

Jenseits der Fachkommunikation – ein Blick in die Zukunft der Terminologie

Von Sue Ellen Wright (aus dem Englischen übersetzt von Detlef Reineke)

Begriffsbestimmung, Benennungszuordnung und hierarchische Wissensmodellierung sind so alt wie die westliche Philosophie. Trotzdem oder vielleicht gerade deshalb bedient man sich dieser Ordnungsmittel im Bereich des webgestützten Wissensmanagements (Stichwort Semantic Web), wenn es um die Nutzung von Ressourcen wie etwa XML-basierten Terminologiedatenbanken, Metadaten-Registern, Katalogsystemen oder Ontologien geht. Der vorliegende Beitrag wirft auf der Grundlage retrospektiver Betrachtungen einen Blick in die Zukunft der Terminologie und zeigt Trends auf, die andeuten, dass die Grundsätze und Methoden der Terminologie in einer wachsenden Zahl von Wissensgemeinschaften (Communities of Practice) Anwendung finden.

Schlagwörter

Terminologie, Begriffssysteme, Metadaten, Wissensressourcen, Semantic Web

Einleitung

Terminologiarbeit ist seit je her Sache von Fachleuten, Wissenschaftsautoren, Technischen Redakteuren, Fachübersetzern, Terminologen sowie Ausbildern in Hochschulen und Industrie. Sie entspringt pragmatischen Anforderungen insbesondere in der Technik: Hier begegnet *Homo faber homo appellans*, dem Namensgeber (*appellavitque Adam nominibus suis cuncta animantia et universa volatilia caeli et omnes bestias terrae* – Und der Mensch [Adam] gab Namen allen Tieren, dem Vieh und den Vögeln des Himmels und allen Tieren des Feldes. Gen. 2.20).

Terminologie ist in diesem Kontext sprachbezogen. Die Bildung von Benennungen sowie deren Beschreibung, Normung und Übersetzung sind Voraussetzung für Textkonsistenz und -kohärenz in der schriftlichen und mündlichen Kommunikation in Wissenschaft, Technik, Handel und Recht. Terminologie macht somit einen Großteil der milliardenschweren Übersetzungsindustrie aus. Zudem lässt sich Terminologie mithilfe der Computertechnik angemessen beschreiben. Diese Wechselbeziehung bestimmt die Entwicklungen im Bereich der Terminologie. So ist etwa die Web 2.0 Community mit ihren interaktiven und gruppenspezifischen Elementen (Stichwort ‚Cloud computing‘) dabei, die Übersetzungslandschaft zu verändern. Im Internet werden Translation Memories zu globalen, multilingualen Online-Korpora zusammengefasst und kollaborative Terminologiediskussions-

gruppen bieten Foren, um terminologisches Wissen auszutauschen. Dabei steht die Beurteilung von Vorschlägen oft stärker im Vordergrund als die Suche nach konkreten Informationen. Übersetzer und Terminologen arbeiten heute auch nur noch selten auf sich allein gestellt, sondern zeitnah und kollaborativ über offene Plattformen im Internet.

Computergestützte Systeme sind zu einem unverzichtbaren Bestandteil des Terminologiemanagements geworden. Gleichzeitig sind diese Systeme, insbesondere webgestützte Systeme, wiederum auf feinstrukturierte terminologische Anwendungen angewiesen. Auf die wechselseitige Abhängigkeit von Technik und Sprache verweist auch der US-amerikanische Zenlehrer, Dichter und Buddhist Norman Fischer, der in einer Ausgabe des IT-Magazins *Wired* aus dem Jahre 1999 mit den Worten zitiert wird: „Die eigentliche Technologie hinter all unseren Technologien ist Sprache.“ [1]. Ob Schlagwörter oder Google-Indizes, Strategien webgestützter Systeme zum Zweck der Suche und Ordnung von Metadaten und Wissen sind am Ende stets auf die Analyse und die Verarbeitung von Benennungen und den zwischen ihnen bestehenden Beziehungen angewiesen. Terminologen ist nur all zu sehr bewusst, dass die Mehrdeutigkeit gemeinsprachlicher Wörter, die Indeterminiertheit individueller Äußerungen und die Unvergleichbarkeit unterschiedlicher Fachsprachen von Wissensgemeinschaften eine effiziente Fachkommunikation erschweren. Dieses Problem lässt sich

ohne Weiteres auf Sprachelemente in Computerumgebungen übertragen. Beispielsweise genannt seien hier die Suchergebnisse von Webbrowsern. Hier sind eine ganze Reihe neuer interdisziplinärer und auf terminologischen Ansätzen basierende Ressourcentypen entstanden, vor allem dort, wo Benennungen oder benennungsähnliche Elemente identifiziert oder zugeordnet, Begriffe festgelegt und systematisiert werden müssen. Vor allem die Verwendung von XML, RDF (Resource Description Framework) und RDFS (RDF Schema) zur Beschreibung von Informationen auf der Grundlage unterschiedlicher methodischer Ansätze in den Wissensgemeinschaften bietet die Möglichkeit universeller Wiederverwendbarkeit und Interoperabilität. Eine breitflächige Umsetzung wird jedoch immer noch durch den Fortbestand inkompatibler Systeme behindert. Noch größere Anstrengungen werden im Bereich der Auflösung von Inkonsistenzen unter Zuhilfenahme von Begriffssystemen unternommen mit dem Ziel, Wissen, Informationen und Daten im World Wide Web schneller abfragen und wiederverwenden zu können.

Wissensordnung in traditionellen terminologischen Begriffssystemen

Die meisten von uns haben wahrscheinlich zum ersten Mal in der Schule in den naturwissenschaftlichen Fächern mit Begriffssystemen Bekanntschaft gemacht. Genannt seien hier Linnés botanische und zoologische Taxonomi-

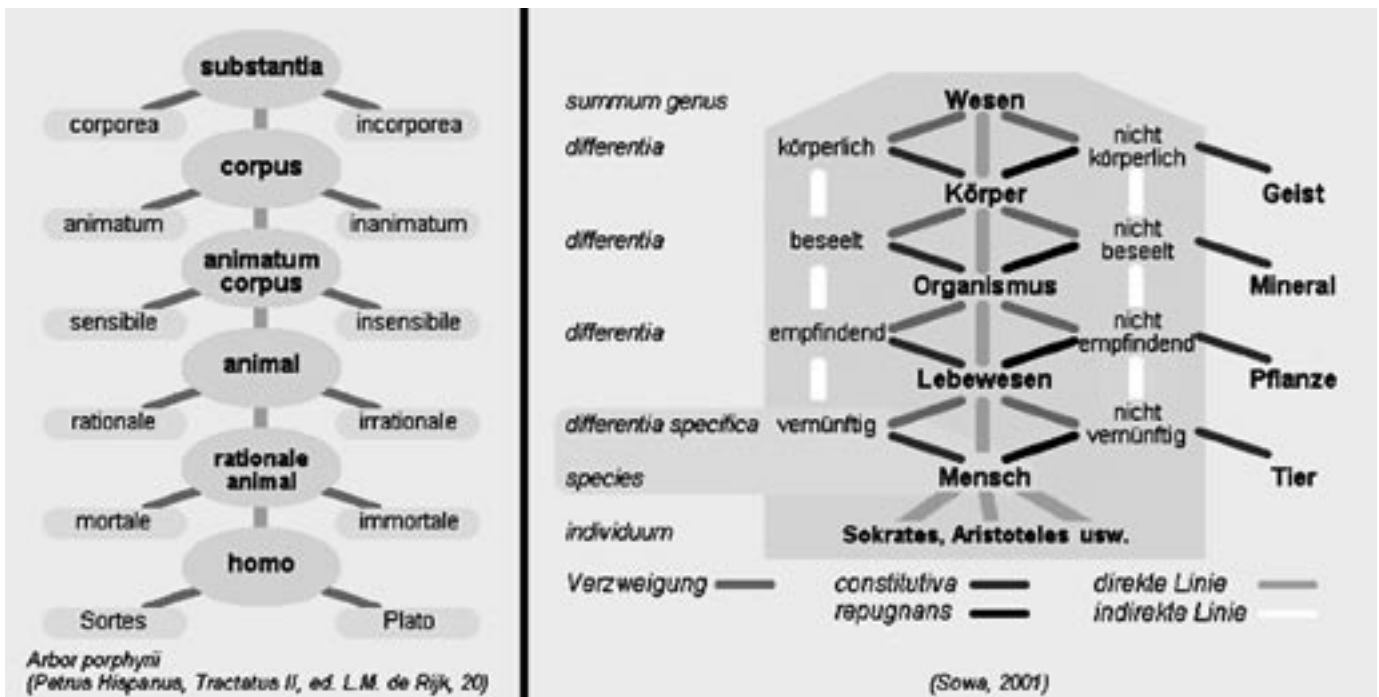


Abbildung 1: Der Porphyriosche Baum des Wissens (Sowa 2001) [2].

en oder Mendelejews Periodensystem der Elemente. Geht man aber in der Geschichte noch etwas weiter zurück, stößt man auf den neoplatonischen Philosophen Porphyrios (* 234 n. Chr. - † im frühen 4. Jahrhundert), dessen Baum des Wissens als Vorläufer der heutigen webgestützten Wissensordnung gilt und vom US-amerikanischen KI-Forscher John F. Sowa nachgestellt wurde (Abb.1).

In Sowas Baumgraphik werden die Beziehungen als Dreiecke dargestellt und gleichen damit den Tripeln, die im Web zur Beschreibung von Wissensbeziehungen in RDF-Notationen verwendet werden. So ist ‚Sokrates‘ über die Eigenschaft ‚individuum‘ mit ‚Mensch‘ verknüpft, ‚Mensch‘ ist über die Eigenschaft ‚species‘ mit ‚Lebewesen‘ verknüpft usw. Jedes Tripel kann für sich allein stehen, sie können aber auch zu kohärenten Systemen miteinander verbunden werden.

Die graphische Darstellung von terminologischen Begriffssystemen in Form von Abbildungen wurde in der Vergangenheit überwiegend manuell vorgenommen. Heute bieten einige Terminologiemanagementsysteme die Möglichkeit, graphische Darstellungen von Begriffssystemen auf der Grundlage definierter Begriffsbeziehungen automatisch zu generieren, so z.B. die von DANTERM für TC 37 gehostete iTerm-Datenbank. iTerm-Baumgraphiken leisten dem Nutzer jedoch nur visuelle Hilfestellung. Sie lassen sich lediglich

als Abbildungen speichern und können nicht von anderen Wissensmanagementsystemen verarbeitet werden. Ein weiteres Manko besteht darin, dass sich einzelne terminologische Einträge nicht dynamisch mit anderen Ressourcen im Web verlinken oder aus dem Web heraus adressieren lassen.

Die Norm ISO/TR 24156 (2008) „Guidelines for using UML notation in terminology work“ geht noch einen Schritt weiter und nutzt die ursprünglich für die Modellierung von Software und anderen Systemen entwickelte Sprache UML (Unified Modeling Language) der Object Modeling Group (OMG) zur Darstellung terminologischer Begriffssysteme. Die Verwendung dieser standardisierten Modellierungssprache ermöglicht die Darstellung tiefstrukturierter, flexibler Begriffsbeziehungen und Merkmale. Gleichzeitig sind kommerzielle, anwenderfreundliche Softwareanwendungen auf dem Markt, die in der Lage sind, graphische Darstellungen zu generieren und, wichtiger noch, zu einem gewissen Grad maschinell zu verarbeiten. Die Verwendung von UML ist allerdings nur ein erster Schritt, um Begriffsmodelle von Wissenssystemen mittels vollautomatischer, standardisierter Webfunktionalitäten darzustellen. Folgen muss nun die Beschreibung von Begriffssystemen anhand einer RDF-basierten Graphensprache, etwa mithilfe eines Dialekts der Web Ontology Language (OWL).

Unterschiedliche Ansätze in den Wissensgemeinschaften

Terminologische Begriffssysteme haben Vieles mit anderen Formen von Wissensordnungssystemen gemeinsam, insbesondere mit Klassifikationssystemen und Taxonomien aus den traditionellen Naturwissenschaften, sowie mit kontrollierten Vokabularen und Thesauri aus der Bibliotheks-, Dokumentations- und Informationswissenschaft. Und gerade hier ist die Unbestimmtheit der Sprache den Terminologen ein Dorn im Auge: sprach- und sachbezogene Vokabulare (wie z.B. Thesauri) werden beide als *Terminologien* bezeichnet, obwohl sie unterschiedliche Merkmale aufweisen. Elaine Svenonius vom Department of Information Studies an der University of California Los Angeles merkt dazu an, dass Termini in Thesauri (gemeint sind eigentlich Deskriptoren) wie etwa ‚Schmetterling‘ sämtliche Bücher zum Thema ‚Schmetterling‘ oder eventuell auch Gemälde, Webseiten



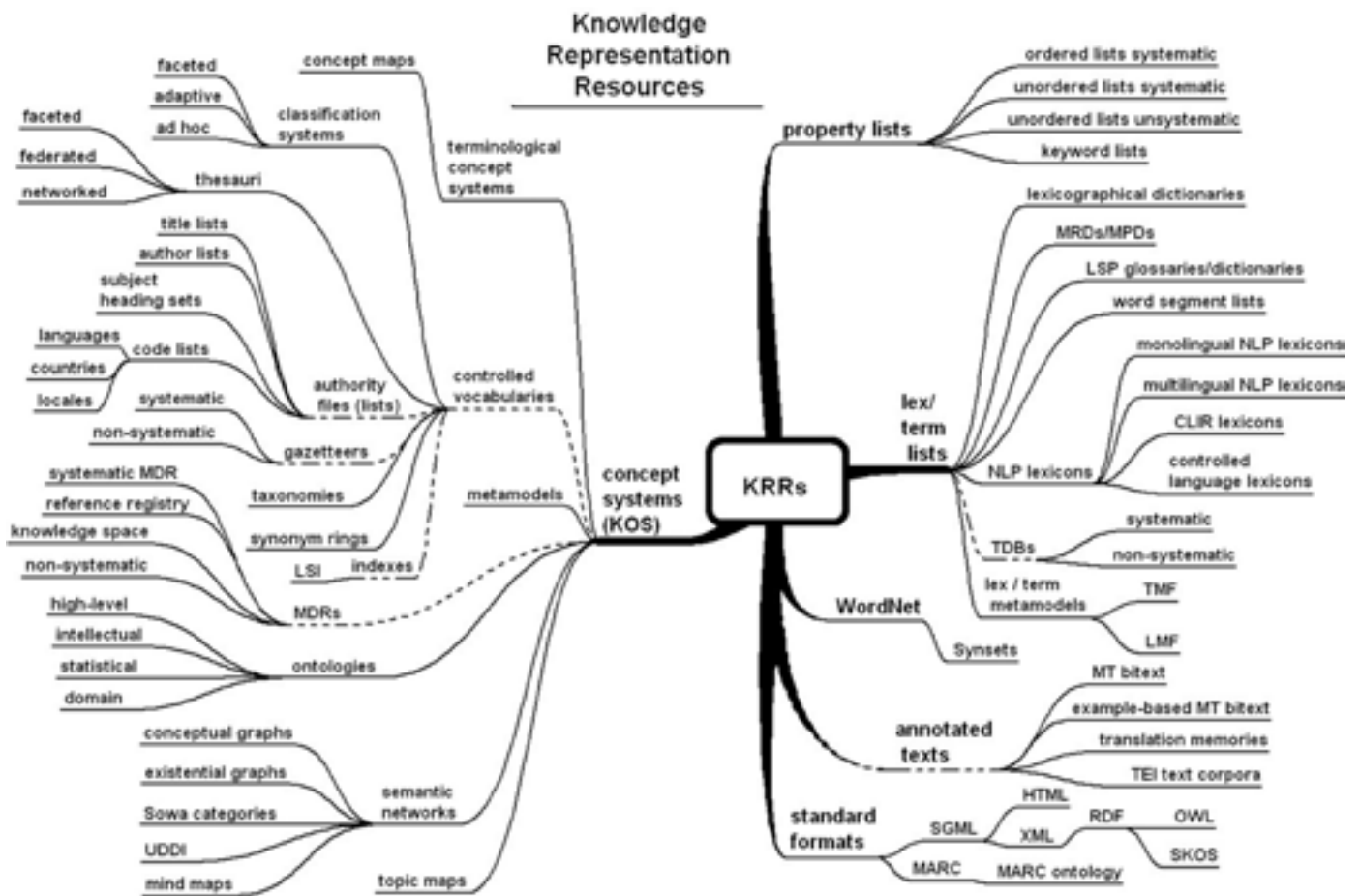


Abbildung 2.: Wissensrepräsentationsressourcen nach Sue Ellen Wright.

oder andere Ressourcen umfassen. Die sprachbezogene Benennung ‚Schmetterling‘ bezieht sich als Ausdrucksseite des entsprechenden Begriffs dagegen auf die Gesamtheit aller realen (sinnlich wahrnehmbaren) oder idealen (gedachten) Schmetterlinge. In der Informationswissenschaft können über Termini *Objekte* oder *Informationen* zum Thema ‚Schmetterling‘ gefunden werden, während Benennungen in der Terminologie dazu dienen, über die *Gegenstände* ‚Schmetterling‘ zu sprechen oder zu schreiben [3].

Der Versuch, diese beiden Ansätze durch eine Cross-Konkordanz in Einklang zu bringen, förderte die Einsicht zutage, dass diese beiden Welten im Kuhnschen Sinne inkommensurabel, also nicht miteinander vergleichbar sind, da beide Systeme – und die jeweiligen Wissensgemeinschaften – unterschiedliche Regeln anwenden. Die Annahme, Informationen ließen sich automatisch zwischen ihnen austauschen, wird somit in Frage gestellt. Die menschlichen Nutzer dieser beiden Systeme stellt dieser Umstand vor keine größeren Probleme, denn sie

erfassen den feinen Unterschied intuitiv oder übersehen ihn einfach. Drückt man diese Systeme jedoch mithilfe von Computern in einer formalen Beschreibungssprache aus, werden die funktionalen Unterschiede sichtbar, die einen automatischen Datenaustausch quasi unmöglich machen.

Eine ähnliche Situation liegt vor, wenn man Metadaten-Register mit Terminologiesammlungen vergleicht. Datenkategorienamen werden als Feldnamen, zugelassene Werte oder Tagnamen in Markup-Sprachen verwendet und sind den Benennungen in terminologischen Ressourcen sehr ähnlich. Bei der Ausarbeitung der ersten Datenkategorienorm ISO 12620 (1999) „Computer applications in terminology – Data categories“ gab es im Technischen Komitee TC 37 Diskussionen darüber, ob die Definitionen von Benennungen in Terminologienormen mit den Definitionen der entsprechenden terminologischen Datenkategorien in den Metadaten-Normen übereinstimmen. Es stellte sich z.B. die Frage, ob es überhaupt notwendig war, eine Datenkategorie */term/* zu spezi-

fizieren, wo doch ‚term‘ bereits in der Norm ISO 1087-1 (2000) „Terminology work – Vocabulary – Part 1: Theory and application“ definiert ist? Erstaunlich auch, dass einige Datenkategoriedefinitionen im Wortlaut anders ausgefallen sind als die Definitionen, die für ähnliche Begriffe in bereits bestehenden Terminologienormen verwendet worden sind.

Nach vielen Jahren hat sich die Einsicht durchgesetzt, dass sich die begriffliche Struktur von Benennungen ändert und gelegentlich neue Definitionen erstellt werden müssen, wenn Benennungen als Datenkategorienamen herangezogen werden. Hinzukommt, dass Datenmodelle für terminologische Einträge nicht angemessen auf Datenkategoriespezifikationen angewendet werden. Ein terminologisches Begriffssystem, das eine Ansammlung von Benennungen enthält, die zugleich auch als Datenkategorienamen verwendet werden, kann eine andere Struktur aufweisen als ein System, das die zwischen den Datenkategorien bestehenden Beziehungen in einem Metadaten-Register abbildet.

So können etwa die einem Begriff untergeordneten Begriffe und die dazugehörigen Benennungen von der Liste zugelassener Werte abweichen, die in einer Pickliste als Datenkategorienamen desselben übergeordneten Begriffs aufgeführt werden.

Zu erwähnen seien aber auch noch andere Ressourcen, in denen, ähnlich wie in der Terminologie, Benennungen oder Namen von Gegenständen spezifiziert, definiert und/oder nach semantischen Kriterien geordnet werden. Dazu gehören u.a.:

- **Genormte Katalogsysteme und Inventory-System-Taxonomien:** Derartige Systeme sind für den e-Commerce-Bereich entwickelt worden, insbesondere für umfangreiche Materialbeschaffungsumgebungen wie etwa bei der NATO. Die eindeutige Beschreibung von Waren, Leistungen, Einzelpersonen und Organisationen in genormten Kennsatz-Registern sichert die Qualität und reduziert die Kosten bei Beschaffung und Management. Einheitliche elektronische Kodierungssysteme stellen die Materialabwicklung, die Lagersteuerung und den Datenaustausch zwischen unterschiedlichen Systemen, Plattformen, Organisationen und Sprachgemeinschaften sicher. Die Normenreihe ISO 8000, die Richtlinien für die Qualität von Daten vorsieht, bietet einen Rahmen für die Kodierung der Syntax und der semantischen Definition von Datenelementwerten [4].
- **Lexika für kontrollierte Sprache:** In Verbindung mit einschränkenden Grammatiken unterstützen solche Lexika die Auflösung homonymer oder polysemer Mehrdeutigkeiten vor allem im Bereich von Single-Source-Autorenwerkzeugen zur Erstellung leicht verständlicher Wartungshandbücher oder ähnlicher Dokumente. Diese Texte genügen den hohen Anforderungen, die an internationalisierte Quelltexte gestellt werden, und lassen sich effizient mithilfe maschineller Übersetzungssysteme in mehrere Sprachen übersetzen [5].
- **Information Mining und Termextraktion:** Zusammen mit Konkordanzwerkzeugen werden Information Mining und Termextraktion im Bereich des Korpusmanagements und der Korpusforschung einge-

setzt, um Informationen aus nicht-kodierten, natürlichsprachigen Texten zu extrahieren und zu ordnen, was wiederum auf die Bedeutung von Terminologie, identifizierten Kollokationen und anderen typischen Sprachelementen schließen lässt. In schriftzeichenbasierten Sprachen ermöglicht die Erstellung von Lexika, in denen die Segmentierung der Benennungen festgelegt ist, den Einsatz natürlichsprachiger Verarbeitungsverfahren.

- **Einfache Relationsregister:** Diese Register, die in Form von speziellen RDF-Datenbanken (RDF-Triplestores) abgelegt werden können, werden außerhalb terminologischer Datenbanken, Metadaten-Register, Textkorpora und anderer Ressourcen zur Wissensrepräsentation aufgebaut und lassen sich über das Internet referenzieren, um semantisch kritische Relationen und Relationssysteme darzustellen und über persistente Identifikatoren mit umfassenderen Wissensressourcen zu verknüpfen.
- **Ontologien:** In RDF-basierten OWL-Dialekten beschriebene Ontologien können komplexere Ordnungssysteme, formal-logische Aussagen und Regeln darstellen, um hochqualitatives Wissensmanagement zu betreiben und automatische Schlussfolgerungen zu ermöglichen. Diese Aufgaben sind ohne derartige Ontologien von Computersystemen nicht zu leisten.
- **Andere mithilfe von SKOS (Simple Knowledge Organization System) beschriebene Wissensorganisationssysteme:** Dazu gehören Thesauri, Klassifikationsschemata, Sacherschließungssysteme und Taxonomien, die zumindest teilweise mit dem Semantic Web Framework kompatibel sind.

Normungsarbeit des Technischen Komitees ISO/TC 37

Trotz der Unterschiede, die zwischen den beschriebenen Ressourcen bestehen, haben sie doch Entscheidendes gemeinsam: Benennungen, benennungsähnliche Namen und Kennsätze (z.B. für Stammdatenobjekte) werden Begriffen zugeordnet, die von Terminologen, Fachleuten und Informationswissenschaftlern unter Berücksichtigung präziser und eindeutiger Kriterien

festgelegt werden. Zu diesem Zweck stehen zwar eine Fülle von Normen und Best Practices zur Verfügung. Die Norm ISO 704 „Terminology – Principles and Methods“ gilt aber weiterhin als die Grundsatznorm schlechthin, um Begriffe in den oben beschriebenen Umgebungen zu bestimmen.

In bestimmten Bereichen der ISO-Normung, der Industrie, der Universitätsforschung u.Ä. können die o.a. Ressourcen aufgrund des betriebenen Forschungsaufwands, der Expertise und des Rufs der (identifizierten) Fachleute, die an der Erstellung der Ressourcen beteiligt sind, als vertrauenswürdig betrachtet werden. Wenn es gelingt, Definitionen über persistente Identifikatoren im Web zu adressieren, lassen sich diese Definitionen heranziehen, um Aussagen zu validieren, die im expandierenden Semantic Web gemacht werden.

Abbildung 3 zeigt die so genannte ‚Semantic Web Torte‘ Webbasiertes Wissen wird demnach in Unicode-Zeichen dargestellt und in XML, RDF und RDFS beschrieben. Häufig werden dabei Regeln und andere Features verwendet, die von Websystemen zur Verfügung gestellt werden. Die oberste Tortenschicht, d.h. ‚Vertrauen‘, lässt sich nur schwer realisieren. Um diesem Problem Herr zu werden, müssten zuverlässige, adressierbare TC 37-konforme Definitionsrepositorien geschaffen werden, die sich dann zur Validierung von Aussagen, die in bestimmten Fachkontexten gemacht werden, referenzieren lassen.

ISOcat, das TC 37 Datenkategorie-Register

In den zehn Jahren nach der Veröffentlichung der Hardcopy-Norm ISO 12620 (1999) mit ihrer festen Zahl von terminologiebezogenen Datenkategorien ist ein offener, dynamischer Webservice entstanden. Nicht nur Terminologen, sondern allen, die in irgendeiner Weise mit sprachbezogenen Ressourcen arbeiten, wird die Möglichkeit geboten,



(Quelle: <http://www.isocat.org>)

Datenkategorien für einen bestimmten Arbeitsbereich auszuwählen oder selbst zu erstellen. Der Unterausschuss

SC 4 des TC 37 z.B. ist gerade dabei, Metadaten-Elemente für Annotati-onsschemata zu definieren, die für die Darstellung und die Validierung von Sprachressourcen, insbesondere in webgestützten Umgebungen, bestimmt sind (morphosyntaktische und syntak-tische Einheiten, semantische Inhalte, menschliche Sprachpraxis, multilin-guale Texte in Translation Memory-, Segmentierungs-, Globalisierungs-, In-ternationalisierungs-, Lokalisierungs- und Übersetzungsanwendungen, Work-flows in Sprachressourcenmanagement usw.). Das Datenkategorie-Register (DCR) des ISOcat-Projekts bietet eine Plattform, über die Sprachexperten und insbesondere Normungsexperten Da-tenkategorie-spezifikationen ausarbeiten und maßgebende Definitionen für fach-sprachspezifische Ressourcen vereinba-ren können [6].

ISOcat ist vielseitig nutzbar. Hier findet man Datenkategorien für die Modellie-rung von terminologischen Datenban-ken oder sprachbezogenen Annotations-

schemata. Datenkategoriegruppen aus dem DCR, so genannte Data Category Selections (DCS), lassen sich innerhalb der DCR identifizieren oder aus dem Register für das Anlegen von Daten-sammlungen und Annotationsschemata exportieren. Mithilfe der Datenkate-gorien können aber auch RDF-Daten-modelle ausgegeben werden, die als Ressource in ontologischen Modellie-rungsverfahren deklariert werden. Die Verwendung persistenter Identifika-toren – man spricht auch von ‚coolen‘ Unified Ressource Identifiern (URI), da das DCR selbst als persistente Re-source konzipiert sein wird – ermöglicht Entwicklern von Relationsregistern und anderen Wissensressourcen, jede einzel-ne Datenkategorie zu referenzieren und mit semantischen Metadaten-Aussagen zu verknüpfen, die sich auf maßgebende ISOcat-Definitionen stützen, vor allem aber auf solche, die bereits als Norm verabschiedet worden sind. Dies dürf-te dazu beitragen, dass der Vertrauens-komponente in der Semantic Web Torte mehr Gewicht verliehen wird.

Literatur

- [1] Fischer, Norman (1999): In: Wired Magazine, 18-21.
- [2] Sowa, John (2001): Arbor porphyrii according to Sowa. [http://commons.wikimedia.org/wiki/File:Arbor_porphyrii_according_Sowa_\(2001\)_GER.png](http://commons.wikimedia.org/wiki/File:Arbor_porphyrii_according_Sowa_(2001)_GER.png).
- [3] Svenonius, Elaine (2000): The Intellectual Foundation of Information Organization: Digital Libraries and Electronic Publishing. Cambridge, Mass. MIT Press.
- [4] ECCMA (2009): electronic commerce management association. <http://www.eccma.org/iso8000home.php>.
- [5] AeroSpace and Defence Industries Association of Europe (2009): FAQ. <http://www.asd-ste100.org/pagina5.htm>.
- [6] ISOcat (2009): Data Category Registry. <http://www.isocat.org>.

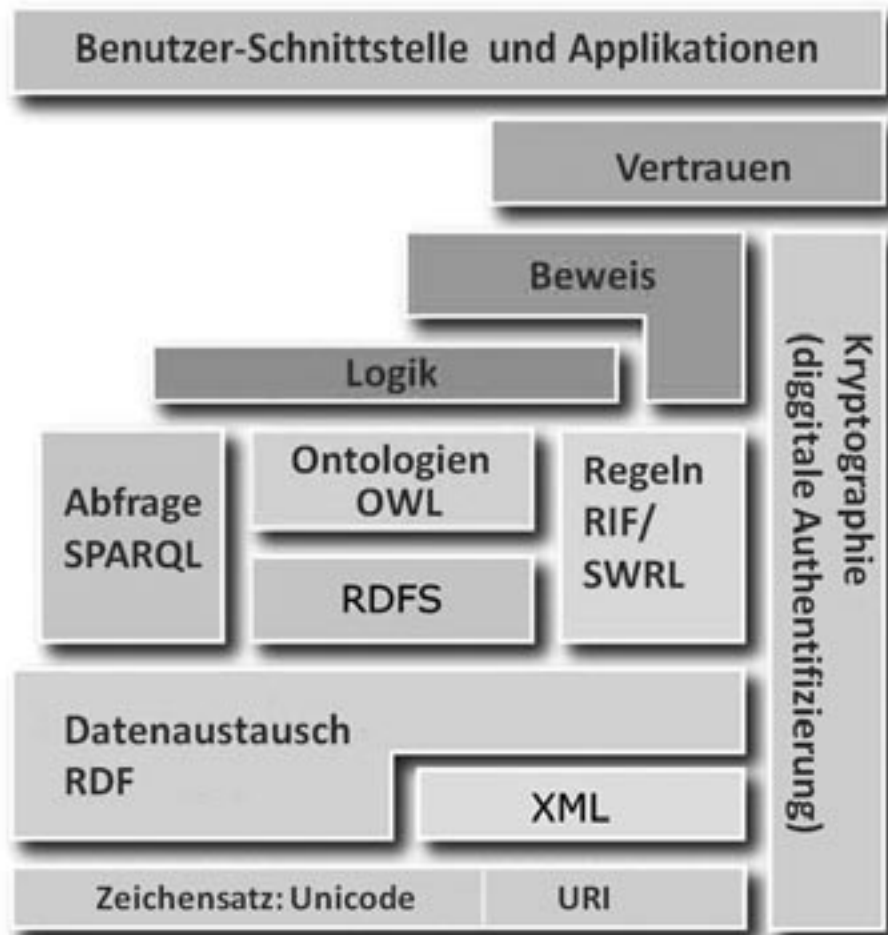


Abbildung 3: Semantic Web Torte (Quelle: <http://www.w3.org/2007/03/layerCake.png>).

Sue Ellen Wright ist Professorin für deutsche Sprache im Institut für Angewandte Sprachwissenschaften an der Kent State University in Kent, Ohio. Sie unterrichtet rechnerunterstützte Anwendungen für Übersetzer und wissenschaftlich-technisches Übersetzen (Deutsch-Englisch). Zusammen mit Prof. Dr. Gerhard Budin hat sie seinerzeit das immer noch richtungsweisende Handbook of Terminology Management herausgegeben. Sie ist zudem stellvertretende Vorsitzende des Unterausschusses 3 „Systeme für die Verwaltung von Terminologie, Wissen und Content“ des technischen Komitees ISO/TC 37 und Vorsitzende des ISOcat DCR-Vorstands.

Kontaktadresse
 Kent State University
 swright@kent.edu
 appling.kent.edu

