3.3.2

Datenmodellierung

Verfahren zur Strukturierung und Organisation von Daten, in der Regel zur Implementierung in ein Datenbankmanagementsystem

3.3.3

Datenmodellierungsvarianz

Abweichung bei der Zuordnung von **Datenkategorien** (3.2.1) zu Datenmodellen infolge unterschiedlicher Philosophien bezüglich der Informationsanordnung in einem **terminologischen Eintrag** (3.1.4)

3.3.4

Metamodell

Datenmodell (3.3.1), das ein oder mehrere andere Datenmodelle näher beschreibt

[ISO/IEC 11179-1:2004, 3.2.20]

3.3.5

Metadaten

Daten, die andere Daten definieren und beschreiben

[ISO/IEC 11179-1:2004, 3.2.16]

3.3.6

globale Informationen

GI

technische und administrative Informationen, die sich auf eine ganze Datensammlung beziehen

[ISO 16642:2003, 3.7]

BEISPIEL Titel der Datensammlung, Änderungshistorie.

3.3.7

Zusatzinformation

Begleitinformation

ZI

Informationen, die den Inhalt von **terminologischen Einträgen** (3.1.4) ergänzen und für die gesamte **terminologische Datensammlung** (3.1.1) gelten

[ISO 16642:2003, 3.7]

BEISPIEL Typische Beispiele für Begleitinformationen sind hierarchische Fachgebietsklassifikationen, Beschreibungen der Institution sowie Bibliographieangaben.

3.3.8

gemeinsame Ressource

Informationsobjekte, auf die von jedem beliebigen terminologischen oder lexikografischen Eintrag in einer terminologischen oder lexikografischen Ressource aus zugegriffen werden kann

BEISPIEL Zu den gemeinsamen Ressourcen gehören in der Regel bibliographische Einträge, Zuständigkeitsangaben, Namensraumbezeichner, häufig als Referenz genutztes Textmaterial, Listen mit geografischen Angaben und externe Dateien wie beispielsweise Grafiken oder Audiodateien.

3.3.9**Sprachebene****LS**

Teil eines **terminologischen Eintrags** (3.1.4), der Informationen zu einer bestimmten Sprache enthält

[ISO 16642:2003, 3.9]

3.3.10**Benennungsebene****TS**

Teil der **Sprachebene** (3.3.9), der Informationen zu einer **Benennung** (3.1.13) enthält

[ISO 16642:2003, 3.15]

3.3.11**Klasse****Objektklasse**

Beschreibung in UML für eine Reihe von Objekten, die dieselben Elemente teilen

3.3.12**Multiplizität**

Anzahl von Instanzen einer **Klasse** (3.3.11), die in Beziehung zu einer Instanz einer anderen Klasse stehen

ANMERKUNG Viele terminologische Einträge können beispielsweise dasselbe Erstellungsdatum haben; ein einzelner terminologischer Eintrag kann allerdings nur ein Erstellungsdatum haben.

3.4 Anwendungsbereiche**3.4.1****Sprachplanung**

bewusste Beeinflussung von menschlichem Verhalten in Hinblick auf Erwerb, Struktur oder funktionale Zuordnung von Sprache

BEISPIEL Zur Sprachplanung gehören unter anderem die Standardisierung von Rechtschreibung und Grammatik, die Festlegung nationaler Amtssprachen, die Schaffung geeigneter Neologismen zur Erweiterung des Ausdrucksspektrums einer Sprache als wissenschaftliches und wirtschaftliches Kommunikationsmittel sowie Maßnahmen zum Schutz der Sprache vor fremden Einflüssen.

3.4.2**deskriptive Terminologearbeit**

Ansatz zur Terminologieverwaltung, der dokumentiert, auf welche Art **Benennungen** (3.1.13) in bestimmten Kontexten verwendet werden, ohne einen bevorzugten Gebrauch anzugeben

3.4.3**präskriptive Terminologearbeit****normende Terminologie**

Ansatz zur Terminologieverwaltung, der einen bevorzugten Gebrauch angibt

3.4.4**normative Terminologearbeit**

Ansatz zur Terminologieverwaltung, der bei der Normungsarbeit oder bei staatlichen Regelungen eingesetzt wird

3.4.5**Übersetzungseditor**

Software, die das Erstellen und Überarbeiten von Übersetzungen erleichtert

3.4.6

regelbasiertes Schreiben

kontrolliertes Schreiben

Erstellen von Texten mit eingeschränktem Wortschatz und begrenzter inhaltlicher Komplexität, um eindeutige, gut verständliche Dokumente zu verfassen

3.4.7

Lokalisierung

L10N

Prozess der sprachlichen und kulturellen Anpassung eines Produkts an die Zielumgebung (Land/Region und Sprache), in der es verwendet und verkauft werden soll

[Localization Industry Standards Association]

4 Terminologieverwaltungssysteme (TVS)

4.1 Allgemeine Beschreibung

Ein TVS ist eine Softwareanwendung, die der Sammlung und Pflege terminologischer Daten dient und den Zugriff auf sie ermöglicht. TVS werden von Übersetzern, Terminologen, technischen Redakteuren und anderen Benutzergruppen eingesetzt. Die Grundlagen der Terminologearbeit, auf denen TVS basieren müssen, sind in ISO 704 festgelegt. Demnach enthalten die mit dem TVS erstellten terminologischen Datenbestände terminologische Einträge, die der Dokumentation von Begriffen und all ihren Benennungen (Synonymen bzw. fremdsprachigen Äquivalenten) dienen.

TVS werden üblicherweise zur Ausarbeitung, Pflege und Verbreitung terminologischer Datensammlungen in einer Vielzahl von Nutzungsumgebungen verwendet. So werden sie unter anderem von nationalen Einrichtungen, Behörden, Interessengemeinschaften, Unternehmen und Einzelpersonen eingesetzt. Je nach organisatorischen Rahmenbedingungen muss ein TVS über eine Vielzahl von Funktionen verfügen, unterschiedliche Aufgaben erfüllen und den Anforderungen der potentiellen Benutzer gerecht werden.

Bei der Planung einer Softwarelösung zur Terminologieverwaltung ist es daher wichtig, die verschiedenen Typen bzw. Klassen von TVS zu berücksichtigen.

4.2 Kommerzielle TVS oder Eigenprogrammierungen

Auf dem Softwaremarkt findet sich ein breites Spektrum an Terminologieverwaltungswerkzeugen. Bei einem Großteil handelt es sich um kommerzielle Produkte, wobei das Angebot an Freeware bzw. Open-Source-Software wächst. Erfüllt eines der Standardsoftwareprodukte die organisatorischen, funktionalen und methodischen Anforderungen der jeweiligen Nutzungsumgebung, kann es ohne weitere Modifizierung direkt verwendet werden. Häufig muss das System jedoch auf die spezifischen Anforderungen des Kunden abgestimmt und entsprechend angepasst werden. Ist keine Anpassung der kommerziellen Softwareprodukte an die Wünsche des Kunden möglich, so muss ein eigenes TVS intern entwickelt werden. Kosten, Wartungsaufwand, Endnutzerunterstützung und Interoperabilität stellen maßgebliche Faktoren bei der Entscheidung dar, ob ein TVS gekauft, geleast oder individuell entwickelt werden soll.

4.3 Vorgegebene oder frei definierbare Eintragsstruktur

Kommerzielle TVS können sowohl über vorgegebene als auch über frei definierbare Datenmodellierungsstrukturen verfügen. Ein TVS mit einer vorgegebenen Zusammenstellung von Datenkategorien, vordefinierten Sprachkombinationen und/oder einem festgelegten Datenmodell muss ausschließlich dann verwendet werden, wenn die Voreinstellungen den spezifischen Anforderungen der Nutzungsumgebung entsprechen. In einem frei definierbaren TVS lassen sich nutzerspezifische Datenkategorien und Eintragsstrukturen anlegen, um die Software den terminologischen Bedürfnissen des Anwenders anzupassen. Hier ist auch eine Umrüstung bei zukünftigen Anforderungsänderungen möglich.

4.4 Desktop-, Client-Server- oder webbasierte TVS

Die einfachste Form des TVS wird als Einzelplatzsystem auf einem lokalen Computer installiert. Mitunter werden die zugrunde liegenden Datenbankdateien auf einem lokal vernetzten Computer gespeichert und von mehreren Client-Computern verwendet. In großen Unternehmen mit vielen Benutzern mit jeweils unterschiedlichen Rollen kann jedoch nur ein TVS mit einer Client-Server-Architektur den unternehmensspezifischen Anforderungen gerecht werden. Die Nutzung eines Client-Server-basierten TVS erfordert die Installation eines Clients auf sämtlichen lokalen Computern sowie die Installation einer Serversoftware, die den parallelen Zugriff verschiedener Clients (mit unterschiedlichen Aufgaben) auf die Datenbankdateien ermöglicht.

Müssen mehrere Anwender weltweit auf dieselben terminologischen Daten zugreifen, stellt ein webbasiertes TVS die einzig gangbare Lösung dar. Hier wird das TVS über einen Standard-Webbrowser aufgerufen und verwendet; die Installation von TVS-spezifischer Software auf dem lokalen Computer entfällt. Bei Einsatz eines webbasierten TVS muss zusätzlich entschieden werden, ob ein eigener Server eingerichtet und verwaltet oder ob das TVS eines Dritten verwendet werden soll. Im zweiten Fall besteht beispielsweise die Wahlmöglichkeit zwischen großen Softwareanbietern und unentgeltlichen Angeboten von webbasierten Umgebungen für gemeinschaftliche Zusammenarbeit. Bevor die Einrichtung und Verwaltung des Servers jedoch an Dritte ausgelagert wird, sollten Faktoren wie Kosten und Erfahrung des Anbieters sowie Sicherheit, Vertraulichkeit und solide Zugänglichkeit der Daten sorgfältig abgewogen werden.

4.5 Eigenständige, integrierte oder hybride TVS

Ein eigenständiges TVS ist ein autonomes Softwarepaket, mit dem Terminologie unabhängig von anderen Anwendungen verwaltet werden kann. Zur Integration der Terminologie aus einem eigenständigen TVS in ein anderes System – z. B. in ein Programm zum regelbasierten Schreiben oder in ein computergestütztes Übersetzungsprogramm (CAT-Tool) – muss die Terminologie in der Regel aus dem TVS exportiert und anschließend in die gewünschte Anwendung importiert werden. Mitunter kann die Terminologie auch mithilfe eines speziell entwickelten anwendungsspezifischen Plug-ins integriert werden. Da eigenständige TVS nicht an eine bestimmte Softwareanwendung gebunden sind, können die in diesen Systemen verwalteten Daten in der Regel besonders flexibel an andere Programme angepasst werden.

Bei einem integrierten TVS ist eine Reihe von Terminologieverwaltungsfunktionen unmittelbar in eine andere Anwendung, z. B. in ein CAT-Tool, eingebunden. Der Benutzer muss in diesem Fall zunächst eine vollständig kompatible Programm-Suite installieren, um das integrierte TVS nutzen zu können. Das integrierte TVS ist mit seinen Funktionen und Merkmalen speziell auf die Software, in die es eingebunden wird, zugeschnitten und ist daher möglicherweise nicht für jedes Anwendungsszenario geeignet. So verfügt beispielsweise ein TVS, das in ein Programm zum regelbasierten Schreiben integriert ist, über Funktionen und Datentypen zur Sprachqualitätssicherung; übersetzungsspezifische Funktionen werden hingegen vermutlich fehlen.

4.6 Einsprachige, zweisprachige oder mehrsprachige TVS

Einsprachige TVS werden hauptsächlich von technischen Autoren und Anwendern verwendet, die normalerweise nicht am Übersetzungsprozess oder der Erstellung mehrsprachiger Dokumente beteiligt sind. Zweisprachige TVS können hingegen einfache Benennungs- bzw. Wortpaare enthalten und arbeiten mitunter nach dem lexikografischen Ansatz, bei dem eine Eins-zu-eins-Entsprechung zwischen einer Benennung A und einer Benennung B ohne Berücksichtigung des Fachgebiets hergestellt wird. Zudem können zweisprachige TVS auch begriffsorientiert sein, d. h. alle Benennungen, die in den beiden Sprachen in einem terminologischen Eintrag gelistet werden, beziehen sich spezifisch auf einen Begriff. Wird eine Benennung mehr als einem Begriff zugeschrieben, wird jeder Begriff in einem gesonderten Eintrag aufgeführt. Mehrsprachige TVS sind in der Regel begriffsorientiert, da die Verwaltung mehrerer Bedeutungen in verschiedenen Sprachen andernfalls eine große Herausforderung darstellen würde. Manche dieser Systeme ermöglichen auch eine zweisprachige Ansicht des mehrsprachigen Terminologiebestands, um das System z. B. in Übersetzungsumgebungen verwenden zu können.

4.7 Datenbankbasierte oder strukturierte textbasierte TVS

Die riesigen Datenmengen, die in einem TVS verwaltet werden müssen, erfordern gewöhnlich den Einsatz einer Datenbank-Engine, die im Hintergrund arbeitet. In Anbetracht der charakteristischen Anforderungen der Terminologieverwaltung – zu denen z. B. eine große Anzahl von Benennungen in mehreren Sprachen, eine Vielzahl von Datenkategorien sowie die (uneingeschränkte) Wiederholbarkeit bestimmter terminologischer Informationen zählen – stellt eine relationale Datenbank die ideale Architektur dar. Auch ein strukturiertes textbasiertes TVS erfüllt viele der charakteristischen Herausforderungen der Terminologieverwaltung, indem die verschiedenen Datentypen wie z. B. in XML gekennzeichnet werden.

4.8 TVS mit einer oder mehreren zugrunde liegenden Datenbanken

Ein TVS kann sämtliche terminologischen Einträge in einer einzelnen Datenbank speichern. Anhand von Attributen kann die Datensammlung in verschiedene Fachgebiete, Bearbeitungsstufen, Kunden und andere Eigenschaften eingeteilt und mittels Filtern virtuell unterteilt werden. Andere Arten von TVS speichern die terminologischen Daten in mehreren physischen Datenbanken, die beispielsweise verschiedene Fachgebiete oder unterschiedliche Kunden repräsentieren. Ist dies der Fall, muss der Anwender entweder eine Datenbank auswählen, die er durchsuchen oder bearbeiten möchte, oder die Anwendung erlaubt einen gleichzeitigen Zugriff auf alle Datenbanken bzw. die Suche darin. Gegebenenfalls kann eine Zugriffshierarchie, d. h. die Reihenfolge der Datenbankabfragen, festgelegt werden. Eine Sonderform stellt hier das föderierte TVS dar, das Daten aus verschiedenen externen Datenbanken zusammenführt, denen möglicherweise unterschiedliche Datenstrukturen zugrunde liegen.

5 Projektphasen

5.1 Übersicht

Wie bereits in 4.2 beschrieben, ist der Kauf bzw. die sonstige Nutzung eines existierenden TVS für bestimmte Zwecke denkbar. Unter Umständen kann es jedoch erforderlich sein, ein neues TVS zu entwickeln. Die Einrichtung eines solchen Systems umfasst die im Folgenden beschriebenen Projektphasen. Dabei spielt es keine Rolle, ob das TVS von einem Drittanbieter oder intern entwickelt wird. Manche der genannten Phasen können im Einzelfall auch entfallen. Es gibt Projekte, die eine iterative Vorgehensweise erfordern: Hier werden die einzelnen Phasen so lange wiederholt, bis das gewünschte Ergebnis erzielt wird. Besonders häufig kommt dies bei agilen Softwareprojekten vor, an deren Entwicklung auch Endbenutzer beteiligt sind.

- Erstellung einer Vorstudie (siehe 5.2)
- Erstellung einer Machbarkeitsstudie (siehe 5.3)
- Durchführung einer Anwendungsfallanalyse (siehe 5.4)

- Festlegung der Systemanforderungen (siehe 5.5)
- Durchführung einer Kosten-Nutzen-Analyse (siehe 5.6)
- Planung des TVS (siehe 5.7)
- Entwicklung des TVS (siehe 5.8)
- Bereitstellung des TVS (siehe 5.9)
- Testbetrieb des TVS (siehe 5.10)
- Auffüllung, Nutzung und Pflege des TVS (siehe 5.11)

Für jede der genannten Phasen sollte ein ausführlicher Arbeits- und Zeitplan mit den entsprechenden zeitlichen Vorgaben erstellt und die Aufgaben den am Projekt beteiligten Mitarbeitern zugewiesen werden. Ergebnisse sollten in Zwischenberichten dokumentiert werden. Aus dem Zeitplan sollte auch klar hervorgehen, wann Entscheidungen über den Fortgang des Projekts (sogenannte Go-/No-go-Entscheidungen) getroffen werden sollen und ob das TVS neu entwickelt, gekauft oder ein bereits existierendes TVS verwendet oder entsprechend modifiziert werden soll. Da im Laufe des Entwicklungsprozesses Anforderungen priorisiert werden, müssen diese Entscheidungen in unterschiedlichen Projektphasen gegebenenfalls erneut getroffen werden.

Die in diesem Kapitel beschriebenen Phasen können auf den Projektumfang und die beteiligten Organisationen (Handelsunternehmen, Behörden, Bildungseinrichtungen, Übersetzungsagenturen usw.) abgestimmt werden.

5.2 Erstellung einer Vorstudie

In der Vorstudie wird der Projektumfang definiert und festgelegt, welche Benutzergruppen, Organisationseinheiten und Systeme berücksichtigt werden sollen. Zu den Projektbeteiligten und Benutzern können Mitarbeiter mit spezifischen Aufgabenbereichen im Unternehmen (Autoren, Übersetzer usw.), Organisationseinheiten (Verlagsabteilungen und Sprachendienste) oder Geschäftspartner, Lieferanten und Kunden gehören. Außerdem können Verantwortliche für artverwandte Software und Systeme – z.B. Dateiverwaltungssysteme, Workflow-Management-Systeme, Übersetzungssoftware und Autorentools – in den Entwicklungsprozess involviert werden.

Einige der hier genannten Prozesseinheiten oder -teile können jedoch auch bewusst nicht berücksichtigt oder auf eine spätere Projektphase verschoben werden. Die Vorstudie liefert auch eine erste Beschreibung des Entwicklungsprozesses des TVS und begründet, warum die Einführung eines TVS notwendig ist. Etwaige alternative Lösungsmöglichkeiten werden aufgezeigt und die Grundanforderungen der potentiellen Anwender beschrieben. Zudem sollte erörtert werden, welche organisatorischen Rahmenbedingungen zur Einrichtung und für den Betrieb des TVS erforderlich sind.

5.3 Erstellung einer Machbarkeitsstudie

Die Machbarkeitsstudie enthält einen genaueren Überblick über die zu erfüllenden Aufgaben sowie eine Aufstellung der für die Durchführung des Projekts und den späteren Betrieb des TVS benötigten finanziellen und personellen Mittel. Dabei sollte auch die technische und organisatorische Einbindung des TVS in die allgemeine Informationsmanagement-Umgebung des Unternehmens berücksichtigt werden. Die Machbarkeitsstudie dient als Grundlage für die endgültige Entscheidung über die Einrichtung eines TVS.

5.4 Durchführung einer Anwendungsfallanalyse

Eine Anwendungsfallanalyse ist notwendig, um sicherzustellen, dass das TVS den Anforderungen der Benutzerzielgruppe entsprechen wird. Diese Phase wird in 6.3 näher erläutert.

5.5 Festlegung der Systemanforderungen

Anhand der Rückmeldungen von Benutzern und Projektbeteiligten werden in dieser Phase die konkreten Anforderungen an das geplante TVS ermittelt und entsprechende Softwarespezifikationen erstellt. Die Hard- und Softwarekomponenten werden anschließend gemäß den aus den Softwarespezifikationen hervorgehenden Grundanforderungen ausgewählt.

Üblicherweise werden die Anforderungen anhand von Benutzerrückmeldungen priorisiert, da mitunter nicht alle Bedürfnisse berücksichtigt werden können (siehe 6.6).

Neben den benutzerorientierten müssen auch die technischen Anforderungen erfüllt werden, die mit der Softwareinfrastruktur der Organisation zusammenhängen. So muss beispielsweise sichergestellt werden, dass das System den Unternehmensstandards sowie dem Sicherheits- und Wartungskonzept entspricht.

Das Datenmodell für das TVS muss festgelegt werden, bevor das System entwickelt oder gekauft wird.

5.6 Durchführung einer Kosten-Nutzen-Analyse

In dieser Phase müssen die Kosten für Einrichtung und Betrieb des TVS sowie der voraussichtliche Nutzen des Systems ermittelt werden. Dabei ist allerdings zu beachten, dass eine finanzielle Rendite erst nach längerer Nutzung des TVS erzielt werden kann. Jedes Geschäftsszenario, das zur Genehmigung der finanziellen Förderung des Projekts herangezogen wird, sollte neben den finanziellen auch die Erfolgspotenziale im nicht-finanziellen (immateriellen bzw. „weichen“) Bereich berücksichtigen. Zu den „weichen“ Erfolgspotenzialen gehören die Verbesserung von Qualität und Kundenzufriedenheit, eine Kostenentlastung durch die Verringerung von Mehrfachaufwand sowie ein effizienteres Arbeiten durch leistungsstärkere Prozesse, z. B. bei Übersetzungsdienstleistungen oder im Kundendienst. Sollten in dieser Phase noch mehrere alternative Handlungsmöglichkeiten in Frage kommen – beispielsweise die Entwicklung eines neuen oder der Kauf eines kommerziellen TVS – sollten Kosten und Nutzen der verschiedenen Optionen abgeschätzt und verglichen werden.

Die durch den Verzicht auf die Einrichtung eines TVS entstehenden Kosten sind ein weiterer wichtiger Punkt, den es zu berücksichtigen gilt. Dazu zählen neben den oben genannten Faktoren zur Kostenentlastung auch weitere Szenarien: Gerichtskosten bei Markenrechtsverletzungen, die ein TVS hätte vermeiden können, der ohne TVS nicht realisierbare Einsatz von regelbasiertem Schreiben sowie die zusätzlichen Schwierigkeiten bei der Einrichtung eines Content-Management-Systems ohne eine festgelegte Terminologie. Dementsprechend sollten bei der Erstellung des Geschäftsszenarios sämtliche aktuellen und zukünftigen Anwendungsmöglichkeiten von Terminologie berücksichtigt werden.

Das Importieren bereits existierender Daten in das neue System kann mit erheblichen Kosten verbunden sein – vor allem, wenn die vorhandenen Datensätze vor dem Import modifiziert werden müssen. Daher sollten bei der Kostenanalyse die Möglichkeiten, die die verschiedenen in Frage kommenden Systeme zum Import von bestehenden Daten bieten, berücksichtigt werden. Eventuell kann beim Anbieter eines existierenden TVS angefragt werden, ob kundenspezifische Konvertierungsroutinen erstellt werden können.

5.7 Systemplanung

In dieser Phase wird entweder intensiv nach Softwareprodukten gesucht, die zur Einrichtung des TVS genutzt werden können, oder die Entwicklung eines neuen TVS vorangetrieben. Finden sich auf dem Markt mehrere Produkte, die für die Implementierung in Frage kommen, müssen diese bezüglich der herausgearbeiteten Anforderungen bewertet und die bestmögliche Lösung – auch unter Berücksichtigung der entstehenden Kosten – ausgewählt werden. Kommt hingegen nur die Neuprogrammierung eines TVS in Frage, muss ein detailliertes Konzept des Softwaresystems, einschließlich der erforderlichen Teilsysteme, entwickelt sowie das organisatorische Umfeld und das notwendige Personal für die Implementierung bereitgestellt werden.

Zusätzlich müssen bei der Planung sämtliche Anforderungen bezüglich des Imports von bestehenden terminologischen Daten berücksichtigt werden.

5.8 Systementwicklung

Ziel dieser Phase ist es, die notwendigen Hard- und Softwarekomponenten für das TVS zu entwickeln. Die Einführung des TVS und seiner Subsysteme wird entweder durch das Programmieren neuer Software oder das Konfigurieren und/oder kundengerechtes Anpassen bereits vorhandener Programme realisiert. Es ist unerlässlich, die Komponenten sowie ihre Funktionsweise innerhalb des Gesamtsystems in der Implementierungsphase zu dokumentieren und zu testen.

5.9 Bereitstellung des Systems

Die Bereitstellung des Systems umfasst mehrere Teilphasen und sollte im Idealfall in die Testphase integriert werden.

Die Erstbereitstellung sollte sich – vor allem in großen Unternehmen – zunächst auf eine kleine ausgewählte Benutzergruppe beschränken. In dieser Phase sollten das System ausgiebig getestet und eventuell auftretende Mängel beseitigt werden. Die an der Erstbereitstellung beteiligten Personen sollten entsprechende Tests durchführen und zunächst mit einer kleinen Auswahl terminologischer Einträge experimentieren, bevor sie mit dem System an großen Projekten arbeiten.

Eine unternehmensweite Bereitstellung sollte erst nach Abschluss der Testphase (siehe 5.10) durchgeführt werden.

5.10 Testbetrieb des Systems

Nach Abschluss der Erstbereitstellung muss ein umfangreicher Test des Gesamtsystems (einschließlich sämtlicher Softwarefunktionen und Anforderungen) erfolgen. Dieser Test muss den gesamten Verarbeitungsprozess einer repräsentativen Menge terminologischer Daten, deren Größe das System fordern kann, umfassen. Für den Fall, dass Störungen im Testbetrieb auftreten, sollten die Testdaten verworfen werden können. Im Testbetrieb sollten sämtliche Aspekte des Terminologieworkflows sowie mehrere Anwendungsszenarien abgedeckt werden. Die Testphase sollte außerdem sämtliche Programmfunktionen (Ein- und Ausgabe, Import, Export, Anzeige etc.) umfassen und Vertreter aller Benutzerrollen (Terminologen, technische Redakteure, Fachgebietsexperten, Übersetzer usw.) einbinden. Manuelle Tests können durch automatisierte Methoden und Routinen erweitert werden. Die so gewonnenen ersten Daten können später eingesetzt werden, um den Nutzen von Systemänderungen zu analysieren.

Bis zum Abschluss der Testphase sollte es Einzelpersonen nicht erlaubt sein, eine eigenständige Nutzungsumgebung des zu testenden TVS einzurichten.

5.11 Auffüllung, Nutzung und Pflege des TVS

Sind Bereitstellung und Testbetrieb des TVS abgeschlossen, wird das System in den Praxiseinsatz überführt. Die vollständige Bereitstellung des Systems umfasst die Schulung von Systemadministratoren und Anwendern, den Aufbau der erforderlichen Infrastruktur zur Pflege des Systems und die Integration des TVS in den gesamten Workflow. Es kann auch notwendig sein, das TVS mit bereits bestehenden Terminologiebeständen aufzufüllen. Vor dem Import sollten die Daten aber geprüft werden, da eventuell Änderungen an Inhalt, Struktur oder Format erforderlich sein können. Diese Aufgabe kann viel Zeit in Anspruch nehmen; es sollte daher der Arbeitsaufwand abgeschätzt und ein entsprechender zeitlicher Rahmen im Gesamtzeitplan festgelegt werden. Weitere Informationen zum Import von Daten finden sich in 12.2.

6 Benutzerzentriertes Design

6.1 Grundlagen

6.1.1 Das benutzerzentrierte Design beruft sich auf Methoden und Grundsätze des Produktdesigns, das sich an den Anforderungen der späteren Benutzer orientiert. Im folgenden Abschnitt wird erklärt, wie Benutzer in den Gestaltungsprozess einbezogen werden können.

Zur potenziellen Benutzergruppe von TVS gehören:

- Anwender, z. B. Terminologen, Übersetzer, Dolmetscher, Fachgebietsexperten, Lehrer, Studierende, technische Redakteure, Werber, Mediengestalter und Werbetexter, Informations- und Wissensingenieure, Language Engineers, Produktentwickler, Techniker etc.;
- Geräte und Anwendungen, die auf terminologische Informationen zugreifen und sie verarbeiten, z. B. „Controlled Authoring Tools“, Rechtschreib- und Grammatikprüfungen, Systeme zur maschinellen Übersetzung, Tools zur Termextraktion, computergestützte Übersetzungsprogramme, Konkordanzsuchen, Suchmaschinen, Textanalyseprogramme etc.

Jede Benutzergruppe hat ihre eigenen Merkmale und spezifischen Anforderungen. Ist das TVS für mehr als eine Benutzergruppe bestimmt, sollte der gemeinsame Nenner bzw. das vorrangige Bedürfnis aller Benutzer ermittelt und gegenüber allen anderen Anforderungen priorisiert werden.

6.1.2 Die Entwickler sollten festlegen,

- ob die Terminologiesammlung normativ, präskriptiv oder deskriptiv sein soll;
- ob es sich um ein fachgebietsübergreifendes oder ein fachbezogenes TVS handeln soll. Bei fachbezogenen TVS muss zusätzlich entschieden werden, ob lediglich ein einzelnes oder mehrere Fachgebiete abgedeckt werden müssen und welche dies sein sollen.

6.1.3 Möglichkeiten zur Übermittlung von Informationen an Benutzer mit unterschiedlichen Anforderungen umfassen:

- verschiedene Anzeigoptionen der terminologischen Daten (Teilansichten; Felder, die erst bei Aufforderung angezeigt werden);
- Verfügbarkeit sämtlicher Informationen in allen Sprachen oder Verfügbarkeit einiger Informationen in nur einer Sprache;
- unterschiedliche Benutzeroberflächen für verschiedene Anwendungsfälle (Übersetzerarbeitsplatz, Unterstützung technischer Redakteure, Interaktion mit anderen computergestützten Anwendungen usw.);
- Zugang von lokaler (interner), nationaler und/oder internationaler Ebene aus;
- Webschnittstelle oder andere Art der Schnittstelle;
- freier öffentlicher Zugang oder auf eine bestimmte Umgebung beschränkter Zugang.

6.1.4 Die Entwickler müssen entscheiden, auf welche Art die Datenbank genutzt werden soll. Bestimmte Informationen sind für einige Benutzer von höherer Relevanz als für andere. So ist beispielsweise für technische Redakteure wichtig, welche Bedeutung ein Fachterminus hat, wie er richtig geschrieben wird (Großschreibung, Silbentrennung etc.) und welche Benennung bei mehreren Synonymen bevorzugt wird. Übersetzer sind hingegen in der Regel daran interessiert, für einen Ausgangsterminus das bevorzugte Äquivalent in einer bestimmten Zielsprache zu finden. Softwareanwendungen, die – wie oben genannt – mit dem TVS interagieren, sollten auf die von ihnen benötigten Informationstypen zugreifen können.

6.2 Schritte zur Umsetzung des benutzerzentrierten Ansatzes

Ein benutzerzentrierter Ansatz bei der Entwicklung eines TVS kann helfen, das System anwenderfreundlich zu machen und die gewünschten Anforderungen zu erfüllen.

Im Folgenden werden die entscheidenden Schritte zur Umsetzung erläutert.

- Ermittlung von Benutzergruppen und -bedürfnissen (siehe 6.3)
- Ermittlung der Endprodukte (siehe 6.4)
- Durchführung einer Aufgabenanalyse und Vorbereitung von Anwendungsfällen (siehe 6.5)
- Erfassung der Systemanforderungen (siehe 6.6)
- Durchführung einer Wettbewerbsanalyse (siehe 6.7)
- Entwicklung und Bewertung eines Design-Prototypen (siehe 6.8)
- Anpassen der Gestaltung an die Benutzerresonanz (siehe 6.9)
- Durchführung eines Beta-Tests (siehe 6.10)

6.3 Ermittlung von Benutzergruppen und -bedürfnissen

Um ein TVS entwickeln zu können, muss zunächst festgelegt werden, wer damit arbeiten wird und welche Benutzerbedürfnisse vorliegen. Potenzielle Anwender sollten daher befragt werden, ob und wie oft sie ein TVS nutzen würden, welche Aufgaben sie damit erledigen würden und welche spezifischen Bedürfnisse sie haben. Zusätzlich können die Entwickler die Resultate der Befragungen in Fokusgruppensitzungen bzw. Einzelgesprächen näher beleuchten. Das TVS sollte so konzipiert werden, dass die in den Umfragen ermittelten Benutzerbedürfnisse befriedigt werden. Dazu sollte in unterschiedlichen Entwicklungsstadien geprüft werden, wie erfolgreich die Anforderungen und Erwartungen der Benutzer umgesetzt wurden.

Bei den Befragungen sollte die Meinung der einzelnen Benutzergruppen entsprechend deren Nutzungshäufigkeit gewichtet werden. Wenn also beispielsweise davon ausgegangen wird, dass 1.000 Übersetzer und 100 Terminologen über das Internet auf das TVS zugreifen werden, dann sollte bei einer Umfrage zum Thema webbasierte Funktionen das Verhältnis von befragten Übersetzern und Terminologen 10 : 1 betragen. Sollten die Terminologen im Vergleich zu den Übersetzern jedoch doppelt so viel Zeit mit der Aktualisierung der Datensammlung verbringen, sollte bei Fragen zur Aktualisierungsfunktion das Verhältnis entsprechend korrigiert werden: $10 \cdot 1 : 1 \cdot 2 = 10 : 2$.

Insbesondere Benutzerbedürfnisse, die scheinbar im Gegensatz zueinander stehen, erfordern spezielle Aufmerksamkeit. Mitunter äußern verschiedene Benutzergruppen gegenläufige Bedürfnisse, die jeweils nur für die eigene Benutzergruppe relevant sind. Für Übersetzer können beispielsweise Kontextsätze (Beispielsätze, die die Benennung enthalten) hilfreich sein, für technische Redakteure haben sie jedoch keinen Nutzen. In solchen Fällen kann man beiden Bedürfnissen gerecht werden, indem benutzerspezifische Ansichten und/oder benutzerspezifische Workflows in das TVS integriert werden.

Onlinebefragungen bieten sich an, um Rückmeldungen von externen Benutzern, die nicht persönlich an den Fokusgruppensitzungen teilnehmen können, zu erhalten. Im Rahmen der Umfragevorbereitung sollte zunächst eine Probebefragung an einem kleinen Teil der Benutzergruppe durchgeführt werden, bevor sämtliche Benutzer befragt werden. Dies hilft sicherzustellen, dass die Fragen und deren Reihenfolge gute Ergebnisse liefern. Die Fokusgruppe der Probebefragung sollte Vertreter aller Benutzergruppen umfassen, sodass tendenziöse Schlussfolgerungen vermieden werden können. Zudem sollten Maßnahmen ergriffen werden, um die Motivation der Benutzer zur Teilnahme an der Umfrage zu sichern.

Wenn die Umfrageteilnehmer Vorschläge einbringen, erhoffen sie sich von deren Umsetzung bestimmte Ergebnisse. Fragen zur Auswertung dieser Ergebnisse sollten aufgenommen werden. So können die von den Umfrageteilnehmern geäußerten Bedürfnisse und ihre Sichtweise darauf bewertet werden. Mehrdeutige Fragestellungen sollten in der Umfrage vermieden werden.

Die Befragung dient nicht nur dazu, die Bedürfnisse der unterschiedlichen Benutzergruppen zu ermitteln, sondern soll auch Informationen zu den nachfolgend aufgeführten allgemeinen Rahmenbedingungen sammeln. Sowohl gegenwärtige als auch zukünftige Anforderungen müssen dabei berücksichtigt werden.

1. Welchem Zweck dient das TVS?

- Soll das TVS beispielsweise die Terminologie in einer oder mehreren Sprachen *standardisieren*, dann sollte die Datensammlung einen *normativen* Charakter haben. Auf diese Weise wird die Verwendung einheitlicher Terminologie gewährleistet. Bezüglich seiner Gestaltung und Methodik sollte das TVS in diesem Fall auf den ISO-Normen zur Standardisierung von Terminologie basieren. Dazu zählen z. B. ISO 704, ISO 1087 und ISO 10241-1. Soll das TVS hingegen die Verwendung von Terminologie *verbessern*, sollte die Datenbank *präskriptiv* organisiert werden. Dies bedeutet, dass Vorzugsbenennungen und nicht bevorzugte Terminologie kenntlich gemacht werden und ein Leitfaden zum Terminologiegebrauch bereitgestellt wird. Dient das TVS dem Zweck, den Gebrauch von Terminologie zu *beschreiben*, ohne dabei Vorgaben zur Verwendung zu machen, handelt es sich um eine *deskriptive* Datensammlung. Bei dieser Art der Terminologieverwaltung werden keine Angaben zu bevorzugten bzw. nicht-bevorzugten Benennungen gemacht.
- Nutzt ein Lokalisierungsdienstleister mit einem großen Kundenstamm die Datensammlung, so hat die Integrität der Kundenterminologie innerhalb des TVS oberste Priorität. Es sollten Datenfelder verwendet werden, mit denen der Ursprung einer Benennung, der Kunde sowie das Projekt, in dem die Benennungen verwendet werden, gekennzeichnet werden können. Eventuell sollten Sicherheitsvorkehrungen getroffen werden, um die vom Kunden zur Verfügung gestellten oder für ihn entwickelten Daten zu schützen. In einigen Fällen sollten für einzelne Kunden gesonderte Datensammlungen angelegt werden.

2. Sind alle Benutzer Mitarbeiter des Unternehmens/der Organisation oder gibt es auch externe Anwender?

- Diese Frage wirkt sich auf die Datenübertragung und den Datenzugang aus. Bei einem netzwerkbasierten TVS müssen beispielsweise besondere Maßnahmen ergriffen werden, um Benutzern außerhalb der Unternehmensfirewall den Zugang zum TVS zu ermöglichen. Sollen sowohl interne als auch externe Benutzer Zugriff auf sämtliche Daten erhalten? Wenn nicht, muss festgelegt werden, welche Daten extern weitergeleitet werden dürfen. Mittels eines speziellen Datenfelds kann beispielsweise eine Einteilung vorgenommen werden. Externe Gruppen, wie z. B. Kunden und Lieferanten, können möglicherweise nicht auf dieselben Terminologieverwaltungswerkzeuge zugreifen. Daher hat die Interoperabilität der Daten einen hohen Stellenwert und muss durch die Einhaltung internationaler Austauschnormen (vgl. ISO 12620, ISO 16642, und ISO 30042) gefördert werden.
- Für den Fall, dass mehrere Datenversionen und -kopien intern und extern vorliegen, sollten die Versionsverwaltung und die globale Aktualisierung der Daten in den Arbeitsablauf und die Zeitplanung integriert werden. Grundsätzlich empfiehlt es sich jedoch nicht, mehrere Datenversionen zu verwalten.

3. Sind alle Benutzer erfahrene Sprachdienstleister (Autoren, Übersetzer, Terminologen etc.) oder gehören sie auch anderen Arbeitsfeldern wie Vertrieb, Verwaltung, Kundenbetreuung, Ausbildung usw. an?
 - Verwenden mehrere Benutzergruppen ein TVS, kann es unter Umständen notwendig werden, einige Informationen nur bestimmten Benutzern anzuzeigen. Wenn viele Benutzer terminologische Einträge erstellen können, benötigen sie dafür eventuell auch unterschiedliche Benutzeroberflächen. Verwaltungspersonal wird in einem Eintrag weniger detaillierte Informationen erfassen wollen, als beispielsweise ein Terminologe; für technische Redakteure müssen Informationen anders dargestellt werden als für Übersetzer.
4. Muss ein anderes Programm – z. B. eine Desktop-Publishing-Software oder ein Translation-Memory-System – auf das TVS zugreifen oder mit ihm interagieren können?
 - In der Entwicklungsphase sollten alle Programme ermittelt werden, die mit den terminologischen Daten interagieren können. Zudem sollte geprüft werden, welche Art von Daten und welche Formate dazu erforderlich sind. Maschinelle Prozesse können Daten benötigen, die menschliche Benutzer üblicherweise nicht brauchen – z. B. die Wortart.
5. Welche Sprachen soll das TVS unterstützen?
 - Unicode ist laut ISO 10646 der empfohlene Zeichenkodierungsstandard, um die Interoperabilität mit anderen Systemen zu gewährleisten und ein möglichst großes Sprachenspektrum abdecken zu können. Sämtliche neuentwickelten TVS müssen Unicode-konform sein.
 - Ob UTF-8 oder höher verwendet wird, hängt sowohl von den geplanten Sprachen als auch vom vorliegenden Betriebssystem ab.
 - Soll das TVS auch bidirektionale Sprachen, z. B. Arabisch oder Hebräisch, verarbeiten, muss es bidirektionale Eingabemethoden und Anzeigeformate unterstützen.
6. Wie viele Sprachen sollte das TVS in Bezug auf den Inhalt bzw. die Benutzeroberfläche unterstützen?
 - Die Gestaltung von ein-, zwei- und mehrsprachigen terminologischen Datenbeständen unterscheidet sich erheblich.
 - Einsprachige Datenbanken werden vorrangig als Bezugsrahmen für Benutzer entwickelt, die in einer einzigen Sprache schreiben oder arbeiten. Zwei- und mehrsprachige Datenbanken richten sich hingegen in erster Linie an Übersetzer und Benutzer, die in mehreren Sprachen kommunizieren. In einsprachigen Datenbanken werden üblicherweise vorwiegend Informationen zur Verwendung einer Benennung erfasst; Definitionen haben tendenziell einen höheren Stellenwert. Zudem werden häufiger verwandte Benennungen und Synonyme aufgeführt. Im Gegensatz dazu ist in zwei- und mehrsprachigen Datenbanken eine Datenkategorie für Kontextsätze üblich, die in einsprachigen Systemen häufig nicht berücksichtigt wird. Zwei- und mehrsprachige Datenbanken verfügen natürlich auch über eine breite Palette an übersetzungsorientierten Datenkategorien: Sprachkennungen, Übersetzungskommentare, Äquivalenzgrad, Regionskennung, übersetzungsspezifische Anmerkungen usw.
 - Ist ein TVS für die ein-, zwei- und mehrsprachige Verwendung bestimmt, sollten unterschiedliche Ansichten entsprechend den einzelnen Benutzertypen und Aktivitäten erstellt werden. In jeder Ansicht können nicht gewünschte Felder verborgen und wichtige Felder hervorgehoben werden.

- Die Entwickler von zwei- bzw. mehrsprachigen TVS sollten außerdem berücksichtigen, ob es eine spezifische Ausgangssprache (Pivotsprache) gibt oder ob die Suchfunktion der Datenbank bi- bzw. multidirektional bezüglich Ausgangs- und Zielsprache sein soll. Gibt es beispielsweise eine festgelegte Ausgangssprache, so kann der Benutzer die Sprache, in der gesucht werden soll, nicht ändern. Die Benutzeroberfläche kann dementsprechend so gestaltet werden, dass die Angaben zur Ausgangssprache immer am oberen Bildschirmrand erscheinen. Bei einem zweisprachigen TVS, das die Suche in beiden Sprachen ermöglicht, sollte die Benutzeroberfläche so gestaltet werden, dass der Benutzer die Ausgangssprache auswählen kann. Die Suchergebnisse könnten sich dynamisch anpassen, sodass die gewählte Ausgangssprache an erster Stelle aufgeführt wird. In einem mehrsprachigen TVS mit Suchfunktionen in allen Sprachen sollte die Gestaltung der Benutzeroberfläche die Auswahl einer Ausgangssprache und einer oder mehrerer Zielsprachen ermöglichen. Wenn bereits im Voraus feststeht, dass die Mehrheit der Benutzer normalerweise eine bestimmte Ausgangssprache auswählen wird, so kann diese Sprache als Standard bei der Suche festgelegt werden, um so die Effizienz zu maximieren. Die Anwender sollten aber stets die Möglichkeit haben, die Sprache auf Wunsch zu verändern.
 - Bei einem mehrsprachigen begriffsorientierten TVS sollten die Entwickler prüfen, wie am sinnvollsten mit quasi-äquivalenten Begriffen verfahren wird: Eine Benennung A kann als Äquivalent zu einer Benennung B in einer Sprache und zu C in einer anderen Sprache betrachtet werden, auch wenn die Begriffe hinter B und C eventuell nicht vollständig äquivalent sind. Im Vorfeld muss daher eine Entscheidung zum Umgang mit unvollständiger Begriffsäquivalenz getroffen werden, da dies Einfluss auf die grundlegende Gestaltung des TVS haben kann. Zum einen bietet sich die Möglichkeit, die Teiläquivalente im selben Begriffseintrag zusammenzufassen und ein Feld für Übersetzungskommentare anzulegen, um die Unterschiede zu verdeutlichen. Zum anderen können für die Begriffe auch gesonderte terminologische Einträge erstellt werden, welche miteinander verlinkt und so als Übersetzungskandidaten gekennzeichnet werden.
7. Welche spezifischen Daten benötigen die einzelnen Benutzertypen?
- Anwender, die einsprachige terminologische Informationen, beispielsweise für regelbasiertes Schreiben, suchen, benötigen in der Regel Begriffsdefinitionen und Informationen zum Gebrauch. Übersetzer verwenden TVS, um zielsprachige Äquivalente zu finden; ihnen reichen meist Kontextsätze anstelle von Definitionen aus. Projektspezifische Kennungen werden nötig, wenn in unterschiedlichen Projekten verschiedene Äquivalente für eine Ausgangsbenennung verwendet werden. Bei großen Datensammlungen, die von unterschiedlichen Anwendern genutzt werden und mehreren Zwecken dienen, müssen eventuell spezielle Felder eingerichtet werden, um die Terminologie je nach Bedarf (z. B. Projekt, Kunde oder Produktkennung) zu kategorisieren. Diese Felder dienen der Unterteilung der Datensammlung und sind daher häufig für den Export von Datenteilmengen unentbehrlich.
8. Wie können die Benutzer Informationen finden und abrufen?
- Muss es für Benutzer möglich sein, Benennungen auch innerhalb anderer Softwareanwendungen, z. B. Textverarbeitungsprogrammen oder Übersetzungseditoren, zu suchen? Wird dazu das Internet oder das Intranet verwendet? Müssen die Anwender von einem mobilen Gerät auf die Daten zugreifen können? Benötigen sie erweiterte Suchfilter oder ist ein einzelner Suchbegriff ausreichend? Können solche Zugangsmedien Performanceprobleme auslösen, die berücksichtigt werden müssen?
 - Mit der Suchfunktion können auch Benennungen gefunden werden, die dem Suchbegriff ähneln (Fuzzy Match) oder ihm bis auf die Deklination entsprechen (Trunkierung). Ganze Benennungsgruppen lassen sich finden, indem bei der Suche nach spezifischen Datenkategorien gefiltert wird, die ihnen gemeinsam sind. Dazu gehören z. B. die Benennungsquelle, eine Produktkennung, eine bestimmte Wortart oder sogar das Datum, an dem der Eintrag erstellt oder zuletzt geändert wurde.

6.4 Ermittlung der Endprodukte

Basierend auf den Ergebnissen der Benutzerumfrage müssen die Entwickler festlegen, welche Endprodukte das TVS herstellen soll, um die Benutzerbedürfnisse zu befriedigen. Die Bandbreite reicht von traditionellen ein-, zwei- oder mehrsprachigen Glossaren, die beispielsweise nur Benennungen und Definitionen umfassen, bis hin zu komplexen maschinenlesbaren Ausgabeformaten, die äußerst granulare Daten erfordern (z. B. morphologische und syntaktische Angaben für maschinelle Übersetzungsprogramme). Einige Endprodukte wie Glossare und Übersetzungswörterbücher erhalten am besten eine benennungsorientierte Struktur, wohingegen z. B. Wörterbücher für das regelbasierte Schreiben sowie Suchmaschinen effektiver mit einer begriffsorientierten Struktur arbeiten. Endprodukte wie Rechtschreibprüfungen benötigen lediglich die Benennung; andere erfordern jedoch eine ganze Reihe terminologischer und lexikalischer Informationen.

Die verschiedenen hergestellten Produkte bedürfen auch unterschiedlicher Exportformate, die von einfachen kommagetrennten Listen bis hin zu detaillierten XML-Auszeichnungen reichen.

Um all diese Faktoren berücksichtigen zu können, müssen die Entwickler eine Bestandsaufnahme aller erwünschten terminologischen Leistungen sowie der dafür benötigten Formate und Datenkategorien durchführen.

6.5 Durchführung einer Aufgabenanalyse und Vorbereitung von Anwendungsfällen

Diese Phase umfasst die eingehende Analyse der Aufgaben, die die zukünftigen Anwender des TVS zurzeit bearbeiten. Daraus werden Anwendungsfälle entwickelt, die zeigen, wie ein Benutzer eine identische Aufgabe – und auch neue Aufgaben – mit dem neuen TVS bewältigt.

Am effektivsten ist die Aufgabenanalyse, wenn dazu eine Fokusgruppensitzung einberufen wird. Im Rahmen dieser Treffen werden die Benutzer gebeten, die mit Terminologie zusammenhängenden Aufgaben zu beschreiben, die sie regelmäßig ausführen, und zu erläutern, mit welchen Problemen sie dabei konfrontiert werden. Zusätzlich müssen die Entwickler Faktoren berücksichtigen, die die Art und Weise, wie Benutzer bestimmte Aufgaben bearbeiten, verändern könnten, und bedenken, welche Auswirkungen dies auf die Aufgaben hätte. Eventuell ist es sinnvoller, jeweils gesonderte Fokusgruppensitzungen für die einzelnen Anwenderzielgruppen zu veranstalten.

BEISPIEL 1 Beispielanalyse einer Aufgabe eines Übersetzers.

Aufgabe: Verwendung einer vorgeschriebenen Benennung in einer Übersetzung

Akteur: externer Übersetzer

Verwendetes Material: Tabelle mit projektspezifischer Terminologie

Verwendete Werkzeuge: Übersetzungseditor

Beschreibung: Übersetzer haben innerhalb des Übersetzungseditors zurzeit lediglich Zugriff auf zweisprachige Wörterbücher, die den Grundwortschatz umfassen. Projektspezifische Wörterbücher werden momentan von den Managern des Übersetzungsprojekts im Tabellenkalkulationsformat erstellt. Die Tabellen werden nicht als integrierte Projektwörterbücher in den Übersetzungseditor importiert, da die Projektmanager nicht über die nötige Zeit verfügen, die erforderlichen Pflichtfelder – Wortart und Projektkennung – hinzuzufügen, und die Datei in ein importierbares XML-Format zu konvertieren. Folglich nutzt der Übersetzer die Tabelle mit der Projektterminologie nur unregelmäßig, sodass eine terminologische Konsistenz nicht gewährleistet werden kann. Für Übersetzer besteht also Bedarf an einer automatisierten Anzeige standardisierter Projektbenennungen und womöglich auch an einer automatischen Überprüfung des Zieltextes auf Einhaltung der spezifisch festgelegten Terminologievorgaben.

Bei der Erstellung von Anwendungsfällen für das neue TVS müssen die Entwickler ermitteln, wer die Aufgabe bearbeitet, was die Grundvoraussetzungen sind, welche spezifischen Daten und Werkzeuge verwendet werden, welche Schritte nacheinander erfüllt werden und ob es eine spezielle Interaktion zwischen dem Benutzer und dem TVS gibt.

BEISPIEL 2 Vereinfachter Anwendungsfall, der demonstriert, wie ein Übersetzer die Aufgabe mit dem neuen TVS lösen könnte.

Aufgabe: Verwendung einer vorgeschriebenen Benennung in einer Übersetzung

Akteur: externer Übersetzer

Verwendete Werkzeuge: Übersetzungseditor, Projektwörterbuch

Voraussetzung: Projektwörterbuch ist im Übersetzungseditor verfügbar

Beschreibung:

1. Während des Übersetzungsvorgangs im Übersetzungseditor stößt ein Übersetzer auf eine Benennung in einem zu übersetzenden Satz, die im Projektwörterbuch enthalten ist (Benennungen im aktiven Übersetzungssegment, die im Projektwörterbuch vorkommen, sollten optisch hervorgehoben werden).
2. Der Übersetzer markiert die entsprechende Benennung.
3. Der Übersetzer klickt auf eine Schaltfläche auf der Benutzeroberfläche des Übersetzungseditors, sodass der terminologische Eintrag aus dem Projektwörterbuch angezeigt wird. Es sollte aber auch eine Aktivierung über einen Klick mit der rechten Maustaste oder eine Tastenkombination möglich sein. Das Projektwörterbuch öffnet sich und zeigt den Eintrag für die markierte Benennung an.
4. Der Übersetzer wählt aus der Liste der Optionen ein passendes Äquivalent aus.
5. Anschließend klickt er auf eine Schaltfläche – oder verwendet eine alternative Aktivierungsmethode – um das ausgewählte Äquivalent in den zu übersetzenden Satz einzufügen.
6. Um das Übersetzungswörterbuch zu schließen und zum Übersetzungseditor zurückzukehren, klickt der Übersetzer erneut auf eine Schaltfläche – oder verwendet eine alternative Aktivierungsmethode.

Verwendet man ein Werkzeug zur Aufgabenmodellierung, können zur Erfüllung der oben genannten Aufgabe zusätzliche Bedingungen und alternative Wege eingesetzt werden. Der Einsatz von visuellen Hilfsmitteln wie Flussdiagrammen ist auch denkbar. Welche Handlungsmöglichkeiten würden sich beispielsweise ergeben, wenn das Projektwörterbuch die markierte Benennung nicht enthalten würde? Was muss getan werden, wenn keine der angezeigten Übersetzungen für den Übersetzer brauchbar ist? Welche alternativen Möglichkeiten ergeben sich für einen internen Übersetzer, der eventuell Zugriff auf weitere Quellen hat? Was würde passieren, wenn die Benennung die Abkürzung einer anderen Benennung wäre oder wenn sie eine Abkürzung hätte? Sollte das Projektwörterbuch stets geöffnet sein und, falls ja, ist eine komprimierte Ansicht des Eintrags notwendig?

6.6 Anforderungen identifizieren und priorisieren

Basierend auf der Aufgabenanalyse und den Anwendungsfällen müssen die Entwickler eine Auflistung der Anforderungen an das TVS erstellen. Die grundlegenden Anforderungen können die folgenden Funktionen umfassen:

- Erstellen und Aktualisieren von terminologischen Einträgen direkt in der Systemschnittstelle,
- Vergleichen von Daten aus zwei oder mehr terminologischen Einträgen,
- Verwalten von Dubletten,
- Importieren und Exportieren von Daten in vordefinierte standardisierte Formate,
- Filtern von Daten anhand einer Vielzahl von Kriterien und
- Veröffentlichen von Daten im Web oder anderweitig.

In einigen Nutzungsumgebungen werden verschiedene Computersysteme für Aufgaben verwendet, deren Bearbeitung durch den Gebrauch von konsistenter Terminologie verbessert werden könnte – beispielsweise das Erstellen von Texten, das Content Management sowie Übersetzungen. Die Komponenten der TVS-Umgebung sollten daher mit den übrigen Systemen vollständig kompatibel sein, um größtmöglichen Nutzen aus den terminologischen Daten für das Unternehmen ziehen zu können. Das TVS kann beispielsweise einem Termextraktionswerkzeug Listen mit bekannten Benennungen zur Verfügung stellen, sodass die Anwendung nur „neue“ Termini effektiv erkennt. Zudem kann das TVS auch über eine automatische Nachschlagefunktion in einer Übersetzungsumgebung verfügen, um die terminologische Konsistenz von Übersetzungen zu erhöhen. Ein System zur Unterstützung des Lokalisierungsworkflows kann außerdem das TVS dynamisch durchsuchen, um relevante Terminologiewörterbücher für ein bestimmtes Lokalisierungsprojekt bereitzustellen. In einer Autorenumgebung kann das TVS helfen, unzulässige Benennungen zu vermeiden sowie Abkürzungen und Akronyme richtig zu verwenden. Wie bereits erwähnt, hat jeder dieser Anwendungsbereiche unterschiedliche Anforderungen an die Terminologieverwaltung, die sorgfältig ermittelt werden sollten.

Sobald die potenziellen Anforderungen feststehen, sollten die Fokusgruppen um eine Einordnung der Anforderungen nach deren Priorität – von der höchsten zur geringsten Bedeutung – gebeten werden. Die Anforderungen, die sich als zwingend notwendig herauskristallisieren, müssen eindeutig festgelegt werden und sind bei der Entwicklung des TVS unverzichtbar.

Die Priorisierung der Anforderungen erfolgt im ersten Schritt standardgemäß anhand der folgenden drei Kategorien: (1) wesentlich, (2) wichtig und (3) optional.

Im zweiten Schritt wird jeder Anforderung ein Wert zwischen 1 und 5 zugeteilt, wobei in diesem Stadium noch vernachlässigt wird, ob die Anforderung als wesentlich, wichtig oder optional eingestuft wurde. Mithilfe dieses Ansatzes lässt sich ein Prioritätenraster anlegen, das die Priorität der einzelnen Anforderungen darstellt.

In manchen Fällen gibt es noch eine weitere Priorisierungskategorie, die die Unternehmensstrategie oder andere strategische oder gesetzliche Anforderungen widerspiegelt. Beispielsweise kann die Unternehmenspolitik die Förderung bestimmter Zugänglichkeitsanforderungen für sensorisch behinderte Anwender enthalten, die unabhängig von den gegenwärtigen Bedürfnissen der Benutzergruppe verpflichtend ist.

Anhand des in 6.1.3 als Beispiel angeführten Anwendungsfalles ist es möglich, die folgenden Gestaltungsanforderungen zu ermitteln. (Die nachfolgende Liste dient lediglich der Veranschaulichung und erhebt keinen Anspruch auf Vollständigkeit.)

- a) Der Übersetzungseditor und das Projektwörterbuch sollten in einer externen Nutzungsumgebung verwendbar sein.
- b) Terminologische Einträge sollten über Projektkennungen verfügen, die vom Übersetzungseditor erkannt werden können.
- c) Es sollte möglich sein, terminologische Einträge in Form eines für den Übersetzungseditor bestimmten Projektwörterbuchs aus der Datensammlung zu exportieren, oder es sollte eine direkte Verbindung zum Übersetzungseditor vorhanden sein. Die Projektkennung sollte daher als Exportfilter oder, bei integrierten Nutzungsumgebungen, als Anzeigefilter verwendet werden können. Außerdem sollte das TVS die notwendigen Pflichtfelder für das Wörterbuch im Übersetzungseditor enthalten.
- d) Es sollte ein direkter Zugriff vom Übersetzungseditor auf das Projektwörterbuch möglich sein.
- e) Der Übersetzungseditor sollte im Projektwörterbuch vorhandene Benennungen automatisch hervorheben.
- f) Der Übersetzungseditor sollte hervorgehobene Benennungen im Ausgangssatz im Projektwörterbuch suchen.

- g) Im Projektwörterbuch sollten Datenfelder sichtbar sein, die dem Übersetzer bei Einträgen mit mehreren möglichen Übersetzungen oder verschiedenen Sätzen die Auswahl erleichtern.
- h) Der Übersetzungseditor sollte so programmiert sein, dass er die ausgewählte Übersetzung in den Zielsatz einfügt. (Diese Abhängigkeit von einem externen Werkzeug sollte erkannt und gemeinsam mit den Entwicklern dieses Programms gelöst werden.)

6.7 Durchführung einer Wettbewerbsanalyse

Um sicherzustellen, dass die ermittelten Aufgaben und Anforderungen nicht zu sehr auf interne Prozesse ausgerichtet sind, sondern auch Terminologieverwaltungspraktiken außerhalb des Unternehmens berücksichtigen, sollte das Entwicklerteam analysieren, wie vergleichbare Unternehmen ähnliche Terminologieaufgaben handhaben. Die Entwickler sollten andere TVS bewerten, um so neue Ideen zur Umsetzung von Funktionen zu erhalten. Bevor mit der Entwicklung eines neuen TVS begonnen wird, sollte das Team untersuchen, ob bereits ein kommerzielles TVS erhältlich ist, das die gewünschten Eigenschaften aufweist. Ist dies der Fall, so muss ermittelt werden, ob der Erwerb bzw. die Anpassung des TVS an die Kundenwünsche eine kostengünstigere Alternative im Vergleich zur Neuentwicklung darstellt.

6.8 Entwicklung und Bewertung eines Prototypen

Mithilfe der in den vorangegangenen Phasen gesammelten Daten müssen die Entwickler einen Prototyp entwickeln, der die Gestaltung und die Funktionen des zukünftigen TVS widerspiegelt. In einer weiteren persönlichen Fokusgruppensitzung sollten die Benutzer den Prototyp anhand seiner Gestaltung bewerten und ihn mit Werkzeugen, mit denen sie momentan arbeiten, vergleichen. Während der gesamten Produktentwicklung sollten Evaluierungen zur Prototypengestaltung durchgeführt werden. Die Benutzerrückmeldungen sollten in Form von Umfragen oder Kommentarformularen gesammelt werden. Diese Daten können dann mit Statistiken aus früheren oder späteren Phasen verglichen werden, um die Gestaltung zu validieren.

Im Entwicklungszeitplan müssen Meilensteine festgelegt werden, an denen das TVS im Entwicklungsstadium mit den Gestaltungsvorgaben verglichen wird. So kann sichergestellt werden, dass die Funktionen gemäß den ursprünglichen Benutzervorgaben entwickelt werden.

6.9 Anpassung der Gestaltung an die Benutzerresonanz

Anhand der Rückmeldungen aus den Evaluierungssitzungen bezüglich der Prototypen müssen die Entwickler die Gestaltung anpassen und so Schwachstellen beseitigen sowie zusätzlich gewünschte Funktionen hinzufügen. Die Feedbacksitzungen sollten nach Bedarf wiederholt werden, bis die Benutzer mit dem Prototypen zufrieden sind.

6.10 Durchführung eines Beta-Tests

Vor dem endgültigen Einsatz des Produkts sollte es von Benutzern getestet werden, um so festzustellen, ob deren Erwartungen erfüllt werden. Gegebenenfalls können dann noch einige abschließende Korrekturen vorgenommen werden, wobei in dieser späten Entwicklungsphase die zentrale Gestaltung nicht mehr wesentlich verändert werden kann. Zur Planung zukünftiger Verbesserungen des Systems ist es auch nützlich, die wesentlichen quantitativen (z. B. die für die Aufgabe aufgewendete Zeit, die Häufigkeit der Inanspruchnahme von Hilfe sowie die Anzahl aufgetretener Fehler) und qualitativen Bewertungskriterien (z. B. Benutzerzufriedenheit) zu untersuchen.

7 Terminologische Datenkategorien

7.1 Einführung in das Thema Datenkategorien

Terminologische Einträge bestehen aus spezifischen Informationseinheiten – den sogenannten *Datenkategorien* – zu denen beispielsweise *Benennungen*, *Definitionen* und *Kontexte* zählen. ISO 12620 definiert die Rahmenbedingungen, anhand derer Datenkategorien festgelegt werden, die der Erfassung von Informationen zu Terminologie und anderen Sprach- und Inhaltsressourcen dienen (z. B. terminologische Datenbestände, lexikografische oder elektronisch verfügbare lexikalische Ressourcen etc.). Die Auswahl der Datenkategorien (DCS) muss von der Thematic Domain Group für Terminologie, die unter Aufsicht von TC 37/SC 3/WG 1 tätig ist, bestimmt werden. In der DCS werden die Datenkategorien aufgelistet, die sich zur Dokumentation von terminologischen Informationen in Computerumgebungen eignen und den Austausch und die Abfrage terminologischer Daten unabhängig von den lokalen Software- oder Hardwareumgebungen, in denen diese Datenkategorien genutzt werden, ermöglichen. Die DCS beinhaltet eine Teilmenge einer umfassenden Datenkategoriesammlung, welche im Verzeichnis der Datenkategorien (DCR, siehe <http://www.isocat.org>) aufgeführt ist und zur Verwendung in linguistischen Ressourcen festgelegt wurde.

Die DCS umfasst mehr als 200 Datenkategorien zur Einteilung terminologischer Daten, die als Feldnamen in Einträgen verwendet werden können. Hinzu kommen weitere Datenkategorien, die als Feldwerte dienen. Die Datenkategorie */grammaticalGender/* (Genus) z. B. wird üblicherweise als Feldname genutzt; die Datenkategorien *masculine* (Maskulinum), *feminine* (Femininum) und *neuter* (Neutrum) stellen hingegen normalerweise Werte dieses Feldes dar.

Die DCS ist umfangreich, da in ihr möglichst alle Datenkategorien aus einem breiten Spektrum an bestehenden terminologischen Datenbeständen erfasst werden sollen. Üblicherweise wird aber in einer Termbank lediglich eine kleine Teilmenge der terminologischen Datenkategorien aus der DCS verwendet. Welche Datenkategorien für eine Terminologiedatenbank ausgewählt werden, hängt von den Vorgaben, Zielen und Benutzerbedürfnissen des jeweiligen Terminologieprojekts ab. Gewöhnlich ist in den meisten Datenbanken eine geringe Anzahl Datenkategorien obligatorisch, während weitere je nach Bedarf hinzugefügt werden können.

In ISO 10241-1 und ISO 12616 werden die folgenden Datenkategorien aufgeführt:

- Benennungen (in jeder beliebigen Sprache);
- benennungsbezogene Datenkategorien (z. B. grammatische Angaben, Register, Status):
 - Klassifizierung von Benennungstypen (z. B. Schreibvarianten, Vollformen, Abkürzungen usw.);
- deskriptive begriffsbezogene Informationen (z. B. Fachgebiet, Definitionen, Kontexte, Beispiele, Anmerkungen und Grafiken);
- administrative und bibliografische Informationen:
 - Kennungen unterschiedlicher Art (z. B. zur Identifikation von Produkten oder Projekten, auf die sich die Benennungen beziehen);
 - Erstellungs- und Änderungsdatum sowie Namen der Autoren, die den Eintrag bzw. Teile davon erstellt oder geändert haben;
 - Eintragsstatus (z. B. vorgeschlagen, in Bearbeitung, konsolidiert);
 - Quellen von Benennungen, Definitionen, Kontexten, Anmerkungen etc.

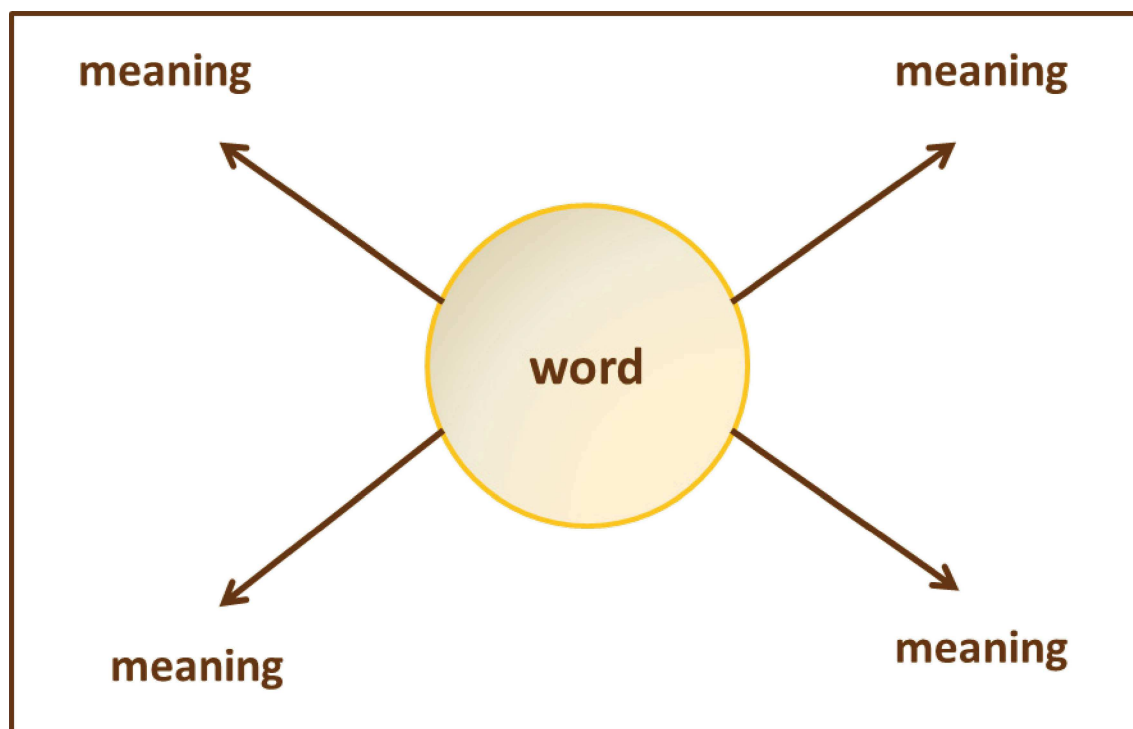
7.2 Grundsätze zur Auswahl und Verwendung von Datenkategorien

7.2.1 Begriffsorientierung

Terminologische und lexikografische Einträge unterscheiden sich in ihrer Struktur dadurch, dass jeder terminologische Eintrag Informationen zu einem einzelnen Begriff (oder manchmal, speziell in mehrsprachigen Datensammlungen, zu quasi-äquivalenten Begriffen) enthält und sämtliche zur Bezeichnung des Begriffs verwendeten Benennungen in beliebig vielen Sprachen aufführt.

Es kann sinnvoll sein, neben der herkömmlichen begriffsorientierten Ansicht als Variante eine lexikalische, benennungsorientierte Struktur anzubieten. Für beide Ansichtsmodi werden weitgehend die gleichen Datenkategorien verwendet, aber auf dem Bildschirm unterschiedlich angeordnet. Die benennungsorientierte Ansicht umfasst sämtliche Bedeutungen eines Wortes und eignet sich daher besonders, um alle Bedeutungen eines Polysems zu erfassen. Im Gegenzug zeigt die terminologische Ansicht sämtliche Benennungen, die einen einzelnen Begriff oder eine Bedeutung bezeichnen, und erleichtert so die Suche nach Synonymen und Übersetzungsäquivalenten.

Die Bilder 1 und 2 veranschaulichen die wesentlichen Unterschiede und Zusammenhänge zwischen einer begriffsorientierten und einer benennungsorientierten Ansicht.

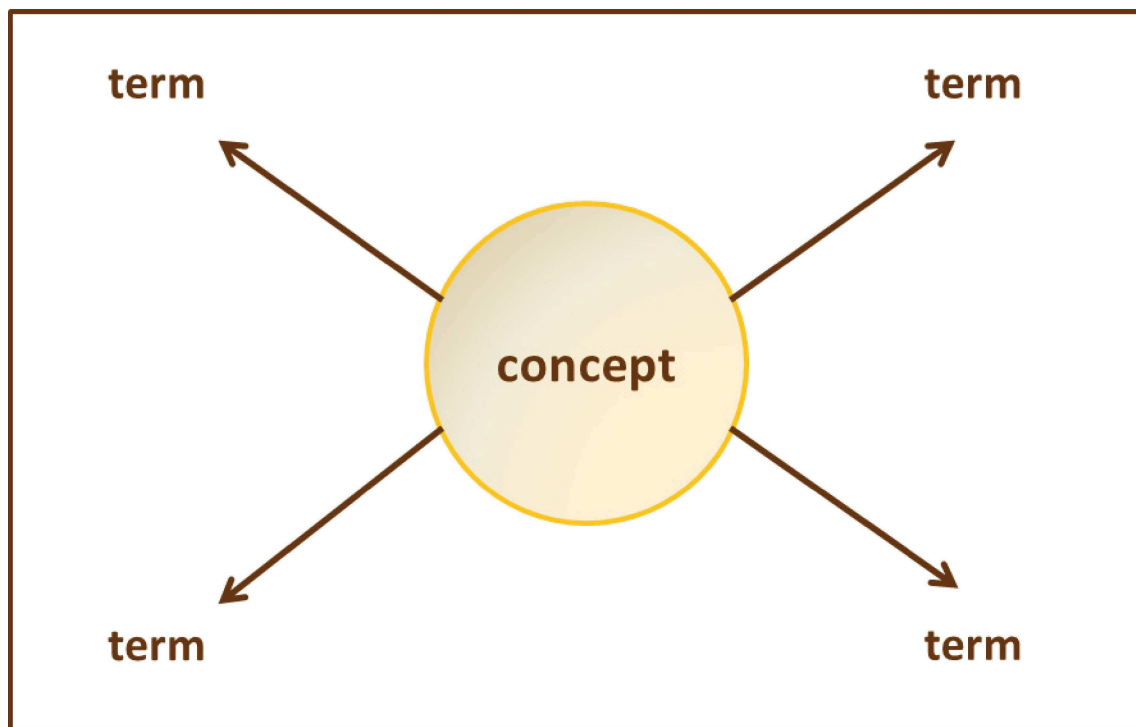


Legende

Word = Wort

Meaning = Bedeutung

Bild 1 — Benennungsorientierte Struktur

**Legende**

Concept = Begriff

Term = Benennung

Bild 2 — Begriffsortorientierte Struktur**7.2.2 Kontextuelle und semantische Datenkategorien**

Häufig ist es schwierig, die Bedeutung einer Benennung lediglich anhand des Wortes selbst zu erfassen. Im Internet veröffentlichte zwei- oder mehrsprachige Glossare, die lediglich Äquivalente – ohne Definitionen, Kontexte und meist auch ohne Übersetzungskommentar oder Fachgebiet – aufführen, sind aus diesem Grund oft nicht verlässlich. Mit so wenigen Zusatzinformationen können Anwender nicht sicherstellen, ob eine zielsprachliche Benennung in einem bestimmten Kontext angemessen ist. Ein gut strukturiertes TVS verfügt daher über Datenkategorien, mit denen zur eindeutigen Bedeutungsbestimmung von Benennungen verschiedenste Informationsarten detailliert gespeichert werden können. Hilfreich sind beispielsweise u. a. Definitionen, Fachgebietsangaben, Kontextsätze und Wortartwerte. Mitunter benötigen Benutzer jedoch tatsächlich einfache Auflistungen von Ausgangssprachlichen und Zielsprachlichen Äquivalenten. Solche Listen sollten jedoch zumindest eine eindeutige semantische Einordnung, z. B. in Form einer Fachgebietsangabe, enthalten, um so zu verdeutlichen, in welchen Kontexten die Benennungen relevant sind. Um ein solches terminologisches Produkt exportieren zu können, sollte ein TVS einen Exportfilter auf eine semantische Datenkategorie (z. B. */subjectField/* (Fachgebiet) o. ä.) anwenden können.

7.2.3 Benennungsautonomie

Nach dem Prinzip der Benennungsautonomie sollte es möglich sein, alle Benennungen in einem terminologischen Eintrag gleichermaßen mit Informationen und Beschreibungen auszustatten. Benennungsautonomie bedeutet, dass sämtliche Synonyme und Äquivalente eines Begriffs mit einem vollständigen Satz von Datenkategorien dokumentiert werden, selbst wenn eine Vorzugsbenennung vorhanden ist. Am besten lässt sich dieser Grundsatz gewährleisten, indem für jede Benennung, die im terminologischen Eintrag aufgeführt werden soll, ein eigenes Datenfeld */term/* (Benennung) vorhanden ist. Mit der Datenkategorie */termType/* (Benennungsart) wird dann angegeben, ob es sich bei der Benennung um eine Vollform, eine Abkürzung, ein Akronym o. ä. handelt. Die Datenkategorie */administrativeStatus/* (administrativer Status) gibt Aufschluss darüber, ob eine Benennung beispielsweise bevorzugt oder

abgelehnt wird. Diverse zusätzliche Datenkategorien können einzelne Benennungen innerhalb eines Eintrags hervorheben und dem Benutzer die Auswahl der passenden Benennung für einen bestimmten Kontext erleichtern.

Werden die Prinzipien der Benennungsautonomie und der Begriffsorientierung erfüllt, so ist eine Datenkategorie */synonym/* (Synonym) nicht mehr erforderlich. Alle Benennungen, die in einem Begriffseintrag dokumentiert werden, sind Synonyme und haben denselben Status als Benennung.

7.2.4 Wiederholbarkeit

Um Informationen zu einem Begriff oder einer Benennung im Begriffseintrag zu erfassen, können Datenkategorien theoretisch beliebig oft wiederholt und mit anderen Datenkategorien verknüpft werden. Beispielsweise sollte die Wiederholbarkeit der Datenkategorie */term/* auf der Sprachebene unbegrenzt sein, um sämtliche Synonyme und Äquivalente in einer bestimmten Sprache dokumentieren zu können. So kann die Datenkategorie */term/* beispielsweise mit einer Vielzahl benennungsbezogener Datenkategorien zusammenhängen, welche die Eigenschaften der Benennung beschreiben. Zur Angabe der Informationsquelle können Text enthaltende Datenkategorien (z. B. */definition/* (Definition); */context/* (Kontext) und */note/* (Anmerkung)) durch die Datenkategorie */source/* (Quelle) ergänzt werden. In dem mehrsprachigen Beispieleintrag in Anhang A wird */term/* gemeinsam mit allen zugehörigen Datenkategorien für jede Benennung in jeder Sprache (Japanisch, Englisch und Deutsch) wiederholt. Im Beispiel wird sie durch */partOfSpeech/* (Wortart), */grammaticalGender/* (Genus; nur bei der deutschen Benennung) und */termType/* ergänzt.

Andererseits kann, je nach Art und Verwendungszweck des TVS, festgelegt werden, dass manche Datenkategorien nicht innerhalb eines Eintrags oder eines Eintragsteils wiederholt werden dürfen. In einem präskriptiven TVS kann die Wiederholbarkeit der Datenkategorie */definition/* beispielsweise auf eine Wiederholung pro Eintrag bzw. eine Wiederholung pro Sprachebene beschränkt sein. In stark deskriptiven TVS kann */definition/* hingegen wiederholt werden. Häufig ist es außerdem vorgesehen, dass in einem bestimmten Begriffseintrag lediglich ein Fachgebiet und ein Wortartwert verwendet werden können. Insgesamt sollte bei der Gestaltung des TVS darauf geachtet werden, die Wiederholbarkeit der Daten auf deren Verwendungszweck sowie die Benutzerbedürfnisse abzustimmen.

7.2.5 Datengranularität

Terminologische Einträge sollten Datenkategorien umfassen, mit denen ein angemessener Grad an Datengranularität erreicht werden kann. Grammatikalische Informationen sollten etwa, wie im folgenden Beispiel, in verschiedenen spezifischen Datenkategorien erfasst werden:

/partOfSpeech/ – Nomen

/grammaticalGender/ – Femininum

/grammaticalNumber/ (Numerus) – Singular

Nicht empfehlenswert ist die im folgenden Beispiel dargestellte Verknüpfung mehrerer Eigenschaften in einer allgemeinen Datenkategorie:

/grammar/ – Nomen, Femininum, Singular

TVS mit mehreren spezifischen Datenkategorien verfügen über eine höhere Granularität als solche mit nur wenigen allgemeinen Datenkategorien. Eine mangelhafte Einteilung der Informationen in Unterkategorien kann dazu führen, dass Daten nur noch schwer abgerufen, manipuliert und ausgetauscht werden können.

7.2.6 Datenelementarität

Das Prinzip der Elementarität von Datenkategorien besagt, dass ein Datenfeld nur eine einzige Information enthalten darf. In dem in 7.2.5 aufgeführten Beispiel enthält die Datenkategorie */grammar/* drei verschiedene Datentypen (Wortart, Genus und Numerus) und verstößt demnach gegen das Prinzip der Elementarität. Gleichmaßen stellt es eine Verletzung der Elementaritätsforderung dar, wenn eine Benennung und deren Abkürzung gemeinsam in einem Datenfeld */term/* aufgeführt werden, wie z. B. bei „Access Control List (ACL)“. Eine solche Vorgehensweise findet sich häufig in veröffentlichten Glossaren, sollte aber in einer Termbank vermieden werden. Bei Nichtbeachtung des Elementaritätsprinzips gestaltet sich die automatische Verarbeitung der meisten Daten als schwierig bis unmöglich. So wäre es beispielsweise ausgeschlossen, eine Datenbank effektiv nach Akronymen zu durchsuchen oder diese zu extrahieren, wenn sich die Akronyme im selben Feld wie deren Vollformen befinden. Gleichmaßen ist das Auftreten von zwei oder mehr Benennungen im selben Datenfeld problematisch für bestimmte CAT-Tools, die Benennungen aus terminologischen Einträgen automatisch in die zu übersetzenden Dokumente einfügen.

Um Entscheidungen bezüglich der Granularität und Elementarität treffen zu können, müssen die daraus resultierenden Vor- und Nachteile sorgfältig abgewogen werden. Positiv ist die verbesserte Informationswiedergewinnung durch das zur Verfügungstellen von Informationen mit einer höheren Granularität, negativ fällt hingegen der oftmals höhere Aufwand durch die Eingabe von Informationen in verschiedene Felder ins Gewicht. In der Entwicklungsphase sollte daher mit besonderer Umsicht ermittelt werden, welche Informationsarten die Benutzer erfassen wollen. Wird beispielsweise versäumt, die passenden Dateneintragsfelder für einzelne Informationen anzulegen, so kann dies zur Nutzung von ungeeigneten Feldern oder zur Eingabe mehrerer unterschiedlicher Informationen in ein einziges Feld führen. Werden jedoch mehr Felder bereitgestellt als tatsächlich benötigt werden, kann dies die Benutzerfreundlichkeit des TVS nachteilig beeinflussen und die Produktivität bei der Dateneingabe verringern.

7.3 Arten von Datenkategorien

7.3.1 Offene, geschlossene und einfache Datenkategorien

Offene Datenkategorien können jeglichen Text enthalten, der ihrer Datenkategoriedefinition entspricht. Die Datenkategorie */term/* beispielsweise sollte nur eine Benennung enthalten, aber keine Definition und keinen Kontext. Sie zählt demzufolge zu den offenen Datenkategorien, da die Benennungen, die erfasst werden können, frei und nicht vorhersehbar sind. Weitere Beispiele für offene Datenkategorien sind */definition/* und */context/*.

Im Gegensatz dazu gibt es andere Datenkategorien, die lediglich einen oder mehrere Werte aus einer begrenzten Anzahl zulässiger Werte – auch zulässige Instanzen genannt – enthalten können. Sie werden als geschlossene Datenkategorien bezeichnet, da ihr Inhalt aus einer geschlossenen Wertemenge besteht. Werden beispielsweise deutschsprachige Benennungen dokumentiert, so kann die Datenkategorie */grammaticalGender/* nur die Werte */masculine/*, */feminine/* oder */neuter/* annehmen. Diese zulässigen Werte bilden den Wertebereich für die Datenkategorie in der deutschen Sprache. Im Französischen hingegen sind nur die Werte */masculine/* und */feminine/* zulässig.

Solche in bestimmten Wertebereichen zulässigen Werte werden als einfache Datenkategorien bezeichnet. Zum Beispiel stellt */feminine/* eine einfache Datenkategorie dar, die als Wert in der geschlossenen Datenkategorie */grammaticalGender/* verwendet werden kann. Einfache Datenkategorien werden innerhalb der jeweiligen Datenkategoriespezifikation für die entsprechenden übergeordneten Datenkategorien erfasst und definiert. Sie werden als einfach bezeichnet, da sie selbst über keinen Inhalt verfügen können. Bei der Planung eines terminologischen Datenbestandes ist es wichtig, die zulässigen Werte, die den geschlossenen Datenkategorien konkret zugeordnet werden, auszuwählen und sie als wählbare Werte (auch Picklistenwerte genannt) im TVS verfügbar zu machen. Ein TVS zur Dokumentation französischer Benennungen beispielsweise benötigt nicht die Datenkategorie */neuter/* als Wert für das Genus, da die französische Sprache über kein sächliches Geschlecht verfügt. Der Gebrauch von Picklisten erleichtert dem Benutzer die Auswahl des entsprechenden Wertes, ohne dass dieser eingegeben werden muss. Außerdem

verhindern Picklisten das Einbringen von Schreibfehlern oder Abwandlungen der Werte in die terminologische Datensammlung. Steht es den Benutzern beispielsweise frei, wie ein männliches Nomen gekennzeichnet werden soll, könnte *masculine*, *masc.* oder einfach *m.* eingetragen werden. Eine einheitliche Darstellung dieser Werte gewährleistet aber die Konsistenz der terminologischen Datensammlung. Diese ist wichtig, um Suchfilter einsetzen und sonstige mit der Datenverwaltung verbundenen Aufgaben erfüllen zu können.

7.3.2 Obligatorische, fakultative, automatische und voreingestellte Datenkategorien

Datenkategorien in einem TVS können entweder obligatorisch oder fakultativ sein. Wurde eines der obligatorischen Felder nicht ausgefüllt, kann ein terminologischer Eintrag in der Regel nicht gespeichert werden. Beispielsweise bedarf jeder terminologische Eintrag mindestens einer Benennung.

Zu den obligatorischen Feldern gehören üblicherweise Fachgebietsangaben oder Wortartwerte, die beide wesentlich für die Unterscheidung von Homografen sein können. Bestimmte Felder als obligatorisch zu kennzeichnen, kann allerdings auch zu Problemen führen. Die Suche nach Definitionen und Kontexten, deren Erörterung und anschließende Aufnahme in das TVS beispielsweise stellen einen erheblichen Zeit- und Arbeitsaufwand dar. Aus diesem Grund ist es möglicherweise produktiver, ein nachträgliches Hinzufügen dieser Informationen zuzulassen.

Der Inhalt einiger Datenkategorien wird nicht von den Benutzern per Hand eingegeben, sondern vom System generiert. Viele TVS weisen den terminologischen Einträgen beispielsweise automatisch Eintragsnummer, Erstellungs- und Änderungsdatum sowie die Namen der Benutzer, die den Eintrag erstellt oder geändert haben, zu. Zum Entwicklungsprozess eines TVS gehört es daher auch, zu entscheiden, welche Eintragsebenen anhand solcher administrativer Informationen dokumentiert werden sollen. Gewöhnlich reicht es nicht aus, die Verwaltungsinformationen lediglich für den Eintrag als Ganzes zu erfassen, da verschiedene Mitarbeiter für unterschiedliche Eintragsteile – und insbesondere für verschiedene Sprachblöcke – zuständig sein können.

Zudem ist es möglich und sogar ratsam, für bestimmte Datenkategorien einen Standardwert festzusetzen. Möchte ein Benutzer etwa während einer Arbeitssitzung ausschließlich Benennungen dokumentieren, die mit einem bestimmten Projekt oder Ausgangstext zusammenhängen, so ist es vorteilhaft, die Werte für die Datenkategorien */source/* oder */projectSubset/* voreinstellen zu können. Sämtliche in der entsprechenden Arbeitssitzung erstellten terminologischen Einträge erhalten dann automatisch diese Werte. Ähnlich verhält es sich, wenn bereits feststeht, dass es sich bei den Benennungen in einer terminologischen Datensammlung hauptsächlich um Substantive handeln wird: Wenn für die Datenkategorie */partOfSpeech/* bereits Substantiv als Wert voreingestellt ist und dies nur im Bedarfsfall – also bei Nicht-Substantiven – geändert werden muss, bedeutet dies eine Zeitersparnis für den Benutzer.

7.3.3 Datenkategorien mit Nur-Lesezugriff, Lese- und Schreibzugriff sowie versteckte Datenkategorien

In einem TVS sollte es möglich sein, auf Datenfeldebene unterschiedliche Zugriffsrechte für Benutzer festzulegen, welche von den jeweiligen Bedürfnissen und Rollen der Benutzer abhängen sollten. Ein sichtbares Datenfeld, das nicht verändert werden kann, ist ein Datenfeld mit „Nur-Lesezugriff“. Ein sichtbares Datenfeld, das verändert werden kann, ist ein Datenfeld mit Lese- und Schreibzugriff. Nicht von allen Benutzern benötigte Datenfelder können versteckt werden.

So ist es beispielsweise empfehlenswert, einem leitenden Terminologen Lese- und Schreibzugriff auf sämtliche Felder zu gewähren. In seinen Zuständigkeitsbereich fällt es zudem häufig, anderen Benutzern Berechtigungsebenen zuzuteilen. Terminologen, die Einträge für bestimmte Sprachen erstellen, sollten über Lese- und Schreibzugriff auf die Datenfelder in ihrer Sprachebene verfügen. Lese- und Schreibzugriff auf Datenfelder zum Eintragsstatus und damit die Befugnis, den Wert auf „konsolidiert“ setzen zu können, sollten hingegen nur ausgewählte leitende Terminologen erhalten. Technischen Redakteuren und Übersetzern wird in der Regel nur auf einzelne Datenfelder Lese- und Schreibzugriff gewährt, um über diese den Terminologen Rückmeldungen geben zu können. Auf die übrigen Datenfelder haben sie meist lediglich Nur-Lesezugriff. Vielfach ist es sinnvoll, administrative Datenkategorien wie */date/* und */responsibility/* zu verbergen, sodass die Mehrheit der Benutzer an einer übersichtlicheren Oberfläche arbeiten kann.

7.3.4 Multimediateien

Die Möglichkeit, Multimediateien (z. B. Grafiken) direkt in terminologische Einträge zu Begriffen einzubinden, die sich auf konkrete Objekte – beispielsweise aus dem Gebiet der Wissenschaft oder Technik – beziehen, kann von Nutzen sein. Die Datenfelder müssen dann entsprechend so eingerichtet werden, dass sie die Speicherung und Anzeige binärer Objekte unterstützen. Das Datenfeld für Grafiken sollte auf der Begriffsebene verfügbar sein, damit die Grafik sprachunabhängig den gesamten Begriff repräsentiert. Allerdings kann auch auf der Sprachebene ein Datenfeld für Grafiken notwendig sein, falls die dem Begriff zugeordneten Objekte in verschiedenen Kulturen im Erscheinungsbild variieren. Wird dieselbe Multimediatei in mehreren terminologischen Einträgen verwendet, so sollte sie aus Effizienzgründen als gemeinsame Ressource gehandhabt werden.

7.3.5 Gemeinsame Ressourcen

In einigen Datenkategorien finden sich Verweise auf andere Ressourcen, z. B. auf andere Dateien (Texte, Grafiken, Audio- oder Videodateien) oder Websites. Häufig werden sie als gemeinsame Ressourcen bezeichnet, da sich in vielen terminologischen Einträgen Verweise auf dasselbe Element befinden können – es wird praktisch von verschiedenen Einträgen gemeinsam genutzt.

Eine weitere Form der gemeinsamen Ressource stellen bibliografische Einträge dar. Üblicherweise wird die Quelle von Benennungen, Definitionen, Kontexten, Anmerkungen, zusätzlichen Informationen und sämtlichen externen Ressourcen (Grafiken, Audio- und Videodateien etc.) dokumentiert. Anstatt aber die gesamte Bibliografieangabe im Eintrag selbst anzugeben, ist es wesentlich effizienter, einen separaten bibliografischen Eintrag zu erstellen und in den Einträgen auf diesen mittels einer bibliografischen Kennung zu verweisen. Dies gilt besonders dann, wenn dieselbe Quellenangabe in mehreren Einträgen verwendet wird. Durch diese Vorgehensweise wird der Grundsatz der Datensparsamkeit gewahrt, da die gleichen Angaben nicht in mehreren terminologischen Einträgen wiederholt werden müssen. Da so jeder bibliografische Eintrag nur noch einmal in der Datensammlung vorkommt, wird außerdem das Aktualisieren von Bibliografieangaben, die in mehreren Einträgen verwendet werden, erleichtert. Eine Ausnahme bildet hier das Zitieren von URLs und URIs, da sie auf Webressourcen verweisen und somit bereits kurze, spezifische Kennungen sind.

Gemeinsame Ressourcen können in einer externen Datenbank oder an einem gesonderten Speicherort in der Termbank selbst gespeichert werden. Den Benutzern sollte beim Durchblättern und Durchsuchen der Terminologiedatenbank der Zugriff auf die gemeinsamen Ressourcen ermöglicht werden, da die enthaltenen Informationen oft ausschlaggebend für die Bewertung eines terminologischen Eintrags sind.

7.3.6 Beziehungen zwischen Einträgen

Oftmals kann es notwendig sein, zu kennzeichnen, dass zwei Benennungen aus verschiedenen Einträgen in einer Beziehung zueinander stehen.

Die Darstellung von hierarchischen Begriffsbeziehungen (z. B. Hyperonymie und Hyponymie) kann in Termbanken sinnvoll sein, mit denen Benutzer derartige Zusammenhänge aus unterschiedlichen Gründen erfassen möchten. Dies kann z. B. bei normativen Datenbanken zutreffen, bei denen die Standardisierung der Terminologie und die Genauigkeit der Definitionen im Vordergrund stehen. Auch für Softwareanwendungen, die solche hierarchischen Daten beispielsweise zum Wissensmanagement und zur Wissensgewinnung verwenden, kann die Darstellung hierarchischer Begriffssysteme von Nutzen sein. In solchen Fällen ist für jede Art der hierarchischen Beziehung eine gesonderte Datenkategorie erforderlich. Mit den entsprechenden Datenfeldern sollte der Benutzer eindeutig auf das Gegenstück der Beziehung verweisen können, bei dem es sich sowohl um einen Begriffseintrag als auch um eine Benennung innerhalb eines Begriffseintrags handeln kann. Als Beispiel ist im Folgenden eine XML-Repräsentation im TBX-Format (vgl. ISO 30042) dieses Datenkategorietyps dargestellt:

```
<descrip type="broaderConceptGeneric" target="cid1234">boat</descrip>
```

Das genannte Beispiel könnte in einem terminologischen Eintrag zu einem bestimmten Bootstyp, z. B. einem „Segelboot“, vorkommen, um von dort auf den Eintrag zum weiter gefassten generischen Oberbegriff „Boot“ zu verweisen. Die spezifische Kennung wird als Wert des Zielattributs angezeigt, welches die Kennung für den die Benennung „Boot“ enthaltenden Begriffseintrag darstellt. Der Inhalt des Elements, also „Boot“, ist keine spezifische Kennung, da das Wort für mehrere unterschiedliche Begriffe stehen und somit in verschiedenen Begriffseinträgen als Benennung vorkommen kann.

Angenommen ein Eintrag mit der Eintragsnummer cid1234 enthält mehrere Benennungen, nämlich „Boot“, „Schiff“ und „Wasserfahrzeug“, so kann es in einem solchen Fall notwendig sein, auf eine konkrete Benennung innerhalb des Eintrags zu verweisen. Möglich wird dies, indem jede Benennung innerhalb des Eintrags wie im folgenden Fall über eine spezifische Kennung verfügt:

```
Concept: cid1234
  boat: tid1234-1
  vessel: tid1234-2
  ship: tid1234-3
```

Möchte der Benutzer in genanntem Beispiel gezielt auf die Benennung „Boot“ verweisen, so würde der Link im TBX-Format wie folgt dargestellt:

```
<descrip type="broaderConceptGeneric" target="tid1234-1">boat</descrip>
```

Begriffssysteme können auch anhand von Notationsnummern rekonstruiert werden, mit denen die Position eines bestimmten Begriffs im Begriffssystem dargestellt wird. Links, die vom Begriffsdiagramm zu den entsprechenden terminologischen Einträgen – und umgekehrt – führen, sind zur Visualisierung von Begriffsbeziehungen äußerst nützlich.

Begriffssysteme können aber auch nicht-hierarchische Begriffsbeziehungen, z. B. assoziative und temporale Beziehungen, einschließen. Zudem gibt es nicht-hierarchische Begriffsbeziehungen, bei denen die Art der Beziehung nicht weiter spezifiziert wird – so etwa bei */relatedConcept/* (verwandter Begriff). In einigen Termbanken wird in solchen Fällen die Kennzeichnung „siehe“ oder „siehe auch“ verwendet. Im Folgenden ist eine beispielhafte Darstellungsweise im TBX-Format aufgeführt:

```
<descrip type="relatedConcept" target="cid2345">sail</descrip>
<ref type="see" target="cid2345">sail</ref>
<ref type="crossReference" target="cid2345">sail</ref>
```

Es empfiehlt sich, solche Begriffsbeziehungen mithilfe der dafür vorgesehenen Felder und Datenkategorien wie in den vorherigen Beispielen darzustellen, um so die Maschinenlesbarkeit zu verbessern. Die Art der Beziehung muss in terminologischen Datensammlungen, die hierarchische Begriffsbeziehungen enthalten, durch die Verwendung der entsprechenden Datenkategorien, z. B. */broaderConcept/* (übergeordneter Begriff), explizit kenntlich gemacht werden.

In einigen Termbanken werden Beziehungen auch in anderen Textfeldern, beispielsweise in einer Definition oder Anmerkung, dargestellt. Tatsächlich stehen Benennungen und Definitionen oft in enger Beziehung zu dem definierten Begriff, und es ist daher eine gängige Praxis, Benennungen als Hyperlink zu markieren und so durch Anklicken direkt den zugehörigen Begriffseintrag zu öffnen. Als Beispiel folgen hierzu zwei Querverweise im TBX-Format aus einer Beispielformatdefinition für die Benennung „Segelboot“.

```
<descrip type="definition">A <hi type="entailedTerm" target="cid1234">boat</hi>
propelled by a <hi type="entailedTerm" target="cid2345">sail</hi>.</descrip>
```

Begriffsbeziehungen können, unabhängig von deren Art, auf der Bedienoberfläche als Hyperlinks dargestellt werden. Durch einen Klick auf den Link können die Benutzer den Zieleintrag direkt öffnen.

7.4 Dateneintragsstrukturen

Termbanken und terminologische Einträge haben – wie in ISO 16642 beschrieben – eine logische Struktur, welche sich in einem hierarchischen Datenmodell widerspiegelt. Die Datenkategorien werden auf verschiedenen Ebenen dieses hierarchischen Modells „verankert“. Diesen Prozess nennt man Datenmodellierung (siehe 8.2).

Eine Termbank besteht aus globalen Informationen (Informationen zum gesamten terminologischen Datenbestand), Zusatzinformationen (beinhalten die gemeinsamen Ressourcen) und terminologischen Einträgen. Jeder terminologische Eintrag verfügt über eine Ebene für begriffsbezogene Datenkategorien (z. B. */conceptIdentifier/* (Begriffskennung), */subjectField/* und */definition/*), an die sich die Sprachebenen anschließen. Jede Sprachebene umfasst eine oder mehrere Benennungsebenen, die die Datenkategorie */term/* sowie benennungsbezogene Datenkategorien (z. B. */partOfSpeech/*, */grammaticalGender/*, */register/* (Register), */termType/* usw.) enthalten. Abschließend können die Benennungsebenen noch in Ebenen für Benennungsbestandteile aufgegliedert werden, in denen Teile von Mehrwortbenennungen oder phraseologischen Einheiten vollständig dokumentiert werden können. Die genannten Strukturen bilden gemeinsam ein Gerüst, an dem die terminologischen Einträge und Datenkategorien verankert sind.

Lässt das Datenmodell dies zu, so können einige Datenkategorien auf mehreren Ebenen vorkommen. Kontextsätze – also Texte, die die Benennung selbst enthalten – können beispielsweise sowohl benennungs- als auch begriffsbezogen sein, da sie Informationen zur Benennung oder auch zum Begriff enthalten können. In manchen Datenbanken sind Definitionen nur auf Begriffsebene zulässig, während sich in anderen Fällen sprachspezifische Definitionen auch auf der Sprachebene finden. Anmerkungen und administrative Datenkategorien können auf fast alle anderen Datenkategorien folgen und daher auf jeder Eintragebene vorkommen.

7.5 Auswahl der Datenkategorien

Die verschiedenen Zielgruppen eines TVS – zu denen nicht nur die Benutzer sondern auch Computeranwendungen (siehe Abschnitt 6) zählen – sowie die verschiedenartigen Endprodukte, die das TVS liefern soll, erfordern eine unterschiedliche Zusammenstellung der terminologischen Datenkategorien. Die in 7.1 genannten Datenkategorien werden zwar häufig verwendet, unter speziellen Umständen – beispielsweise bei der Softwarelokalisierung, für bestimmte Unternehmensanwendungen, die Terminologiestandardisierung oder für die mitunter von staatlichen Institutionen vorgenommene Sprachplanung – können aber auch weitere Datenkategorien erforderlich sein. Bei der Entwicklung eines TVS ist es daher wichtig, frühzeitig alle Datenkategorien zu ermitteln, die für den angestrebten Verwendungszweck benötigt werden. Um entscheiden zu können, ob der Kauf bzw. die Modifizierung eines bestehenden Systems der Entwicklung eines neuen TVS vorzuziehen ist, muss geprüft werden, ob ein bereits vorhandenes TVS alle erforderlichen Datenkategorien abdecken kann. Müssen bei einem existierenden TVS umfangreiche Ergänzungen oder Änderungen der Datenkategorien vorgenommen werden, so kann dies hohe Kosten verursachen. Zudem gibt es TVS, bei denen aufgrund ihrer Gestaltung keine Modifizierungen vorgenommen werden können.

Sind Altdaten (auch Bestandsdaten genannt) vorhanden, die in das TVS integriert werden sollen, so muss hier zunächst die Zusammenstellung der Datenkategorien beurteilt werden, um die für das neue TVS benötigten Datenkategorien festzulegen (siehe 8.5.). Es sollte darauf geachtet werden, dass die Bezeichnungen der Datenkategorien des bestehenden und des neuen TVS einheitlich sind.

In Anhang A findet sich eine Liste häufiger Datenkategorien, gefolgt von einem mehrsprachigen terminologischen Eintrag für Japanisch, Englisch und Deutsch.

7.6 Übersetzungsspezifische Datenkategorien

Benennungen, die in einem Eintrag in verschiedenen Sprachen aufgeführt werden, gelten als Übersetzungsäquivalente. Allerdings sind mitunter die Begriffe, auf die sich die Benennungen beziehen, nicht vollständig äquivalent zueinander. In einem solchen Fall kann es notwendig sein, die Unterschiede zwischen zwei quasi-äquivalenten Begriffen anzugeben oder festzuhalten, unter welchen Bedingungen sie als äquivalent gelten. Hierzu wird die Datenkategorie */degreeOfEquivalence/* (Äquivalenzgrad) verwendet.

In der Datenkategorie */falseFriend/* (falscher Freund) können Benennungen dokumentiert werden, die häufig – vielfach aufgrund morphologischer Ähnlichkeit – für semantische Äquivalente anderer Benennungen gehalten werden; tatsächlich haben die Benennungen jedoch eine unterschiedliche Bedeutung. Beispiel hierfür ist die französische Benennung *sortir*: Sie bedeutet nicht *sortieren* sondern *hinausgehen*. Zwei Benennungen, die als falsche Freunde gelten, müssen selbstverständlich in gesonderten Einträgen dokumentiert werden, da sie unterschiedliche Begriffe repräsentieren. Ein Querverweis mittels der Datenkategorie */falseFriend/* weist den Benutzer noch einmal ausdrücklich darauf hin, diesen potentiellen Fehler zu vermeiden. Es folgt ein Beispiel für einen solchen Link im TBX-Format, wie er in einem terminologischen Eintrag für „sortieren“ vorkommen könnte. In diesem Beispiel ist tid5678 die spezifische Kennung der Benennung „sortir“:

```
<termNote type="falseFriend" target="tid5678">sortir</termNote>
```

Eine weitere nützliche Datenkategorie stellt */transferComment/* (Übersetzungskommentar) dar. Sie kann zusätzlichen Aufschluss darüber geben, unter welchen Bedingungen quasi-äquivalente Begriffe verwendet werden können.

7.7 Präskriptive Datenkategorien

Eine Reihe von Datenkategorien dient der Festlegung von Terminologie. Mit der Datenkategorie */administrativeStatus/* kann beispielsweise angegeben werden, ob eine Benennung innerhalb einer bestimmten Arbeitsumgebung, z. B. in einem Wirtschaftsunternehmen, *bevorzugt*, *erlaubt* oder *abgelehnt* ist. Normgebende Instanzen, beispielsweise Normungsorganisationen und staatliche Institutionen zur Sprachplanung, können eine ähnliche Datenkategorie nutzen: die */normativeAuthorization/* (normative Zulassung). Sie weist auf normative Nutzungsbeschränkungen hin, etwa für den Fall, dass eine Benennung bereits in einer offiziellen Rechtsvorschrift definiert wurde.

7.8 Arbeitsorganisatorische Datenkategorien

Terminologische Einträge werden normalerweise in mehreren Phasen erstellt, welche mit der Datenkategorie */elementWorkingStatus/* (Bearbeitungsstand des Elements) kenntlich gemacht werden können. Die Werte *starter* (Anfangseintrag), *working* (in Bearbeitung) und *consolidated* (konsolidiert) spiegeln dabei die Vollständigkeit und den Zulassungsstatus des Eintrags wider. Die Datenkategorie */language-planningQualifier/* kann Werte wie *recommended* (empfohlen), *nonstandardized* (nicht genormt), *proposed* (vorgeschlagen) und *new* (neu) annehmen, um den Status eines terminologischen Eintrags in einem Zulassungsverfahren im Rahmen der Sprachplanung zu illustrieren. Mit der Datenkategorie */processStatus/* (Bearbeitungsstand) können grundlegende Angaben zum Bearbeitungsstand gemacht werden: *unprocessed* (unbearbeitet), *provisionallyProcessed* (vorläufig) und *finalized* (abgeschlossen).

7.9 Standardisierte Datenkategorienamen und Datenkategoriebegriffe

Wie bereits in 7.1 erwähnt, wurden die Namen der Datenkategorien sowie deren jeweilige Bedeutung im Verzeichnis der Datenkategorien (DCR) vereinheitlicht. Um den ISO-Normen zu entsprechen, darf ein TVS Datenkategorien nur entsprechend ihrer in der Auswahl der Datenkategorien des DCR von ISO/TC 37 (siehe 7.1) festgelegten Bedeutung und Beschreibung einführen und nutzen. Gegebenenfalls ist es jedoch auch möglich, in einem bestimmten TVS vom DCR abweichende Datenkategorienamen zu verwenden, sofern eine eindeutige Zuordnung zu den Datenkategorien im DCR möglich ist. In einem solchen Fall kann der Administrator des TVS die alternativen Bezeichnungen im DCR registrieren. So ist z. B. die Bezeichnung

„sample sentence“ für die Datenkategorie */context/* oder die Bezeichnung „grammatical category“ für die Datenkategorie */partOfSpeech/* in einem TVS möglich. Die Unterstützung verschiedener anwendungsspezifischer Datenkategorienamen kann auch dann wichtig werden, wenn ein TVS in einer anderen Sprache als Englisch entwickelt wird. In einem französischen TVS kann beispielsweise die Bezeichnung „*partie du discours*“ für die Datenkategorie */partOfSpeech/* stehen. Mitunter werden neue Datenkategorien auch von Entwicklern zum DCR hinzugefügt und können dort für die zukünftige Standardisierung vorgeschlagen werden.

8 Datenmodellierung

8.1 Das terminologische Metamodell

Datenmodelle liefern den formalen Rahmen für die Strukturierung von Daten in einem TVS. Mit terminologischen Metamodellen werden einheitliche komplexe Architekturen zur Verfügung gestellt, die bei Modellen einer Fachrichtung die Interoperabilität gewährleisten sollen.

Wie bereits in 7.4 erwähnt, legt ISO 16642 ein Rahmenmodell fest, welches Anleitung zu den Grundprinzipien der Darstellung und Aufzeichnung von Daten in einer terminologischen Datensammlung gibt. Das Rahmenmodell umfasst ein Metamodell sowie Methoden zur Definition einer XML-basierten (eXtensible Markup Language) terminologischen Auszeichnungssprache (TML). Das in ISO 16642 vorgestellte Metamodell erleichtert den Austausch von terminologischen Daten und kann zur Analyse einer bestehenden bzw. zur Entwicklung einer neuen terminologischen Datensammlung genutzt werden.

Spezifische Datenmodelle, die die Grundlage für TVS darstellen, müssen folglich auf dem in ISO 16642 vorgegebenen generischen Datenmodell basieren. Das terminologische Metamodell basiert seinerseits auf den in ISO 704 beschriebenen Methoden und Grundsätzen der Terminologieverwaltung für die Erstellung terminologischer Einträge.

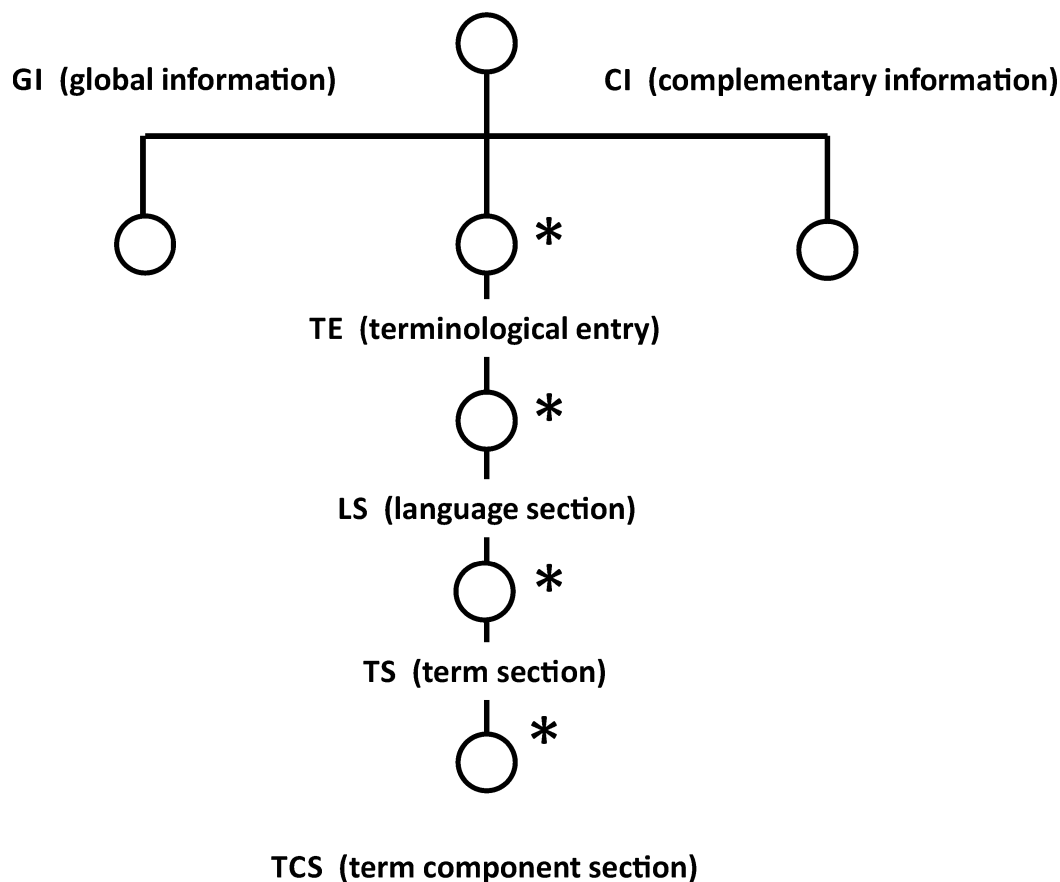
Das Metamodell ist ein abstraktes konzeptuelles Datenmodell, das die grundlegende Hierarchie der Informationsebenen darstellt, denen jede terminologische Datensammlung entsprechen muss: (1) globale Informationen über die Sammlung, (2) sämtliche Zusatzinformationen (gemeinsame Ressourcen, auf die in der gesamten Sammlung Bezug genommen werden kann) und (3) eine bestimmte Anzahl terminologischer Einträge.

Jeder Eintrag erfüllt drei Funktionen:

- er beschreibt einen Begriff bzw. zwei oder mehr quasi-äquivalente Begriffe in einer oder mehreren Sprachen;
- er dokumentiert die Benennungen, die die Begriffe bezeichnen;
- er beschreibt die Benennungen selbst.

Jeder Eintrag kann über mehrere Sprachebenen, jede Sprachebene über mehrere Benennungsebenen verfügen. Sämtliche Datenelemente in einem Eintrag können mit verschiedenen deskriptiven und administrativen Informationen verbunden sein. Zusätzlich gibt es diverse gemeinsame Ressourcen, auf die in ein oder mehreren Einträgen verwiesen werden kann. Zu diesen Ressourcen zählen Bibliografieangaben, Ontologiebeschreibungen sowie Bilder, die der Illustration von Begriffen dienen. Weitere Informationen zu gemeinsamen Ressourcen finden sich in 7.3.5.

Außerdem enthält ISO 16642 eine schematische Darstellung der Struktur eines terminologischen Datenbestandes, die in Bild 3 übernommen wurde.

TDC (terminological data collection)**Legende**

TDC (terminological data collection) = TDC (Terminologischer Datenbestand)

GI (global information) = GI (Globale Informationen)

CI (complementary information) = ZI (Zusatzinformationen)

TE (terminological entry) = TE (Terminologischer Eintrag)

LS (language section) = LS (Sprachebene)

TS (term section) = TS (Benennungsebene)

TCS (term component section) = TCS (Benennungskomponentenebene)

Bild 3 — Schematische Darstellung des terminologischen Metamodells aus ISO 16642

8.2 Datenmodellierung zur Begriffsorientierung

Einer der bedeutendsten Grundsätze der Terminologieverwaltung ist die Begriffsorientierung: Ein terminologischer Eintrag betrachtet ausschließlich einen Begriff. In einem mehrsprachigen Eintrag bezeichnen in der Regel alle Benennungen sprachübergreifend denselben Begriff. Jedoch können sich in den verschiedenen Sprachen die Eigenschaften des Begriffs unterscheiden. Für solche Fälle kann festgehalten werden, dass sprachspezifische Begriffe, die im gleichen Eintrag aufgeführt werden, als Quasi-Äquivalente gelten.

In Wissenschaft und Technik kommt es zwischen Benennungen in verschiedenen Sprachen mitunter zu großen begrifflichen Übereinstimmungen (siehe Bild 4), welche häufig so ausgeprägt sind, dass die Benennungen allgemein als Bezeichnungen für ein und denselben Begriff wahrgenommen werden. In den Sozial- und Kulturwissenschaften unterscheiden sich hingegen einige Begriffe je nach Sprache leicht voneinander. Dies kann selbst dann der Fall sein, wenn die Benennungen, die diese sprachspezifischen Begriffe bezeichnen, gemeinhin als funktional äquivalent gelten, also für den gleichen Übersetzungszweck verwendet werden können. Zudem können in einer Sprache Synonyme vorkommen, sodass jede Sprachebene mehrere Benennungen enthalten kann. Unterscheiden sich die Synonyme leicht in ihrer Bedeutung oder Anwendung, gelten auch sie als Quasi-Äquivalente. Um solche geringfügigen Bedeutungsunterschiede darzustellen, gibt es verschiedene Vorgehensweisen. Eine Methode ist die Entwicklung eines Datenmodells, das Begriffsvarianten zulässt. Des Weiteren kann ein homogenes Begriffsmodell genutzt werden, in dem Bedeutungsunterschiede auf Datenkategorieebene dokumentiert werden.

Bezeichnen zwei Benennungen in einer Sprache denselben Begriff, so werden sie als Synonyme gekennzeichnet und müssen im selben terminologischen Eintrag verzeichnet werden. Synonyme können sich jedoch bezüglich ihres Stils oder Registers unterscheiden; eine der Benennungen kann beispielsweise in bestimmten Kontexten passender sein als die andere. Diese Unterschiede müssen in einer geeigneten Datenkategorie, z. B. */administrativeStatus/* oder */register/* festgehalten werden.

Bild 4 zeigt als Beispiel einen mehrsprachigen Eintrag, der mehrere Benennungen in zwei Sprachen umfasst, die alle den Begriff *Protokollstapel* bezeichnen. Ähnlich wie viele Begriffe aus der modernen Wissenschaft und Technik hat sich diese Besonderheit der Internetarchitektur weltweit entwickelt und ist daher über die Kulturgrenzen hinaus bekannt und diskutiert. Die zugehörigen Benennungen sind in den meisten Sprachen im Grunde Lehnübersetzungen der ursprünglichen englischen Bezeichnung *protocol stack*. Hätte sich die Benennung in den verschiedenen Kulturen jeweils einzeln herausgebildet, so wäre das metaphorische Bild hinter der Benennung in den unterschiedlichen Sprachen möglicherweise ein anderes gewesen. Im Englischen wird in eindeutigen Kontexten häufig die Kurzform *stack* anstelle der Vollform der Benennung verwendet. Außerdem tritt die Paraphrase *stack of interconnected protocols* gelegentlich als eine Art Erklärung oder Synonym auf. In Bezug auf die Begriffsintension gibt es in den verschiedenen Sprachen nahezu keine Bedeutungsnuancen oder -abweichungen, sodass alle Benennungen problemlos ohne eine Übersetzungsanmerkung im selben Eintrag aufgeführt werden können.

Entry level
Entry number: 7
Subject: Protocols

English

Definition: particular software implementation of a computer networking protocol suite

Source Identifier: http://en.wikipedia.org/wiki/Protocol_stack

protocol stack

Source Identifier: IEE 802.11

Part of Speech: noun

Context: Another standard architectural model that is often used to describe a network protocol stack is the OSI reference model. This model consists of a seven layer protocol stack.

Source Identifier: <http://www.cisco.com>

stack

Source Identifier: <http://www.w3.org/DesignIssues/Stack.html>

Part of Speech: noun

stack of interconnected protocols

Source Identifier: <http://www.w3.org/DesignIssues/Stack.html>

Part of Speech: noun

German

Definition: konzeptionelle Architektur von mehreren Protokollschichten, die direkt aufeinander aufbauen

Source identifier: <http://www.esk.fraunhofer.de>

Protokollstapel

Source Identifier: <http://www.winfuture-forum.de.html>

Part of speech: noun

Grammatical Gender: masculine

Context: Daten, die über ein Netz übertragen werden, werden von einem Netzprotokoll des Protokollstapels nach dem anderen verarbeitet.

Source Identifier: <http://de.wikipedia.org/wiki/Protokollstapel>

Legende

Entry level = Eintragsebene

Entry number = Eintragsnummer

Subject = Fachgebiet

Protocols = Protokolle

English = Englisch

Definition = Definition

Source Identifier = Quellenkennung

Part of Speech = Wortart

Noun = Nomen

Context = Kontext

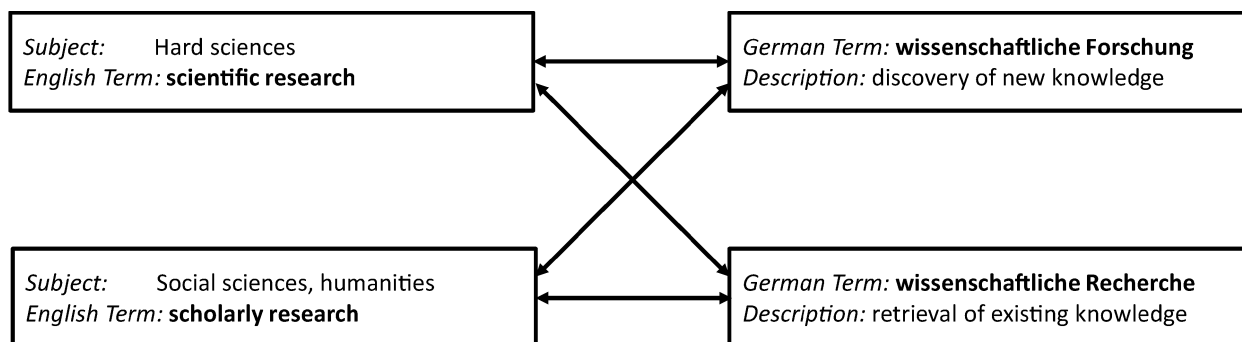
German = Deutsch

Grammatical gender = Genus

Masculine = Maskulinum

Bild 4 — Mehrsprachiger terminologischer Eintrag zur Dokumentation eines einzelnen Begriffs

Wie bereits in 7.6 beschrieben, bezeichnen verschiedensprachige Benennungen, die für praktische Zwecke als äquivalent gelten, in manchen Fällen nicht genau denselben Begriff. Dieses Phänomen der Begriffsverzweigung wird in Bild 5 veranschaulicht. Der Begriff, der der deutschen Benennung *wissenschaftlich* zugrunde liegt, wird – ins Englische übersetzt – entweder zu *scientific* (bei empirischen Naturwissenschaften) oder zu *scholarly* (in den Geisteswissenschaften). Parallel lässt sich die englische Benennung *research* aufklären: Sie kann im Deutschen entweder mit *Forschung* (Schaffung neuen Wissens) oder mit *Recherche* (Abruf und Einbeziehung vorhandenen Wissens) übersetzt werden.



Legende

Subject = Fachgebiet

Hard sciences = Naturwissenschaften

English term = Englische Benennung

Social sciences, humanities = Sozial- und Geisteswissenschaften

German term = Deutsche Benennung

Description = Beschreibung

discovery of new knowledge = Gewinnung neuer Erkenntnisse

retrieval of existing knowledge = Abruf vorhandenen Wissens

Bild 5 — Beispiel für die Äquivalenzbeziehung zwischen zwei englischen und zwei deutschen Begriffen

Bild 6 zeigt, dass durch die Kombination der genannten Begriffe vier neue Begriffe mit unterschiedlichen Bedeutungsnuancen entstehen, die möglicherweise in vier separaten terminologischen Einträgen dokumentiert werden sollten.

Entry 1:

Subject: Hard sciences
English Term: scientific research
German Term: wissenschaftliche Forschung
Description: discovery of new knowledge

Transfer comment:

The concept *scientific research* is partially equivalent to two German concepts, *wissenschaftliche Forschung* and *wissenschaftliche Recherche*, because the term *wissenschaftlich* refers to both the hard sciences and to the social sciences and the humanities.

Entry 3:

Subject: Social sciences, humanities
English Term: scholarly research
German Term: wissenschaftliche Forschung
Description: discovery of new knowledge

Transfer comment:

The concept *wissenschaftliche Forschung* is partially equivalent to the two concepts in English, *scientific research* and *scholarly research*, because a distinction is made in German between original scholarship that creates new knowledge (*Forschung*) and the retrieval of existing knowledge (*Recherche*).

Entry 2:

Subject: Hard sciences
English Term: scientific research
German Term: wissenschaftliche Recherche
Description: retrieval of existing knowledge

Transfer comment:

The concept *scientific research* is partially equivalent to two German concepts, *wissenschaftliche Forschung* and *wissenschaftliche Recherche*, because the term *wissenschaftlich* refers to both the hard sciences and to the social sciences and the humanities.

Entry 4:

Subject: Social sciences, humanities
English Term: scholarly research
German Term: wissenschaftliche Recherche
Description: retrieval of existing knowledge

Transfer comment:

The concept *wissenschaftliche Recherche* is partially equivalent to the two concepts in English, *scientific research* and *scholarly research*, because a distinction is made in German between original scholarship that creates new knowledge (*Forschung*) and the retrieval of existing knowledge (*Recherche*).

Legende

Entry = Eintrag

Subject = Fachgebiet

Hard sciences = Naturwissenschaften

English term = Englische Benennung

German term = Deutsche Benennung

Description = Beschreibung

Discovery of new knowledge = Gewinnung neuen Wissens

Transfer comment = Übersetzungskommentar

Bild 6 — Einträge, die die Informationen zu den Begriffen in Bild 5 zusammenfassen

Werden vier einsprachige Begriffseinträge erstellt, so kann der Benutzer entscheiden, welche Einträge miteinander verlinkt werden sollen, um Quasi-Äquivalente zu kennzeichnen. Diese Herangehensweise ist für die Darstellung von Begriffen auf der Sprachebene besser geeignet, als alle Sprachen in einem Eintrag zusammenzufassen. Allerdings wird der Ansatz von einem Großteil der im Handel erhältlichen TVS nicht verfolgt. In mehrsprachigen Einträgen könnten aber Übersetzungskommentare verwendet werden, um die hier gezeigten Beziehungen zu erläutern.

8.3 Anwendungsorientierte Vorgehensweise

Eine terminologische Datensammlung kann auf zwei Arten verwaltet werden: zum einen über ein relationales Datenbankmanagementsystem, zum anderen mittels der Speicherung als strukturierte Dokumente in einem – üblicherweise XML-basierten (eXtensible Markup Language) – Dateiformat mit Auszeichnung. Ist die Entscheidung zwischen einem Datenbankmanagementsystem und einer Auszeichnungsdatei gefallen, kann ein Datenmodell entwickelt werden. Bei einer relationalen Datenbank wird zur Beschreibung eines Datenmodells gewöhnlich ein ER-Diagramm (Entity Relationship) oder ein UML-Diagramm (Unified Modelling Language) verwendet. Bei einem XML-Dokument wird dafür in der Regel eine Dokumenttypdefinition (DTD) oder ein XML-Schema erstellt; beide können in einem Diagramm mit Baumstruktur dargestellt werden.

Um der vorliegenden Internationalen Norm zu entsprechen, müssen sämtliche XML-Formate und relationalen Datenbank Anwendungen für terminologische Daten auf den folgenden Grundlagen basieren:

- dem in ISO 16642 festgelegten Metamodell,
- einem mittels eines UML- oder ER-Diagramms bzw. einer XML-Dokumenttypdefinition oder eines XML-Schemas definierten Datenmodell und
- einer Auswahl von Datenkategorien, die aus dem in ISO 12620 festgelegten Verzeichnis für Datenkategorien stammen.

Bei der Definition der Datenstruktur eines TVS gibt es zwei unterschiedliche Herangehensweisen, die davon abhängen, ob das TVS Altdaten enthalten wird. Für die Vorgehensweise bei einem TVS mit Altdaten, siehe 8.5.

8.4 Beispiele für Datenmodellierungen

In einem ersten Schritt werden die Datenkategorien anhand der in Abschnitt 7 beschriebenen Grundsätze ausgewählt. Anschließend wird in einem zweiten Schritt die Datenstruktur mittels eines UML-Diagramms definiert. Abschließend werden im dritten Schritt die Datenkategorien ihrer entsprechenden Position in dieser Struktur zugeordnet. Das so entstandene Modell stellt die Grundlage für die Einrichtung des TVS dar.

Sollen die Daten in einer relationalen Datenbank gespeichert werden, kann das erstellte UML-Diagramm als Grundlage für die Ausarbeitung der Datenbanktabellen verwendet werden. Das UML-Diagramm kann auch als Grundlage für eine Dokumenttypdefinition oder ein XML-Schema dienen, wenn die Daten mithilfe eines XML-basierten TVS gesammelt werden.

Es folgt die beispielhafte Auflistung einer Zusammenstellung von Datenkategorien, die in den folgenden Abschnitten verwendet wird, um die Entwicklung eines Datenmodells für ein hypothetisches Beispiel-TVS zu beschreiben:

- Fachgebiet;
- Definition;
- Anmerkung;
- Benennung;
- Wortart;
- Quellenkennung (kombinierbar mit *Definition*, *Benennung*, *Kontext*, *Anmerkung*);
- Kontext;
- Übersetzungskommentar;
- Urheber;
- Ursprungsdatum.

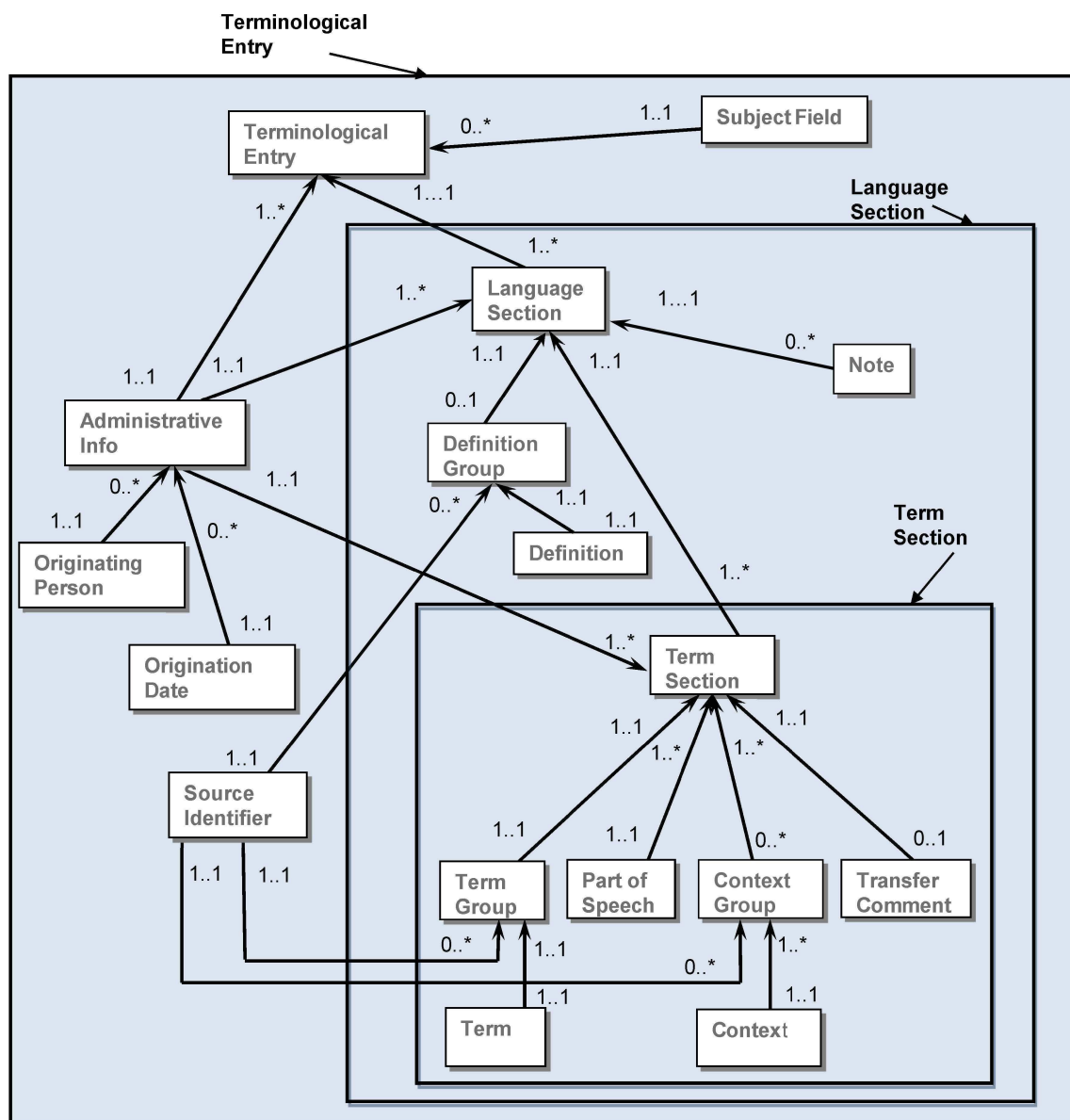
Bild 7 zeigt ein UML-Diagramm, das die oben genannten Datenkategorien zur möglichen Implementierung eines TVS umfasst. Klassen werden durch Rechtecke, Beziehungen (Assoziationen) zwischen den Klassen durch Linien gekennzeichnet. Zu jeder Klasse werden Angaben zur Multiplizität, also der Anzahl der an der Beziehung beteiligten Objekte der Klasse, gemacht. Zudem wird aufgeführt, ob diese Beziehung optional oder obligatorisch ist. In UML werden die folgenden Symbole verwendet:

- 1 genau ein Objekt ist an der Beziehung beteiligt
- 0..1 höchstens ein Objekt ist an der Beziehung beteiligt
- 1..* mindestens ein Objekt ist an der Beziehung beteiligt
- 0..* beliebig viele Objekte sind an der Beziehung beteiligt

Bei dem Modell in Bild 7 umfasst jeder terminologische Eintrag eine oder mehrere Sprachebenen. Jeder terminologische Eintrag wird einem Fachgebiet zugeordnet, ein Fachgebiet kann aber mit beliebig vielen terminologischen Einträgen verknüpft sein. Im vorliegenden Beispiel kann jede Sprachebene lediglich eine Definition beinhalten. Das bedeutet, dass zwischen der Sprachebene und der Datenkategorie */definition/* eine 1:1-Beziehung besteht. Die Null bedeutet, dass */definition/* optional ist. Die gleiche Beziehung besteht zwischen der Sprachebene und der Datenkategorie */note/*. Häufig sind Definitionen obligatorisch, und oft ist es sinnvoll, mehrere Definitionen speichern zu können, die sich beispielsweise an unterschiedliche Zielgruppen richten. Außerdem kann das Speichern von verschiedenen Versionen einer Definition nützlich sein, während sie die einzelnen Zulassungsstufen durchläuft. Voraussetzung hierfür wäre, dass die Multiplizitätsangabe zwischen der Sprachebene und der Datenkategorie */definition/* von 1..1 auf 0..* geändert würde. Im vorgestellten Modell ist auch lediglich eine Datenkategorie */sourceldentifier/* für jede */definition/* vorgesehen. In Fällen, in denen eine aus einer bestimmten Quelle stammende Definition von einem Terminologen verändert oder umformuliert wird, würde es sich jedoch empfehlen, sowohl die ursprüngliche Quellenkennung als auch den Namen des Terminologen als Quelle zu speichern. Laut dem in ISO 16642 festgelegten Metamodell kann sich eine Definition auch auf der Begriffsebene (terminologischer Eintrag, TE) befinden und sich so auf alle Sprachebenen beziehen. Dieser Ansatz kommt im Modell in Bild 7 jedoch nicht zur Anwendung.

Ob bestimmte Begriffe in zwei oder mehr Sprachen äquivalent sind, kann anhand ihrer Eigenschaften ermittelt werden. Sind zwei Begriffe vollständig äquivalent, so sind sie identisch und gelten als ein Begriff. Werden dagegen quasi-äquivalente Begriffe gemeinsam in einem Eintrag dokumentiert, so bedeutet dies, dass sie bis zu einem gewissen Grad dieselben Eigenschaften aufweisen und die zugehörigen Benennungen daher in Übersetzungen als Äquivalente verwendet werden können. Die Datenkategorie */transferComment/* gibt Aufschluss darüber, in welcher Äquivalenzbeziehung zwei so eng verbundene Benennungen stehen, die in einem terminologischen Eintrag gespeichert wurden. Beispielsweise kann es eine Anmerkung über die spezifischen begrifflichen Merkmale geben, die bei einer englischen und einer deutschen Benennung übereinstimmen. Eine weitere Anmerkung kann die Übereinstimmungen zwischen einer englischen und einer französischen Benennung aufzeigen.

Eine Benennung kann aus einer oder mehreren Quellen stammen, und eine Quelle kann mehreren Benennungen zugeordnet werden. Zwischen */term/* und */sourceIdentifier/* besteht daher eine n:m-Beziehung. Auch zwischen den Datenkategorien */term/* und */context/* besteht eine n:m-Beziehung, da eine Benennung in mehreren Kontexten vorkommen sowie ein Kontext mehrere Benennungen beinhalten kann. Im hier vorgestellten Modell kommt die Datenkategorie */context/* nicht immer vor, da sie optional ist. Des Weiteren zeigt das Modell in Bild 7 der Einfachheit halber lediglich eine Information (*/partOfSpeech/*), die mit der Datenkategorie */term/* verbunden ist, obwohl ein terminologischer Eintrag auch eine Vielzahl anderer Informationen umfassen könnte. Da sich jeder Wert der Datenkategorie */partOfSpeech/* auf mehrere Benennungen beziehen kann, besteht zwischen den Klassen */partOfSpeech/* und */termSection/* eine 1:n-Beziehung.



Legende

Terminological Entry = Terminologischer Eintrag

Language Section = Sprachebene

Term Section = Benennungsebene

Bild 7 — UML-Diagramm für grundlegende Datenkategorien

Soll das TVS mithilfe eines relationalen Datenbankmanagementsystems eingerichtet werden, so kann das Datenmodell in Bild 7 zur Ausarbeitung von Datenbanktabellen verwendet werden. Bild 8 zeigt mehrere Tabellen, die entsprechend dem Modell in Bild 7 erstellt wurden.

Die Bezeichnungen, die für die Tabellen und Spalten verwendet wurden, entsprechen denen des terminologischen Metamodells aus ISO 16642 sowie den im DCR enthaltenen Datenkategorien.

TerminologyEntry			
#TE_ID	Subject_field_ID		
SubjectField			
#Subject_field_ID	Subject_field		
LanguageSection			
#LS_ID	LANGUAGE	TE_ID	
TermSection			
#TS_ID	LS_ID	Transfer_Comment	
Definition			
DEFINITION_ID	DEFINITION	LS_ID	SOURCE_IDENTIFIER_ID

Bild 8 — Tabellenstruktur, die gemäß einigen Klassen aus dem in Bild 7 dargestellten Modell erstellt wurde

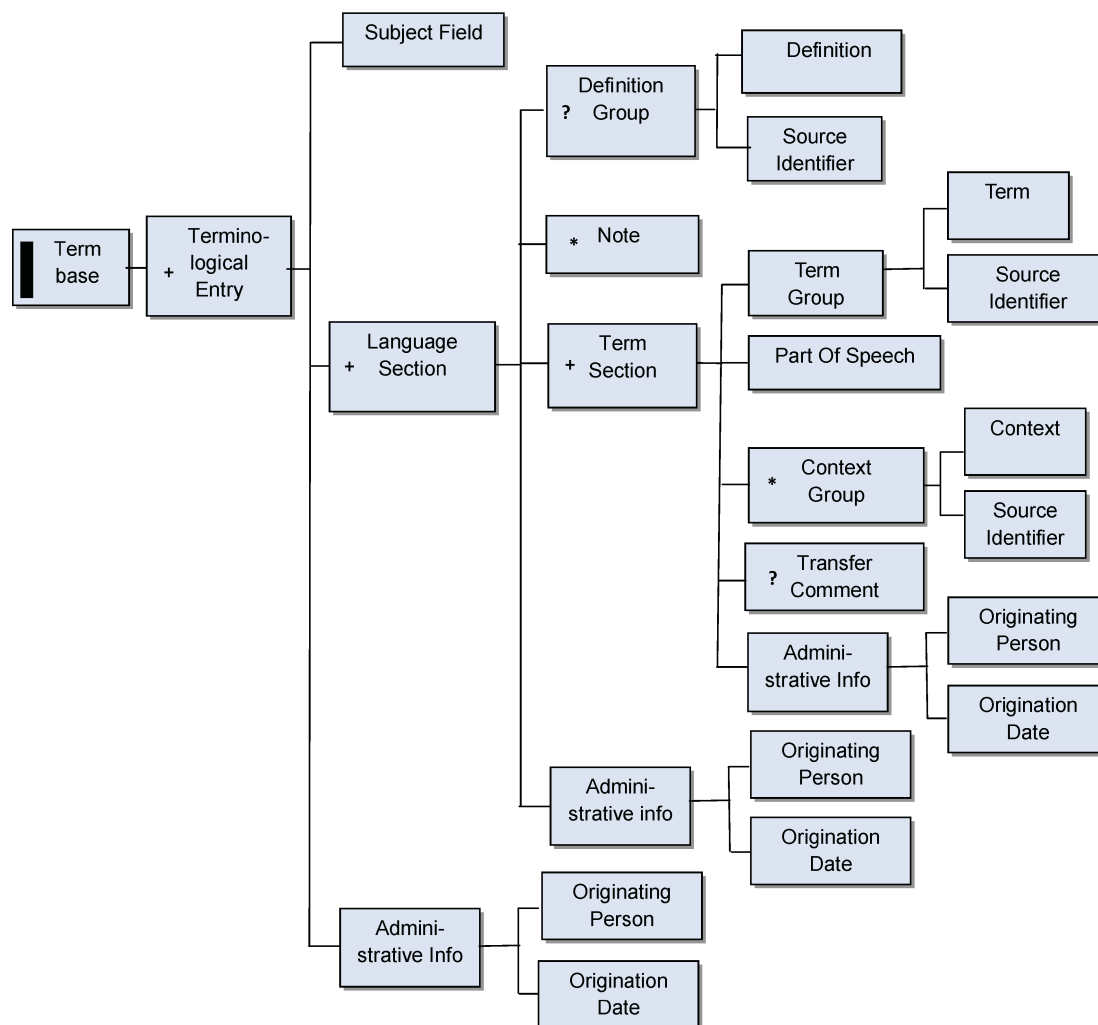
Bild 9 zeigt eine dem UML-Diagramm in Bild 7 entsprechende Dokumenttypdefinition (DTD).

Die DTD spiegelt eine Struktur wider, in der ein Eintrag einen Begriff oder zwei oder mehr quasi-äquivalente Begriffe umfasst, ohne Aufschluss über deren Äquivalenzbeziehung zu geben. Für ein normgerechtes XML-Vokabular sollten sich die Anwender auf ISO 30042 sowie die in ISO 16642 genannten Formate (einschließlich Geneter) berufen. Im nachfolgenden Beispiel wird zu Demonstrationszwecken ein fiktives XML-Vokabular definiert.

<!--Element	Termbase	(TerminologicalEntry+)>
<!--Element	TerminologicalEntry	(SubjectField, LanguageSection+, AdministrativeInfo)>
<!--Element	SubjectField	(#PCDATA)>
<!--Element	LanguageSection	(DefinitionGroup?, Note*, TermSection+, AdministrativeInfo)>
<!--Element	DefinitionGroup	(Definition, SourceIdentifier)>
<!--Element	Definition	(#PCDATA)>
<!--Element	SourceIdentifier	(#PCDATA)>
<!--Element	Note	(#PCDATA)>
<!--Element	TermSection	(TermGroup, PartOfSpeech, ContextGroup*, TransferComment?, AdministrativeInfo)>
<!--Element	TermGroup	(Term, SourceIdentifier)>
<!--Element	Term	(#PCDATA)>
<!--Element	PartOfSpeech	(#PCDATA)>
<!--Element	ContextGroup	(Context, SourceIdentifier)>
<!--Element	Context	(#PCDATA)>
<!--Element	TransferComment	(#PCDATA)>
<!--Element	AdministrativeInfo	(OriginatingPerson, OriginationDate)>
<!--Element	OriginatingPerson	(#PCDATA)>
<!--Element	OriginationDate	(#PCDATA)>

Bild 9 — DTD

Mit einigen XML-Tools können die Benutzer eine DTD oder ein XML-Schema erstellen und diese dann in ein Diagramm (hierarchisches Baumdiagramm) übertragen. Alternativ kann auch ein Diagramm erstellt werden, das anschließend in einer XML-Syntax dargestellt wird. Bild 10 zeigt ein hierarchisches Baumdiagramm, das die DTD in Bild 9 widerspiegelt.



Symbols:

■	root
?	zero or one
+	one or more
*	zero, one or more
no marking	exactly one

Legende

Symbols = Symbole

Root = Basis

zero or one = kann max. einmal vorkommen

one or more = muss min. einmal vorkommen

zero, one or more = kann beliebig oft vorkommen

no marking = keine Kennzeichnung

exactly one = kommt genau einmal vor

Bild 10 — Hierarchisches Baumdiagramm, das die DTD in Bild 9 widerspiegelt

Bild 11 zeigt den in Bild 4 dargestellten Eintrag mit XML-Tags, der entsprechend der DTD aus Bild 9 und dem hierarchischen Baumdiagramm aus Bild 10 erstellt wurde.