

# Ein Ontologie basiertes Modell für Indexierung und Retrieval

Winfried Gödert

Fachhochschule Köln, Institut für Informationswissenschaft

In diesem Beitrag wird ausgehend von einem ungelösten Problem der Informationserschließung ein Modell vorgestellt, das die Methoden und Erfahrungen zur inhaltlichen Dokumenterschließung mittels kognitiv zu interpretierender Dokumentationssprachen mit den Möglichkeiten formaler Wissensrepräsentation verbindet. Die Kernkomponente des Modells besteht aus der Nutzung von Inferenzen entlang der Pfade typisierter Relationen zwischen den in Facetten geordneten Entitäten innerhalb einer Wissensrepräsentation zur Bestimmung von Treffermengen im Rahmen von Retrievalprozessen. Es werden die möglichen Konsequenzen für das Indexieren und Retrieval diskutiert.

## 1. Einführung

Inhaltliche Erschließung wird nicht mehr als ein Gebiet mit großem Potenzial für methodische Innovationen gesehen. Bei den in der Praxis eingesetzten Verfahren ist ein Abschied von elaborierten Methoden der intellektuellen Indexierung zur Repräsentation des im Dokument vorhandenen Inhalts, der Aboutness zu verzeichnen. Einen breiten Erhalt intellektueller Erschließung gibt es allenfalls im Umfeld des Social tagging auf einer vergleichsweise niedrigen Ebene der Differenzierung der zu repräsentierenden Inhalte.

Anders stellt sich die Situation auf der Seite des Information Retrieval dar. Hier lässt sich eine Entwicklung von Methoden und Verfahrensweisen verzeichnen, die das Grundmodell des Booleschen Retrievals um verschiedenen Komponenten wie Link-topologische oder probabilistische Ansätze erweitern und so zu Verfeinerungen der Treffermengenbildung und -anzeige gelangt, die ohne menschliche Arbeitsschritte algorithmisch erreicht werden können.

Weniger Entwicklung hat auf der Seite der Indexierung stattgefunden. Die Suche nach automatisierbaren Vorgängen hat zu Methoden zur Identifikation von Token in Kombination mit linguistischen und statistischen Vorgehensweisen geführt. Die Übereinstimmung der ermittelten Token mit Wörtern, Begriffen oder thematischen Kontexten nimmt mit der Reihenfolge der vorstehend genannten Konzepte ab und begrenzt den Findeerfolg häufig auf Formen eines zeichenorientierten Matchings zwischen Sucheingabe und gespeicherten Daten. Die in der jüngeren Vergangenheit entwickelten Methoden mit der besten Eignung zur Disambiguierung im Indexierungsergebnis und dem höchsten Potenzial zur Bildung präziser Treffermengen, Kombinationen aus Facettenansätzen mit syntaktischer Indexierung, sind wegen ihres intellektuellen Aufwandes am wenigsten Bestandteil der aktuellen Informationssysteme geworden. Präzisierung im Indexierungsergebnis wird durch Kombinationen linguistischer und statistischer Verfahren zur Bildung und zur gerankten Präsentation von Treffermengen ersetzt.

Es stellt sich die Frage, ob die ehemals als Probleme gesehenen Aufgaben der inhaltlichen Erschließung als gelöst angesehen werden dürfen. Ein näherer Blick zeigt, dass dem nicht so ist. Es lohnt daher, eine Identifikation der noch nicht gelösten Probleme der inhaltlichen Dokumenterschließung vorzunehmen. Damit soll die Frage verbunden werden, welche Chancen gesehen werden dürfen, sie durch weitere Verbesserungen auf der Retrievalseite zu lösen bzw. welche Probleme voraussichtlich allein durch Maßnahmen auf der Erschließungsseite gelöst werden können.

Einfache Beispiele können eine Illustration der offenen Fragen geben. Betrachten wir etwa die beiden Suchinteressen:

- Gesucht werden Dokumente über den *Wandertrieb von Singvögeln*
- Gesucht werden Dokumente über *Singvögel mit Wandertrieb*

Welche Unterschiede bestehen zwischen den beiden Fragestellungen mit Relevanz für eine semantische Repräsentation, die Indexierung sowie das Retrieval? In der ersten Fragestellung sind *Singvögel* und *Wandertrieb* behandelte Gegenstände und es liegt eine gegenseitige Bezugnahme vor. Zum präzisen Auffinden des Themas ist also die Indexierung *beider* Begriffe und eine Kennzeichnung der zwischen ihnen bestehenden Beziehung *im* Dokument erforderlich, um eine *koextensive* Inhaltsrepräsentation zu erreichen. In der Sprache der Indexierungsmethoden heißt dies, dass eine Form *syntaktischen Indexierens* und der Einsatz von *Syntaxoperatoren* für das Retrieval erforderlich ist. Allein die Methode des *gleichordnenden Indexierens* sowie rein *postkoordinierendes Retrieval* kann die entsprechende Präzision nicht garantieren<sup>1</sup>. Um die beiden Begriffe indexieren zu können, müssen sie als Elemente in einer entsprechenden Dokumentationssprache enthalten sein.

Im zweiten Suchinteresse soll der *Wandertrieb* nicht behandelter Gegenstand sein, darf also für die zu findenden Dokumente nicht indexiert worden sein. In diesem Fall ist der *Wandertrieb* nur einschränkende Eigenschaft bei der Auswahl der für eine Treffermengenbildung zu selektierenden Arten von *Singvögeln* und damit der für die Suche zu berücksichtigenden Elemente der Dokumentationssprache. Die Verankerung einer solchen Eigenschaft kann allein in der Wissensrepräsentation erfolgen und muss durch eine geeignete Gestaltung der Retrievalumgebung für einen Suchvorgang abrufbar gemacht werden.

Verallgemeinert kann die Situation unserer Beispiele folgendermaßen beschrieben werden: Es werden Dokumente über Gegenstände gesucht, indem Wissen (Eigenschaften) über diese Gegenstände als Selektionskriterium eingebracht wird, das nicht indexiert werden darf, weil es nicht behandeltes Thema des Dokuments ist.

Im Rahmen der klassischen Indexierungs- und Retrievalmodelle lässt sich eine derartige Unterscheidung nicht vornehmen. Folgende Eigenschaften kennzeichnen diese Modelle: Das Finden von Datensätzen wird an das Vorhandensein von Indextermen gebunden. In geringem Umfang wird die Struktur der verwendeten Dokumentationssprache - in der Regel die Hierarchie - genutzt, um Sammelböden von Indextermen für die Suche zu bilden. Hilfsmittel zur Unterscheidung der beiden genannten Fragestellungen werden nicht angeboten. Nach derzeitiger Kenntnis ist kein Weg sichtbar, Methoden der automatischen Text-Analyse mit statistischen oder linguistischen Ansätzen zur Behandlung derartiger Probleme zu nutzen. Unsere Frage wird daher sein, ob sich die genannte Unterscheidung durch eine Verbindung von Wissensrepräsentation mit Indexierung und Retrieval lösen lässt?

Unsere Antwort besteht aus einem Ja, indem eine Erweiterung des dokumentbezogenen Indexierungs- und Retrievalmodells vorgenommen wird. Diese Erweiterung kann charakterisiert werden als eine Kombination aus

---

<sup>1</sup> Wir verzichten darauf, alle im Text erwähnten grundlegenden Fachausdrücke näher zu erläutern und verweisen statt dessen auf die einschlägige Basisliteratur, z.B.: Gödert, W., K. Lepsky u. M. Nagelschmidt: Informationserschließung und Automatisches Indexieren: ein Lehr- und Arbeitsbuch. Berlin: Springer, 2011.

- Indexierung von Dokumenten ist eine Angabe über die *Aboutness* von Dokumenten, d. h. ein Indexierungsterm wird nur zugeteilt, wenn das Dokument von dem entsprechenden Sachverhalt handelt;
- Die Wissensrepräsentation enthält "Wissen" durch die Angabe von Eigenschaften der enthaltenen Entitäten, d. h. in der Wissensrepräsentation sind spezifische Zusammenhänge zwischen den Begriffen durch die Angabe von Eigenschaften und Relationen hinterlegt;
- Inferenzen über das hinterlegte Wissen und die Indexierungsterme werden zur Bildung von Treffermengen genutzt, d. h. es werden nicht allein die indexierten Entitäten, sondern auch die Beziehungen zwischen ihnen zur Basis gemacht, um Treffermengen zu bilden.

## 2. Das Ontologie basierte Modell für Indexierung und Retrieval

Wir legen unseren Überlegungen eine Ausgangssituation zugrunde, die durch die Abb. 1 veranschaulicht wird. Auf einzelne Bestandteile wird in der nachfolgenden Diskussion Bezug genommen.

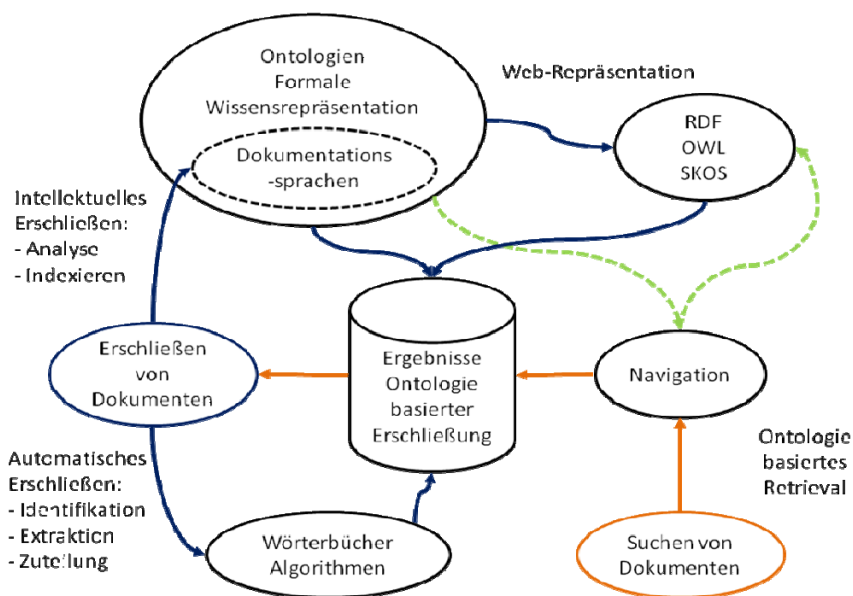


Abb. 1: Ontologie-basiertes Konzept für Erschließung und Retrieval

Zur Beschreibung unseres Modells ist es erforderlich, eine nähere Betrachtung

- typisierter Relationen in der Wissensrepräsentation zur Modellierung spezifischer Beziehungen zwischen den Begriffen
- und der Formen von Inferenzen im Information Retrieval zur Bildung von Treffermengen

durchzuführen und beide Gesichtspunkte miteinander zu verknüpfen. Dies wird nachfolgend geschehen. Wir knüpfen dabei an eine Darstellung an, die bereits an anderer Stelle gegeben wurde<sup>2</sup>.

Zur Vereinfachung der Sprechweise treffen wir nachstehende Vereinbarungen. Die Menge aller strukturierten terminologischen Repositorien (Thesauri, Klassifikationssysteme), die für Zwecke der Indexierung genutzt werden können, soll als *Wissensrepräsentationen* bezeichnet und im Folgenden nur dann näher unterschieden werden, wenn es der Argumentationskontext erfordert. Die Wissens-

<sup>2</sup> Gödert, W.: Semantische Wissensrepräsentation und Interoperabilität: Teil 1: Interoperabilität als Weg zur Wissensexploration. Teil 2: Ein formales Modell semantischer Interoperabilität. In: Information - Wissenschaft und Praxis. 61(2010) H.1, S.5-18 (T.1) S.19-28 (T.2).

repräsentationen bestehen aus Elementen - *Entitäten* genannt - und einer Struktur, die durch die zwischen den Elementen ausgewiesenen Beziehungen bestimmt ist. Die Verbindung mit der Sicht traditioneller Dokumentationssprachen macht es erforderlich, eine Vereinbarkeit zwischen intellektueller Interpretation der Entitäten auf der einen Seite und ihrer formalen definitorischen Charakterisierungen durch Angabe von Merkmalen und relationalen Eigenschaften auf der anderen Seite anzustreben. Als strukturbildende Eigenschaften werden die üblichen Arten von Begriffsbeziehungen im Zusammenwirken mit Aspektsichten zur Herstellung facettierter Strukturen gesehen.

## 2.1 Inferenzen als Hilfsmittel des Information Retrieval

Die Benutzung eines Indexterms zur Bildung von Treffermengen kann als einfachste Form eines Schlussfolgerns gesehen werden. Weitere Beispiele für Schlussfolgerungsprozesse werden durch die Benutzung der bekannten Werkzeuge des Information Retrieval wie etwa Trunkierung oder die Benutzung der Booleschen Operatoren geboten. Eine Veranschaulichung dieser Situationen bietet die Abb. 2.

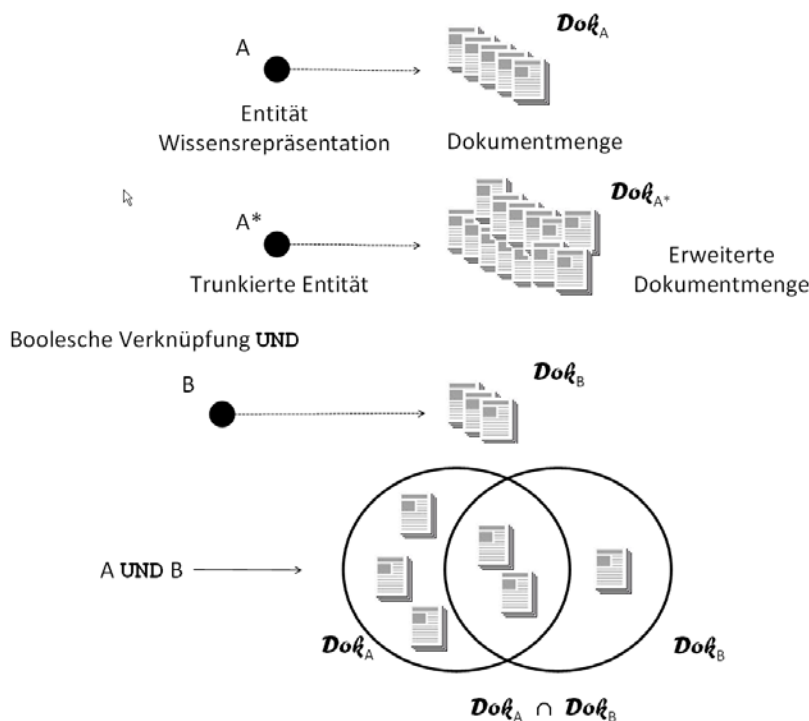


Abb. 2: Beispiele für Inferenzoperationen zur Bildung von Treffermengen

Ausgehend von einem Operieren auf der Menge der Entitäten wird durch die Verbindung, die durch den Prozess des Indexierens zur Dokumentkollektion hergestellt wurde, eine Treffermenge für eine Fragestellung erzeugt.

Eine Erweiterung dieses Schlussfolgerns besteht aus Inferenzen entlang der Relationspfade, die in der Dokumentationssprache angelegt sind. Allgemein bekannt ist hierfür die Bildung thematischer Ausschnitte durch Nutzung hierarchischer Zusammenhänge, wie dies die Abb. 3 für ein einfaches Beispiel zeigt.

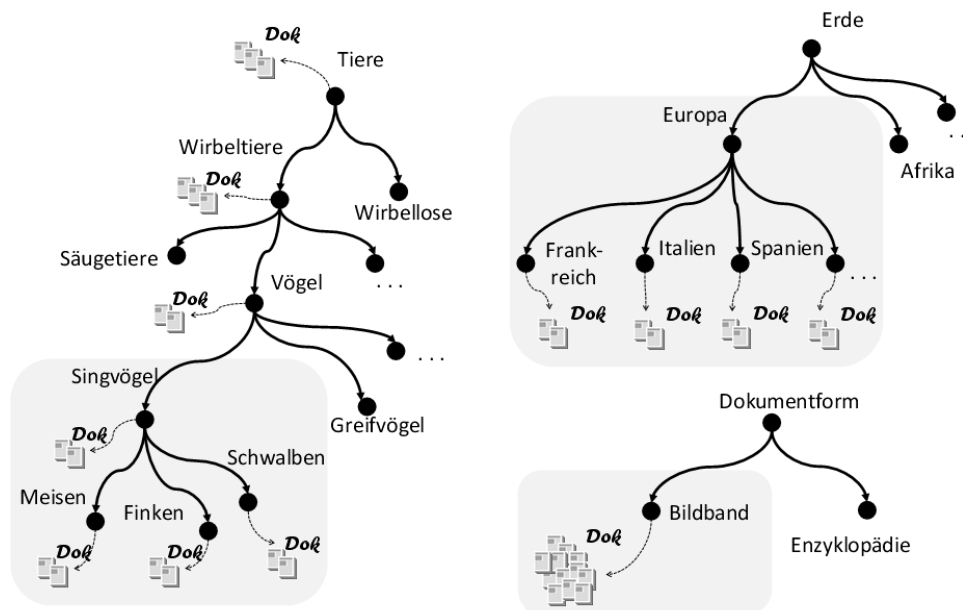


Abb. 3: Beispiele für Facettenbildung und hierarchische Inferenzoperationen

Sind die Entitäten der Wissensrepräsentation aspektorientiert strukturiert, lässt sich die Bildung thematischer Ausschnitte für jeden Aspekt getrennt durchführen und eine Vorbereitung für anschließende aspektorientierte Suchprozesse gestalten. Die Abb. 3 zeigt auch hierfür ein einfaches Beispiel.

Soll sich diese Vorgehensweise über mehrere Stufen der Relationenpfade erstrecken, ist eine saubere Ausarbeitung der entsprechenden Hierarchien erforderlich. Als Kriterien für die begriffliche Strukturierung formuliert, ist diese Anforderung am besten zu erfüllen, wenn keine Polyhierarchien gebildet werden und eine strikte Beachtung des Kriteriums erfolgt, dass jeder hierarchisch untergeordnete Sachverhalt *alle* Merkmale des übergeordneten aufweist. Dann ist gesichert, dass von einer Vererbung der Merkmale über mehrere Stufen eines Relationspfades ausgegangen werden kann.

Die formale Anforderung einer solchen Merkmalsvererbung wird über die Eigenschaft der Transitivität beschrieben. Sie besagt abstrakt, dass bei Vorliegen der gültigen Aussagen  $a \sim b$  und  $b \sim c$  für die drei Elemente  $a, b, c$  und die Relation  $\sim$  geschlossen werden kann, dass auch  $a \sim c$  eine gültige Aussage ist.

Übertragen auf ein Verständnis von Hierarchie zwischen begrifflichen Entitäten bedeutet dies, dass aus

Entität B ist Unterbegriff zu Entität A

und

Entität C ist Unterbegriff zu Entität B

formal folgt:

Entität C ist Unterbegriff zu Entität A.

Es verdient hervorgehoben zu werden, dass diese Eigenschaft der Relation zukommt und nicht inhaltlich an Eigenschaften der Entitäten A, B und C geprüft werden muss. Wenn die Relation die Eigenschaft der Transitivität besitzt, dann gilt sie für alle Entitäten, die innerhalb der Wissensrepräsentation durch die Relation verbunden sind. Damit ist ein wichtiger Unterschied in der Betrachtung von Dokumentationssprachen und formalen Wissensrepräsentationen beschrieben. Im Rahmen von

Dokumentationssprachen werden Zusammenhänge zwischen Entitäten in der Regel inhaltlich an den Eigenschaften der Entitäten festgemacht und weder als Merkmalsbesitz formal ausgewiesen noch als allgemeine Eigenschaft der Relationen gesehen. In real existierenden Dokumentationssprachen trifft man so immer wieder auf die Situation, dass die Transitivität zwischen hierarchisch verbundenen Begriffen mal vorhanden ist und mal nicht. Will man das aus Wissensrepräsentationen bekannte Potenzial des Schlussfolgerns entlang der Relationenpfade auf das Information Retrieval übertragen, kann diese Situation keinen Bestand haben, sondern es müssen die entsprechenden formalen Anforderungen an die Definition der Entitäten und der Relationen beachtet werden. Ziel muss es sein, inhaltlich gewünschte Vererbbarkeit von Merkmalen der Entitäten so auf die formalen Relationsstrukturen zu übertragen, dass Schlussfolgerungsprozesse entlang der Pfade gestaltet werden können, um nachfolgend als allgemeine Retrievaltools eingesetzt zu werden.

Die Eigenschaft der Vererbbarkeit von Merkmalen hat Vorteile für die formale Gestaltung von Begriffsordnungen. Die Abb. 4 zeigt ein einfaches Beispiel für eine Begriffsordnung, in der jede zu berücksichtigende Eigenschaft jedem Begriff spezifisch zugeordnet worden ist.

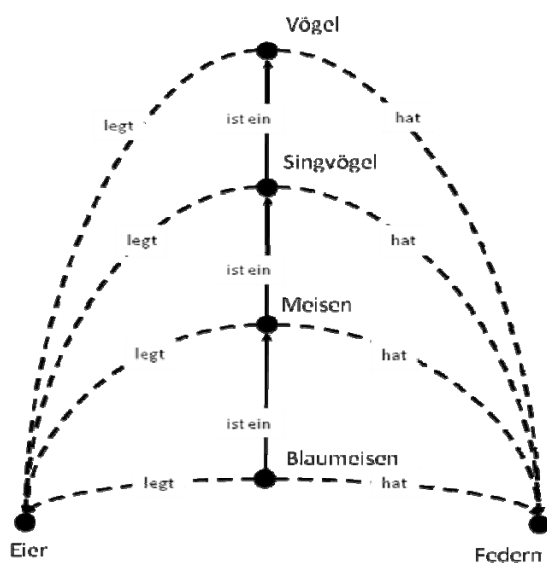


Abb. 4: Beispiel für eine Wissensordnung ohne Vererbung der Eigenschaften

Die Abb. 5 zeigt dieselbe Begriffsordnung in einer formalen Darstellung mit Merkmalsvererbung.

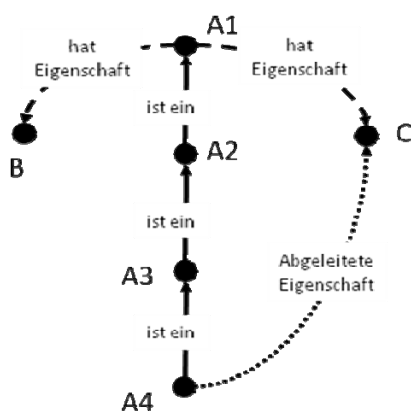


Abb. 5: Wissensordnung aus Abb. 4 mit Merkmalsvererbung

Die Darstellung kann folgendermaßen gelesen werden:

$A_n$  ist ein  $A_{n-1}$

$A_1$  hat Eigenschaft C

$\Rightarrow A_n$  hat Eigenschaft C

Diese Vorgehensweise entspricht der Zielsetzung einer Wissensrepräsentation, das repräsentierte Wissen formal interpretierbar zu machen und über die ausgewiesenen Relationen Schlussfolgerungen ableiten zu können. Aus der Gestaltung von Expertensystemen ist bekannt, dass diese Vorgehensweise an Grenzen stößt - im Beispiel der Abb. 4 etwa zu sehen an einer möglichen Berücksichtigung des Begriffs *Pinguin* - die in Wissensrepräsentationen durch Ausnahmeklauseln behandelt werden müssen. Die Vereinbarkeit zwischen kognitiver Interpretation und Zuweisung formaler Attribute bleibt eine spezifische Herausforderung für die Gestaltung von Wissensrepräsentationen, die für die Zwecke der Indexierung und des Information Retrieval verwendet werden sollen. Unser Ansatz vermeidet dabei die Notwendigkeit, Ausnahmeklauseln angeben zu müssen.

## 2.2 Typisierte Relationen in der Wissensrepräsentation

Ein möglicher Weg zur Verbesserung der formalen Eigenschaften der Relationen in Dokumentationssprachen<sup>3</sup> wird durch eine Erweiterung des Relationenspektrums beschrieben, das als Grundlage für die Beziehungen genommen wird, die zwischen der Entitäten einer Wissensrepräsentation ausgewiesen werden. Ausgangspunkt unserer Diskussion sollen die drei Relationstypen sein, die aus den klassischen Dokumentationssprachen bekannt sind: die *Synonymie*, die *Hierarchie* und die *Assoziation*. Für unsere Diskussion interessiert besonders, welche Aussagen sich über die Möglichkeiten des Schlussfolgerns machen lassen, wenn Relationen gleichen oder unterschiedlichen Typs zu einem Relationspfad verbunden werden. Betrachten wir zunächst die in Frage kommenden Relationen mit ihren jeweiligen Eigenschaften.

### *Synonymie- oder Äquivalenzrelation*

Formale Wissensrepräsentationen machen keinen Unterschied zwischen unterschiedlichen Bezeichnungen einer Entität. Da ihr Aufgabenbereich nicht an Indexierungszwecken orientiert ist, wird keine Unterscheidung in Indexierungs- und Zugangsvokabular benötigt. Der semantische Inhalt einer Entität wird durch jede ihrer sprachlichen Repräsentationen zum Ausdruck gebracht. Die Unterscheidung in Indexierungs- und Zugangsvokabular ist jedoch traditioneller und wichtiger Bestandteil von Dokumentationssprachen und soll im Rahmen unserer Überlegungen beibehalten werden. In die Überlegungen zur Transitivität braucht dieser Relationstyp jedoch nicht spezifisch einbezogen zu werden, da er immer nur als erste Relation eines Pfades vorkommen kann und die Eigenschaften des resultierenden Pfades von den Eigenschaften der anderen beteiligten Relationen abhängen.

### *Hierarchierelationen*

Das Konzept Hierarchierelationen lässt die Unterscheidung in mehrere Arten von Relationen zu. Die *Abstraktionsrelation* ist dabei der Modellfall, der mit der strikten definitorischen Vorgabe verbunden ist, dass ein Unterbegriff *alle* Merkmale des Oberbegriffs sowie mindestens ein weiteres besitzt. Häufig unausgesprochen wird dabei mitgedacht, dass diese weiteren Merkmale einem gemeinsamen

---

<sup>3</sup> Aus der Vielzahl an Darstellungen zur Gestaltung von Dokumentationssprachen und den dabei verwendeten Relationstypen geben wir hier nur an:

ISO 25964: Thesauri and interoperability with other vocabularies: Part 1: Thesauri for information retrieval. Part 2: Interoperability with other vocabularies. Geneva: ISO, 2011- 2013.

Aspektraum oder einer kategorial bestimmten Facette entstammen müssen, wir nennen dies nachfolgend auch den generischen Kontext. Nur die strikte Beachtung dieser Bedingung bietet eine problemfreie Kombination der inhaltlich gedachten begrifflichen Über- und Unterordnung mit der formalen Eigenschaft der Transitivität. Wird der generische Kontext für die Merkmalsbestimmung gewechselt, kommt es zu dem bekannten Phänomen der Polyhierarchien, die häufig in Präkombinationen münden. Als Beispiel hierfür sei genannt:

Tisch, Holztisch, Glastisch, Küchentisch, Wohnzimmertisch, Jugendstiltisch, Schreibtisch, Wickeltisch, Beistelltisch

Jeder Versuch, allein diese Begriffe in einer Monohierarchie zu ordnen, von einer Erweiterung gar nicht zu reden, muss an den angegebenen Gründen scheitern. Daher kann auch eine die Aspekte wechselnde Subordination, wie man sie etwa in präkombinierten Systematiken häufig antrifft, wie sie aber bei extensiver Kompositionsbildung auch in Thesauri vorkommt, keine geeignete Voraussetzung für Inferenzbildungen sein. Eine für Inferenzbildung geeignete Auflösung solcher Strukturen ist allein durch eine Facettierung mit Merkmalskonsistenz innerhalb jedes Hierarchiestrangs möglich. Dies kann dazu führen, dass auch die Abstraktionsbeziehung entsprechend einer Zweckbindung der Wissensstruktur in Typen unterschieden werden muss<sup>4</sup>.

Der zweite wichtige Relationstyp zur Hierarchiebildung wird durch die *Ganzes-Teil-Beziehung* beschrieben. Da im materiellen Sinn ein Teil niemals alle Eigenschaften des zugehörigen Ganzen haben kann, lässt sich dieses Kriterium nicht zur Feststellung einer Transitivität über mehrere Stufen benutzen. Beachtet man den durch den Ausgangsbegriff beschriebenen begrifflichen Kontext für alle beteiligten Begriffe einer Ganzes-Teil-Relation und schließt damit Mehrdeutigkeiten aus, so ist auch der Teil eines Teils (eines Teils ...) eines Ganzen ein Teil des Ganzen. Der Relation kann damit formal die Eigenschaft der Transitivität zugeordnet werden.

Zusätzlich betrachtet werden muss die Bildung von Relationspfaden in Kombination mit anderen Relationen. Dabei wird die Reihenfolge der Pfadbildung eine wichtige Rolle spielen.

Eine lineare Ordnung wird auch über die chronologische Beziehung ausgedrückt, die als *Früher-Später-Beziehung* Bestandteil mancher Dokumentationssprachen ist. Falls Zeiträume gebildet werden, kann daraus ein hierarchischer Zusammenhang abgeleitet werden, der bei gleicher Richtung des Zeitpfeils für transitive Schlussfolgerungen nutzbar ist.

### *Assoziationsrelationen*

Die dritte Gruppe Relationen in Dokumentationssprachen wird durch die *Verwandtschafts-* oder *Assoziationsrelationen* gegeben. Auch wenn sie in Thesauri beispielsweise nur durch ein Relationskürzel vertreten sind, verbirgt sich dahinter in der Regel ein ganzes Bündel unterschiedlicher Typen, die aus Vereinfachungsgründen in der Regel jedoch nicht näher spezifiziert werden. Typisches Merkmal ist der beabsichtigte Nutzen für Retrievalvorgänge, indem die kognitive Interpretation der als Suchvokabular verwendeten Begrifflichkeiten durch Hinweise auf Begrifflichkeiten mit ähnlicher inhaltlicher Bedeutung unterstützt wird. Schon die Bezeichnung des Relationstyps kann als Hinweis auf diese Absicht gesehen werden. Viele der Begrifflichkeiten eines solchen Ähnlichkeitsclusters lassen sich daher als *Quasi-Synonyme* kennzeichnen. Dieses Cluster muss für die begriffliche

---

<sup>4</sup> Vgl. z.B.: Boteram, F.: Semantische Relationen in Dokumentationssprachen: vom Thesaurus zum semantischen Netz. Köln: Fachhochschule, Institut für Informationswissenschaft, 2008. 94 S. (Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft; Bd. 54). Auch unter: <http://www.fbi.fh-koeln.de/institut/papers/kabi/volltexte/band054.pdf>.



Modellierung einer Dokumentationssprachen durch Synonymsetzung oder das Ausweisen als Verwandte Begriffe aufgelöst werden<sup>5</sup>. Entscheidungen über das Vorliegen einer Assoziation zwischen Begriffen werden allenfalls unter Konsistenzgesichtspunkten durch Kriterien unterstützt, nicht aber durch formale Vorgaben. Daher kann im Allgemeinen keine Transitivität entlang der Pfade der Assoziationsrelationen erwartet werden. Analysen in bestehenden Dokumentationssprachen bestätigen diese Erwartung. Wir geben in Abb. 6 davon einen Eindruck anhand einer Zusammenstellung Verwandter Begriffe unterschiedlicher Pfadlänge aus dem *Infodata Thesaurus* ausgehend vom Deskriptor *Dokumentationssprache*<sup>6</sup>.

Deskriptor: Dokumentationssprache				
VB Pfadlänge 1	VB Pfadlänge 2	VB Pfadlänge 3	VB Pfadlänge 4	VB Pfadlänge 5
Ordnungssystem	Anordnungstechnik	Datenanordnung	Dateiaufbau	Benutzerführung
	Hierarchie	Permutation	Datenbankaufbau	Browsing
		Relation	Datenstruktur	Datei
			Hypertext	Datenfeld
			KWIC	Datenverknüpfung
			KWOC	Elektronisches Publizieren
			Registererstellung	Entity-Relationship
				Erstellung
				Formatierung
				Index
				Indexierungsverfahren
				Kategorienschema
				Maschinelle Registererstellung
				Metadaten
				Multimedial
				Objekt-orientiert
				Register
				Relational
				Struktur
				Volltextspeicherung
				Wörterbucherstellung

Abb. 6: Verwandte Begriffe zu *Dokumentationssprache* über mehrere Stufen im *Infodata-Thesaurus*

Es liegen eine Reihe von Studien vor, die eine feinere Strukturierung der Assoziationsrelation zur Gestaltung von Dokumentationssprachen und zur Anwendung für Erschließung und Retrieval unter inhaltlichen Gesichtspunkten vorschlagen<sup>7</sup>. Mit unserem Versuch einer Verknüpfung zwischen inhalt-

<sup>5</sup> Vgl. die Darstellung in: Gödert, W. et. al. (Fußnote 1), S. 31-35.

<sup>6</sup> Vgl.: INFODATA Thesaurus. Version: Okt. 2000. Potsdam: Informationszentrum für Informationswissenschaft und -praxis, 2000.

<sup>7</sup> Vgl.: Tudhope, D., H. Alani u. C. Jones: Augmenting thesaurus relationships: possibilities for retrieval. In: Journal of digital information, vol. 1(2001), no. 8. URL: <http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Tudhope/>. Green, R., C.A. Bean: Aligning systems of relationships. In: Knowledge organization, information systems and other essays: Professor A. Neelameghan Festschrift. Ed. by K.S. Raghavan and K.N. Prasad. New Delhi: Ess Ess Publications, 2006, S. 111-128.

Bean, C.A., R. Green (Eds): Relationships in the organization of knowledge. Boston: Kluwer Academic, 2001.

licher Interpretation und der Betrachtung des Nutzens formaler Schlussfolgerungsprozesse entlang der Relationenpfade wollen wir an diese Vorschläge anknüpfen. Zu diesem Zweck haben wir die vorliegenden Vorschläge zur Typisierung noch einmal kondensiert und in ein eigenes Schema gebracht, das in Abb. 7 wiedergegeben wird.

- Rohstoff / Voraussetzung
- Methode
- Handlung, Anwendung (Tätigkeit)
- Handlung, Anwendung (Automatischer Prozess)
- Handlung, Anwendung (Hilfsmittel)
- Handlung, Anwendung (Produkt)
- Handelnder / Nutzer / Anwender (Person)
- Handelnder / Nutzer / Anwender (Institution)
- Sachgebiet, Anwendungs-, Bezugsbereich

Abb. 7: Inventar typisierter assoziativer Beziehungen

Mit diesem Inventar lassen sich nun die Ergebnisse unserer Analysen in tabellarischen Übersichten zusammenstellen. Die erste Tabelle (Abb. 8) enthält dabei Aussagen zur Transitivität bei Verbindung von Relationen gleichen Typs.

Relationstyp	Relation 1	Relation 2	Transitivität
Äquivalenz	Synonym	Synonym	O
Hierarchie	Abstraktion, generischer Kontext	Abstraktion, generischer Kontext	+
	Ganzes / Teil	Ganzes / Teil	+
	Abstraktion, generischer Kontext	Ganzes / Teil	-
	Ganzes / Teil	Abstraktion, generischer Kontext	-
Chronologischer Kontext	Früher / Später	Früher / Später	+
	Später / Früher	Später / Früher	+
	Früher / Später	Später / Früher	-
	Später / Früher	Früher / Später	-
Assoziation	Unspezifische Assoziation	Unspezifische Assoziation	-
	Rohstoff / Produkt	Rohstoff / Produkt	+
	Kausalität (Ursache – Wirkung)	Kausalität (Ursache – Wirkung)	+
	Handelnder (Person) - Handlung	Handelnder (Person) - Handlung	-
	Handelnder (Institution) - Handlung	Handelnder (Institution) - Handlung	-
	Handelnder (Person) - Produkt	Handelnder (Person) - Produkt	-
	Handelnder (Institution) - Produkt	Handelnder (Institution) - Produkt	-

---

Michel, D. (1997): Taxonomy of Subject Relationships. URL: <http://web2.ala.org/ala/alctscontent/CCS/committees/subjectanalysis/subjectrelations/msrscu2.pdf>. Part of: Final Report to the ALCTS/CCS Subject Analysis Committee. June 1997. URL: <http://web2.ala.org/ala/alctscontent/CCS/committees/subjectanalysis/subjectrelations/finalreport.cfm>.

	Handlung / Produkt	Handlung / Produkt	-
+: Transitivität liegt vor -: Transitivität kann nicht erwartet werden O: Im Rahmen klassischer Dokumentationssprachen nicht zugelassen / eingesetzt			

Abb. 8: Transitivität bei gleichem Relationstyp

Die meisten Aussagen zum Vorliegen der Transitivität erklären sich von selbst oder durch Zuhilfenahme einfacher Beispiele, für einige sollen ergänzende Erläuterungen gegeben werden. Für die Synonymrelation kann Transitivität in Dokumentationssprachen nicht untersucht werden, da ein Synonym niemals sowohl Quelle als auch Ziel für eine Relation sein kann. Entsprechend erfolgte die Kennzeichnung in der Abb. 8.

Die Tabelle der Abb. 9 enthält Aussagen zur Transitivität bei Verbindung von hierarchischen und chronologischen Relationen unterschiedlichen Typs. Es wiederholt sich ein entsprechender Sachverhalt für die Verbindung von Synonymen mit anderen Relationen; Synonyme als Verweisungsformen können in Dokumentationssprachen niemals als Ziel einer Relation auftauchen.

Relationstyp	Relation 1	Relation 2	Transitivität
Hierarchie	Synonym	Abstraktion, generischer Kontext	+
	Synonym	Ganzes / Teil	+
	Abstraktion, generischer Kontext	Synonym	O
	Abstraktion, generischer Kontext	Synonym	O
	Ganzes / Teil	Synonym	O
	Ganzes / Teil	Synonym	O
Chronologischer Kontext	Synonym	Früher / Später	+
	Synonym	Später / Früher	+
	Früher / Später	Synonym	O
	Später / Früher	Synonym	O
	Abstraktion, generischer Kontext	Früher / Später	+
	Abstraktion, generischer Kontext	Später / Früher	+
	Früher / Später	Abstraktion, generischer Kontext	+
	Später / Früher	Abstraktion, generischer Kontext	+
	Ganzes / Teil	Früher / Später	+
	Ganzes / Teil	Später / Früher	+
	Früher / Später	Ganzes / Teil	+
	Später / Früher	Ganzes / Teil	+
+: Transitivität liegt vor -: Transitivität kann nicht erwartet werden O: Im Rahmen klassischer Dokumentationssprachen nicht zugelassen / eingesetzt			

Abb. 9: Transitivität bei Verbindung unterschiedlicher Typen hierarchischer Relationen

Die Übergänge aus hierarchischen in assoziative Strukturen sind in den Relationengrafiken der Abb. 10 und Abb. 12 modellhaft dargestellt. Es wurde schon erwähnt, dass es für Aussagen über die Transitivität beim Übergang von einer Struktur in die andere wesentlich ist, die Reihenfolge der

Verknüpfung der Relationen zu beachten. Abb. 10 zeigt zunächst den Übergang aus einer assoziativen in eine hierarchische Struktur.

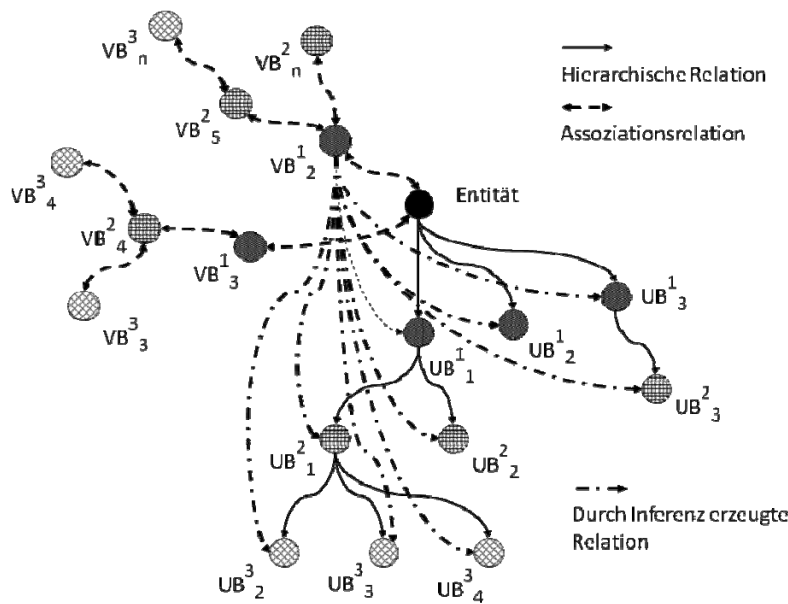


Abb. 10: Übergang aus einer assoziativen in eine hierarchische Relationenstruktur

Eine Interpretation dieser Situation kann wie folgt vorgenommen werden. Ein Übergang von einem Verwandten Begriff zu einem anderen kann nur inhaltlich mit seinem Nutzen für einen Retrievalvorgang begründet werden und ist damit an Bewertungseigenschaften gebunden, die außerhalb der Möglichkeiten einer Formalisierung innerhalb der Begriffsstruktur liegen. Die Fortsetzung in eine hierarchische Struktur mit Aspektdifferenzierung, die ihrerseits durch Transitivität zwischen den Ebenen gekennzeichnet ist, ist daher im Allgemeinen möglich. Abb. 11 zeigt daher die transitiven Verbindungen unterschiedlicher typisierter assoziativer Relationen als erster Relation mit hierarchischen Relationen als zweiter Relation in Form einer Tabelle.

Relationstyp	Relation 1	Relation 2	Transitivität
Assoziation	Unspezifische Assoziation	Abstraktion, generischer Kontext	+
	Unspezifische Assoziation	Ganzes / Teil	+
	Unspezifische Assoziation	Früher / Später *	+
	Rohstoff / Produkt	Abstraktion, generischer Kontext	+
	Rohstoff / Produkt	Ganzes / Teil	+
	Rohstoff / Produkt	Früher / Später *	+
	Handlung / Produkt	Abstraktion, generischer Kontext	+
	Handlung / Produkt	Ganzes / Teil	+
	Handlung / Produkt	Früher / Später *	+
	Handelnder (Person) - Handlung	Abstraktion, generischer Kontext	+
	Handelnder (Person) - Handlung	Ganzes / Teil	+
	Handelnder (Person) - Handlung	Früher / Später *	+
	Handelnder (Institution) - Handlung	Abstraktion, generischer Kontext	+

	Handelnder (Institution) - Handlung	Ganzes / Teil	+
	Handelnder (Institution) - Handlung	Früher / Später *	+
	Kausalität (Ursache - Wirkung)	Abstraktion, generischer Kontext	+
	Kausalität (Ursache - Wirkung)	Ganzes / Teil	+
	Kausalität (Ursache - Wirkung)	Früher / Später *	+
	Handelnder (Person) - Produkt	Abstraktion, generischer Kontext	+
	Handelnder (Person) - Produkt	Ganzes / Teil	+
	Handelnder (Person) - Produkt	Früher / Später *	+
	Handelnder (Institution) - Produkt	Abstraktion, generischer Kontext	+
	Handelnder (Institution) - Produkt	Ganzes / Teil	+
	Handelnder (Institution) - Produkt	Früher / Später *	+

\* Auch: Später / Früher

Abb. 11: Transitivität bei Verbindung unterschiedlicher Typen assoziativer mit hierarchischen Relationen

Abb. 12 zeigt den Übergang aus einer hierarchischen in eine assoziativen Struktur. Eine formal bestimmte transitive Fortsetzung kann aus den schon vorstehend genannten Gründen für diesen Fall im Allgemeinen nicht erwartet werden.

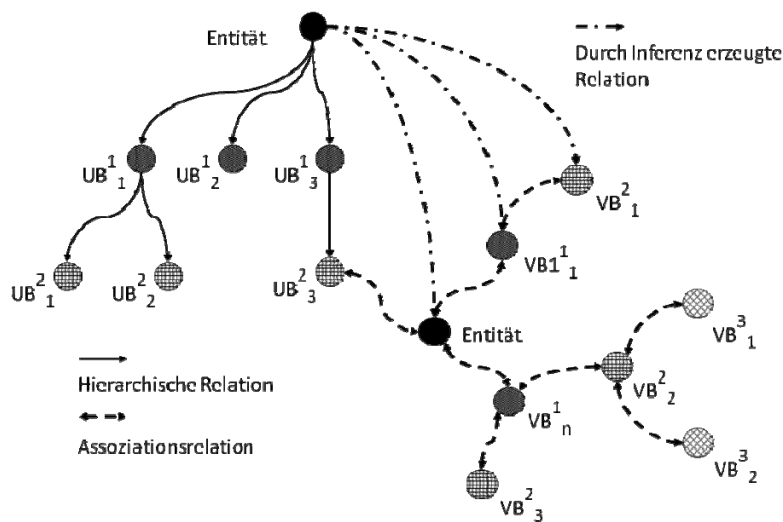


Abb. 12: Übergang aus einer hierarchischen in eine assoziative Relationenstruktur

Auf die Angabe der nicht-transitiven Verbindungen unterschiedlicher typisierter assoziativer Relationen wollen wir generell verzichten.

Als Ergebnis der Analysen kann festgehalten werden, dass bei Typisierung der Assoziationsrelation mehr Möglichkeit zur formalen Inferenz entlang der Relationenpfade einer begrifflichen Struktur vorhanden sind, als das aus dem gewohnten Kontext von Dokumentationssprachen erwartet werden kann.

Ist man an der Fortsetzung des aufgezeigten Weges interessiert, bleibt es eine Aufgabe, das Gebäude aus typisierten Relationen und zugehörigen formalen Eigenschaften weiter auszuarbeiten.

Es soll jetzt aufgezeigt werden, welchen Nutzen die Kombination aus typisierten Relationen und Inferenzen für Retrievalvorgänge haben kann.

### 2.3 Typisierte Relationen und Inferenzen im Information Retrieval

Aufbauend auf der vorgestellten Analyse wollen wir unser Modell für ein Ontologie basiertes Indexieren und Retrieval vorstellen. Dieses Modell besteht aus einer Kombination von Elementen der traditionellen Indexierung und der formalen Wissensrepräsentation. Die Entitäten als begriffliche Repräsentanten sollen in Facetten geordnet sein. Innerhalb der Facetten soll die begriffliche Strukturierung auf transitiv hierarchische Relationen begrenzt werden. Zwischen den Entitäten unterschiedlicher Facetten werden typisierte assoziative Relationen ausgewiesen, deren formale Eigenschaften für Inferenzprozesse näher charakterisiert werden müssen. Die formalen Eigenschaften der Wissensrepräsentation gestatten es, Bedingungen zur Entitätenauswahl formulieren. Hierfür werden die ausgewiesenen Relationen genutzt. Unter Zuhilfenahme von Inferenzprozessen lassen sich Treffermengen erzeugen, ohne dass alle Begriffe der Suchformulierung indexiert worden sein müssen.

Zur Illustration des Modells betrachten wir unsere eingangs angegebenen Beispiele von Fragestellungen. Diese lauteten:

- Gesucht werden Dokumente über den *Wandertrieb von Singvögeln*
- Gesucht werden Dokumente über *Singvögel mit Wandertrieb*

Es wurde bereits festgehalten, dass das traditionelle Indexierungs- und Retrievalmodell keine Lösung für die zweite Fragestellung anbietet. Wir wollen daher nun klären, welche Wissensmodellierung mit unserer Herangehensweise möglich wird und welchen Beitrag eine solche Modellierung zur Lösung unseres Problems leisten kann.

Eine gemeinsame Indexierung mit den Entitäten *Singvögel* und *Wandertrieb* kann für das zweite Thema nicht zutreffend sein. Mit unserem Ansatz betrachten wir zunächst in Abb. 13 eine Wissensmodellierung, die neben einer hierarchisch geordneten Taxonomie der Vogelarten typisierte Relationen zur Angabe weiterer Eigenschaften der berücksichtigten Entitäten benutzt. Diese Repräsentation ist für unsere Problemstellung ebenfalls nicht tauglich, da die Aussage *Singvögel hat Wandertrieb* faktisch unrichtig ist, ebenso wie die Aussage *Singvögel fressen Körner*. Die Unrichtigkeit entsteht aus der Interpretation als Aussage über *alle* Singvögel und der nicht vorhandenen Möglichkeit, die Aussage auf die zutreffenden Arten zu spezifizieren. Es gibt Singvögel mit Wandertrieb, aber eben auch solche ohne; es gibt Singvögel, die Körner fressen, aber auch solche, die dies nicht tun.

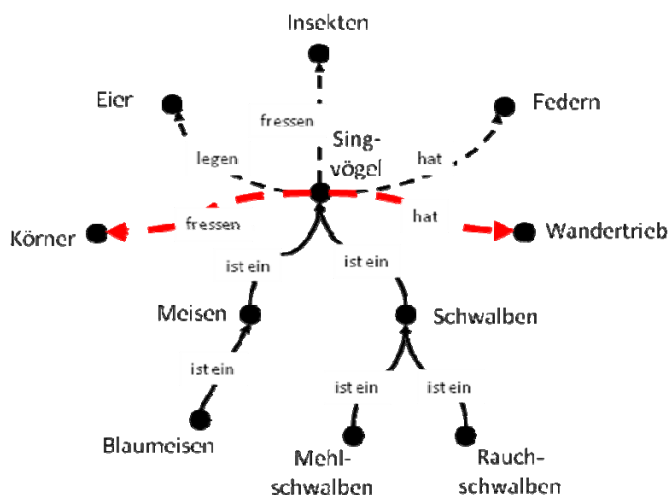


Abb. 13: Wissensrepräsentation mit falscher Eigenschaftszuordnung

Damit der Weg, Eigenschaften über typisierte Relationen anzugeben, nicht zu fehlerhaften Aussagen führt, muss auf die richtigen Anknüpfungsstellen geachtet werden. Als Regel kann gelten, dass das Ausweisen der typisierten Relationen nicht für jeden hierarchischen Strang einer Taxonomie vollständig geschehen darf (auch einzelne Säugetierarten verfügen über einen Wandertrieb), sondern dass die Modellierung gemäß der Prinzipien aspektorientierter Facettierung und selektiv erfolgen muss. So müsste für unser Beispiel neben einer Taxonomie-Facette eine Verhaltens-Facette berücksichtigt werden.<sup>8</sup> Weist man die notwendigen Beziehungen für jede Entität spezifisch aus, so ist ein erheblicher Aufwand erforderlich und es entsteht ein Geflecht an Beziehungen, das großes Potenzial zur Unüberschaubarkeit in sich birgt. Die Abb. 14 gibt für unser Beispiel einen exemplarischen Eindruck.

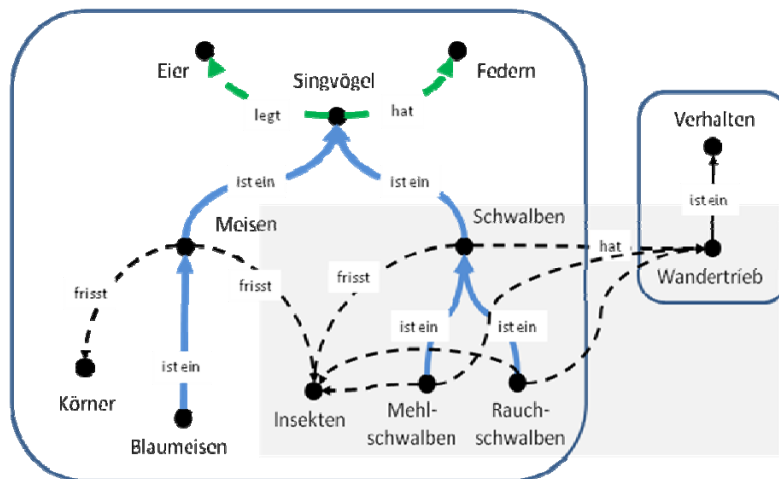


Abb. 14: Wissensrepräsentation mit typisierten Eigenschaftszuweisung ohne Inferenzen

Mit Hilfe von Inferenzen entlang der Pfade hierarchischer Strukturen kann dieser Unüberschaubarkeit entgegen gewirkt werden. Die durch eine typisierte Relation ausgewiesene Beziehung besteht dann nicht allein zwischen den direkt beteiligten Entitäten, sondern sie überträgt sich formal auf alle jeweiligen Unterbegriffe. Die inhaltliche Stimmigkeit der formalen Eigenschaft muss dabei durch Wahl der jeweils richtigen Entitäten für das Herstellen der Verbindung sicher gestellt werden. In

<sup>8</sup> Für eine geeignete Darstellung der hier zugrunde gelegten Prinzipien zur Herstellung facettierter Ordnungsstrukturen sei verwiesen auf: Buchanan, B.: Bibliothekarische Klassifikationstheorie. München: Saur, 1989.

unserem Beispiel heißt dies, es muss die jeweils zutreffende Entität des höchsten hierarchischen Niveaus gefunden werden. Diese Lösung ist in der Abb. 15 angedeutet.

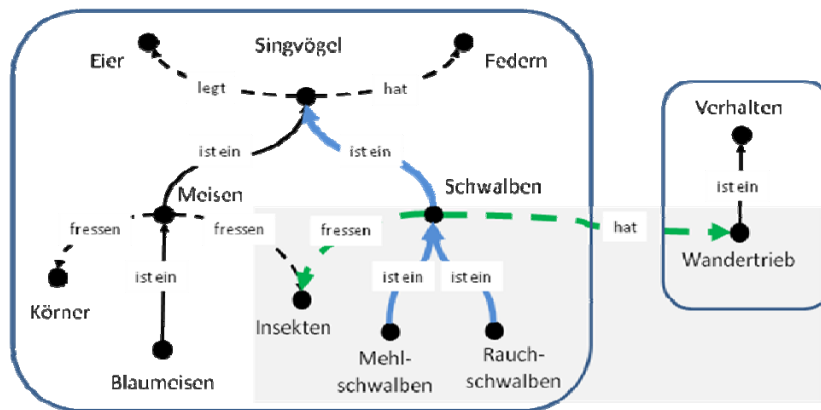


Abb. 15: Wissensrepräsentation mit Angabe aspektorientierter typisierter Relationen und mit Inferenzen

Mit diesen Gestaltungsoptionen lässt sich eine Lösung für unser Problem der Fragestellung *Gesucht werden Dokumente über Singvögel mit Wandertrieb* angeben. Für diese Antwort brauchen weder *Singvögel* noch *Wandertrieb* als indexierte Entitäten mit Dokumenten verbunden zu sein. Wir suchen ja nach Dokumenten, die von bestimmten Entitäten (*Singvögel*) mit bestimmten Eigenschaften (*haben Wandertrieb*) handeln, ohne dass diese Eigenschaften selbst behandelte Gegenstände sein müssen. Unsere Wissensrepräsentation enthält in der Taxonomie-Facette die Kenntnis, welche Arten von *Singvögeln* es gibt. Eine typisierte Relation stellt für die zutreffenden Entitäten die Verbindung zur Eigenschaft *hat Wandertrieb* her. Eine Inferenz über die hierarchischen Relationspfade gestattet es, die Menge aller in Frage kommenden Singvögel zu bilden und damit eine Recherche durchzuführen. Eine Antwort auf unsere erste Fragestellung - *Gesucht werden Dokumente über den Wandertrieb von Singvögeln* - würde wie im klassischen Indexierungs- und Retrievalmodell voraussetzen, dass *Wandertrieb* als indexierte Entität gesucht werden kann und dass die Retrievalumgebung die postkoordinierende Suche beider Entitäten gestattet. Wir können also zusammenfassen, dass die Kombination aus Verwendung typisierter Relationen und formalen Inferenzen die Lösung eines Problems gestattet, das mit den klassischen Methoden allein nicht gelöst werden kann.

### 3 Ein Prototyp

Um die hier diskutierten Möglichkeiten innerhalb eines Informationssystems interaktiv abfragbar machen zu können, wurde am Institut für Informationswissenschaft der Fachhochschule Köln ein experimenteller Prototyp<sup>9</sup> entwickelt, dessen Bestandteile und Möglichkeiten zur Veranschaulichung unserer Ergebnisse hier vorgestellt werden sollen. Für diesen Prototyp wurde keine idealtypische Situation konstruiert, sondern es wurde auf bereits existierende Daten zurückgegriffen. Diese Daten bestehen aus ca. 14.000 Dokumenten, die der Datenbank *Literatur zur Informationserschließung*<sup>10</sup> entnommen wurden. Der inhaltliche Schwerpunkt dieser Dokumente korrespondiert zu einem Auszug aus dem *ASIST-Thesaurus*<sup>11</sup>, der als Topic map<sup>12</sup> unter der Software *Ontopia*<sup>13</sup> umgestaltet wurde,

<sup>9</sup> Ein besonderer Dank geht an Matthias Nagelschmidt und Jens Wille, die die inhaltliche und technische Einrichtung der Suchumgebung vorgenommen und damit erste Experimente mit dem hier vorgeschlagenen Modell ermöglicht haben.

<sup>10</sup> Vgl.: <http://ixtrieve.fh-koeln.de/LitIE/>.

<sup>11</sup> Vgl.: ASIS&T Thesaurus of Information Science, Technology and Librarianship. Eds: Redmond-Neal, A./M.M.K. Hlava. 3rd ed. Medford, NJ: Information Today, 2005.

<sup>12</sup> Vgl. für das Konzept einer Topic map: ISO 13250: Topic maps. Information technology. Geneva: ISO, 1999.



indem die vorhandenen Deskriptoren des Ausschnitts mit typisierten Relationen eines erweiterten Relationeninventars nachrelationiert wurden. Als formale Repräsentationsumgebung benutzt *Ontopia* das Format *XTM*, dessen formale Spezifikation<sup>14</sup> zur Erweiterung des Relationenumfangs genutzt werden kann. Wir zeigen dies nachfolgend am Beispiel der Deklaration für die zu unterscheidenden Rollen *is method of* und *is adopting* der typisierten assoziativen Relation *Methodology*:

```
<!-- Präambel 1-A: Typed associative relations -->
<!-- Relation -->
<topic id="Methodology">
  <baseName>
    <instanceOf>
      <topicRef xlink:href="#AssociativeRelation"/>
    </instanceOf>
    <baseNameString>Associative Relation (Methodology)</baseNameString>
  </baseName>
</topic>

<!-- role 1 -->

<topic id="isMethodOf">
  <baseName>
    <instanceOf>
      <topicRef xlink:href="#MethodologyRelationMember"/>
    </instanceOf>
    <baseNameString>is method of</baseNameString>
  </baseName>
</topic>

<!-- role 2 -->

<topic id="isAdopting">
  <baseName>
    <instanceOf>
      <topicRef xlink:href="#MethodologyRelationMember"/>
    </instanceOf>
    <baseNameString>is adopting</baseNameString>
  </baseName>
</topic>
```

Die Verbindung zwischen zwei Entitäten - *indexing* und *abstracting\_and\_indexing\_service\_bureaus* durch die Relation *Methodology* mit dem Rollenpaar *isAdopting* und *isMethodOf* hat dann die Gestalt:

```
<!-- Präambel 2: Definition of entity roles of typed relations -->
<association>
  <instanceOf>
    <topicRef xlink:href="#Methodology"/>
  </instanceOf>
  <member>
```

---

Pepper, S.: The TAO of topic maps: finding the way in the age of infoglut. URL: <http://www.ontopia.net/topicmaps/materials/tao.html>.

Garshol, L.M.: Metadata? Thesauri? Taxonomies? Topic maps!: making sense of it all. In: Journal of information science, vol. 30(2005), no.4, pp.378-391. Auch unter: <http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html>.

Park, J./Hunting, S. (Eds): XML topic maps: creating and using topic maps for the Web. Boston: Addison-Wesley, 2003.

<sup>13</sup> Vgl.: <http://www.ontopia.net>.

<sup>14</sup> Vgl.: Pepper, S./Moore, G. /TopicMaps.Org Authoring Group (2001): XML Topic Maps (XTM) 1.0: TopicMaps.Org Specification. URL: <http://www.topicmaps.org/xtm/1.0/xtm1-20010806.html>.

```

<roleSpec>
  <topicRef xlink:href="#isAdopting"/>
</roleSpec>
<topicRef xlink:href="#abstracting_and_indexing_service_bureaus"/>
</member>
<member>
  <roleSpec>
    <topicRef xlink:href="#isMethodOf"/>
  </roleSpec>
  <topicRef xlink:href="#indexing"/>
</member>
</association>

```

Wir verzichten auf die Angabe weiterer modellierungstechnischer Details und zeigen in Abb. 16 eine grafische Visualisierung für den Deskriptor *Index languages*.

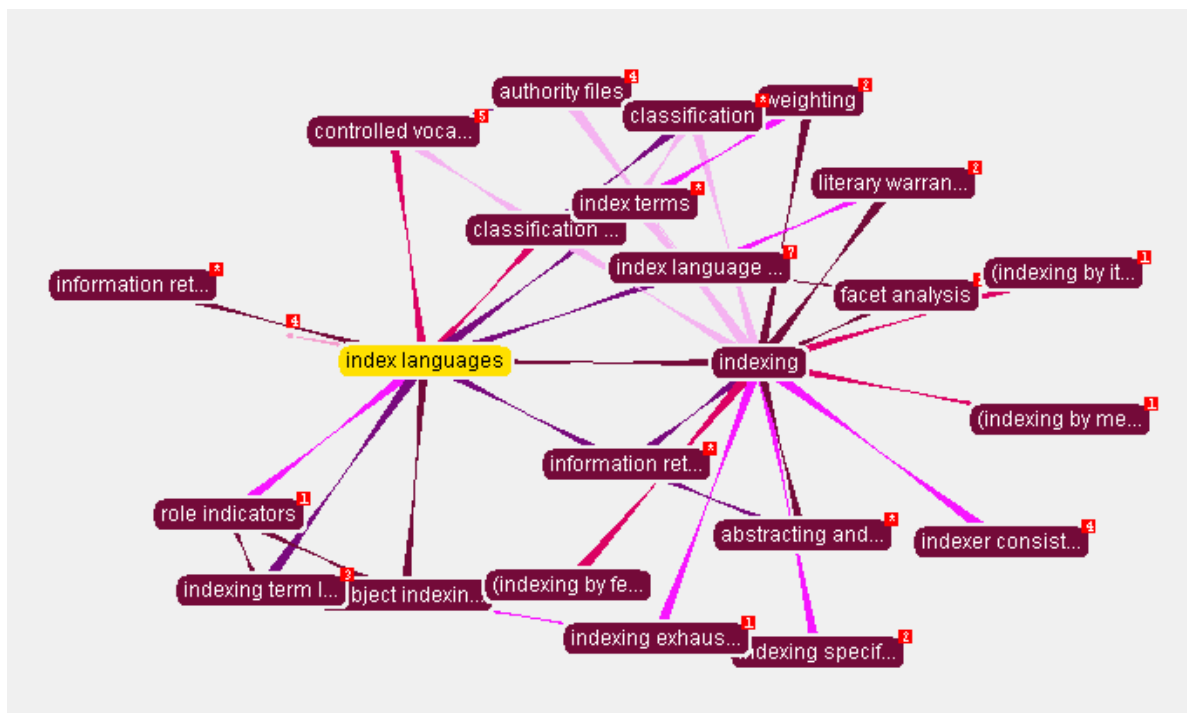


Abb. 16: *Ontopia* Topic map Auszug des *ASIST-Thesaurus* für den Deskriptor *Index languages*

Mit Hilfe eines Web-Interfaces<sup>15</sup> und Benutzung der *Ontopia* spezifischen Abfragesprache *tolog*<sup>16</sup> können Suchfragen formuliert werden, die nicht allein Entitäten berücksichtigen, sondern auch die zwischen ihnen bestehenden Beziehungen als Selektionskriterien zu spezifizieren gestatten. Die Abb. 17 zeigt das entsprechende Interface.

<sup>15</sup> Vgl.: <http://ixtrieve.fh-koeln.de/ghn>.

<sup>16</sup> Garshol, L.M.: Tolog: a topic map query language. URL: <http://www.ontopia.net/topicmaps/materials/tolog.html>.  
Tolog language tutorial (2007). Ver. 3.4. URL: <http://www.ontopia.net/omnigator/docs/query/tutorial.html>.

[Topics \(XTM\)](#) | [Documents \(DBM\)](#) | [New query](#)

```
import "http://psi.ontopia.net/tolog/string/" as s
CUSTOM INFERENCE RULES (optional)
select $[TOPIC] from YOUR QUERY?
```

YOUR QUERY [\[sample\]](#):

```
Methodology($TOPIC : isMethodOf, indexing :
isAdopting)?
```

CUSTOM INFERENCE RULES (optional) [\[sample\]](#):

[ [tolog tutorial](#) | [tolog reference](#) ] Daten absenden

Topics (4):

- [facet analysis](#) (9)
- [index languages](#) (0)
- [literary warrant](#) (0)
- [weighting](#) (3)

Documents (12):

- [Devadason, F.J.: Ranganathan's idea of facet analysis in action \(1986\)](#)
- [Coetzee, P.C.: theory of logistic facet analysis \(1968\)](#)
- [Broughton, V.: Faceted classification as a basis for knowledge organization in a digital environment – the Bliss Bibliographic Classification as a model for vocabulary management and the creation of multi-dimensional knowledge structures \(2001\)](#)
- [Kim, M.; Kang, S.; Lim, J.: Automatic user preference learning for personalized electronic program guide applications \(2007\)](#)
- [Bonnici, L.J.; Kim, J.; Burnett, K.; Miksa, S.D.: development of a facet analysis system to identify and measure the dimensions of interaction in online learning \(2007\)](#)
- [Beghtol, C.: facet concept as a universal principle of subdivision \(2006\)](#)
- [Small, H.; Zitt, M.: Modifying the journal impact factor by fractional citation weighting – the audience factor \(2008\)](#)
- [LaBarre, K.: Facets, search and discovery in next generation catalogs – informing the future by revisiting past understanding](#)
- [Facets: a fruitful notion in many domains – special issue on facet analysis \(2008\)](#)
- [Skare, R.: Complementarity - a concept for document analysis? \(2009\)](#)
- [Kumar, V.; Neelameghan, A.; Deokattey, S.: method for developing a domain ontology – a case study for a multidisciplinary subject \(2010\)](#)
- [LaBarre, K.: Facet analysis \(2010\)](#)

Abb. 17: Such-Interface für den als Topic map gestalteten *ASIST-Thesaurus* und zugeordneten Dokumenten

Nach Durchführung einer Suche mit *tolog* werden die Topics angezeigt, die den Bedingungen der formulierten Fragestellung entsprechen. Die mit diesen Topics indexierten Dokumente werden als Treffermenge generiert. Für die Bewertung der Ergebnisse verdient beachtet zu werden, dass die Indexierung der 14.000 Dokumente mit Deskriptoren des *ASIST-Thesaurus* auf der Basis automatischer Indexierung mit Hilfe der Software *Lingo*<sup>17</sup> erfolgte. Die inhaltliche Indexierungsqualität kann damit gegenüber einer intellektuellen Indexierung suboptimal sein.

Zur besseren Illustration unseres Ansatzes diskutieren wir zwei Beispiele, in denen die Entität *Indexing* mit unterschiedlichen Relationen zu anderen Entitäten versehen ist und die Ergebnismengen über die Relationen gefiltert werden sollen. Abb. 17 zeigt die Anfrage nach Dokumenten, in denen der Topic *indexing* als *methodic procedure* behandelt wird. Als *tolog*-Formulierung lautet die Anfrage (linke Seite in Abb. 17):

```
Methodology($TOPIC : isMethodOf, indexing : isAdopting)?
```

Die Abb. 17 zeigt die ermittelten Topics unterhalb der Überschrift *Topics* (obere rechte Seite). Für unsere Fragestellung werden vier Topics ermittelt:

```
facet analysis
index languages
literary warrant
weighting
```

<sup>17</sup> Vgl.: Lepsky, K., J. Vorhauer: Lingo - ein open source System für die Automatische Indexierung deutschsprachiger Dokumente. In: ABI-Technik, vol. 26(2006), H.1, S. 18-28.  
Vgl. auch: <http://lex-lingo.blogspot.com>.

Dieses Ergebnis basiert auf der vorgenommenen Modellierung des *ASIST Thesaurus* als Topic map, wie es die Abb. 18 für den Topic *Indexing* und die Relation *Methodology* zeigt:

Relations (5):

- Associative Relation (Characterization)
  - is characterizing
    - indexer consistency
    - indexing exhaustivity
    - indexing specificity
- Associative Relation (Methodology)
  - is adopting
    - facet analysis
    - index languages
    - literary warrant
    - weighting
  - is method of
    - abstracting and indexing service bureaus
- Associative Relation (Production)
  - is producing
    - information retrieval indexes
- Associative Relation (Ūsāgē)
  - is using
    - authority files
    - classification
    - classification schemes
    - index terms
- Hierarchical Relation
  - Broader Term
    - (indexing by feature indexed)
    - (indexing by item indexed)
    - (indexing by method used)

Abb.18: Typisierte Relationen für den Topic *Indexing* der *ASIST* Topic map

Unter der Überschrift *Documents* wurden 12 Dokumente angezeigt, die mit einem oder mehreren der vier Topics indexiert wurden (untere rechte Seite der Abb. 17).

Unser zweites Beispiel fragt nach Hilfsmitteln zur Indexierung. Der entsprechende Topic ist wieder *Indexing*, die typisierte Relation ist *usage*. Die *tolog*-Formulierung lautet:

```
Usage($TOPIC : isInstrumentOf, indexing : isUsing)?
```

Die ermittelten Topics lauten in diesem Fall:

```
authority files
classification
classification schemes
index terms
```

Sie korrespondieren zur Modellierung der Abb. 18, die im Kasten mit gestrichelten Linien zu sehen ist. Die resultierende Ergebnismenge besteht aus 565 Dokumenten<sup>18</sup>.

Als Ergebnis dieses ersten Suchbeispiels kann festgehalten werden, dass es die Verwendung typisierter Relationen in der Wissensmodellierung gestattet, für differenzierte Suchanfragen unterschiedliche Mengen von Topics auszuwählen. Damit können unterschiedliche Treffermengen gebildet werden, die jeweils die Dokumente für die unterschiedlich spezifizierten Suchinteressen beinhalten.

Der Prototyp gestattet es auch, die in unserem Modell vorgestellte Kombination einer Auswahl typisierter Relationen mit Inferenzen über die hierarchische Struktur einer Wissensrepräsentation am Beispiel der Topic map zu demonstrieren. Dies soll als zweites Suchbeispiel vorgestellt werden, das zunächst alle hierarchisch untergeordneten Topics zu *controlled\_vocabularies* einbindet. Hierfür müssen *Ontopia* spezifische Inferenzregeln, die sogenannten *Custom Inference Rules* benutzt werden. Das Such-Interface bietet hierfür einen eigenen Eingabebereich an, der in Abb. 19 in der linken unteren Hälfte zu sehen ist.

Abb. 19: Such-Interface für die ASIST Topic map mit *Custom Inference Rules*

Vordefiniert sind die nachstehend wiedergegebenen Regeln, die über den Link "sample" abgerufen werden können. Bei Bedarf können sie modifiziert oder durch eigene Regeln ersetzt werden.

```
direct-narrower-term($A, $B) :-
    HierarchicalRelation($A : broaderTermMember,
                        $B : narrowerTermMember).

strictly-narrower-term($A, $B) :- {
    direct-narrower-term($A, $B) |
    direct-narrower-term($A, $C), strictly-narrower-term($C, $B)
}.

narrower-term($A, $B) :- {
```

<sup>18</sup> Die Suche kann durchgeführt werden unter: <http://ixtrieve.fh-koeln.de/ghn/>.

```

    $A = $B | strictly-narrower-term($A, $B)
}.

narrower-term-1($A, $B) :- {
    $A = $B | direct-narrower-term($A, $B)
}.

narrower-term-2($A, $B) :- {
    narrower-term-1($A, $B) |
    narrower-term-1($A, $C), narrower-term-1($C, $B)
}.

narrower-term-3($A, $B) :- {
    narrower-term-2($A, $B) |
    narrower-term-2($A, $C), narrower-term-1($C, $B)
}.

direct-broader-term($A, $B) :-
    direct-narrower-term($B, $A).

strictly-broader-term($A, $B) :-
    strictly-narrower-term($B, $A).

broader-term($A, $B) :-
    narrower-term($B, $A).

broader-term-1($A, $B) :-
    narrower-term-1($B, $A).

broader-term-2($A, $B) :-
    narrower-term-2($B, $A).

broader-term-3($A, $B) :-
    narrower-term-3($B, $A).

```

Unter Zuhilfenahme dieser Regeln lassen sich nun Suchanfragen formulieren, die sowohl typisierte Relationen filtern als auch Inferenzen entlang der Hierarchiepfade beinhalten. Das erste Beispiel lautet in der *tolog*-Syntax:

```

Production($TOPIC : isProducing, $PRODUCT : isProductOf),
narrower-term(controlled_vocabularies, $PRODUCT)?

```

Damit sollen Topics für eine Suche nach *Voraussetzungen zur Erstellung von Dokumentations-sprachen* gesucht und werden. Dies entspricht der Modellierung innerhalb der Topic map durch den Topic *controlled\_vocabularies* und der Richtung *isProducing* der Relation *production*. Der Typ der Dokumentationssprache soll nicht unterschieden werden, was zur hierarchischen Inferenz Anlass gibt. Es werden fünf Topics ermittelt, denen 687 Dokumente zugeordnet sind:

```

automatic indexing (96)
index language construction (0)
subject heading lists (45)
subject headings (530)
vocabulary control (41)

```

Die zweite Frage lautet in der *tolog*-Syntax:

```

Production($TOPIC : isProductOf, $PRODUCT : isProducing),
narrower-term(controlled_vocabularies, $PRODUCT)?

```

Mit dieser Formulierung sollen Topics für eine Suche nach *Produkten, die mit der Hilfe von Dokumentationssprachen erstellt wurden*, selektiert werden. Dies entspricht der Modellierung innerhalb der Topic map durch den Topic *controlled\_vocabularies* und der Richtung *isProductOf* der Relation *production* unter Einschluss der hierarchischen Inferenz. Es werden zwei Topics ermittelt, denen 502 Dokumente zugeordnet sind:

```
authority files (140)
thesauri (372)
```

Weitere Beispiele lassen sich durch Benutzung des bereitgestellten Such-Interfaces bearbeiten.

#### 4 Bewertung und Ausblick

Wir haben im Rahmen unserer Diskussion gezeigt, wie ein Ontologie basiertes Modell für Indexierung und Retrieval aussehen kann, das typisierte Relationen zur verfeinerten Repräsentation des semantischen Wissens und Inferenzprozesse zur Durchführung von Suchprozessen benutzt. An einem Prototypen wurde gezeigt, wie man sich die Aufgabenverteilung zwischen Indexierung und Auswahl von Suchtermen vorstellen kann und welche Optionen das Modell für die Gestaltung von Indexierungs- und Retrievalsystemen bietet.

Zum Abschluss unserer Diskussion wollen wir auf einen weiteren Gesichtspunkt aufmerksam machen, für den nur vorläufige Ergebnisse vorgelegt werden können. Vielleicht regen diese ersten Ergebnisse aber zu weiteren Untersuchungen an.

Mit der Verbindung facettierter Strukturen und der Herstellung typisierter Relationen zwischen den Entitäten unterschiedlicher Facetten gelingt es, Sachverhalte auszudrücken, die üblicherweise durch Hilfsmittel des syntaktischen Indexierens zum Ausdruck gebracht werden. Wir wollen die Zusammenhänge andeuten.

Die von uns gewählten typisierten Relationen sind typische Elemente von *Phase relations* oder *Citation orders*, wie sie beispielsweise im Kontext der *Colon Classification*, der *Bliss Classification*, 2nd ed. oder des Indexierungssystems *PRECIS* entwickelt wurden<sup>19</sup>. Damit rückt die Frage in das Blickfeld, ob durch die beschriebene Vorgehensweise Sachverhalte repräsentiert und recherchierbar gemacht werden können, für die man sonst Syntaxelemente wie *Phase relations* oder *Citation orders* benutzt, um die im Dokument enthaltenen a posteriori Relationen auszudrücken. Abb. 20 zeigt eine prototypische Situation einer facettierten Wissensordnung.

---

<sup>19</sup> Vgl.: Ranganathan, S.R. et al.: *Colon Classification: basic and depth version: Vol.1: Schedules for classification*. 7th ed. Bangalore: Sarada Ranganathan Endowment for Library Science, 1987.  
Bliss Bibliographic Classification - BC2: Introduction and auxiliary schedules. Eds. J. Mills and V. Broughton. 2nd ed. London: Butterworths, 1992.  
Austin, D.: *PRECIS: a manual of concept analysis and subject indexing*. 2nd ed. London: British Library, 1984.  
Austin, D.: Prospects for a new general classification. In: *Journal of librarianship*, 1(1969), S. 149-169.  
Austin, D.: The new general faceted classification: an outline of the CRG scheme. In: *Catalogue and index*, 1969, no. 14, S.11-13.

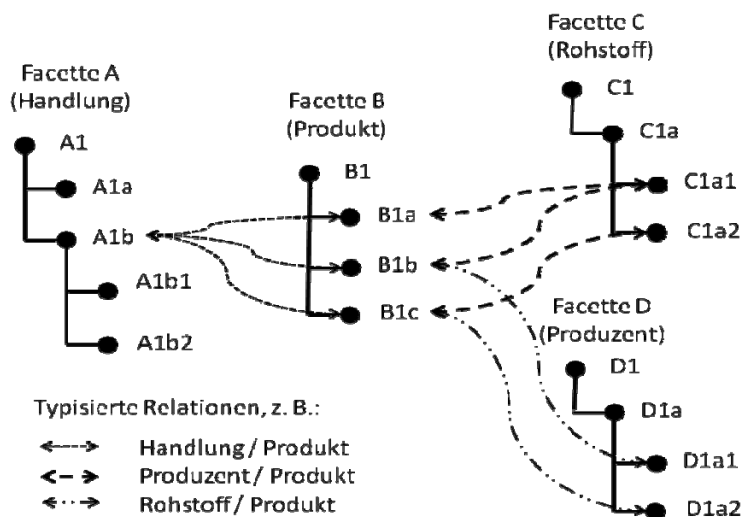


Abb. 20: Facettierte systematische Struktur mit typisierten Relationen zwischen den Facetten

Im Modell des syntaktischen Indexierens wird ein dokumentspezifischer Zusammenhang zwischen Entitäten der einzelnen Facetten durch ein synthetisiertes Indexat zum Ausdruck gebracht, das gemäß einer Citation order gebildet wird, z. B. hier in Form einer synthetischen Notation *D1a1B1bA1b*. In unserem Modell wäre ein solcher Zusammenhang über die Verbindungen darstellbar, die über die typisierten Relationen in der Wissensrepräsentation modelliert sind. Damit ist nicht die vollständige Flexibilität syntaktischen Indexierens zur freien Synthese dokumentspezifischer Sachverhalte erreichbar, in Abhängigkeit von der ausgewiesenen Relationsstruktur lassen sich aber eine Vielzahl von Themen ausdrücken, die mit den Mitteln des gleichordnenden Indexierens und des Booleschen Retrievals nicht spezifisch darstellbar sind. Damit lassen sich die bisher vorgeschlagenen Ansätze zur Nutzung facettierter Strukturen in Retrieval-Umgebungen<sup>20</sup> um neue Ideen und Gestaltungsoptionen ergänzen.

Im gegenwärtigen Stand der Ausarbeitung des Modells kann noch keine abschließende Antwort auf die eingangs des Abschnitts gestellte Frage gegeben werden. Weitere Untersuchungen können auf diese Fragen vielleicht Antworten finden und damit möglicherweise dem Thema des koextensiven Indexierens neue Impulse geben.

**Autor** Prof. Winfried Gödert  
 Fachhochschule Köln  
 Institut für Informationswissenschaft  
 Postanschrift: Gustav-Heinemann-Ufer 54, D-50968 Köln  
 Hausanschrift: Claudiusstr.1, D-50678 Köln  
 winfried.goedert@fh-koeln.de

<sup>20</sup> Vgl. etwa: Gödert, W.: Facet classification in online retrieval. In: International classification. 18(1991) no.2, S.98-109.

Broughton, V.: Faceted classification as a basis for knowledge organization in a digital environment: the Bliss Bibliographic Classification as a model for vocabulary management and the creation of multi-dimensional knowledge structures. In: New review of hypermedia and multimedia. 7(2001) no.7, S.67-102.

Tunkelang, D.: Faceted search. San Rafael, CA: Morgan and Claypool, 2009.