

coreon

Knowledge meets language.

Unleash the knowledge in your terminologies



Visually explore terms
through taxonomies.

Drive quality through
collaborative work in
semantic context.

Complement language tools,
tune enterprise search,
mine big data.

Easy start. Work right
inside any browser.

coreon.com

Nachruf auf Prof. Dr. Reiner Arntz



Letztes Jahr, am 10. Dezember, einem Montagnachmittag, meldete sich Reiner Arntz am Telefon: Nein, 100 Seiten Selbstbericht könne er diese Woche nicht mehr Korrektur lesen, aber zwei Kapitel könnten es schon sein.

Ja, morgen und per Mail würde reichen – bis Freitag dann.

„Was auf den Schreibtisch kommt, wird gelesen“, so eines seiner zahlreichen Bonmots und zugleich Ausdruck seines Arbeitsethos. Er arbeite nicht schnell, dafür aber lange und andauernd.

Reiner Arntz, geboren 1943 in Beckum (Westfalen), studierte nach dem Abitur 1962 zunächst in Bonn Romanistik und Jura und ab 1963 in Saarbrücken am damaligen Dolmetscher-Institut der Universität des Saarlandes. Bereits vor seinem Abschluss als Diplom-Übersetzer (1965 Spanisch, Französisch, Recht, 1966 Portugiesisch und Niederländisch) unterrichtete er ab 1965 am gleichen Institut Portugiesisch, ab 1966 Niederländisch. 1968 wurde er Lektor für Niederländisch und Spanisch mit dem Schwerpunkt Rechtssprache und wurde 1975 zum akademischen Rat und 1976 zum Ober-rat ernannt. Parallel dazu führte er sein Jurastudium fort, das er 1973 nach der Promotion als Volljurist abschloss.

In den 70er Jahren lernte Eberhard Tanke, der damalige Leiter des Sprachendienstes bei Siemens, den jungen Dozenten der Universität des Saarlandes kennen, der bereits damals ein gutes Dutzend Sprachen beherrschte. Er lud ihn zu einem Praktikum ein. Bis 1979 schlossen sich daran weitere Aufenthalte bei Übersetzungs- und Terminologiebüros an (u. a. Niederländisches Außenministerium in Den Haag, Bundessprachenamt in Köln; EU in Luxemburg, Philips AG in Eindhoven). Ziel war, das damals neue [und noch exotische] Gebiet „Terminologielehre“ in der Praxis kennenzulernen und als Fach für die Lehre aufzubauen.

1980 nahm Reiner Arntz den Ruf auf eine Professur für Romanische Sprachwissenschaft an der Universität Hildesheim an und baute den Diplomstudiengang „Fachübersetzen“ auf, den er später mit Beharrlichkeit und großem Engagement in den Studiengang „Internationale Fachkommunikation“ und in den „Bologna-Zeiten“ in neue Bachelor- und Masterstudiengänge überführte.

Von 1981 bis 1985 und von 2001 bis 2007 war er Geschäftsführender Leiter des Instituts für Angewandte Sprachwissenschaft der damaligen Hochschule und späteren Universität Hildesheim, von 1986 bis 1993 sowie von 1997 bis 2001 Dekan des Fachbereichs „Sprachen und Technik“.

In den neunziger Jahren wirkte er maßgeblich am Aufbau der Europäischen Akademie Bozen mit. Er hatte 1996/1997 die Vertretungsprofessur für Fachkommunikation und Fachsprachenpädagogik an der Wirtschaftsuniversität Dänemark inne, war von 2005 bis 2009 wissenschaftlicher Direktor des Fremdsprachenzentrums der Hochschulen im Lande Bremen und übernahm 2010 eine Gastprofessur an der Universität Sakarya in der Türkei.

Nach seiner Emeritierung 2008 hatte er sich, neben seinen vielfältigen Tätigkeiten, darunter auch als Kurator des Sprachen & Dolmetscher Institut (SDI München), auf die Länder verlegt, in denen ihn Menschen und wissenschaftliche Aufbauarbeit interessierten. Neben Südamerika waren das die Türkei sowie Russland und Weißrussland.

An jenem Montagnachmittag des 10. Dezember 2012 schien Reiner Arntz zwar etwas müde, war aber – wie so häufig, wenn er anrief – guter Dinge. Er erzählte von seinen Aufenthalten in Russland und Weißrussland, seinen letzten Vorträgen, seinen Plänen für die Türkei und von seinem Altershobby, dem Reiten auf seinem Pferd. Am frühen Abend des 10. Dezember 2013 ist Reiner Arntz, für alle unvermutet und völlig unerwartet, gestorben. Er fehlt.

Prof. Dr. Felix Mayer

Inhalt

Editorial

- 3 Nachruf auf Prof. Dr. Reiner Arntz
Felix Mayer

Kurznachrichten

- 6 Produkte, Unternehmen, Personen, Bücher

Schwerpunktthemen

- 9 Terminologie und Multilinguales Semantic Web in Symbiose –
TBX, RDF und OWL
Dagmar Gromann, Nathan Rasmussen und Alan Melby
- 15 Interdisziplinäre Terminologie – über die Konzeption einer
Ontologie für den Exzellenzcluster „Bild Wissen Gestaltung“
Christian Stein
- 20 Übersetzungsorientierte Terminografie – neue und adaptierte
Datenkategorien für ISOcat
Georg Löckinger

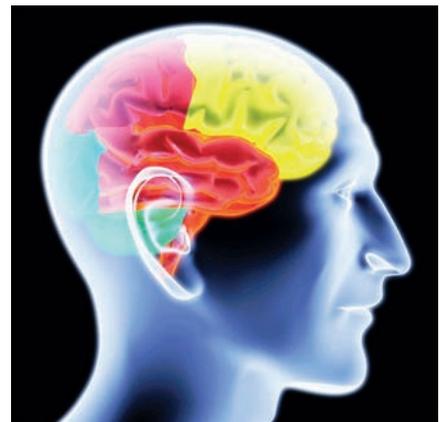
Datenfluss im Web



Terminologie ist überall. Sauber definierte Begriffe und Begriffssysteme sind mittlerweile auch im Bereich des Semantic Web ein Muss. Schnittstellen zwischen herkömmlichen Terminologieformaten wie TBX und formalen Beschreibungssprachen wie RDF wird daher in der Zukunft eine größere Bedeutung zukommen.

Von TBX nach RDF und zurück.....9

Interdisziplinäres Labor



Wie kommen Fachleute aus verschiedenen Fachbereichen auf einen gemeinsamen begrifflichen Nenner? Christian Stein von der Humboldt Universität zu Berlin zeigt auf, wie Ontologien im Exzellenzcluster „Bild Wissen Gestaltung“ modelliert werden..

Fachübergreifende Ontologien.....15

Impressum

eDITion
9. Jahrgang. eDITion erscheint zweimal im Jahr (Februar/August).
ISSN 1862-023X

Herausgeber

Deutscher Terminologie-Tag e.V. (DTT)
www.dttev.org

Redaktionsleitung

Dr. Detlef Reineke
Tel. 0034 928 415705
redaktion.edition@dttev.org

Wissenswertes

Dr. Nicole Keller
Tel. 06221 419036
wissenswertes.edition@dttev.org

Aus den Verbänden

Maryline Hernandez
Tel. 02642 9057727
verbaende.edition@dttev.org

Redaktionsbeirat

Prof. Dr. Gerhard Budin, Mark D.
Childress, Prof. Dr. Petra Drewer, Dr.
Christian Galinski, Ursula Fähndrich,

Dorina Frey, John D. Graham, Prof. Dr.
Felix Mayer, Angelika Ottmann, Prof. Dr.
Uwe Reinke, Prof. Dr. Klaus-Dirk Schmitz,
Prof. Dr. Sue Ellen Wright, Jutta Zeumer

Satz und Druck

SDK Systemdruck Köln
GmbH & Co. KG, Köln
www.sdk-koeln.de

Gestaltung

Detlef Reineke

Titelblatt

Detlef Reineke und Lydia Alarcón Queralt

Bezugspreis

Einzelheft €10,- zzgl. Versandkosten.
Anfragen richten Sie bitte an Detlef
Reineke (redaktion.edition@dttev.org).
Für Mitglieder des DTT ist eDITion im
Mitgliedsbeitrag enthalten. Informatio-
nen zur Mitgliedschaft und zum Verband
finden Sie unter www.dttev.org.

Abonnement

€20,- zzgl. Versandkosten. Anfragen rich-
ten Sie bitte an Detlef Reineke (redaktion.
edition@dttev.org). Abonnements laufen
mindestens ein Kalenderjahr und werden

Toolschau



Die Transline Deutschland Dr.-Ing. Sturz GmbH hat mit Termflow eine mobile, webbasierte Eigenentwicklung auf den Markt gebracht. Neben herkömmlichen Funktionen hat das Reutlinger Unternehmen bei der Entwicklung besonderen Wert auf managementspezifische Features gelegt wie Planung, Steuerung, Prozesskopplung, Schnittstellenmanagement und Controlling. Ein Bericht von Nicole Keller.

Management groß geschrieben....26

Terminologie pflegen



Untersuchungen haben gezeigt, dass ab einer bestimmten Zahl von terminologischen Einträgen und Verantwortlichen die Qualität von Termbanken abnimmt. Eine adäquate Datenbankpflege kann dann oft nicht mehr geleistet werden. Im Rahmen des LISE-Projekts wurden Strategien und Lösungen entwickelt, um die Qualität gewachsener Terminologieressourcen sicherzustellen.

LISE-Projekt.....31

Wissenswertes

- 26 Einfache Terminologieabstimmung mit Termflow
Nicole Keller
- 31 LISE – Investitionsschutz für gewachsene Terminologieressourcen
Michael Wetzel, Tanja Wissik und Elena Chiocchetti
- 33 SAP-Terminologie auf SAPterm.com verfügbar
Mark D. Childress

Aus den Verbänden

- 25 DTT-Seminar „Finanzberichte nach IFRS/DRS. Eine Einführung für Übersetzer und Terminologen“
- 32 DTT-Seminar „Terminologearbeit. Grundlagen, Recherche, Werkzeuge, Prozesse“
- 34 Neuer DTT-Vorstand
- 34 Arbeitstreffen des DIT in Karlsruhe

jeweils um ein weiteres Jahr verlängert, wenn nicht mindestens bis einen Monat vor Jahresende eine schriftliche Kündigung bei der Redaktionsleitung erfolgt.

Beiträge

Beiträge schicken Sie bitte bis zum 15. Mai (August-Ausgabe) bzw. 15. November (Februar-Ausgabe) an Detlef Reineke (redaktion.edition@dttev.org). Die formalen und organisatorischen Anforderungen entnehmen Sie bitte den Autorenrichtlinien, die Sie unter www.dttev.org beziehen können. Für unverlangte Einsendungen

wird keine Haftung übernommen. Die Verfasser erklären sich zudem mit einer nicht sinnentstellenden redaktionellen Bearbeitung ihrer Beiträge einverstanden. Beiträge, die mit vollständigem Namen oder Initialen des Verfassers gekennzeichnet sind, stellen nicht in jedem Fall die Meinung der Redaktion dar.

Copyright

© Die Zeitschrift ‚eDITion‘ und alle in ihr enthaltenen Beiträge und Abbildungen sind urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen

des Urheberrechtsgesetzes ist ohne Zustimmung des Verlags unzulässig und strafbar. Dies gilt vor allem für Vervielfältigungen, Übersetzungen, Mikroverfilmungen sowie die Einspeicherung und Bearbeitung in elektronischen Medien.

Anzeigen, Beilagen, Kurznachrichten und Leserbriefe schicken Sie bitte an Detlef Reineke (redaktion.edition@dttev.org). Redaktionsschluss ist der 15. Mai (August-Ausgabe) bzw. der 15. November (Februar-Ausgabe). Die formalen und organisatorischen Anforderungen entnehmen Sie bitte den

entsprechenden Richtlinien, die Sie unter www.dttev.org beziehen können.

Druckauflage

500 Exemplare

Kurznachrichten



Die **FH Oberösterreich** betreibt an vier Standorten in Oberösterreich (Hagenberg, Linz, Steyr und Wels) Studiengänge mit den Studienschwerpunkten Medien und Kommunikation, Management, Informatik, Gesundheit und Soziales, Wirtschaft und Technik, Umwelt und Energie sowie Technik. Die Fakultät für Technik und Umweltwissenschaften auf dem Campus Wels bietet u. a. das berufs begleitende, sechssemestrige Bachelorstudium „**Produktdesign und Technische Kommunikation (PDK)**“ an. In diesem Studiengang werden neben einer umfassenden Ausbildung im Bereich der Technischen Kommunikation und Dokumentation grundlegende Kompetenzen im Design von mechanischen Produkten und Softwaretools sowie fundierte Kenntnisse über

Ergonomie und Bedienungs-freundlichkeit vermittelt.

Zahlreiche Unternehmen wie der österreichische Telekommunikationskonzern emporia oder der Spritzgussmaschinenhersteller ENGEL GmbH arbeiten eng mit der Fakultät für Technik und Umweltwissenschaften zusammen. So wurde von Welser PDK-Studentinnen im Rahmen eines Praxisprojektes die bestehende Anleitung für das emporiaClick-Handy analysiert, einem Gebrauchstauglichkeitstest unterzogen und im Hinblick auf die Zielgruppe optimiert. Die ENGEL GmbH setzt auf das Know-how des Studienganges insbesondere in den Bereichen Terminologiemanagement und Lokalisierung, die zentrale Inhalte des Curriculums darstellen.

www.fh-ooe.at/pdk



Die **KERN AG, Sprachdienste** hat auf der tek-

om-Frühjahrstagung das Online-Terminologieportal **term4client™** vorgestellt. Mit der Eigenentwicklung können Unternehmen abteilungs- und länderübergreifend Fachbegriffe und deren Benennungen in allen Sprachen verwalten. Zu jedem Terminus können umfangreiche und benutzerdefinierte Metadaten zur Grammatik, Definitionen oder Verwendungsempfehlungen sowie erklärende Bilder hinterlegt werden. Der Zugriff auf die Terminologiebestände erfolgt über einen Web-Browser mit benutzerungebundenen Lese- oder benutzergebundenen Lese- und/oder Schreibrechten, die nicht nur eine flexible Recherche in der hinterlegten Terminologiedatenbank, sondern – ein Schreibrecht



Terminologiedienstleister

Dienstleister	Kontakt	E-Mail	Web	Dienstleistungen
dbterm terminologieservice	Diana Brändle	office@dbterm.de	www.dbterm.de	<ul style="list-style-type: none"> Beratung Training Praktische Terminologiearbeit
Fleury & Fleury Consultants	Isabelle Fleury	info@fleuryfleury.com	www.fleuryfleury.com	<ul style="list-style-type: none"> Beratung Schulung Implementation
Kaleidoscope GmbH	Klaus Fleischmann	klaus@kaleidoscope.at	www.kaleidoscope.at	<ul style="list-style-type: none"> quickTerm = MultiTerm + Terminologieworkflow
KERN AG, Sprachdienste	Maren Zieger	maren.zieger@e-kern.com	www.e-kern.com	<ul style="list-style-type: none"> Fachübersetzungen Terminologiemanagement Training & Beratung
Terminology Research & Development	Detlef Reineke	detlef_reineke@yahoo.es		<ul style="list-style-type: none"> Datenaustausch TBX Weiterbildung

vorausgesetzt – auch die Pflege bereits bestehender und die Ergänzung neuer Terminologieeinträge erlaubt. Die KERN AG erstellt in Abstimmung mit dem Auftraggeber und dessen Ländergesellschaften die kundenspezifische multilinguale Terminologiedatenbank, richtet die Website ein und verwaltet die Datenbestände auf abgesicherten Servern. Der Aufbau und das Hosten der Datenbankbestände seitens des Auftraggebers entfallen. Weil Auftraggeber und Sprachdienstleister gemeinsam auf die Terminologie zugreifen, wird die korrekte Wortwahl bereits im Ausgangstext sowie die Qualität und terminologische Konsistenz der beauftragten Übersetzungen sichergestellt.

www.e-kern.com



Der österreichische Softwareentwickler und Dienstleister **Kaleidoscope GmbH** hat vor dem Hintergrund Sozialer Medien und des Web 2.0 zahlreiche Neuerungen in seine Anwendung **quickTerm 5.3** integriert; denn Terminologiearbeit in Unternehmen dreht sich oft um die zentrale Frage: „Wie begeistere ich Kollegen und finde möglichst viele Mitstreiter für die gemeinsame Sache?“ Daher bietet quickTerm 5.3 zahlreiche Möglichkeiten, das Mitmach-Web für die eigene Terminologiearbeit zu nutzen. Ein „Term of the Week“ oder ein Terminologie-Quiz machen z. B. das Terminologieangebot im Intranet attraktiver und verankern es im Bewusstsein der Benutzer. Terminologen können den Kollegen wichtige Termeinträge empfehlen, die ihrerseits Termeinträge einfach „ liken “ können. Die Suchfunktionen lassen sich durch Anbieter wie Leo, Pons, Linguee, oder firmeneigene Datenbestände usw. erweitern. quickTerm 5.3 wartet mit erweiterten, interaktiven Workflows, einfacherer Anpassbarkeit an die individuellen Bedürfnisse und das jeweilige Cor-

porate Design auf. Völlig neu ist das „Dashboard“, das Steuerungszentrum von quickTerm: Versionierung, Änderungsverfolgung sowie ein genauer Überblick über die historische Entwicklung, Statistiken zu den Zugriffen oder zu den Freigabeprozessen sind hier nur einige der neuen Anwendungsmöglichkeiten.

www.kaleidoscope.at



Die **Transline Deutschland Dr.-Ing. Sturz GmbH** veranstaltet Seminare und Workshops im Bereich Übersetzungs- und Terminologiemanagement. Am 12.09.2013 findet das **Seminar „Management und Controlling von Terminologie im unternehmensweiten Kontext“** in Überlingen am Bodensee statt. Im Seminar erfahren die Teilnehmer, wie und wann sich der Einsatz von Terminologiemanagement-Systemen in Unternehmen rechnet und durch welche Kennzahlen die Erfolge messbar gemacht werden können. Zusätzlich werden konkrete Praxisbeispiele aus dem betrieblichen Alltag eingebunden und im Dialog mit den Referenten besprochen. Ziel des Seminars ist es, den Teilnehmern Grundlagen für ihre Kosten-Nutzen-Berechnung zu vermitteln, um die Effizienz ihres Terminologiemanagements zu steigern. Mehr Informationen und ein Anmeldeformular stehen auf www.transline.de/infocenter/termine bereit.

www.transline.de



Das Berliner Softwareunternehmen **Coreon** hat unter gleichem Namen eine Software für multilinguales Wissensmanagement entwickelt. Mittels eines neuartigen visuellen Ansatzes verbindet Coreon taxonomische mit terminologischen Methoden. Die in Markteinführung

stehende, neu entwickelte Software wendet sich an Organisationen, die ihre Datenbestände durch Struktur und Bedeutung aufwerten und somit deren Wert erweitern wollen. Anstatt Datenbankeinträge isoliert zu pflegen, ermöglicht die taxonomische, systematische Sicht eine visuelle und kontextbezogene Navigation: Eltern- und Kindbegriffe (z. B. Bildschirm – Flachbildschirm – LCD Anzeige) stehen nicht mehr alphabetisch weit getrennt, sondern semantisch verknüpft beieinander. Umgekehrt werden durch begriffsorientierte Terminologiedaten solche Taxonomien – wie auch Klassifikationssysteme oder Nomenklaturen – multilingual und z. B. zur Synonymie- oder Doublettenkontrolle intelligent nutzbar. Durch Coreon wird aus einer Terminologieressource eine multilinguale Wissensdatenbank. Sprachübergreifende Netze von Begriffen („concept maps“) gewährleisten nicht nur effiziente Terminologieverwaltung, sondern spielen eine Schlüsselrolle für organisationskritische, wertschöpfende Lösungen, wie Interoperabilität (Kopplung durch Abgleich mehrerer Begriffssysteme) oder die intelligente, multilinguale Analyse von großen Textdaten (z. B. Big Data Recherche-Verfeinerung mittels semantischer Schlussfolgerung). Die Plug-In- und installationsfreie Web-Applikation befindet sich seit Juni 2013 in der Beta-Phase. Die Geschäfte der eigens gegründeten Coreon GmbH werden von Jochen Hummel und Michael Wetzel geleitet.

www.coreon.com



Die neue Version des Terminologieverwaltungssystem **LookUp** der **D.O.G. Dokumentation ohne Grenzen GmbH** integriert neben zusätzlichen Filtermöglichkeiten auch drei neue Exportformate, mit denen sich z. B. dreidimensionale Modelle ohne größeren Konvertierungsaufwand in zweidimensionale

Excel-Tabellen exportieren lassen. Der Benutzer aktiviert die Felder auf Begriffs- sowie auf Sprach- und Benennungsebene, die er exportieren möchte, und entscheidet sich für eine der drei Varianten: eine Zeile pro Begriff und eine Benennung pro Zelle; eine Zeile pro Begriff und mehrere Benennungen pro Zelle (feste Anzahl von Spalten) und mehrere Zeilen pro Begriff bei einer Sprache pro Spalte. Diese Formate können auch wieder importiert werden.

www.dog-gmbh.de



Mit **[i]-correct**, dem Onlineportal von **itl** in einem Kooperationsprojekt entwickelt wurde, lassen sich Lektoratsprozesse webbasiert steuern und optimieren. Lektoren können mittels Weblink und Passwort auf die zu prüfenden SDLXLIFF-Dateien zugreifen und Korrekturen direkt in diesen Dateien ausführen, ohne dass Datenpakete verschickt werden müssen und ohne dass das Trados-native Format verlassen werden muss. Dieses Verfahren verringert die Gefahr, dass Elemente verändert oder zerstört werden und außerdem wird sichergestellt, dass alle Korrekturen direkt in das Translation-Memory übertragen werden. Komplexe Übersetzungsprojekte haben in der Praxis gezeigt, dass die Lektoratsarbeit im Onlineportal die Lektoren zu einem besseren Verständnis des gesamten Übersetzungsprozesses bringt, da sie derselben Segmentstruktur unterliegen wie die Übersetzer.

www.itl.eu

TERMDAT, die Terminologie-Datenbank der **schweizerischen Bundesverwaltung**, ist seit März 2013 vollumfänglich und uneingeschränkt im Internet abrufbar. TERMDAT wird seit über 20 Jahren von der Sektion Terminologie der Zentralen Sprachdienste der Bundeskanzlei betrieben und durch Datenbestände der mehrsprachigen Kantone ergänzt. Die Datenbank umfasst rund 400.000

Einträge, grossenteils in den vier Landessprachen Deutsch, Französisch, Italienisch und Rätoromanisch sowie in Englisch. Die systematisch erarbeitete Terminologie ist in thematische Sammlungen eingeteilt und stammt vorwiegend aus den Fachgebieten Recht und Verwaltung, aber auch aus zahlreichen weiteren Bereichen wie Bildung, Politik, Wirtschaft, Finanzen, Informatik, Telekommunikation, Verteidigung, Umwelt oder Verkehr.

Wegen wachsender Nachfrage wurde TERMDAT 2009 im Internet zugänglich gemacht, allerdings vorerst mit 100.000 Einträgen. Wer von ausserhalb der Bundesverwaltung auf den gesamten Datenbestand zugreifen wollte, musste für eine öffentliche Stelle tätig sein und ein Zertifikat beantragen.

Die vollständig erneuerte Anwendung ist nun erstmals in vollem Umfang frei zugänglich und kann auch über mobile Geräte konsultiert werden. Einige wesentliche Änderungen im Überblick: Validierte Einträge und Einträge in Bearbeitung sind neu in einer einzigen Datenbank zusammengefasst, farblich besser gekennzeichnet und leichter unterscheidbar. Ein übersichtlicheres aufgesplittertes Eintragsformat hat das bisherige kompakte Format ersetzt. Die vordefinierten Standardeinstellungen erlauben eine einfachere und schnellere Abfrage, auch ohne vorgängige Registrierung oder persönliches Konto. Die Suche lässt sich auf zusätzliche Felder (Definition, Anmerkung, Kontext, Quellen) ausweiten. Die Suchergebnisse werden in einer ausführlichen Trefferliste mit wählbarem Format angezeigt, die sämtliche Synonyme und Äquivalente in den gewählten Ausgangs- und Zielsprachen enthält. Die Treffer sind anhand farbiger Wappen auf den ersten Blick als Bundes- oder kantonale Einträge erkennbar. Ein verbessertes Erfassungsmodul erleichtert schliesslich die Erstellung und Bearbeitung von Einträgen. Weitere Optimierungen sind geplant.

www.termdat.ch



Ilona Wallberg ist seit Oktober 2010 bei **beo Gesellschaft für Sprachen und Technologie mbh** in leitender Position beschäftigt. Zum 01.05.2013 wird Frau Wallberg zur weiteren Geschäftsführerin berufen. Zu diesem Termin wird Herr Thomas Wedde als Mitglied der Geschäftsführung ausscheiden. Wir danken Thomas Wedde an dieser Stelle für die gute Zusammenarbeit in den letzten knapp 4 Jahren. Wir wünschen ihm alles Gute für seine neuen Aufgaben.

www.beo-doc.de



Across Systems GmbH hat mit der Version v5.7 des **Across Language Servers** neue Funktionen für

Übersetzungsprozesse eingeführt und vorhandene Features erweitert. Eine wesentliche Neuerung ist die webbasierte Lösung zur Erstellung und Verarbeitung von Terminologievorschlügen. Mithilfe des neuen Moduls können sich alle Mitarbeiter, die Across über einen Webbrowser nutzen, mit eigenen Vorschlägen am Aufbau der Firmenterminologie beteiligen. Die „Track Changes“-Funktion gewährleistet höhere Transparenz, die besonders den Anforderungen streng regulierter Arbeitsbereiche entgegenkommt, denn hierdurch können DOCX-Dokumente inklusive nachverfolgter Änderungen in die Zielsprachen übertragen werden. Zu den Neuheiten gehört auch ein Modul, das umfassende Lieferantenvergleiche erlaubt und Projektmanager bei der Preiskalkulation und Kostenkontrolle unterstützt. Zusätzliche Funktionen bei der Lokalisierung von Displaytexten, wie das Hinterlegen von Kommentaren, und die Unterstützung von XLIFF-Dateien runden den Leistungsumfang ab. Kurze Video-Präsentationen zu ausgewählten Funktionen finden Interessenten auf dem YouTube-Kanal von Across (Across-LanguageServer).

www.across.net

Terminologie und Multilinguales Semantic Web in Symbiose – TBX, RDF und OWL

Dagmar Gromann, Nathan Rasmussen und Alan Melby

Terminologische Ressourcen sind für Übersetzer wichtig, liegen jedoch in vielen unterschiedlichen Formaten vor. Der Industriestandard TBX (TermBase eXchange) ermöglicht den standardisierten Austausch dieser Ressourcen. Das Semantic Web bietet eine Vielzahl von maschinenlesbaren Daten mit expliziter Semantik und Technologien wie etwa das Resource Description Framework (RDF) oder die Web Ontology Language (OWL), die sich für Terminologien nutzen lassen. Im vorliegenden Beitrag soll insbesondere der von den Autoren entwickelte Konvertierungsmechanismus TBX/RDF vorgestellt werden.

Einleitung

Terminologie existiert seit vielen Jahren in Theorie und Praxis, hat jedoch bisher den Weg ins Semantic Web (SW) noch nicht gefunden. Das SW vernetzt gleichermaßen Daten und Dokumente mithilfe von Graphen und Begriffen, die auf Formalismen, etwa dem Resource Description Framework (RDF) oder der Web Ontology Language (OWL), aufbauen. Das Ziel des SW ist es, diese Daten und Dokumente mithilfe expliziter Semantik maschinenlesbar und für automatische Inferenzen durch einen Reasoner zugänglich zu machen. Eine Anbindung terminologischer Ressourcen an das SW bietet Übersetzern und Terminologen die Möglichkeit, Unmengen an strukturierten Daten mit Wissen in lokalen Termbanken zu verknüpfen und so der SW- sowie der Linked Data-Gemeinschaft die Nutzung terminologischer Ressourcen und Expertise zu ermöglichen. Stellt man Termbanken in maschinenlesbaren SW-Technologien dar, können automatisierte Schlussfolgerungen aus den darin enthaltenen terminologischen und wissensbasierten Fakten gezogen werden. Der XML-basierte ISO- und Industriestandard TBX (ISO 30042: 2008) erleichtert den Austausch von Termbanken. Die Darstellung von Datenkategorien und Definitionen bietet allerdings keine Möglichkeit, automatisierte Inferenzen zu realisieren. Aus

diesem Grund arbeiten die Autoren an der Entwicklung des Konvertierungsmechanismus TBX/RDF, der es ermöglicht, Terminologien und SW einander näher zu bringen. Der ontologiebasierte Formalismus TBX/RDF ist eine weitgehend direkte Übersetzung der bestehenden XML-Struktur von TBX nach RDF, wodurch das zugrunde liegende Datenmodell des TBX-Standards erhalten bleibt. Eine Standardisierung von TBX/RDF ist derzeit noch nicht angedacht.

Begriffsorientierte Termbanken enthalten terminologische Einträge, in denen Begriffe eines Fachgebiets und deren Bezeichnungen dokumentiert werden. Terminologische Ressourcen können dabei mithilfe von TBX-Dialekten wie TBX-Default, TBX-Basic, TBX-Glossary oder anderen benutzerdefinierten Dialekten dargestellt werden, die sich nach Anzahl, Platzierung und den festen Wertemengen der verwendeten Datenkategorien voneinander unterscheiden. Die Wiederverwendbarkeit terminologischer Ressourcen und die Interoperabilität von TBX-Dialekten können durch die Verwendung semantischer Technologien wie etwa RDF und OWL erhöht werden, denn diese Technologien ermöglichen es, „Dinge“ (Ressourcen) im Web, also etwa auch die in der Online-Registrierungsdatenbank ISOcat abgelegten Datenkategorien, eindeutig und einheitlich durch sogenannte Uniform Resource Identifier (URI)

zu referenzieren und zu identifizieren. TBX/RDF reduziert somit die Notwendigkeit, für jeden existierenden, zukünftigen und benutzerdefinierten TBX-Dialekt eine eigene Konvertierungssoftware zu entwickeln, da deren Darstellung in RDF einen automatisierten Vergleich der Dialekte erheblich erleichtert. Im Umkehrschluss bedeutet das aber auch, dass TBX/RDF die Erstellung benutzerdefinierter TBX-Dialekte fördern kann.

Terminologien lassen sich durch TBX/RDF in semantische Formalismen konvertieren (TBX zu TBX/RDF), die Konvertierung ist aber auch in entgegengesetzter Richtung möglich (TBX/RDF zu TBX). Die hierarchische Struktur eines TBX-Dokuments wird in einem RDF-Graphen neu erstellt (z. B. Abb. 1 oder Beispiel 2), wodurch das Datenmodell des TBX-Standards erhalten bleibt und auf einfache Weise wieder in TBX zurückkonvertiert werden kann. Dadurch soll Terminologen der Weg zu RDF erleichtert und ihre praktische Erfahrung in Form von Termbanken der SW-Gemeinschaft eröffnet werden. Die Beibehaltung des bestehenden TBX-Datenmodells unterstützt dieses Ziel.

TMF, TML und TBX

In der ISO 16642:2003 wird ein Rahmen (Terminological Markup Framework = TMF) für die Beschreibung terminologischer Markup-Sprachen

(TML) festgelegt. TBX ist eine Familie terminologischer Markup-Sprachen, zu der z. B. die oben bereits genannten Dialekte TBX-Default, TBX-Basic und TBX-Glossary gehören. Die TBX-Grundstruktur entspricht weitgehend dem TMF-Metamodell und unterteilt sich in Begriffs-, Sprach-, Bezeichnungs- und Bezeichnungskomponentenebene. Grundstruktur und Datenkategorien sind die beiden zentralen Module von TBX. Die Datenkategorien basieren auf dem ISO-Standard 12620 aus dem Jahr 1999 und werden seit 2009 als Einträge einer global einheitlichen Registrierungsdatenbank online angeboten (www.isocat.org). Die ambivalente Platzierung der Datenkategorien auf verschiedenen strukturellen Ebenen stellt eine Modellierungsherausforderung dar. Beispielsweise kann eine Definition sowohl auf Begriffs-, Sprach-, aber auch auf der Bezeichnungsebene verwendet werden. Bei der wörtlichen Übersetzung der TBX-Elemente in RDF wird die vorhandene Struktur übernommen. In weiterer Folge ist jedoch auch eine logische Darstellung des TBX-Datenmodells in OWL geplant, wozu die derzeitige XML-Struktur mit expliziter Semantik in formaler Logik erstellt werden muss.

Einige andere Elemente der TBX-Grundstruktur könnten von einer leicht veränderten Granularität profitieren. Beispielsweise wird die Bezeichnungsebene nach „term information group“ (<tig>) und „nested term information group“ (<ntig>) unterteilt, wobei das <ntig>-Element eine größere Verschachtelungstiefe zur näheren Beschreibung der Komponenten einer Bezeichnung zur Verfügung stellt (Beispiel 1).

```
<tig id="C001en1">
  <term>federated database</term>
  <termNote type="partOfSpeech"
  >noun</termNote>
</tig>
```

Beispiel 1: Bezeichnungsebene eines TBX-Eintrags

Die Einführung eines Elternelements <termSection>, also eines Kindelements zu <langSet>, würde die Unterscheidung nach <tig> und <ntig> ersetzen. Beispiel 2 zeigt den TBX-Eintrag aus Beispiel 1 in RDF, wobei eine Datenkategorie mit der Bezeichnung „term section“ anstelle eines <tig>-Elements verwendet wird. Zur Veranschaulichung zeigt Beispiel 2 die ID und Bezeichnung (angefügt mithilfe eines Kommentars #) jeder verwendeten Datenkategorie. Namensraum-Präfixe wie etwa „cat“ oder „ex“ dienen der Abkürzung des URI und werden am Beginn der RDF-Datei definiert.

```
ex:C001en1 a cat:DC-5250; # term
section
  tbx:textValue "federated database";
  cat:DC-1345 [ # part of speech
    tbx:simpleDatcat cat:DC-1333.
    # noun
  ].
```

Beispiel 2: RDF-Darstellung der TBX-Bezeichnungsebene aus Beispiel 1 in Turtle-Notation

Die RDF-Darstellung von TBX-Einträgen ermöglicht in manchen Fällen eine feinere Granularität, wie etwa die Verwendung der Datenkategorie */quasi-synonym/* oder */synonym/*, erfordert aber in jedem Fall die explizite Darstellung impliziter TBX-Inhalte. Beispielsweise nimmt die ISO 30042:2008 implizit an, dass Einträge auf derselben Bezeichnungsebene, also in einem Eintrag, synonym sind, außer explizit anders angegeben. Daher wird die Angabe der Datenkategorie */synonym/* laut ISO 30042:2008 nicht empfohlen. Für eine Darstellung dieser Tatsache in semantischen Formalismen ist es erforderlich, einen expliziten Verweis auf die ISO Datenkategorie */synonym/* anzugeben. TBX/RDF kann auch benutzerdefinierte Datenkategorien wie etwa */quasi-synonym/*, die nicht in der ISO 30042:2008 spezifiziert sind, darstellen.

SW-Umgebungen

Das SW ist eine Evolution des bestehenden Webs. Ziel ist es, Daten für Mensch und Maschine lesbar zu machen. Durch das Verlinken von Daten gelangt der Benutzer aufgrund ihrer Gleichartigkeit von einem Datensatz zu vielen anderen mit identem Kontext. Natürliche Sprache ermöglicht es, diese Daten mit wahrnehmbaren oder gedachten Gegenständen zu verbinden und stellt diese für menschliche Benutzer verständlich dar. Um Wissen Maschinen zugänglich zu machen, benötigt das SW bestimmte Formalismen. Die richtungsweisende Organisation hinter der Weiterentwicklung von Standards für SW-Formalismen und das SW ist das World Wide Web Consortium (W3C), das auch Empfehlungen für RDF(S) und OWL definiert. RDF Schema (RDFS) wird weitgehend als Formalismus zur Darstellung terminologischer Ontologien klassifiziert, also auf taxonomischen Hierarchien basierend, wohingegen OWL als Repräsentationsformat für formale Ontologien betrachtet wird, die ein ausgeprägtes logisches Fundament aufweisen, aus dem automatische Schlussfolgerungen gezogen werden können.

RDF versieht Ressourcen im Web mit URIs, auch wenn diese nicht direkt im Web abrufbar sind, wie beispielsweise Personen. RDF besteht aus Ressourcen, die

- jede Art von mittels URI identifizierten Daten sein können
- Properties, die Attribute oder Beziehungen von Ressourcen kennzeichnen
- und Statements (Aussagen), die der Property einer bestimmten Ressource einen Wert zuweisen.

Ähnlich einem Satz besteht ein Statement aus dem Tripel Subjekt, Prädikat und Objekt, z. B. *TBX* (RDF-Ressource und Subjekt) *repräsentiert* (RDF-Property und Prädikat) *Terminologie* (RDF-Ressource und Objekt). RDF fokussiert auf die Lesbarkeit dieser Informationen sowie deren Austausch

durch Anwendungen. Genauere Beschreibungen von RDF erfordern ein RDF Schema (RDFS), wie etwa die Bestimmung von Restriktionen und die Ordnung in einer Klassenhierarchie. RDFS gilt daher als die einfachste Form der Ontologiesprachen. RDFS ermöglicht die Beschreibung eines Vokabulars für RDF, z. B. bestimmte Klassen oder Arten von Ressourcen. Linked Data basiert auf RDF sowie RDFS und repräsentiert Best Practices zur Veröffentlichung und Vernetzung strukturierter Daten im Netz [1]. RDF in Attributes (RDFa) bietet eine Reihe von Attributen zur Einbettung maschinenlesbarer, strukturierter RDF-Statements in visuellen Informationen im Web, z. B. in HTML-Webseiten oder XML-Strukturen, um deren automatisierte Verarbeitung zu erleichtern.

Die Web Ontology Language (OWL) ist eine Erweiterung von RDF in einer formalen Beschreibungssprache mit einer formalen Semantik. Natürliche Sprache kann in Ontologien durch die RDFS-Elemente `rdfs:label` und `rdfs:comment` sowie RDF URI-Referenzen dargestellt werden. Benennungen sind jedoch ausschließlich in `rdfs:label`-Elementen zu finden, da die anderen Elemente Beschreibungen, Kommentare oder natürlichsprachliche IDs enthalten. Ein Ontologie-Element kann mehrere `rdfs:label`-Elemente enthalten, deren Sprache jeweils durch den Wert eines `xml:lang`-Attributs gekennzeichnet ist. Sowohl auf RDF als auch auf OWL können sogenannte Reasoner angewandt werden, um automatisierte Schlussfolgerungen zu ziehen und Fakten zu extrahieren. OWL ist auf formaler Logik begründet wie etwa Beschreibungslogik und ist daher ausdrucksstärker als RDF. Dennoch basieren sowohl OWL als auch die neuere Empfehlung OWL 2 auf RDF und OWL-Ontologien können auch als RDF-Dateien ausgetauscht werden. Da nicht jede Anwendung den gleichen Bedarf an logischen Ausdrücken hat, gibt es drei Unterkategorien

von OWL (OWL-Lite, OWL-DL und OWL-Full) und weitere drei OWL 2 Profile, die syntaktische Einschränkungen zu Beschreibungslogik darstellen.

TBX/RDF

Die Entwicklung von TBX/RDF beruht auf einer relativ direkten Übersetzung des existierenden XML-basierten TBX Standards nach RDF, ermöglicht jedoch auch die entgegengesetzte Richtung. Nach einer erfolgreichen Einführung terminologischer Ressourcen in die RDF-Welt können Modellierungsdiskussionen forciert werden, da RDF eine vereinfachte Adaptierung vorhandener Informationen in anderen Modellen ermöglicht. Eine optimale Interaktion zwischen TBX und RDF erfordert einen Datentransfer in beide Richtungen. Die TBX-zu-RDF Richtung wird durch ein neues XML-Schema ermöglicht, das TBX in der Struktur ähnlich ist, jedoch RDFa 1.1 Attribute beinhaltet. RDFa bettet RDF in Form von Attributen in ein XML-Dokument ein. Die Anordnung dieser Attribute wird bei TBX/RDF durch die TBX-Dokumentstruktur vorgegeben. Ebenso bestimmen die Datenkategorien jedes TBX-Dialekts die in RDFa zu verwendenden Attribute, damit existierende TBX-Dokumente direkt mit RDFa annotiert werden können. Die RDF-Topologie wird durch die Anordnung der Attribute bestimmt sowie die RDF-Prädikate über die verwendeten URIs festgelegt werden, wodurch das annotierte Dokument das gewünschte RDF-Ergebnis durch die Verarbeitung in einem allgemeinen RDFa-Distiller erzeugt.

Als destillierte RDF-Daten werden terminologische Datenkategorien nicht mit textuellen Labels wie bei TBX-Dialekten dargestellt, sondern mit global einheitlichen URIs (idealerweise URIs von ISOcat), wie in Abb. 1 dargestellt. Das bedeutet, dass RDF-Daten frei von Dialekten sind und verwendete Datenkategorien eindeutig identifiziert werden können,



term4client™

Online-Terminologie-datenbank

Terminologierecherche
und -verwaltung

Ein- oder Mehrsprachig

Flexible Rechteverwaltung

☎ (069) 75 60 73-0

www.e-kern.com
www.kerntraining.com

KERN AG, Sprachendienste

Kurfürstenstraße 1, 60486 Frankfurt/Main
Fax: +49 (0) 69-75 60 73-56 49
info@e-kern.com

Aachen · Augsburg · Berlin · Bielefeld · Bochum · Bonn
Braunschweig · Bremen · Darmstadt · Dortmund · Dresden
Düsseldorf · Duisburg · Essen · Frankfurt/Main · Freiburg
Friedrichshafen · Hamburg · Hannover · Heilbronn · Ingolstadt
Kaiserslautern · Karlsruhe · Kassel · Kiel · Köln · Leipzig
Mainz · Mannheim · Mönchengladbach · München · Münster
Nürnberg · Offenbach · Saarbrücken · Schweinfurt · Stuttgart · Ulm · Weil a. Rh. · Wiesbaden · Würzburg · Wuppertal

International: Amsterdam · Hong Kong · London · Lyon
New York · Paris · San Francisco · Wien

einschließlich benutzerdefinierter Datenkategorien und nicht nur jene in der TBX-Spezifikation, wodurch TBX/RDF einen verlustfreien Datenaustausch ohne Rücksprache (blind interchange) weiter unterstützt. Abb. 1 zeigt eine englische „Language Section“ von TBX/RDF mit zwei Synonymen und einer Definition. Ressourcen werden durch Ellipsen dargestellt, Properties in Rot auf den Pfeilen und Rechtecke stellen Werte, z. B. natürlichsprachliche Bezeichnungen, dar.

Reasoner, also Anwendungen um automatische Inferenzen zu ziehen, sind für die Konvertierung von terminologischen Daten in RDF zurück zu TBX essenziell. Natürlich entspricht nicht jeder RDF-Graph dem TBX-Datenmodell und dialektneutrale Daten müssen mit den für den entsprechenden TBX-Dialekt erforderlichen natürlichsprachlichen Bezeichnungen ausgestattet werden. Ein Großteil der Konvertierung kann in RDF mithilfe zweier Ontologien in OWL ermöglicht werden, die dem Reasoner die Bestimmung der erforderlichen RDF-Fragmente erleichtern und die dialekt-spezifischen Informationen liefern. Dieser Vorgang kann wie folgt beschrieben werden: Eine Ontologie entspricht der Grundstruktur von TBX. Sie bestimmt ausschließlich jene Fragmente des RDF-Graphen, welche den Strukturelementen von TBX/RDF in der begriffsorientierten Gliederung entsprechen. Die zweite Ontologie wurde speziell für den bestimmten Zieldialekt von TBX erstellt, liefert menschenlesbare sowie dialekt-

neutrale Identifikatoren und ordnet den TBX-Elementen (z. B. <descrip>, <admin>, etc.) Datenkategorien zu. Sie kann auch andere Restriktionen enthalten, wie etwa die Einschränkung, dass jede Bezeichnung durch mindestens eines der folgenden Elemente beschrieben werden muss: Wortklasse, Definition, Kontextbeispiel. Jedoch muss die Ontologie nur die entsprechenden Elemente, die für jeden Dialekt auf einen bestimmten URI-Namensraum beschränkt sind, zum RDF-Graphen hinzufügen und Informationen verschiedener Dialekte können konfliktfrei im selben Graphen überlagert werden. Die Ontologie der Grundstruktur wird letztlich erneut verwendet um sicherzustellen, dass die zugewiesenen TBX-Elemente den Restriktionen des zugrunde liegenden Dokumentenschemas entsprechen. Wenn das der Fall ist, werden sie als exportbereit markiert.

An dieser Stelle sind alle erforderlichen Informationen für die Erzeugung eines TBX-Dokuments im RDF-Graphen durch bestimmte Eigenschaften gekennzeichnet, entweder im Namensraum des Dialekts oder im TBX-Namensraum. Diese Informationen können beim Export des Graphen ausfindig gemacht werden (in einer adäquaten Notation wie etwa Grit oder nur als Datenstruktur) und an ein entsprechendes System übergeben werden, um das ganze TBX-Dokument zu erstellen. Ausschlaggebend ist, dass der RDF-zu-TBX-Code lediglich den Namensraum der dialekt-spezifischen Informationen benötigt und generisch

sein kann, da die Entscheidungen bezüglich des Dialekts bereits in RDF getroffen wurden. Im Allgemeinen kann TBX/RDF den Austausch von Daten zwischen TBX-Dialekten vereinfachen und soll die (gut durchdachte) Verwendung von benutzerdefinierbaren Bestandteilen von TBX fördern.

TBX und Logik

TBX/RDF ist ein erster Schritt der Annäherung der TBX- und SW-Landschaft, dem viele weitere folgen können. Die eingangs erwähnte Logik von OWL ist in der Lage, zwei wichtige Aspekte zu erfüllen. Einerseits können Einschränkungen auf bestimmte TBX-Elemente, die dem Datenmodell bzw. Dialekt zugrunde liegen, maschinenlesbar dargestellt werden. Beispielsweise kann die Einschränkung, dass entweder ein <tig>- oder <ntig>-Element in einem terminologischen Eintrag enthalten sein muss, mit owl:disjointWith definiert werden. Andererseits kann eine logisch formal definierte Ontologie auch eine Kontrollinstanz für die Kompatibilität einzelner Termbanken in RDF mit den Grundanforderungen von TBX sicherstellen, um eine nahtlose Integration verschiedener Ressourcen zu gewährleisten. Automatisches Reasoning und automatische Inferenzen sind auch in TBX/RDF möglich, jedoch liefert eine axiomatisierte, also logisch restriktiv deklarierte Ontologie, aufgrund der klar und explizit definierten Semantik bessere Ergebnisse, weshalb TBX/RDF auch zwei OWL-Ontologien verwendet.

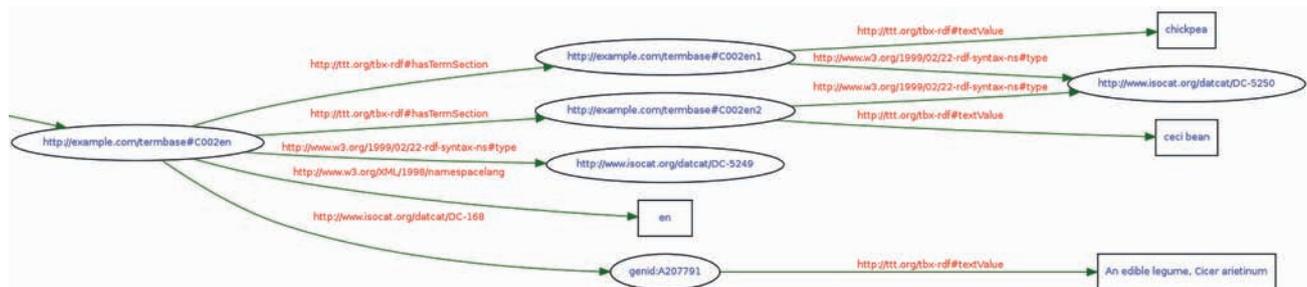


Abb. 1: TBX/RDF LangSection mit zwei TermSections und Definition

Eine bestehende Ontologie zu einem bestimmten Fachgebiet kann nicht einfach zu einer Terminologie umgewandelt werden, da sich die Art der Beziehungen zwischen den Begriffen sowie die Begriffsbildung in beiden Ressourcen unterscheiden [2]. Weiters kann eine bestehende Terminologie in XML oder TBX/RDF nicht einfach in eine formale (axiomatisierte) Ontologie übergeleitet werden, da die entsprechenden Axiome weder in XML noch in einer wörtlichen Übersetzung von XML nach RDF vorhanden sind. Dazu müssen die Dialekte erst formal in einer OWL-Ontologie definiert werden. Eine weitgehende Akzeptanz von TBX/RDF würde die Darstellung der Termbanken in Logik (OWL Ontologien) erleichtern und bietet einen ersten wichtigen Schritt in Richtung Kompatibilität des SW und Terminologien.

Interoperabilität und Wiederverwendbarkeit

Stellt man strukturierte terminologische Daten in semantischen Technologien dar, erleichtert das den konfliktfreien Austausch dieser Daten zwischen unterschiedlichen TBX-Dialekten, aber auch über Termbanken hinweg. Änderungen dieser Dialekte können daher leichter in verschiedenen TBX/RDF-Ressourcen übernommen werden. Da ISO eine Aktualisierung des TBX-Standards plant, stellt diese vereinfachte Integration und Adaptierung bestehender Ressourcen eine gute Lösung für den Übergang dar. Weder TBX noch TBX/RDF sind systemgebunden, weshalb Ressourcen in diesen Formaten von jedem Benutzer und jeder Natural Language Processing (NLP)-Applikation wiederverwendet werden können. Durch die Anbindung von TBX an das SW werden die Integration und die automatisierte Verarbeitung verschiedener Ressourcen durch die zahlreichen verfügbaren SW Anwendungen noch zusätzlich vereinfacht.

TBX ist nicht das einzige Datenmodell, das in RDF implementiert

werden kann. Andere RDF-Benutzer möchten vielleicht ein lexikographisches Modell darstellen, wie etwa jenes in OLIF, das primär nach Lexemen unterteilt ist und Bedeutungen darunter gruppiert [4]. Besteht eine grundsätzliche Übereinstimmung von zwei Modellen in RDF, so kann diese in einem Reasoner programmiert werden, damit Schlussfolgerungen eines Modells aus den Schlussfolgerungen des anderen Modells abgeleitet werden können. Dadurch erhöht sich die Verfügbarkeit von Daten und die Notwendigkeit für die Entwicklung benutzerdefinierter Konvertierungssoftware, wie etwa jene für das TBX-Glossary Projekt [4], wird reduziert.

Literatur

- [1] Bizer, Christian; Heath, Tom; Berners-Lee, Tim (2009): Linked Data – The Story So Far. In: International Journal on Semantic Web and Information Systems, Vol. 5 (Issue 3), 1-22.
- [2] Declerck, Thierry; Gromann, Dagmar (2012): Combining three Ways of Conveying Knowledge: Modularization of Domain, Terminological, and Linguistic Knowledge in Ontologies. In: Proceedings of the 6th International Workshop on Modular Ontologies (WoMO) 2012: Graz, 28-39.
- [3] Melby, Alan K. (2012): Terminology in the age of multilingual corpora. In: The Journal of Specialised Translation, Issue 18.
- [4] Wright, Sue Ellen; Melby, Alan K.; Rasmussen, Nathan; Warburton, Kara (2010): TBX Glossary: A Crosswalk between Termbase and Lexbase Formats. In: 2010 Association for Machine Translation in the Americas (AMTA) Conference.



Dagmar Gromann

ist wissenschaftliche Mitarbeiterin an der Wirtschaftsuniversität Wien und verfasst ihre

durch Gerhard Budin betreute Dissertation zur Schnittstelle terminologischer und semantischer Technologien mit einem besonderen Schwerpunkt auf Ontologien. Sie hat ein Magisterstudium der Anglistik/Amerikanistik an der Karl-Franzens-Universität Graz sowie ein Bachelorstudium der Wirtschaftsinformatik an der Universität Wien abgeschlossen.

Kontaktadresse

Wirtschaftsuniversität Wien
dgromann@wu.ac.at
www.wu.ac.at



Alan K. Melby

ist Professor für Linguistik am Provo-Campus der Brigham Young University. Er arbeitet seit

den 1980ern intensiv an der Entwicklung von Austauschformaten im Bereich der Terminologie mit, einschließlich der Version P3 der Text Encoding Initiative Richtlinien und terminologischer ISO-Standards in TC37.

Kontaktadresse

Department of Linguistics and English Language
Brigham Young University
akmtrg@byu.edu
www.byu.edu



Nathan Rasmussen

hat Computerlinguistik studiert und promoviert derzeit am Institut für Sprachwissenschaften

der Ohio State University. Er arbeitet bereits seit 2008 auf dem Gebiet der Terminologie und im Bereich des TBX-Standards.

Kontaktadresse

The Ohio State University
nrass@ling.ohio-state.edu

BOOST YOUR MULTITERM



quickTerm ©

MultiTerm + quickTerm
= Term Lifecycle

Terminology Workflow · Approval and Translation of Terminology · Change Management
Term Request Workflow · Access Through Windows, Browser, Tablet, Smartphone
Social Terminology · Term of the Week
Term Quiz · Liked and Recommended Entries

Interdisziplinäre Terminologie

Über die Konzeption einer Ontologie für den Exzellenzcluster „Bild Wissen Gestaltung“

Christian Stein

Im November ist an der Humboldt Universität zu Berlin der Exzellenzcluster „Bild Wissen Gestaltung“ gestartet, der sich selbst als „interdisziplinäres Labor“ bezeichnet. Die Idee hinter diesem Großprojekt ist es, unterschiedlichste wissenschaftliche Disziplinen in ungewöhnlichen und neuen Zusammenstellungen an Fragen unserer Zeit arbeiten zu lassen. Dabei kommen nicht nur Geistes-, Natur- und Technikwissenschaften zusammen, sondern erstmals gleichberechtigt auch Designer und Gestalter. Insgesamt sind über 25 Disziplinen beteiligt. So arbeiten in einzelnen Basisprojekten beispielsweise Architekten, Kunsthistoriker, Informatiker, Chemiker, Kulturwissenschaftler und Interaction Designer zusammen an einer gemeinsamen Fragestellung.

Mit zu Hochzeiten über 200 Mitarbeitern wurde mit dem Exzellenzcluster eine Forschungskonfiguration geschaffen, die einzigartig ist und Gelegenheit bietet, das Zusammenarbeiten und die Kommunikationsstrukturen in solchen massiv interdisziplinären Konstellationen zu untersuchen. Darin liegt eines der Hauptinteressen des Clusters: Was ist Interdisziplinarität? Wo funktioniert sie tatsächlich? Und wo nicht? Wann ist sie wirklich hilfreich und zielführend und wann verkommt sie zu einem Schlagwort? Schließlich: Wie kann man das Gelingen der Kommunikation zwischen den Disziplinen sicherstellen bzw. verbessern? Damit stehen eine genuin terminologische Fragestellung und gleichermaßen eine extreme Kommunikationssituation im Zentrum dieses Forschungsgrößprojektes.

Die Erfahrungen des ersten halben Jahres Clusterarbeit zeigen, dass Terminologiediskussionen einen außergewöhnlich großen Raum einnehmen. So ergibt es sich sehr häufig, dass verwendete Benennungen von anderen Disziplinen aufgegriffen, diskutiert und mit disziplinspezifischen Verständnisweisen und Assoziationen angereichert werden. Dabei ist ein generelles Bemühen spürbar,

in höheren Abstraktionsschichten wieder einen gemeinsamen, überdisziplinären Bedeutungskern zu finden, der die jeweiligen disziplinspezifischen Begriffe inkludiert. Dabei werden aktiv Autohyponyme konstruiert (Termini, deren Benennungen identisch mit den Benennungen ihrer Oberbegriffe sind, beispielsweise ist „Tag“ auch der Oberbegriff für „Tag und Nacht“). Diese Autohyponyme entspringen dem Bestreben, einem gemeinsamen Verständnis auf sprachlicher Ebene Gestalt zu verleihen, können dabei jedoch sehr abstrakt werden und sind ohne ihre disziplinär eingeschränkten Bedeutungen nicht wirklich handhabbar. Begriffsinhalt und -umfang werden dann sehr groß, sodass im Cluster von „großen Begriffen“ gesprochen wird. Beispielsweise wurde der Begriff zu der Benennung „Bild“ so ausgeweitet, dass mathematische, informatische, kulturwissenschaftliche, kunsthistorische und biologische Perspektiven einbezogen wurden. Ein Bild wurde so vorläufig als „virtuelles oder physisches Objekt mit zumindest minimaler intentionaler menschlicher Modifikation“ definiert. Unter diese Definition fällt der Faustkeil ebenso wie die bildgebenden Verfahren eines Elektronenrastermikroskops.

Auch wenn diese Bemühungen um große Begriffe abstrakt erscheinen mögen, so haben sie sich dennoch für den interdisziplinären Dialog als wertvoll erwiesen. Dabei geht es tatsächlich weniger darum, ein eindeutiges Verständnis zu erzeugen, als die eigenen Denkstrukturen und Perspektiven produktiv aufzubrechen und in einem gemeinsamen Gesamtbild zu verorten. In vielen Fällen hat diese Herangehensweise nach der ersten Irritation tatsächlich zu einem besseren Verständnis des eigenen und fremden Denkens geführt. Selten jedoch konnte sich bisher auf eine spezifische Definition geeinigt werden, die alle Beteiligten vollständig unterstützen. Auch das kann als Terminologiearbeit verstanden werden – allerdings mit einer völlig anderen Zielsetzung als in der klassischen Terminologiearbeit. Dennoch bleibt es das Ziel, auch dieses Arbeiten an Terminologie zu formalisieren, zu ordnen und beschreibbar zu machen.

Die Basisprojekte „Virtuelle und reale Architektur des Wissens“ und „Shaping Knowledge“ setzen sich mit den Möglichkeiten einer solchen Formalisierung auseinander. Als besonders hilfreich haben sich dabei die Erkenntnisse aus dem *iglos*-Projekt (Akronym für „intelligentes Glossar“)

Content Optimization Software

Wir bieten:

- › Rechtschreib- und Grammatikprüfung
- › Stil- und Terminologieprüfung
- › Multilinguale Terminologieverwaltung
- › Wiederverwendbare Textmodule
- › Leichte Bedienbarkeit

Ihr Nutzen:

- › Reduzierte Übersetzungskosten
- › Präzise und konsistente Inhaltserstellung
- › Suchmaschinenoptimierte Inhalte
- › Automatisiertes Reporting



Ihre Inhalte,
nur besser

der Technischen Universität Braunschweig erwiesen, mit dem der Exzellenzcluster kooperiert. Das *iglos*-Team hat jahrelang die Kommunikationsprozesse in interdisziplinären Teams im ingenieurwissenschaftlichen Kontext untersucht und darauf basierend ein eigenes Terminologiemanagementsystem entwickelt (www.iglos.de). Dieses setzt auf terminologische Ontologien statt der reinen Sammlung und Listung von Benennungen, Definitionen und ein paar Metadaten. Ontologien sind vielfach vernetzte Wissensstrukturen, die auf Basis sehr einfacher Ausdrücke komplexe Zusammenhänge abbilden können.

Die grundlegende Struktur, die allen Ontologien zugrunde liegt, ist das Tripel, also die Verbindung aus einem Subjekt, einem Prädikat und einem Objekt. Aus der Kombination von vielen verschiedenen dieser Tripel lassen sich dann komplexe Aussagen treffen und diese mit existierendem Wissen verbinden. Eine verbreitete Möglichkeit, Ontologien zu beschreiben sind die beiden Ontologiesprachen RDF (Resource Description Framework) und OWL (Web Ontology Language), die beide vom W3C (World Wide Web Consortium) standardisiert sind. In diesen Sprachen liegen viele Milliarden Tripel im Web zur rechnergestützten Verarbeitung bereit. So stellt beispielsweise die DBpedia die formal strukturierten Daten der Wikipedia maschinenlesbar und kostenlos in RDF bereit [1]. Mit diesen Daten lassen sich auch selbst erstellte Ontologien relativ einfach verbinden, um zusätzliche Informationen bereitzustellen. In einer terminologischen Ontologie könnte man also beispielsweise Aussagen treffen wie die folgende, hier in der Turtle-RDF-Syntax:

```
ex:animals rdf:type skos:Concept .  
ex:animals skos:prefLabel „animals“@en .  
ex:animals skos:altLabel „creatures“@en .  
ex:animals skos:prefLabel „animaux“@fr .  
ex:animals skos:altLabel „créatures“@fr .
```

Durch „rdf:type“ und dessen Verbindung mit „skos:Concept“ wird

ausgesagt, dass es einen Begriff „ex:animals“ gibt. Dieser Begriff hat im Englischen („@en“-Suffix) die Vorzugsbenennung „animals“ und die alternative Benennung „creatures“. Im Französischen (@fr-Suffix) ist die Vorzugsbenennung „animaux“ und die alternative Benennung „créatures“. Über diese einfachen Zusammenhänge hinaus lassen sich noch beliebig viele weitere Prädikate nutzen oder selbst definieren. Dies ist einer der großen Vorteile von Ontologien: Sie zu ergänzen, zu ändern und zu erweitern ist wesentlich einfacher als bei fest definierten Datenformaten. Außerdem sind die Daten nicht nur von Menschen, sondern auch von Computern versteh- und berechenbar. Das *iglos*-Projekt hat über 50 weitere Relationstypen gefunden und definiert, die sich in verschiedenen Fällen als sinnvoll erwiesen haben. Dazu gehören beispielsweise Abkürzungen, falsche Freunde, Ober- und Unterbegriffe, Kontext, Anwendungsgebiet, Polysemie und Homonymie, aber auch Bestandsbeziehungen, Input-Output-Folgen, Weisungs- und Besitzhierarchien, Verortung, funktionale Abhängigkeit, gemeinsame Verwendung, Instanzierung und viele mehr.

So wird es möglich, die semantischen Zusammenhänge von terminologischen Einträgen als Netzwerk zu modellieren und zu visualisieren, die für ein tatsächliches Begriffsverstehen oft unerlässlich sind. Ein solches semantisches Netz ist auch in der Lage, verschiedene Verständnisräume zu modellieren, die sich nicht ohne Weiteres vereinheitlichen lassen. Das gleiche Problem ist bei der Terminologiearbeit in Unternehmen bereits seit Langem bekannt: Eine Unternehmensterminologie kann im Idealfall in sich konsistent sein, wenn eine zentrale Terminologieabteilung die Kontrolle und Pflege übernimmt. Bei unternehmensübergreifenden Projekten, wie sie immer häufiger werden, müssten die Terminologien der beteiligten Unternehmen eigentlich vereinheitlicht bzw. kombiniert werden, was

in der Praxis meist viel zu aufwendig ist. Bereits in diesem Szenario gibt es also Verständnis- bzw. Terminologieräume, die gleichberechtigt und unter Umständen widersprüchlich nebeneinanderstehen. Eine ganz ähnliche Situation ist im Exzellenzcluster in potenziert Form vorhanden: viele Akteure, viele Disziplinen und viele verschiedene Terminologien.

Für das Erforschen dieser Kommunikationssituation wurde daher die Entscheidung getroffen, eine Cluster-Ontologie aufzubauen, die mehr als nur die rein terminologische Ebene berücksichtigt. Dabei werden Sprachkonstrukte aus bestehenden Ontologien wie beispielsweise Dublin Core, SKOS, FOAF und CIDOC CRM wiederverwendet. Eine wichtige Überlegung dabei ist, dass sich die Adaption und Verwendungsweise von Terminologie nur dann erfassen lässt, wenn man sich ihre Anwendung durch Personen oder in Quellen und ihre Kookkurrenzen, also das (wiederholte) gemeinsame Auftreten verschiedener Termini in Texten, in diesen Umgebungen ansieht. Die Beschreibung von Verständnisräumen erfordert es, die reale Anwendung von Terminologie zu untersuchen und nicht nur die gewollte: Wer verwendet sie? Wer kommuniziert mit wem damit? Wel-

che Quellen verwenden sie? Welche Person rezipiert welche Quellen? Wie sind diese Quellen verbunden? Welche Themen und Communities sind durch welche Terminologie charakterisierbar? Und schließlich: Wer versteht wen unter welchen Umständen überhaupt richtig? Wo kann ein Forscher für ihn interessante Quellen finden? Und mit wem kann er sich über bestimmte Themen unterhalten? Diese Fragen und viele mehr haben die Gestaltung der Ontologie wesentlich beeinflusst und die zentrale Rolle der Terminologie in dieser gefestigt.

Vier zentrale Entitäten organisieren gemäß den oben genannten Überlegungen das Schema der Ontologie: *Personen, Quellen, Themen* und *Termini*. Unter Personen sind zunächst die konkreten Mitarbeiter des Clusters zu verstehen; hier werden aber auch angegliederte Personen, virtuelle Akteure sowie Personengruppen, Projekte und Teams modelliert. So entsteht zunächst ein Personennetzwerk, das Informationen darüber enthält, wer in welchen Organisationseinheiten mit wem steht. Die kontinuierlich durchgeführte Selbstbeobachtung der Clustermitarbeiter erlaubt aber auch eine Modellierung der Kommunikationsverbindungen zwischen den Personen. So kann angegeben werden, wer wie

umfangreich mit wem kommuniziert und auf welchem Wege. Diese Informationen werden teils automatisch mit technischen Mitteln und teils durch empirische Beobachtung erfasst.

Aber nicht nur Personen und deren Verbindungen untereinander werden modelliert, sondern auch die Objekte, mit denen sie umgehen. Diese sind in der Ontologie als „Quellen“ gefasst. Quellen können beispielsweise Bücher oder Zeitschriften sein, die Personen lesen. Genauso kann es sich aber auch um produzierte Texte handeln. Neben Texten werden als Quellen auch Zeichnungen, Fotos und Bilder, Modelle und andere Datenformate gefasst. Eine Quelle ist also ein beliebiges, physisches oder digitales Objekt, mit dem interagiert wird. Quellen können auch Beziehungen untereinander aufweisen, beispielsweise im Sinne von Zitation oder Beeinflussung. Auch diese Beziehungen können in der Ontologie modelliert werden.

Der Sprung auf die inhaltliche Ebene geschieht mithilfe der „Themen“. Themen fungieren als zusammenfassende Charakterisierung von zusammenhängenden Inhalten. Typische Themen sind beispielsweise Sachgebiete, Domänen, Ausbildungsgänge und Ähnliches. Themen können in Baumstrukturen aufgebaut werden,



Jetzt mit
**Acolada
UniTerm**

Finalyser TERM CHECK

Effizienter Terminologie-Check zum günstigen Preis.

Ob Bedienungsanleitung, Dokumentation oder Katalog – überall spielt die konsequente, einheitliche Verwendung von Begriffen die entscheidende Rolle für das richtige Verständnis komplexer Sachverhalte. Finalyser TERM CHECK ist die wirkungsvolle und kostengünstige Lösung, mit der Sie in allen Dokumenten ganz einfach die Konsistenz der Terminologie sicherstellen können. TERM CHECK integriert starke Funktionen zur Terminologieüberprüfung direkt in Ihre Standardsoftware. Alle Mitarbeiter nutzen dabei dieselbe zentrale Excel- oder Acolada UniTerm Terminologiedatenbank. Mit TERM CHECK gewährleisten Sie die einheitliche Verwendung von Begrifflichkeiten und sparen viel Zeit und Geld. Am besten gleich herunterladen: die kostenlose TERM CHECK Testversion!

Fragen Sie die Terminologie-Experten von SQUIDDS, was sie für Sie tun können: **+49 (0) 911 / 21 53 47 - 22**

www.finalyser.de
www.squidds.de
www.workflowblog.de

SQUIDDS
PEOPLE.PRODUCTS.PASSION.

sodass sich allgemeinere und speziellere Themen anlegen und miteinander verbinden lassen. So kann man Themen zur Charakterisierung von Kompetenzen von Personen verwenden. Oder sie werden genutzt, um Quellen zu klassifizieren und so Aussagen über die darin behandelten Inhalte zu treffen. Dabei ist es wahrscheinlich, dass Personen oder Quellen jeweils mehrere Themen zugeordnet bekommen. Die Flexibilität der Ontologie ermöglicht es, verschiedene existierende Klassifizierungssysteme zu integrieren. Hier wird beispielsweise auf das Open Directory [2] zurückgegriffen werden, das eine der umfangreichsten Themenklassifikationen beinhaltet.

Der aus terminologischer Perspektive vielleicht interessanteste Teil der Ontologie ist der des Terminus. Unter Terminus wird im Folgenden „das zusammengehörige Paar aus einem Begriff und seiner Benennung als Elemente einer Terminologie“ verstanden, wie es die DIN 2342:1992 definiert. Die Uminterpretation der Benennung „Terminus“ in der DIN 2342:2011 als Synonym zu „Benennung“ wird hier nicht gefolgt. In der Ontologie des Clusters werden Termini analog zu der Modellierung des *iglos*-Systems vorgenommen [3]. Jeder Terminus wird als die Kombination von Benennung, Definition und Varietät modelliert. Eine Varietät ist dabei der Verständnisraum, innerhalb dessen der Terminus verstanden wird. Eine Varietät kann beispielsweise eine bestimmte Disziplin wie die Informatik sein, eine Unterdisziplin wie die Graphentheorie, eine Unternehmenssprache, bis hin zur spezifischen Projektsprache, deren Termini nur noch ein eng umrissener Personenkreis im gleichen Sinne versteht. In einem interdisziplinären Szenario wie dem des Clusters lässt sich eine Monosemierung im Sinne einer für alle verbindlichen Festlegung aus den oben geschilderten Gründen kaum durchführen. Dies ist allerdings auch in weit weniger komplexen Szenarien mitunter schwierig. Man denke im Unternehmenskontext an die heterogenen An-

forderungen an Vorzugsbenennungen aus den Marketingabteilungen und den F&E-Abteilungen. Da Benennungen sehr häufig in unterschiedlichen Varietäten vorkommen, dort jeweils jedoch mit unterschiedlichen Begriffen verbunden werden, erlaubt die geschilderte Modellierung eine Strukturierung auch dieser Fälle.

In der Ontologie des Clusters werden Termini auf unterschiedlichen Ebenen modelliert. Zum einen wird den einzelnen Basisprojekten die Möglichkeit gegeben, terminologische Festlegungen für ihre Bereiche vorzunehmen und sich über die anderer zu informieren. Dazu wird wahrscheinlich ebenfalls das *iglos*-System zum Einsatz kommen. Zum anderen werden in größeren Arbeitsgruppen die oben genannten „großen Begriffe“ diskutiert und theoretisiert werden. Die in diesem Rahmen zu diskutierenden Benennungen beginnen mit dem Namen des Clusters selbst – „Bild, Wissen, Gestaltung“ und „Interdisziplinarität“. Andere wichtige große Begriffe formieren sich um die Benennungen „Struktur“, „Modell“, „Architektur“, „Raum“ oder „Code“. Es leuchtet schnell ein, dass diese Bemühungen zwar auch kurze und prägnante Definitionen anstreben, diese aber zwingend umfangreicherer Erläuterungen bedürfen. Darüber hinaus reicht es in diesen Fällen nicht aus, eine einzelne Definition zu finden – vielmehr ist es erforderlich eine ganze Menge dazugehöriger Termini zusammenzubringen, zu strukturieren und zu definieren. Für ein solches Vorgehen sind Ontologien besonders geeignet, da sie die Verbindungen und Zusammenhänge zwischen den Termini modellieren können. Auch für diese Aufgabe wird wahrscheinlich das *iglos*-System zum Einsatz kommen, da es den Nutzer mit visuellen Modellierungswerkzeugen bei der Strukturierung unterstützen kann. Der dritte Anwendungsbereich erstreckt sich auf die Analyse tatsächlicher Terminologieverwendung in den Quellen des Clusters. Ein Großteil der verwendeten Textquellen wird wenn nötig

digitalisiert und zentral verwaltet werden. Dazu kommt das Open-Source-System *Zotero* [4] zum Einsatz, das eine direkte RDF-Schnittstelle besitzt. So kann aus der Ontologie direkt auf die Quellen verlinkt werden. Mit verschiedenen Text-Mining-Methoden werden die Digitalisate dann untersucht. Dazu gehört beispielsweise die Topic Detection, die es ermöglicht, Quellen bestimmten Themen zuzuordnen und den Grad der Relevanz der Quelle für ein Thema zu berechnen. Den Themen in der Ontologie sind dazu Termini zugeordnet, die signifikant für das jeweilige Thema sind. Die Ontologie lernt über die Analysen automatisch neue Termini, die zu einem gegebenen Thema gehören. Alle diese Entitäten und die Relationen zwischen ihnen werden je nach Bedarf in Unterklassen spezialisiert. Das in Abb. 1 dargestellte Modell zeigt nur die abstrakten Basisentitäten und die allgemeinen Relationen zwischen ihnen.

Hier ist jedoch gut zu sehen, wie vielfältig vernetzt sich bereits auf dieser Ebene eine Forschungslandschaft präsentiert: „Kommunikation“ findet zwischen Personen statt, diese haben „Kompetenzen“ zu bestimmten Themen und sprechen bzw. verstehen eine „persönliche“ Auswahl von „Terminologie“. Sie arbeiten mit „Quellen“, indem sie sie rezipieren oder produzieren. „Quellen“ haben implizite oder explizite „Referenzen“ auf andere Quellen, behandeln verschiedene abstrakte oder konkrete „Themen“ (→ „Exploration“) und benutzen eine eigene Auswahl an „Quellen-Terminologie“, die sich in ihnen finden lässt. Themen wiederum können sich zu häufig zusammen auftretenden Themengruppen gruppieren (→ „Frame“) und haben eine charakteristische „Terminologie“. Jeder „Terminus“ wiederum ist durch seine Relationen mit anderen, definierenden Termini „definiert“. So ergibt sich eine Struktur, die zur formalen Beschreibung von Forschungsaktivitäten dienen kann und auch dem einzelnen Forscher viel Information über die eigene Arbeitsweise zurückgibt.

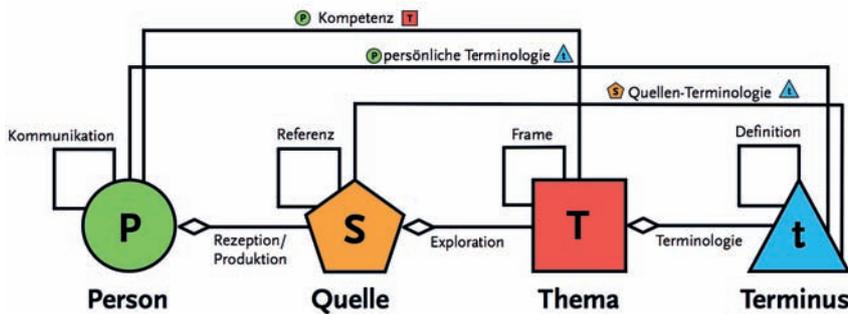


Abb. 1: Abstrakte Basisentitäten und allgemeine Relationen

In Abb. 2 sind einige angestrebte Daten zur persönlichen Information der einzelnen Forscher dargestellt. So können alle erfassten rezipierten Quellen dargestellt, aber auch Quellen vorgeschlagen werden, die ähnliche Themen behandeln und im Cluster bereits verfügbar sind. Ebenso können Personen vorgeschlagen werden, die über Kompetenzen zu solchen Themen verfügen. Zumindest in Bezug auf die Rechneraktivitäten lassen sich auch Informationen zur Arbeitsweise und Produktivität erfassen. Schließlich besteht Zugang zu der eigenen Terminologie der Person bzw. des Projekts sowie zu der der anderen. So lassen sich auch Überlappungen und unterschiedliche Definitionen erkennen.

Das Projekt ist mit der vorgeschlagenen Ontologie und den dazugehörigen Datenerfassungs- und Verarbeitungstechnologien sicherlich sehr ambitioniert und die zu erwartenden Schwierigkeiten nicht zu unterschätzen: Daten werden trotz Methodenpluralität niemals vollständig sein, die Technologien müssen zusammenarbeiten, der Datenschutz muss maximal gesichert werden und die Bereitstellung einer nutzerfreundlichen Anwendungsumgebung muss gelingen. Auf der anderen Seite ist der Cluster mit vielen gut ausgebildeten und motivierten Wissenschaftlern und Entwicklern ausgestattet und hat 5 Jahre Zeit zu zeigen, dass auf Basis einer formalisierten Selbstbeobachtung neue Erkenntnisse über die

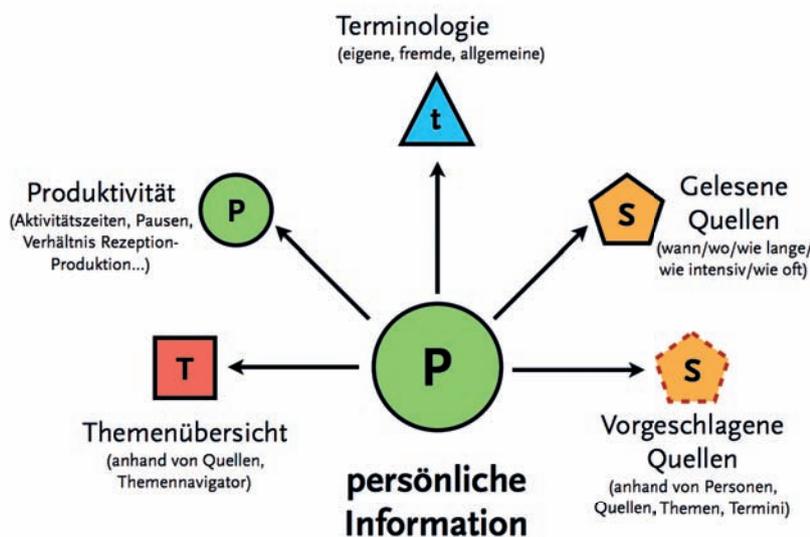


Abb. 2: Darstellung der persönlichen der beteiligten Forscher

reale Arbeitsweise gewonnen werden können. Das Ziel ist es dabei nicht, individuelle und disziplinspezifische Arbeitsweisen über einen Kamm zu scheren – im Gegenteil: Gerade die Heterogenität der Arbeitsweisen besser zu verstehen und mit einer empirischen Datenbasis zu versehen ist nach unserer Überzeugung die Voraussetzung für eine bessere, vielfältigere und innovativere Forschungsarbeit. Wenn sich die Rahmenbedingungen interdisziplinärer Forschung auf Basis der gewonnenen Erkenntnisse auch nur etwas verbessern lassen, ist bereits viel gewonnen – denn interdisziplinäre Strukturen werden immer häufiger und nichts verschwendet so viele Ressourcen wie ungünstige Bedingungen und falsche Annahmen.

Am 1. Juni fand die offizielle Eröffnung des Exzellenzclusters „Bild Wissen Gestaltung“ statt. Hunderte Besucher haben die Projekte begutachtet, sich informiert und angeregt diskutiert. Nun heißt es, die Pläne umzusetzen. Was daraus wird und wie sich die Arbeit entwickelt, kann man unter www.interdisciplinary-laboratory.hu-berlin.de verfolgen.

Literatur

- [1] DBpedia (2013). <http://dbpedia.org>.
- [2] dmoz – open directory project (2013). <http://www.dmoz.org>.
- [3] Schnieder, Lars; Stein, Christian; Schielke, Arno (2011): Terminologiemanagementsysteme der nächsten Generation – Schlüssel für den Fachwortschatz. In: eDITion – Fachzeitschrift für Terminologie. Ausgabe 1-2011, Köln: SDK, S. 26-31.
- [4] zotero (2013). <http://www.zotero.org>.



Dr. Christian Stein hat an der Technischen Universität Braunschweig Germanistik und Informatik studiert und arbeitet derzeit an der Humboldt Universität zu Berlin. Forschungsschwerpunkte liegen in den Bereichen Informations- und Wissensmanagement, Semantic Web und Semiotik.

Kontaktadresse

Humboldt Universität zu Wien
christian.stein@hu-berlin.de

Übersetzungsorientierte Terminografie – neue und adaptierte Datenkategorien für ISOCat

Georg Löckinger

ISOCat, frei verfügbar unter www.isocat.org, ist ein internationales Verzeichnis von Datenkategorien, das für den Aufbau und die Gestaltung verschiedenster Sprachressourcen genutzt werden kann. Den historischen Kern von ISOCat bilden Datenkategorien für die übersetzungsorientierte Terminografie. Im vorliegenden Beitrag werden neue und adaptierte Datenkategorien vorgeschlagen, die sich auf aktuelle Forschungsergebnisse stützen.

Einleitung

Professionelle Übersetzer(innen) benötigen für ihre Tätigkeit viele verschiedene Arten fachsprachlicher Informationen. Diese wurden im Idealfall anhand der entsprechenden Anforderungen ausgewählt und aufbereitet, bevor sie professionellen Übersetzer(inne)n zur Verfügung gestellt werden. Im vorliegenden Beitrag soll, aufbauend auf Forschungsergebnissen [1], skizziert werden, welche neuen bzw. adaptierten Datenkategorien in ISOCat für die übersetzungsorientierte Terminografie erforderlich sind, und wie diese anhand eines Modells neuartiger übersetzungsorientierter Fachwörterbücher eingesetzt werden könnten. Somit sind die folgenden Ausführungen auch als Anstoß zur wissenschaftlich fundierten Weiterentwicklung übersetzerischer Hilfsmittel im Hinblick auf die konkret erforderlichen fachsprachlichen Datenkategorien zu verstehen.

Modell neuartiger übersetzungsorientierter Fachwörterbücher

Die im nächsten Abschnitt beschriebenen Datenkategorien leiten sich von einem abstrakten Modell neuartiger übersetzungsorientierter Fachwörterbücher ab (s. Abb. 1). Die Grobstruktur dieses Modells geht auf das in der ISO 16642:2003 [2] festgelegte Metamodell zurück, das als allgemeine Grundlage für die Modellierung und

Gestaltung terminologischer Datenbestände herangezogen werden kann.

Vom Modell abgeleitete Datenkategorien

Während die meisten Bestandteile des Modells in Abb. 1 (beispielsweise „Definition“, „Anmerkung“ usw.) bereits jetzt semantisch und funktionell brauchbare Datenkategorie-Entsprechungen in ISOCat haben, treten bei der Ableitung der erforderlichen Datenkategorien vom Modell auch einige Lücken zutage.

Die betreffenden Bestandteile innerhalb des Kästchens „Zusätzliche Informationen“ sind folgende: „Meta-Information“, „Alphabetischer Auszug“, „Bibliografische Angaben“, „Textkorpus“ sowie „Multimedialer Inhalt“ (zusätzlich auch auf Eintragungsebene) und „Begriffsplan“. Weiters handelt es sich um die Bestandteile „Enzyklopädische Angabe“ (Eintragungsebene, Sprachebene und Benennungsebene) und „Grammatikalische Angabe“ (Benennungsebene).

„Meta-Information“

„Meta-Information“ entspricht einer Datenkategorie, die den jeweiligen terminologischen Datenbestand anhand der für professionelle Übersetzer(innen) bedeutenden Eigenschaften beschreibt, beispielsweise in Bezug auf:

- die abgedeckten Sprachen

- die Methoden, nach denen die fachlexikografische/terminografische Arbeit gestaltet wurde
- das zugrunde liegende Textkorpus oder die zugrunde liegenden Textkorpora (falls zutreffend)
- die Zielgruppe(n) des terminologischen Datenbestandes
- den Zweck des terminologischen Datenbestandes
- die Eingrenzung des gewählten Fachgebiets oder Geltungsbereichs
- den Umfang
- die Mikro-, Medio- und Makrostruktur
- die effiziente Nutzung sowie
- die Kompatibilität mit anderen Sprachressourcen [1].

In ISOCat gibt es derzeit keine entsprechende Datenkategorie. Allerdings sollten die ISOCat-Themengruppen „Lexicography“ und „Terminography“ um eine solche ergänzt werden: In vielen Fällen fehlen diese Informationen in lexikografischen oder terminologischen Nachschlagewerken, sind jedoch von überragender Bedeutung für professionelle Übersetzer(innen). Diese müssen regelmäßig entscheiden, welche der vielen verfügbaren Sprachressourcen sie am sinnvollsten als Hilfsmittel zur Lösung eines terminologischen Problems heranziehen sollen. Die oben genannten Informationen können die Entscheidung wesentlich erleichtern; umgekehrt ergeben sich oftmals unnötige und unproduktive Leerläufe in der Recher-

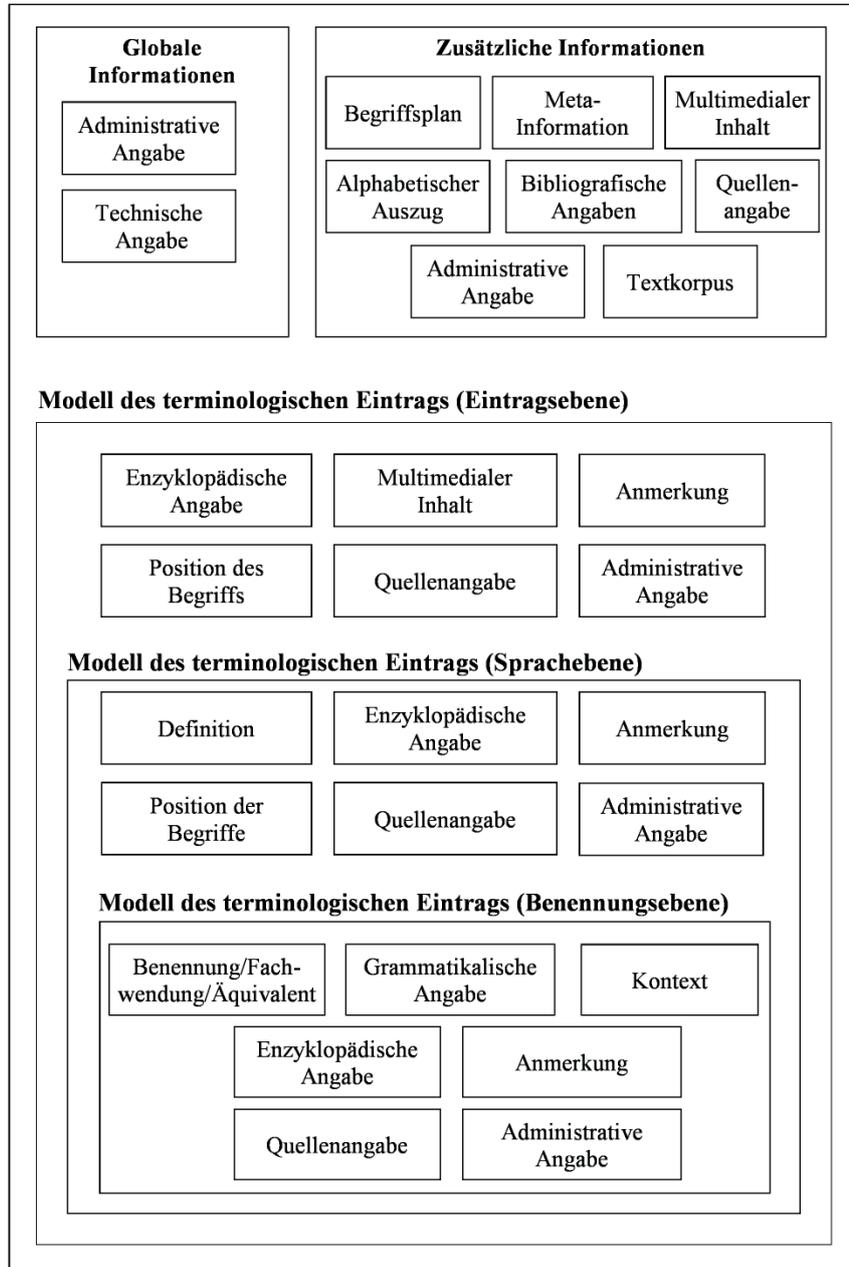
Modell neuartiger übersetzungsorientierter Fachwörterbücher**(Ebene: terminologische Datenbasis)**

Abb. 1: Modifiziertes abstraktes Modell neuartiger übersetzungsorientierter Fachwörterbücher; nicht formalisierte Darstellung [3]

che, weil diese Informationen nicht zur Verfügung stehen, was auch heute noch ein Manko vieler Sprachressourcen darstellt.

In ISOcat könnten diese Überlegungen in Anlehnung an die ISO 12620:2009 [4] durch Einführung einer neuen Datenkategorie */meta information/* (vom Typ „container“) umgesetzt werden, die für die Umsetzung des oben abgebildeten Modells

folgende weitere Datenkategorien beinhalten müsste:

- */language identifier/* (besteht bereits, s. <http://www.isocat.org/datcat/DC-279>)
- */compilation method/* (neu, vom Typ „open data category“)
- */note/* (besteht bereits, s. <http://www.isocat.org/datcat/DC-382>, würde zur Beschreibung des Text-

Wie geht's, wo steht's?

Nicht weniger als diese zwei Fragen stellen sich uns immer wieder, wenn wir die **technische Dokumentation** komplexer Produkte **übersichtlich** und **verständlich** gestalten und **präzise** in Worte fassen.

Wir unterstützen Sie auf den Gebieten der **technischen Redaktion** sowie der **Risikobeurteilung** und entwickeln für Sie **Software zur Prozessoptimierung** – wie unsere Terminologieverwaltungssoftware TippyTerm®.



SysKon
Systemlösungen GmbH

www.SysKon.com

korpus bzw. der Textkorpora dienen)

- */target group/* (neu, vom Typ „open data category“)
- */purpose/* (neu, vom Typ „open data category“, könnte durch Überarbeitung einer gleichnamigen bestehenden Datenkategorie festgelegt werden, s. <http://www.isocat.org/datcat/DC-1657>)
- */subject field/* oder */domain/* (besteht bereits, s. <http://www.isocat.org/datcat/DC-489>, müsste eventuell adaptiert werden)
- */number of [terminological entries/lexical entries/head words/...]/* (neu, vom Typ „open data category“)
- */structure/* (neu, vom Typ „open data category“)
- */instructions for use/* (neu, vom Typ „open data category“)
- */compatibility/* (neu, vom Typ „open data category“)

Die einzelnen neuen Datenkategorien könnten in ISOcat folgendermaßen definiert werden:

- */compilation method/*
way of creating and designing a language resource

- */target group/*
audience for which a language resource has been created and designed
- */purpose/*
activity for which a language resource has been created and designed
- */number of [terminological entries/lexical entries/head words/...]/*
size of a language resource, expressed by figures that describe how many lexical or terminological units the language resource contains
- */structure/*
arrangement of information in a language resource
- */instructions for use/*
practical information on how to exploit a language resource
- */compatibility/*
property that expresses how a language resource can be combined with external content and tools

„Textkorpus“

Während in ISOcat eine Datenkategorie namens */text corpus/* besteht, gibt es keine davon umfasste Daten-

kategorie, die sich auf die einzelnen Dateien eines Textkorpus bezieht. Hier bietet es sich an, eine neue Datenkategorie */text corpus file/* vom Typ „open data category“ sinnvoll in ISOcat einzuführen. Dies wäre beispielsweise möglich unter Zugrundelegung der bestehenden Datenkategorie */corpus trace/* (s. <http://www.isocat.org/datcat/DC-161>), deren Definition zu diesem Zweck jedoch angepasst werden müsste. Für die Umsetzung des Modells neuartiger übersetzungsorientierter Fachwörterbücher sollten sowohl */text corpus/* als auch */text corpus file/* u. a. mit der bestehenden Datenkategorie */language identifier/* (s. <http://www.isocat.org/datcat/DC-279>) beschrieben werden können.

„Alphabetischer Auszug“

Derzeit beinhaltet ISOcat keine Datenkategorie für alphabetisch geordnete einsprachige oder mehrsprachige Listen mit Benennungen, die beispielsweise aus vollständigen terminologischen Einträgen extrahiert wurden. Daher sollte eine neue Datenkategorie */alphabetical extract/* vom Typ „open data category“ in ISOcat eingeführt werden. Für die Zwecke



**Sie konzentrieren sich auf die Inhalte.
Wir bieten Ihnen die Infrastruktur.**

Terminologiemanagement auf die einfache Art.

des oben abgebildeten Modells könnte man */alphabetical extract/* als Art von */terminology list/* [5] vom Typ „closed data category“ auffassen, eine Datenkategorie, die wiederum durch drei weitere Datenkategorien beschrieben werden könnte:

- */monolingual/* (vom Typ „simple data category“, könnte durch Überarbeitung einer gleichnamigen bestehenden Datenkategorie festgelegt werden, s. <http://www.isocat.org/datcat/DC-2980>)
- */bilingual/* (vom Typ „simple data category“, könnte ebenfalls durch Überarbeitung einer gleichnamigen bestehenden Datenkategorie festgelegt werden, s. <http://www.isocat.org/datcat/DC-2981>)
- */multilingual/* (vom Typ „simple data category“, könnte ebenfalls durch Überarbeitung einer gleichnamigen bestehenden Datenkategorie festgelegt werden, s. <http://www.isocat.org/datcat/DC-2982>).

Die zwei neuen Datenkategorien könnten in ISOcat wie folgt definiert werden:

- */terminology list/*
list of (potential) terms in one or several languages that has been compiled from texts or dedicated language resources that contain more detailed information
- */alphabetical extract/*
alphabetically ordered terminology list created on the basis of a lexical or terminological resource that holds more detailed information

Bei der Umsetzung des Modells neuartiger übersetzungsorientierter Fachwörterbücher sollten beide Datenkategorien u. a. mit der bestehenden Datenkategorie */language identifier/* (s. <http://www.isocat.org/datcat/DC-279>) beschrieben werden können. Darüber hinaus ist je nach den Eigenschaften der Inhalte (statisch/dynamisch) eine Verwendung in Kombination mit Datenkategorien vorstellbar, die in der einen oder anderen Art auf die jeweilige Sprachkombinationsrichtung Bezug nehmen.

„Multimedialer Inhalt“ und „Begriffsplan“

ISOcat beinhaltet derzeit mehrere Datenkategorien, die multimediale Inhalte beschreiben: */figure/* (s. <http://www.isocat.org/datcat/DC-229>), */audio/* (s. <http://www.isocat.org/datcat/DC-90>), */video/* (s. <http://www.isocat.org/datcat/DC-533>), */table/* (s. <http://www.isocat.org/datcat/DC-500>) und */other binary data/* (s. <http://www.isocat.org/datcat/DC-395>). Allerdings fehlt eine Entsprechung zur äußerst bedeutenden allgemeineren Datenkategorie */nontextual illustration/*, die im Datenkategorieverzeichnis der (mittlerweile zurückgezogenen) ISO 12620:1999 [6] noch enthalten war. Im Hinblick auf die übersetzungsorientierte Terminografie, für die multimediale Inhalte immer bedeutender werden, sollte diese Datenkategorie (vom Typ „closed data category“) beispielsweise mit folgender Definition und Bezeichnung in ISOcat eingeführt werden:

- */multimedia information/*
static or dynamic content that depicts objects or concepts only or partly in nontextual form

Auch fehlen in ISOcat Datenkategorien zur Beschreibung von Begriffsplänen oder Begriffsmodellen. Eine neue Datenkategorie */concept diagram/* vom Typ „simple data category“ sollte in ISOcat aufgenommen werden, genauso wie eine neue Datenkategorie */concept model/*, ebenfalls vom Typ „simple data category“. Zu diesem Zweck werden folgende Definitionen vorgeschlagen:

- */concept diagram/*
graphic representation of a concept system [7]
- */concept model/*
concept diagram formed by means of a formal language [8]

Im Zusammenhang mit dem Modell neuartiger übersetzungsorientierter Fachwörterbücher sollten alle drei genannten Datenkategorien u. a. mit der bestehenden Datenkategorie */language identifier/* (s. <http://www.isocat.org>).



70% Bis zu 70% der Fehler in der Technischen Dokumentation entstehen, weil falsche Terminologie verwendet wird.¹

85% 85% der Arbeitnehmer verwenden eine andere Benennung als Kollegen in anderen Abteilungen, obwohl sie von ein und derselben Sache sprechen.²

4-6% Terminologie macht 4% - 6% eines Textes bezogen auf die Gesamtzahl der Wörter aus.³

9 Min Suchanfragen können um bis zu 9 Minuten reduziert werden, wenn eine strukturierte, relevante Terminologiedatenbank eingesetzt wird.⁴

Weitere Informationen:
www.sdl.com/de/einheitliche-unternehmenskommunikation

1. Jörg Schütz and Rita Nübel, MULTIDOC project.
2. Tekom-Studie "Successful terminology management in companies", Straub und Schmitz, 2010.
3. Champagne
4. Calculation based empirical findings and appraisals. Silvia Cerrella Bauer.

org/datcat/DC-279) beschrieben werden können. Für die Umsetzung des Modells sind darüber hinaus auch die bereits bestehenden Datenkategorien */thesaurus/* (s. <http://www.isocat.org/datcat/DC-4040>), */ontology/* (s. <http://www.isocat.org/datcat/DC-4042>) und */taxonomy/* (s. <http://www.isocat.org/datcat/DC-4039>) von großem Interesse.

„Bibliografische Angaben“

In übersetzungsorientierten Fachwörterbüchern sollte an zentraler Stelle ein Quellenverzeichnis abgelegt sein, auf das in den einzelnen terminologischen Einträgen verwiesen werden kann. Eine entsprechende Datenkategorie ist derzeit nicht Teil von ISOcat, kann aber leicht aus dem Datenkategorieverzeichnis der ISO 12620:1999 übernommen werden:

- */bibliographic list/* (vom Typ „open data category“) a list of bibliographic entries embedded in a document [6]

Für das Modell neuartiger übersetzungsorientierter Fachwörterbücher wäre diese Datenkategorie in Kombination mit der bestehenden Datenkategorie */source/* (s. <http://www.isocat.org/datcat/DC-471>) zu verwenden, die die einzelnen Quellenverzeichniseinträge abdecken würde.

„Enzyklopädische Angabe“

Während ISOcat mehrere Datenkategorien namens */encyclopedic information .../* enthält, ist keine davon für die übersetzungsorientierte Terminografie brauchbar. Daher wird vorgeschlagen, folgende neue Datenkategorie vom Typ „open data category“ in ISOcat speziell für die Themengruppe „Terminology“ aufzunehmen; möglicherweise kann sie auch in der Themengruppe „Lexicography“ eine Lücke schließen:

- */encyclopaedic information/* domain-specific background information that relates to the concept,

term, and/or object concerned or to the relevant professional discourse and provides more detailed information than a definition

Ausgehend vom obigen Modell, könnte diese neue Datenkategorie in Kombination mit */note/* (s. <http://www.isocat.org/datcat/DC-382>) eingesetzt werden: Detaillierte fach(sprach)liche Informationen, die sich auf verschiedene terminologische Ebenen (Begriff, Benennung, Gegenstand) oder auf den betreffenden Fachdiskurs beziehen und über die Informationen in Definitionen hinausgehen, sind auch für professionelle Übersetzer(innen) von überragender Bedeutung.

„Grammatikalische Angabe“

Derzeit sind in ISOcat zum Thema Grammatik mehrere Datenkategorien zu finden, die jedoch alle semantisch nicht das beschreiben, was für die übersetzungsorientierte Termi-

Professionelle Terminologieverwaltung einfach gemacht



- Webbasiert seit 2003 für maximale Zugänglichkeit
- Grenzüberschreitendes und produktives Arbeitsumfeld durch Workflows und Grid View Editing
- Schnellere Rendite durch Integration in MS Office, XTM, SDL Trados Studio, Ontram, ...

Mehr Informationen auf interverbumtech.de



BERLIN STOCKHOLM LINKÖPING CHICAGO SINGAPUR

nografie in Bezug auf allgemeine grammatikalische Angaben hilfreich wäre. Allerdings gab es bereits in der ISO 12620:1999 eine einschlägige Datenkategorie */grammar/*, die dort als Überkategorie für */grammatical gender/*, */grammatical number/* usw. fungierte. Folglich sollte eben diese Datenkategorie in ISOcat eingepflegt werden (als „open data category“ oder als „container“):

- */grammar/*
grammatical information about a term [6]

Wenngleich es in vielen Fällen besser sein mag, mit feineren Datenkategorien wie */grammatical gender/*, */grammatical number/* usw. zu arbeiten, so käme die Verfügbarkeit einer allgemeineren Datenkategorie */grammar/* sicherlich den Bedürfnissen punktueller übersetzungsorientierter Terminologearbeit sehr entgegen. Dabei bildet */grammar/* den abstrakten Überbau für die feineren Datenkategorien */grammatical gender/*, */grammatical number/* usw., ohne jedoch ein eigenes Datenfeld bereitstellen zu müssen.

Fazit

ISOcat ist ein sehr umfangreiches internationales Verzeichnis von Datenkategorien, das jedoch aus der Sicht der übersetzungsorientierten Terminografie gewisse Lücken und Verbesserungsmöglichkeiten aufweist. Aus diesem Grund wurden die entsprechenden Datenkategorien von einem abstrakten Modell neuartiger übersetzungsorientierter Fachwörterbücher abgeleitet und zur Einführung oder Adaptierung in ISOcat vorgeschlagen.

Literatur

- [1] Löckinger, Georg (2013): Entwicklung und empirische Erprobung eines Modells neuartiger übersetzungsorientierter Fachwörterbücher. Unveröffentlichte Dissertation, Universität Wien.
- [2] ISO 16642:2003. Computer applications in terminology – Terminological markup framework. Genf: ISO.
- [3] Löckinger, Georg (2011): Von TMF in Richtung UML: in drei Schritten zu einem Modell des übersetzungsorientierten Fachwörterbuchs. In: Hedeland, Hanna; Schmidt, Thomas; Wörner, Kai (Hrsg.): Multilingual Resources and Multilingual Applications. Proceedings of the Conference of the German Society for Computational Linguistics and Language Technology (GSCL) 2011. Arbeiten zur Mehrsprachigkeit (Working Papers in Multilingualism) Folge B, Nr. 96. University of Hamburg, S. 197-201. Online verfügbar unter <http://www.corpora.uni-hamburg.de/gscl2011/downloads/AZM96.pdf>.
- [4] ISO 12620:2009. Terminology and other language and content resources – Specification of data categories and management of a Data Category Registry for language resources. Genf: ISO.
- [5] ISO 22128:2008. Terminology products and services – Overview and guidance. Genf: ISO.
- [6] ISO 12620:1999. Computer applications in Terminology – Data categories. Genf: ISO.
- [7] ISO 1087-1:2000. Terminology work – Vocabulary – Part 1: Theory and application. Genf: ISO.
- [8] ISO/DIS 24156-1:2013. Graphic notations for concept modelling in terminology work – Part 1: Guidelines for using UML notation in terminology work. Genf: ISO.



Dr. phil. Georg Löckinger ist promovierter Fachübersetzer und Terminologe für die Sprachen Deutsch,

Englisch, Russisch und Französisch. Als (Gast-)Forscher arbeitet er am Zentrum für Translationswissenschaft der Universität Wien sowie am Institut für Corpuslinguistik und Texttechnologie der Österreichischen Akademie der Wissenschaften. In seiner Forschung beschäftigt er sich vor allem mit übersetzungsorientierter Fachlexikografie/Terminografie und mit der Entwicklung modernster Sprachtechnologie-Werkzeuge.

Kontaktadresse
terminologist@gmail.com

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

DTT-Seminar
Finanzberichte nach IFRS/DRS
Eine Einführung für Übersetzer und Terminologen

© pixelio.de

Termin:
Freitag, 29.11.2013, 9 Uhr – 17.30 Uhr,
Samstag, 30.11.2013, 9 Uhr – 17.30 Uhr

Ort:
Best Western Hotel Köln
Bennoplatz 2, 51103 Köln
bestwestern-koeln.de

Zielgruppen:

- Übersetzer, Dolmetscher, Terminologen
- Sprachübergreifend arbeitende Fachleute
- Studierende im Bereich Übersetzen / Dolmetschen

Inhalte:

- Vermittlung der Grundsätze der IFRS sowie des Aufbaus von IFRS-Abschlüssen / Konzernfinanzberichten von börsennotierten deutschen Unternehmen
- Analyse der zweisprachigen Terminologie von IFRS-Abschlüssen
- Analyse der in den Standards und Interpretationen des IASB verwendeten Terminologie

Teilnahmegebühr (inkl. Seminarunterlagen, Kaffeepausen und Mittagessen):

- Teilnahme an beiden Tagen: 400 € für Nicht-Mitglieder bzw. 280 € für Mitglieder von DTT, ADÜ Nord, ASTTI, ATICOM, BDÜ, DGI, tekem, Universitas
- 190 Euro für Mitglieder DTT, ADÜ Nord, ASTTI, ATICOM, BDÜ, DGI, tekem, Universitas

Veranstalter: Deutscher Terminologie-Tag e.V.
Information bei: fortbildung@dttev.org
Anmeldeformular und Programm unter: www.dttev.org

Stand: 1.8.2013, Änderungen vorbehalten

Einfache Terminologieabstimmung mit Termflow

Nicole Keller

Termflow ist ein rein webbasiertes System, mit dem komplexe Terminologie-Workflows abgebildet und die verschiedenen Aufgaben im Abstimmungsprozess genau gesteuert und bei Bedarf automatisiert werden können. Hierbei bedarf es keiner Installation oder Pflege vonseiten des Systemnutzers, da diese Aufgaben vom Systemanbieter Transline übernommen werden.

Termflow steht in verschiedenen Produktpaketen und für verschiedene Benutzergruppen mit unterschiedlichen Rechten zur Verfügung. Für die Beschreibung der Funktionalitäten in diesem Artikel wird das Professional-Paket und der Benutzer Administrator verwendet, da nur so alle Funktionalitäten und Rechte beschrieben werden können. Kleinere Pakete oder andere Benutzergruppen haben in bestimmten Bereichen Einschränkungen bei der Bedienung und Konfiguration von Termflow.

Administration und Einrichten des Systems

Vor dem ersten Einsatz des Systems müssen zunächst einige Entscheidungen bezüglich der Konfiguration getroffen und diese entsprechend im

System umgesetzt werden. Grundsätzlich muss festgelegt werden, ob eine Terminologie in einer oder mehreren Termbanken abgelegt werden soll. Der Einfachheit halber wird für diesen Artikel nur eine Termbank (Testinstanz) verwendet. Auf Wunsch des Kunden kann eine Termbank mit bereits vorhandenen Daten kostenpflichtig vorbefüllt werden, damit eine Grundlage für eine Erweiterung des Terminologiebestands vorhanden ist.

Für die vorhandene Termbank müssen zunächst die Sprachen festgelegt werden, in denen Termini vorgeschlagen bzw. eingegeben werden dürfen. Anschließend müssen in den Termbankeigenschaften die Datenfelder der Eintragsstruktur definiert werden. Diese sogenannten Attribute können wahlweise auf Begriffs- oder Benennungsebene angelegt und als Textfeld, Auswahlliste oder Anhang klassifiziert werden. Über Vorlagen kann der Administrator nun definieren, welche Datenfelder auf den einzelnen Ebenen erscheinen sollen, wenn ein Vorschlag eingereicht wird. Die Eintragsstruktur enthält systemseitig eine zusätzliche Ebene: die Sprachebene. Dieser Bereich enthält

bereits die Datenfelder „Definition“ und „Anmerkung“ und kann nicht um weitere benutzerdefinierte Felder erweitert werden.

Nach der Konfiguration der Termbank müssen die Gruppen und deren Rechte definiert werden. Hierbei wird neben dem Gruppennamen festgelegt, auf welche Bereiche oder Aktionen eine Benutzergruppe zugreifen darf (z. B. „Projektmanagement“ oder „Termextraktion starten“). Darüber hinaus gibt es auf Begriffs- und Benennungsebene unterschiedliche Status (z. B. „vorgeschlagen“, „angenommen“, „abgelehnt“ usw.), die den einzelnen Gruppen als „sichtbar“ oder „versteckt“ zugeordnet werden. So kann gezielt gesteuert werden, welche Informationen für eine Benutzergruppe überhaupt verfügbar sind.

Termflow verfügt bereits über eine umfangreiche Auswahl an Benutzergruppen, die entweder angepasst oder aber um zusätzliche benutzerdefinierte Gruppen erweitert werden kann. Dabei können die Rechte einer Gruppe von der Bearbeitung aller Funktionen bis hin zum reinen Leser (Enduser) reichen, der die Termbank lediglich durchsuchen kann und neue Vorschläge machen darf.

ID	Gruppenname:	Rechte	Sichtbare Begriffsstatus	Sichtbare Benennungsstatus	Aktionen	
320	Admins	<input checked="" type="checkbox"/> Administrator <input checked="" type="checkbox"/> Projektmanager <input checked="" type="checkbox"/> Alle Workflows starten <input checked="" type="checkbox"/> Direkt editieren <input checked="" type="checkbox"/> Stoppwortliste pflegen <input checked="" type="checkbox"/> Termextrakt starten <input checked="" type="checkbox"/> Statistiken anzeigen	<input checked="" type="checkbox"/> vorgeschlagen <input checked="" type="checkbox"/> in Bearbeitung <input checked="" type="checkbox"/> in Freigabe <input checked="" type="checkbox"/> freigegeben <input type="checkbox"/> gelöscht	<input checked="" type="checkbox"/> versteckt <input checked="" type="checkbox"/> vorgeschlagen <input checked="" type="checkbox"/> vorgeschlagen (bevorzugt) <input checked="" type="checkbox"/> angenommen <input checked="" type="checkbox"/> in Bearbeitung freigegeben <input checked="" type="checkbox"/> bevorzugt <input checked="" type="checkbox"/> abgelehnt <input type="checkbox"/> gelöscht	<input checked="" type="checkbox"/> in Freigabe <input checked="" type="checkbox"/> Freigabe <input type="checkbox"/> unklar <input checked="" type="checkbox"/>	

Abb. 1: Bearbeitung einer Gruppe und Zuordnung der Rechte

Die einzelnen Personen, die mit Termflow arbeiten, werden nach der namentlichen Anlage und der Zuweisung von Benutzernamen und Passwörtern den einzelnen Benutzergruppen zugeordnet. Abschließend müssen die Workflows nach den individuellen Bedürfnissen angepasst oder neu erstellt werden. In diesem Bereich steht ebenfalls eine große Auswahl an bereits angelegten Workflows zur Verfügung, die direkt verwendet werden können. Die Erstellung eines neuen Workflows bzw. die Anpassung eines bestehenden wird folgendermaßen vorgenommen:

1. Eingabe eines Workflow-Namens
2. Auswahl der im Workflow enthaltenen Aufgaben: Hierfür stehen folgende Aufgaben zur Verfügung: „Bereinigen“, „Akzeptieren“, „Kommentieren“, „Attribuieren“, „Freigeben“, „Übersetzen“, „Ändern“. Diese können in einer beliebigen Reihenfolge und beliebig häufig hinzugefügt werden.
3. Zuordnung einer oder mehrerer Gruppen zu einer Aufgabe
4. Zuweisung der unterschiedlichen Benennungsstatus für eine Aufgabe

Zusätzlich kann definiert werden, dass ein sogenannter Auslöser einen bestimmten Workflow bedingt. Als Auslöser stehen derzeit „Vorschlag“, „Termextraktion“ und „Änderungsvorschlag“ zur Verfügung. Grundsätzlich kann ein Benutzer mit dem

entsprechenden Recht alle verfügbaren Workflows starten, wenn er z. B. die Übersetzung einer Benennung in eine bestimmte Sprache anstoßen möchte.

Anlegen und Bearbeiten eines Projekts

Es gibt insgesamt vier Möglichkeiten, um einen Vorschlag in Termflow anzustoßen:

1. Direkter Vorschlag einer neuen Benennung, die noch nicht in der Termbank vorhanden ist: Wird für ein Suchwort in Termflow kein Treffer gefunden, bietet das System dem Benutzer automatisch an, die Benennung als Vorschlag einzureichen. In einem neuen Fenster werden dann die gemäß des Workflows und der Eintragsstruktur (obligatorische Felder) definierten Informationen vom Benutzer abgefragt. Am Ende der Eingabe wird ein neues Projekt erstellt, das über die Projektübersicht jederzeit von den Verantwortlichen eingesehen und bearbeitet werden kann.
2. Termextraktion aus einem externen Dokument (PDF, XLF, TMX, TXT, ZIP): Die Termextraktion dient häufig dazu, einen grundlegenden Datenbestand in der Termbank zu erstellen, der dann weiter aufgebaut werden kann; sie ist aber auch eine gute Vorgehensweise, um einen bestehenden Datenbestand schnell zu erweitern.

FLEURY & FLEURY
CONSULTANTS

Effiziente Strukturen für die globale Kommunikation

Qualität
 Innovation XML Akquise
 Wettbewerbsfähigkeit Transfer Partner
 Redaktionsleitfaden Testing Backsourcing SLA
 Translation Memory Datenprozess Best Practice
 Marktexploration Preisdruck Standards Termintreue
 Styleguide CMS Leistungen Recherche Leistungsoptimierung
 EN 15038 Infrastrukturplanung Sprachen Effizienz Q-Checker
 Vertragsgestaltung Prozessmanagement Übersetzung Outsourcing
 ISO 9000 Nachhaltigkeit Informationsfluss Kundenmanagement
 Change Management Toolsauswahl Technologie Standardisierung
 Workflow Virtuelle Teams Wiederverwendung Kundenorientierung
 Service Wissensmanagement Prozesse Vernetzung Dienstleistungsauswahl
 Automatisierung Controlling Trends Autorenssysteme Niederlassungen
 Organisation Teamleitung Professionelle Kommunikation Schnittstellen
 Terminologiemangement Ressourcenmanagement Weiterbildung
 Marktpositionierung Qualitätssicherung Kommunikationsmanagement
 Gute Geschäftsbeziehung Produktivität Dokumentation Durchlaufzeit
 Auftreten Preisgestaltung Fördermöglichkeiten Verhandlungen
 Anforderungsermittlung Mitarbeiterführung Servicemanagement
 Kundenzufriedenheit CAT-Tools Ergebnisicherung Globalisierung
 Branchenentwicklung Kundengespräche Projektmanagement
 Geschäftsführung Strategie Lokalisierung Marktübersicht
 Post Editing Copywriting Copywriting Maschinenübersetzung
 Interkulturelle Zusammenarbeit Marktforschung
 Risikomanagement ROI Konfliktlösung
 CI Übersetzungsmanagement
 Benchmarking

Wir beraten Sie.
Expertise. Erfahrung. Unabhängigkeit.

www.fleuryfleury.com




Für die Termextraktion können zunächst vordefinierte Filtereinschränkungen angegeben („Wörter aus Kleinbuchstaben ignorieren“, „Minimale Zeichenanzahl“, „Mindestanzahl an Vorkommnissen“,

Workflows

Workflow	Aufgaben	Gruppen	Optionen	Aktionen	Verfügbare Aufgaben
Vorschlag	Bereinigen Akzeptieren	<input type="checkbox"/> Admins <input type="checkbox"/> Endbenutzer <input type="checkbox"/> Freigeber <input checked="" type="checkbox"/> Terminologen	Benennungsstatus <input type="checkbox"/> unverändert <input checked="" type="checkbox"/> abgelehnt <input checked="" type="checkbox"/> vorgeschlagen	<input type="button" value="Speichern"/> <input type="button" value="Zurück"/>	Bereinigen Akzeptieren Kommentieren Attribuieren Freigeben Übersetzen Ändern

Abb. 2: Beispiel-Workflow „Vorschlag“

„Anzahl der Beispielsätze“) und wahlweise positive und negative reguläre Ausdrücke definiert werden. Für die erste automatische Bereinigung des Extraktionsergebnisses werden die Transline- und die Kunden-Stoppwortliste (kundenspezifische Wörter, alle abgelehnten Benennungen und alle bereits in der Termbank vorhandenen Benennungen) mit einbezogen, um das bestmögliche Ergebnis zu erzielen. Nach der Extraktion kann die Termkandidatenliste direkt in Termflow bereinigt, aber auch wahlweise als CSV-Datei exportiert werden.

3. Import einer externen Liste mit Vorschlägen (CSV-Dateien): Nach dem erfolgreichen Import der CSV-Datei erfolgt der Projektstart analog zur Termextraktion. Auf diese Weise können auch externe Vorschlagslisten schnell in den Abstimmungsprozess von Termflow gebracht werden.
4. Direkte Eingabe (wenn die entsprechenden Rechte vorhanden sind): Auf diese Weise wird ein Abstimmungs- und Freigabeprozess umgangen und das System

wechselt direkt in den Bearbeitungsmodus für einen neuen Eintrag. Hier werden die im Vorfeld in der Eintragsstruktur definierten Felder angezeigt, die mit Informationen gefüllt werden können/müssen.

Bei der Erstellung eines Projekts bzw. bei der Abarbeitung der einzelnen Aufgaben gibt es grundsätzlich die Möglichkeit einer automatischen Benachrichtigung per E-Mail, wobei der Benachrichtigungstext für die einzelnen Aufgabentypen individuell angepasst werden kann.

Über die Projektübersicht werden die angelegten Projekte weiter bearbeitet. Je nach Workflow gibt es eine oder mehrere Aufgaben, die verschiedenen Benutzergruppen zugeordnet werden können. Das folgende Beispiel (Abb. 3) zeigt die Aufgaben „Bereinigen“ und „Akzeptieren“. Wird ein Vorschlag final akzeptiert oder abgelehnt, sind der Workflow und damit das Projekt abgeschlossen und das positive Ergebnis wird automatisch in der Termbank gespeichert. Wird eine Benennung abgelehnt, wird sie automatisch auf die Kunden-

Stoppwortliste übertragen und bei einer neuen Extraktion nicht mehr angezeigt.

Suchen und filtern

Die Suche in der Termbank gestaltet sich sehr einfach und intuitiv. Der Benutzer gibt die Benennung einfach in das entsprechende Suchfeld ein und bekommt dann eine Liste mit Treffern angezeigt. Dabei stehen folgende Suchmöglichkeiten zur Verfügung:

- Suche mit Asterisk * (term*, *term, *term*)
- Autovervollständigung des Suchworts während des Tippens für bereits vorhandene Benennungen
- Anzeige des gesamten Terminologiebestands und alphabetische Suche
- Suche in allen in einem Eintrag vorhandenen Feldern

Es lassen sich folgende Filtermöglichkeiten einstellen:

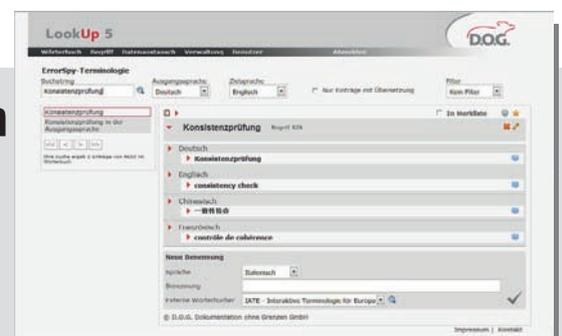
- Ohne Filter: Je nach Benutzereinstellung werden grundsätzlich alle verfügbaren Sprachen unabhängig

LookUp 5

Das Online Terminologieverwaltungssystem

Die ideale Lösung für Dienstleister und Firmen

- Begriffsorientiert
- Import- und Export-Funktion (u. a. TBX, CSV)
- Beliebige Anzahl von Feldern und Attributen
- Benutzerrechte: Lesen, Kommentieren, Ändern
- Benachrichtigungsfunktion bei Änderungen per E-Mail
- Mögliche Integration in ErrorSpy für die Qualitätssicherung von Übersetzungen



**Festpreis oder günstige Mietlösung
Auf eigenem oder externem Server**

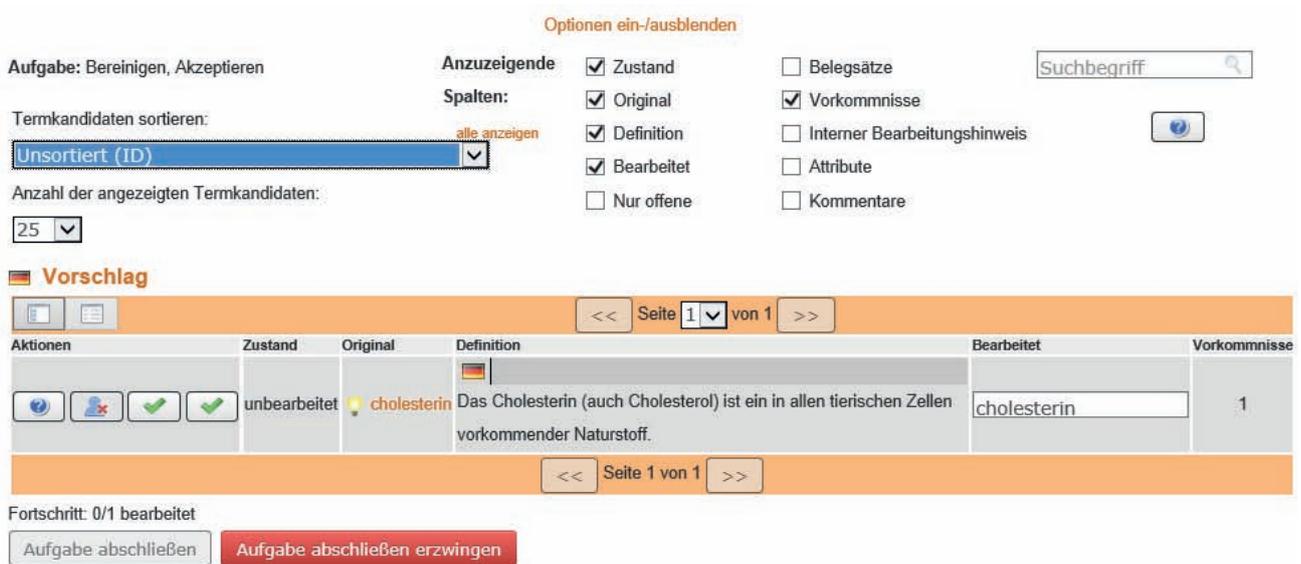


Abb. 3: Bearbeitung eines neuen Vorschlags

des Begriffs- oder Benennungsstatus durchsucht.

- Filtern nach Sprache: Es kann nur in einer, mehreren oder allen Sprachen gleichzeitig gesucht werden.
- Filtern nach Begriffs- oder Benennungsstatus: Nur Einträge mit einem bestimmten Status werden durchsucht.
- Anzeige bestimmter Attribute: Hier kann festgelegt werden, ob

in der Ergebnisliste direkt auch bestimmte Zusatzinformationen angezeigt werden sollen oder lediglich die Benennung selbst.

Bearbeitung eines Eintrags

Wie oben bereits erwähnt, steht für die Bearbeitung der einzelnen Einträge ein Bearbeitungsmodus zur Verfügung. Vorausgesetzt ein Benutzer hat

die entsprechenden Rechte, kann er bestehende Einträge innerhalb eines Workflows, aber auch direkt, ohne weiteren Abstimmungsprozess bearbeiten. Die Bearbeitung gliedert sich in drei separate Ebenen: Begriffs-, Sprach- und Benennungsebene, die jeweils eine eigene Bearbeitungsoberfläche haben. Bei der Bearbeitung stehen die in den Termbankeigenschaften definierten Felder zur



Online-Terminologie-Management für Übersetzungsdienstleister, Unternehmen und Organisationen

TERM

Verfügung. Zusätzlich gibt es auf der Begriffsebene die Möglichkeit, Begriffe zu verknüpfen (Verlinkung zwischen Einträgen), Begriffe zusammenzuführen oder Begriffe zu löschen.

Auf der Sprachebene können mehrere Definitionen in unterschiedlichen Sprachen hinzugefügt werden, wobei jede Definition auch mit einer Notiz versehen werden kann. Auf Benennungsebene stehen neben der eigentlichen Bearbeitung folgende Funktionen zur Verfügung:

- **Anfrage stellen:** Es öffnet sich eine Art E-Mail-Fenster, in dem eine Anfrage zu der eben aufgerufenen Benennung an einen bestimmten Benutzer geschickt werden kann.
- **Zur weiteren Bearbeitung markieren (Zwischenspeicher):** So können verschiedene Benennungen zunächst im Zwischenspeicher abgelegt werden und es wird abschließend entschieden, welche Benennungen gegebenenfalls in ein Projekt zur weiteren Bearbeitung überführt werden.

- **Änderung vorschlagen:** Hier können direkt Änderungen zu einer bestehenden Benennung angegeben werden.

Außerdem können weitere Benennungen als Synonyme oder Übersetzungen in anderen Sprachen direkt eingeben oder vorgeschlagen werden.

Fazit

Termflow dient als unabhängige Abstimmungsoberfläche für Terminologieprozesse und ist toolunabhängig. Die Ergebnisse können anschließend in jedes beliebige System überführt werden, sodass sie beispielsweise in einem Translation-Memory-System aktiv genutzt werden können. Die unterstützten Exportformate sind TXT, CSV und TBX, wobei über TBX ebenfalls ein sogenannter Differenzexport ausgeführt werden kann, bei dem nur die innerhalb eines bestimmten Zeitraums geänderten Benennungen exportiert werden.

Das System ist leicht einzurichten und bietet dem Benutzer mit dem Professional Paket auch eine maxima-

le Flexibilität bei der Konfiguration des Systems und der Abbildung der kundenspezifischen Abläufe. Über die automatische Benachrichtigungsfunktion und Weitergabe von Workflow-Aufgaben können Projekte ohne weiteres Eingreifen abgearbeitet und abgeschlossen werden. Das Versenden von Fragen und Antworten oder das Hinterlegen von Kommentaren erleichtert dabei die Kommunikation zwischen den Projektbeteiligten. Am Ende kann der Systemnutzer auch auf ein statistisches Auswertungstool zugreifen, über das z. B. Aussagen zu der Häufigkeit von Suchbegriffen, aber auch über die graphische Auswertung, z. B. darüber, wie viele Benennungen pro Sprache es gibt, gemacht werden.

Dr. Nicole Keller

Universität Heidelberg
Seminar für
Übersetzen und Dolmetschen
nicole.keller@
iued.uni-heidelberg.de
www.iued.uni-heidelberg.de



Rhein-Main-Hallen, Wiesbaden, 6.–8. November 2013

Jahrestagung

mit **tcworld**
conference 2013

Mit der weltgrößten Messe für
Technische Kommunikation!
• Mehr Fläche durch zusätzliche Halle
• Viele neue Aussteller
• Aktuelle Trends und Neuheiten

Themenbereiche: Redaktionssysteme • Katalogerstellung • Lokalisierung und Übersetzung • Terminologie • Sprachtechnologie • User Assistance • Offene technische Standards • Gesetze, Normen, Richtlinien • Informationsdesign • International Management • Junge Technische Redakteure • Karriere und Persönlichkeitsentwicklung • Personal- und Kostenmanagement • Professionelles Schreiben • Visuelle Kommunikation • Wissenschaftliche Beiträge zur Technischen Kommunikation • Mobile Dokumentation • Soziale Netzwerke

www.tekom.de/tagung

LISE – Investitionsschutz für gewachsene Terminologieressourcen

Das diesen Sommer auslaufende Projekt Legal Language Interoperability Services (LISE) wurde im Rahmen des ICT PSP Programms ermöglicht. Ziel ist, die Qualität von bereits bestehenden terminologischen Ressourcen auf unterschiedlichen Ebenen zu verbessern und eine Hilfestellung bei interinstitutioneller kollaborativer Terminologearbeit zu bieten. Das Projekt wurde von einem Konsortium von akademischen und institutionellen Partnern sowie privaten Unternehmen unter der Leitung des Zentrums für Translationswissenschaft der Universität Wien durchgeführt (www.lise-termservices.eu).

Das LISE-Projekt zeigt auf, dass ab einer bestimmten Menge von terminologischen Einträgen die Datenbestände nicht mehr überschaubar sind und nicht mehr alleine durch intellektuelle Arbeitsschritte manuell effizient gepflegt und aktualisiert werden können. Erschwerend kommt hinzu, dass viele verschiedene Personen über lange Zeiträume hinweg an einer Ressource arbeiten, was zu einer mangelnden Einheitlichkeit führt: nicht alle Sprachen und Fachbereiche sind gleichmä-

ßig abgedeckt; Doubletten und Einträge mit Rechtschreibfehlern treten auf.

Während die terminologischen Ressourcen jahrelang wachsen, wird es immer schwieriger, eine anhaltend hohe Qualität sicherzustellen. Dies liegt auch daran, dass Datenbanken nicht nur „moderat“ zunehmen, also Eintrag für Eintrag, sondern, dass durch organisationale Umstrukturierungen, Abteilungs- oder Unternehmensfusionen und internationale Kooperationen früher getrennt gepflegte Terminologieressourcen immer öfter in eine gemeinsame Datenbank importiert und von dort weiter bearbeitet werden müssen. Ein Hersteller von Kamera- und Videogeräten hatte ursprünglich vielleicht zwei Datenbanken, eine für klassische Fotoapparate, eine für Videokameras. Diese Grenzen verschwinden heute, denn digitale Fotoapparate können nun alle auch Videos aufnehmen. Die Technologien und deren Fachbegriffe wachsen also zusammen, so auch die jeweiligen Terminologiedatenbanken mit der Folge, dass Einheitlichkeit und Übersichtlichkeit der Datenbank oft abnehmen (siehe auch Doubletten-Problematik).

Solche Entwicklungen erfordern also neue Lösungsansätze, um nicht nur Eintrag für Eintrag zu überprüfen, sondern um die Gesamtqualität einer Ressource wiederkehrend zu messen und zu optimieren.

LISE stellt Terminologieverantwortlichen Werkzeuge sowie einen Leitfaden bereit, um die Qualität von Terminologiesammlungen gezielt zu verbessern. Umfangreiche Forschungsarbeiten zu realen Prozessabläufen in der Terminologieerstellung und -verwaltung im Bereich der Rechts- und Verwaltungsterminologie stellen einen Kernbereich des Forschungsprojektes dar.

Für den Leitfaden wurden u. a. die Terminologieverantwortlichen der Ernährungs- und Landwirtschaftsorganisation der Vereinten Nationen, des Auswärtigen Amtes in Berlin, der EU-Institutionen und von termcat befragt. Ein Ergebnis ist, dass die konkreten Arbeitsprozesse in der Terminologie viel weniger standardisiert sind als in der Literatur dargestellt. Gerade was die Terminologieerstellung betrifft, unterscheiden sich z. B. die EU-Institutionen viel stärker voneinander als angenommen.

Der Terminologie-Service in LISE wurde vom schwedischen Sprachtechnologie-Unternehmen ESteam AB entwickelt. Die Ausgangsbasis ist eine bereits bestehende terminologische Ressource. Deren Daten werden exportiert und dann auf verschiedene Merkmale hin analysiert. Die möglichen Fehler und Lücken mit passenden Korrekturvorschlägen werden den Anwendern in speziell angepassten Werkzeugen präsentiert. Diese haben nun die Möglichkeit, die Vorschläge anzunehmen oder abzulehnen. Auf einer Diskussions-Plattform können schwierige oder strittige Fälle thematisiert und zur Abstimmung gebracht werden; diese Plattform wurde explizit für LISE entwickelt, um dem organisationsübergreifenden Charakter

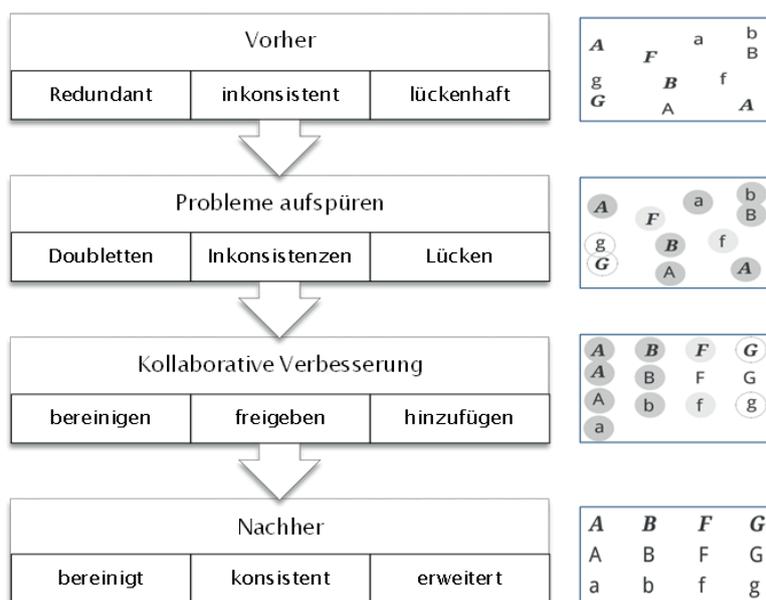


Abb. 1: Schritte zur Qualität

gerecht zu werden. Danach werden die Daten zurück in die ursprüngliche Datenbank importiert. Am Ende dieses Prozesses hat man eine bereinigte, konsistentere und erweiterte terminologische Ressource.

Terminologische Ressourcen können hinsichtlich unterschiedlicher Kriterien verbessert werden: Datenbereinigung mit „Cleanup“, Eintragsgruppierung mit „Omeo“, multilinguale Erweiterung mit „Fillup“. Cleanup spürt potenzielle formale Fehler auf: Rechtschreibfehler und Inkonsistenzen (z. B. Umlautverwendung), technische Fehler (z. B. englische Termini als deutsche klassifiziert), Redundanzen (doppelte Einträge oder nur in Schreibung unterschiedliche Termini) sowie unzureichende Domänenattributierung („Arbeitslosengeld“ nur klassifiziert unter „Finance“ aber nicht auch unter „Social Security“). Omeo fasst getrennte gleichwohl semantisch ähnliche bzw. identische Einträge zusammen, um deren Vergleich, Beschreibung, Klassifizierung und auch Erweiterung/Übersetzung in andere Sprachen effizienter zu gestalten. Fillup scannt Translation Memorys und macht darauf basierende Benennungsvorschläge für bestehende Einträge mit noch fehlenden Sprachen.

Für das LISE-Projekt wurde als Fallstudie die interinstitutionelle Datenbank der EU, IATE, herangezogen. Sie ist das Resultat der Zusammenführung von unterschiedlichen terminologischen Datenbanken und eignet sich daher sehr gut als Fallstudie. IATE beinhaltet mehr als 11 Millionen Termini, in ca. 1,5 Millionen Einträgen und 22 Sprachen. Aus praktischen Gründen wurde indes nur eine Teilmenge der IATE-Daten verwendet. Fachgebietsangaben dienten als Auswahlkriterium: Die Wahl fiel auf die Einträge aus dem Sozialrecht, ca. 96.000 Benennungen in 21 Sprachen. Und tatsächlich fanden sich in alleine diesem Bereich durch die Analyse 1324 Eintragsdoubletten, über 5000 mögliche Rechtschreibfehler und fast 1500 falsche Sprachklassifikationen. Ferner hat die Analyse ergeben, dass sehr viele Einträge nur ein oder zwei

Sprachen abdecken. Hier zeigt sich das Potenzial für Fillup.

Diese theoretischen Erkenntnisse wurden auch während zweier Workshops mit Terminologieverantwortlichen von IATE bestätigt: „So ein Aufräumgerät bräuchten wir dringend!“ – so ein Teilnehmer in Brüssel. Ebenso wurde der semi-automatische Ansatz der Technologie hochgeschätzt, nämlich dass die Software lediglich Vorschläge macht, während stets ein Experte die Entscheidungshoheit über Datenänderungen behält.

In LISE liegt der Fokus zwar auf Rechts- und Verwaltungsterminologie, die methodischen Erkenntnisse und Werkzeuge lassen sich aber auf andere Domänen umlegen. Der Leitfaden zu Terminologieprozessen und Qualitätsmethoden ist ab Juli 2013 öffentlich zugänglich. Die Werkzeuge können von den EU-Institutionen im Rahmen von Service-Leistungen nun direkt eingesetzt werden. Die der Software zugrunde liegenden Algorithmen wird ESteam weiterentwickeln, um sie effizient auch für andere Anwender außerhalb IATE zur Verfügung zu stellen. Die Markt- und Technologie-Erkenntnisse aus dem LISE-Projekt führten letztlich auch zur Gründung der Coreon GmbH, einer Software-Firma, welche sich in Partnerschaft mit ESteam dediziert auf die Qualität und Potenziale von Wissensressourcen fokussiert.

Michael Wetzel
michael@esteam.se



Tanja Wissik
tanja.wissik@univie.ac.at



Elena Chiochetti
elena.chiochetti@eurac.edu



DTT-Seminar Terminologiearbeit Grundlagen, Recherche, Werkzeuge, Prozesse



© pixelto.de

Termin:
11. Oktober 2013, 9.00 – 17.30 Uhr

Ort:
Leonardo Hotel Karlsruhe
Ettlinger Straße 23, 76137 Karlsruhe
www.leonardo-hotels.de

Zielgruppen:

- Übersetzer, Dolmetscher, Terminologen
- Technische Redakteure
- Sprachübergreifend arbeitende Fachleute
- Studierende im Bereich Übersetzen / Dolmetschen / Technische Redaktion

Inhalte:

- Grundlagen (Prof. Dr. Petra Drewer, Hochschule Karlsruhe)
- Recherche (Prof. Dr. Klaus-Dirk Schmitz, Fachhochschule Köln)
- Werkzeuge (Prof. Dr. Klaus-Dirk Schmitz, Fachhochschule Köln)
- Prozesse (Dr. Rachel Herwartz, TermSolutions)

Die Teilnahmegebühr schließt die Seminarunterlagen, die Kaffeepausen und das Mittagessen ein:

- 260 Euro für Nicht-Mitglieder
- 190 Euro für Mitglieder DTT, ADÜ Nord, ASTTI, ATICOM, BDÜ, DGI, tekom, Universitas

Veranstalter: Deutscher Terminologie-Tag e.V.

Information bei: fortbildung@dttev.org

Anmeldeformular und Programm unter:
www.dttev.org

Stand: 1.8.2013, Änderungen vorbehalten

SAP-Terminologie auf SAPterm.com verfügbar

Weltweit gibt es eine ständige Nachfrage nach SAP-Terminologie, nicht nur bei SAP. Auch Kunden, Partner und Übersetzer verwenden die SAP-Standardterminologie und ihre Definitionen, um die Qualität von Oberflächen, Dokumentation oder Schulungen sicherzustellen, korrekte Übersetzungen zu liefern und die Kosten von SAP-Projekten und -produkten zu reduzieren.

SAP prägt auch die Sprache von ERP-Software weltweit und unterstreicht, dass SAP-Terminologie der Branchenstandard ist. So wurden Begriffe für „Customer Relationship Management“ erstmals im Rahmen von SAP-Projekten ins Türkische, Ungarische und weitere Sprachen übersetzt. SAP muss daher nicht lange überlegen, ob sie ihre Terminologie anderen zur Verfügung stellen möchte. Es lohnt sich, Terminologie mit Externen zu teilen. Weil Terminologie ein kritischer Projektfaktor ist, hat SAP seit 1990 verschiedene Auslieferungsoptionen getestet und diese den externen

Kunden und Partnern angeboten, um die Nachfrage nach Terminologie zu friedenzustellen. Die verschiedenen Lösungen, von gedruckten Wörterbüchern bis hin zu CDs und Dateien, sind nie hundertprozentig akzeptabel gewesen. Ein Problem war z. B., dass der ganze Inhalt der zentralen SAPterm-Terminologiedatenbank nie zur Verfügung stand, da weder alle Sprachen noch alle Einträge mitgeliefert werden konnten. Es gab auch Bedenken, als trotz Stillschweigevereinbarungen SAP-Terminologiedateien im Internet erschienen.

Im Laufe eines zweijährigen Projekts hat SAP Language Services (SLS) www.sapterm.com entwickelt und gab diese Webseite Anfang März 2013 für die Öffentlichkeit frei. Ob in der Entwicklung, im Vertrieb oder bei Übersetzern – die Website bietet Nutzern einen kostenlosen, einfachen und direkten Zugriff auf offizielle SAP-Terminologie. Die Oberfläche auf Basis von SAP WebDynpro bietet Funktionen für das Suchen und Filtern von Inhalten in der Termi-

nologiedatenbank SAPterm. Sieht man sich die SAPterm-Einträge in der Detailansicht an, findet man Informationen zum Themengebiet und Softwarerelease sowie Definitionen, Abkürzungen, Synonyme und Zusatzinformationen. Teilweise stehen für die Begriffe Übersetzungen in über 40 Sprachen zur Verfügung.

SLS aktualisiert SAPterm.com ständig mit neuer Terminologie aus der zentralen SAPterm-Datenbank. Darüber hinaus können Anwender über SAP Terminology Community der SAP Feedback zur Terminologie geben und über Fragen der Terminologieverwaltung diskutieren. Außerdem ist vorgesehen, SAPterm.com als Plattform für weitere Innovationen im Bereich Terminologieauslieferung und -bereitstellung zu nutzen.



Mark D. Childress
Mark.childress@sap.com

Terminologie

Certificate of Advanced Studies (CAS), 15 ECTS-Punkte

- Terminologielehre und Terminologearbeit
- Terminologiewerkzeuge
- Projekt- und Prozessmanagement Terminologie

Dozierende: Klaus-Dirk Schmitz (FH Köln), Rachel Herwartz (TermSolutions), Beate Früh (TermSolutions), Christian Kriele (ZHAW), Martin Kappus (ZHAW)

Start: September 2013

www.linguistik.zhaw.ch/terminologie

ZHAW, Departement Angewandte Linguistik

Maren Runte, Arbeitsstelle Fachkommunikation und Terminologie, Theaterstrasse 15c, 8401 Winterthur, Tel. +41 58 934 60 88, maren.runte@zhaw.ch

Zürcher Fachhochschule

Neuer DTT-Vorstand

Auf der diesjährigen DTT-Jahresmitgliederversammlung in Köln (19. April) wurde der Vorstand neu gewählt. Detlef Reineke war neun Jahre lang Geschäftsführer und wechselt in das Amt des stellvertretenden Vorstandsvorsitzenden. Er wird auch weiterhin als Herausgeber und treibende Kraft der DTT-Fachzeitschrift eDITion tätig sein und sich zusätzlich um den Auftritt und das Marketing des DTT kümmern.

Als Nachfolgerin von Detlef Reineke wurde Annette Weilandt gewählt. Sie studierte an der Universität Leipzig, wo sie derzeit zum Thema Ter-

minologiemangement promoviert. Zudem hat sie mehrere Jahre bei der Daimler AG in diesem Bereich gearbeitet und ist Autorin mehrerer Artikel und Präsentationen zu Terminologie-Themen.

Auch auf der DTT-Geschäftsstelle hat es personelle Veränderungen gegeben. Inke Raupach verlässt die Geschäftsstelle in Richtung Luxemburg, um bei der EU als Übersetzerin zu arbeiten. Wir danken ihr an dieser Stelle für die geleistete Arbeit und wünschen ihr weiterhin viel Erfolg. Ihre Nachfolge tritt Olga Vodopyanova an. Sie hat in Russland Linguistik und inter-

kulturelle Kommunikation studiert und an der FH Köln den BA Mehrsprachige Kommunikation und den MA Terminologie und Sprachtechnologie abgeschlossen. Derzeit ist sie wissenschaftliche Mitarbeiterin am Institut für Informationsmanagement der FH Köln.

Der Vorstand freut sich auf den Zuwachs und fühlt sich in seinem Anliegen gestärkt, weiterhin im Sinne der DTT-Mitglieder arbeiten zu können.

Mark D. Childress
childress@dttev.org



Von links nach rechts: Vorsitzender: Mark D. Childress, stellv. Vorsitzende: Klaus-Dirk Schmitz, Ursula Fähndrich, Detlef Reineke, Schatzmeisterin: Angelika Ottmann, Geschäftsführerin: Annette Weilandt, Leiterin der Geschäftsstelle: Olga Vodopyanova

Arbeitstreffen des DIT in Karlsruhe

Am 15.02.2013 trafen sich die Mitglieder des Deutschen Instituts für Terminologie (DIT) und einige Vorstandsmitglieder des Deutschen Terminologie-Tages (DTT) zu einem äußerst produktiven und angenehmen Arbeitstreffen an der Hochschule Karlsruhe. Zu besprechen gab es vieles, denn nach einer längeren Zeit der Ruhe erweckt der neue Vorstand des DIT (Petra Drewer, Donatella Pulitano, Felix Mayer) nun viele Projekte zu (neuem) Leben.

Ein Projekt, das beiden Vorständen sehr am Herzen liegt, ist die Ausschreibung des DTT/DIT-Förderpreises. Der Vorstand des DTT hat in Abstimmung mit seinen Fachbeiräten im DIT beschlossen, alle zwei Jahre einen Förderpreis für herausragende

Leistungen auf dem Gebiet der Terminologiewissenschaft und der Terminologiearbeit zu verleihen. Erstmals wird der DTT-Förderpreis im Jahr 2013 ausgeschrieben; die Preisverleihung wird auf dem DTT-Symposium 2014 stattfinden, auf dem die Preisträgerin oder der Preisträger auf Einladung des Vereins einen Vortrag halten wird. Auf dem Arbeitstreffen wurden die genauen Modalitäten festgelegt und Verantwortliche benannt. Nun erwarten alle DIT-Mitglieder (insbesondere diejenigen, die in der Jury tätig werden) die hoffentlich zahlreich eingehenden Bewerbungen (www.dttev.org/Foerderpreis_DTT.html).

Das DIT arbeitet außerdem an der Aktualisierung des erfolgreichen Best-Practice-Ordners des DTT, der in die

2. Auflage gehen soll. Und so wurde auf dem Arbeitstreffen ein Schlachtplan entworfen, der es ermöglichen soll, bis 2014 eine komplette Überarbeitung zu realisieren. Die bisherigen 6 Themenbausteine werden wo nötig aktualisiert, und auf vielfältigen Wunsch hin werden zwei Module zu Themen ergänzt, die das Publikum besonders interessieren und so den Ordner abrunden: die Wirtschaftlichkeit und das Urheberrecht.

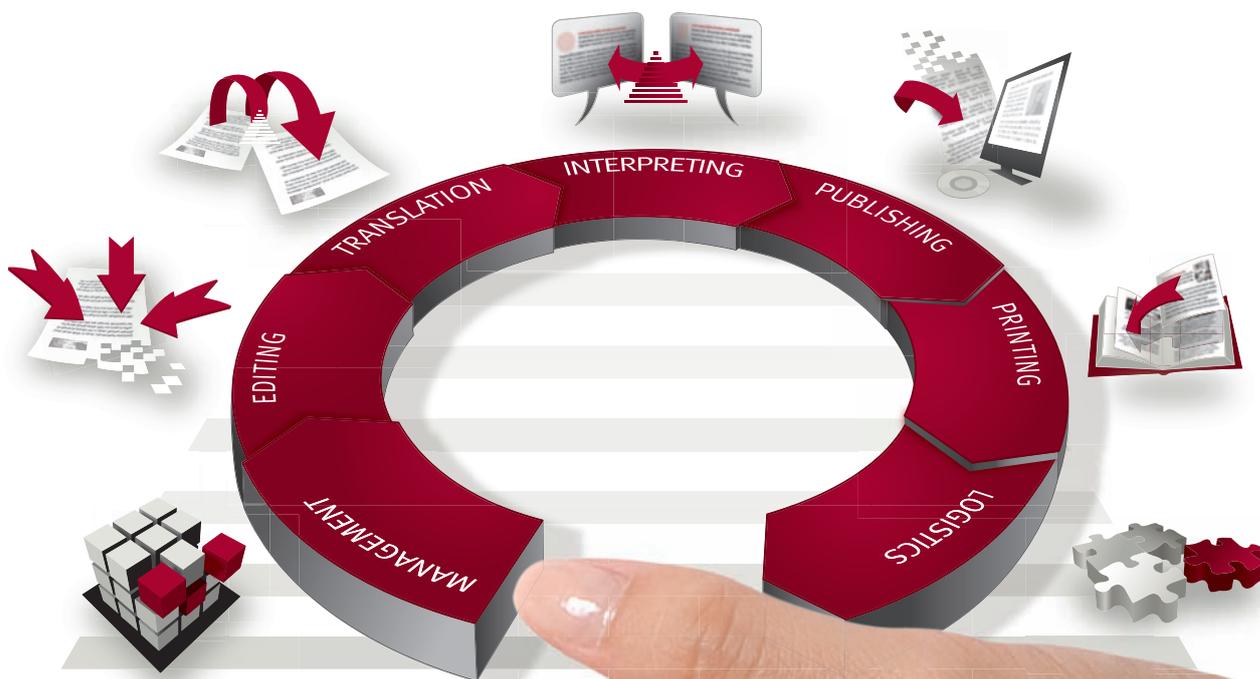
Petra Drewer
drewer@ditev.org



Content creation · Translation/Localization · DTP/Prepress
Illustration/Animation · Automated multi-channel publishing
Interactive information retrieval · Process control & automation
IT services · Training · Consulting ...

... STAR solutions & services

The added value to your corporate
product information lifecycle
management.



Industry solutions from a single source.