

## New Comer Corner

Pea Dunckel\*

# Zusammenführung mehrerer Thesauri zu einem gemeinsamen Begriffssystem

## Probleme und Lösungsansätze

<https://doi.org/10.1515/iwp-2017-0052>

**Zusammenfassung:** Im Zeitalter der digitalen Vernetzung und im Zuge zunehmend interdisziplinärer und internationaler Forschungsbestrebungen gewinnt die Interoperabilität von Informationssystemen stärker an Bedeutung. Ein wichtiger Schritt, um diese Interoperabilität zu erreichen und einen zentralen Zugang zu verteilten Systemen zu ermöglichen, liegt in der Zusammenführung der jeweiligen Erschließungsinstrumente der Informationssysteme. Eine Integration der verschiedenen kontrollierten Vokabulare erlaubt einen Datenaustausch zwischen den diversen Informationssystemen. Außerdem ermöglicht und vereinfacht sie eine systemübergreifende Recherche über mehrere Datenbanken, ohne dass eine Suchanfrage mehrmals formuliert oder umformuliert werden muss. Es existieren unterschiedliche Methoden und Instrumente sowie Strukturmodelle, mit denen eine Thesaurusintegration durchgeführt werden kann. Die am weitesten verbreitete Methode, das Mapping, ist sogar als ISO-Norm formuliert. Bei einer Thesauruszusammenführung können diverse Probleme auftreten. Jedoch gibt es Lösungsansätze, mit denen den Herausforderungen begegnet werden kann.

**Deskriptoren:** Thesaurus, Methode, Modell, Integriert, Informationsaustausch

### Joining of several thesauri into a collective system of concepts

Issues and solution approaches

**Abstract:** In a time of digital networks and in the context of increasing interdisciplinary and international research efforts, the interoperability of information systems is gaining importance. To achieve interoperability and to provide a central access to allocated systems, it is a vital step to join the vocabularies of each information system. The integrati-

on of the different controlled vocabularies allows data exchange between the multiple information systems. Furthermore, it makes possible and simplifies a research across systems without the need to repeatedly formulate or rephrase search queries. There are different methods and instruments as well as structural models with which a thesaurus integration may be implemented. The most commonly used method, the mapping, is even drafted as ISO standard. During a thesaurus integration, various issues may arise. However, solution approaches exist to meet the challenges.

**Descriptors:** Thesaurus, Method, Model, Integrated, Information Exchange

### La fusion de plusieurs thesauri en un système conceptuel commun

Problèmes et solutions

**Résumé:** À l'ère des réseaux numériques et dans le cadre des efforts de recherche de plus en plus interdisciplinaires et internationaux, l'interopérabilité des systèmes d'information gagne en importance. Une étape importante pour parvenir à cette interopérabilité et fournir un accès centralisé à des systèmes distribués, réside dans la fusion des outils d'exploitation des systèmes d'information. Une intégration des différents vocabulaires contrôlés permet l'échange de données entre les divers systèmes d'information. En plus, elle permet et simplifie la recherche dans plusieurs bases de données, tous systèmes confondus, sans que la recherche doive être formulée ou reformulée plusieurs fois. Il existe différentes méthodes, outils et modèles structurels qui permettent une intégration de thesauri. La méthode la plus largement utilisée, le mapping (c'est-à-dire la mise en correspondance des données), est même décrite sous forme d'une norme ISO. Lors de la fusion de plusieurs thesauri, divers problèmes peuvent surgir. Cependant, il existe des solutions qui permettent de relever les défis.

**Descripteurs:** Thésaurus, Méthode, le Modèle Intégré, le Partage de l'Information

---

\*Kontaktperson: Pea Dunckel, FH Potsdam, Fachbereich Informationswissenschaften, Kiepenheuerallee 5, 14469 Potsdam, E-Mail: pea.dunckel@fh-potsdam.de

## 1 Vorwort: Einleitende Bemerkungen und Begriffsklärung

Informationssysteme stehen meist dezentral und isoliert voneinander zur Verfügung, was eine system- und disziplinübergreifende Suche erschwert. Mit zunehmender Interdisziplinarität von Forschungsprojekten und vermehrter internationaler Zusammenarbeit gewinnt eine Interoperabilität der Systeme jedoch an Bedeutung. Ein möglicher Weg, um diese Interoperabilität zu erreichen, besteht darin, die Erschließungsinstrumente der jeweiligen Informationssysteme zusammenzuführen und somit einen zentralen Zugang zu ermöglichen. Durch die Zusammenführung mehrerer Thesauri kann ein Datenaustausch zwischen den Vokabularen, Systemen und Fachgebieten stattfinden. Zudem können Unterschiede zwischen den Vokabularen aufgehoben bzw. vorteilsbringend genutzt werden. Es wird eine bessere Vermarktung der Systeme erreicht, und die Suche kann erleichtert, optimiert sowie erweitert werden. Der Anwender formuliert seine Suche einmal mit dem ihm bekannten Thesaurus und die Suchbegriffe werden in die anderen Vokabulare übersetzt, um eine Suche über unterschiedliche Datenbanken, Fachgebiete und Terminologien zu ermöglichen. Demzufolge kann der Sucherfolg sowohl in Bezug auf den Recall als auch auf die Precision verbessert werden. Dieses „Cross-Searching“ in mehreren Quellen gewinnt vor allem im digitalen Zeitalter stärker an Bedeutung, wo vernetzte Ressourcen stetig zunehmen und die Anwender Interoperabilität erwarten.<sup>1</sup>

Verschiedene Publikationen, die sich primär auf konkrete Vokabularzusammenführungs-Projekte beziehen, setzen sich mit den damit verbundenen Fragestellungen auseinander. Bisher fehlt es jedoch an einer umfassenden theoretischen Abhandlung. Dieser Beitrag soll zur Schließung dieser Lücke beitragen und legt sein Augenmerk auf die Zusammenführung von Thesauri. Im Zentrum stehen Verfahren, die eine Zusammenführung ermöglichen, sowie eine Strukturierung von Problemfällen und die Ausarbeitung von Lösungsansätzen, mit denen den Problemen einer Thesauruszusammenführung begegnet werden kann.

Bevor das eigentliche Thema der Thesauruszusammenführung behandelt wird, sollen wesentliche Begriffe definiert werden, die im weiteren Verlauf dieses Beitrags auftreten. Sowohl für die Thesaurusbildung als auch für die Zusammenführung von Thesauri ist es notwendig, ein-

deutig zwischen einer Bezeichnung und dem mit ihr verbundenen Konzept (Begriff) zu unterscheiden. Hierzu wird auf die von Ferdinand de Saussure und Umberto Eco geprägten Termini Signifikant und Signifikat zurückgegriffen, wobei ersterer für die Zeichenkette steht und letzterer die mentale Vorstellung bzw. das Konzept bezeichnet.<sup>2</sup>

Signifikate, die jeweils durch einen Deskriptor repräsentiert werden, sind die Grundelemente eines Thesaurus. Ein Thesaurus bildet verschiedene Beziehungen ab und baut ein semantisches Begriffsnetz auf. Bei den Beziehungen handelt es sich zum einen um hierarchische und assoziative Relationen zwischen Signifikaten. Zum anderen definiert ein Thesaurus in Form von Äquivalenzrelationen Beziehungen zwischen Signifikaten und ihren jeweiligen Signifikanten.<sup>3</sup> Die abgebildeten Relationen sind hauptsächlich paradigmatischer Natur. Syntagmatische Relationen sind nur rudimentär als Assoziationsrelationen enthalten.

Mit Hilfe der begrifflichen und terminologischen Kontrolle gelingt es einem Thesaurus eine Retrieval-relevante Sprache zu erstellen.<sup>4</sup> So werden synonyme Signifikate zu einem Signifikat zusammengefasst, Signifikate eindeutig definiert und Ambiguitäten aufgelöst.

## 2 Wesentliche Methoden und Instrumente

Vokabulare können auf unterschiedliche Art zusammengeführt werden. Mapping bzw. das Erstellen von Crosskordanzanzen, Merging sowie der Einsatz des W3C-Standard SKOS (Simple Knowledge Organization System)<sup>5</sup> sind wesentliche Methoden bzw. Instrumente, die im Zusammenhang mit Integrationsversuchen verschiedener Vokabulare genannt werden.

Mapping wird als ein Prozess definiert, bei dem Terme, Begriffe und hierarchische Beziehungen zweier oder mehrerer Thesauri nach Äquivalenzen geprüft werden und ausgehend von einem Quellthesaurus mit Entitäten eines Zielthesaurus in Beziehung gebracht werden.<sup>6</sup> Bei einem Mapping bleiben die beteiligten Begriffssysteme in ihrer Originalform bestehen. Es werden lediglich Querverbindungen, sogenannte Mappingrelationen, zwischen ihren

<sup>1</sup> Vgl. Hahn (2004, 48ff.), McCulloch, Shiri & Nicholson (2004) und Zeng & Chan (2004, 377ff.).

<sup>2</sup> Vgl. Eco (1977, 30ff.).

<sup>3</sup> Vgl. ISO 25964-1 (2011, 44ff. und 57ff.).

<sup>4</sup> Vgl. Burkart (1997, 163).

<sup>5</sup> Vgl. Miles & Bechhofer (2009).

<sup>6</sup> Vgl. Doerr (2001), Liang & Sini (2006) und ISO 25964-2 (2013, 11f.).

Konzepten abgebildet und es wird von einem Vokabular auf das andere zugegriffen.

Diese Abbildungen der Verweise zwischen verschiedenen Thesauruskonzepten werden auch als Crosskonkordanzen bezeichnet. Crosskonkordanzen sind „intellektuell und manuell erstellte Verbindungen, die Äquivalenz, Hierarchie und Verwandtschaft zwischen Termen zweier kontrollierter Vokabulare über Relationen bestimmen“.<sup>7</sup> In der Regel sind diese Verbindungen bilateral, also in beide Richtungen verlaufend. Das heißt, dass Konzepte aus Thesaurus A mit Konzepten aus Thesaurus B verbunden sind und umgekehrt. Jedoch ist es nicht zwingend notwendig, dass diese bilateralen Beziehungen auch symmetrisch sind. So ist es möglich, dass ein Deskriptor aus Thesaurus A, wie „Computer“, auf das Konzept „Information System“ in Thesaurus B gemappt wird, während dasselbe Konzept „Information System“ in Thesaurus B beim umgekehrten Mapping hingegen auf ein anderes Konzept in Thesaurus A, wie beispielsweise auf „Data base“ verweist.<sup>8</sup>

Bei einem Mapping entstehen demnach als Ergebnis Crosskonkordanzen bzw. Mappingrelationen zwischen verschiedenen Vokabularen, während die ursprünglichen Systeme erhalten bleiben. Im Gegensatz dazu zielt das Merging auf eine tatsächliche Integration verschiedener Systeme.<sup>9</sup> Als Ergebnis von Merging entsteht ein neues, permanent kombiniertes Vokabular. Duplikate werden entfernt und die ursprünglichen Thesauri aufgelöst.<sup>10</sup>

Dadurch, dass bei einem Merging die originären Vokabulare eliminiert werden, können in ihnen enthaltene Informationen verloren gehen, da sie gegebenenfalls für das zusammengeführte System als irrelevant eingestuft und deshalb nicht übernommen werden. Eine Erweiterung des Suchraums ist somit nur bedingt gegeben. Zudem können die Ausgangsthesauri nicht mehr getrennt voneinander für eine Suche genutzt werden. Werden Vokabulare aufeinander gemappt, bleiben sie dagegen in ihrer eigentlichen Form bestehen. Das hat den Vorteil, dass die Vokabulare sowohl einzeln als auch in ihrer Kombination für die Recherche genutzt werden können. Dies gestaltet die Suche flexibel. Durch die Einbeziehung von Mappingrelationen bei der Suche wird eine tatsächliche Erweiterung des Suchraums erreicht und ein Cross-Searching ermöglicht. Aus diesen Gründen findet das Mappingverfahren in der Praxis die meiste Beachtung.<sup>11</sup>

<sup>7</sup> Mayr & Petras (2008, 5).

<sup>8</sup> Vgl. Mayr & Petras (2008, 5).

<sup>9</sup> Vgl. Liang & Sini (2006, 2).

<sup>10</sup> Vgl. Hedden (2012, 23).

<sup>11</sup> Vgl. McCulloch, Shiri & Nicholson (2004).

### 3 Strukturmodelle

Je nach Ziel und Ausgangssituation des Vorhabens kann die Zusammenführung von Thesauri nach unterschiedlichen Strukturmodellen erfolgen. In der Fachliteratur werden u. a. das Satellitenmodell, der Makrothesaurus, das Switching und das Direct-linked Modell definiert.<sup>12</sup> Die strukturellen Modelle sind als Grundgerüst für eine Vokabularzusammenführung zu verstehen, auf dem aufgebaut werden kann. Auch ist eine Mischung der Modelle denkbar.<sup>13</sup>

Beim Satellitenmodell werden aus einem bestehenden Thesaurus, der für verschiedene Fachbereiche die allgemeine Terminologie zur Verfügung stellt, die für jeden einzelnen Fachbereich als relevant erachteten Teile als neue Satelliten- oder Mikrothesauri extrahiert.<sup>14</sup> ISO 25964-2 bezeichnet das Satellitenmodell als Hub-Struktur und gibt eine abweichende Definition. Aus bestehenden Thesauri, die zusammengeführt werden sollen, wird ein Thesaurus als Hub gewählt. Zu diesem Hub-Thesaurus werden die weiteren Thesauri als Satelliten in eine oder beide Richtungen gemappt.<sup>15</sup> Das Modell des Makrothesaurus bildet die Umkehrung des Satellitenmodells. Er stellt ebenfalls eine Suprastruktur dar, die jedoch dadurch entsteht, dass ähnliche Konzepte aus verschiedenen Thesauri in allgemeinere Themenkategorien zusammengefasst werden.<sup>16</sup> Die Switching Language bzw. das Intermediate Lexicon<sup>17</sup> besteht aus Codes oder Notationen. Sie dienen als Vermittler zwischen den einzelnen Thesauri, die zusammengeführt werden sollen.<sup>18</sup> Für diesen Zweck müssen wechselseitige Äquivalenzen zwischen der Switching Language und den beteiligten Vokabularen hergestellt werden.<sup>19</sup> Beim Direct-linked Modell werden die Vokabulare bilateral direkt aufeinander gemappt. Jeder Thesaurus wird sowohl als Quelle als auch als Ziel des Mappings herangezogen.<sup>20</sup>

<sup>12</sup> Für eine detailliertere Beschreibung vgl. Lancaster & Smith (1983, 64ff.), Zeng & Chan (2004) und ISO 25964-2 (2013, 16ff.).

<sup>13</sup> Vgl. Zeng & Chan (2004, 387).

<sup>14</sup> Vgl. Lancaster & Smith (1983, 62).

<sup>15</sup> Vgl. ISO 25964-2 (2013, 17f.).

<sup>16</sup> Vgl. Lancaster & Smith (1983, 64ff.).

<sup>17</sup> Vgl. Lancaster & Smith (1983, 57f.).

<sup>18</sup> Vgl. Zeng & Chan (2004, 384f.).

<sup>19</sup> Vgl. Lancaster & Smith (1983, 57f.).

<sup>20</sup> Vgl. ISO 25964-2 (2013, 17).

## 4 Mappingverfahren und Mappingrelationen nach ISO 25964-2

Entsprechend der Nomenklatur der Thesaurusrelationen wird das Mappingverfahren nach der ISO-Norm 25964-2 in drei Hauptmappingtypen unterteilt: Äquivalenzmapping, hierarchisches Mapping und assoziatives Mapping.<sup>21</sup>

Die Äquivalenzrelation kennzeichnet ähnliche Konzepte zweier Thesauri. Sie wird weiterhin in die exact simple equivalence (Symbol: =EQ), die inexact simple equivalence (Symbol: ~EQ) und die partial simple equivalence sowie in die intersecting compound equivalence (Symbol: EQ+) und die cumulative compound equivalence (Symbol: EQ |) differenziert. Somit ist es bei einem Mapping möglich, verschiedene Äquivalenzgrade zu definieren und auch 1:n Relationen zu verwenden.

Hierarchische Mappingrelationen stellen Oberbegriffs- (Symbol: BM) bzw. Unterbegriffsbeziehungen (Symbol: NM) zwischen zwei Konzepten dar. Ein assoziatives Mapping wird eingesetzt, wenn keine der zuvor beschriebenen Mappingarten zutrifft, zwei Konzepte aber dennoch in ihrer Bedeutung so miteinander assoziiert werden können, dass es für die Suche relevant erscheint.

Für die Durchführung eines Mappings empfiehlt ISO 25964-2 diesen Ablauf: Zunächst ist das Strukturmodell zu wählen, nach dem das Mapping erfolgen soll. Zudem muss festgelegt werden, welche Mappingrelationen angewendet werden und ob das Mapping bilateral oder nur in eine Richtung durchgeführt wird. Wenn ein Mapping in beide Richtungen erfolgen soll und kann, ist es sinnvoll das Mapping zuerst systematisch in die eine Richtung zu erstellen und anschließend in die andere. Sind an einem Mapping mehr als zwei Vokabulare beteiligt, muss dieser Vorgang so lange wiederholt werden, bis jeder beteiligte Thesaurus als Quellthesaurus auf allen weiteren Thesauri abgebildet wurde. Es wird empfohlen, jeweils ein Thesauruspaar zu betrachten und Konzept für Konzept passende Mappingrelationen zwischen dem Quell- und dem Zielthesaurus zu identifizieren.

Werden Mappings in beide Richtungen vorgenommen, können die Ergebnisse aus dem bereits durchgeführten Mapping von Thesaurus A nach Thesaurus B hilfreiche Anhaltspunkte für das umgekehrte Mapping liefern. Exakt äquivalente Mappingrelationen, wie „antennae =EQ antennas“, können beispielsweise ohne weitere Prüfung für das umgekehrte Mapping eingetragen werden. Bei den

übrigen Mappingtypen können die identifizierten Relationen zwar für die Gegenrichtung vorgeschlagen werden, müssen aber auf ihre Korrektheit geprüft und gegebenenfalls korrigiert oder auch verworfen werden.

In den letzten Jahren sind zahlreiche Veröffentlichungen von Mappingprojekten erschienen.<sup>22</sup> Zum Teil nutzen die Projekte Mappingrelationen wie sie die ISO-Norm 5964 bzw. die Nachfolgenorm ISO 25964-2 definiert<sup>23</sup> oder sie weichen leicht von den Vorgaben der beiden Normen ab.<sup>24</sup> Doch in der Regel werden Äquivalenzmappings und hierarchische Mappingrelationen verwendet.<sup>25</sup> Auch das assoziative Mapping wird in der Praxis berücksichtigt.<sup>26</sup> Compound Mappings werden dagegen nicht immer zugelassen.<sup>27</sup> Vermehrt werden bei Mappingprojekten zudem die Mappingrelationen verwendet, die der SKOS-Standard bietet.<sup>28</sup>

## 5 Probleme und Lösungsansätze des Thesaurusmapping

Bei der Zusammenführung mehrerer Thesauri lassen sich Problematiken und die entsprechenden Lösungen auf unterschiedlichen Betrachtungsebenen feststellen. Diese sind:

- die Signifikatebene, also die Ebene des Konzepts,
- die Signifikantebene, also die Ebene der Zeichenkette,
- die signifikat- und signifikantbezogene Ebene,
- die strukturelle Ebene.

### 5.1 Ebene der signifikatbezogenen Herausforderungen

Viele der signifikatbezogenen Probleme bei der Zusammenführung verschiedener Begriffssysteme lassen sich auf die Diskursabhängigkeit von Thesauri zurückführen. Die Domäne, in der ein Thesaurus verwendet wird, hat einen großen Einfluss auf die Bedeutung von Konzepten.

<sup>22</sup> Vgl. z. B. Doerr (2001), CARMEN (AP 12) (1999-2002), CrissCross (2011), Gulbrandsen et al. (2015), Liang & Sini (2006), Mayr & Petras (2008) und Vizine-Goetz et al. (2004).

<sup>23</sup> Vgl. z. B. Gulbrandsen et al. (2015) und Mayr & Petras (2008).

<sup>24</sup> Vgl. z. B. Doerr (2001).

<sup>25</sup> Vgl. z. B. Gulbrandsen et al. (2015), Liang & Sini (2006) und Mayr & Petras (2008).

<sup>26</sup> Vgl. z. B. Mayr & Petras (2008).

<sup>27</sup> Vgl. Gulbrandsen et al. (2015, 8).

<sup>28</sup> Vgl. z. B. Gulbrandsen et al. (2015, 7), Liang & Sini (2006, 5) und Mayr & Petras (2008, 19).

<sup>21</sup> Hier und weiter vgl. ISO 25964-2 (2013).

Je nach Kontext und Zielgruppe können Signifikate unterschiedlich gebraucht und interpretiert werden.<sup>29</sup> Das Konzept „Statistik“ wird sich beispielsweise in einem Psychologie-Thesaurus auf die Wahrscheinlichkeitstheorie beziehen, während es in der Soziologie eher die Demoskopie meinen wird.<sup>30</sup> Ähnlich unterschiedlich wird bei solch divergierenden Bedeutungsnuancen auch die jeweilige hierarchische Einordnung ausfallen.<sup>31</sup> Ein kunstgeschichtlicher Thesaurus könnte das Konzept „Pferd“ beispielsweise unter dem Oberbegriff „Kunstmotiv“ einordnen, während ein naturwissenschaftlicher Thesaurus es den Säugetieren zuordnen würde. Ebenfalls können Konzepte innerhalb eines Fachbereichs als synonym betrachtet werden, die in einem anderen Gebiet streng unterschieden werden. Die Konzepte „Marketingstrategie“ und „Marketingmethode“ gelten beispielsweise in dem landwirtschaftlichen AGROVOC-Thesaurus als quasisynonym, wobei „Marketingmethode“ der Deskriptor ist und „Marketingstrategie“ ein Nichtdeskriptor. Dieser Umstand verleitet dazu, eine exakte Äquivalenzrelation zwischen dem Konzept „Marketingmethode“ im AGROVOC-Thesaurus und dem Konzept „Marketingstrategie“ in einem betriebswirtschaftlichen Zielthesaurus herzustellen. In einem betriebswirtschaftlichen Zusammenhang weisen die Konzepte jedoch erhebliche Bedeutungsunterschiede auf.<sup>32</sup>

Daher ist es unabdingbar, die Bedeutung der Konzepte in ihrem jeweiligen Kontext eingehend zu prüfen, um sie in einem Mapping korrekt zu verbinden. Eine solche Prüfung lässt sich anhand der Relationen der Konzepte innerhalb der jeweiligen Thesauri durchführen, da diese das Konzept definieren.<sup>33</sup> Sollten Scope Notes oder Definitionen vorhanden sein, sollten auch diese konsultiert werden. In manchen Fällen kann es auch sinnvoll sein, bestimmte Konzepte bzw. Bedeutungen durch ein compound Mapping mit dem Operator NOT explizit aus einer Mappingrelation auszuschließen, wie „Marketingstrategie =EQ (Marketing + Strategie) NOT Marketingmethode“.

## 5.2 Ebene der signifikantbezogenen Herausforderungen

Auf der Signifikantenebene betrachtet, stellt die größte Herausforderung bei der Zusammenführung mehrerer Thesauri wohl die unterschiedliche Verwendung von Präkom-

bination, Präkoordination und Postkoordination innerhalb der Begriffssysteme dar. Jedes dieser Verfahren zur Zusammenführung komplexer Sachverhalte weist Vor- und Nachteile für die Erschließungs- und Retrievalfunktion eines Thesaurus auf, die sich auch auf ein Mappingvorhaben auswirken können. Je stärker sich die zusammenzuführenden Vokabulare in der Verwendung von Präkombination und Präkoordination bzw. Postkoordination unterscheiden, desto umständlicher wird eine Integration.<sup>34</sup>

Ein Quellthesaurus verwendet beispielsweise den präkombinierten Deskriptor „Arbeitnehmermotivation“, der im Zielthesaurus nicht vorhanden ist. Er kann jedoch durch die Kombination mehrerer Konzepte ausgedrückt werden. Die Mappingrelation kann durch eine assoziative 1:n Verbindung dargestellt werden: „Arbeitnehmermotivation RM Arbeitnehmer + Motivation“.<sup>35</sup> Bei präkombinierten Deskriptoren muss zudem berücksichtigt werden, dass je nach Thesaurus invertierte Wortstellungen, wie „Anerkennung, soziale“, oder nicht invertierte Wortstellungen, wie „soziale Anerkennung“ bevorzugt werden können.<sup>36</sup>

Da ein präkombiniertes Vokabular die Indexierung zwar erleichtern kann, aber gleichzeitig die Freiheiten bei der Erschließung einschränkt und das Begriffssystem sehr groß werden lässt, wählen viele Thesauri einen Kompromiss zwischen den Verfahren und greifen auch auf Prä- oder Postkoordination zurück.

Thesauri, die für prä- und postkoordinierte Erschließungsverfahren ausgearbeitet wurden, verfügen über ein kleineres Vokabular und bieten mehr Freiheiten bei der Erschließung. Doch auch hier lassen sich Schwierigkeiten bei einem Mappingvorhaben identifizieren. Insbesondere die semantische Zerlegung, die einen Signifikanten in seine Bedeutungseinheiten trennt, die aber nicht zwingend in dem Ausgangssignifikanten enthalten sein müssen,<sup>37</sup> kann in diesem Zusammenhang zu Problemen führen. Das Konzept „Thermometer“ lässt sich beispielsweise in die semantischen Bestandteile „Temperature“, „Measurement“ und „Instrument“ zerlegen. Es kann demnach vorkommen, dass ein Thesaurus den Deskriptor „Thermometer“ verwendet, während ein anderer stattdessen die genannten drei Konzepte kombiniert, um das Signifikat „Thermometer“ darzustellen.<sup>38</sup> Es lässt sich demnach nicht ohne weiteres von einem Deskriptor auf seine semantischen Bestandteile schließen. Bei einem Mapping können USE-Verweise, wie „Thermometer USE Temperature + Measure-

<sup>29</sup> Vgl. Lancaster & Smith (1983, 74) & Vizine-Goetz et al. (2004).

<sup>30</sup> Vgl. Hahn (2004, 52).

<sup>31</sup> Vgl. Doerr (2001).

<sup>32</sup> Vgl. Lauser et al. (2008, 50 f.).

<sup>33</sup> Vgl. Liang & Sini (2006, 5) und Lauser et al. (2008, 47).

<sup>34</sup> Vgl. Lancaster & Smith (1983, 54).

<sup>35</sup> Vgl. Hahn (2004, 55).

<sup>36</sup> Vgl. Zeng & Chan (2004, 387).

<sup>37</sup> Vgl. Bertram (2005, 223).

<sup>38</sup> Vgl. Lancaster & Smith (1983, 55).

ment + Instrument“,<sup>39</sup> Aufschluss darüber geben, welche Deskriptorenkombinationen für die Darstellung von nicht vorhandenen Deskriptoren genutzt werden sollen. Die entsprechende Mappingrelation wird analog dazu wie folgt gebildet: „Thermometer =EQ Temperature + Measurement + Instrument“. Ohne den USE-Verweis wäre es bei einem Mapping allerdings deutlich aufwendiger, die passenden Konzepte zu finden.

Ebenso schwierig kann es sein, aus einer Postkoordination das Konzept zu rekonstruieren, das damit ausgedrückt werden soll, wie bei der Postkoordination „factory + grinding“, die im Art & Architectural Thesaurus (AAT) für das Signifikat „mill“ verwendet wird.<sup>40</sup> Weitere Problemfälle können mehrdeutige Kombinationen, wie die Postkoordination „Baum + Stamm“, darstellen, die sowohl für das Konzept „Stammbaum“ als auch für das Konzept „Baumstamm“ stehen kann.<sup>41</sup> Entsprechend schwierig wird es demnach, in einem Mapping passende Konzepte für eine Prä- oder Postkoordination zu finden. Auch in diesem Fall ist es für ein Mappingvorhaben hilfreich, wenn der betreffende Thesaurus von einem Nichtdeskriptor auf die zu verwendende Postkoordination verweist. Ebenfalls sollten prä- bzw. postkoordinierte Thesauri Regeln bereitstellen, die bestimmen, auf welche Weise Kombinationen vorzunehmen bzw. Konzepte zu zerlegen sind. Sind solche Regeln vorhanden, sollten sie bei einem Mappingvorhaben konsultiert werden. Sind keine zu verwendenden Kombinationen durch die Thesaurushersteller vorgegeben, muss bei einem Mapping geprüft werden, ob dennoch Konzepte existieren, die miteinander kombiniert werden können, um die Bedeutung des zu mappenden Konzepts wiederzugeben. Wurden die Bedeutungen von prä- bzw. postkoordinierten Signifikaten identifiziert, können sie in der Regel durch exakte, inexakte, hierarchische oder assoziative one-to-many Mappings aufgelöst werden,<sup>42</sup> so bei „hacker RM computers + crime“<sup>43</sup> oder „Autowerkstatt =EQ Automobil + Werkstatt“.<sup>44</sup>

Eine weitere Problematik auf der Signifikantenebene stellt die Wahl der Vorzugsbenennung dar. So kann es sein, dass dasselbe Signifikat innerhalb mehrerer Thesauri von unterschiedlichen Signifikanten repräsentiert wird. Deskriptoren können sich in ihrer Schreibweise unterscheiden oder es können je nach Thesaurus Abkürzungen oder

individuelle bzw. eigen konstruierte Varianten bevorzugt werden.<sup>45</sup>

Welcher Signifikant als Deskriptor eingesetzt wird, kann u. a. von der bevorzugten Sprachform, dem Kontext oder der Zielgruppe abhängen. Ein biologischer Thesaurus kann die wissenschaftliche Bezeichnung „Fringilla Coelebs“ als Deskriptor wählen, während ein Thesaurus einer anderen Domäne den synonymen Signifikanten „Buchfink“ als Vorzugsbenennung einsetzt.<sup>46</sup> Werden bei einem Mappingvorhaben jedoch nur die Deskriptoren berücksichtigt, kann es passieren, dass Äquivalenzrelationen aufgrund unterschiedlicher Vorzugsbenennungen übersehen werden. Daher ist es notwendig, auch die Nichtdeskriptoren bei einem Mapping mit einzubeziehen, um weitere mögliche Äquivalenzen zwischen den Konzepten der Thesauri zu finden.

Sind an einem Mapping mehr als zwei Thesauri beteiligt, können über die jeweiligen Nichtdeskriptoren zudem Äquivalenzrelationen zwischen den Vokabularen identifiziert werden, die andernfalls eventuell nicht erkannt worden wären. Kommt in Thesaurus A beispielsweise der Deskriptor „Fleischer“ vor, der in Thesaurus B Nichtdeskriptor der Vorzugsbenennung „Metzger“ ist, und in Thesaurus C kommt ebenfalls der Deskriptor „Metzger“ vor, „Fleischer“ jedoch nicht, ist es dennoch plausibel zu schlussfolgern, dass der Deskriptor „Fleischer“ aus Vokabular A mit dem Konzept „Metzger“ aus Vokabular C äquivalent ist. Diese Äquivalenz lässt sich nur über Thesaurus B herleiten. Relationen, die auf diesem Weg identifiziert werden, werden Cross-Referenzen genannt.<sup>47</sup> Wie das weiter oben genannte Beispiel der Konzepte „Marketingstrategie“ und „Marketingmethode“ jedoch zeigt, können derartige Cross-Referenzen auch zur Herleitung falscher Synonymie führen.

### 5.3 Ebene der signifikat- und signifikantbezogenen Herausforderungen

Des Weiteren können sich Thesauri in ihren Disambiguierungsverfahren unterscheiden. So ist es möglich in einem Thesaurus ambigie Begriffe durch einen Qualifikator eindeutig zu machen, während in einem anderen Thesaurus der Kontext die Begriffsbedeutung klärt. Bei einem Mapping kann dieser Umstand zu Problemen führen, da mehr-

<sup>39</sup> Vgl. Neville (1970, 322).

<sup>40</sup> Vgl. Doerr (2001).

<sup>41</sup> Vgl. Burkart (1997, 167).

<sup>42</sup> Vgl. ISO 25964-2 (2013, 32).

<sup>43</sup> Vgl. Mayr & Petras (2008).

<sup>44</sup> Vgl. Doerr (2001).

<sup>45</sup> Vgl. Doerr (2001) und Neville (1970, 328ff.).

<sup>46</sup> Vgl. Doerr (2001).

<sup>47</sup> Vgl. Lancaster & Smith (1983, 57).

deutige Begriffe gegebenenfalls verwechselt werden. Es kann sein, dass ein Architektur-Thesaurus mit „Schiff“ das Kirchenschiff meint, während es in einem nautischen Thesaurus für das Wasserfahrzeug steht.<sup>48</sup> Werden diese beiden Konzepte über ein Äquivalenzmapping miteinander verbunden und ein Nutzer verwendet eines davon anschließend für eine Suchanfrage, wird die Ergebnismenge zahlreiche Treffer enthalten, die für den Nutzer irrelevant sind.

Bei ambigen Konzepten ist es demnach nicht möglich, das Mapping allein auf den Signifikanten zu stützen. Korrekte Mappingbeziehungen können auch hier erst nach einer genauen Prüfung der Bedeutung der Konzepte erstellt werden, wobei vor allem die Oberbegriffe der mehrdeutigen Konzepte Aufschluss geben. Hat beispielsweise das nicht disambiguierte Konzept „Bremse“ in Thesaurus A den Oberbegriff „Fahrzeugteil“ und in Thesaurus B den Oberbegriff „Insekt“, wird offensichtlich, dass zwischen den Konzepten keine Mappingrelation erstellt werden kann. Stimmen die übergeordneten Konzepte jedoch überein, kann dagegen von einem exakten Mapping ausgegangen werden. Sollte ein Thesaurus eine bestimmte Bedeutung eines mehrdeutigen Signifikanten aufführen, ein anderer aber nicht, ist je nach Strukturmodell und Anwendungsbereich zu entscheiden, ob die fehlende Bedeutung bei der Zusammenführung berücksichtigt oder ausgeschlossen werden soll.

## 5.4 Ebene der strukturbezogenen Herausforderungen

Verschiedene Thesauri können eine differierende Granularität und Struktur aufweisen. Bei einer Thesaurusintegration stellt insbesondere die heterogene Spezifität eine Herausforderung dar.<sup>49</sup> Die Unterschiede in der Spezifität eines Vokabulars werden unter anderem durch divergierende Anwendungsgebiete und Themenfelder der Vokabulare verursacht. Ebenfalls kann ein Thesaurus detaillierter ausfallen als ein allgemein gefasster Thesaurus, wenn er für die Erschließung eines spezifischeren Gebietes genutzt werden soll. In einem eher generellen Thesaurus genügt es beispielsweise zwischen den Konzepten „Pflanzenfresser“ und „Fleischfresser“ zu unterscheiden. Ein speziellerer Thesaurus würde die Pflanzenfresser noch weiter in „Blattfresser“, „Holzfresser“ usw. differenzieren.

Dies führt besonders dann zu Schwierigkeiten, wenn ein weniger spezifisches Vokabular auf ein spezifischeres abgebildet werden muss. Denn für sehr spezielle Konzepte wird es in einem allgemein gehaltenen Vokabular voraussichtlich wenig exakte Entsprechungen geben.<sup>50</sup> Das gilt vor allem für das Hub-Modell und das direkte Mapping, die jeweils bilateral durchzuführen sind. Bei dem Modell des Makrothesaurus, der die Konzepte als eher allgemeine Themenfelder zusammenfasst, spielt die unterschiedliche Spezifität eine weniger wichtige Rolle, muss aber dennoch beachtet werden.

Grundsätzlich kann man Differenzen in der Spezifität bei einer Thesaurusintegration mit einem hierarchischen oder einem cumulative compound equivalence Mapping begegnen. Existiert beispielsweise in Thesaurus A das speziellere Konzept „Falke“, in Thesaurus B jedoch nur das allgemeinere Konzept „Greifvogel“, ist ein Äquivalenzmapping folglich nicht möglich. Das Konzept „Falke“ kann jedoch über ein broader Mapping mit dem Oberbegriff „Greifvogel“ verbunden werden.<sup>51</sup> Das umgekehrte Mapping von Thesaurus B zu Thesaurus A kann dementsprechend durch ein narrower Mapping aufgelöst werden: „Greifvogel NM Falke“. Existieren in Thesaurus A noch weitere spezielle Konzepte, wie „Fischadler“ oder „Habicht“, wäre eine alternative Möglichkeit ein one-to-many Mapping mit dem Booleschen Operator OR: „Greifvogel EQ Falke | Fischadler | Habicht“. Bei einer Suche mit Konzepten, die durch derartige Mappingrelationen miteinander verbunden sind, wird allerdings die Precision der Ergebnismenge reduziert, da stets auch Dokumente gefunden werden, die allgemeiner bzw. spezifischer sind als die Suchanfrage. Sucht ein Nutzer von Thesaurus B beispielsweise nach Dokumenten, die Greifvögel im Allgemeinen thematisieren, wird er durch das cumulative compound mapping auch Dokumente finden, die sich speziell mit Falken, Habichten oder Seeadlern befassen. Wird bei einer Zusammenführung die Struktur eines Makrothesaurus gewählt, könnten alle zuvor genannten Konzepte beispielsweise über die allgemeineren Oberbegriffe „Vogel“ oder sogar „Tier“ zusammengefasst werden. Dies würde ebenfalls die Precision verringern.

Finden sich für ein Konzept aus dem Quellthesaurus weder exakt noch hierarchisch passende Konzepte im Zielthesaurus, kann gegebenenfalls eine inexakte Äquivalenzrelation wie „child protection ~ EQ safeguarding children“<sup>52</sup> hergestellt werden.

<sup>48</sup> Vgl. Neville (1970, 323 f.).

<sup>49</sup> Vgl. z. B. Lancaster & Smith (1983, 53), Neville (1970), McCulloch, Shiri & Nicholson (2004) und Zeng & Chan (2004, 387).

<sup>50</sup> Vgl. Lancaster & Smith (1983, 54).

<sup>51</sup> Vgl. Lauser et al. (2008, 51).

<sup>52</sup> Vgl. ISO 25964-2 (2013, 28).

Ein weiterer Faktor auf struktureller Ebene, der die Machbarkeit einer Thesaurusintegration beeinflussen kann, ist die Verwendung unterschiedlicher Hierarchieformen innerhalb der Vokabulare.<sup>53</sup> Bei einer Zusammenführung von monohierarchischen und polyhierarchischen Thesauri muss darauf geachtet werden, dass keine falschen Verweise erzeugt werden. So führt ein polyhierarchischer Thesaurus A das Konzept „dekorative Waffe“ sowohl unter dem Oberbegriff „Waffe“ als auch unter dem Oberbegriff „Dekoration“. Der monohierarchische Thesaurus B ordnet das Konzept nur unter „Dekoration“ ein, nicht in der ebenfalls vorhandenen Hierarchie „Waffe“. Ein korrekt durchgeführtes bilaterales Mapping zwischen A und B sollte dementsprechend die folgenden Relationen erzeugen:

**Tabelle 1:** Beispiel-Mapping zwischen poly- und monohierarchischen Thesauri.

Richtung	Relation
A nach B	Dekorative Waffe <Dekoration> = EQ Dekorative Waffe <Dekoration>
	Dekorative Waffe <Waffe> BM Waffe
B nach A	Dekorative Waffe <Dekoration> = EQ Dekorative Waffe <Dekoration>

In beide Richtungen kann ein exaktes Mapping zwischen den Konzepten hergestellt werden, die dem Konzept „Dekoration“ als Unterbegriffe zugeordnet sind. Das Konzept, das in Thesaurus A zudem den Oberbegriff „Waffe“ aufweist, wird über ein broader Mapping mit dem Konzept „Waffe“ in Thesaurus B verbunden. Dabei sollte keine Relation zwischen den Konzepten „Dekorative Waffe <Waffe>“ aus Thesaurus A und „Dekorative Waffe <Dekoration>“ aus Thesaurus B hergestellt werden, da die Konzepte nicht exakt übereinstimmen und die gezeigten Mappingrelationen geeignetere Verbindungen darstellen.

Es wird deutlich, dass auch hier zunächst die thesauruspezifische Bedeutung der betreffenden Signifikate zu prüfen ist, um sie korrekt aufeinander abzubilden. Gleiches gilt, wenn ein Thesaurus auch partitive Relationen verwendet, ein anderer dagegen nur generische Relationen. Es muss entschieden werden, welche Hierarchiearten in dem Mapping oder dem verbundenen System erlaubt sind. So stellt ISO 25964-2 auch Mappingrelationen für partitive Hierarchien zur Verfügung.

## 6 Fazit

Um eine Thesauruszusammenführung umzusetzen, müssen verschiedene Aspekte berücksichtigt und Entscheidungen getroffen werden. An erster Stelle steht die Wahl der geeigneten Methoden und Instrumente. Möglichkeiten der Zusammenführung sind sowohl das Mappingverfahren oder die Erstellung von Crosskonkordanzen als auch das Mergingverfahren. In der Praxis findet allerdings in erster Linie das Mappingverfahren Anwendung. Denn mit einem Mapping können die Ziele einer Thesauruszusammenführung, wie der Datenaustausch, die Suchraumerweiterung und das Cross-Searching besonders effektiv erreicht werden.

Eine Thesauruszusammenführung kann je nach Ziel und Zweck des Vorhabens im Rahmen unterschiedlicher Strukturmodelle durchgeführt werden. Die Modelle sind nicht ausschließlich auf die hier beschriebene Form festgelegt, sondern sind auch abwandelbar oder miteinander kombinierbar. Es zeigt sich, dass jedem dieser Modelle eine Form von Mappingverfahren zugrunde liegt. Demzufolge beruhen alle Strukturmodelle auf einer Identifikation ähnlicher Vokabularkonzepte, die je nach Modell präziser oder allgemeiner ausfällt.

Für ein Mappingvorhaben können äquivalente, hierarchische und assoziative Mappingrelationen festgelegt werden, die den Relationen innerhalb eines Thesaurus entsprechen. Wie unterschiedliche Mappingprojekte zeigen, finden alle genannten Mappingtypen Anwendung in der Praxis.

Die Herausforderungen, die bei einem Thesaurusmapping auftreten können, beziehen sich auf verschiedene Ebenen, die Signifikatebene, die Signifikantebene und die strukturelle Ebene. Herausforderungen, die sowohl die Signifikat- als auch die Signifikantebene betreffen, sind ebenfalls möglich.

Es wird deutlich, dass eine Thesauruszusammenführung mit zahlreichen Problemen verbunden ist, die unter Berücksichtigung verschiedener Aspekte jedoch gelöst werden können.

Der erste Schritt ist das Erkennen und Verstehen der Schwierigkeiten, die bei einem Thesaurusmapping auftreten können. Dies erfordert zum einen Kenntnisse der semantischen Problematiken der natürlichen Sprache. Zum anderen ist es nötig, die Methoden und Prozesse der Thesauruserstellung sowie die Struktur und die Funktionen der internen Thesaurusrelationen zu verinnerlichen. Denn bei einer Thesauruszusammenführung sind ähnliche Herausforderungen wie bei einer Thesauruserstellung zu erwarten.

Des Weiteren gilt es für die Lösung der Probleme und damit für die Durchführung eines erfolgreichen Mapping

<sup>53</sup> Vgl. Doerr (2001).



verschiedene Voraussetzungen zu erfüllen. Um die Mappingrelationen korrekt einzusetzen, müssen Funktion und Nutzen eindeutig definiert und erfasst werden. Zudem braucht es für die Wahl der passenden Mappingrelationen und somit auch für die Qualität eines Mapping ein vollständiges und exaktes Verständnis der Semantik der zu mappenden Konzepte.<sup>54</sup> Für ein solches Verständnis ist eine intensive und sichere Verwendung der internen Thesaurusrelationen unerlässlich, da diese ein Konzept definieren. Ebenso hilfreich für ein Mappingvorhaben ist die Erfahrung bei der Arbeit mit den einzelnen Thesauri und die Kenntnis ihrer spezifischen Besonderheiten, wie Kombinationsregeln oder Hierarchietypen.

All jene Umstände machen ein Mapping zu einem sehr aufwendigen und anspruchsvollen Vorhaben, das einen hohen intellektuellen Einsatz erfordert.<sup>55</sup> Es existieren zwar automatisierte Verfahren, die einen Teil der Arbeit übernehmen können, doch können sie weder alle aufgeführten Problemfälle erkennen noch eigenständig lösen.<sup>56</sup> Eine Kontrolle durch den Menschen ist zwingend notwendig. Aus diesem Grund sollte der Mensch die zentrale ausführende Instanz bei einem Mapping bleiben. Denn ein qualitativ hochwertiges Mapping kann nur von Experten durchgeführt werden, die sowohl über Kenntnisse in der Erstellung und im Umgang mit Thesauri als auch ein Verständnis für die Probleme und Lösungen eines Mappingvorhabens verfügen.

Werden alle Probleme erkannt und gelöst und ein Mapping erfolgreich durchgeführt, bringt es einen großen Mehrwert für die Nutzer eines zusammengeführten Systems. Bei der Evaluation ihres Mappingprojekts konnten Philipp Mayr und Vivien Petras diesen Mehrwert belegen. Dafür wurden die Suchergebnisse, die bei der Suche mit nur einem der beteiligten Vokabulare erhalten wurden, mit den Ergebnissen verglichen, die bei einer Recherche unter Einbeziehung der Mappingrelationen gefunden wurden. Der Vergleich ergab eine deutliche Verbesserung sowohl der Precision als auch des Recalls der Treffermenge, die durch eine Suche mit dem gemappten System erreicht wurde.<sup>57</sup>

Um diesen Nutzen zu erhalten, ist eine Pflege der Mappingrelationen unabdingbar. Die beteiligten Thesauri müssen regelmäßig auf Änderungen geprüft werden und diese gegebenenfalls in das Mapping übernommen werden. Nur so lässt sich die Qualität eines Mapping auf Dauer bewahren.

<sup>54</sup> Vgl. Lauser et al. (2008, 47).

<sup>55</sup> Vgl. Doerr (2001) und ISO 25964-2 (2013, 38).

<sup>56</sup> Vgl. hierzu z. B. Lauser et al. (2008), Hellweg et al. (2001) oder ISO 25964-2 (2013, 40).

<sup>57</sup> Vgl. Mayr & Petras (2008, 17ff.).

## Literatur

- Bertram, Jutta: Einführung in die inhaltliche Erschließung. Grundlagen – Methoden – Instrumente. Würzburg: Ergon, 2005.
- Burkart, Margarete: Thesaurus. In: Buder, Marianne; Rehfeld, Werner; Seeger, Thomas & Strauch, Dietmar [Hrsg]: Grundlagen der praktischen Information und Dokumentation. Ein Handbuch zur Einführung in die fachliche Informationsarbeit, 4. völlig neu gefasste Ausgabe. München: Saur, 1997, S. 160–179.
- CARMEN. Content Analysis, Retrieval and MetaData: Effective Networking. Abschlussbericht des Arbeitspakets 12 (AP 12). Crosskonordanzen von Klassifikationen und Thesauri. Regensburg. (1999-2002). <https://epub.uni-regensburg.de/10091/31.5.2017>.
- CrissCross: Verknüpfung der Schlagwortnormdatei mit der Dewey-Dezimalklassifikation. (2011). <https://ixtrieve.fh-koeln.de/crisscross/swd-ddc-verknuepfung.html> [31.5.2017].
- Doerr, Martin: Semantic Problems of Thesaurus Mapping. In: *Journal of Digital Information*, Vol 1, No 8 (2001). <https://journals.tdl.org/jodi/index.php/jodi/article/view/31/32> [31.5.2017].
- Eco, Umberto: Zeichen. Einführung in einen Begriff und seine Geschichte. Frankfurt am Main: Suhrkamp, 1977.
- Gulbrandsen, Are Dag; Heggø, Dan Michael O.; Knutsen, Unni & Seland, Grete: Towards a general Norwegian Thesaurus? Subproject „Methodology for mapping Humord to WebDewey“. (2015). [http://www.uio.no/for-ansatte/enhetssider/ub/prosjekter/mapping-mot-webdewey/delte-dokumenter/methodology\\_mapping\\_final\\_report.pdf](http://www.uio.no/for-ansatte/enhetssider/ub/prosjekter/mapping-mot-webdewey/delte-dokumenter/methodology_mapping_final_report.pdf) [31.5.2017].
- Hahn, Klaus: Probleme der Integration digitaler Bibliothekssysteme: Semantische Heterogenität bei datenbankübergreifenden Recherchen. In: Informationszentrum Sozialwissenschaften [Hrsg]: Sharing Knowledge: Scientific Communication. 9. Kongress der IuK-Initiative der Wissenschaftlichen Fachgesellschaft in Deutschland. Bonn: Informationszentrum Sozialwissenschaften, 2004, S. 47–57.
- Hedden, Heather: Mapping, Merging, and Multilingual Taxonomies. SLA (Special Libraries Association) 2012 Conference Presentation. (2012). [http://www.hedden-information.com/Mapping\\_Merging\\_&\\_Multilingual\\_Taxonomies.pdf](http://www.hedden-information.com/Mapping_Merging_&_Multilingual_Taxonomies.pdf) [31.5.2017].
- Hellweg, Heiko; Krause, Jürgen; Mandl, Thomas; Marx, Jutta; Müller, Matthias N.O.; Mutschke, Peter & Strötgen, Robert: Treatment of Semantic Heterogeneity in Information Retrieval. Bonn: Informationszentrum Sozialwissenschaften, 2001. [http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis\\_reihen/iz\\_arbeitsberichte/ab\\_23.pdf](http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/iz_arbeitsberichte/ab_23.pdf) [31.5.2017].
- ISO 25964-1: Information and documentation – Thesauri and interoperability with other vocabularies – Part 1: Thesauri for information retrieval. (2011).
- ISO 25964-2: Information and documentation – Thesauri and interoperability with other vocabularies – Part 2: Interoperability with other vocabularies. (2013).
- Lancaster, F. Wilfrid & Smith, Linda C.: Compatibility Issues Affecting Information Systems and Services. Paris: Unesco, 1983. <http://unesdoc.unesco.org/images/0005/000584/058439EB.pdf> [31.5.2017].
- Lauser, Boris; Johannsen, Gudrun; Caracciolo, Caterina; Hage, Willem Robert van; Keizer, Johannes & Mayr, Philipp: Comparing human and automatic thesaurus mapping approaches in the agricul-

- tural domain. In: Greenberg, Jane [Hrsg.]; Klas, Wolfgang [Hrsg.]: *Metadata for semantic and social applications: proceedings of the International Conference on Dublin Core and Metadata Applications, 22-26 September 2008*. Göttingen: Univ.-Verl. Göttingen, 2008, S. 43–53. <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-46958-1> [31.5.2017].
- Liang, Anita C. & Sini, Margherita: Mapping AGROVOC and the Chinese Agricultural Thesaurus: Definitions, tools, procedures. In: *New Review in Hypermedia and Multimedia*, 12 (1) (2006), S. 51–62. <http://www.fao.org/3/a-ag862e.pdf> [31.5.2017]. (Die Paginierung der online-Version ist abweichend von den hier angegebenen Seitenzahlen.)
- Mayr, Philipp & Petras, Vivien: Crosskonkordanzen: Terminologie Mapping und deren Effektivität für das Information Retrieval. In: *The International Federation of Library Associations and Institutions (IFLA) [Hrsg.]: 74th IFLA General Conference and Council*. (2008). <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-315066> [31.05.2017].
- McCulloch, Emma; Shiri, Ali & Nicholson, Dennis: Challenges and issues in terminology mapping: a digital library perspective. In: *The Electronic Library*, 23 (6) (2004), S. 671–677. <http://eprints.rclis.org/5829/> [31.5.2017].
- Miles, Alistair & Bechhofer, Sean: SKOS Simple Knowledge Organization System. Reference. W3C Recommendation 18 August 2009. (2009). <https://www.w3.org/TR/2009/REC-skos-reference-20090818/> [31.5.2017].
- Neville, H. H.: Feasibility Study of a Scheme for reconciling Thesauri covering a common Subject. In: *Journal of Documentation*, Vol 26, Iss 4 (1970), S. 313–336.
- Vizine-Goetz, Diane; Hickey, Carol; Houghton, Andrew & Thompsen, Roger: Vocabulary Mapping for Terminology Services. In: *Journal of Digital Information*, Vol 4, No 4 (2004). <https://journals.tdl.org/jodi/index.php/jodi/article/view/114/113> [31.5.2017].
- Zeng, Marcia Lei & Chan, Lois Mai: Trends and Issues in Establishing Interoperability Among Knowledge Organization Systems. In: *Journal of American Society for Information Science and Technology*, 55 (5) (2004), S. 377–395.

**Pea Dunckel**

FH Potsdam

Fachbereich Informationswissenschaften

Kiepenheuerallee 5

14469 Potsdam

Bild: © Melanie Hauke

[pea.dunckel@fh-potsdam.de](mailto:pea.dunckel@fh-potsdam.de)

Pea Dunckel hat im März 2017 ihr Bachelorstudium im Fach Information und Dokumentation an der Fachhochschule Potsdam (FHP) abgeschlossen und studiert nun im Masterstudiengang Informationswissenschaften der gleichen Hochschule. Ihr Interesse gilt den Methoden, Modellen und Herausforderungen der Thesauruszusammenführung.