

Stefan Keil, Darmstadt

Terminologie Mapping: Grundlagen und aktuelle Normungsvorhaben

Der Artikel basiert auf einer Bachelorarbeit mit dem Titel „Terminologie Mapping: Grundlagen und aktuelle Standardisierung“, die im Sommersemester 2011 an der Hochschule Darmstadt geschrieben wurde. Er befasst sich mit den Grundlagen des Terminologie Mappings, also der Erstellung von Querverbindungen zwischen verschiedenen kontrollierten Vokabularen und der Standardisierung der Mapping-Verfahren. Die Grundlagen sollen anhand der Erkenntnisse aus Doerrs zentraler Publikation aus dem Jahre 2001 (Doerr 2001) dargestellt werden. Darauf folgt ein Blick in den Bereich der Standardisierung, der gerade in den letzten Jahren eine starke Entwicklung durchlaufen hat, was dazu führt, dass immer mehr aktualisierte nationale und internationale Standards herausgegeben werden, die Terminologie Mapping in verschiedener Weise als Mittel zur Erlangung von Interoperabilität empfehlen. Ebenso soll das „Simple Knowledge Organization System“ (SKOS), ein Standard des W3Cs zur Repräsentation von kontrollierten Vokabularen, thematisiert werden. Dieser hat das Ziel, kontrollierte Vokabulare im Semantic Web darzustellen. Obwohl sich bei der Modellierung von klassischen kontrollierten Vokabularen (überwiegend Thesauri) in SKOS noch einige Probleme abzeichnen, stellt dieser Standard einen wichtigen Schritt für kontrollierte Vokabulare aus der Fachinformation in das Semantic Web dar. Abschließend wird ein Fazit zur Standardisierung gezogen sowie ein Ausblick auf die Möglichkeiten der Einbindung von Terminologie Mapping im Semantic Web via SKOS gegeben.

Deskriptoren: Thesaurus, Kreuzkonkordanz, Normung, Rechnernetz, SKOS

Terminology Mapping: Basics and recent standardization

The following article is based on a Bachelor thesis with the title “Terminologie Mapping: Grundlagen und aktuelle Standardisierung” which was written in summer 2011 at Hochschule Darmstadt. The thesis focuses on the basics of terminology mapping which is the creation of cross-linking between different controlled vocabularies. Foundations of terminology mapping can be referred to the seminal publication of Doerr (2001). The second part

is describing the recent standardization efforts which have advanced in the last years. More and more standards on national and international levels have been published which take terminology mapping into account as a method of establishing interoperability. The standard “Simple knowledge organization system” (SKOS) by the W3C is also a topic. This standard aims at representing controlled vocabularies for the Semantic Web. Even though there are still some problems in modeling classic controlled vocabularies (especially thesaurus) in SKOS, this standard can be seen as a huge step forward for vocabularies out of the specialized information domains towards the Semantic Web. In the end we come to a conclusion about the standardization efforts and take a look on the possibilities to integrate terminology mappings in the Semantic Web.

Keywords: semantic web, thesaurus, mapping, standards, SKOS

La cartographie de terminologie: Principes de base et projets actuels de normalisation

L'article est basé sur une thèse de baccalauréat intitulée «La cartographie de terminologie: Principes de base et projets actuels de normalisation» qui a été écrite pendant le semestre d'été 2011 à la Hochschule (école supérieure) de Darmstadt. Cette thèse traite des principes de base de cartographie de terminologie, à savoir l'établissement de liens entre divers vocabulaires contrôlés, et la normalisation du processus de cartographie. Les principes de base sont illustrés au travers des résultats de la publication centrale de Doerr de l'année 2001 (Doerr 2001). Ensuite, l'auteur jette un regard sur le domaine de la normalisation, qui a connu ces dernières années des développements importants conduisant à la publication d'un nombre croissant de normes nationales et internationales qui recommandent diverses possibilités de cartographie de terminologie comme moyen d'interopérabilité. Il explique également le „Simple Knowledge Organization System“ (SKOS), une norme du W3C pour la représentation des vocabulaires contrôlés. Cette norme a comme but de présenter les vocabulaires contrôlés dans le Web sémantique. Bien que l'utilisation du SKOS pour la modélisation

de vocabulaires contrôlés classique pose encore quelques problèmes, cette norme constitue une étape importante pour la représentation de vocabulaires contrôlés dans le domaine de l'information spécialisée dans le Web sémantique. Enfin, l'auteur tire des conclusions sur la normalisation et il décrit les perspectives pour les possibilités d'intégration de la cartographie de terminologie dans le Web sémantique via SKOS.

Mots-clés: thésaurus, vocabulaire contrôlé, norme, cartographie de terminologie, Web sémantique, SKOS

1 Einleitung

Gerade im Bereich der Fachinformation, der in Deutschland durch Bibliotheken und Fachinformationseinrichtungen bedient wird, ist die Wichtigkeit von kontrollierten Vokabularen überwiegend in Form von Klassifikationen und Thesauri umstritten. Während Klassifikationen Wissensbereiche nach Klassen strukturieren, helfen Thesauri, als Mittel der verbalen Erschließung, Mehrdeutigkeiten in der Sprache zu kontrollieren und aufzulösen.

Doch gerade durch die rasante Entwicklung im Bereich des World Wide Webs scheint die Menge an fachgebundenen Vokabularen und universalen Systemen zur Wissensordnung allein nicht auszureichen. Nutzerinnen und Nutzer erwarten, wenn sie eine Recherche beginnen, immer häufiger „One-Stop“-Suchen wie sie Websuchmaschinen anbieten, denn die große Auswahl an verschiedenen Angeboten erschwert die Informationssuche. Wie Mayr (2010, S. 15) beschreibt, sind diese Angebote innerhalb eines Fachgebiets oftmals nicht zusammengefasst und somit nicht über einen einzigen Zugang abrufbar. Der Nutzer ist also gezwungen, sein Informationsbedürfnis über eine Anfrage an verschiedene Systeme zu stellen und dieses entsprechend der zugrunde liegenden Vokabulare oftmals neu zu formulieren. „Dies erhöht die kognitive Belastung des Nutzers bei der Recherche und führt zwangsläufig zu längeren und unterbrochenen Recherchen. Häufig ist die Konsequenz, dass Nutzer aus Bequemlichkeitsgründen ausschließlich mit Internetsuchmaschinen recherchieren und die fachwissenschaftlich erschlossenen Literaturdatenbanken nicht mehr konsultiert werden.“ (Mayr 2010, S. 15) Gerade dieser direkte Vergleich von Fachangeboten mit Internetsuchmaschinen erzeugt Erwartungen an Fachinformationsdienstleistungen und Bibliothekskataloge, die oftmals nicht befriedigend erfüllt werden.

Virtuelle Fachbibliotheken im Sinne von hybriden Bibliotheken mit einem gemischten Bestand aus elektronischen und gedruckten Daten (vgl. Krause 2003, S. 19) versuchen, diese Nutzerbedürfnisse zu befriedigen, können aber die erwartete Menge an Datenbeständen, die gleichzeitig abgefragt werden soll, häufig noch nicht bereitstellen.

Die Schwierigkeit ist hierbei nicht mehr das generelle Zusammenfassen der verteilten Angebote in einem System, sondern der Schritt, dabei auch die Präzision der Suche, die durch kontrollierte Vokabulare gestützt wird, im übergreifenden System zu erhalten. Dies führt zu Überlegungen zur Verknüpfung der einzelnen Vokabulare, mit denen die einzelnen Sammlungen erschlossen wurden, um die ursprüngliche Qualität der Suche beizubehalten.

Würde man diese nicht berücksichtigen, so ist durch den Verlust der Semantik, die durch die Vokabulare hergestellt wird, damit zu rechnen, dass die Suche in einem größeren Suchraum wesentlich ungenauer und schwieriger werden würde. Ein Mittel, um dieses Problem zu lösen, sind Terminologie Mappings, die als semantische Verbindungen zwischen den Klassen und Termen der jeweiligen Systeme eine Brücke darstellen, die es ermöglicht, die Vorteile der oftmals intellektuell erstellten Systeme weiterhin zu gewährleisten.

Krause, der (intellektuell) erstelle Mappings als (Teil-)Lösung für die semantische Heterogenität, also der Existenz und der Nutzung von unterschiedlichen kontrollierten Vokabularen, sieht, plädiert dafür, diesen Zustand, der als die Nutzung von verschiedenen Standards und (Quasi-)Standards beschrieben werden kann, zu akzeptieren. Seine These „Standardisierung von der Heterogenität her denken“ stimmt mit der aktuellen Entwicklung innerhalb der Standardisierung überein. Die aktuellen Revisionen verschiedener nationaler und internationaler Standards, die Thesauri und andere kontrollierte Vokabulare betreffen, zeigen, dass Interoperabilität zunehmend mehr Beachtung findet.

Terminologie Mapping, als Mittel der Heterogenitätsbehandlung, wird im Folgenden nochmals von Grund auf erläutert.

2 Terminologie Mapping: Definition

Doerr (2001) gibt für den Term „Thesaurus Mapping“ eine Basisdefinition, die auch als Definition für das allgemeinere Terminologie Mapping gelten kann: „We regard thesaurus mapping as the process of identifying terms,

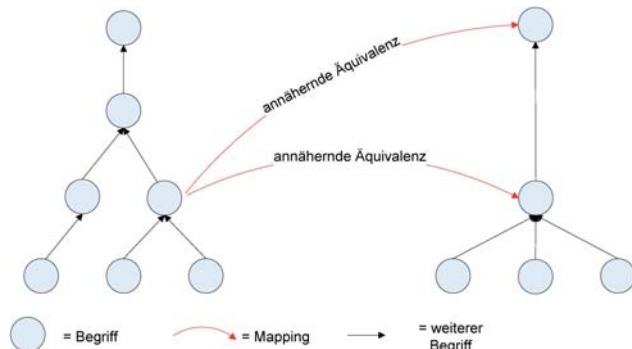


Abb. 1: Annähernde Äquivalenz abgeleitet von Doerr (2001).

concepts and hierarchical relationships that are approximately equivalent“.

Abbildung 1 verdeutlicht diese Definition nochmals. Verschiedene Begriffe, beziehungsweise Konzepte, werden als annähernd äquivalent identifiziert und ermöglichen so eine Aussage über Überschneidungen der Begriffe, genauer gesagt, deren Begriffsinhalten. Ohne detaillierte Spezifizierung sind die Mappings aber noch sehr schwach hinsichtlich ihrer Aussagekraft. Während die Spezifizierung der Beziehungen im Teil 5 folgt, soll der Ansatz des Concept-based Mappings nach Doerr geklärt werden.

3 Concept-based Mapping

Das Concept-based Mapping nähert sich der Thematik in Form einer Vagheits-Problemstellung. Im klassischen Information Retrieval besteht das Problem der Vagheit zwischen Nutzeranfrage (user query) und den durch kontrollierte Terme repräsentierten Dokumenten. Es handelt sich dabei um die terminologischen Unterschiede zwischen Nutzervokabular und Indexierungsvokabular. Durch das Hinzukommen einer weiteren Vagheit, die zwischen den vernetzten Vokabularen auftritt, da ihre Begriffe und Terme nicht immer exakt bedeutungsgleich sind, entsteht eine weitere Unschärfe, die es beim Retrieval in heterogenen Beständen zu beachten gibt. Hellweg et al. (2001, S. 7) verdeutlichen dies mit einem Zwei-Schritt-Modell.

Der erste Schritt stellt die bereits angesprochene Vagheit zwischen Anfragevokabular und Indexierungsvokabular dar. Durch statistische Verfahren des Mappings, die eine Ähnlichkeit zwischen den (Such-)Terminen und dem kontrollierten Vokabular errechnen und dadurch eine entsprechende Annäherung erzeugen, kann diese Vagheit minimiert werden (vgl. Mayr 2010, S. 60f; Mayr, Mutschke, Petras 2008). Der zweite Schritt bezieht sich

auf die Vagheit, die zwischen den verschiedenen Dokumentbeständen beziehungsweise konkreter deren Vokabularen entsteht. Diese Vagheit ist vor allem vier Gründen zuzuschreiben, die Doerr wie folgt benennt:

- **Es werden verschiedene Terme genutzt:** Dies bezieht sich vor allem auf das Problem der Synonymie. Durch verschiedene redaktionelle Überlegungen fallen Entscheidungen bezüglich der Synonymie und Quasi-Synonymie auch in gleichen oder ähnlichen Fachgebieten verschieden aus. Ebenso nimmt der Grad der Prä- und Postkoordination einen starken Einfluss auf die Auswahl der Terme.
- **Die Abdeckungsbereiche der Sammlungen (Scopes) beziehungsweise deren Vokabulare sind verschieden:** Vor allem Thesauri sind Werkzeuge der Erschließung, die eng mit den zu ordnenden Sammlungen verknüpft sind. Je nach Entwicklung des Scopes der Sammlung, variieren Eigenschaften wie zum Beispiel die Erschließungstiefe. Hier wird der oftmals zweckorientierte Charakter deutlich, der (bei Thesauri) Vorrang gegenüber dem Anspruch universaler Wissensordnung in einem Fachgebiet hat.
- **Es existieren verschiedene semantische Strukturen:** Die Prägung von Begriffen und deren Begriffsinhalten variiert. Gerade bei Thesauri aus verschiedenen Sprachen ergeben sich abweichende Begriffsinhalte. Neben diesem Aspekt zählt auch die unterschiedliche Handhabung der Facetten dazu. Meint der Begriff Universität das Gebäude oder die Einrichtung als Lehr- und Forschungsstätte? Je nach Sichtweise werden diese Begriffe, Terme und Klassen an anderen Stellen des Ordnungssystems angesiedelt.
- **Unterschiede hinsichtlich der verwendeten semantischen Relationen:** Generelle Überlegungen zur Definition von Begriffen beeinflussen auch die Anwendungen der Relationen. Doerr betont aber vor allem, dass monohierarchische Bestrebungen oftmals dazu führen, dass unabhängig vom vorherigen Punkt auch verschiedene Entscheidungen getroffen werden, was die Stellung eines Begriffs in einem Ordnungssystem betrifft, um Polyhierarchien zu vermeiden. Diese Stellung beeinflusst die Relationen entsprechend.

Um mit dieser Heterogenität der Systeme umgehen zu können, entwickelt Doerr die Theorie des Concept-based Mappings, die sich auf das Prinzip der Beschreibungslogik stützt.

Ausgangspunkt ist die Annahme, dass Deskriptoren durch ihre Begriffe identifiziert werden. Diese werden wiederum durch die Menge an Dokumenten, die sie beschreiben, geprägt. Folgt man dieser Definition, so ist der bestimmende Ausgangspunkt für den Begriffsinhalt eines Deskriptors jeweils die Menge der Dokumente, die die folglich untergeordneten Instanzen (Begriffe und Deskriptoren) prägt. Daher ist es möglich das Mapping beziehungsweise das Identifizieren von Äquivalenzen zwischen Begriffen und Relationen als Abgleichen von Dokumentenmengen anzusehen. Daraus leitet Doerr folgende Prinzipien für das „concept-based Mapping“ ab:

- Ein Begriff wird auf die mit ihm assoziierte Menge der Objekte gemapped, die er somit richtig beschreibt.
- Die mit dem Begriff verbundenen weiteren und engeren Begriffe sind zusammenfassende Mengen oder Teilmengen von Objekten des Terms. Der weitere Begriff fasst also seine engeren Begriffe und damit die zugeordneten Objekte zusammen.
- Das Mapping zwischen zwei Begriffen wird durch die Beziehung der ihnen zugeordneten Objektmengen bestimmt.

Diese Definition von Mapping ist strikt und kann mathematischer durchgeführt werden als zum Beispiel die Definition von Crosskonkordanzen nach Krause und beinhaltet keine Einschränkung hinsichtlich der methodischen Umsetzung (intellektuell oder automatisch), die Krause explizit vornimmt. Krause beschreibt Crosskonkordanzen als „intellektuell erstellte Verbindungen zwischen Ter- men zweier Thesauri oder Klassifikationen“ (Krause 2004, S. 81).¹ Das Mapping durch Crosskonkordanzen wird also auf der Begriffsebene zwischen den Vokabularen durchgeführt und beschäftigt sich nicht explizit mit den Dokumentenmengen, die mit den Deskriptoren indexiert wurden sind. Es steht also die Interpretation eines Deskriptors durch seine Anwendung auf eine konkrete Dokumentenmenge gegen die Interpretation eines Deskriptors als Begriff mit seinem beabsichtigten Begriffsinhalt. Es ist aber anzumerken, dass bei einer konsistenten Zuteilung von Deskriptoren bei der jeweiligen Indexierung in beiden Vokabularen die Überschneidung der Dokumentenmengen eine logische Konsequenz ist.

4 Relationen

Eine weitere Besonderheit bei Doerrs Publikation ist die Definition eines Sets an Relationen, die bei einem Mapping angewendet werden können. Diese Relationen basieren auf den Äquivalenz-Relationen der ISO 5964 (somit auch der DIN 1463 – Teil 2, die nahezu identisch ist) und finden sich in dieser und oftmals auch leicht abgewandelter Form in einer Vielzahl von Mapping-Projekten wieder.

Wie bereits in 3 dargelegt, kann das Erkennen von Äquivalenzen nur wirklich aussagekräftig sein, wenn die Art der Äquivalenz auch genauer spezifiziert wird. Doerr schlägt folgende Relationen, in Anlehnung an die ISO 5964, vor:

- Die genaue Äquivalenz bleibt in dieser Form bestehen, ist aber in Sinne von Concept-based Mapping als „gleiche Objektmengen zugeordnet“ zu verstehen.
- Die vergleichsweise unschärferen Teil-Äquivalenz sollte in „weitere Äquivalenz“ und „engere Äquivalenz“ geändert werden. (Teilmenge und zusammenfassende Menge)
- Die unscharfe Äquivalenz sollte sich als „Überschneidungen mit“ verstehen.
- Die Äquivalenz zwischen einer und mehreren Bezeichnungen sollte spezifiziert werden als zusammen gesetzte Äquivalenz („compound equivalence“) mit Verwendung der booleschen Operatoren AND, OR sowie NOT. Die vorher aufgeführten Abstufungen wie „genaue“, „Teil-“, und „weitere“ oder „engere“ Äquivalenz können hierbei ebenso angewendet werden.

Auffallend sind die expliziteren Kennzeichnungen der weiteren beziehungsweise engeren Äquivalenz sowie die Einführung von booleschen Operatoren innerhalb der „compound equivalence“. Die Einführung der Operatoren rechtfertigt Doerr mit der Aussagekraft der Beziehungen. Da es die meisten Systeme dem Nutzer ermöglichen, komplexe Anfragen mittels boolescher Operatoren zu stellen, ist es sinnvoll, dass das Mapping diese komplexen Anfragen umsetzen kann. Andernfalls könnten Nutzer komplexe Anfragen stellen, die der Übersetzungsmechanismus nicht verarbeiten kann. Diese Argumentation ist wiederum stark von seiner Definition des „Concept-based Mapping“ geprägt und der darin angesprochenen Sichtweise, dass Mappings anhand der Dokumentenmengen, die mit dem entsprechenden Deskriptor indexiert wurden, vollzogen werden. Der Einsatz der booleschen Operatoren ist also analog zur Anwendung im Information Retrieval

¹ Der Unterschied ergibt sich vor allem auch daraus, dass Krause Crosskonkordanzen nur als einen Teil zur Lösung der Heterogenität zwischen Beständen benennt. Näheres in Krause (2003, 2004).

zu sehen, bei dem sich die Ergebnismengen bei komplexen Anfragen auch aus Schnitt- oder Vereinigungsmengen ergeben.

5 Methodische Empfehlung

Aufgrund der Heterogenität zwischen den verschiedenen Vokabularen, stellt Doerr fest, dass es keine Garantie für das Vorhandensein von Äquivalenzrelationen für alle Deskriptoren aus zwei verschiedenen Vokabularen gibt. Fehlt eine entsprechende Relation, so wäre die Konsequenz, dass Teile der Datenbestände nicht abgefragt werden können, weil die Übersetzung nicht stattfindet. Des Weiteren ist zu bedenken, dass bei einem Einstieg mit einem Deskriptor aus einem Thesaurus nicht ersichtlich ist, inwiefern die Suche durch das Nicht-Vorhandensein einer exakten Äquivalenzbeziehung beeinflusst wird. Führt die Übersetzung durch das Mapping über eine hierarchisch sehr starke Umformung auf einen weit entfernten weiteren Begriff im Zielvokabular und es werden somit mehrere Stufen in der Hierarchieleiter übersprungen, so wird ein sehr extremer Trefferzuwachs festzustellen sein, der sehr zulasten der Precision fallen kann. Um diese Problematiken zu beseitigen, beziehungsweise etwas berechenbarer zu machen, stellt Doerr folgende Regeln auf (im Folgenden verkürzt):

1. Wenn es nicht möglich ist einen Begriff c aus dem Thesaurus A mit einer exakten Äquivalenz abzubilden, dann sollte für diesen Begriff mindestens eine weitere und eine engere Äquivalenz gefunden werden
2. Die weitere Äquivalenz sollte minimal sein
3. Die engere Äquivalenz sollte maximal sein
4. Der Nutzer sollte die Möglichkeit haben seine Anfrage zu spezifizieren, falls diese Änderungen vor einem niedrigen Recall oder einer niedrige Precision bewahrt

Basierend auf diesen Regeln leitet Doerr folgende Zustände für Mappings ab. Wenn die erste Regel erfüllt ist, dann kann man von einem kompletten Mapping sprechen. Es ist also gewährleistet, dass jeder Term aus einem Vokabular A in das Vokabular B übersetzt wird. Für den Fall, dass keine „exakte Äquivalenz“-Relation festgestellt werden kann und auch keine weitere beziehungsweise engere Äquivalenz-Beziehung möglich ist, müsste das „universal concept“ eingesetzt werden, das nach Doerr das gesamte Zielvokabular (Vokabular B) umfasst. Dies ist natürlich

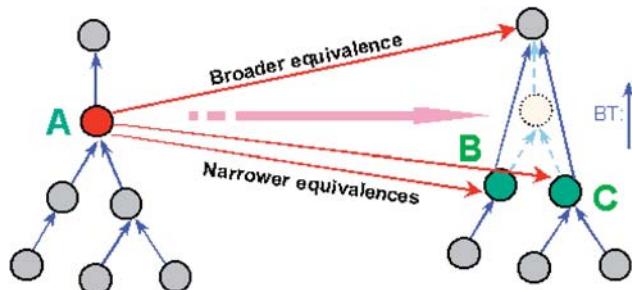


Abb. 2: Abbildung eines Terms durch Mapping aus Doerr 2001.

ein Fall, der zu vermeiden ist, obwohl Doerr pragmatische Lösungen vorschlägt, wie einen „query refinement cycle“, der dem Nutzer die Möglichkeit bietet, seine Anfrage zu ändern.

Die zweite und auch die dritte Regel zielen auf die Genauigkeit des Mappings ab. Doerr spricht bei der Einhaltung dieser Regeln von einem optimalen Mapping („optimal Mapping“).

Regel Vier stellt nochmals die Forderung nach einer Möglichkeit zur manuellen Anfrageverfeinerung dar.

Was die Nutzung der Beziehungen bedeutet und wie sich die Regeln auf das Mapping auswirken, verdeutlicht Abbildung 2.

Diese Abbildung verdeutlicht viele der genannten Aspekte. Man erkennt die Abbildung des Terms A über insgesamt drei Relationen zu Termen des Zielvokabulars. Die engeren Äquivalenz-Relationen zu den Termen B und C stellen eine zusammengesetzte, engere Äquivalenz dar, die zum Beispiel mit dem Operator „and“ verknüpft ist. Da hierbei auch durch die Mappings die spezifischen, semantischen Beziehungen beachtet werden, spricht Doerr von „semantic term inclusion“ (übersetzt sinngemäß „semantische Term-Aufnahme“). Man erkennt, dass sich durch die engere und weitere Äquivalenz-Beziehungen ein theoretisches Abbild des Begriffs A ergibt, das zwischen den entsprechenden weiteren und engeren Begriffen liegt. Im Vergleich zu Abbildung 1 lassen sich wesentlich konkretere Aussagen treffen.

Diese Grundlagen finden sich in vielen der durchgeführten Mapping-Projekte in verschiedenen Formen und Abwandlungen wieder. Auch im folgenden Teil, der sich mit der Thematik der Standardisierung beschäftigt, werden starke Ähnlichkeiten und auch Übereinstimmungen zu finden sein.

6 Standardisierung

Die Standardisierung von komplexen Sachverhalten ist immer ein relativ schwieriges Unterfangen. Gerade bei Standards, deren Inhalte nicht zwingend verpflichtend sind. So enthalten Industriestandards oftmals Anforderungen, deren Erfüllung notwendig ist, um zum Beispiel ein anwendbares Produkt zu erstellen. So haben bei der Herstellung von Druck- und Schreibpapier fest definierte Größen, wie zum Beispiel DIN A4 (210 × 297 mm) im deutschen Raum, nicht nur empfehlenden Charakter. Stellt man Papier her, das keiner genormten Größe entspricht, ist es unwahrscheinlich, dass man Abnehmer findet, weil es sich nicht für Drucker und Ähnliches eignet. Anders verhält es sich bei einem Standard der sich mit Terminologie Mapping befasst. Hierbei handelt es sich eher um eine Anleitung, die Methoden und Grundlagen aus der Forschungspraxis aufbereitet und die Methoden, die sich als gangbar und auch zielführend erwiesen haben, darstellt, damit Käufer des Standards diese nachvollziehen können. Man könnte auch sagen, dass versucht wird, eine größtmögliche Nachnutzung der aus den (Forschungs-) Projekten gewonnenen Erkenntnisse, zu erreichen. Bedenkt man, dass die Zielgruppe in diesem Fall nicht nur Experten aus dem Fachgebiet sind, sondern potenzielle Interessenten auch in staatlichen Institutionen und der freien Wirtschaft zu finden sind, wird deutlich, dass die Verständlichkeit einen sehr hohen Stellenwert hat. Oftmals sind Mappings ein Mittel zur Datenintegration von Altbeständen oder zur Verknüpfung von verschiedenen Informationssystemen, die zusammengelegt werden. Dextre Clarke (2011) deutet an, dass in solchen Fällen zwar häufig die technischen Möglichkeiten gegeben sind, um auch komplexe Mappings umzusetzen, aber oftmals das Verständnis bei der Durchführung (häufig auch unter Zeitdruck) zu voreiligen Relationen führt, die überwiegend Äquivalenz-Beziehungen zwischen homographisch identischen Termen beinhalten. Dies verdeutlicht auch, dass für annähernd exakte Ergebnisse ein Verständnis für die verbale und auch klassifikatorische Erschließung Voraussetzung ist.

Daher ist die Standardisierung von Mappings ein thematischer Komplex, der nicht unabhängig von den Beziehungen innerhalb eines kontrollierten Vokabulars betrachtet werden sollte. Wie aus Doerrs Definition und ähnlichen Projekten hervorgeht, wurde sich bereits in der Praxis an den Beziehungstypen der internen Struktur von kontrollierten Vokabularen (vorwiegend Thesauri) orientiert. Daher ist es nicht verwunderlich, dass die Standardisierung der Mapping-Beziehungen bei der Revision der

Normen, die Thesauri und andere Vokabulare betreffen, ebenfalls mit behandelt wird. Schließlich ist Mapping ein Thema, das als Mittel zur Interoperabilität von Vokabularen diesen angegliedert ist. Leider war es nur möglich, den amerikanischen Standard (ANSI/NISO Z39.19) einzusehen. Die Einsicht in „ISO 25964-2 Thesauri and interoperability with other vocabularies – Part 2: Interoperability with other vocabularies“, die sich in der Entwicklung befindet, war nicht gegeben. Daher soll der Weg zur Standardisierung und der Stand des Entwurfes der ISO 25964-2 anhand von zwei Publikationen (Dextre Clarke 2007, 2011) dargestellt werden, die sich mit der Thematik befassen und auch Einblick in den (damaligen) Entwicklungsstand geben.

Betrachtet man zu Beginn die Ausgangssituation, so stellt man fest, dass zum einen Handlungsbedarf nötig ist (oder war) und zum anderen auch in den letzten Jahren einige Schritte unternommen wurden. Beispielsweise sind die deutsche Norm DIN 1463-1 „Erstellung und Weiterentwicklung von Thesauri; Einsprachige Thesauri“ mit ihrer letzten Revision 1987 sowie der zugehörige Teil DIN 1463-2 „Erstellung und Weiterentwicklung von Thesauri; Mehrsprachige Thesauri“ von 1993 bezeichnend für die seinerzeitige Lage der Normung. Durch die Entwicklung des World Wide Web und der damit einhergehenden neuen Anforderungen an Thesauri ergibt sich ein Nachholbedarf. Gedruckte Thesauri, die in den alten Revisionen noch im Fokus stehen, finden heute immer seltener Anwendung und werden gerade von Nutzern der Systeme, die keine Informationsexperten sind, oftmals ignoriert. Von Interoperabilität ist im damaligen Entwurf, im heutigen Sinne, noch keine Rede. Es sei angemerkt, dass diese beiden Dokumente fast identisch mit der ISO 2788 „Guidelines for the establishment and development of monolingual thesauri (1986)“ und der ISO 5964 „Guidelines for the establishment and development of multilingual thesauri (1985)“ sind.

Die jüngsten Entwicklungen finden sich in Großbritannien mit dem fünfteiligen British Standard 8723, der schrittweise in den Jahren 2005 bis 2008 veröffentlicht wurde, sowie dem US-amerikanischen Standard ANSI/NISO Z39.19, der in der aktuell gültigen Form 2005 verabschiedet wurde.²

Beide Standards behandeln Interoperabilität und sprechen die Thematik des Mappings an. Interoperabilität wird innerhalb des Z39.19 in einem sehr weiten Rahmen angesprochen. Nach der Verdeutlichung, warum diese

² Details zur Entwicklung (vor allem des Britischen Standards) in Dextre Clarke 2007.

nötig ist, werden verschiedene Methoden zur Realisierung kurz vorgestellt und mit Beispielen verdeutlicht, aber im Rahmen der eigentlichen Norm nicht weiterführend behandelt. Für weiterführende Informationen ist Appendix D „Methods for Achieving Interoperability“ angefügt, der auf einer Publikation von Zeng und Chan (2004) basiert. Mapping spielt hier eine eher untergeordnete Rolle, weshalb wohl auch keine konkreten Mapping-Relationen für diesen Anwendungsfall eingeführt werden.

Der britische Standard BS 8723 ist in fünf Dokumente gegliedert. Einschlägig im Sinne des Artikels ist vor allem Teil vier „BS 8723 Part 4: Interoperability between vocabularies“. Erstmals wird die Standardisierung von Mapping-Beziehungen konkretisiert und im Vergleich zum ANSI/NISO Z39.19 tiefergehend behandelt. Der Begriff Mapping wird als „statement of the relationships between the terms, notations or concepts of one vocabulary and those of another“ (Dextre Clarke 2007, S. 449) definiert. Im Vergleich zu Doerrs Publikation sei angemerkt, dass im Standard nicht nur Thesauri im Fokus stehen, sondern auch weitere kontrollierte Vokabulare, wie Klassifikationen. Beachtet man die bisherigen Projekte, wie KoMoHe (Mayr; Petras 2008), Renardus (Becker; Neuroth 2002) oder Criss-Cross (Hubrich 2008; Müller; Jacobs; Mengel 2010), wird deutlich, dass es sinnvoll ist, ein Mapping nicht nur zwischen gleichen Typen von Vokabularen durchzuführen.

Darauf aufbauend werden drei Modelle zur Erlangung von Interoperabilität dargestellt. Das erste Modell „structural unity“ bezieht sich relativ eng auf multilinguale Thesauri, die im Wesentlichen die gleiche Struktur besitzen. Dies bezieht sich auch auf die Hierarchieleitern für einzelne Begriffe. Variieren diese, so ist keine Symmetrie mehr gegeben und man spricht von asymmetrischen, multilingualen Thesauri (vgl. Zeng 2009, S. 11). Die weiteren Modelle „non-equivalent pairs“ und „backbone model“ entsprechen Methoden, die sich ebenso in Zeng (2004) und auch in der ANSI/NISO Z39.19 finden. Hierbei spielt das Mapping die übergreifende Rolle. Die Methode des „non-equivalent pairs“ (BS 8723), beziehungsweise des „direct mapping“ (Zeng 2004) bedeutet, dass alle Vokabulare, die in das Mapping involviert sein sollten, direkt mit jedem anderen teilnehmenden Vokabular vernetzt werden. Das „backbone model“ (BS 8723), auch „switching“ (Zeng 2004) genannt, schlägt vor, die beteiligten Vokabulare über ein Vokabular zu vernetzen, das als Vermittler und Übersetzer zwischen allen Beteiligten fungieren kann. Ein sehr anschauliches Beispiel hierfür ist das Projekt Renardus, bei dem mehrere Klassifikationssysteme über die DDC verbunden wurden. Universelle kontrollierte Vokabulare bieten sich für ein solches Vorhaben

an, weil sie viele Fachgebiete umfassen und es somit ermöglichen, Vokabulare aus verschiedenen Fachgebieten zu verbinden.

Die grundlegenden Mapping-Beziehungen im BS 8723 sind Äquivalenz, Hierarchie und Assoziation. Die hierarchischen Beziehungen werden mit „narrower“ oder „broader“ spezifiziert, während die Äquivalenz-Beziehung eine einfache Beziehung zwischen zwei Termen ist oder auch durchaus aus einer 1:n-Relation, also einer zusammengefassten Äquivalenz, bestehen darf. Abstufungen der Äquivalenz-Beziehung werden jedoch nur angesprochen und nicht als fester Bestandteil der Beziehungen eingeführt.

Nachdem zwei nationale Standards, die im letzten Jahrzehnt eine Revision erfahren haben, mit dem Fokus auf das Mapping und dessen Relationen umrisse wurden, folgt die ISO 25964 „Thesauri and Interoperability with other vocabularies“.

Die ISO 25964 „Thesauri and Interoperability with other vocabularies“ hat, wie bereits beschrieben, das Ziel, die ISO 2788 und ISO 5964 abzulösen.³ Dextre Clarke (2011) bemerkt zudem, dass es sich bereits um eine Revision des BS 8723 handelt. Obwohl entschieden wurde, sich am BS 8723 zu orientieren, wird es sich um einen zweiteiligen Standard handeln. Teil eins ISO 25964-1 „Thesauri for information retrieval“ soll alle Themenbereiche für Thesauri abdecken inklusive eines Datenmodells, relevante Protokolle und Austauschformate. Teil zwei ISO 25964-2 „Interoperability with other vocabularies“ betrifft die Interoperabilität mit anderen Vokabularen und soll auch Anleitungen zum Mapping zwischen Thesauri und anderen Vokabularen beinhalten (vgl. National Information Standards Organization: Project ISO 25964).

Mapping wird innerhalb des Standards als „relationship between a concept in one vocabulary and one or more concepts in another“⁴ (Dextre Clarke 2011a) beschrieben. Die Grundtypen für Mapping-Beziehungen zeigt Tabelle 1.

³ Alle Darstellungen bzgl. der ISO 25964 basieren überwiegend auf einer Publikation von Dextre Clarke aus dem Jahr 2011, die ein Konferenzbeitrag zur Tagung „Concepts in Context 2010“ beschreibt. Abweichungen dieser Beschreibungen zum erscheinenden Standard sind möglich.

⁴ Diese, sowie alle folgenden Beschreibungen der ISO 25464 beziehen sich auf einen Entwurf, der zum Zeitpunkt des Schreibens der vorliegenden Publikation (Dextre Clarke 2011b) noch nicht verabschiedet war. Daher ist es möglich, dass der endgültige Standard vom hier Beschriebenen abweicht.

Mapping type	Tag	Example
Equivalence	EQ	laptop computers EQ notebook computers
Hierarchical	NM	roads NM streets
	BM	streets BM roads
Associative	RM	journals RM magazines

Tabelle 1: Mapping-Beziehungen ISO 25964-2 aus (Dextre Clarke 2011, 2011a).

Die in Spalte 2 gezeigten Tags sind an die üblichen Thesauruskürzel „BT“, „NT“ und „RT“ angelehnt und stehen für „Equivalence“ (Äquivalenz), „Narrower Mapping“ (engeres Mapping oder engere Äquivalenz), „Broader Mapping“ (weiteres Mapping oder weitere Äquivalenz) und „Related Mapping“ (assoziatives Mapping). Die Äquivalenz-Beziehung, die die Stellung einer (Quasi-)Synonym-Beziehung innerhalb eines Thesaurus einnimmt, unterscheidet sich jedoch geringfügig von dieser. Während die Kürzel USE und UF Äquivalenzbeziehungen zwischen Termen, damit also Synonymie, ausdrücken, handelt es sich bei einem Äquivalenz Mapping um eine Beziehung, die eine Äquivalenz zwischen Begriffen ausdrückt, die als gleichwertig anzusehen sind. Die Äquivalenz-Beziehung benötigt somit keinen Antagonisten, der den Umkehrschluss darstellt. Bei einer Beziehung mit USE, die auf einen Deskriptor verweist, wird in einem Thesaurus immer ein entsprechender UF-Verweis bei dem zugehörigen Nicht-Deskriptor verlangt (vgl. Dextre Clarke 2011).

Neben diesen Grundtypen können weitere Spezifikationen der Beziehungen angegeben werden, wobei diese als optional gekennzeichnet sind und nicht immer gefordert werden.

Zum einen handelt es sich dabei um die „compound“-Äquivalenz, also die Äquivalenz zwischen einem Begriff und mehreren anderen Begriffen. Sie entspricht Doerrs Defintion und seiner Terminologie (siehe unter 5. „Relationen“) und fand zum Beispiel in ähnlicher Form auch im KoMoHe-Projekt (Mayr, Petras 2008) Anwendung. Für das Verständnis könnten durchaus, wie bei Doerr, die booleschen Operatoren „and“ und „or“ für die Auszeichnung einer zusammengesetzten Äquivalenz genutzt werden. Um aber Verwirrungen zu vermeiden wurden die Tags „+“ für „and“ (intersecting) und „|“ für „or“ (cumulative) eingefügt.

Auch die Hierarchie-Relation kann als generisch, partitiv und als hierarchische Instanz-Relation gekennzeichnet werden, die den auch in Thesauri möglichen Relationen entsprechen (vgl. Dextre Clarke 2011a).

Nachdem bisher Normen von klassischen Normungsorganen wie der „International Standard Organization“ vorgestellt wurden, soll ein weiterer Standard aus einer anderen Community angesprochen werden. Das W3C (World Wide Web Consortium) ist eine internationale Gemeinschaft, die es sich zur Aufgabe gemacht hat, durch die Entwicklung von Protokollen und Richtlinien, das Web so zu gestalten, dass alle Möglichkeiten, die es bieten kann, auch ausgeschöpft werden (vgl. [w3.org Mission Statement](http://w3.org/Mission-Statement)).

Durch die Entwicklung des Standards „Simple Knowledge Organization System“ (kurz: SKOS) hat diese Community eine Möglichkeit geschaffen, kontrollierte Vokabulare einheitlich im Web bzw. dem Semantic Web darzustellen und somit auch einen wichtigen Beitrag zum Austausch von kontrollierten Vokabularen und auch deren Interoperabilität geleistet. SKOS wird im Folgenden kurz beschrieben, wobei der Fokus auf den möglichen Beziehungen liegt, die innerhalb der Vokabulare und auch zwischen ihnen angewendet werden können.

SKOS erhebt nicht den Anspruch, ein neues Vokabular darzustellen, und versucht auch nicht, die etablierten kontrollierten Vokabulare zu ersetzen. „The aim of SKOS is not to replace original conceptual vocabularies in their initial context of use, but to allow them to be ported to a shared space, based on a simplified model, enabling wider re-use and better interoperability.“ (W3C SKOS – Primer 2009)

Daher überrascht es nicht, dass die Beziehungen, die in SKOS angewendet werden können, keine großen Unterschiede zu den bereits bekannten aufweisen.

Mapping-Beziehung Erklärung

skos:closematch	„closematch“ drückt eine semantische Ähnlichkeit zwischen zwei Begriffen aus unterschiedlichen KOS aus. (nicht transitiv)
skos:exactmatch	„exactmatch“ ist eine Verstärkung der „closematch“-Beziehung. Es wird davon ausgegangen, dass die Ähnlichkeit der Begriffe noch ausgeprägter ist. (transitiv)
skos:broadmatch	„broadmatch“ entspricht einer üblichen „weiterer Begriff“ bzw. „weiteres Mapping“-Beziehung.
skos:narrowmatch	„narrowmatch“ entspricht einer üblichen „engerer Begriff“ bzw. „engeres Mapping“-Beziehung
skos:relatedmatch	„related match“ stellt eine assoziative Beziehung zwischen zwei Begriffen dar.

Tabelle 2: Mapping-Beziehungen in SKOS, basierend auf (W3C SKOS – Primer 2009, Dextre Clarke 2011).

Tabelle 2 stellt alle in SKOS zugelassenen Mapping-Beziehungen kurz dar. Als Besonderheit ist die Differenzierung zwischen „closematch“ und „exactmatch“ anzusehen, die sich hauptsächlich in der Eigenschaft transitiv bzw. nicht-transitiv zu sein niederschlägt. Dies lässt deutlich erkennen, dass man sich bei der Erarbeitung des Standards sehr bewusst an technischen Rahmenbedingungen orientiert hat. Ist eine Beziehung transitiv, so ergibt sich bei einer Beziehung zwischen drei Begriffen A1, A2 und A3, die durch A1 „skos:exactmatch“ A2 und A2 „skos:exactmatch“ A3 in Relation zu einander stehen, automatisch die dritte Beziehung A1 „skos:exactmatch“ A3. Je nach Implementierung des Mappings, können solche Feinheiten einen sehr großen Unterschied für das Retrieval beziehungsweise die Transformation der Begriffe bedeuten.

Nachdem in den aktualisierten klassischen Standards darauf hingewiesen wurde, dass die Annahme der Symmetrie von Beziehungen zwischen Begriffen aus verschiedenen Vokabularen nicht immer zutreffend ist, soll in diesem Zusammenhang auf zwei Besonderheiten von SKOS hingewiesen werden.

SKOS ist so modelliert, dass, sofern kein inverses Mapping vorliegt, die Symmetrie der Beziehung angenommen wird (vgl. Dextre Clarke 2011). Vor allem bei compound-Mappings erweist sich diese Annahme oftmals als fehlerhaft und kann nicht immer als korrekt angesehen werden. Dies führt aber auch zur weiteren Feststellung, dass SKOS in der grundlegenden Form keine compound-Mappings vorsieht. Wie Mayr, Zapilko und Sure (2010, S. 5f.) darstellen, bietet für eine Modellierung mit zusammengesetzten Begriffen, die Nutzung einer Erweiterung (in diesem Fall „eXtension for Labels“) eine Lösung. Durch das Fehlen von zusammengesetzten Begriffsbeziehungen im Standard SKOS scheint die Annahme der Symmetrie weniger problematisch für die Qualität des Mappings. Dennoch empfiehlt sich eine sorgfältige Prüfung.

Die angesprochene Erweiterung „eXtension for Labels“ spielt auch bei einer weiteren Besonderheit eine große Rolle. Wie Mayr, Zapilko und Sure (2010) detailliert darstellen, ist SKOS konzept-basiert bzw. begriffsbasiert, was bei der Modellierung von Vokabularen wie Thesauri eine gewisse Hürde birgt. Wie allgemein gültig sind innerhalb eines Thesaurus Term-Term-Beziehungen eines der wichtigsten Mittel zur terminologischen Kontrolle (Festlegung von (Quasi-)Synonymen und bevorzugten Benennungen). Die Erweiterung ermöglicht es, auch „Use“- und „Used-for“-Beziehungen innerhalb der SKOS-Modellierung darzustellen.

Wie man aus dem Beschriebenen nun ableiten kann, stellt SKOS keine wesentliche Neuerung hinsichtlich der

Prinzipien von kontrollierten Vokabularen dar. Seine Stärke bzw. Besonderheit ist die Repräsentation in einem Format, das für Semantic-Web-Anwendungen, und generell für den Austausch von kontrollierten Vokabularen über das Web, geeignet ist. Die angesprochenen Problematiken sind, wie einige Versuche in der Praxis zeigen, lösbar und eine Weiterentwicklung des Standards wird in der Fachgemeinschaft intensiv diskutiert. Es ist abzusehen, dass weitere Änderungen durch den Austausch zwischen Informationseinrichtungen (Fachinformationszentren, Bibliotheken u. a.) und der Webcommunity erfolgen, die die angesprochenen Problematiken lösen.

7 Fazit

Wie die Darstellungen zur Standardisierung in diesem Artikel gezeigt haben, sind Dokumentationssprachen in Form von kontrollierten Vokabularen immer noch ein wichtiges Thema. Auffällig ist aber, dass der Fokus nicht mehr nur auf den Deskriptoren und ihren Beziehungen untereinander liegt, sondern viel Aufmerksamkeit der Interoperabilität gewidmet wird, um eine gewinnbringende Nutzung unterschiedlicher Vokabulare bei einer Suchanfrage zu erreichen. Die durchgeführte Evaluation der Crosskonkordanzen als Mittel zur Heterogenitätsbehandlung (Mayr 2010) zeigt, dass Terminologie Mapping durchaus ein geeignetes Mittel ist, um dieses Ziel zu realisieren.

Nachdem im engeren Bereich der kontrollierten Vokabulare durch langjährige Forschungsarbeit bereits viele Erkenntnisse gewonnen wurden, ist das Terminologie Mapping noch in einer relativ frühen Phase. Die Standardisierung verläuft aber bereits nach einem ähnlichen Muster wie bei Thesauri. Durch verschiedene Projekte, die Prototypen wie zum Beispiel den TEST-Thesaurus hervorgebracht haben, wird versucht, allgemein gültige Regeln und Empfehlungen zu erarbeiten, die für nachfolgende Bemühungen in diesem Bereich genutzt werden können (vgl. Aitchison; Dextre Clarke 2004, S. 6ff.).

Normung und Standardisierung sind wichtige Grundlagen, um eine große Interoperabilität der Vokabulare zu erreichen. Hierbei spielt es vorerst keine Rolle, ob diese Standards von klassischen Normungsgremien stammen oder in Gremien wie dem W3C entstehen. Leider bergen klassische Standards eine große Hürde, die sich im Preis äußert. Die hohen Preise stellen eine Barriere dar, die die Nutzung einschränken kann. Neue Finanzierungsmodelle, die zu niedrigeren oder überhaupt keinen Kosten führen, wären hierbei von Vorteil.

Die neueren Entwicklungen, wie das Semantic Web oder konkreter die Linked Open Data-Initiative, bieten neue Methoden, um Mappings zu nutzen und ein Netz aus Daten entstehen zu lassen.

Die stetig wachsende Linked Open Data-Cloud (LOD-Cloud)⁵ zeigt, dass Bibliotheken und Informationseinrichtungen sich an dieser Entwicklung beteiligen. Überwiegend handelt es sich bei den bisher publizierten Daten um Normdaten, wie Vokabulare. Doch die Deutsche Nationalbibliothek (DNB) plant weitere Veröffentlichungen mit bibliographischen Daten (Altenhöner et al. 2010). Diese Veröffentlichungen tragen zur Qualität der Daten bei und Altenhöner et al. sprechen vom Aufbau eines „Global Cultural Graph“, der wohl in Anlehnung an Tim Berners Lees „Giant Global Graph“ einen Teil der Daten kennzeichnen soll, die überwiegend aus Bibliotheken, Informations- und Kultureinrichtungen stammen und durch strikte Regeln eine große Verlässlichkeit in mehreren Punkten aufweisen.

Dies führt zu Überlegungen, inwiefern das entstehende Netz aus Daten unterschiedlicher Herkunft und Erstellung unter anderem für das Retrieval genutzt werden kann.

Altenhöner et al. (2010) stellen hierbei vor allem die Verlässlichkeit und Persistenz der Daten in den Vordergrund und verweisen auf ein Schalenmodell, das in den Grundzügen dem von Krause (1996) ähnelt.

Krause (1996, S.18) beschreibt allgemein ein Modell, das beginnend von einem Kernbereich, der „hohe Relevanz [besitzt und] tief und qualitativ hochwertig erschlossen [ist]“, weitere Schalen von niedrigerer Relevanz und Erschließungstiefe definiert. Dieser Ansatz ermöglicht es bei einschlägigen virtuellen Fachbibliotheken auch Quellen einzuschließen, die bei einem Modell, bei dem nur gleichwertig erschlossene Daten beachtet werden, nicht eingebunden werden würden.

Altenhöner et al. (2011, S. 70) definieren im Sinne dieses Modells drei Schalen, die von einem stabilen Kern mit kontrollierter Qualität über „kontrollierte Qualität“ zu einem offenen und dynamischen Erschließungsraum übergehen, der nutzergenerierte Daten enthält und auch von diesen hinsichtlich der Qualität kontrolliert wird.

Zapilko und Sure (2009) beschreiben den Transfer des Schalenmodells in das Semantic Web und beschäftigen sich neben den technischen Bedingungen und Umsetzungsschritten auch mit der Frage, inwiefern dieser Ansatz eine Auswirkung auf die Text-Fakten-Integration haben könnte. Hintergrund ist hierbei, dass in vielen

wissenschaftlichen Bereichen die Sekundäranalyse von Primärdaten (Fakten) ein wichtiger Teilbereich ist. Daher zählt die zeitgleiche Recherche nach Text-Daten und Fakten-Daten ebenso zum Nutzeranspruch an eine „One-Stop“-Suche. Die verschiedenen Metadaten und oftmals auch die verschiedenen Vokabulartypen (Thesauri und Klassifikationen) erschweren die Zusammenführung beider Datenarten. Ontologien und der Schalenmodellansatz werden von Zapilko und Sure als viel versprechende Lösung angesehen.

Neben dem Mehrwert der Interoperabilität und den Überlegungen, die als „linked-data“ publizierten Daten durch die Modellierung des Suchraums sinnvoll zu nutzen, bietet die Repräsentation der kontrollierten Vokabulare in SKOS zudem neue Möglichkeiten, automatisierte Mappings zu erstellen.

Hiermit beschäftigt sich unter anderem die Ontology Alignment Evaluation Initiative⁶, deren jährliche Evaluationen Entwicklungen im Bereich der automatischen Ontologie-Abgleichung darstellen.

Literatur

- Altenhöner, Reinhard; Hannemann, Jan; Kett, Jürgen: Linked Data aus und für Bibliotheken: Rückgratstärkung im Semantic Web. In: Semantic Web & Linked Data – Elemente zukünftiger Informationsstrukturen. In: Ockenfeld, Marlies (Hrsg.): Semantic Web & Linked Data: Elemente zukünftiger Informationsinfrastrukturen; 1. DGI-Konferenz; 62. Jahrestagung der DGI, Frankfurt am Main, 7.–9. Oktober 2010 ; Proceedings, Frankfurt am Main: Dt. Gesellschaft für Informationswissenschaft u. Informationspraxis, 2010, S. 67–75.
- Aitchison, Jean; Dextre Clarke, Stella (2004): The Thesaurus: A Historical Viewpoint, with a Look to the Future. In: Cataloging and Classification Quarterly (37), S. 5–21.
- Becker, Hans Jürgen; Neuroth, Heike (2002): Crosssearchen und crossbrowsen von „Quality-controlled Subjekt Gateways im EU-Projekt Renardus“. In: Zeitschrift für Bibliothekswesen und Bibliographie 49 (3), S. 133–146.
- Dextre Clarke, Stella (2007): Evolution towards ISO 25964: an international standard with guidelines for thesauri and other types of controlled vocabulary. In: Information Wissenschaft & Praxis 58 (8), S. 441–445.
- Dextre Clarke, Stella (2010a): Types of mapping recommended in ISO 25964, and the question of reciprocity [Folien]. Online: <http://www.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2010/presentations/dextreclarke.pdf> [14. 10. 2011].
- Dextre Clarke, Stella (2011): In Pursuit of Interoperability: Can We Standardize Mapping Types? In: Boteram, F., Goedert, W.,

5 <http://richard.cyganiak.de/2007/10/lod/imagemap.html>.

6 Weitere Informationen zum Ontology Matching unter <http://ontologymatching.org> und zu Evaluierungen verschiedener Algorithmen unter <http://oaei.ontologymatching.org/>.

- Hubrich, J. (Hrsg.): Concepts in Context – Cologne Conference on Interoperability and Semantics in Knowledge Organization; 2010 Jul 19-2010 Jul 20; Köln, Germany. Würzburg: Ergon Verlag, 2011. Präsentation online: http://linux2.fbi.fh-koeln.de/cisko2010/praesentationen/2010-07-20_dextre Clarke.ppt [30. 12. 2011].
- Dextre Clarke, Stella (2011a): ISO 25964: A Standard in support of KOS interoperability. [Preprint]. Online: <http://www.iskouk.org/conf2011/papers/dextreclarke.pdf> [14. 10. 2011].
- Doerr, Martin (2001): Semantic Problems of Thesaurus Mapping. In: Journal of Digital Information 1 (8). Online: <http://journals.tdl.org/jodi/article/viewArticle/31/32>, [14. 10. 2011].
- Hellweg, Heiko; Krause, Jürgen; Mandl, Thomas; Marx, Jutta; Müller, Matthias N.O.; Mutschke, Peter; Strötgen, Robert (2001): Treatment of Semantic Heterogeneity in Information Retrieval. IZ-Arbeitsbericht Nr. 23. InformationsZentrum Sozialwissenschaften. Online: http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/iz_arbeitsberichte/ab_23.pdf?download=true [14. 10. 2011].
- Hubrich, Jessica (2008): CrissCross: SWD-DDC-Mapping. In: Mitteilungen der Vereinigung österreichischer Bibliothekarinnen und Bibliothekare 61 (3), S. 50–58. Online: <http://www.univie.ac.at/voeb/fileadmin/Dateien/Publikationen/VOB-Mitteilungen/vm6120083.pdf> [14. 10. 2011].
- Krause, Jürgen (2004): Konkretes zur These, die Standardisierung von der Heterogenität her zu denken. In: Zeitschrift für Bibliothekswesen und Bibliographie 51 (2), S. 76–89.
- Krause, Jürgen; Niggemann, Elisabeth; Schwänzl, Roland (2003): Normierung und Standardisierung in sich verändernden Kontexten: Beispiel Virtuelle Fachbibliotheken. In: Zeitschrift für Bibliothekswesen und Bibliographie 50, S. 19–28.
- Krause, Jürgen: Shell Model, Semantic Web and Web Information Retrieval. S. 95–106. In: Harms, Ilse; Luckhardt, Heinz-Dirk; Giessen, Hans W. (Hrsg.): Information und Sprache: Beiträge zu Informationswissenschaft, Computerlinguistik, Bibliothekswesen und verwandten Fächern; Festschrift für Harald H. Zimmermann. München: Saur. ISBN 3-598-11754-X.
- Mayr, Philipp; Mutschke, Peter; Petras, Vivien (2008): Reducing semantic complexity in distributed Digital Libraries: treatment of term vagueness and document re-ranking. Online: http://www.ib.hu-berlin.de/~mayr/arbeiten/LR_Mayr-Mutschke-Petras-final.pdf [14. 10. 2011].
- Mayr, Philipp; Petras, Vivien (2008): Crosskonkordanzen: Terminologie Mapping und deren Effektivität für das Information Retrieval. Online: http://archive.ifla.org/IV/ifla74/papers/129-Mayr_Petras-trans-de.pdf [14. 10. 2011].
- Mayr, Philipp (2010): Information Retrieval. Mehrwertdienste für Digitale Bibliotheken. Crosskonkordanzen und Bradfordizing. Bonn: GESIS (GESIS-Schriftenreihe, 5), ISBN 978-3-86819-012-0.
- Mayr, Philipp; Zapilko, Benjamin; Sure, York (2010): Ein Mehr-Thesauri-Szenario auf Basis von SKOS und Crosskonkordanzen. Online: <http://www.ib.hu-berlin.de/~mayr/arbeiten/Oberhof2010.pdf> [14. 10. 2011].
- Müller, Katrin; Jacobs, Jan-Helge; Mengel, Tina (2010): Insights and Outlooks: A Retrospective View on the CrissCross Project [Folien]. Online: http://linux2.fbi.fh-koeln.de/cisko2010/praesentationen/2010-07-19_jacobs_mueller_mengel.ppt [14. 10. 2011].
- National Information Standards Organization: Project ISO 25964. Online: <http://www.niso.org/workrooms/iso25964> [14. 10. 2011].
- ANSI/NISO Z39.19, 2005: Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. Online: http://www.niso.org/kst/reports/standards/kfile_download?id%3Austring%3Aiso-8859-1=Z39-19-2005.pdf&pt=RkGKiXzW643YeUaYUqZ1BFwDhIG4-24RJbcZBWg8uE4vWdpZsDs4RjLz0t90_d5_ymGsj_IKVaGZww13HuDIYn5U74Ydfa-3TffjxYQ25QrtR8PONuJLqvxo-l0Nir5 [14. 10. 2011].
- Normenausschuss Bibliotheks- und Dokumentationswesen (2002): DIN 1463. Erstellung und Weiterentwicklung von Thesauri – Teil 2: Mehrsprachige Thesauri. In: Bibliotheks- und Dokumentationswesen. Gestaltung und Erschließung von Dokumenten, Bibliotheksmanagement, Codierungs- und Nummerierungssysteme, Bestandserhaltung in Archiven und Bibliotheken: Normen. 1. Aufl., Stand der abgedr. Normen: November 2001. Berlin; Wien; Zürich: Beuth, S. 17–36.
- W3C (2009): W3.org – SKOS Primer. Online: <http://www.w3.org/TR/2009/NOTE-skos-primer-20090818/> [14. 10. 2011].
- W3C Mission Statement: w3.org – Consortium Mission Statement [14. 10. 2011].
- Zapilko, Benjamin; Sure, York (2009): Transferring the shell model to the semantic web and the impact on text-fact-integration. In: ICSD 2009 – International conference for digital libraries and the semantic web: proceedings; Trento, Italy, September 2009, Trent: Univ. of Trento, S. 1–7.
- Zeng, Marcia Lei; Chan, Lois Mai (2004): Trends and Issues in Establishing Interoperability Among Knowledge Organization Systems. In: Journal of the American Society for Information Science and Technology 55 (5), S. 377–395.
- Zeng, Marcia Lei (2009): ISO 25964: Thesauri and Interoperability with Other Vocabularies [Folien]. Online: <http://www.slideshare.net/mzeng/iso-25964-thesauri-and-interoperability-with-other-vocabularies> [14. 10. 2011].

Eingereicht am 4. November 2011, angenommen am 28. November 2011



Stefan Keil B.Eng.
Hochschule Darmstadt
Campus Dieburg
Studiengang Informationswissenschaft
am Fachbereich Media
Max-Planck-Straße 2
64807 Dieburg
stefan.keil87@googlemail.com
www.iu.h-da.de

Stefan Keil ist Absolvent des Bachelorstudiengangs Information Science & Engineering/Informationswissenschaft an der Hochschule Darmstadt, aktuell im ersten Semester des entsprechenden Masterstudiengangs; drei Monate Praktikum bei der GESIS im Sommer 2010; aktuell Tutor in der Lehre für Semantik; wissenschaftliche Hilfskraft im AIIRE-Projekt an der Hochschule Darmstadt; Interessensgebiete: kontrollierte Vokabulare, Inhaltserschließung, Semantic Web.