

# Vom Stringmatching zur Begriffsexploration: das Potential integrierter begrifflicher Interoperabilität

Jessica Hubrich

Fachhochschule Köln, Institut für Informationsmanagement, Köln

## *Zusammenfassung:*

Als Reaktion auf die mit fortschreitender Internationalisierung sich wandelnden Anforderungen an Informationsräume und Wissensorganisationssysteme haben sich verschiedene Modelle semantischer Interoperabilität herausgebildet, die darauf abzielen, mittels einer Verbindung von Begriffen verbreiteter Dokumentationssprachen eine Grundlage für effizientere Recherchen in heterogenen Informationsräumen zu schaffen. Diese können sowohl in Hinblick auf die in ihnen enthaltenen Eigenschaften zur Unterstützung von Suchprozessen als auch nach Art der Datenhaltung differenziert werden. Im Artikel werden die unterschiedlichen Interoperabilitätskonzepte und ihr jeweiliger Mehrwert für den thematischen Zugriff auf Informationen vorgestellt. Das Potential integrierter begrifflicher Interoperabilität insbes. zur Unterstützung von explorativen Suchprozessen wird unter Rückgriff auf die in dem DFG-Projekt *CrissCross*<sup>1</sup> erstellten Mappings zwischen der Schlagwortnormdatei (SWD) und der Dewey-Dezimalklassifikation (DDC) veranschaulicht.

## *1 Thematische Suche*

Ausgehend von dem traditionellen Information-Retrieval-Modell werden thematische Suchen häufig vereinfacht als einstufige Prozesse verstanden, die mit der Eingabe einer natürlich-sprachigen Suchanfrage beginnen und über einen automatischen Abgleich von Zeichenketten (Stringmatching) zu einer Anzeige von Informationsressourcen führen und damit enden. Der Mehrwert von Wissensorganisationssystemen, die auch für die inhaltliche Erschließung verwendet werden, wird dabei in der Bereitstellung eines normierten Vokabulars gesehen, durch das ein standardisierter Zugriff auf Information gegeben ist und mittels dem die *Precision* der Treffermengen gesteigert wird. Dies setzt jedoch voraus, dass die Informationssuchenden einerseits mit dem verwendeten Vokabular vertraut sind und idealerweise die Gesamtheit der für die Themenbeschreibung relevanten Begriffe kennen, andererseits bereits vor Formulierung der Suchanfrage eine genaue Vorstellung des Themas in seiner ganzen Komplexität inklusive aller relevanten (Neben-)Aspekte besitzen.

Selbst in homogenen Informationsräumen kann keine der genannten Prämissen uneingeschränkt als zutreffend gelten. Infolgedessen greifen auch die darauf aufbauenden Überlegungen zu Anforderungen an Wissensorganisationssysteme zur Unterstützung thematischer

---

<sup>1</sup> CrissCross Homepage: <http://linux2.fbi.fh-koeln.de/crisscross/index.html>.

Rechercheprozesse zu kurz. Ein erweitertes Verständnis von *thematischer Suche* wird notwendig, das dieselbige als einen mehrdimensionalen Prozess begreift, der sich aus einer Vielzahl von Einzelrecherchen zusammensetzen, aus einer Aufeinanderfolge unterschiedlicher Suchstrategien bestehen und sowohl automatische als auch kognitive Vorgänge umfassen kann.

Der Einstieg in die Suche erfolgt entweder mit Zeichenketten (Wörtern) oder mit Begriffen. Die darauf aufbauenden Rechercheprozesse können dementsprechend als *wortbasierte* bzw. *begriffliche* Suchen bezeichnet werden<sup>2</sup>. Wissensorganisationssysteme zielen darauf ab, begriffliche Suchen zu unterstützen, die das thematisch Gemeinte als Ausgangspunkt nehmen. Ihr adäquater Einsatz erfordert, dass die Informationssuchenden Einblick in alle zur Bestimmung des begrifflichen Bedeutungsumfangs erforderlichen Informationen erhalten und alternative Benennungen<sup>3</sup> im Retrieval ebenso berücksichtigt werden wie die primären Repräsentationsformen von Begriffen.<sup>4</sup>

Die Suche im engeren Sinn ist i.d.R. durch eine Kombination von automatischen Retrievalprozessen, die auf geeigneten Matchingparametern aufsetzen, und kognitiven Explorationsprozessen<sup>5</sup> geprägt. Die Explorationsprozesse können den automatischen Retrievalprozessen sowohl vor- als auch nachgeordnet sein und sind darauf ausgerichtet, (weiterführende) Informationen und Informationsressourcen zu finden, die für die zugrunde liegende thematische Fragestellung relevant sind. Sie basieren auf einer Mensch-Maschine-Interaktion und können durch eine angemessene Integration von Wissensorganisationssystemen in die Retrievalumgebung wesentlich unterstützt werden.

Wissensorganisationssysteme fördern vor allem Prozesse *begrifflicher Exploration*, d.h. die Erkundung des durch eine Dokumentationssprache repräsentierten Wissensraums. Die Ausweisung eines möglichst lückenlosen, hinreichend differenzierten Netzes von a priori Beziehungen zwischen Begriffen hilft den Informationssuchenden, sich einen Überblick über die in der Erschließungspraxis verwendeten Begrifflichkeiten zu verschaffen und herauszufinden, mit welchen Begriffen das dem Suchinteresse zugrunde liegende Thema am besten beschrieben werden kann. Die aktive Auseinandersetzung mit dem Vokabular kann zudem zu einer Bewusstwerdung verschiedener Aspekte des Themas beitragen und ggf. auch zu einer Spezifizierung desselben führen. Anschließende Recherchen werden dadurch effizienter.

Vom Stringmatching bis hin zur Wissensexploration, so lässt sich stichpunktartig das Spektrum der thematischen Suchprozesse beschreiben, das idealerweise auch in heterogenen Informationsräumen durch entsprechende – die Funktionalität einzelner Wissensorganisationssysteme ergänzende – semantische Interoperabilitätsmodelle unterstützt werden soll. In welcher Form dies den einzelnen Ansätzen zur Heterogenitätsbehandlung gelingt, wird im Folgenden erörtert.

---

<sup>2</sup> Vgl. Gödert 2007, 233.

<sup>3</sup> Alternative Benennungen sind in verbalen Begriffssystemen wie Thesauri die ausgewiesenen Synonyme. In Retrievalumgebungen können auch die in manchen Klassifikationssystemen ausgewiesenen Indexterme wie alternative Benennungen behandelt werden.

<sup>4</sup> Der Begriff „primäre Repräsentationsformen“ bezieht sich im Fall von Klassifikationen primär auf die die Begriffe repräsentierenden Notationen, im Fall von verbalen Begriffssystemen wie Thesauri auf den jeweiligen Deskriptor bzw. die Vorzugsbenennung / Ansetzungsform.

<sup>5</sup> Zu explorativen Suchprozessen siehe auch White/Roth 2009.

## ***2 Modelle semantischer Interoperabilität***

Kennzeichnend für heterogene Informationsräume ist die Vielzahl sowie die sprachliche, strukturelle und typologische Unterschiedlichkeit der zur inhaltlichen Beschreibung von Informationsressourcen eingesetzten Wissensorganisationssysteme. Sowohl umfassende begriffliche Recherchen, die auf automatischen Retrievalvorgängen fußen, als auch explorative Suchprozesse können nur unzureichend unterstützt werden. Zudem repräsentieren die in den einzelnen Erschließungsinstrumenten enthaltenen Begrifflichkeiten und Strukturen jeweils *einen* spezifischen Blickwinkel, der der perspektivischen Vielfältigkeit der verwendeten Systeme und der den Raum nutzenden Informationssuchenden entgegen steht.

Die Herstellung einer Verbindung zwischen Begriffen unterschiedlicher Dokumentations-sprachen erweitert die Funktionalität einzelner Wissensorganisationssysteme und lässt sie auf Informationen anderer Erschließungsinstrumente zugreifen, durch die zugleich der Zugang zu anders erschlossenen Informationsressourcen gewährleistet wird. Der sich dadurch dauerhaft ergebende Mehrwert für thematische Suchprozesse wird dabei wesentlich von den zugrunde liegenden Mappingstrategien sowie der Datenpflege und der Art der Datenhaltung beeinflusst.

### ***2.1 Stufenmodell semantischer Interoperabilität***

Unter dem Gesichtspunkt, auf welcher Grundlage die Verknüpfung von Begriffen unterschiedlicher Wissensorganisationssysteme erfolgt und welche Eigenschaften sich daraus für die Unterstützung sowohl automatischer als auch kognitiver (explorativer) Suchprozesse ergeben, lassen sich drei Stufen semantischer Interoperabilität unterscheiden.

#### ***2.1.1 Wortbasierte Interoperabilität***

Semantische Interoperabilitätsmodelle der untersten Stufe setzen bei den in Dokumentations-sprachen enthaltenen Wörtern bzw. Zeichenketten an und berücksichtigen häufig nur die primären Repräsentationsformen von Begriffen. Die erstellten Verbindungen sind ungerichtet und werden durch den Einsatz von automatischen Matchingverfahren unter Berücksichtigung verschiedener Ähnlichkeitsparameter erzeugt. Verbreitet sind vor allem quantitativ-statistische Ansätze der Heterogenitätsbehandlung, die Interoperabilität auf der Basis dokumentbezogener Term-Kookurenzen schaffen und damit die Indexierungsebene miteinbeziehen.<sup>6</sup>

Die erstellten Verbindungen sind darauf ausgerichtet, automatische sprachübergreifende thematische Recherchen zu ermöglichen und den *Recall* zu erhöhen. Die unter Rückgriff auf die Mappings initiierte Suchraumexpansion besteht aus erweiterten wortbasierten Suchen, die nicht von dem eigentlichen semantischen Gehalt des aus einer Dokumentations-sprache stammenden ursprünglichen Suchbegriffs ausgehen und damit mit einer geringeren *Precision* gekoppelt sind. Begriffliche Suchen, die auf das thematisch Gemeinte zielen, bleiben im Rahmen automatischer Suchprozesse aufgrund der Unschärfe der erstellten Verbindungen weiterhin auf einzelne Dokumentationssprachen und die mit ihnen erschlossenen Informationsressourcen beschränkt. Explorative Suchen, die über den Begriffsraum eines einzelnen Wissensorganisationssystems hinausgehen, werden durch dieses Interoperabilitätsmodell nicht unterstützt.

---

<sup>6</sup> Zu automatischen Verfahren der Heterogenitätsbehandlung vgl. u.a. Wang et al. 2009, Krause 2004, 81ff.

### 2.1.2 Begriffliche Interoperabilität

Begriffliche Interoperabilitätsmodelle enthalten die bestmöglichen Verbindungen zwischen den in den unterschiedlichen Dokumentationssprachen ausgewiesenen Begriffen ungeachtet der jeweiligen Erschließungspraxis. Sie basieren auf intellektuellen Methoden der Heterogenitätsbehandlung und berücksichtigen die Strukturen und typologischen Charakteristika der einzelnen Wissensorganisationssysteme. Elaborierte Modelle begrifflicher Interoperabilität zeichnen sich durch gerichtete Mappingrelationen aus, die präzise Angaben darüber machen, welches Vokabular als Ausgangs- und welches als Zielvokabular dient und damit Hinweise auf ihren adäquaten Einsatz in Retrievalumgebungen geben.

Verknüpfungen zwischen typologisch ähnlichen Erschließungsinstrumenten werden häufig bidirektional bzw. bilateral modelliert und implizieren eine gewisse Austauschbarkeit von Begriffen in Retrievalszenerarien.<sup>7</sup> *CrissCross* als eines der wenigen Projekte, die sich zum Ziel gesetzt haben, Brücken zwischen typologisch unterschiedlichen Systemen zu bauen, verfolgt hingegen einen unidirektionalen Verknüpfungsansatz. Den Begriffen der deutschen Schlagwortnormdatei (SWD) werden Notationen der Dewey-Dezimalklassifikation (DDC) zugeordnet, wobei die SWD als Ausgangs- und die DDC als Zielvokabular dient.<sup>8</sup> Die anvisierten Retrievalszenerarien sehen entsprechend vor, dass die SWD-Schlagwörter als erweitertes Zugangsvokabular zur DDC bzw. zu DDC-erschlossenen Informationsressourcen fungieren. Die intendierte Suchraumexpansion erfolgt somit ausschließlich in Hinblick auf das Zielvokabular.

Sowohl bidirektionale als auch unidirektionale begriffliche Interoperabilität ermöglicht eine inhaltsbezogene Erweiterung einer begrifflichen Suchanfrage, die auf dem mit dem verwendeten Begriff thematisch Gemeinten aufbaut. In heterogenen Informationsräumen werden so auch begriffliche Suchen unterstützt. Im Rahmen automatischer Rechercheprozesse fördern die begrifflichen Mappings die *Precision* der Treffermenge. Zudem schaffen sie eine solide Basis für explorative Suchprozesse, die über den Begriffsraum eines einzelnen Wissensorganisationssystems hinausgehen.

### 2.1.3 Differenzierte Interoperabilität

Differenzierte Interoperabilitätsmodelle geben Auskunft über die Art der zwischen den Begriffen der unterschiedlichen Dokumentationssprachen erstellten Verbindungen und erhöhen damit die Expressivität und Funktionalität der Mappingrelationen. Anwendungsorientierte Formen der Spezifikation fördern das (automatische) Ranking von Treffermengen in heterogenen Informationsräumen und können zur Steuerung von *Recall* und *Precision* beitragen. Begriffsbezogene Spezifizierungen erleichtern hingegen den Übergang von einem Wissenssystem zu einem anderen und unterstützen damit die kognitive Erkundung des durch mehrere Begriffssysteme repräsentierten gesamten Wissensraums. Ambitionierte Ansätze erfordern ein über die Basisrelationen (äquivalenter Begriff, Oberbegriff, Unterbegriff, verwandter Begriff) hinausgehendes Inventar von typisierten Mappingrelationen, dessen Mehrwert sich indes erst im Zusammenhang mit einer lückenlosen, formallogischen und typisierten Relationierung der einzelnen Wissensorganisationssysteme voll entfalten kann.

---

<sup>7</sup> Ein prominentes Beispiel hierfür ist mit den im Projekt *Kompetenzzentrum Modellbildung und Heterogenitätsbehandlung (KoMoHe)* erstellten Mappings gegeben. Zu KoMoHe vgl. u.a. Mayr/Petras 2008.

<sup>8</sup> Die CrissCross-Vorgehensweise wird detailliert dargestellt in Hubrich (2008).

Die in *CrisisCross* vergebenen *Determiniertheitsgrade* zielen vor allem auf ein Ranking DDC-erschlossener Informationsressourcen und tragen zudem zur Steuerung von *Recall* und *Precision* in Hinblick auf DDC-erschlossene Dokumente bei.<sup>9</sup> Ihre Vergabe orientiert sich an den in der DDC inherenten Thema-Klasse-Beziehungen, wie sie in Hier-auch-Anweisungen<sup>10</sup> oder Einschließlich-Hinweisen<sup>11</sup> explizit gemacht werden. Diese sind mit spezifischen Implikationen für die Klassifizierungspraxis verbunden, deren Berücksichtigung in Retrievalszenerarien zu einer positiven Beeinflussung von *Recall* und *Precision* beitragen kann. So verweisen Hier-auch-Anweisungen auf Themen, die wesentliche Übereinstimmung mit dem Klasseninhalt aufweisen und bei denen aufgrund dessen das Anhängen von Hilfstabellen erlaubt ist, d.h. dokument-spezifische synthetische Notationen gebildet werden dürfen, wohingegen dies bei Themen, die in Einschließlich-Hinweisen stehen, verboten ist. Für Retrievalszenerarien bedeutet dies zum Beispiel, dass ein automatischer Einbezug von dokumentspezifischen untergeordneten Notationen mittels Rechtstrunkierung zur Erhöhung des *Recalls* in Hinblick auf DDC-erschlossene Dokumente zweckmäßig ist, wenn eine wesentliche Übereinstimmung des Grundthemas vorliegt. Bei einer Suche nach DDC-erschlossenen Dokumenten über eine verbale Suchanfrage in Form eines Schlagworts kann ein solcher Prozess automatisch unter Rückgriff auf die *Determiniertheitsgrade* initiiert werden.

Die Möglichkeiten einer auf den *CrisisCross*-Relationen aufbauenden rein begriffsbezogenen Typisierung von Mappingrelationen werden im Rahmen der theoretischen Reflexionen des *CrisisCross*-Projekts in Zusammenarbeit mit dem an der Fachhochschule Köln durchgeführten Projekt *RESEDA*<sup>12</sup> ausgelotet. Eine solche Typisierung würde gerade angesichts der in *CrisisCross* praktizierten *One-to-Many*-Methode, derzufolge aufgrund der Polysemie der SWD-Schlagwörter einem Schlagwort mehrere Notationen zugeteilt werden können, den Übergang von der SWD zu der DDC und den Umgang mit den begrifflichen Unterschieden vor allem im Rahmen kognitiver Suchprozesse erheblich erleichtern. Erste Ansätze hierzu werden in Boteram & Hubrich 2010 vorgestellt.

## 2.2 Datenpflege und Modelle der Datenhaltung

Ein dauerhafter Mehrwert für Suchszenarien ist mit den zwischen unterschiedlichen Begriffssystemen erstellten Verbindungen nur mit adäquater Datenpflege und Datenhaltung gegeben. Gerade Modelle begrifflicher und differenzierter Interoperabilität erfordern elaborierte Konzepte für den Umgang mit begriffsbezogenen Bedeutungsänderungen in Wissensorganisations-systemen, durch die sich auch die Beziehung zwischen Begriffen unterschiedlicher Dokumentations-sprachen ändern kann. Ihre Tauglichkeit kann langfristig zudem nur gewährleistet werden, wenn die Verknüpfungen nicht auf der Ebene der Klartextbezeichnungen (d.i. Wortebene) erfolgen, sondern auf der Grundlage von eindeutigen Identifikationsmerkmalen wie Identifikationsnummern oder – gerade auch in Hinblick auf neue Technologien und die Verwendbarkeit der Daten im Rahmen des *Semantic Web* – auf der Basis von *URIs* (*Uniform resource identifier*).

---

<sup>9</sup> Die mit *CrisisCross* gegebenen Rankingmöglichkeiten werden ausführlich beschrieben in Hubrich (2010).

<sup>10</sup> Vgl. DDC-Glossar in DDC 22 (2005), lxxxiii.

<sup>11</sup> Vgl. ebda., lxxxi.

<sup>12</sup> RESEDA-Homepage: <http://linux2.fbi.fh-koeln.de/reseda/index.html>

Am weitesten verbreitet ist die Speicherung der zwischen verschiedenen Erschließungsinstrumenten erstellten Verknüpfungen in einer eigenständigen, externen Datenbank. Um die aktuelle Gültigkeit der Mappings zu garantieren, muss immer wieder ein Abgleich mit den Ursprungsdaten, d.i. mit den Begriffen der verknüpften Dokumentationssprachen, stattfinden. Die Einbettung der Mappingdaten in Retrievalumgebungen ermöglicht eine automatische Erweiterung des Suchraums. Explorative Suchprozesse können erst dann unterstützt werden, wenn zusätzlich auch die verwendeten Wissensorganisationssysteme bereit gestellt werden.

Dem Modell der externen Datenhaltung, das auch charakteristisch für Crosskonkordanzen ist, steht das Modell der integrierten Datenhaltung gegenüber, in dem die Mappingdaten als integraler Bestandteil einer Dokumentationssprache gespeichert werden. Die Pflege der Verknüpfungen wird dadurch erleichtert, da zumindest in Hinblick auf *eine* Erschließungssprache alle hierfür notwendigen Informationen mit den Mappingdaten zur Verfügung stehen. Die Integration der Verknüpfungen in Retrievalumgebungen geht zugleich mit dem Einbezug eines Wissensorganisationssystems einher. So können sowohl automatische Suchprozesse als auch kognitive Recherchen unterstützt werden, wobei Prozesse der begrifflichen Exploration zunächst auf das eine System beschränkt bleiben. Allerdings kann dieses, gerade wenn es über kein lückenloses Relationsgefüge verfügt, durch das zusätzliche integrierte Fremdelement (d.i. die integrierte Verbindung zu einem anderen Begriffssystem) indirekt an Struktur gewinnen. Das ursprüngliche Erschließungsinstrument wird dadurch leistungsfähiger.

Die integrierte Datenhaltung in *CrisisCross* zeichnet sich dadurch aus, dass die Mappings mit einem Zeitstempel gekennzeichnet sind, durch den die Gültigkeit der erstellten Verknüpfungen zwischen SWD und DDC verifiziert werden können und auf denen später eine gelenkte Recherche aufsetzen kann. Eine Aktualisierung der Mappings in Anschluss an Änderungen seitens der DDC oder SWD ist geplant; Mappings, die nicht mehr aktuelle Gültigkeit beanspruchen können, werden mit Angabe des Gültigkeitszeitraums in einem separaten Feld gespeichert. Die Verbindung zur DDC wird über die expressive DDC-Notation hergestellt, die auch als eine wesentliche Komponente einer DDC-spezifischen URI fungiert.<sup>13</sup>

### ***3 Exploration mittels integrierter begrifflicher Interoperabilität***

Die in *CrisisCross* praktizierte integrierte begriffliche Interoperabilität erhöht die Funktionalität der SWD zur Unterstützung der kognitiven Erkundung des SWD-spezifischen Wissensraums. Zudem bietet sie in heterogenen Informationsräumen eine gute Grundlage für Prozesse der umfassenden Begriffsexploration, indem sie den bestmöglichen Übergang von SWD zu DDC schafft und damit den Umgang mit diesen beiden Dokumentationssprachen erleichtert.

Bei der SWD handelt es sich um ein historisch gewachsenes Erschließungsinstrument, das zwar Thesaurusrelationen aufweist, eine durchgehende, lückenlose Relationierung jedoch missen lässt.<sup>14</sup> Die DDC-Notationen, die in Gefolge des *CrisisCross*-Projekts zu einem Bestandteil der SWD werden, haben eine strukturbildende Wirkung auf die SWD, durch die das vor-

---

<sup>13</sup> Vgl. auch Panzer (2008).

<sup>14</sup> In einer unveröffentlichten Studie, die 2004 im Auftrag der Deutschen Bibliothek an der FH Köln durchgeführt wurde, konnte ermittelt werden, dass 18% der Schlagwörter über gar keine Relation verfügten und 34% der Schlagwörter ohne Ober-, Unter- oder verwandte Begriffe waren. Zwar konnten seitdem bereits einige Relationen „nachgetragen“ werden, der grundsätzliche Tatbestand ist indes weitgehend unverändert geblieben.

handene Relationsnetz erweitert und infolgedessen eine verbesserte Navigation in der SWD ermöglicht wird. SWD-Begriffe werden über ihren Bezug zu den zugeordneten DDC-Notationen zueinander in Beziehung gesetzt und zu Schlagwortclustern zusammengefasst. Zusätzliche Verbindungen zwischen Schlagwörtern bzw. den durch die DDC-Notationen sich bildenden Schlagwortclustern sind indirekt mittels der hierarchieabbildenden Eigenschaft der DDC-Notationen gegeben, über die untergeordnete bzw. übergeordnete Klassen anhand der Ziffernfolge der Notation und der Notationslänge identifiziert werden können. Unrelationierte Schlagwörter, die bisher nur als Einzelbegriffe auffindbar waren, werden nun als Cluster in der SWD recherchierbar. Welcher Mehrwert dadurch entsteht, tritt deutlich durch das in Abb. 1 visualisierte Beispiel aus dem Bereich der Biologie<sup>15</sup> zutage, das zweimal denselben Ausschnitt aus der SWD zeigt – einmal mit Anzeige der SWD-Relationen und einmal mit Anzeige der SWD-Relationen sowie zusätzlich der Relationen, die durch die SWD-DDC-Verknüpfungen entstehen.

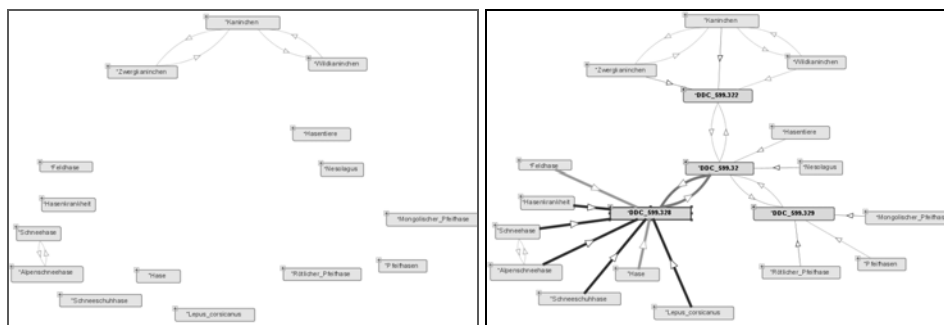


Abb. 1: Clustering von unrelatierten Schlagwörtern durch CrissCross-Mappings

Die *CrissCross*-Mapping unterstützen jedoch nicht nur eine erweiterte SWD-Begriffsexploration, sondern tragen zudem zu einer Differenzierung des SWD-Relationsgefüges bei. Die Verwendung einer verhältnismäßig kleinen Anzahl von Relationstypen (Synonym, Ober-/ Unterbegriff, verwandter Begriff) führt in der SWD teils dazu, dass die Verbindung zwischen einem Schlagwort und einer großen Anzahl anderer Schlagwörter durch einen einzelnen Relationstyp charakterisiert ist. In der SWD sind zum Beispiel zum Schlagwort „Jagd“ drei verwandte Begriffe sowie zahlreiche Unterbegriffe ausgewiesen, die zur Unterstützung von kognitiven begrifflichen Suchen zunächst nur unsortiert oder alphabetisch angezeigt werden können. Unter Rückgriff auf die den einzelnen Schlagwörtern zugeordneten expressiven DDC-Notationen lassen sich die Unterbegriffe in verschiedene Cluster von thematisch näher stehenden Begriffen unterteilen. Formal erfolgt dadurch indirekt bereits eine Spezifizierung der Relation, die als Grundlage für eine weitergehende Typisierung der Relation dienen kann. In dem konkreten Fallbeispiel können die DDC-Klassenbenennungen als ein Hilfsmittel zur

<sup>15</sup> Bis vor kurzem wurde auf die Ausweisung hierarchischer Relationen in der biologischen Nomenklatur verzichtet. Erst die in diesem Jahr herausgegebene 5. Ergänzungslieferung der 3. überarb. u. erw. Ausgabe der RSWK sieht die Angabe bestimmter hierarchischer Verweisungen vor (vgl. RSWK 2009, 135). Daher sind die meisten Schlagwörter der SWD, die sich auf Tiere beziehen, relationslos.

Beschreibung der Relationstypen herangezogen werden: so bezieht sich ein Teil der Unterbegriffe von „Jagd“ auf *Jagdmethoden*, ein anderer Teil wiederum auf *Jagd auf einzelne Tierarten* (vgl. Abb. 2). Eine schnellere Einschätzung bezüglich der Relevanz der Unterbegriffe in Hinblick auf das Suchinteresse wird möglich.

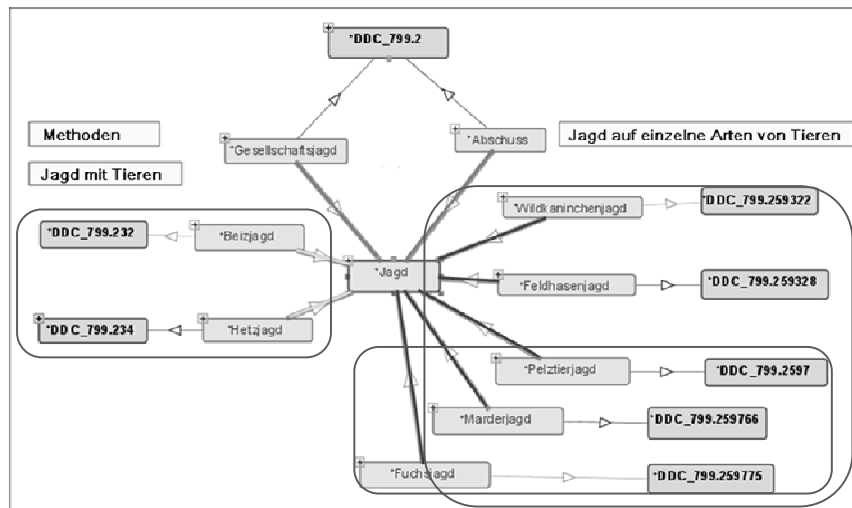


Abb. 2: Differenzierung von SWD-Relationen durch CrissCross-Mappings

Die Integration sowohl von SWD als auch von DDC in Retrievalumgebungen führt zu einem weiter verzweigten Relationsnetz, das über den perspektivisch begrenzten Wissensraum der SWD hinausgeht und der Multiperspektivität einer umfassenden explorativen Suche in heterogenen Informationsräumen entgegenkommt. Die DDC stellt neben den sich in den Notationen widerspiegelnden hierarchischen Verbindungen zwischen Klassen auch weitere Bezüge zwischen Klassen einerseits und Themen und Klassen andererseits her, die eine wichtige Orientierung für Prozesse der Begriffsexploration bieten können. Manche Beziehungen sind zur Zeit zwar nur indirekt in Klassifizierungshinweisen o.ä. enthalten, es ist aber zu erwarten, dass mit der ontologischen Weiterentwicklung der DDC diese stärker explizit gemacht werden und damit den Informationssuchenden besser zur Verfügung stehen.<sup>16</sup> Mit der Angabe und Bereitstellung von Bestandteilen von synthetischen Notationen, die Teil der DDC-Wissensbasis sind, ergeben sich zusätzliche Möglichkeiten für Explorationsprozesse. Im Rahmen des *CrissCross*-Projekts werden zum Zwecke der Erstellung der bestmöglichen Verbindung zwischen Schlagwort und DDC-Notation<sup>17</sup> auch synthetische Notationen erstellt, die zusammen mit Notationsteilen, Klassenbenennungen und Registereinträgen zunächst institutionell in MelvilClass<sup>18</sup> gespeichert werden. Auf Dauer sollen sie in die Wissensbasis der deutschen DDC eingehen und für Prozesse der Begriffsexploration bereit gestellt werden.

<sup>16</sup> Vgl. u.a. Panzer & Zeng 2009, Green & Panzer 2010

<sup>17</sup> Diese Vorgehensweise wird im Kontext des Mappingprozesses auch als *Deep-Level-Mapping* bezeichnet.

<sup>18</sup> Vgl. <http://melvil.d-nb.de/>



#### 4 Ausblick

Thematische Suchprozesse in heterogenen Informationen erfordern adäquate Konzepte zur Verbindung und Nutzung der in den verwendeten Erschließungsinstrumenten enthaltenen Informationen unter Berücksichtigung der mit ihnen gegebenen Charakteristika. Das mit Mappings gegebene Potential richtet sich nach dem zugrundeliegenden Interoperabilitätsmodellen, die jeweils mit bestimmten Implikationen für die Suche verbunden sind. Mit dem Projekt *CrossCross* wurde ein elaborierter Ansatz der Heterogenitätsbehandlung detaillierter vorgestellt, der sowohl automatische begriffliche Suchraumerweiterungen ermöglicht als auch zusätzliche Strukturen zur kognitiven Erkundung des Begriffsraums bietet und damit den Anforderungen an ein Interoperabilitätsmodell, das in Ergänzung zu den vorhandenen Wissensorganisationssystemen alle Suchprozesse vom Stringmatching bis hin zur Begriffsexploration unterstützt, gerecht wird.

Umfassende Prozesse der Begriffsexploration, die sich auf den durch alle in dem Informationsraum verwendeten Dokumentations Sprachen repräsentierten Wissensraum beziehen, können jedoch nicht allein durch geeignete Mappingdaten unterstützt werden, die als Brücke zwischen den unterschiedlichen Wissenssystemen fungieren. Eine zusätzliche Unterstützung, die vor allem den Umgang mit den sich in den unterschiedlichen Strukturen widerspiegelnden lokalen Perspektiven erleichtert, ist notwendig. Diese kann durch ein umfassendes Wissensorganisationssystem gegeben werden, bei dem der Übergang von einer Dokumentations-sprache zu einer anderen über ein internationalisiertes Kernsystem erfolgt (vgl. Abb. 3).<sup>19</sup>

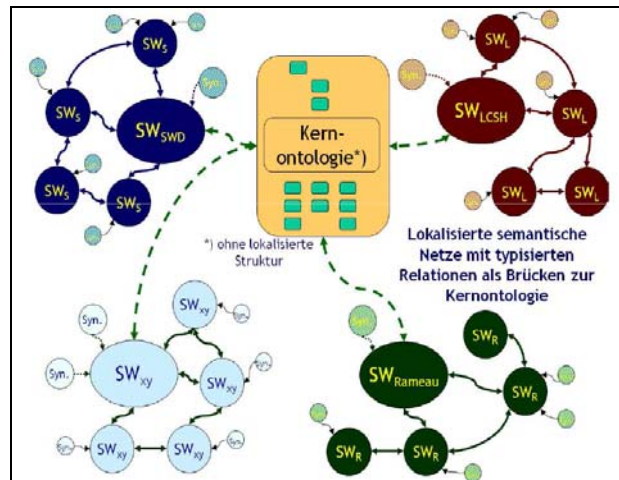


Abb. 3: Modell eines umfassenden Wissensorganisationssystems zur Unterstützung der Begriffsexploration

<sup>19</sup> Vgl. Gödert 2008.

## Literatur

- Boteram, F.; Hubrich J.: Specifying Intersystem Relations: Requirements, Strategies, Issues. In: Knowledge Organization (2010). [erscheint in Kürze]
- Dewey-Dezimalklassifikation und Register. Begr. v. Melvil Dewey. DDC 22. Deutsche Ausgabe. Hrsg. von Der Deutschen Bibliothek. Bd. 1. München: Saur, 2005.
- Gödert, W.: Multilingualität und Lokalisierung zur Wissenserkundung oder vom Nutzen semantischer Netze für das Information Retrieval. In: Weigel, Harald (Hrsg.): Wa(h)re Information. 29. Österreichischer Bibliothekartag Bregenz 2006. Graz-Feldkirch: Neugebauer, 2007. (Schriften der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare; 2), S. 233-242.
- Gödert, W.: Ontological Spine, localization and multilingual access. Some reflections and a proposal. In: New Perspectives on Subject Indexing and Classification. Essays in Honour of Magda Heiner-Freiling. Leipzig u.a.: Deutsche Nationalbibliothek 2008, S. 233-240.
- Green, R.; Panzer, M.: The Ontological Character of Classes in the Dewey Decimal Classification. In: Paradigms and conceptual systems in KO: proc. Eleventh int. ISKO conference, Rome, 23-26 February 2010. [wird in Kürze veröffentlicht]
- Hubrich, J.: Thematische Suche in heterogenen Informationsräumen auf der Basis von unidirektionaler begrifflicher Interoperabilität. In: The Ne(x)t Generation. 30. Österreichischer Bibliothekartag, Graz 15. - 18. September 2009. Graz 2010. [Tagungsband erscheint in Kürze]
- Hubrich, J.: CrissCross. SWD-DDC-Mapping. In: Mitteilungen der Vereinigung österreichischer Bibliothekarinnen & Bibliothekare 61 (2008), Heft 3, S. 50-58. <<http://www.univie.ac.at/voeb/php/downloads/vm6120083.pdf>> (abgerufen am 01.12.2009).
- Mayr, P.; Petras, V.: Cross-concordances: terminology mapping and its effectiveness for information retrieval. In: World Library and Information Congress: 74th IFLA General Conference and Council. 10-14 August 2008, Québec, Canada. <[http://archive.ifla.org/IV/ifla74/papers/129-Mayr\\_Petras-en.pdf](http://archive.ifla.org/IV/ifla74/papers/129-Mayr_Petras-en.pdf)> Eine deutsche Übersetzung ist unter dem Titel „Crosskonkordanzen: Terminologie Mapping und deren Effektivität für das Information Retrieval“ erschienen. <[http://archive.ifla.org/IV/ifla74/papers/129-Mayr\\_Petras-trans-de.pdf](http://archive.ifla.org/IV/ifla74/papers/129-Mayr_Petras-trans-de.pdf)> (abgerufen am 01.12.2009)
- Krause, J.: Konkretes zur These, die Standardisierung von der Heterogenität her zu denken. In: Zeitschrift für Bibliothekswesen und Bibliographie 51 (2004), 2, S. 76-89.
- Panzer, M.: Cool URIs for the DDC: Towards Web-Scale Accessibility of a Large Classification System. In: Greenberg, J. and Klas, W. (eds.): Metadata for semantic and social applications: proceedings of the International Conference on Dublin Core and Metadata Applications, Berlin, 22-26 September 2008. Göttingen: Univ.-Verl. 2008. S. 183-90. <<http://dcpapers.dublincore.org/ojs/pubs/article/view/932/928>> (abgerufen am 01.12.2009)
- Panzer, M.; Zeng, M.: Modeling Classification Systems in SKOS: Some Challenges and Best-Practice Recommendations. In: Proceedings of the International Conference on Dublin Core and Metadata Applications 2009, Seoul, Korea, 12-16 October 2009. S. 3-14. <<http://dcpapers.dublincore.org/ojs/pubs/article/view/974/944>> (abgerufen am 01.12.2009)
- Regeln für den Schlagwortkatalog (RSWK). 3., überarbeitete und erweiterte Auflage auf dem Stand der 5. Ergänzungslieferung April 2009. Hrsg. von der Arbeitsstelle für Standardisierung. Leipzig. Frankfurt am Main, Berlin: Deutsche Nationalbibliothek, 2009. <[http://files.d-nb.de/pdf/rswk\\_gesamtausgabe.pdf](http://files.d-nb.de/pdf/rswk_gesamtausgabe.pdf)> (abgerufen am 01.12.2009)
- Wang, S.; Isaac, A.; Schopman, B.; Schlobach, S.; van der Meij, Lourens: Matching multilingual subject vocabularies. In: Agosti, M. u.a. (Hrsg.): Research and Advanced Technology for Digital Libraries. 13<sup>th</sup> European Conference, ECDL 2009, Corfu, Greece, September 27- October 2, 2009. Proceedings. Berlin, Heidelberg: Springer, 2009, S. 125-137. <<http://www.few.vu.nl/~aisaac/papers/telplus-ecdl09.pdf>> (abgerufen am 01.12.2009).
- White, R. W.; Roth, R. A.: Exploratory Search. Beyond the Query-Response Paradigm. [San Rafael, Calif.]: Morgan & Claypool, 2009. (Synthesis Lectures on information concepts, retrieval and services; 3)