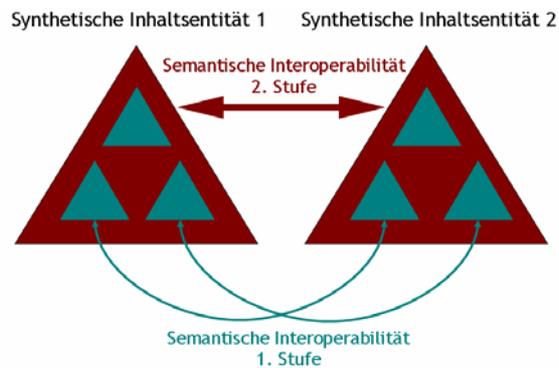
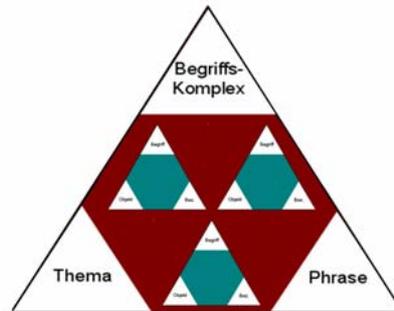
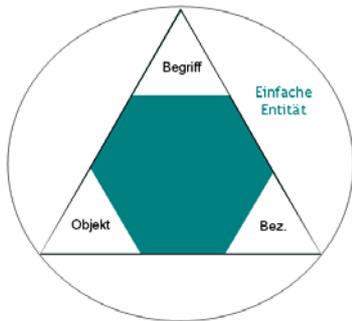


# Semantische Wissensrepräsentation und Interoperabilität

Teil 1: Interoperabilität als Weg zur Wissensexploration

Teil 2: Ein formales Modell semantischer Interoperabilität

Winfried Gödert



Fachhochschule Köln  
Institut für Informationswissenschaft

## Inhalt

<b>1. Einführung</b> .....	2
<b>2. Retrievalkonzepte / -paradigmen</b> .....	2
2.1 <i>Abgleich von Wörtern</i> .....	2
2.2 <i>Begriffliches Suchen</i> .....	3
2.3 <i>Exploration des semantischen Umfelds</i> .....	3
2.4 <i>Thematische Exploration</i> .....	3
<b>3. Semantische Interoperabilität: Kognitive oder maschinelle Interpretation</b> .....	4
<b>4. Anwendungsmodelle semantischer Interoperabilität</b> .....	<b>5</b>
4.1 <i>Konkordanz</i> .....	5
4.2 <i>Citation Pearl Growing</i> .....	8
4.3 <i>Semantische Brücken zu einem Kernsystem</i> .....	11
<b>5. Exkurs: Auswahl der Kernontologie</b> .....	16
<b>6. Anreicherung der lokalisierten Semantischen Netze</b> .....	17
<b>7. Interoperabilität, maschinelle Interpretation und Semantic Web</b> .....	21
<b>8. Retrievalszenarios zur Wissensexploration</b> .....	23
<b>9. Wissensexploration als Konzept für Zukunftsszenarios</b> .....	29
<b>10. Road map zur semantischen Exploration</b> .....	<b>30</b>
<b>11. Typen von Inhaltsentitäten</b> .....	<b>31</b>
<b>12. Typen der Inhaltsentitäten auf der Ebene der Dokumentationssprachen</b> .....	<b>33</b>
12.1 <i>Prinzipien der Begriffs- und Klassenbildung</i> .....	34
12.2 <i>Präkombination, Schlüsselung, Facettierung und Synthese</i> .....	40
<b>13. Typen der Inhaltsentitäten auf der Ebene der dokumentations sprachlichen Repräsentation der Dokument-Gegenstände</b> .....	<b>42</b>
<b>14. Interoperabilitätstypen zwischen Inhaltsentitäten</b> .....	<b>44</b>
14.1 <i>Semantische Interoperabilität 1. Stufe (Begriffliche Interoperabilität)</i> .....	44
14.2 <i>Semantische Interoperabilität 2. Stufe (Thematische Interoperabilität)</i> .....	44
14.3 <i>Interoperabilität zwischen Begriffen und Klassen verschiedener Komplexitätsstufen</i> .	46
<b>15. Die pragmatische Dimension bei Begriffs- und Klassenbildung sowie Konsequenzen für Interoperabilität</b> .....	<b>48</b>

# Semantische Wissensrepräsentation und Interoperabilität

## Teil 1: Interoperabilität als Weg zur Wissensexploration

Dieser aus zwei Teilen bestehende Beitrag setzt Methoden der semantischen Wissensrepräsentation in Verbindung zur Gestaltung von Retrievalsystemen für begriffliche Recherchen und insbesondere für die Wissensexploration. Ausgehend von heterogenen Erschließungssituationen werden Konzepte vorgestellt, wie durch Maßnahmen zur Herstellung von Interoperabilität ein Beitrag zur Lösung der Heterogenitätssituation geleistet werden soll. Basierend auf einem im zweiten Teil vorgestellten formalen Modell zum besseren Verständnis von semantischer Interoperabilität wird ein Vorschlag für ein Gesamtsystem entwickelt, das aus einer Kernontologie und lokalisierten semantischen Netzen mit erweitertem Relationenumfang besteht. Die Möglichkeiten zur Recherche und Exploration in einem solchen Gesamtrahmen werden skizziert.

### 1. Einführung

„Standardisierung von der Heterogenität her denken“<sup>1</sup>, so lautete eine Aufforderung, die das Denken über die Gestaltung von Dokumentations- und Retrievalsystemen während der letzten Jahre beeinflusst hat. So richtig es ist, die Heterogenität der inhaltlichen Erschließung zum Ausgangspunkt für Gestaltungsüberlegungen zu nehmen, so wenig selbstverständlich ist es, das mit der genannten Aussage verbundene Schalenmodell als alleinigen Lösungsweg zu sehen. Um einen Zugang zu alternativen Ansätzen zu bekommen, ist es empfehlenswert, die den Suchfragen zugrunde liegenden Suchinteressen genauer zu analysieren und daraus eine Typologie zu entwickeln. In diesem Beitrag wird zunächst eine solche Typologie und dann ein Lösungsansatz für die Überwindung von heterogenen Erschließungssituationen vorgestellt, der primär der Verbindung von Retrievalansätzen mit Möglichkeiten zur Wissenserkundung gewidmet ist und dazu Gedanken und Techniken der Wissensrepräsentation und der semantischen Interoperabilität aufgreift.

### 2. Retrievalkonzepte / -paradigmen

Die Gestaltung von Dokumentations- und Retrievalsystemen folgt gewissen Paradigmen, die sich an Verbindungen von Methoden der Indexierung mit den Typen von Suchfragenformulierungen sowie dem Stand konzeptioneller Vorstellungen und technischer Realisierung von Retrievalumgebungen orientieren. Nachfolgend wird ein Stufenplan der Möglichkeiten vorgestellt, der sich sowohl am bereits Machbaren als auch am Wünschbaren der vorhersehbaren Möglichkeiten orientiert.

#### 2.1 Abgleich von Wörtern

Die einfachste und grundlegendste Methode der Suche basiert auf dem Abgleich einzelner Wörter. Dieses Suchen über eine nicht weiter zu spezifizierende Volltextsuche anhand eines ein-

---

<sup>1</sup> Vgl. z.B.: **Krause, J., E. Niggemann u. R. Schwänzl:** *Normierung und Standardisierung in sich verändernden Kontexten Beispiel: Virtuelle Fachbibliotheken.*

In: Zeitschrift für Bibliothekswesen und Bibliographie. 50(2003) H.1, S.19-28.

**Krause, J.:** *Konkretes zur These, die Standardisierung von der Heterogenität her zu denken.*

In: Zeitschrift für Bibliothekswesen und Bibliographie. 51(2004) H.2, S.76-89.

**Krause, J.:** *Shell Model, Semantic Web and Web Information Retrieval.*

In: Information und Sprache: Beiträge zu Informationswissenschaft, Computerlinguistik, Bibliothekswesen und verwandten Fächern. Festschrift für Harald H. Zimmermann. Herausgegeben von Ilse Harms, Heinz-Dirk Luckhardt und Hans W. Giessen. München: Saur 2006. S.95-106.

fachen Zeichenkettenabgleichs oder durch einen Metadaten unterstützten Indexaufbau lässt sich technisch leicht realisieren und benötigt keine Aufbereitung der zu durchsuchenden Dokumente oder die Verwendung einer Dokumentationssprache. Allerdings ist die für einen erfolgreichen Retrievalprozess notwendige Vorhersagbarkeit des vorhandenen oder zur Erschließung verwendeten Vokabulars erschwert. Typische Beispiele sind die heute geläufigen Suchmaschinen basierten Retrievalumgebungen. Auch bei der Verwendung komplexer und umfangreicher Dokumentationssprachen ist eine solche Beschränkung der Suchstrategie auf die einzelnen Wörter aufgrund der möglicherweise hohen Spezifität der verwendeten Begriffe problematisch. Eine bewusst gewollte Unschärfe und Ausweitung der Suche lässt sich lediglich über eine Trunkierung der zu suchenden Zeichenkette erreichen, was eine gewisse Variation der Anfrage erlaubt – eine gezielte inhaltliche Suche ist nicht möglich.

## 2.2 *Begriffliches Suchen*

Eine inhaltsbezogene Erweiterung der Suche und der Anfrageformulierung mittels einer begrifflichen Suche, die den Übertritt auf die Begriffsebene und damit auf das thematisch Gemeinte ermöglicht, erfordert die umfassende Verknüpfung aller möglicher Sucheinstiege für einen bestimmten Begriff mit der für die Erschließung verwendeten Vorzugsbenennung anhand ausgewiesener Synonymie-Relationen. Diese müssen in entsprechenden Begriffssystemen in Form von Synonymwörterbüchern oder als Teil von Dokumentationssprachen definiert und für die Suche hinterlegt werden. Für die Zusammenführung und Integration mehrerer Systeme ist die umfassende und präzise Ausweisung solcher Synonymie-Relationen ebenfalls von entscheidender Bedeutung.

## 2.3 *Exploration des semantischen Umfelds*

Sollen nicht nur alternative Sucheinstiege in Form von Synonymen vorgehalten sondern auch die Exploration des gesamten semantischen Umfelds ermöglicht werden, ist eine differenzierte Relationierung anhand geeigneter Begriffsbeziehungen notwendig. Diese vorab definierten semantischen Relationen, welche als a priori, d.h. dokumentunabhängig bestehende Zusammenhänge Bestandteil eines Gegenstandsbereichs und der ihn abbildenden Dokumentationssprache sind, bilden ein präzise zu definierendes Inventar, das die Möglichkeiten einer semantischen Modellierung wesentlich bestimmt. Art und Umfang der Typisierung entscheiden über die Leistungsfähigkeit der Retrievalumgebung. Ambitioniertere Vorstellungen erfordern in jedem Fall ein über die üblichen Thesaurus-Konventionen hinaus gehendes Inventar. Unter dem Gesichtspunkt des noch zu diskutierenden Konzepts semantischer Interoperabilität zeigt sich, dass die Bestimmung und Definition eines Inventars geeigneter Relationstypen nicht nur für die Gestaltung einzelner Dokumentationssprachen und Wissensrepräsentationssysteme, sondern auch für die Verknüpfung mehrerer solcher Systeme von Bedeutung ist.

## 2.4 *Thematische Exploration*

Thematische Exploration bricht sowohl mit der Vorstellung, dass dem Recherchierenden alle begrifflichen Ausdrucksweisen der gewünschten Themenbausteine als auch die verschiedenen Möglichkeiten zur Herstellung syntaktischer Ausdrücke zum Zeitpunkt der Recherche bereits bewusst sein müssen und postuliert, dass das Aufzeigen neuer Zusammenhänge den Rechercheverlauf und die Ergebnismenge positiv beeinflusst. Nach gängigen Vorstellungen liegt jeder Anfrage an ein Information-Retrieval-System ein klar umrissener und präzise verbalisierter Informationsbedarf zugrunde. Geht man aber davon aus, dass Information-Retrieval-Systeme auch die thematische Exploration eines Gegenstandsbereichs ermöglichen sollen, beispielsweise zur Unterstützung der Klärung des Informationsbedarfs und dessen Umsetzung in eine vom System zu verarbeitende Anfrage, so sind weitere strukturelle und semantische Eigenschaften

der verwendeten Dokumentationssprachen sowie der Retrievalsysteme notwendig. Der Retrievalprozess wandelt sich vom einstufigen Such- und Findvorgang zum Explorationsprozess mit nachfolgender Treffermengenbildung.

Zur Formulierung komplexer thematischer Anfragen im Sinne der thematischen Erkundung ist daher nicht allein die begriffliche Exploration entlang zahlreicher semantischer Relationen innerhalb der Dokumentationssprache erforderlich, sondern es ist auch die Durchführung syntaktischer Operationen mit Elementen des Vokabulars notwendig. Es müssen dabei die dokumentspezifischen a posteriori Beziehungen berücksichtigt werden können, die zweckmäßigerweise aspekt- oder facettenorientiert innerhalb der Indexierungsumgebung bereitgestellt und im Suchprozess abgerufen werden können.

Die Realisierung derartiger Überlegungen setzt eine umfangreiche Relationierung sowie ein leistungsstarkes System syntaktischer Operatoren mit Möglichkeiten zum selektiven Einsatz ausgewählter Relationen voraus. Noch lässt sich kein Beispiel angeben, in dem diese Zielsetzung vollständig realisiert wäre, insofern handelt es sich bei dieser Vorstellung zunächst um eine Projektion für zukünftige Entwicklungen.

Im Vordergrund der nachfolgenden Betrachtungen steht daher die Verbesserung der Retrieval-situation unter Berücksichtigung der Heterogenitätsbedingung und der Zielsetzung einer semantischen Wissenserkundung. Eine wichtige Rolle spielt dabei ein genaueres Verständnis semantischer Interoperabilität.

### 3. Semantische Interoperabilität: Kognitive oder maschinelle Interpretation

Mit dem Ausdruck *Semantische Interoperabilität* begegnet einem ein schillernder Begriff, wie neuerdings im Umfeld des sog. *Semantic Web* häufiger. Mal soll damit Ähnlichkeit und Austauschbarkeit von Begriffen gemäß bedeutungsmäßigem Inhalt ausgedrückt werden, mal der fehlerfreie Ersatz oder die fehlerfreie Verknüpfung semantischer Daten zwischen Maschinen. Dementsprechend lässt sich zunächst nicht recht erkennen, ob mit der Verwendung von *Semantischer Interoperabilität* ein Problem beschrieben werden soll oder ob damit im Sinne von Konzepten und Methoden Hilfsmittel zur Lösung von Problemen angegeben werden. Für Betrachtungen im Rahmen der Dokumenterschließung wird immer ein Mittelweg zwischen kognitiver und maschineller Betrachtung erforderlich sein, da einerseits dem maschinellen Austausch und der Nachnutzbarkeit von Daten Rechnung getragen werden muss, andererseits aber auch die kognitive Interpretation der Daten zur eigentlichen Zweckbestimmung inhaltserschließender Aktivitäten gehört. Das Beschreiten eines solchen Mittelweges ist nicht nur schwieriger als jeder der einzelnen Wege, sondern zieht auch Konsequenzen für Repräsentationsmodelle semantischer Daten nach sich, die nicht mehr allein – wie im klassischen Ansatz der Dokumenterschließung – als durch Indexierer oder Benutzer interpretiert gedacht werden dürfen, sondern die insbesondere im Bereich der Beziehungen zwischen den dokumentationssprachlichen Repräsentanten von Inhaltsentitäten<sup>2</sup> formale Eigenschaften aufweisen müssen, um Inferenzen entlang der durch die Relationen definierten Wege zu gestatten.

Illustriert an einem einfachen Beispiel würde dies heißen: In einer durch die Begriffe Bäcker, Backen, Brot, Weißbrot und Kuchen gegebenen Beziehungsstruktur (vgl. Abb.1).

---

<sup>2</sup> Der Ausdruck Inhaltsentität wird als Verallgemeinerung für alle Arten von inhaltskennzeichnenden Ausdrücken benutzt; Näheres dazu in Teil 2 dieses Beitrages: *Semantische Wissensrepräsentation und Interoperabilität: Teil 2: Ein formales Modell semantischer Interoperabilität*.

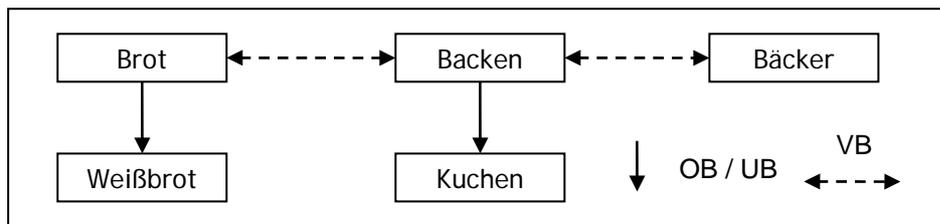


Abb.1: Inferenzen in Beziehungsstrukturen

wäre formal nur modelliert, dass der Bäcker Brot und Kuchen backt. Die Aussage, dass der Bäcker auch Weißbrot backt, müsste entweder explizit durch weitere Beziehungspfade ausgedrückt werden (die ‚klassische‘ Thesaurus-Vorgehensweise) oder über eine formal-logische Inferenz erschlossen werden, indem der Merkmalsbesitz der einzelnen Begriffe genauer spezifiziert wird und der Ober- / Unterbegriffsrelation die Eigenschaft der Vererbbarkeit zugeordnet wird. Letzterer Weg ist in den Modellierungsstrategien der Wissensrepräsentation üblich. Er ist zunächst mit mehr Aufwand verbunden und besitzt nicht die intuitive kognitive Interpretierbarkeit (sind beispielsweise Torten Sorten von Kuchen, die ge„backen“ werden?), besitzt jedoch mehr Potenzial für die Aussagekraft des insgesamt vorhandenen semantischen Gehalts einer Repräsentationsstruktur als auch für die Erstellung von Retrievalsystemen.

In jedem Fall ist der Ausdruck *Semantische Interoperabilität* erläuterungsbedürftig. Im zweiten Teil dieses Beitrages wird genauer angegeben, welches Verständnis von *Semantischer Interoperabilität* hier zugrunde gelegt wird und wie dieses Verständnis unter Berücksichtigung der oben beschriebenen Vereinbarkeit zwischen kognitiver und maschineller Interpretation durch ein formales Modell beschrieben werden kann.

Im Rahmen des *Reseda*-Projekts an der Fachhochschule Köln<sup>3</sup> wird der Frage nachgegangen, wie sich die Anforderungen der semantischen Interpretierbarkeit und logischen Formalisierung in einem Repräsentationsmodell verbinden lassen, das möglichst optimale Unterstützung für Retrievalzwecke bietet. Wertvolle Anregungen konnten dabei aus der Arbeit am *CrissCross*-Projekt zur semantischen Anreicherung des Zugangsvokabulars der *DDC* durch Verknüpfung mit der *Schlagwortnormdatei (SWD)*<sup>4</sup> gewonnen werden.

#### 4. Anwendungsmodelle semantischer Interoperabilität

In diesem Abschnitt sollen zunächst zwei Modelle vorgestellt werden, die unter der Überschrift *Semantische Interoperabilität* für den Einsatz zur praktischen Realisierung in Retrievalumgebungen diskutiert werden. Unter Bezugnahme auf das im zweiten Teil dieses Beitrages angegebene formale Modell zur Semantischen Interoperabilität wird anschließend eine dritte Vorgehensweise vorgestellt, die weniger die semantische Gleichwertigkeit in den Vordergrund stellt, sondern vielmehr das Retrievalkonzept auf der Basis von Brücken zwischen semantischen Satellitensystemen und einem ontologischen Kernsystem entwickelt.

##### 4.1 Konkordanz

Wie durch das eingangs gegebene Zitat gezeigt, wird die Lösung des Heterogenitätsproblems gerne darin gesehen, Verbindungen zwischen den Elementen der zur Erschließung verwendeten

<sup>3</sup> Reseda steht für Repräsentationsmodelle semantischer Daten; vgl.: <http://www.f03.fh-koeln.de/fakultaet/iim/forschungsprojekte/aktuelle/>.

<sup>4</sup> Vgl.: <http://linux2.fbi.fh-koeln.de/crisscross/>.

Dokumentationssprachen herzustellen, gerne *Crosswalks* oder *Cross-Konkordanzen* genannt<sup>5</sup>. So nahe liegend dieser Ansatz zunächst anmutet, er besitzt eine Reihe von Eigenschaften, die näher betrachtet werden müssen.

Bei der Herstellung der Konkordanzen werden Natur oder Qualität der Verbindungen im Allgemeinen nur schwach oder gar nicht differenziert. Der Maßstab für Interoperabilität ist in der Regel eine intellektuelle Bewertung des allgemein semantischen Gehalts der beteiligten Inhaltsentitäten außerhalb ihrer Fixierung im Beziehungsgefüge der jeweiligen Dokumentationssprache. Auch die strukturellen Eigenschaften der beteiligten Dokumentationssprachen werden in der Regel nicht im Mappingprozess typisiert, ggf. wird versucht, Unvergleichbarkeiten durch Hilfskonstruktionen wie gedachte Boolesche Verknüpfungen auszugleichen. Ein Beispiel für diese Vorgehensweise zeigen die nachstehenden Abbildungen (Abb.2 und 3) aus dem *KoMoHe*-Projekt.<sup>6</sup>

Tabelle 1: Beispiel für Crosskonkordanz-Relationen			
Thesaurus Sozialwissenschaften	Relation	Relevanz	Standard Thesaurus Wirtschaft
Abgaben	=	h	Gebühr
Deutsche Bundesbank	=+	h	Zentralbank + Deutschland
Abitur	<	m	Bildungsabschluss
Entschuldung	^	h	Schuldenerlass
Katastrophe	>	g	Naturkatastrophe
Pädagogische Faktoren	0		

Abb.2: Crosskonkordanzen (1) aus dem *KoMoHe*-Projekt

<sup>5</sup> Verfolgt wird dieser Ansatz beispielsweise im *KoMoHe*-Projekt:  
**Mayr, P., V. Petras:** *Building a Terminology Network for Search: the KoMoHe project.*  
 In: Metadata for semantic and social applications : proceedings of the International Conference on Dublin Core and Metadata Applications, Berlin, 22 - 26 September 2008, DC 2008: Berlin, Germany / ed. by Jane Greenberg and Wolfgang Klas. Göttingen: Univ.-Verl. 2008. S.177-182,  
 vgl. auch: <http://www.gesis.org/forschung-lehre/programme-projekte/informationswissenschaften/projektuebersicht/komohe/?L=http://www.gesis.org/>.

<sup>6</sup> Die Abbildung wurde entnommen aus:  
**Mayr, P., A.-K. Walter:** *Einsatzmöglichkeiten von Crosskonkordanzen.*  
 In: <http://www.gesis.org/Information/Forschungsuebersichten/Tagungsberichte/Vernetzung/Mayr-Walter.pdf>.

**Tabelle 2: Beispiel für Crosskonkordanz-Relationen ausgehend von dem Deskriptor „Biologieunterricht“ des Thesaurus Sozialwissenschaften.**

Biologieunterricht	<	Unterricht	DZI
Biologieunterricht	<	Unterricht	Standard Thesaurus Wirtschaft
Biologieunterricht	=	Biologieunterricht	Schlagwortnormdatei
Biologieunterricht	<+	Biology + Teaching	CSA
Biologieunterricht	=+	Naturwissenschaftlicher Unterricht + Biologie	Psyndex Terms
Biologieunterricht	=+	Fachunterricht/Unterrichtsfach + Biologie	IBLK
Biologieunterricht	=+o	Biologie + Schulfach	BISp-Liste
Biologieunterricht	<+o	Biologie + Unterrichtsstunde	BISp-Liste
Biologieunterricht	<+	Biologie + Schule	DZA
Biologieunterricht	^+	Biologie + Unterricht	FES

Abb.3: Crosskonkordanzen (2) aus dem *KoMoHe*-Projekt

Dabei bedeuten:

- = Äquivalenzrelation: für Terme, die das gleiche Konzept bezeichnen
- < Oberbegriffsrelation: für Terme, die in einer Hierarchiebeziehung stehen (Teil-Ganzes, Abstraktion)
- > Unterbegriffsrelation: wie Oberbegriffsrelation
- ^ Ähnlichkeitsrelation: für Terme, die ähnliche oder verwandte Konzepte bezeichnen
- 0 keine Entsprechung im Zielthesaurus

h, m, g - hoch, mittel, gering

Ein weiteres Beispiel dieser Vorgehensweise, das noch stärker an der Verknüpfung verschieden-sprachiger Vokabularien orientiert ist und damit eine Verbindung zwischen Interoperabilität und Multilingualität herstellen will, findet sich im *MACS*-Projekt, das Verknüpfungen zwischen der *Schlagwortnormdatei (SWD)*, den *Library of Congress Subject Headings (LCSH)* und dem *Répertoire d'Autorité-Matière Encyclopédique et Alphabétique Unifié (Rameau)* herstellt<sup>7</sup> (vgl. Abb.4).

<sup>7</sup> Vgl.: **Landry, P.:** *MACS: multilingual access to subject and link management: Extending the Multilingual Capacity of TEL in the EDL Project.*  
 In: <http://www.edlproject.eu/workshop/programme.php>.  
 Vgl. auch: <http://macs.cenl.org>.

**macs**  
multilingual access to subjects

**Search Results**  
FHKöln2/reader/Work

Readers  
Annotators  
Search  
Links  
List Terms  
Help  
Logout

Searched for **theater** in SWD.  
Search retrieved 48 links in 0.68 seconds.

RAMEAU	LCSH	SWD	Domains	MAC'S Link	Dups
<a href="#">Théâtre AND Bibliographie</a>	<a href="#">Theater -- Bibliography</a>	<a href="#">Theater AND Bibliographie</a>	790	<a href="#">MACS0000194</a>	
<a href="#">Théâtre -- Objets de collection</a>	<a href="#">Theatrical paraphernalia</a>	<a href="#">Theater AND Sammlung</a>	790	<a href="#">MACS0000265</a>	
<a href="#">Théâtre AND Prix et récompenses</a>	<a href="#">Theater -- Awards</a>	<a href="#">Theater AND Kulturpreis</a>	790	<a href="#">MACS0000193</a>	
<a href="#">Théâtre AND Aspect moral</a>	<a href="#">Theater -- Moral and ethical aspects</a>	<a href="#">Theater AND Moral</a>	--	<a href="#">MACS0000209</a>	
<a href="#">Théâtre AND Expositions</a>	<a href="#">Theater -- Exhibitions</a>	<a href="#">Theater AND Ausstellung</a>	790	<a href="#">MACS0000199</a>	
<a href="#">Théâtre AND Biographies</a>	<a href="#">Theater -- Biography</a>	<a href="#">Theater AND Biographie</a>	790	<a href="#">MACS0000195</a>	
<a href="#">Théâtre -- Comptes rendus</a>	<a href="#">Theater reviews</a>	<a href="#">Theater AND Aufsatzsammlung</a>	790	<a href="#">MACS0000219</a>	
<a href="#">Théâtre -- Moyen Age</a>	<a href="#">Theater -- History -- Medieval, 500-1500</a>	<a href="#">Theater AND Mittelalter</a>	790	<a href="#">MACS0000201</a>	
<a href="#">Théâtre -- Programmes</a>	<a href="#">Theater programs</a>	<a href="#">Theater AND Programmheft</a>	--	<a href="#">MACS0000239</a>	
<a href="#">Théâtre -- Publics</a>	<a href="#">Theater audiences</a>	<a href="#">Theater AND Publikum</a>	--	<a href="#">MACS0000233</a>	

1/5 Next Page

Abb.4: Crosskonkordanzen aus dem MACS-Projekt

Die häufige Verwendung des Booleschen AND bei diesen Verknüpfungen ist ein Indiz für die strukturellen Unterschiede der berücksichtigten Vokabularien: auf der einen Seite die im Sachschlagwortbereich stärker an Einzelbegrifflichkeiten orientierte *SWD* – auf der anderen Seite die mit starken Präkombinationsstrukturen versehenen *LCSH*.

Ein weiterer Aspekt verdient Beachtung: Für die dauerhafte Tauglichkeit einmal erstellter Konkordanzen ist eine datentechnische Rahmenbedingung wichtig: Werden die Konkordanzrelationen allein auf der Ebene der Klartextbezeichnungen (etwa in externen Dateien) vorgenommen oder unter Berücksichtigung der jeweiligen Datensatz-Identnummern? Soll der Konkordanzprozess spätere Änderungen in den jeweiligen Vokabularien automatisch integrieren können, ist allein der zweite Weg sinnvoll. Dieselbe Problematik zeigt sich im Kontext des Semantic Web, wenn semantische Entitäten mitsamt ihres Strukturkontextes über stabile Adressen wie URIs anderen Web-Anwendungen zur Verfügung stehen sollen. Im besten Fall würden alle beteiligten Dateien in einem Editionssystem gepflegt, was aber gleichzeitig die Vorstellung einer supranationalen multilingualen Superdatei aufscheinen lässt und damit fiktional wird. Der Vorstellung einer dynamischen Verbindung mehrerer Vokabularien werden damit enge Grenzen gesetzt, die allein schon die Frage nach alternativen Vorgehensweisen aufwerfen.

#### 4.2 Citation Pearl Growing

Ein weiterer Vorschlag zur Heterogenitätsbehandlung wurde durch eine Methode eingebracht, die den Namen *Citation Pearl Growing* bekommen hat<sup>8</sup>. Ausgangspunkt ist das zunehmende

<sup>8</sup> Vgl. z.B.: **Riesthuis, G.J.A.:** *Subject searching in merged catalogues: a plea for redundancy.*

parallele Vorhandensein von Daten verschiedener Erschließungssprachen an Titeldaten und der Wunsch, diese Daten gemeinsam für Recherchen nutzen zu können. Es geht dabei also nicht um die Feststellung einer kontextfreien semantischen Vergleichbarkeit der Elemente von Dokumentationssprachen, sondern der Interoperabilitätsgedanke wird durch das gemeinsame Vorhandensein von Erschließungsdaten an einem Titeldatensatz eingebracht. Um sich das Verfahren vorzustellen, betrachte man den Datensatz der folgenden Abb.5<sup>9</sup>.

<b>Handbook on ontologies.</b> Staab, S. u. R. Studer (Hrsg.). Berlin: Springer, 2004. XV, 660 S. ISBN 3-540-40834-7. (International handbook on information systems)	
<b>DDC:</b>	<a href="#">003/.54</a> -- dc22
<b>LCC:</b>	Q387.H36 2004
<b>RVK:</b>	<a href="#">ST 300 Informatik/Monographien/Künstliche Intelligenz/Allgemeines</a> <a href="#">QP 345 Wirtschaftswissenschaften/Allgemeine Betriebswirtschaftslehre/ Unternehmensführung/Organisation/Informationswesen. Informationssysteme</a> <a href="#">ST 302 Informatik/Monographien/Künstliche Intelligenz/Expertensysteme; Wissensbasierte Systeme</a> ST 265
<b>BK:</b>	<a href="#">54.72 ; Künstliche Intelligenz</a> <a href="#">54.62 ; Datenstrukturen</a>
<b>GHBS:</b>	<a href="#">TVU</a> <a href="#">TWY</a> TVUR <a href="#">TVB</a> <a href="#">TZH</a> <a href="#">TYP</a> TXV
<b>RSWK:</b>	Informationssystem ; Wissenstechnik ; Ontologie <Wissensverarbeitung> ; Aufsatzsammlung Semantic Web ; Wissensbasiertes System ; Ontologie <Wissensverarbeitung> ; Aufsatzsammlung Informationssystem ; Konzeptionelle Modellierung ; Ontologie <Wissensverarbeitung> ; Aufsatzsammlung
<b>RSWK Einzel-SW:</b>	<a href="#">Informationssystem</a> <a href="#">Wissenstechnik</a> <a href="#">Ontologie &lt;Wissensverarbeitung&gt;</a> <a href="#">Aufsatzsammlung</a> <a href="#">Semantic Web</a> <a href="#">Wissensbasiertes System</a> Konzeptionelle Modellierung
<b>LCSH:</b>	<a href="#">Knowledge representation (Information theory)</a> Conceptual structures (Information theory) <a href="#">Ontology</a> <a href="#">Database design</a> <a href="#">Expert systems (Computer science)</a>

Abb.5: Datensatz mit Erschließungsdaten unterschiedlicher Systeme

In: New perspectives on subject indexing and classification: essays in honour of Magda Heiner-Freiling. Red.: K. Knull-Schlomann, u.a. Leipzig: Deutsche Nationalbibliothek 2008. S.257-260.

**Riesthuis, G.J.A., A.-M. Colenbrander-Dijkman:** *Subject access to central catalogues: incompatibility issues of library classification systems and subject headings in subject cataloguing.*

In: Die Klassifikation und ihr Umfeld: Proc. 10. Jahrestagung der Gesellschaft für Klassifikation, Münster, 18.-21.6.1986. Hrsg.: P.O. Degens. Frankfurt: Indeks 1986. S.112-141. (Studien zur Klassifikation; Bd.17)

<sup>9</sup> Die für die Diskussion verwendeten Daten wurden über den KVK recherchiert, den entsprechenden bibliografischen Daten zugeordnet und in einer eigenen Datenbank mit der Software *MIDOS*, die über eine Verlinkungs-Funktionalität verfügt, suchbar gemacht.

Die Idee des *Citation Pearl Growing* unterstützt nun den Wechsel zwischen den Erschließungsdaten verschiedener Systeme und will damit nicht allein die Möglichkeiten des Sucheinstiegs erweitern und Möglichkeiten zur Suchfortsetzung über Hyperlinks anbieten, sondern es soll auch eine Lösung für den Umstand angeboten werden, dass nicht alle Datensätze mit Daten aller Erschließungssysteme ausgestattet sind. Im angegebenen Beispiel sieht man dies bei den Schlagwörtern *Ontologie <Wissensverarbeitung>* der *RSWK* und *Ontology* der *LCSH*. Die jeweils zugehörigen Treffermengen lauten:

#### Ontologie <Wissensverarbeitung>

1. **Stuckenschmidt, H.:** *Ontologien: Konzepte, Technologien und Anwendungen*. Berlin: Springer, 2009. 160 S.
2. *Handbook on ontologies*. Staab, S. u. R. Studer (Hrsg.). Berlin: Springer, 2004. XV, 660 S.
3. *Semantic Web: 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, November 11-15, 2007*: proceedings. Eds.: K. Aberer et al. Berlin: Springer, 2007. XXVII, 973 S.
4. **Harmelen, F. van; Stuckenschmidt, H.:** *Information sharing on the semantic web*. Berlin: Springer, 2005. XIX, 276 S.

#### Ontology

1. *Towards the semantic Web: ontology-driven knowledge management*. Eds.: J. Davies u.a. Chichester: Wiley, 2003. XX, 288 S.
2. **Raskin, V.; Nirenburg, S.:** *Ontological semantics*. Cambridge, MA: MIT Press, 2004. XXI, 420 S.
3. *Handbook on ontologies*. Staab, S. u. R. Studer (Hrsg.). Berlin: Springer, 2004. XV, 660 S.

An dieser Stelle kann die Grundsatzfrage der Relevanz der einzelnen Titel für eine konkrete Fragestellung nicht vertieft werden<sup>10</sup>, eine sinnvolle Erweiterung der Treffer kann man aber feststellen. Fraglich ist, ob die beiden SW als semantisch interoperabel anzusehen sind, eine Entscheidung, die bei der weitgehend kontextfreien Gestaltung sowohl der *SWD* als auch der *LCSH* nur schwer zu entscheiden ist. Hier ergibt sich der Zusammenhang auf der Basis der gemeinsamen Zuteilung zu einem Dokument, also einem Aspekt der dokumentbezogenen Aboutness und nicht allein eines kontextfreien Vergleichs von Begriffsinhalten.

Dass die Methode bei einfacher Anwendung ihre Grenzen haben wird, sieht man sehr schnell an der dem Ausgangsbuch ebenfalls zugeteilten Notation 54.72 ; Künstliche Intelligenz der Basisklassifikation. In der Treffermenge dieser Notation liegen auch Titel wie

**Cleve, J.; Lämmel, U.:** *Künstliche Intelligenz*: mit 50 Tabellen, 43 Beispielen, 208 Aufgaben, 89 Kontrollfragen und Referatsthemen. 3., neu bearb. Aufl. München: Hanser, 2008. 348 S.

Dieser Titel besitzt eine viel allgemeinere Aboutness als der Ausgangstitel. Der Charakter des jeweiligen Erschließungssystems kann natürlich nicht außer Acht bleiben, wenn man es in die Bildung von Treffermengen mit der Methode des *Citation Pearl Growing* einbinden will. Je stärker die Erschließung auf der Basis koextensiver Repräsentationen erfolgt, desto günstiger wird die Ausgangssituation sein.

Man sieht dies deutlich am Beispiel der Notation TWY Management-Informationssysteme (MIS), Datenbanken der *GHBS*-Systematik, deren Treffermenge (15 Titel) u.a. auch die folgenden Dokumente enthält:

---

<sup>10</sup> Ebenso wenig sollen hier Überlegungen angestellt werden, ob die Zuteilung der inhaltskennzeichnenden Merkmale möglicherweise auf Fehlentscheidungen beruhen, die Indexierung also ggf. fehlerhaft ist.

1. **Kammer, M.:** *Literarische Datenbanken: Anwendungen der Datenbanktechnologie in der Literaturwissenschaft*. München: Fink, 1996. 259 S.
2. **Dadam, P.:** *Verteilte Datenbanken und Client/Server-Systeme: Grundlagen, Konzepte und Realisierungsformen*. Berlin: Springer, 1996. XII,416 S.
3. *Handbuch Internet-Suchmaschinen: Nutzerorientierung in Wissenschaft und Praxis*. Lewandowski, D. (Hrsg.). Heidelberg: Akademische Verlagsgesellschaft AKA, 2009. VIII, 409 S.

Interessant ist in diesem Zusammenhang die Verbindung zur *DDC*, die mit ihren Regeln zur Indexierung stärker den Gesichtspunkt der Koextensivität beachtet. In ihrer Klassenstruktur besitzt die *DDC* auf der Basis des ebenfalls beachteten Prinzips des *literary warrant* aber deutliche Eigenschaften, die eher auf ein Sammelsurium schließen lassen, als dass die Trennschärfe vorhanden wäre, die bei strikter Beachtung klassifikatorischer Merkmale zur Klassenbildung erwartet werden dürfte. Hier lautet die entsprechende Treffermenge für die zugeordnete Notation 003/.54 Informationstheorie

1. **Ritchie, L.D.:** *Information*. Newbury Park, Calif.: Sage Publ., 1991. VIII,75 S.
2. *The mathematical theory of information*. Kähre, J. (Hrsg.). Berlin: Springer US, 2002. 520 S.
3. **Völz, H.:** *Information verstehen: Facetten eines neuen Zugangs zur Welt*. Wiesbaden: Vieweg, 1994. X,184 S.
4. **Hodgson, J.P.E.:** *Knowledge representation and language in AI*. Chichester, UK: Horwood, 1991. 230 S.
5. *Handbook on ontologies*. Staab, S. u. R. Studer (Hrsg.). Berlin: Springer, 2004. XV, 660 S.

Auch hier ist eine inhaltliche Erweiterung des Ausgangskontextes erkennbar, wobei aufgrund von Umfang und Zustandekommen der benutzten Datenbank keine verbindlichen Aussagen über den repräsentativen Charakter der Ergebnisse gemacht werden können; dies muss weiteren Untersuchungen vorbehalten bleiben.

Die Entscheidung über inhaltliche Ähnlichkeit fällt bei der Vorgehensweise des *Citation Pearl Growing* nicht auf der Basis von Überlegungen hinsichtlich Gleichheit oder Verschiedenheit der Inhaltselemente als Bestandteile der Dokumentationssprachen, sondern auf der Basis von Indexierungsentscheidungen verschiedener Personen und in unterschiedlichen Institutionen mit eigener Zweckbindung. Ein Einfluss auf die Konsistenz muss hierbei angenommen werden. Es wäre eine interessante Frage für Untersuchungen, ob geeignete Ranking-Verfahren in eine gesamte Treffermenge eine sinnvolle Ordnung und Abfolge bringen können. Das gerne gegen den Einsatz von Ranking-Verfahren für bibliografische Daten ins Feld geführte Argument der Armut an beschreibenden Wörtern und Texten ist im vorliegenden Fall durch die Vielzahl der vorhandenen Erschließungsdaten entkräftet.

#### 4.3 Semantische Brücken zu einem Kernsystem

Bei der Diskussion von möglichen Alternativen zu vorstehenden Vorgehensweisen kann insbesondere der Aspekt der Multilingualität einen wesentlichen Gesichtspunkt beitragen, da vielfach mit einer anderen Sprache auch eine mehr oder weniger stark veränderte Konzeptualisierung von Objekten und Sachverhalten verbunden ist, die von Faktoren gespeist wird, denen noch nachzugehen sein wird. Analysiert man in den einzelnen Dokumentationssprachen stärker, welche Rückschlüsse aus den zwischen den Begriffen hergestellten Beziehungen auf die Bedeutung der Begriffe gezogen werden können, so stellt man deutliche Abhängigkeiten zur Herkunft und Zweckbindung der Dokumentationssprache fest, die nicht nur die Auswahl der Begriffe beeinflusst, sondern insbesondere die Art und Anzahl der zwischen ihnen hergestellten bzw. auch nicht hergestellten Beziehungen. Insbesondere zeigt sich deutlich eine Wechselwirkung zwischen dem jeweiligen Beziehungsgefüge und dem Verständnis der Begriffe. Eine kontextfreie Interpretation wird hiermit deutlich in Frage gestellt.

Betrachten wir als ein Beispiel das Thema Gesetzgebung in seiner Strukturierung in der *SWD* und den *LCSH*<sup>11</sup> (Abb.6).

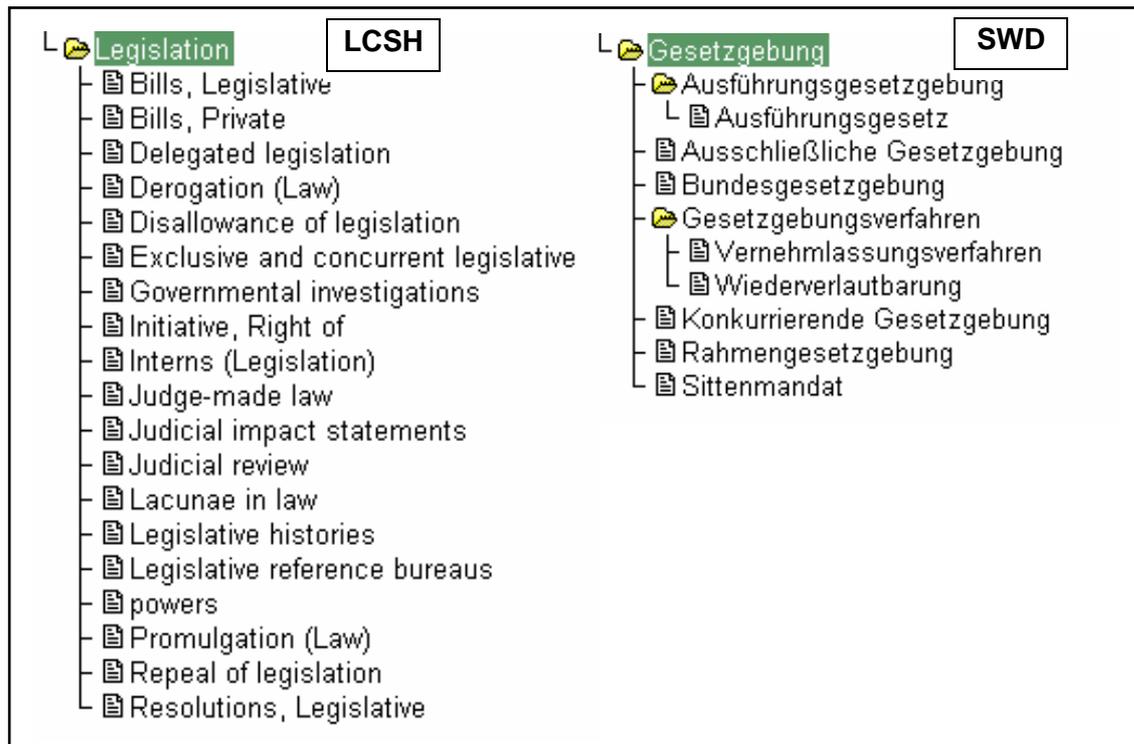


Abb.6: Das Thema *Gesetzgebung* in seiner Strukturierung in der *SWD* und den *LCSH* (1)

Lässt sich eine Feststellung treffen, ob *Legislation* und *Gesetzgebung* als semantisch interoperabel anzusehen sind? Die sprachliche Ebene allein – also z.B. eine Übersetzung – kann für eine Antwort nicht ausreichend sein, damit würde der gesamte normierende Ansatz einer Dokumentationssprache ausgeklammert. Da weder die *SWD* noch die *LCSH* umfassend darüber Auskunft geben, welche quasi-synonymen Begriffe für das Verständnis von *Legislation* bzw. *Gesetzgebung* mitgedacht werden sollen, bleibt allein die Analyse der weiteren Relationen, die zu den beiden Begriffen ausgewiesen werden (Abb.7).

<sup>11</sup> Die Visualisierung der Zusammenhänge erfolgt unter Zuhilfenahme der Software *MIDOS\_Thesaurus*; vgl.: <http://www.progris.de>.



- Historische Entwicklungen und Zusammenhänge
- Ethnische Themen
- Religiöse Themen
- Juristische Themen
- Nationale Organisationsformen
- Politische Strukturen
- Erziehungs- und Bildungssystem
- Alltagskulturelle Themen (Sport, Haushalt, Hobby, Brauchtum, ...)
- Fauna und Flora

Die hier vertretene These ist, dass derartige Sachverhalte auch jetzt schon in den jeweiligen Dokumentationsprachen spezifisch – wenn auch meist ohne gestalterische Absicht – ausgewiesen sind, dass sie es wert sind, erhalten zu werden und nicht einem Crosswalk zum Opfer fallen, sondern vielmehr für ein spezifisches Retrieval zur Verfügung gestellt werden sollten. Dementsprechend wäre nach einem Weg zu suchen, der semantische Interoperabilität über multilinguale, multinationale und multidisziplinäre Dokumentationsprachen hinweg zu gestalten gestattet.

Ein solcher Vorschlag wurde bereits im Zusammenhang mit der Erstellung der deutschen Ausgabe der *Dewey Decimal Classification* eingebracht.<sup>13</sup> Dieser Vorschlag basierte auf der Feststellung, dass eine Klassifikation mit jeder Übersetzung in andere Sprachen eine jeweils neue Sichtweise gegenüber den vorherigen Ausgaben eröffnet: Je mehr bei der Strukturierung des Stamm-Systems darauf geachtet wird, nur „universale“ Strukturen zu berücksichtigen, desto mehr können bei Erweiterung der Klassenstruktur in den übersetzten Systemen oder bei der Gestaltung des Zugangsvokabulars Gesichtspunkte einer Lokalisierung eingebracht werden. Ergebnis wäre ein entlokalisierendes universales Kernsystem mit einem Kranz lokalisierter Systeme, die nicht allein Übersetzung des Kernsystems sind, sondern in ihrer Struktur das Lokalisierungsgebiet der jeweiligen Sprache berücksichtigen. Ein Modell dieser Vorstellung wird durch die Grafik der Abb.8 gegeben.

---

<sup>13</sup> **W. Gödert, M. Preuss:** *Anforderungen an ein Klassifikationssystem in einer globalisierten Welt.* Vortrag anlässlich des DDC Workshops in Frankfurt a.M., 20. April 2005. Folien der Präsentation unter: <http://www.ddc-deutsch.de/publikationen/pdf/workshop2005-goedert-preuss.pdf>.

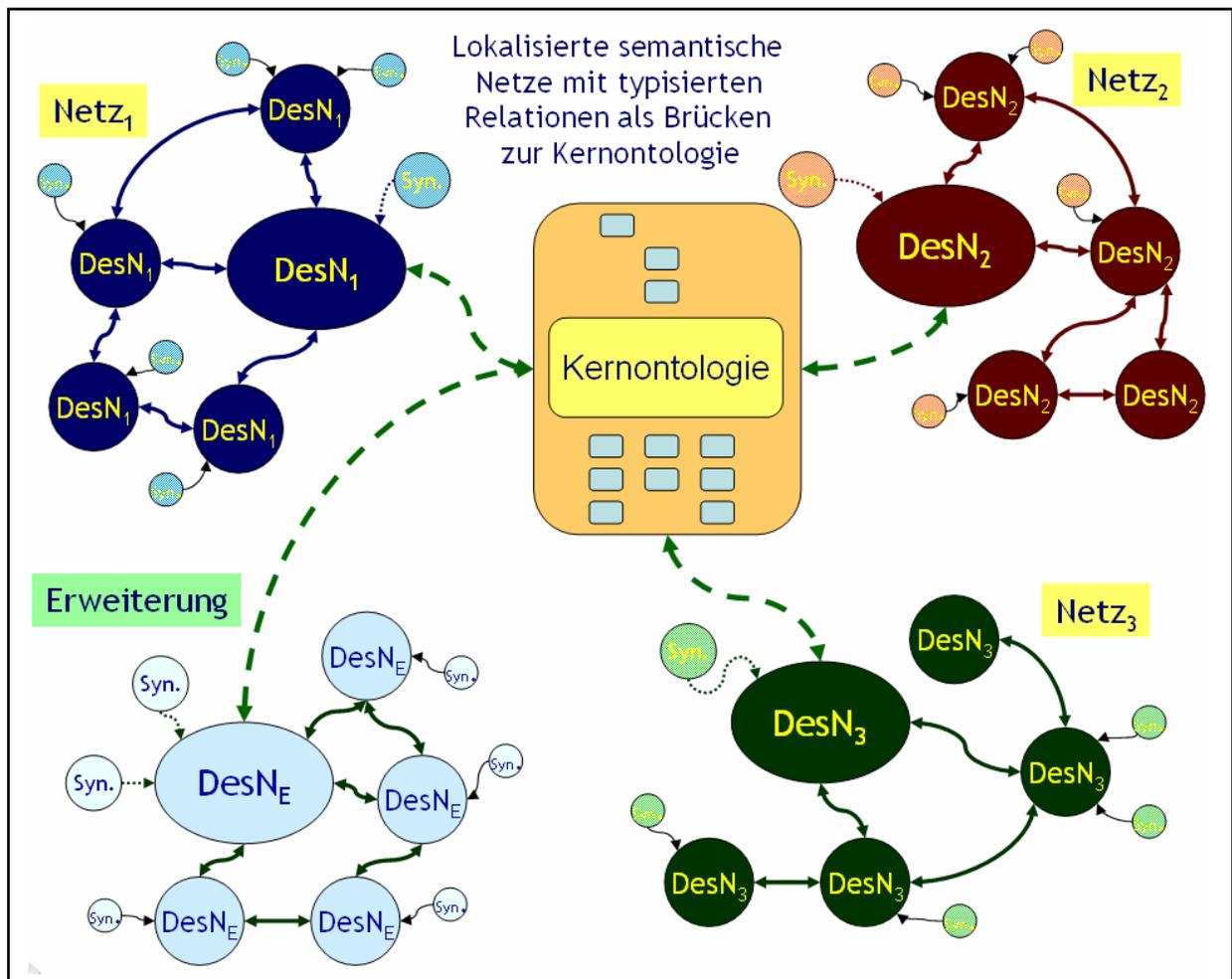


Abb.8: Kernontologie mit lokalisierten semantischen Netzen

Bei entsprechender Gestaltung der beteiligten Dokumentationssprachen ist dieses Modell durch folgende Eigenschaften gekennzeichnet:

- Die Inhaltseinstitäten und die zwischen ihnen ausgewiesenen Relationen lassen sich sowohl kognitiv als auch maschinell interpretieren;
- Die zwischen den Inhaltseinstitäten ausgewiesenen Relationen erfüllen formal-logische Anforderungen, so dass Inferenzen entlang der Kanten des Netzes möglich sind;
- Die Kernontologie kann als Einstieg in Themen eines Netzes mit nicht vertrauter Lokalisierung und Sprache dienen, wenn man mit dem Netz der vertrauten Sprache und Lokalisierung gestartet ist;
- Lokalisierungsgesichtspunkte sind vollständig in den jeweiligen Netzen abgebildet;
- Die Terme der Netze haben die Funktion eines Registervokabulars für die Klassen und die Struktur der Kernontologie;
- Weitere Netze (auch in weiteren Sprachen) lassen sich problemlos mit der Kernontologie verbinden, ohne dass es Rückwirkungen auf vorhandene Netze oder deren Verbindung mit der Kernontologie gibt.

In diesem Modell wird die Zielsetzung aufgegeben, Inhaltseinstitäten der beteiligten Netze untereinander oder die Verbindung der Netz-Entitäten zur Kernontologie als konkordant anzusehen. Die Verbindung einer Netz-Entität zu einer Entität der Kernontologie soll vielmehr als eine Art Brückenbildung gesehen werden, die einem Nützlichkeitsgedanken für die Gestaltung von Retrievalumgebungen folgt. Erkenntnisse aus den Überlegungen zur semantischen Interoperabilität, die im Teil 2 dieses Beitrages angegeben werden, legen es nahe, die Verbindungen zwischen den Inhaltseinstitäten der Netze und den Knoten der Ontologie durch Interoperabilitätstypen zu kennzeichnen, die ebenfalls wieder Nutzen für Retrievalgestaltungen haben können. Die Arbeiten am *CrissCross*-Projekt mit einer feineren Differenzierung der Verbindungen zwischen *SWD*-Schlagwörtern und *DDC*-Notationen (sog. Determiniertheitsgrade) haben bereits gezeigt, welcher Nutzen daraus ggf. im Retrieval – etwa durch die Gestaltung von Ranking-Verfahren – gezogen werden kann<sup>14</sup>. Nachstehend wird dieser Aspekt noch einmal aufgegriffen, wenn weitere positive Eigenschaften des Modells für die Gestaltung von Retrievalumgebungen diskutiert werden.

## 5. Exkurs: Auswahl der Kernontologie

An dieser Stelle taucht die schon viel diskutierte Frage auf, welches (Klassifikations-) System die beste Eignung besitzt, als Kernontologie verwendet zu werden. Es liegt zunächst nahe, diese Frage anhand der strukturellen Eigenschaften zu entscheiden, die auf der Basis der vorstehenden Diskussion gefordert werden müssen<sup>15</sup>:

1. Klare begriffliche Strukturierung, in der insbesondere die Hierarchiestränge über die Eigenschaft der Vererbbarkeit verfügen und somit keine präkombinierten Klassen aufweisen;
2. Vollständige Entlokalisierung aller repräsentierten Sachverhalte;
3. Gestaltung eines – ggf. mehrsprachigen – Zugangsvokabulars für die Kernontologie als semantisches Netz, das die nicht für die Hierarchiebildung zu nutzenden assoziativen Relationen in Form eines typisierten Relationeninventars nutzt.

Damit kommt allein das Konzept der Facettenklassifikation in Frage. Verfügbar ist ein derartiges System gegenwärtig nicht. Die Erstellung eines neuen System erfordert ein groß angelegtes Projekt; die Chancen auf Realisierung dürften unterschiedlich beurteilt werden, ausgeprägt groß sind sie wohl nicht. Insbesondere wenn man daran interessiert ist, in überschaubaren Zeiträumen auch Verbesserungen im Dokumentretrieval zu realisieren, wird man sich für ein System interessieren, mit dem bereits möglichst viele Dokumente erschlossen wurden, und zwar im internationalen Rahmen. Damit verkleinert sich die Zahl der in Frage kommenden Kandidaten unmittelbar und die *Dewey Decimal Classification* bietet sich als ernst zu nehmendes System an. Unterstützt wird diese Wahl durch folgende Faktoren:

1. Die *DDC* liegt inzwischen in einer Reihe von Übersetzungen vor<sup>16</sup>, so dass nicht allein der multilinguale Zugang über das jeweilige Registervokabular verbreitert wurde, sondern jeder Übersetzungsvorgang hat bereits Aspekte eingebracht, die den Gedanken der Lokalisierung

---

<sup>14</sup> Vgl. **Hubrich, J.:** *CrissCross: SWD-DDC-Mapping*.

In: *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare*. 61(2008) H.3, S.50-58.

<sup>15</sup> Kriterien wie Aktualität, Differenzierungsgrad und Klassenzahl, fachliche Angemessenheit können an dieser Stelle außer Acht bleiben, da sie nicht das funktionale Potenzial zur Herstellung von Retrievalumgebungen beeinflussen.

<sup>16</sup> Vgl.: <http://www.oclc.org/dewey/about/translations/default.htm>;

unterstützen, die bislang aus Gründen der Editions politik aber nicht voll ausgearbeitet werden konnten<sup>17</sup>;

2. Denkt man bei den Maßnahmen zur Heterogenitätsbehandlung an Suchvorgänge, die Unterstützung bei Konzeptualisierung und Wortwahl benötigen, so ist hierfür das Vorhandensein einer systematischen Struktur und Erschließung erforderlich, die den gesamten Dokumentbestand erfasst. Es erscheint bei gegenwärtiger Ressourcenlage ausgesprochen unrealistisch zu sein, diesen Zustand intellektuell herbeiführen zu wollen. Allein Methoden des automatischen Klassifizierens können hier ein Beitrag leisten. Aus methodischen Gründen ist es auch hierfür nützlich, über einen möglichst großen Bestand an bereits erschlossenen Dokumenten zu verfügen, um die Verfahren optimieren zu können. Die im *Colibri*-Projekt<sup>18</sup> gewonnenen Erfahrungen lassen nicht nur hoffen, dass die entwickelten Methoden zur Dekomposition synthetischer *DDC* einen wertvollen Beitrag zur Verbesserung der Retrieval-situation leisten werden, sondern dass das Ziel erreichbar ist, noch nicht erschlossene Bestände automatisch mit *DDC*-Notationen auszustatten, die wiederum nutzbringend für Retrievalvorgänge verwendet werden können.

## 6. Anreicherung der lokalisierten Semantischen Netze

Sein volles Potenzial kann das hier vorgestellte Modell erst entfalten, wenn die mit der Kernontologie verbundenen semantischen Netze entsprechend gestaltet werden. Bisherige Thesauri oder Schlagwortvokabulare sind durch eine schwach ausgeprägte Relationierung gekennzeichnet, sowohl was die Typenvielfalt als auch den Umfang der Relationen insgesamt betrifft. Der Kernbestand an Relationen, Synonymie, Ober- / Unterbegriffe, Verwandte Begriffe wird eher selten dadurch erweitert, dass die generische von der partitiven Relation differenziert wird, chronologische oder Instanz-Zusammenhänge sichtbar gemacht werden. Die Vielfalt der assoziativen Beziehungen wird fast nie spezifiziert. Dies schränkt in der Regel auch die Möglichkeit ein, Inferenzen entlang der Relationspfade zu ziehen, da die entsprechenden logischen Eigenschaften nicht im Vokabular enthalten sind. Inferenzen zu ziehen, setzt meist die Analyse des Einzelfalls voraus, was eine maschinelle Prozedur unmöglich macht.

Das Prinzip einer möglichen Vorgehensweise wird durch die Abb.9 am Beispiel einer facettierten Struktur dargestellt<sup>19</sup>. Innerhalb der Facetten wird auf der Basis der klassifikatorischen Merkmale eine Hierarchie gebildet. Die darüber hinaus als nützlich angesehenen assoziativen Relationen werden gemäß einer zu bildenden Typologie in den angeschlossenen semantischen Netzen ausgewiesen. Idealerweise sollten auch die formalen Eigenschaften (z.B. Eignung für Inferenzen) in der Modellierung berücksichtigt werden.

<sup>17</sup> Vgl. aber: **Mitchell, J.S.:** *Locality and universality in the DDC*.

In: [http://www.nb.admin.ch/slb/slb\\_professionnel/projektarbeit/00729/01615/01675/index.html?lang=de](http://www.nb.admin.ch/slb/slb_professionnel/projektarbeit/00729/01615/01675/index.html?lang=de). [Vortrag während der EDUG-Konferenz der European DDC users' group am 11.06.2007 in Bern].

<sup>18</sup> Vgl.: **Reiner, U.:** *Automatische DDC-Klassifizierung von bibliografischen Titeldatensätzen*.

In: [http://www.gbv.de/vgm/info/biblio/01VZG/06Publikationen/2009/pdf/pdf\\_3921.pdf](http://www.gbv.de/vgm/info/biblio/01VZG/06Publikationen/2009/pdf/pdf_3921.pdf).

[Präsentation zum Vortrag anlässlich des 98. Deutscher Bibliothekartag in Erfurt, 2009].

**Reiner, U.:** *DDC-based search in the data of the German National Bibliography*.

In: *New perspectives on subject indexing and classification: essays in honour of Magda Heiner-Freiling*. Red.: K. Knull-Schlomann, u.a. Leipzig: Deutsche Nationalbibliothek 2008. S.121-129.

<sup>19</sup> Für die Bedeutung der verwendeten Nomenklatur – z.B.  $K^1_3(m_1m_3)$  sei auf Teil 2, Abschnitt 12.1, insbes. Abb.29 dieses Beitrages verwiesen.

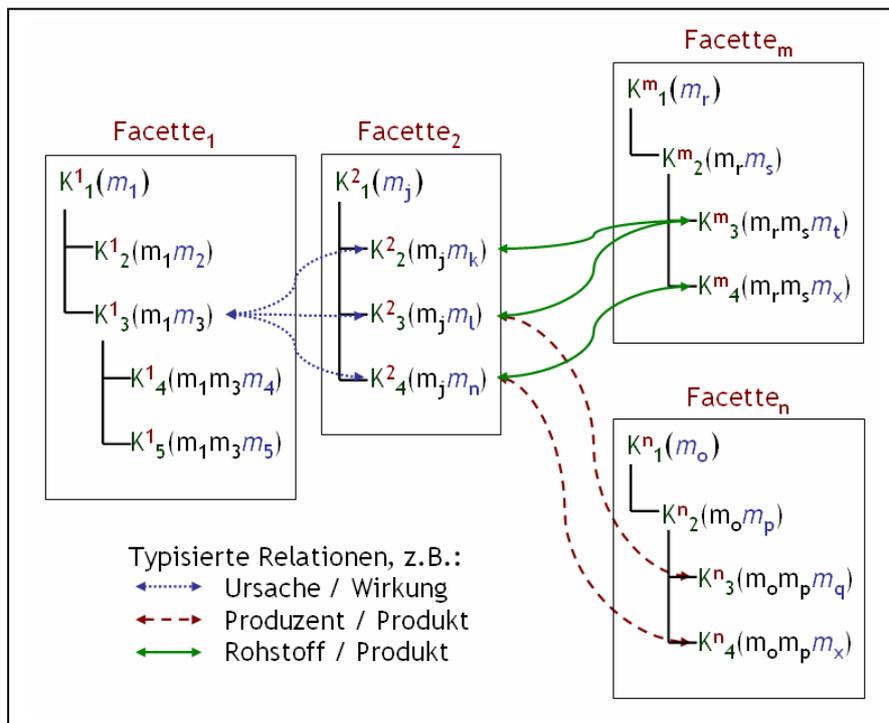


Abb.9: Typisierte Relationen innerhalb der systematischen Struktur

Mit diesen strukturellen Voraussetzungen lassen sich beispielsweise Fragen wie die folgenden beantworten:

1. Welche Produkte aus Facette<sub>2</sub>, die Spezialprodukte zu  $K^2_1$  sind, werden mit dem Rohstoff  $K^m_3$  aus Facette<sub>m</sub> hergestellt?
2. Welche davon werden vom Produzenten  $K^n_3$  aus Facette<sub>n</sub> hergestellt?

Die Antwort auf Frage 1 erfordert einen zweistufigen Prozess zur Ergebnisbildung. Zunächst müssen für die Ermittlung der Spezialprodukte von  $K^2_1$  alle Klassen ermittelt werden, die  $K^2_1$  hierarchisch untergeordnet sind. Im zweiten Schritt, müssen darunter die ermittelt werden, die in einer Rohstoff / Produkt – Beziehung zu Klassen der Facette<sub>m</sub> stehen. Als Ergebnis ergeben sich die Klassen  $\{K^2_2, K^2_3\}$ . Die Antwort auf die Frage 2 erfordert die Selektion der Klassen aus  $\{K^2_2, K^2_3\}$ , die in einer Produzent / Produkt – Relation zu Klassen der Facette<sub>n</sub> stehen. Als Ergebnis ergibt sich die Klasse  $K^2_3$ .

Welches Potenzial erweiterte Relationeninventare für Retrievalzwecke haben können, wurde bereits im Besucherinformationssystem der *EXPO 2000* demonstriert<sup>20</sup>. Im Rahmen einer Diplomarbeit an der Fachhochschule Köln hat *Felix Boteram* eine genauere Analyse bisheriger Ansätze zur Typisierung semantischer Relationen vorgenommen und daraus einen Vorschlag zur gestuften Gestaltung solcher Inventare abgeleitet<sup>21</sup>, der hier für eine Differenzierung der generischen Relation angewendet auf den Gegenstandsbereich *Theater* wiedergegeben werden soll (Abb.10)<sup>22</sup>.

<sup>20</sup> **Schmitz-Esser, W.:** *EXPO-INFO 2000: Visuelles Besucherinformationssystem für Weltausstellungen*. Berlin: Springer 2000.

<sup>21</sup> **Boteram, F.:** *Semantische Relationen in Dokumentationssprachen: vom Thesaurus zum semantischen Netz*. Köln: Fachhochschule, Institut für Informationswissenschaft 2008. (Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft; Bd. 54).

<sup>22</sup> Das vollständige Relationeninventar findet sich a.a.O., S.80-81.



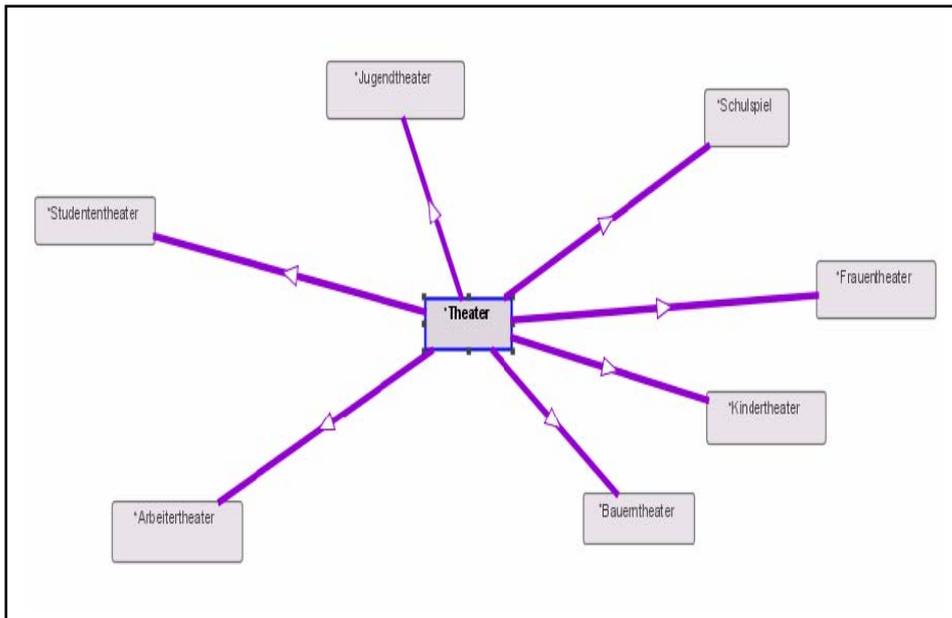


Abb.12: Reduktion des Netzes zu *Theater* auf alle *Unterbegriffen gemäß Genre*

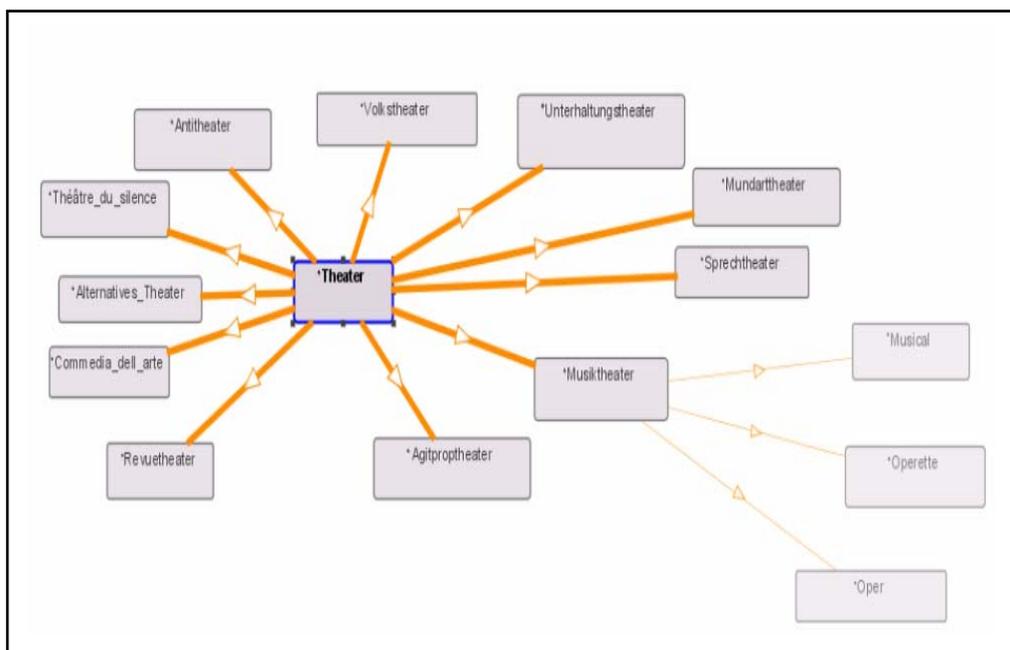


Abb.13: Reduktion des Netzes zu *Theater* auf alle *Unterbegriffen gemäß Darsteller*

Im Unterschied zur klassischen Vorgehensweise müssten die Begriffe der Abbildungen 12 und 13 nicht über eine intellektuelle Analyse einzeln aufgesucht werden, sondern sie lassen sich über einen maschinellen Selektionsprozess durch Auswahl des entsprechenden Relationstyps zur Verfügung stellen.

Die genaue Bestimmung der Relationstypen nach Art und Umfang sowie die Frage ihrer Berücksichtigung in der Kernontologie bzw. den angeschlossenen Begriffsnetzen bleibt zunächst ein noch näher zu untersuchender Gegenstand. Die angegebenen Beispiele dienen der Illustration und geben keine endgültigen Zustände wieder.

Neben dem hier wiedergegebenen Ansatz finden sich in der Literatur weitere Untersuchungen zur Entwicklung von Relationeninventaren, die in hinreichender Allgemeinheit gegenüber einzelnen Gegenstandsbereichen für die Gestaltung großer Dokumentationssprachen genutzt werden könnten<sup>23</sup>. Es stellt sich ja prinzipiell immer die Frage, welcher Weg der sinnvollere ist: eine Dokumentationssprache nach neuen Prinzipien neu zu konzipieren und zu erarbeiten oder eine vorhandene Dokumentationssprache gemäß veränderter Prinzipien umzuarbeiten. Angesichts des Umfangs, den ein Vokabular haben müsste, das zur fachlich differenzierten Erschließung eines disziplinär nicht begrenzten Dokumentenbestandes geeignet ist und das neben der Fachterminologie auch Umgangssprache enthält – die *SWD* als umfangreichstes deutsches Vokabular dieser Art verfügt über ca. 160.000 Sachbegriffe – lässt es als vollständig unrealistisch erscheinen, ein solches Vokabular mitsamt neuer differenzierter Relationierung neu zu erstellen. Erfahrungen aus dem *CrissCross*-Projekt lassen den Rückschluss zu, dass die Nachrelationierung der *SWD* – ggf. unter Beimischung anderer Vokabularien und unter Einsatz linguistischer Tools – zwar aufwendig wäre, aber erfolgreich zum Abschluss gebracht werden könnte.

## 7. Interoperabilität, maschinelle Interpretation und Semantic Web

Es wurde bereits betont, dass es ein wesentlicher Bestandteil des hier vorgestellten Konzepts ist, eine Verbindung zwischen kognitiver Interpretation der Inhaltselemente in Such- und Findprozessen und der formalen Repräsentation zur maschinellen Verarbeitung in Retrievalkontexten zu finden. Dabei spielt die intellektuelle Interpretierbarkeit des Vokabulars für den suchenden Nutzer die entscheidende Rolle, um in Such-, aber auch in Wissenserkundungsprozessen erfolgreich agieren zu können. Die hier diskutierten großen Normdateien zeichnen sich alle durch die Eigenschaft aus – anders als Fachthesauri oder gar spezielle Vokabulare aus dem Bereich der Wissensrepräsentation – auch allgemeinsprachliche Ausdrücke zu berücksichtigen. Dieser Vorzug sollte für den Einstieg in die Kernontologie erhalten bleiben. Nachteilig ist es jedoch, dass die Inhaltselemente in den Dateien meist nicht hinreichend charakterisiert sind (z.B. durch definitorische Hinweise), um ihren Begriffsinhalt und –umfang zweifelsfrei zu erschließen. Hier ist Nachbesserungsbedarf vorhanden.

Das Nebeneinander von Allgemein- und Fachsprache oder auch von wissenschaftlicher Taxonomie bzw. nutzenorientierter Subordination schafft dabei eine weitere grundsätzlich schwierige Situationen für die Relationierung, wie sich am Beispiel folgender Frage gut zeigen lässt: Welche Relationen weist man aus zwischen den Begriffen Pilze, Lamellenpilze, Röhrenpilze, Giftpilze und Speisepilze? In einem gewissen Sinn handelt es sich um einen Begriff mit vier Unterbegriffen. Genauere Betrachtung zeigt aber, dass es zwei verschiedene dichotome Kriterien gibt, die zur Ober- / Unterbegriffsbestimmung herangezogen werden können, so dass für

---

<sup>23</sup> Vgl.: **Tudhope, D., H. Alani u. C. Jones:** *Augmenting thesaurus relationships: possibilities for retrieval*. In: *Journal of digital information*. 1(2001) no.8. [<http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Tudhope/>].  
**ALA / Subcommittee on Subject Relationships/Reference Structures:** *Final Report to the ALCTS/CCS Subject Analysis Committee*. June 1997. Unter:  
<http://www.ala.org/ala/alctscontent/catalogingsection/catcommittees/subjectanalysis/subjectrelations/finalreport.htm>.  
Vgl. auch: **Green, R.:** *Relationships in knowledge organization*. In: *Knowledge organization*. 35(2008) nos.2/3, S.150-159.  
**Green, R.:** *Semantic types, classes, and instantiation*. In: *Knowledge organization for a global learning society: Proceedings of the 9th International ISKO Conference, 4-7 July 2006, Vienna, Austria*. Hrsg.: G. Budin, C. Swertz u. K. Mitgutsch. Würzburg: Ergon Verlag 2006. S.151-157. (*Advances in knowledge organization*; vol.10).

konkrete Pilzsorten (z.B. Pfifferling oder Steinpilz) meist zwei der vier möglichen Relationen (ist essbar, ist giftig, hat Lamellen, hat Röhren) zutreffen. Abb.14 zeigt die entsprechende Struktur.

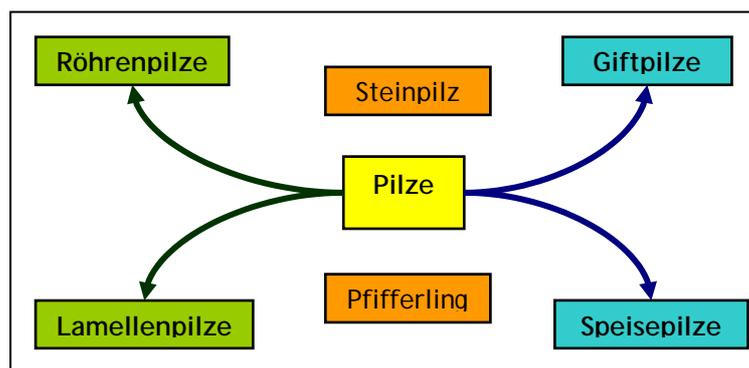


Abb.14: Dilemma der Relationierung umgangs- und fachsprachlicher Terme

Aus unserer Diskussion um ein erweitertes Relationeninventar geht hervor, dass diese im Allgemeinen multivalente Situation nicht wie häufig praktiziert durch die Zuweisung mehrerer Ober- / Unterbegriffsrelationen (Polyhierarchie) bearbeitet werden sollte, sondern entsprechend der ja eigentlich vorhandenen Aspektbindung durch spezielle Formen des Relationeninventars, um dann auch wieder maschinelle Verarbeitbarkeit zu gewährleisten<sup>24</sup>. Die Parallelität zwischen umgangs- und fachsprachlichen Termen im Vokabular kann so erhalten werden.

Für die maschinelle Interoperabilität ergeben sich zwei Gesichtspunkte. Der eine betrifft grundsätzlich die wünschenswerte logische Ausarbeitung der Beziehungsstruktur, um eine Basis für maschinelles Schlussfolgern zu haben. Dieser Gesichtspunkt wurde bereits diskutiert. Der andere Gesichtspunkt betrifft die Wege, wie die Vokabularien in Web-Umgebungen zur Verfügung gestellt werden können. Angesichts des Aufwandes zur Erstellung solcher Tools wird das Interesse an einer Nachnutzung vorhandener Dateien immer größer. Will man diesen Weg beschreiten, so sind besondere Repräsentationsanforderungen an die jeweiligen Inhaltseinstitäten zu beachten, z.B. im Rahmen des *RDF / OWL* Framework des Semantic Web. Die Realisierung dieser Wege wird für die *DDC*<sup>25</sup> oder für die *SWD*<sup>26</sup> unter Anlehnung an das *Simple Knowledge Organisation System (SKOS)*<sup>27</sup> bereits diskutiert und für Einzelanwendungen vorbereitet. Es ergäbe sich die attraktive Möglichkeit, die umfangreichen Vokabularien mit den Suchinterfaces anderer Webangebote zu verbinden. Es wäre allerdings zu wünschen, dass die im *SKOS*-Modell derzeit vorhandene Orientierung am klassischen Thesaurus-Konzept mit seiner eingeschränkten Zahl von Relationen noch einmal überdacht wird. Eine Erweiterung des Modells zu Berücksichtigung

<sup>24</sup> Näheres dazu in Teil 2 dieses Beitrages.

<sup>25</sup> **Panzer, M.:** *Cool URIs for the DDC: towards Web-scale accessibility of a large classification system.*  
In: Metadata for semantic and social applications : proceedings of the International Conference on Dublin Core and Metadata Applications, Berlin, 22 - 26 September 2008, DC 2008: Berlin, Germany / ed. by Jane Greenberg and Wolfgang Klas. Göttingen: Univ.-Verl. 2008. S.183-190.

<sup>26</sup> **Jahns, Y.:** *20 years SWD: German subject authority data prepared for the future.*  
In: <http://www.ifla2009satellitflorence.it/meeting2/program/assets/Jahns.pdf>.

**Svensson, L.G.:** *Unified access: a semantic Web based model for multilingual navigation in heterogeneous data sources.*

In: New perspectives on subject indexing and classification: essays in honour of Magda Heiner-Freiling. Red.: K. Knull-Schlomann, u.a. Leipzig: Deutsche Nationalbibliothek 2008. S.241-248.

<sup>27</sup> Vgl.: <http://www.w3.org/2004/02/skos/>.

**Voß, J.:** *Vom Social Tagging zum Semantic Tagging.*

In: Good tags - bad tags: Social Tagging in der Wissensorganisation. Hrsg.: B. Gaiser, u.a. Münster: Waxmann 2008. S.175-186.

(Medien in der Wissenschaft; Bd.47)

weiterer Relationstypen wäre auch hier im Hinblick auf Interoperabilität und für die Zwecke semantischer Exploration geboten.

## 8. Retrievalszenarios zur Wissensexploration

Eingangs waren die *Exploration des semantischen Umfeldes* und die *thematische Exploration* als Paradigmen für die Gestaltung von Systemen zur Wissensexploration diskutiert worden. Der sprachlich erweiterte Zugriff auf eine Kernontologie mit Übergang zu lokalisierten Wissensstrukturen bietet hierfür Realisierungsoptionen. In einem solchen – optional zu wählenden – Modus würde eine Sucheingabe nicht unmittelbar durch Ausgabe einer Treffermengen beantwortet. Die Wörter der Sucheingabe würden auf ein linguistisch normalisiertes Vokabular treffen und eine Weiterleitung in die Kernontologie bewirken. Mit dieser Kernontologie oder den verbundenen lokalisierten Netze wäre die Möglichkeit zu Wissensnavigation und –exploration gegeben. Erst danach würde die Treffermenge generiert, deren Sortierung durch die diskutierten Interoperabilitätstypen zwischen den Inhaltsentitäten als Ranking-Parameter beeinflusst würde. Abb.15 veranschaulicht diese Überlegungen:

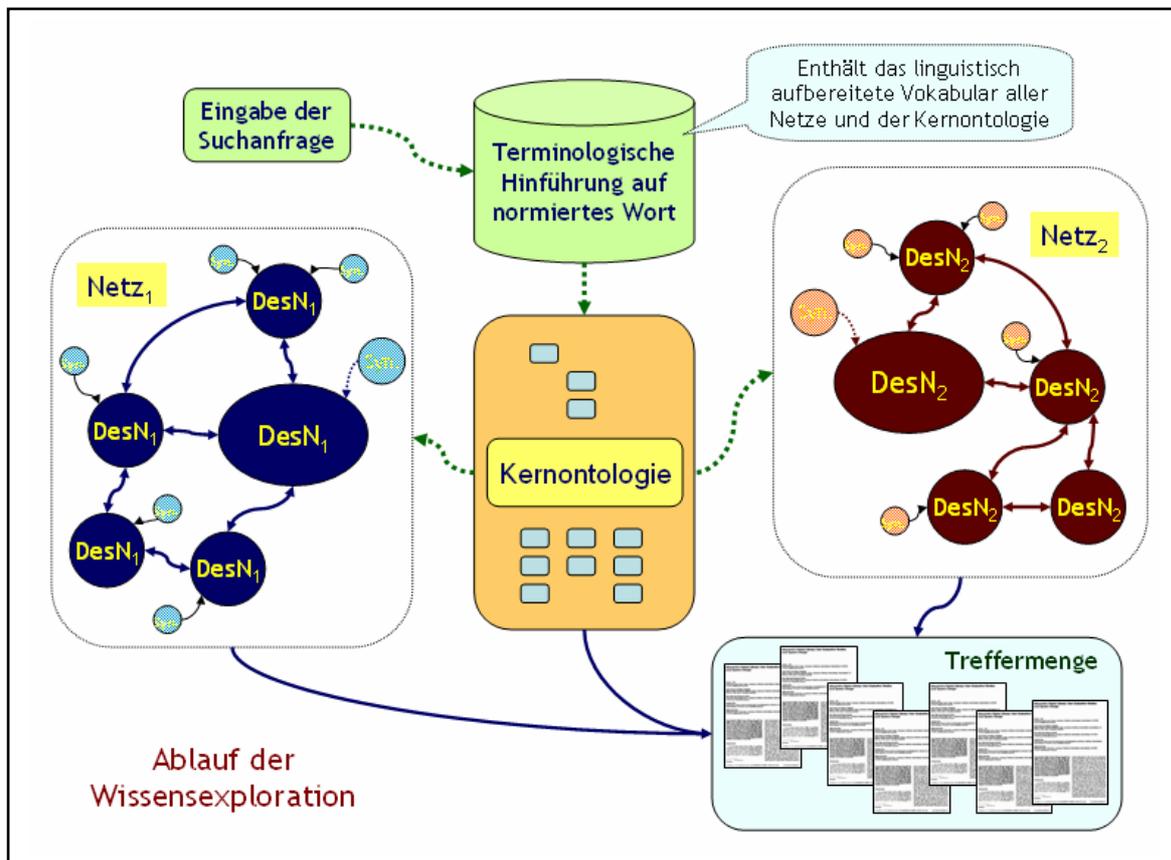


Abb.15: Schematisches Modell für Wissensexploration im Retrieval

Für unsere Diskussion spielt es dabei keine Rolle, ob die linguistische Normalisierung als Funktion an die Sucheingabe geknüpft wird, als Bestandteil des Ontologie- oder Netzvokabulars vorgehalten oder durch automatische Indexierung mit den Dokumenten verbunden wird. Wichtig ist allein, dass die Benutzung von nicht im normierten Vokabular enthaltenen Wortvarianten nicht den Rechercheerfolg gefährden. Ebenso wichtig ist der Gesichtspunkt, dass jede Form der linguistischen Bearbeitung sprachabhängig zu erfolgen hat, indem nicht allein unterschiedliche Wörterbücher sondern auch die zur jeweiligen Sprache passenden Verfahren eingesetzt werden.

Da es noch keine funktionstüchtige Retrievalumgebung gibt, die Wissensexploration im hier beschriebenen Sinn beinhaltet, kann zur Veranschaulichung der vorgestellten Ideen nur eine beispielhafte Demonstration gegeben werden, die im Rahmen des *CrissCross*-Projekts mit Daten der *DDC Deutsch* und der *SWD* entstanden ist. Für die Aussagekraft des Beispiels bleibt zu berücksichtigen, dass die relationale Reichhaltigkeit der *SWD* eng begrenzt ist.

Ausgangspunkt ist eine Datenbank mit ca. 330.000 Dokumenten und heterogener Erschließungssituation (alle Dokumente haben *DDC*-Notationen, der größte Teil *RSWK*-Schlagwortdaten). *CrissCross*-Mappings verbinden *DDC*-Notationen und *RSWK*-Daten. Die Verbindung dieser Mappings ist zunächst als unidirektional anzusehen: von *RSWK*-Daten als Einstiegsvokabular in die *DDC*-Struktur. Die für unsere Überlegungen ebenfalls interessante umgekehrte Richtung ist nicht Gegenstand der Projektarbeiten und bedürfte einer Überarbeitung der *SWD*-Struktur, um die im Rahmen dieses Beitrages skizzierten Potenziale voll zur Entfaltung zu bringen. Die genannten Rahmenbedingungen lassen die Simulation verschiedener Retrievalszenarios zu, die Elemente einer semantischen Exploration aufweisen<sup>28</sup>.

Betrachten wir zunächst ein Beispiel: Ausgangspunkt ist das Suchinteresse Gemüseanbau. Die direkte Worteingabe ergibt auf verschiedene Suchkategorien verteilt das Ergebnis:

Schlagwort	Titel-StW	Freitext
0	6	13

Nachschlagen in der semantischen Struktur zeigt, dass die zulässige Ansetzungsform des Schlagwortes in der *SWD* Gemüsebau lautet. Die Verwendung dieses Wortes ergibt folgende Treffermengen:

Schlagwort	Titel-StW	Freitext
11	14	31

Bezieht man alle zu Gemüsebau in der *SWD* relationierten Unterbegriffe mit in die Suche ein, so erhält man 20 Treffer in der Schlagwortkategorie.

Wechselt man über einen Registereintrag zur *DDC Deutsch*, so befindet man sich im systematischen Kontext, wie er in Abb.16 wiedergegeben wird:

<sup>28</sup> Eine erste Vorstellung dieser Ergebnisse erfolgte auf dem Bibliothekartag 2009 in Erfurt. Vgl.: **Boteram, F., W. Gödert u. J. Hubrich:** *Thematische Recherche und Interoperabilität: Wege zur Optimierung des Zugriffs auf heterogen erschlossene Bestände.* Unter: <http://www.opus-bayern.de/bib-info/volltexte/2009/782/>.

600	Technik, Medizin, angewandte Wissenschaften
630	Landwirtschaft
633-635	Einzelne Kulturpflanzen
<b>635</b>	<b>Gartenpflanzen (Gartenbau) ; Gemüse</b>
635.0284	Geräte, Ausstattung, Materialien
635.0285	Gartenbau—Computereinsatz
635.029	Gartenbau - Unternehmen - Verzeichnis
635.04	*Anbau, Ernte, verwandte Themen
635.05	Gartenbau—fortlaufende Sammelwerke
...	
635.1-635.8	Gartengemüse
635.9	Blumen und Zierpflanzen

Abb.16: Systematischer Kontext zu Gemüseanbau in *DDC Deutsch*

Eine Suche mit der Notation 635 ergibt 198 Treffer (wenn als alleinige Notation gesucht) und 351 Treffer (wenn als Teil einer synthetischer Notation gesucht wird). Über die Relevanz der erzielten Treffer ist damit natürlich noch nichts gesagt. In unserer Ausgangskonfiguration ist lediglich sichergestellt, dass die Suche alle Dokumente der Datenbank berücksichtigt hat und dass die Treffer in einem der Suchfrage entsprechenden Kontext liegen. Mit der Klasse verknüpft sind 12 Registerbegriffe und eine Reihe von Schlagwörtern der *SWD* gemäß folgender Interoperabilitätskennzeichnungen:

- 4 *SWD*-Schlagwörter mit hoher Relevanz
- 27 *SWD*-Schlagwörter mit mittlerer Relevanz
- 8 *SWD*-Schlagwörter mit geringer Relevanz

Schon diese Kennzeichnung gestattet 2 Optionen zur weiteren Prozessierung des Suchablaufes und der Ergebnisanzeige. Einmal könnte die Kennzeichnung als Relevanzcharakterisierung verstanden und zu einer sortierten Ergebnisanzeige verwendet werden. Abb.17 zeigt eine nach Schlagwörtern sortierte Ergebnisliste, die eigentliche Sortierung nach Relevanz kann wegen mangelnder Softwareunterstützung nicht gezeigt werden. Zum anderen könnte aus den zugeordneten Schlagwörtern eine Auswahl getroffen und ein selektives Ergebnis erzeugt werden (Abb.18).

1. Notation	Titel	1. SW (1.Kette)	2. Notation
635	Top-Ideen für Ihren Garten : kleine Sitzplätze für Genießer ; Gartenteiche	!04019311X!; s; Gartengestaltung	
635	Gärten zum Wohlfühlen : so gestalten Sie Ihre Wellness-Oase / [Red. und	!04019311X!; s; Gartengestaltung	
635	Sichtschutz im Garten / Tanja Ratsch	!04019311X!; s; Gartengestaltung	
635	Gartenprojekte : Terrassen, Wege, Wassergärten, Zäune und mehr ; [Sch	!04019311X!; s; Gartengestaltung	717
635	Wir pflanzen eine Laube : Bauen mit lebenden Gehölzen / Hermann Fritz	!040193128!; s; Gartenlaube	
635	Handbuch Samengärtnerei : Sorten erhalten, Vielfalt vermehren, Gemüse	!040200698!; s; Gemüse	
635	Vielfalt aus der Samentüte : säen - pflegen - staunen / Natalie Faßmann	!040200698!; s; Gemüse	
635	Alte Gemüsearten neu entdeckt : Schätze aus dem Bauergarten ; [über	!040200728!; s; Gemüsebau	641.65
641.5	Mein Gartenkochbuch : ernten, kochen und genießen im Jahreslauf / Julie	!040279960!; s; Jahreslauf	635
641.65	Gemüse-Garten-Kochbuch : mit 160 saisonalen Rezepten durch das Jahr	!040279960!; s; Jahreslauf	635
641.65	Gemüse-Garten-Kochbuch : mit 160 saisonalen Rezepten durch das Jahr	!040279960!; s; Jahreslauf	635
634	Veredeln : Obst- und Ziergehölze, Rosen und Kübelpflanzen / Peter Kloc	!040430588!; s; Obstgehölze	635
635	Kleine Gärten anlegen / Phil Clayton. [Fotos Mark Winwood. Übers. Wieb	!040734919!; s; Kleingarten	
635	Gärtnern leicht gemacht : Balkon - Terrasse - Mini-Garten ; perfekt planen	!040734919!; s; Kleingarten	

Abb.17: Sortierung der Ergebnismenge gemäß der mit der DDC-Klasse verbundenen Schlagwörter

Exotisches Gemüse:	1 Treffer
Feldgemüse:	1 Treffer
Gemüse:	36 Treffer
Gemüsebau:	11 Treffer
Gemüsegarten:	23 Treffer
Gemüseproduktion:	1 Treffer

Abb.18: Ergebnisse der veränderten Suchfragen durch Auswahl verknüpfter Schlagwörter

Verallgemeinert man die vorgestellten Ergebnisse, so lassen sich auf der Basis der angegebenen Ausgangssituation folgende Retrievalsszenarios angeben:

### *Szenario 1*

Die Schlagwörter und Synonyme der Netze dienen als erweitertes Registervokabular und ermöglichen den Wechsel von der verbalen Sucheingabe in die systematische Struktur der Ontologie (Abb.16 und Abb.19). Die Ergebnismengen bei Suchen über die Ontologie können nach verschiedenen Kriterien geordnet ausgegeben werden. Zuordnung mehrerer Notationen kann über Auswertung der Interoperabilitätsparameter für die jeweilige Zuordnung ebenfalls für das Ranking genutzt werden.

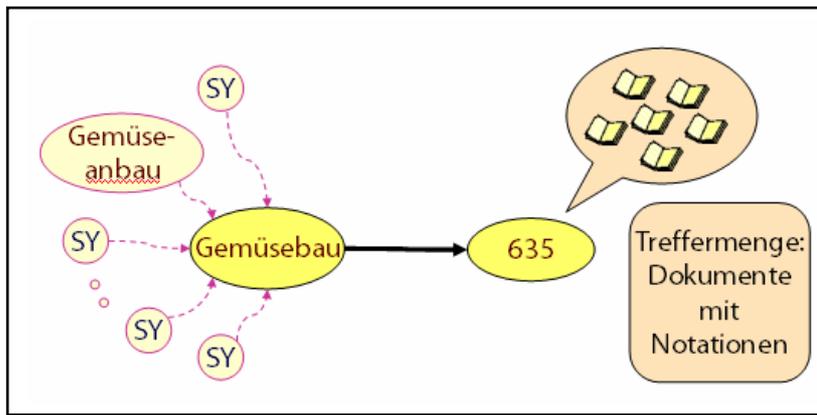


Abb.19: Das Netzvokabular dient als Registervokabular für die Ontologie

### Szenario 2

Die mit den Notationen der Ontologie verknüpften Schlagwörter werden ggf. unter Nutzung der zugeteilten Interoperabilitätskriterien oder der Struktur des Netzes für differenzierte Suchen und Bildung von Treffermengen genutzt (Abb.20).

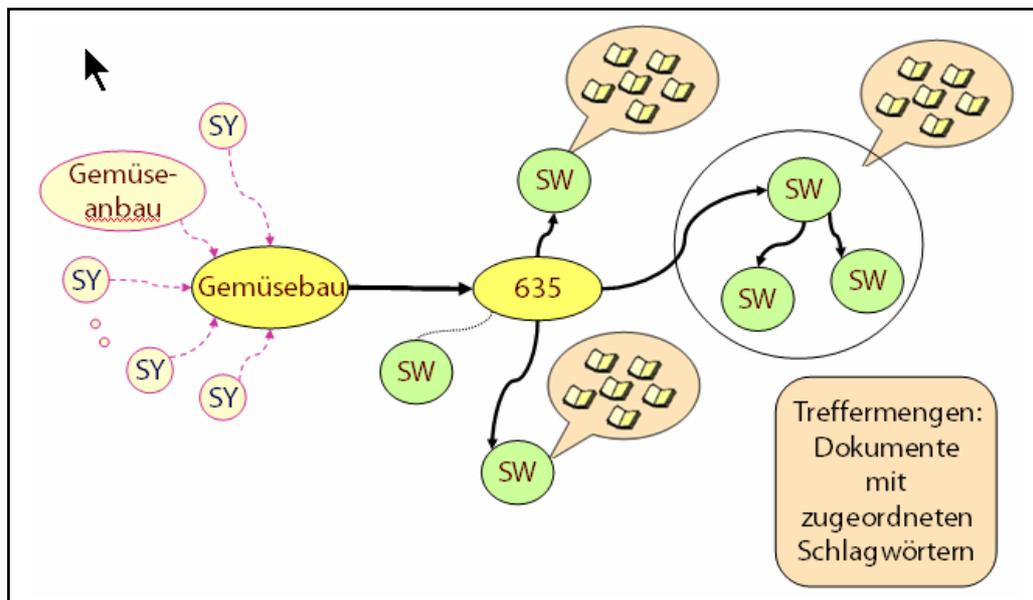


Abb.20: Übergang von der Ontologie zu den zugeordneten Schlagwörtern

### Szenario 3

In den lokalisierten Netzen sind differenzierte Relationierungen modelliert (Abb.21); es ist ein Übergang aus der Kernontologie in das entsprechende Netz möglich, es kann eine Auswahl der entsprechenden Relationen erfolgen und es können auf der Basis der zugehörigen Schlagwörtern Treffermengen gebildet werden (Abb.22).

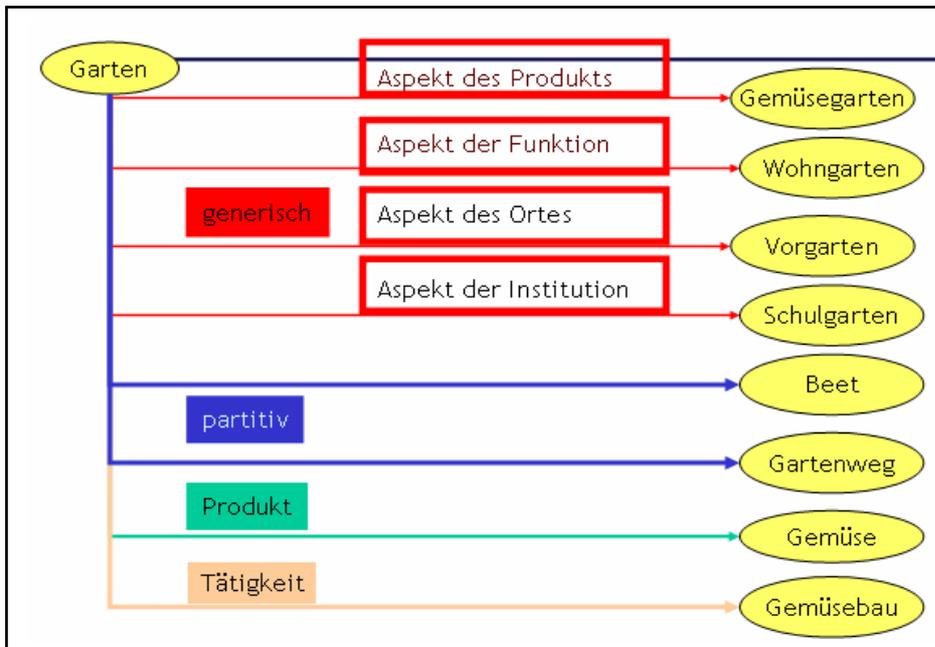


Abb.21: Typisiertes Relationeninventar

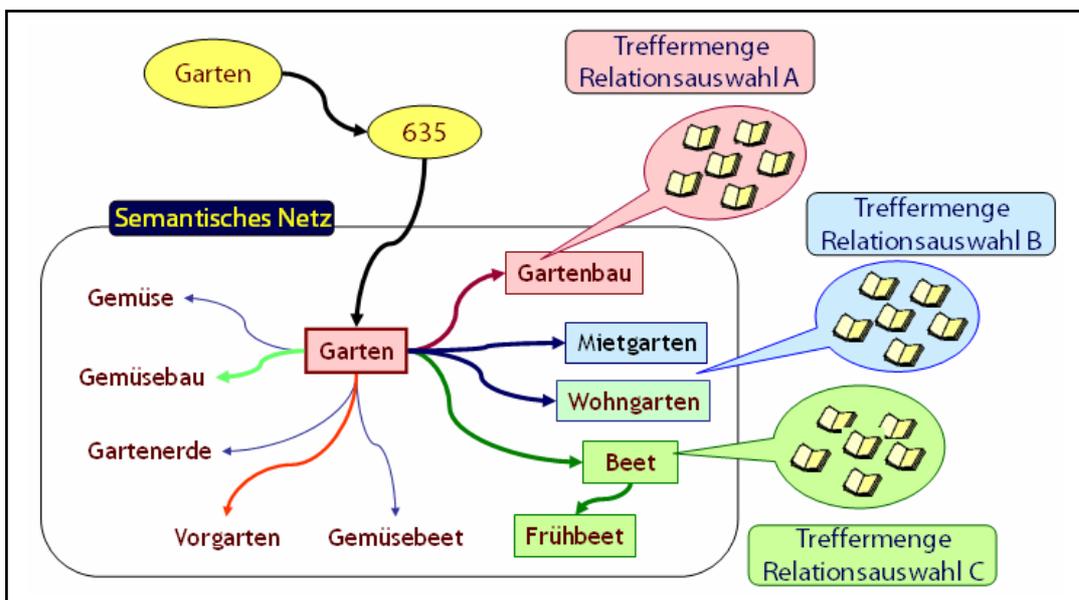


Abb.22: Selektive Treffermengenbildung nach Auswahl typisierter Relationen

Dieses Szenario war bereits in der allgemeinen Diskussion unter dem Aspekt Lokalisierung am Beispiel *Gesetzgebung* diskutiert worden.

Es sein noch einmal betont, dass die vorgenannte Darstellung nicht das gesamte Spektrum des hier vorgestellten allgemeinen Konzeptes berücksichtigen konnte, die SWD entspricht in ihrer gegenwärtigen Relationierung in keinsten Weise den hier vorgestellten Ideen; insbesondere das Szenario 3 wird durch die gegenwärtige Realität der SWD nicht unterstützt. Es konnte auf der Basis einer experimentellen Retrievalsituation lediglich eine Vorstellung vom Potenzial des Vorschlages gegeben werden.

## 9. Wissensexploration als Konzept für Zukunftsszenarios

Der hier diskutierte Vorschlag, Wissensexploration stärker als bisher als Paradigma für die Gestaltung von Retrievalumgebungen zu sehen, steht im Wettbewerb zu anderen Vorschlägen, die die Zukunft des Information Retrieval stärker geprägt sehen durch die Entwicklung automatischer, insbesondere statistisch basierter Verfahren oder kollaborativer Ansätze wie etwa dem Social tagging. Insofern bietet sich eine vergleichende Betrachtung dieser Vorschläge an.

Gegen den Einsatz automatischer Verfahren können keine Einwände erhoben werden, gerade im hier diskutierten Kontext wird durch die geeignete Verbindung von intellektueller Strukturierung mit automatisch durchgeführter Terminologienormierung ein entscheidendes Plus gegenüber allen anderen Ansätzen erzeugt. Gutes Beispiel hierfür ist der Einsatz computerlinguistischer Verfahren zur Rückführung von Indextermen oder von Sucheingaben auf die jeweils zulässigen Ansetzungsformen.

In kollaborativen Vorgehensweisen darf zu Recht das Potenzial für Verbesserungen gesehen werden; fraglich ist jedoch, wie nachhaltig solche Bemühungen gestaltet werden können. Es stellt sich die Frage, mit welcher Zuverlässigkeit davon ausgegangen werden darf, dass derartige Anstrengungen einen für die verfolgte Zielsetzung notwendigen Zeitraum überdauern. Hier sind Zweifel angebracht. Ohnehin kann man inzwischen relativierende Einschätzungen über die Erwartungen an ein Social tagging lesen. So schreibt etwa *Marc Buzinkay*: "Bibliotheken werden um Social Tagging nicht herum kommen, da schlichtweg die Kapazitäten für eine inhaltliche Erschließung fehlen. Stattdessen wäre es von Notwendigkeit, Normdateien zum Bestandteil des Tagging zu machen. Verfehlt wäre es, auf die bibliothekarische Erschließung zu verzichten, weil das zu Lasten des Verhältnisses von Recall und Precision gehen würde."<sup>29</sup>

Für unsere Diskussion ist wichtiger, dass Recall und Precision ohnehin nur 2 – und nicht alle – Bestimmungsgrößen sind, die zur Bewertung von Such- und Findesystemen genutzt werden können. Bei dem hier diskutierten Vorschlag geht es im Kern um einen dem Findevorgang vorgeschalteten Explorationsprozess, der auf der Basis von Strukturen erfolgen muss, die zwischen Inhaltseinstitäten bestehen. Diese Strukturen können nach gegenwärtiger Einschätzung niemals allein durch automatischen Prozesse generiert werden. Sie als Ergebnis eines freien kollaborativen Prozesses zu sehen, fällt doch schwer. Die Herstellung dieser Strukturen ist zunächst ein zutiefst intellektueller Vorgang, der mindestens die Vertrautheit mit verschiedenen Sachverhalten (Wissen um die Inhalte und Fragestellungen eines Sach- oder Fachgebiets und der zwischen ihnen bestehenden Zusammenhänge) und die Einhaltung bestimmter Spielregeln voraussetzt (z.B. Kenntnis und Berücksichtigung der zur Strukturierung der Inhaltseinstitäten notwendigen Regeln). Beides kann durch freie Kollaboration nur schwerlich sicher gestellt werden. Die einmal erzeugten und hoffentlich dann auch gepflegten Strukturen können in vielfältiger Weise für eine Nachverwertung (automatisch oder kollaborativ) genutzt werden.

Sicher kann man in den hier vorgestellten Komponenten zu Recht einen großen Aufwand sehen. Hält man die Idee der Wissensexploration für eine sinnvolle Maßnahme zur Gestaltung von Retrievalvorgängen, so würde der Verzicht auf diesen Aufwand bedeuten, dass er an anderer Stelle geleistet werden muss (z.B. in Form einer vollständigen Verlagerung auf den suchenden und das Retrievalsystem befragenden Nutzer) und dass die Retrievalsysteme um eine mögliche sinnvolle Eigenschaft ärmer wären.

---

<sup>29</sup> Unter: <http://www.buzinkay.net/blog-de/2009/05/buchrezension-good-tags-bad-tags/>.

## 10. Road map zur semantischen Exploration

Zum Abschluss seien noch einmal die Punkte zusammen gefasst, die notwendig zu erfüllen sind, um das vorstehend diskutierte Gesamtkonzept zu realisieren:

- Entwicklung und Vereinbarung einer Kernontologie (Neuaufbau und Umbau eines bestehenden Systems);
- Entwicklung eines erweiterten Relationeninventars zur Repräsentation semantischer Inhaltentitäten;
- Aufbau lokalisierter semantischer Netze gemäß dieses erweiterten Relationeninventars;
- Entwicklung eines Inventars semantischer Interoperabilitätstypen zur Verbindung der Netze mit der Kernontologie;
- Sprachliche Anreicherung des Zugangsvokabulars zur Kernontologie und den verbundenen Netzen, z.B. durch automatische Indexierung oder vergleichbare linguistische Bearbeitung der Sucheingaben;
- Vollständige Ausstattung aller Dokumente mit Notationen der Kernontologie, ggf. durch Einsatz von Verfahren der automatischen Klassifizierung;

Auf die weiteren wichtigen Gesichtspunkte, die die konkreten Erfordernisse zur Repräsentation der Daten für ihre Nachnutzung und die Optionen zur Herstellung funktional geeigneter Retrievalumgebungen zum Gegenstand hätten, muss hier nicht weiter eingegangen werden.

# Semantische Wissensrepräsentation und Interoperabilität

## Teil 2: Ein formales Modell semantischer Interoperabilität

In diesem zweiten Teil wird ein formales Modell zur semantischen Interoperabilität als Brücke zwischen formaler Modellierung und intellektueller Interpretation vorgestellt, das ein besseres Verständnis der zentralen Begriffe von semantischer Ähnlichkeit und Verschiedenheit von Begriffen und Klassen sowohl als elementare Inhaltselemente von Dokumentationssprachen als auch als synthetische Repräsentationen von Dokumentinhalten ermöglichen soll.

### 11. Typen von Inhaltselementen

Im Folgenden soll ein Modell vorgeschlagen werden, das im Sinne der im Abschnitt 3 angesprochenen Verbindung von formaler Modellierung, kognitiver Interpretation und maschineller Verarbeitung geeignet ist, ein Verständnis von semantischer Interoperabilität zu entwickeln und das über einfache Begriffe hinaus auch komplexe Sachverhalte und die synthetisierten Ergebnisse von Erschließungsvorgängen berücksichtigt.

Um Überlegungen zur Vergleichbar- oder Austauschbarkeit semantischer Elemente machen zu können, muss eine Beschreibung derartiger Elemente in einem Objektverständnis entwickelt und eine Sprachregelung gefunden werden. Die zu betrachtenden Objekte werden hier *Inhaltselemente* genannt<sup>30</sup> und repräsentieren sowohl *Begriffe* als auch *Klassen*. In einem ersten Schritt werden Typen von Inhaltselementen angegeben und es werden Formen der zwischen diesen Elementen bestehenden Beziehungen diskutiert.<sup>31</sup> Es erweist sich dabei als notwendig, verschiedene Komplexitätsstufen zu unterscheiden, da Inhaltserschließung nicht nur mit elementaren Begrifflichkeiten, sondern insbesondere mit daraus aufgebauten und als Inhalt von Dokumenten ermittelten Sachverhalten zu tun hat. Komplexe Sachverhalte, zunächst intuitiv verstanden als Aussagen, die durch Verbindung mehrerer Begrifflichkeiten gebildet werden, treten dabei an verschiedenen Stellen auf:

- in den Dokumenten als behandelte Themen oder Gegenstände,
- möglicherweise in der benutzten Dokumentationssprache als präkombinierte Elemente,
- in der dokumentationssprachlichen Repräsentation der Dokument-Gegenstände als synthetisierte Aussagen.

So ergeben sich 3 Typen inhaltsbezogener Entitäten, die Bezugspunkte für Betrachtungen der semantischen Interoperabilität sein können:

1. die Dokumentinhalte,
2. die Elemente von Dokumentationssprachen,
3. die dokumentationssprachlichen Repräsentationen von Dokumentinhalten.

Alle Inhaltselemente werden in Anlehnung an das semiotische Dreieck nach dem Schema der Abb.23 symbolisiert.

<sup>30</sup> Aus Gründen der sprachlichen Vereinfachung wird im Text statt *Inhaltselement* auch das Wort *Entität* benutzt.

<sup>31</sup> Die Darstellung lehnt sich an das noch im Entwurfsstadium befindliche FRSAD-Modell an; vgl.: IFLA Working Group on Functional Requirements for Subject Authority Records (FRSAR) (Hrsg.): Functional Requirements for Subject Authority Data (FRSAD): a conceptual model. 2nd Draft 2009-06-10. Unter: <http://nkos.slis.kent.edu/FRSAR/report090623.pdf>. 69 S.

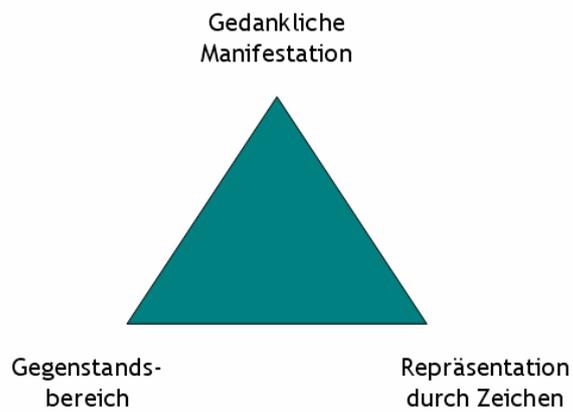


Abb.23: Allgemeines Semiotisches Dreieck

Es sollen dabei 2 Strukturtypen von Aussageentitäten unterschieden werden:

1. ein einfaches materielles oder immaterielles Objekt der Wahrnehmung oder Konzeptualisierung mit seiner Repräsentation durch eine Bezeichnung, im Folgenden *Einfache Inhaltsentität (EI)* – zur sprachlichen Vereinfachung auch *Einfache Entität* – genannt und symbolisiert durch Abb.24.

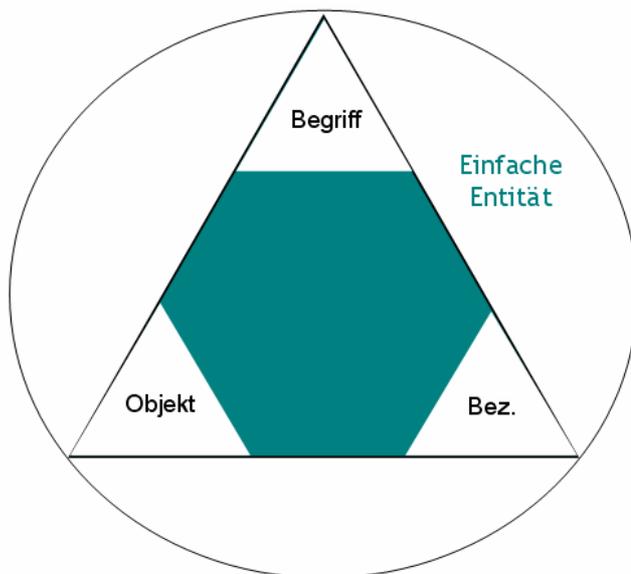


Abb.24: *Einfache Entität* als Semiotisches Dreieck

2. ein komplexes Objekt, ein Thema, an dessen inhaltlicher Interpretation erkennbar mehr als eine *Einfache Entität* beteiligt ist, mit seiner Konzeptualisierung und Repräsentation durch eine Bezeichnungs-Phrase, im Folgenden *Komplexe Inhaltsidentität (KI)* – oder *Komplexe Identität* – genannt und symbolisiert durch Abb.25.

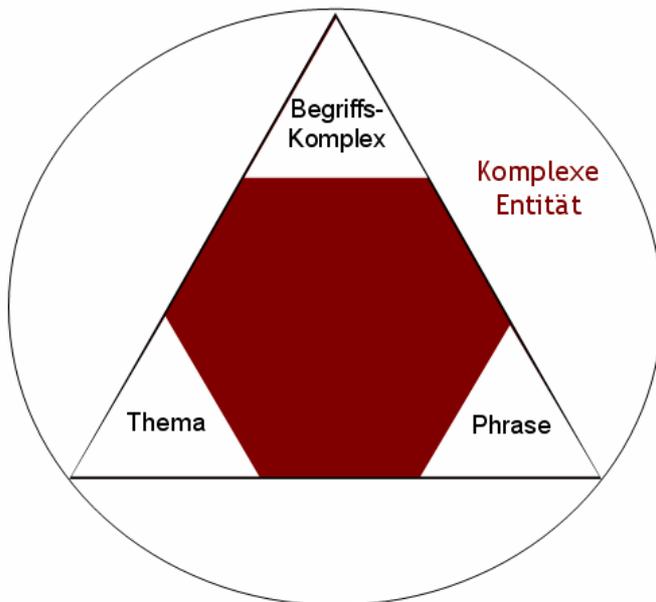


Abb.25: *Komplexe Entität* als Semiotisches Dreieck

Inhalte von Dokumenten, die in ihnen behandelten Themen und getroffenen Aussagen dürfen im Allgemeinen als komplex betrachtet werden. Die Komplexität ist nicht allein Folge syntaktischer Faktoren bei der Bildung zusammengesetzter Aussagen, sondern ebenso Resultat semantischer Zuordnungen. Einen Beleg hierfür liefert der Umstand, dass sich die Begrifflichkeiten, die den Inhalt konstituieren, häufig verschiedenen kategorialen Gesichtspunkten, wie Raum, Zeit, Sache, aber auch Agens, Wirkung, Eigenschaft zuordnen lassen. Dokumentbezogene Inhaltsentitäten gehören im Allgemeinen also zum Typ der komplexen Inhaltsentitäten. Ihre Verbindung mit den Inhaltsentitäten der Dokumentations Sprachen kann auf unterschiedliche Art und Weise erfolgen und muss nun näher betrachtet werden.

## 12. Typen der Inhaltsentitäten auf der Ebene der Dokumentations Sprachen

Inhaltsentitäten auf der Ebene der Dokumentations Sprachen zeigen bei genauerer Betrachtung einen Doppelcharakter. Zunächst besitzen sie Elementcharakter als Bestandteile eines größeren Systems. Andererseits zeigt die inhaltliche Betrachtung dieser Elemente jedoch häufig, dass sie aus begrifflichen Komponenten zusammengesetzt sind, die in der Regel aus pragmatischen Anwendungsgründen in Form von präkombinierten Ausdrücken in die Dokumentations Sprache integriert werden. Wir haben es bei Inhaltsentitäten von Dokumentations Sprachen also sowohl mit einfachen als auch mit komplexen Inhaltsentitäten zu tun. Terminologisch sei vereinbart, dass die Benennungen *Einfach* und *Komplex* auf der inhaltlichen Ebene der vor-dokumentations sprachlichen Festlegung angesiedelt sind, die noch einzuführenden Benennungen *Elementar*, *Präkombiniert* und *Synthetisch* auf der Ebene der Dokumentations Sprache. Der Zusammenhang wird durch das Diagramm der Abb.26 veranschaulicht.

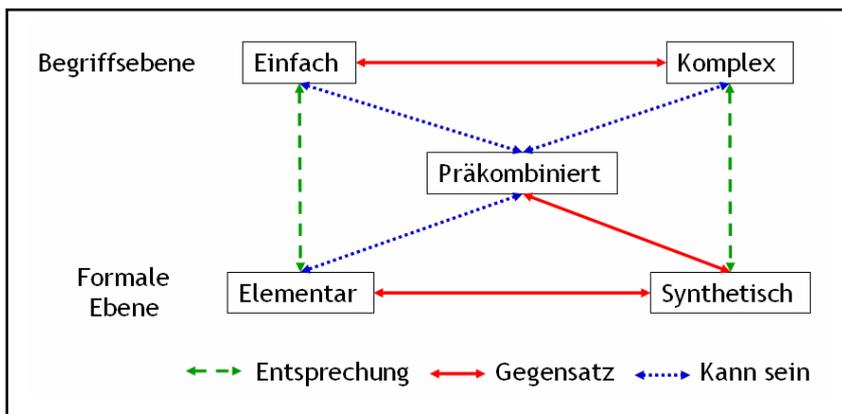


Abb.26: Sonderstellung der präkombinierten Inhaltselemente

Die Lesart des Diagramms ist: Formale Elemente einer Dokumentationssprache können auf *einfachen Begriffen* basieren, dann heißen sie als Bestandteile der Dokumentationssprache *elementar*. Basieren sie auf *komplexen Begriffen* (d.h. mit Merkmalen aus mehreren Aspektbereichen), dann heißen sie *präkombiniert*. Komplexe Begriffe können aber auch durch Einsatz syntaktischer Regeln aus elementaren Bestandteilen zusammengesetzt werden, die formalen Elemente der Dokumentationssprache heißen dann *synthetisch*.

Um die Brücke zwischen formaler und kognitiver Interpretation zu schlagen und eine formale Behandlung semantischer Interoperabilität zu ermöglichen, soll nun das für Dokumentationssprachen übliche Begriffs- und Klassenverständnis näher betrachtet werden.

### 12.1 Prinzipien der Begriffs- und Klassenbildung

Will man eine präzisierende Beschreibung der Begriffe *Begriff* und *Klasse* geben, so muss man entscheiden, welcher von beiden die Basis des anderen sein soll? In dokumentations Sprachlichen Betrachtungen hat sich allgemein durchgesetzt, *Begriff* als elementar und *Klasse* als daraus abgeleitet zu betrachten. So baut die DIN 32705 auf der DIN 2330 auf und enthält entsprechend als Definitionen<sup>32</sup>:

1. „Eine *Klasse* ist die Zusammenfassung derjenigen Begriffen, die mindestens ein identisches Merkmal haben.“
2. „Ein *Merkmal* ist ein Begriffselement, das durch Aussage über die Eigenschaft eines Gegenstandes festgelegt wird.“
3. „Ein *Klassem* (Synonym: *Klassifikatorisches Merkmal*) ist dasjenige gemeinsame Merkmal von Begriffen, das zur Bildung einer Klasse benutzt wird und diese von anderen Klassen unterscheidet.“

Zur Entwicklung des Begriffsverständnisses geht man im Rahmen dieser gegenstandsbezogenen, analytischen Begriffstheorie davon aus, dass sich kognitive Vorstellungen von realweltlichen oder abstrakten *Objekten* sich in Form von Begriffen fixieren lassen, die zum Zwecke der Kommunikation mit anderen Menschen in Zeichensystemen als *Bezeichnungen* oder *Be-*

<sup>32</sup> DIN 32705: *Klassifikationssysteme: Erstellung und Weiterentwicklung von Klassifikationssystemen*. Berlin 1987.

*nennungen* (Bezeichnungen in Sprache) formuliert werden können. In der DIN 2330<sup>33</sup> lautet entsprechend die Präzisierung für *Begriff*:

„Die gedankliche Zusammenfassung von individuellen Gegenständen zu gedachten allgemeinen Gegenständen führt zu Denkeinheiten, die als *Begriffe* bezeichnet werden.“

Der Prozess der *Zusammenfassung* wird dabei auf die Feststellung von *Merkmalen* gestützt, die an den oder für die Begriffe festgestellt werden, indem Aussagen über die den Begriffen zugrunde liegenden Objekte gemacht werden:

„Die Zusammenfassung vollzieht sich auf der Basis von Merkmalen. Unter Merkmalen versteht man die Eigenschaften eines einzelnen individuellen Gegenstandes oder diejenigen Eigenschaften von individuellen Gegenständen, die zur Analyse der entsprechenden Begriffe herangezogen werden.“<sup>34</sup>

Dem Konzept *Merkmal* kommt also in den begrifflichen Fixierungen der Konzepte *Begriff* und *Klasse* eine entscheidende Bedeutung zu. Es ist daher verwunderlich, dass die Ausweisung der die begrifflichen Entitäten in Dokumentationssprachen konstituierenden Merkmale weder Eingang in die Normen- und Regelwerke gefunden hat<sup>35</sup>, noch in nennenswertem Umfang in den real existierenden Dokumentationssprachen anzutreffen ist. Selbst in der besonders nahe liegenden Frage der Hierarchiebildung in Klassifikationssystemen und der Vererbung von Inhalten über mehrere Hierarchiestufen, wird in der Regel das die Klassen formal bestimmende klassifikatorische Merkmal nicht explizit ausgewiesen. Eine derartige Vorgehensweise als formale Modellierung würde heute in der Verbindung zwischen maschineller und kognitiver Interpretation der Inhaltselemente und der zwischen ihnen bestehenden Relationen Vorteile bieten, wie in Teil 1 dieses Beitrages diskutiert wird.

Das hiermit vorgestellte Begriffsverständnis hat manche Kritik und Modifikation erfahren<sup>36</sup>, insbesondere der pragmatische Aspekt der tatsächlichen *Begriffsverwendung*, der für die Dynamisierung von Dokumentationssprachen eine große Bedeutung besitzt, kommt in dem Konzept zu kurz und bleibt für die Modellbildung ein Desiderat. Wir greifen diesen Gesichtspunkt in Abschnitt 15 noch einmal auf. Aus Gründen der technischen Einfachheit in der Darstellung der Interoperabilitätsfragen wollen wir für das Folgende zunächst bei dem statischen Verständnis der Begriffsbildung und der Rückführung der DIN 32705 bleiben. Als Diagramm lässt sich das Verständnis von *Klasse* durch Abb.27 veranschaulichen.

---

<sup>33</sup> DIN 2330: *Begriffe und Benennungen*. Berlin: Beuth 1979, Abschnitt 3.1.

<sup>34</sup> DIN 2330, Abschnitt 3.2.

<sup>35</sup> Vgl. ergänzend zu den schon genannten Normen auch: DIN 1463: *Erstellung und Weiterentwicklung von Thesauri: Teil 1: Einsprachige Thesauri* (1987); *Teil 2: Mehrsprachige Thesauri* (1993). Berlin: Beuth 1987-1993.

<sup>36</sup> Vgl. z.B. das Themenheft: *Ethik und Sozialwissenschaften*. 7(1996) H.1 mit einem Impulsartikel sowie den darauf antwortenden Beiträgen:

**Dahlberg, I.:** *Zur 'Begriffskultur' in den Sozialwissenschaften: lassen sich ihre Probleme lösen?*.

In: *Ethik und Sozialwissenschaften*. 7(1996) H.1, S.3-13.

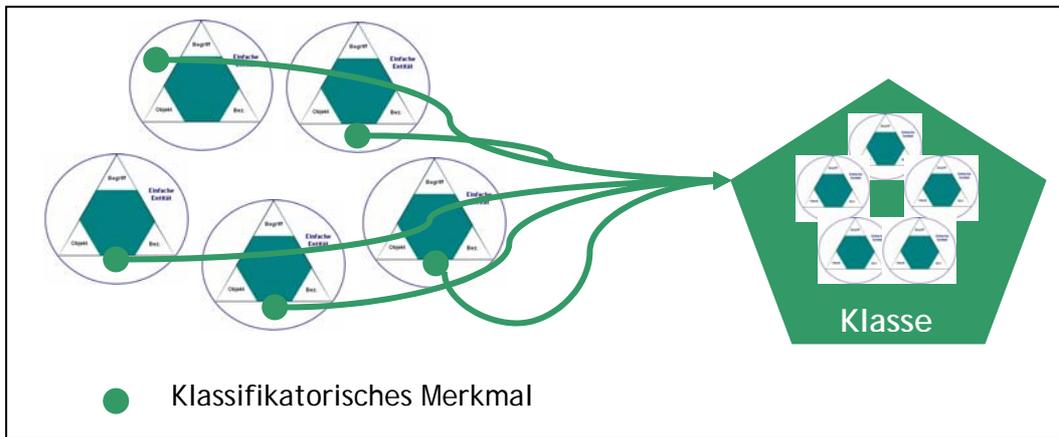


Abb.27: Klassenbildung mit Begriffen und gemeinsamem klassifikatorischem Merkmal

Tritt ein weiteres Merkmal hinzu und gibt es Klassen, die sowohl Begriffe mit dem alten als auch Begriffe mit dem neuen Merkmal beinhalten, so sieht man das Verhältnis der Klassen zueinander als hierarchisch an (Abb.28).

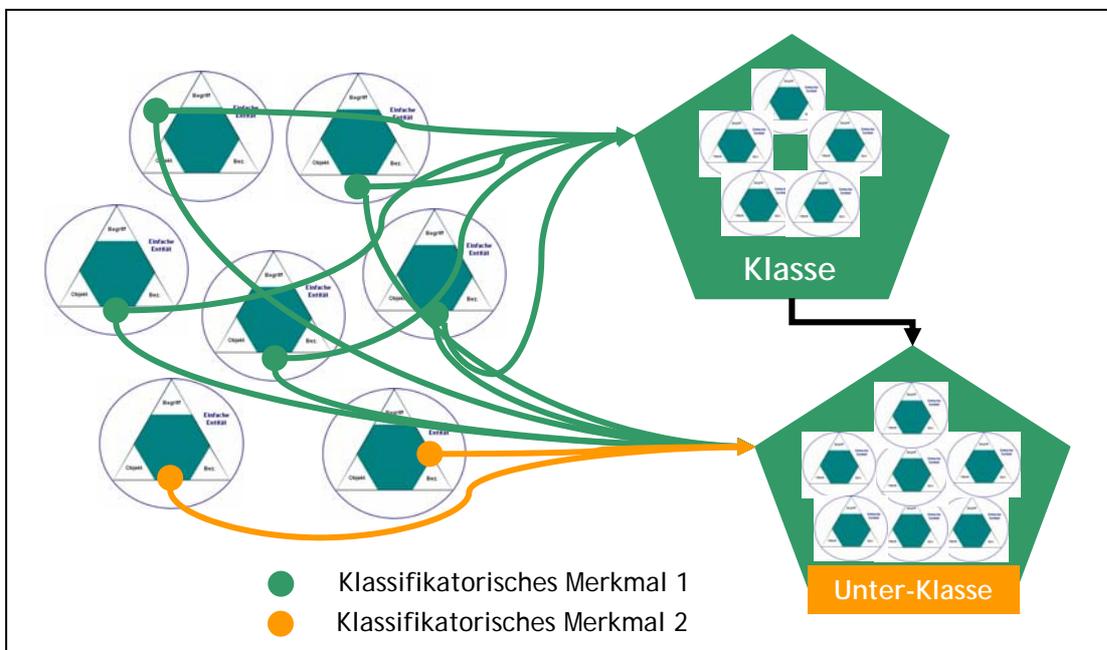


Abb.28: Hierarchie zwischen Klassen

Diese Vorstellung entspricht dem Verständnis der Ober- / Unterbegriffsrelation, die durch die Eigenschaft charakterisiert ist, dass ein Unterbegriff alle Merkmale des Oberbegriffs sowie mindestens ein zusätzliches besitzt. Aus der Betrachtung der der Klassen zugeordneten klassifikatorischen Merkmale lassen sich die hierarchischen Struktureigenschaften des Systems herleiten (vgl. Abb.29).

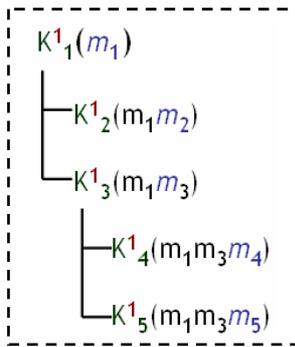


Abb.29: Hierarchie in der klassifikatorischen Struktur

So ist etwa die Klasse  $K_4^1$  der Klasse  $K_3^1$  hierarchisch untergeordnet, weil die jeweiligen Merkmalsmengen  $\{m_1m_3m_4\}$  sowie  $\{m_1m_3\}$  sich nur in einem hinzukommenden Merkmal unterscheiden, ansonsten aber identisch sind. Für die nachfolgende Diskussion sein schon festgehalten, dass diese Klassen, deren klassifikatorische Merkmale aus nur einem Aspektbereich hervorgehen, *Elementare Klassen* heißen sollen.

Die Basierung der Schlussfolgerung auf Basis der Merkmalsmenge hat den Vorteil, dass sie sich auch maschinell und nicht nur durch Vorhandensein einer strukturabbildenden Notation oder Auswertung zusätzlicher Hierarchieinformationen herstellen lässt.

Innerhalb von Klassifikationssystemen, die für die Zwecke der Dokumenterschließung benutzt werden, wird in den seltensten Fällen ein klassifikatorisches Merkmal explizit angegeben. In der Regel muss man es über die intellektuelle Analyse der Klassenbedeutungen erschließen. Betrachten wir ein Beispiel. Sucht man in der *DDC* über das Register nach Zoologie, so wird man auf die Klasse 590 geführt (Abb.30).

Zoologie	590
----------	-----

Abb.30: Registereintrag *Zoologie* in der *DDC*

Die Klasse 590 trägt die Benennung *Tiere* mit einem Hinweis auf weitere Begriffe, die Inhalt der Klasse sind (Abb.31).

<b>590</b>	<b>Tiere</b>
Hier auch: Naturgeschichte und deskriptive Biologie von Tieren, Zoologie, fächerübergreifende Werke über Tiere	

Abb.31: Klasse 590 in der *DDC*

Für unsere Interessen einer genaueren Interoperabilitätsanalyse stellt sich die Frage: Ist die Klasse 590 interoperabel mit dem Begriff *Zoologie*? Die Frage wäre einfach zu beantworten, könnte man sich an einem eindeutigen klassifikatorischen Merkmal orientieren, das zur Klassenbildung benutzt wurde. Da ein solches nicht explizit angegeben ist, bleiben nur Hilfsargumente durch weitere Betrachtungen der hergestellten Strukturierung. Man findet man unter 591 zunächst die in Abb.32 zu sehenden subordinierten Klassen.

<b>591</b>	<b>Einzelne Themen in der Naturgeschichte von Tieren</b>
.3	<b>Genetik, Evolution, Jungtiere</b>
.4	<b>Anpassung des Körpers</b>
.5	<b>Verhalten</b>
.6	<b>Verschiedene Gruppen nichtverwandter Tiere</b>
.7	<b>Tierökologie, für einzelne Lebensräume charakteristische Tiere</b>
.9	<b>Behandlung von Tieren nach einzelnen Kontinenten, Ländern, Ortschaften</b>

Abb.32: Klasse 591 in der DDC

Die Klassen können nicht als taxonomische Spezialisierung angesehen werden, sondern als verschiedene Aspekt-Sichtweisen. Abb.33 zeigt dann die Klassen, die der taxonomischen Spezifizierung gewidmet sind.

<b>590</b>	<b>Tiere (Zoologie)</b>
591	Einzelne Themen in der Naturgeschichte
592	Evertebrata (Wirbellose)
593	Wirbellose des Meeres und der Meeresküste
594	Mollusca (Weichtiere), Tentaculata (Kranzfühler)
595	Arthropoden (Gliederfüßer)
596	Chordata (Chordatiere)
597	Wechselwarme Wirbeltiere; Pisces (Fische)
598	Aves (Vögel)
599	Mammalia (Säugetiere)

Abb.33: Klassenübersicht zu 590 in der DDC

Ist die Klasse 590 nun interoperabel mit dem Begriff Zoologie? Führt man intellektuell die inhaltliche Analyse der Klasse durch, indem man alle subordinierten Klassen betrachtet, so sieht man, dass zwischen den subordinierten Klassen 591 und 592 – 599 ein Aspektwechsel stattfindet. Im Unterschied hierzu war die Hierarchie zwischen Klassen in Abb.27 – Abb.29 durch die Eigenschaft geprägt, dass das zusätzliche Merkmal demselben Aspekt zuzuordnen ist. Der Aspektwechsel in Abb.33 zeigt sich deutlich durch die vorhandene implizite Facettierung, die zur dokumentspezifischen Notationssynthese benutzt wird, z.B. in 591.5 (vgl. Abb.34).

<b>.5</b>	<b>†Verhalten</b>	Hier auch: Tierpsychologie, Verhaltensanpassung, Verhaltensforschung (Ethologie) Klassifiziere vergleichende Psychologie zwischen Menschen und Tieren in 156; klassifiziere Physiologie des Nervensystems in 573.8; klassifiziere umfassende Werke über Anpassung bei Tieren in 591.4 <i>Siehe Praxishilfe bei 302–307 vs. 156</i>
.51	Allgemeine Themen zum Verhalten	Begrenzt auf die nachfolgend Genannten Klassifiziere auf einzelne Verhaltensweisen bezogene allgemeine Themen in 591.53–591.59; klassifiziere Spiel in 591.563
.512	†Instinkt	
.513	†Intelligenz	
.514	†Lernen	
.518	†Nächtliches Verhalten	Hier auch: Nachttiere, nachtaktive Tiere

Abb.34: Implizite Facettierung in der DDC-Struktur für die Klasse 591.5

Noch deutlicher wird diese Vorgehensweise an anderen Stellen der DDC zum Ausdruck gebracht, wenn explizite Syntheseanweisungen gegeben werden (Abb.35):

<b>573</b>	<b>Einzelne physiologische Systeme bei Tieren, regionäre Histologie und Physiologie bei Tieren</b>	Hier auch: umfassende Werke über einzelne physiologische Systeme Abgesehen von Abweichungen bei bestimmten Einträgen, hänge an jede durch * gekennzeichnete Unterklasse an wie folgt:
01–1	Standardschlüssel; vergleichende Physiologie; das System bei einzelnen Tierarten	Hänge die Ziffern an, die 571 in 571.01–571.1 folgen, z.B. vergleichende Physiologie des Systems 1, das System bei Säugetieren 19
2	Aktivität eines physiologischen Systems innerhalb eines anderen Systems	Klassifiziere Hormone in 374
21	Kreislauf innerhalb des Systems	
2101–211	Standardschlüssel; vergleichende Physiologie; Kreislaufsystem bei einzelnen Tierarten	Hänge an 21 die Ziffern an, die 571 in 571.01–571.1 folgen, z.B. Kreislaufsystem bei Säugetieren 2119
213	Anatomie und allgemeine biologische Kreislaufprozesse	Hänge an 213 die Ziffern an, die 571 in 571.3–571.9 folgen, z.B. Anatomie des Kreislaufs 2133, Entwicklung des Kreislaufs 2138
25	Integument des Systems	Hier auch: Membranen um ein Organ, auch wenn sie kein Bestandteil des Organs sind, z.B. Perikard 573.1725 Hänge an 25 die Ziffern an, die 21 in 2101–213 folgen, z.B. Zytologie des Integuments eines Organs 2536
27	Muskeln des Systems	Hänge an 27 die Ziffern an, die 21 in 2101–213 folgen, z.B. Biophysik von Muskeln 2734

Abb.35: Implizite Facettierung in der DDC

Klassen wie im Beispiel der *DDC-Klasse 590*, bei denen Merkmale unterschiedlicher Aspekte gemeinsam als klassifikatorische Merkmale zur Klassenbildung verwendet werden, nennt man im Allgemeinen *präkombinierte Klassen*. Will man Interoperabilitätsbetrachtungen zwischen Begriffen und Klassen auf derartige präkombinierte Klassen erweitern, so spielt also der Kontext oder die Aspektierung eine wesentliche Rolle. Nimmt man für die *DDC-Klasse 590* eine aspektorientierte Merkmalsanalyse vor, so kommt man zu einer Darstellung wie sie in Abb.36 gegeben wird.

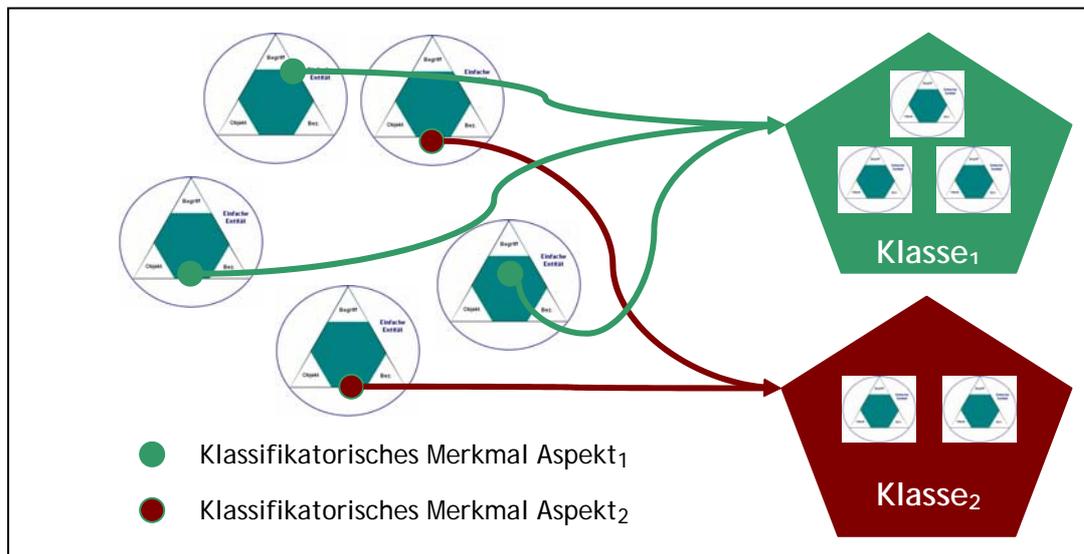


Abb.36: Aspektorientierte Klassenbildung

Die konsequente Beachtung der aspektorientierten Trennung klassifikatorischer Merkmale führt zur Facettenklassifikation, die Beschränkung auf generelle Aspekte, die prinzipiell auf alle Klassen des Systems anwendbar sind (inhaltlich gesehen z.B. Raum, Zeit, Form, Sprache) auf (allgemeine) Schlüsselungen.

## 12.2 Präkombination, Schlüsselung, Facettierung und Synthese

Zur Vermeidung von Polyhierarchien ist es sinnvoll, die zu berücksichtigenden Merkmale nach Aspekten zu differenzieren und die entstehenden Klassen in Facetten zu gliedern (vgl. Abb.37).

Wie bereits im Kontext der Abb.29 angegeben, soll als terminologische Festlegung vereinbart werden, dass Klassen, die durch Berücksichtigung eines klassifikatorischen Merkmals aus nur einem Aspektbereich hervorgehen, *Elementare Klassen* heißen, alle anderen Klassen sollen *Komplexe Klassen* heißen. In der Facettenstruktur kommen komplexe Klassen nicht vor.

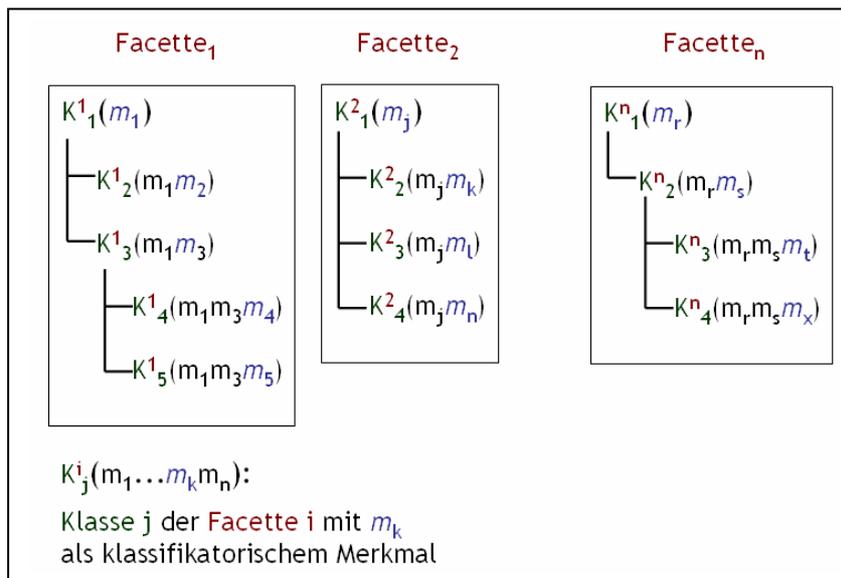


Abb.37: Facettengliederung systematischer Strukturen

Komplexe Sachverhalte sind nicht als Klassen in der Struktur repräsentiert, sondern sie werden als Verbindung mehrerer Klassen gedacht, die sich als Bausteine aus einer Aspektanalyse (analytischer Teil der Facettenklassifikation) der Sachverhalte ergeben (vgl. Abb.38).

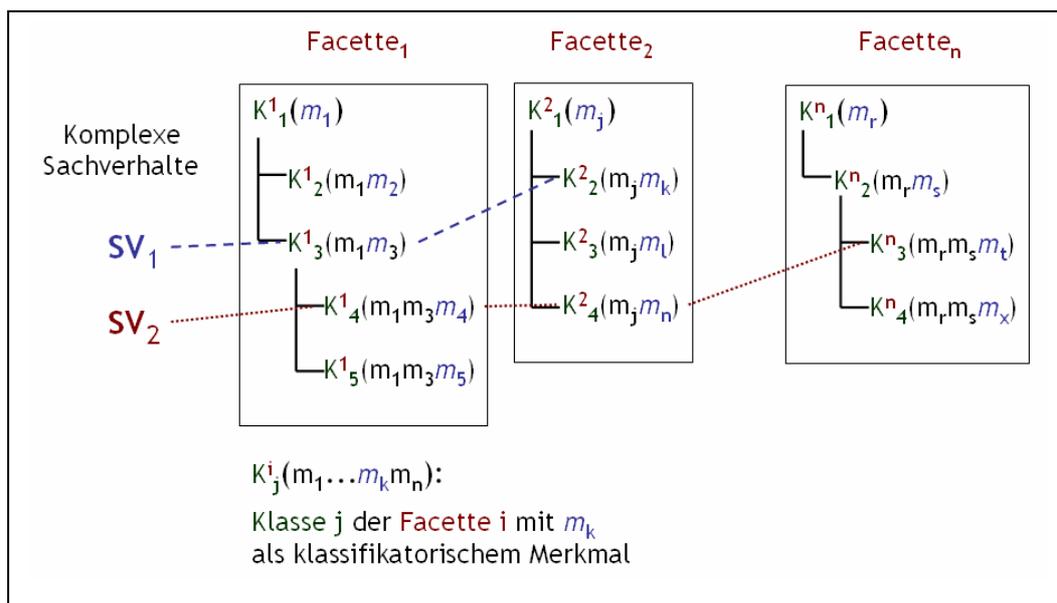


Abb.38: Repräsentation komplexer Sachverhalte durch aspektorientierte Elementar-Klassen

Der komplexe Sachverhalt  $SV_1$  erfordert für seine Repräsentation die Klassen  $K^1_3$  der Facette<sub>1</sub> und die Klasse  $K^2_2$  der Facette<sub>2</sub>, entsprechend der Sachverhalt  $SV_2$  die Klassen  $K^1_4$  der Facette<sub>1</sub>,  $K^2_4$  der Facette<sub>2</sub> und  $K^n_3$  der Facette<sub>n</sub>. Im Rahmen der Facettenklassifikation wird die Synthese zu einer Gesamtaussage durch die Citation order gesteuert.

Das hier am Beispiel klassifikatorischer Strukturen ausführlicher diskutierte Phänomen der Abbildung von inhaltlicher Komplexität der Dokument-Gegenstände durch präkombinierte Strukturen der Inhaltselemente als Bestandteile einer Dokumentationssprache hat seine Widerspiegelung in verbalen Dokumentationssprachen, wenn etwa wie im Beispiel der *Library of Congress Subject Headings (LCSH)* Elemente gebildet werden, wie

- Book industries and trade – Germany – History
- Book industries and trade – Handbooks, manuals, etc.
- Cataloging of audio-visual materials – United States – Handbooks, manuals, etc.
- Cataloging of nonbook materials – United States – Handbooks, manuals, etc.

oder wenn aspektübergreifende Komposita anzutreffen sind, wie es besonders eindrucksvoll in der *Methode Gülich*<sup>37</sup> zu sehen ist:

- Eisenbahnbetriebswirtschaft
- Parlamentskontrolle der Wehrpolitik
- Parlamentsmitgliederbibliographie

So kann es auch bei den Inhaltselementen verbaler Dokumentationssprachen zu einem Nebeneinander von Repräsentanten einfacher Begriffe aus Dokumentinhalten mit Repräsentanten für komplexe Dokument-Gegenstände kommen, wie es in Anlehnung an die Darstellung der Abb.26 in Abb.39 gezeigt wird.

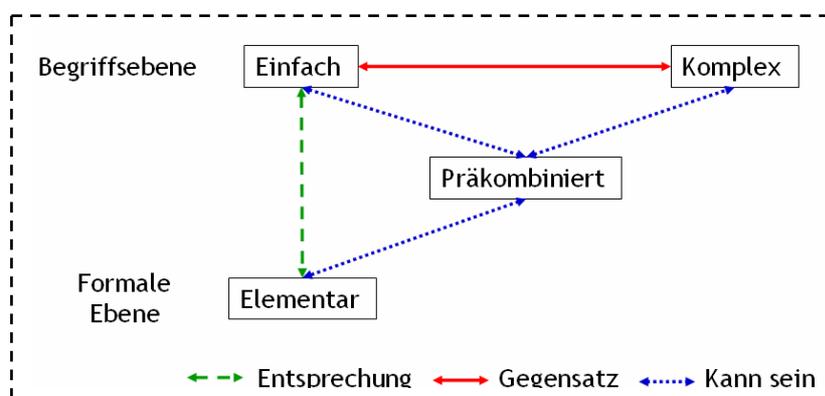


Abb.39: Sonderstellung der präkombinierten Inhaltselemente in Dokumentationssprachen

### 13. Typen der Inhaltselemente auf der Ebene der dokumentations sprachlichen Repräsentation der Dokument-Gegenstände

Betrachten wir nun, wie sich die Aufgabe der Repräsentation von – im Allgemeinen komplexen – Dokument-Gegenständen durch Inhaltselemente auf die typologische Beschreibung der Inhaltselemente auswirkt. Erfolgt die Repräsentation im Sinne eines gleichordnenden Indexierens ohne jede Verknüpfungssyntax, so ergibt sich keine Notwendigkeit für neue Überlegungen. Werden jedoch unter Anwendung gewisser syntaktischer Konnektoren a posteriori aus den Elementen der Dokumentationssprache größere Einheiten gebildet, die als Ganzes eine höhere Aussagekraft haben sollen als die gedachte Boolesche Verknüpfung der Elemente, so resultieren Einheiten, die hier als *Synthetische Inhaltselemente (SI)* bezeichnet werden sollen (vgl. Abb.40).

<sup>37</sup> Die Beispiele sind dem Thesaurus der Bibliothek des Deutschen Bundestages, Ausgabe 1975 entnommen. Vgl. auch: **Winter, E.E.:** *Das Gülich-System und sein Standort im Rahmen einer formalen Theorie des realen Kataloges: dargestellt am Aufbau des Sachkataloges der Bibliothek des Institutes für Weltwirtschaft an der Universität Kiel.* Frankfurt a.M.: 1975.

**Scheerer, H.:** *Gülich online: Die Retrokonversion des Zettelkatalogs der Bundestagsbibliothek.* In: Bibliotheksdienst. 41(2007) H.8, S.893-901.

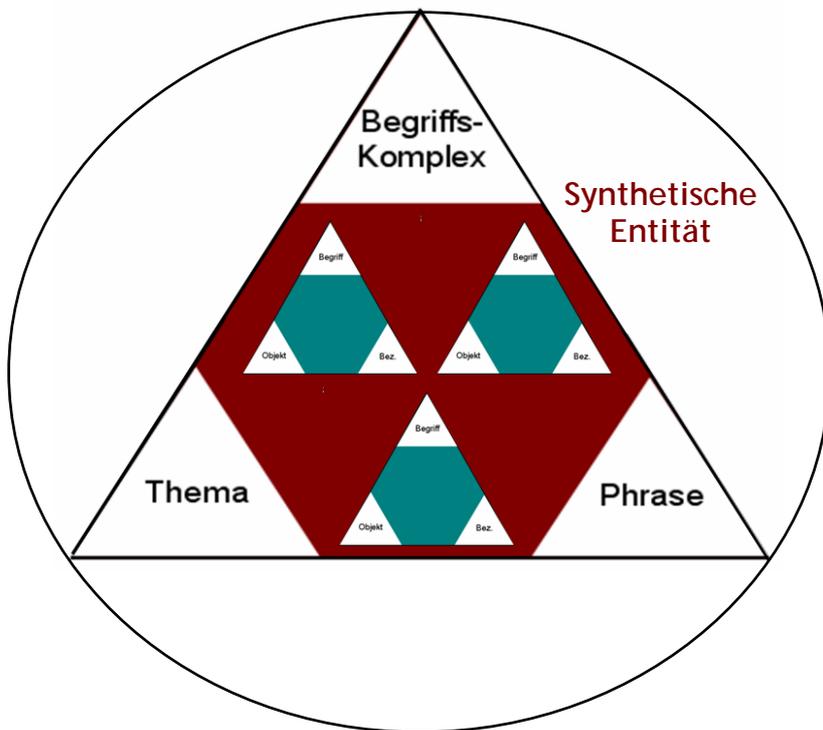


Abb.40: Symbolisierung *Synthetischer Entitäten* in Anlehnung an das Semiotische Dreieck

In erster Linie werden diese *Synthetische Inhaltsentitäten (SI)* als aus *Einfachen Inhaltsentitäten* gebildet gedacht:

$$SI = EI_1 \otimes EI_2 \otimes \dots \otimes EI_n,$$

$\otimes$  steht dabei als Symbol für die syntaktischen Verknüpfungsregeln, die innerhalb der jeweiligen Dokumentationssprache zur Synthese zugelassen sind. Aufgrund der expliziten Verknüpfungsregeln ergibt sich für die *Synthetischen Inhaltsentitäten* ein erhöhtes Maß, sie als zusammengesetzt aus Einheiten zu denken, als dies üblicherweise bei *Komplexen Inhaltsentitäten* beachtet wird. So müssen etwa Komposita je nach Vorgabe einer expliziten Syntax – ggf. unter Heranziehung einer Facettierung – als synthetisierte Inhaltsentitäten betrachtet und nicht zu den Bestandteilen der Dokumentationssprache als Begriffe gerechnet werden, die nach Maßgabe der Kompositionsbildung spezifizierte Begriffe sind. In diesem Verständnis steht der Ausdruck *Insektenatmung* als sprachliche Verschmelzung der beiden kategorial unterscheidbaren Begriffe *Insekt* (Facette Tier) und *Atmung* (Facette Physiologie) nicht für einen Unterbegriff zu *Atmung*, sondern für eine synthetische Inhaltsentität *Atmung von Insekten*, die gemäß der Regeln zur Kompositionsbildung der deutschen Sprache auch als *Insektenatmung* repräsentiert werden kann. Die kategoriale Trennung erlaubt nachfolgend die dokumentbezogene Synthese mit anderen Entitäten und macht die Repräsentation entsprechender komplexer Dokument-Gegenstände unabhängig vom evtl. Vorhandensein entsprechender präkombinierter Elemente.

Wie bereits ausgeführt, enthalten konkrete Dokumentationssprachen häufig jedoch auch präkombinierte Inhaltsentitäten als Elemente, die im hier gegebenen Modell als *Komplexe Inhaltsentitäten (KI)* interpretiert werden müssen. Daher lautet die vollständige Beschreibung *Synthetischer Inhaltsentitäten*:

$$SI = EI_1 \otimes EI_2 \otimes \dots \otimes EI_n \otimes KI_1 \otimes \dots \otimes KI_m.$$

Neben den  $EI_i$  und  $KI_j$  tragen auch die Verknüpfungen  $\otimes$  zum Gesamtinhalt bei, sodass aus derselben Menge Komponenten durch verschiedene Syntaxvarianten unterschiedliche Synthetische Inhomogenitäten erzeugt werden können:

$$SI^1 = EI_1 \otimes_1 EI_2 \neq EI_1 \otimes_2 EI_2 = SI^2.$$

Die Repräsentation komplexer Dokumentinhalte erfordert synthetische Notationen, gebildet aus den elementaren Klassen der jeweiligen Facetten, z.B.

$$\text{DokInh}_1 = K^1_{3(m_1 m_3)} \otimes K^2_{3(m_j m_i)} \otimes K^n_{3(m_r m_s m_t)}$$

## 14. Interoperabilitätstypen zwischen Inhomogenitäten

Gemäß der vorstehenden Typisierung von Inhomogenitäten können nun Typen von Beziehungen zwischen den Inhomogenitäten verschiedener Dokumentationssprachen unterschieden werden, um zu einer Beschreibung von *Semantischer Interoperabilität* zu gelangen, die inhaltliche Vergleichs- oder Austauschbarkeit der Elemente zweier oder mehr Dokumentationssprachen berücksichtigt. Die Inhomogenitäten einer Dokumentationssprache sind dabei einerseits durch einen mehr oder weniger formal präzisierten semantischen Gehalt und andererseits durch ein mehr oder weniger stark differenziertes Beziehungsinventar charakterisiert.

### 14.1 Semantische Interoperabilität 1. Stufe (Begriffliche Interoperabilität)

Die einfachste Form ist die Beziehung zwischen zwei *Einfachen Inhomogenitäten*, nachfolgend auch *Semantische Interoperabilität 1. Stufe* genannt. Im Allgemeinen wird hierbei ein Muster von Gleichheit, Ähnlichkeit oder Aufeinanderbezogenheit entwickelt, das sich am Merkmalsbesitz der beteiligten Begriff hinsichtlich Inhalt und Umfang orientiert und sowohl auf den semantischen Gehalt als auch das Beziehungsgeflecht der Inhomogenitäten Bezug nimmt.

Die häufig zu beobachtende Schwierigkeit des Vergleich einer *Einfachen Inhomogenität* mit einer präkombinierten Inhomogenität, erfährt in diesem Modell seine Bestätigung. Es wird erkennbar, dass es sich dabei um den Versuch handelt, einen Vergleich zwischen *Einfachen* und *Komplexen Inhomogenitäten* anzustellen. Im Hinblick auf Gesichtspunkte einer inhaltlichen Interoperabilität kann ein solcher Vergleich niemals mit einer Gleichheitsaussage enden, das Ergebnis kann allenfalls eine Aussage über Stufen einer Übereinstimmung sein.

### 14.2 Semantische Interoperabilität 2. Stufe (Thematische Interoperabilität)

Die *Semantische Interoperabilität 2. Stufe* (auch *Thematische Interoperabilität*) soll als inhaltlicher Vergleich *Synthetischer Inhomogenitäten* verstanden werden. Komplexe Inhomogenitäten können bei der Betrachtung unberücksichtigt bleiben, da Interoperabilitätsbetrachtungen nur im dokumentations sprachlichen Kontext eine Rolle spielen und wir die Komplexität auf dieser Ebene durch die *Synthetischen Inhomogenitäten* beschreiben (vgl. Abb.41).

Synthetische Inhaltsentität 1      Synthetische Inhaltsentität 2

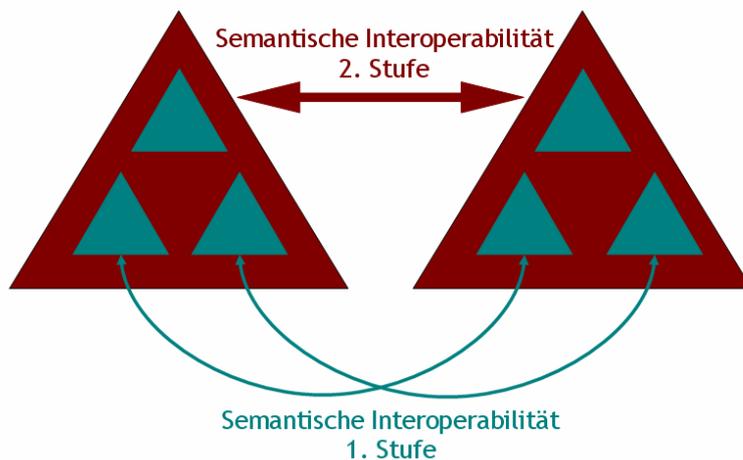


Abb.41: Modell der Semantischen Interoperabilität 2. Stufe

Die Feststellung von Gleichheit, Ähnlichkeit oder Differenz der beteiligten *Synthetischen Inhaltsentitäten* soll im Modell auf die entsprechenden Aussagen über die *Einfachen Inhaltsentitäten* zurückgeführt werden. Mit dieser Vorgehensweise wäre zunächst der Fall behandelt, der die Synthese als additiv aus Komponenten aufgebaut denkt. Außer für den Fall einer komponentenweisen Gleichheit ergeben sich im Allgemeinen aber schon hier verschiedene Fälle von Unterschiedlichkeit. Einige Beispiele sollen diese Sachverhalte formal verdeutlichen.

Dazu seien  $SI^1$  und  $SI^2$  je eine *Synthetische Inhaltsentität*:

*Beispiel 1:*

$$SI^1 = EI^1_1 \otimes EI^1_2 \otimes EI^1_3,$$

$$SI^2 = EI^2_1 \otimes EI^2_2 \otimes EI^2_3.$$

Sofern Interoperabilität in allen Komponenten festgestellt wird und auch die Verknüpfungen vergleichbar sind, wird man die Interoperabilität der beiden synthetischen Identitäten feststellen.

*Beispiel 2:*

$$SI^1 = EI^1,$$

$$SI^2 = EI^2_1 \otimes EI^2_2.$$

Bei diesem Beispiel handelt es sich um den Prototypen des bereits diskutierten Sachverhalts, *einfache* oder *präkombinierte* (also komplexe) *Inhaltsentitäten* einer Dokumentationsprache hinsichtlich ihrer Interoperabilität zu vergleichen. Im Allgemeinen kann dieses Ergebnis nicht vollständig positiv lauten, da inhaltlich meist eine Enthaltensein-Situation vorliegen dürfte.

*Beispiel 3:*

$$SI^1 = EI^1_1 \otimes EI^1_2 \otimes EI^1_3,$$

$$SI^2 = EI_1^2 \otimes EI_2^2 \otimes EI_3^2; \text{ mit } EI_1^1 = EI_1^2, EI_2^1 > EI_2^2, \text{ aber } EI_3^1 < EI_3^2.$$

Durch die Zeichen  $>$  (bzw.  $<$ ) soll ein begrifflicher Unterschied ausgedrückt sein, der die Interoperabilität zwischen den jeweiligen Komponenten verneint. Wie ist die Konsequenz für die Interoperabilität der synthetischen Entitäten zu bewerten? Hebt sich etwa der Richtungswechsel auf? Formal handelt es sich um eine unentscheidbare Situation, die formale Modellierung hat mit diesem Fall die Grenzen ihrer Ausdrucksfähigkeit erreicht. Maschinelle Ansätze zur Interoperabilitätsfeststellung wäre dieser Fall ebenfalls als Grenze gesetzt. Eine Klärung kann allein inhaltlich erfolgen. Im Allgemeinen dürfte im Fall des Beispiels 3 inhaltlich eine Überlappungssituation vorliegen.

Durch Kombination der gegebenen Beispiele lassen sich weitere Fälle konstruieren, die das volle Spektrum der bei Interoperabilitätsüberlegungen anzutreffenden Enthaltenseins- und Überlappungssituationen abdecken und einen weiteren Beitrag für die Grenzen maschineller Verarbeitung liefern.

Für die Feststellung von Gleichheit oder Verschiedenheit *Synthetischer Inhaltsentitäten* kann die Rückführung auf *Einfache* oder *Komplexe Inhaltsentitäten* allerdings keine vollständigen Aussagen garantieren. Wie bereits oben festgestellt wurde, lassen sich mit den gleichen Komponenten unterschiedliche Synthesen erzeugen, indem verschiedene Syntaxelemente eingesetzt werden, z.B.:

$$SI^1 = EI_1^1 \otimes_1 EI_2^1 \neq EI_1^1 \otimes_2 EI_2^1 = SI^2$$

$\otimes_1$  und  $\otimes_2$  sind dabei zwei unterschiedliche Syntax-Konnektoren (z.B. für gerichtete Handlungen), die bei sonst gleichen Inhaltsentitäten unterschiedliche Sachverhalte generieren (Muster: Hund beißt Briefträger – Briefträger beißt Hund). Bei entsprechender Modellierung wäre dieser Fall allerdings einer maschinellen Verarbeitung zugänglich.

Vollständige Interoperabilitätsfeststellungen müssen im Allgemeinen also auch die eingesetzte Syntax der jeweiligen Dokumentationssprache berücksichtigen.

### 14.3 Interoperabilität zwischen Begriffen und Klassen verschiedener Komplexitätsstufen

Die vorstehend zunächst an Begriffen orientierten Überlegungen können nun als Frage nach der Interoperabilität als Gleichheit oder Verschiedenheit zwischen *Begriffen* und *Klassen* übertragen werden. Neben den semantischen Eigenschaften der zu berücksichtigenden Entitäten kommen dabei auch Struktureigenschaften der jeweiligen Bezugssysteme zum tragen. Es müssen verschiedene Fälle unterschieden werden:

- Interoperabilität zwischen *Begriffen*
- Interoperabilität zwischen *Klassen*
- Interoperabilität zwischen *Begriffen* und *Klassen*

Die Interoperabilität zwischen Begriffen wurde bereits vorstehend behandelt. Werden Klassen so verstanden, dass sie Begriffe auf der Basis eines gemeinsamen klassifikatorischen Merkmals zusammenfassen, so kann die Frage der Interoperabilität zwischen Klassen auf die Interoperabilität zwischen Begriffen zurückgeführt werden. Insbesondere gewinnt man so ein Modell für die Frage der Interoperabilität präkombinierter Klassen, indem man sie hinsichtlich Interoperabilität wie synthetische Inhaltsentitäten betrachtet.

Wir gehen also von der Situation aus, dass Einfache Begriffe innerhalb der Dokumentations-  
sprache formal durch *Elementare Inhaltsentitäten* (*Elementardeskriptoren* oder *Elementare  
Klassen*) repräsentiert werden. Die Repräsentanten aller anderen Aussagen sind *Komplexe  
Inhaltsentitäten* und werden innerhalb der Dokumentations-  
sprache formal durch *Synthetische  
Inhaltsentitäten*, *Präkombinierte Klassen* oder *Synthetische Klassen* repräsentiert.

Interoperabilitätsbetrachtungen zwischen Klassen verschiedener Systematiken – und erst recht  
zwischen Entitäten einer verbalen Dokumentationssprache und Klassen einer Systematik –  
müssen die Unterschiede in den Prinzipien des Klassenbildungsprozesses berücksichtigen. Die in  
der vorstehenden Diskussion aufgetretenen verschiedene Typen von Klassen lassen sich in ge-  
wisser Weise nach dem Grad ihrer Komplexität charakterisieren. Als *Elementare Klassen*  
können die Klassen angesehen werden, die nach dem Muster der Abb.27 – Abb.29 durch Be-  
nutzung eines klassifikatorischen Merkmals aus nur einem Aspektbereich hervorgegangen sind.  
Alle anderen Klassen sollen als *Komplexe Klassen* bezeichnet werden. Komplexität meint dabei  
die Benutzung mehrerer klassifikatorischer Merkmale aus mehreren Aspektbereichen. Die  
Bildung solcher komplexer Klassen kann verschiedene Ursachen haben. Beispielsweise kann  
gewünscht werden, sie als festen Bestandteil des Systems zu verankern, weil man weiß oder ver-  
mutet, dass sie genau in dieser Komplexität häufiger verwendet werden. In diesem Fall wären es  
*Präkombinierte Klassen*. Für die Frage, ob eine derartige Klasse als interoperabel zu einer  
anderen Klasse oder einer verbalen Entität angesehen werden kann, muss eine solche Klasse als  
aus verschiedenen Inhaltsentitäten zusammengesetzt gedacht werden (vgl. Abb.42).

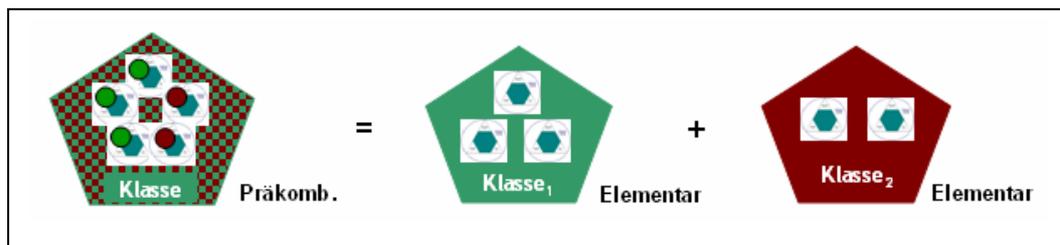


Abb.42: Präkombinierte Klassen als komplexe Zusammensetzungen

Damit entsprechen aus Interoperabilitätsbetrachtungen die präkombinierten Klassen den  
*Synthetischen Klassen*, die als dokumentspezifisch orientierte Klassen zur koextensiven Inhalts-  
repräsentation gebildet werden (vgl. Abb.43 und 44).

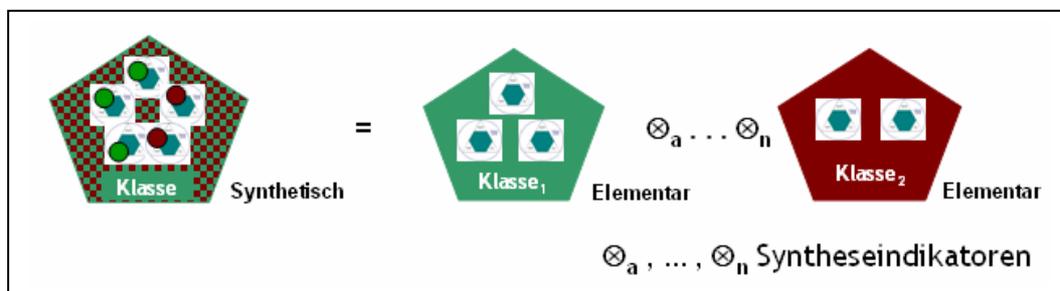


Abb.43: Synthetische Klassen als Zusammensetzungen elementarer Klassen

Oder allgemein:

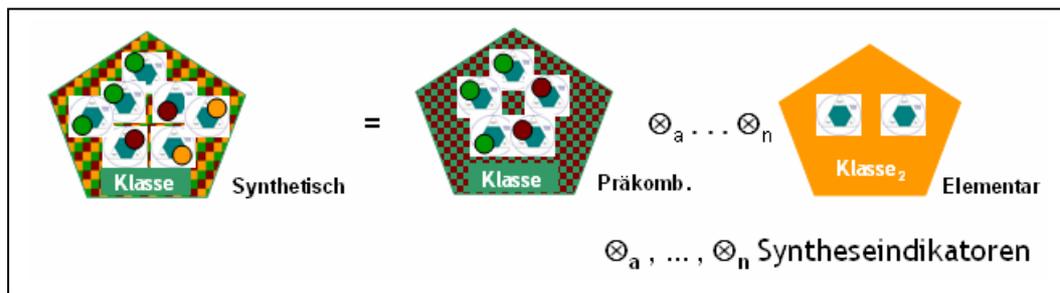


Abb.44: Synthetische Klassen als Zusammensetzungen elementarer oder präkombinierter Klassen

Sofern zwischen den Facetten syntaktische Verbindungen hergestellt werden sollten, um verschiedene Syntagmata auszudrücken, bedarf es wie schon im diskutierten allgemeinen Fall des Einsatzes von zusätzlichen Indikatoren:

$$\text{DokInh}_1 = K^1_4(m_1 m_3 m_4) \otimes_1 K^2_3(m_j m_l) \neq K^1_4(m_1 m_3 m_4) \otimes_2 K^2_3(m_j m_l) = \text{DokInh}_2$$

Im Unterschied zu den präkombinierten Klassen legen synthetische Klassen aber ihre Synthesestellen und Regeln zur Synthese durch Schlüsselungs- oder Facettenindikatoren offen und können damit auch eine bessere Basis für Interoperabilitätsentscheidungen anbieten, wie zuvor schon bei geeigneter Repräsentation auch für maschinelle Entscheidungen.

## 15. Die pragmatische Dimension bei Begriffs- und Klassenbildung sowie Konsequenzen für Interoperabilität

Das bisher vorgestellte Modell zur dokumentations sprachlichen Repräsentation von inhaltlichen Aussagen als Inhaltselementen auf der Basis von Begriffselementen oder daraus aufgebauten synthetischen Verbindungen muss bei genauerer Betrachtung durch zwei Eigenschaften charakterisiert werden, für deren Unabdingbarkeit keine zweifelsfreie Begründung gegeben werden kann:

1. Es wird die Existenz von elementaren Bausteinen und das Vorhandensein von Möglichkeiten zum Aufbau von Aussagen durch ein Inventar syntaktischer Elemente vorausgesetzt;
2. Es wird eine stationäre Beschreibung ohne Integration dynamischer Elemente zur Berücksichtigung des Gebrauchs der Begriffe und entsprechender Veränderungen gegeben.

Beide Eigenschaften kann man zunächst als notwendig für eine Formalisierung des Modells ansehen. Es ist jedoch interessant, nach Leitlinien zu suchen, in welche Richtung die Modelle weiter entwickelt werden müssten, um sie von den genannten Einschränkungen zu befreien.

Die Fundierung des Modells auf fixierbare Entitäten, die Vorstellung der Möglichkeit, Begriffe durch Angabe ihrer Eigenschaften zu charakterisieren und Klassen durch Selektion von Begriffen mit gemeinsamen Merkmalen zu bilden, wird man so schnell nicht aufgeben können. Mindestens für die maschinelle Verarbeitung als adressierbare Entitäten scheint eine solche Vorstellung nicht verzichtbar. Empfindlich ist jedoch der Verzicht auf den Aspekt der zeitlichen Veränderung im Verständnis der Entitäten.

Bei der Gestaltung von Dokumentations sprachen wie Thesauri oder Klassifikationssystemen werden verschiedene Verfahrensweisen besprochen, um die Zeitgebundenheit der jeweiligen Entitäten zu beschreiben. Üblich ist beispielsweise eine Datumsangabe bei Einführung oder

Löschung des entsprechenden Deskriptors oder der Klasse. In Klassifikationssystemen werden Verbindungen zwischen den Notationen verschiedener Versionen häufig durch Konkordanztabellen hergestellt. In beiden Fällen werden die Deskriptoren oder Klassen manchmal durch eine eigene Form der Relation miteinander verbunden. In der *Schlagwortnormdatei (SWD)* erfolgt dies z.B. durch die sog. *Chronologische Form (CF)*, die allerdings auf die Verwendung für Geografika und Körperschaften begrenzt ist und nicht für den wünschenswerten Verwendungs- und Bedeutungsgebrauch von Sachbegriffen eingesetzt werden kann.

Wichtig werden alle diese Überlegungen im Zusammenhang mit der Einführung permanenter Adressen von dokumentations sprachlichen Elementen in der Architektur des Semantic Web.<sup>38</sup>

Es wurde bereits der Vorschlag gemacht, die im Begriffsgebrauch sichtbar werdende pragmatische Dimension als vierten Bestandteil in der Modellbildung einer Begriffstheorie zu berücksichtigen und das semiotische Dreieck zu einem semiotischen Tetraeder zu erweitern. Abb.45 zeigt eine grafische Form des semiotischen Tetraeders in Anlehnung an die Vorschläge der *FRISCO*-Gruppe<sup>39</sup>.

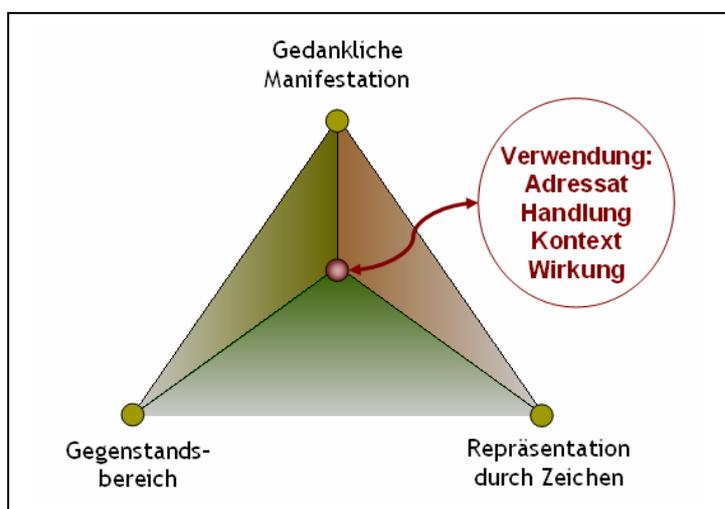


Abb.45: Semiotischer Tetraeder

Im Rahmen unserer Erörterungen ist nun interessant, ob aus dieser Erweiterung Konsequenzen für Interoperabilitätsbetrachtungen abgeleitet werden können. Als erste Konsequenz kann zwischen *diachronischem* und *synchronischem Verwendungunterschied* unterschieden werden. In diesem Fall würde die vierte Ecke des Tetraeders als Repräsentant für den pragmatischen Verwendungswandel über den Zeitablauf interpretiert.

<sup>38</sup> Vgl. z.B. das Konzept für die DDC: **Panzer, M.:** *Cool URIs for the DDC: towards Web-scale accessibility of a large classification system.*

In: *Metadata for semantic and social applications : proceedings of the International Conference on Dublin Core and Metadata Applications, Berlin, 22 - 26 September 2008, DC 2008: Berlin, Germany / ed. by Jane Greenberg and Wolfgang Klas. Göttingen: Univ.-Verl. 2008. S.183-190.*

<sup>39</sup> **Hesse, W.:** *Information: das Soma des ausgehenden Jahrhunderts?*

In: *Ethik und Sozialwissenschaften. 9(1998) H.2, S.212-215.*

**Hesse, W., A. Verrijn-Stuart:** *Towards a theory of information systems: the FRISCO approach.*

In: [http://public.beuth-hochschule.de/~giak/arbeitskreise/softwaretechnik/publikationen/EJC\\_00-fin.pdf](http://public.beuth-hochschule.de/~giak/arbeitskreise/softwaretechnik/publikationen/EJC_00-fin.pdf).

Vgl. auch: <http://cs-exhibitions.uni-klu.ac.at/index.php?id=445>.

FRISCO = FRamework of Information System COncepts.

Damit lässt sich eine Überlegung in Analogie zur Behandlung von Quasi-Synonymen formulieren: Gibt die pragmatische Dimension Anlass, eine inhaltlich so wesentlich veränderten Begriff anzunehmen, dass er auch durch eine neue Inhaltsentität repräsentiert werden muss oder verbleibt die Veränderung im Rahmen eines Synonym-Umfeldes? Wenn eine neue Inhaltsentität ausgeprägt wird, welche Art von Relation wird zwischen den dann verschiedenen Inhaltsentitäten hergestellt? In Abb.46 sind die beiden Vorgehensweisen grafisch dargestellt.

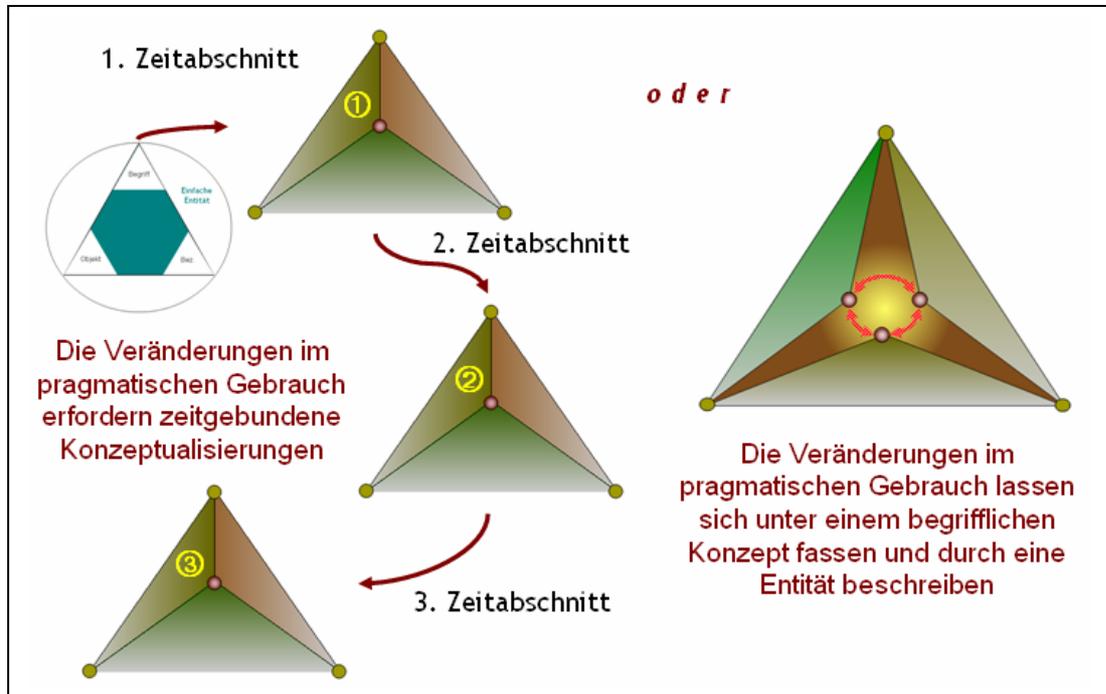


Abb.46: Zeitbindung der pragmatischen Begriffsverwendung

Der Fall des *synchronischen Verwendungsunterschieds* lässt sich nicht in gleicher Weise behandeln, da hierbei ja ein Nebeneinander zur gleichen Zeit stattfindet. Dieser Fall entspricht also der Entwicklung unterschiedlicher Begriffe auf der Basis verschiedener Aspekte und ihrer Repräsentation durch unterschiedliche Entitäten, wie es schon im Fall der aspektorientierten Klassenbildung diskutiert wurde (vgl. Abb.36), hier neu dargestellt in Abb.47.

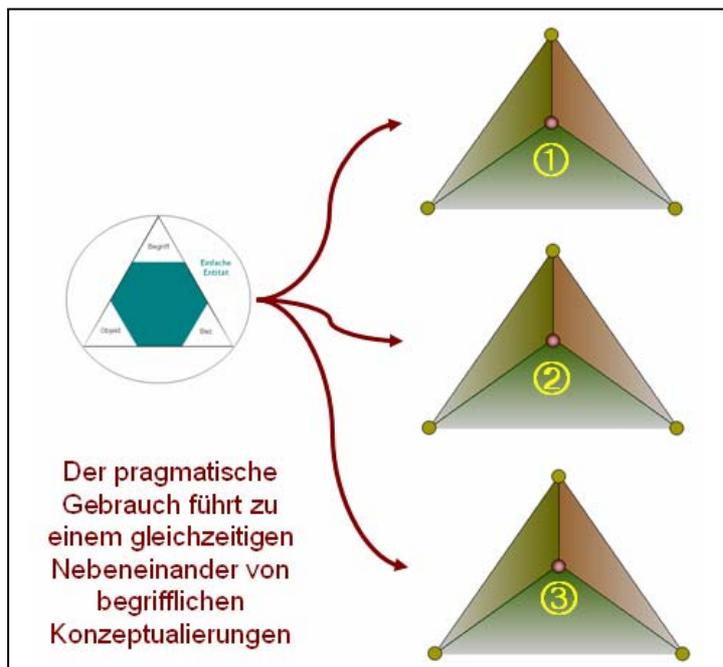


Abb.47: Synchrones Nebeneinander pragmatisch unterschiedener Inhaltsentitäten

Das Ausweisen dieses Nebeneinanders von pragmatisch orientierten Konzeptualisierungen würde es gestatten, in der Erschließung genauer Sichtweisen zu berücksichtigen, wie sie etwa von bestimmten Schulen entwickelt wurden. Bislang ist jedoch weder dies noch die Verbindung der Entitäten durch eigene Relationen übliche Praxis bei der Gestaltung von Dokumentations-sprachen oder im Erschließungsprozess. Es muss zunächst offen bleiben, konkrete Modelle dafür anzugeben, wie aus der Berücksichtigung der pragmatischen Dimension Nutzen in einem RetrievalszENARIO gezogen werden könnte. Ein erster Ansatzpunkt könnte das Ausweisen einer spezifisch typisierten Relation und die Bildung von Treffermengen entlang dieser Relation sein. Dieser Ansatz würde sich in die Diskussionen des Abschnitt 6 einbetten, in dem der Nutzen einer Typisierung von Relationen in erweiterten semantischen Netzen zur Wissensrepräsentation vorgestellt wurde.