

Begriffe und semantische Relationen in der Wissensrepräsentation

Wolfgang G. Stock, Düsseldorf

Begriffsorientiertes Information Retrieval bedarf einer informationswissenschaftlichen Theorie der Begriffe sowie der semantischen Relationen. Ein Begriff wird durch seine Intension und Extension sowie durch Definitionen bestimmt. Dem Problem der Vagheit begegnen wir durch die Einführung von Prototypen. Wichtige Definitionsarten sind die Begriffserklärung (nach Aristoteles) und die Definition über Familienähnlichkeiten (im Sinne Wittgensteins). Wir modellieren Begriffe als Frames (in der Version von Barsalou). Die zentrale paradigmatische Relation in Wissensordnungen ist die Hierarchie, die in verschiedene Arten zu gliedern ist: Hyponymie zerfällt in die Taxonomie und die einfache Hyponymie, Meronymie in eine ganze Reihe unterschiedlicher Teil-Ganzes-Beziehungen. Wichtig für praktische Anwendungen ist die Transitivität der jeweiligen Relation. Eine unspezifische Assoziationsrelation ist bei den angepeilten Anwendungen wenig hilfreich und wird durch ein Bündel von generalisierbaren und fachspezifischen Relationen ersetzt. Unser Ansatz fundiert neue Optionen der Anwendung von Wissensordnungen in der Informationspraxis neben ihrem „klassischen“ Einsatz beim Information Retrieval: Erweiterung von Suchanfragen (Anwendung der semantischen Nähe), automatisches Schlussfolgern (Anwendung der terminologischen Logik in Vorbereitung eines semantischen Web) und automatische Berechnungen (bei Funktionalbegriffen mit numerischen Wertangaben).

Concepts and semantic relations in knowledge representation.

Concept-based information retrieval is in need of an information science theory of concepts and semantic relations. A concept is determined by its intension and extension as well as by definition. We meet the problem of vagueness by the introduction of prototypes. Important kinds of definition are concept explanations (after Aristotle) and the definition of family resemblances (in the sense of Wittgenstein). We model concepts as frames (in the version of Barsalou). The most important paradigmatic relation in knowledge organization systems is the hierarchy, which is to be arranged into different kinds. Hyponymy consists of taxonomy and simple hyponymy, meronymy of a lot of different part-whole-relations. For practical applications the transitivity of the given relation is important. An unspecific associative relation is only little helpful with our focused applications and should be replaced by generalizable and domain-specific relations. Our approach founds new options of applications of knowledge organization systems in information practice beside „classical“ applications in information retrieval: query expansion (application of semantic proximity), automatic reasoning (application of description logic as a preliminary of the semantic Web) and automatic computations (in the case of functional concepts and numeric values).

Eine Wissensordnung besteht aus Begriffen und semantischen Relationen, die eine Wissensdomäne terminologisch abbilden. In der Wissensrepräsentation unterscheiden wir mit den Nomenklaturen, Klassifikationssystemen, Thesauri, Ontologien und – als Grenzfall einer Wissensordnung – den Folksonomies fünf Ansätze von Wissensordnungen (Stock & Stock, 2008). Wissensdomänen sind abgrenzbare thematische Bereiche, beispielsweise eine wissenschaftliche Dis-

ziplin, eine Wirtschaftsbranche oder die Sprache eines Unternehmens. Ziel einer Wissensordnung in der Informationspraxis ist es, den Retrievalprozess zu unterstützen. Wir wollen dem Nutzer Begriffe zur Suche und zum Stöbern anbieten, automatisch indexieren, automatisch Suchanfragen erweitern, automatisch auf der Basis der eingesetzten Wissensordnung im sog. „semantischen Web“ Schlüsse ziehen und beim Vorliegen numerischer

Werte bei gewissen Relationen ebenfalls automatisch Berechnungen durchführen. In diesem Artikel geht es um Probleme von Wissensordnungen im Allgemeinen und von beliebigen Wissensdomänen. Wie müssen wir Begriffe und Relationen repräsentieren, damit die Ziele erreichbar werden? Hierzu erweitern wir den Status quo der informationswissenschaftlichen Auffassung von Begriffen um die Ansätze der Begriffserklärung (nach Aristoteles) und der Familienähnlichkeit (nach Wittgenstein), führen Synkategoriemata (nach Menne) ein, thematisieren Vagheit sowie Prototypen (mit Rosch) und modellieren Begriffe als Frames (im Sinne von Barsalou). Bei den Relationen (das sind strukturelle Invarianten von Attributen in Frames) ist es wichtig, auf Transitivität bei allen Arten von Hyponymie und Meronymie zu achten. Zudem scheint es geboten, soweit wie möglich auf unspezifische Assoziationsrelationen zu verzichten, um stattdessen mit jeweils spezifischen Begriffsbeziehungen zu arbeiten. Uns geht es hier um eine neue informationswissenschaftliche Sicht auf Begriffe, die zwar auf bekannte und bewährte Theorien und Modelle aufsetzt, die aber auch geeignet ist, alle Vorteile von Knowledge Organization Systems (KOS) und dabei insbesondere von Ontologien für die Informationswissenschaft und -praxis nutzbar zu machen. Wenn wir so etwas wie das „Semantic Web“ schaffen wollen, müssen wir uns notgedrungen Gedanken über den Begriff des Begriffs machen, denn hier liegt der Schlüssel zu jeglicher Semantik (Hjørland, 2007). Zu diesem Zwecke sind auch Ausflüge in die Allgemeine Sprachwissenschaft, die Philosophie, die Kognitionswissenschaft und in die Informatik sinnvoll.

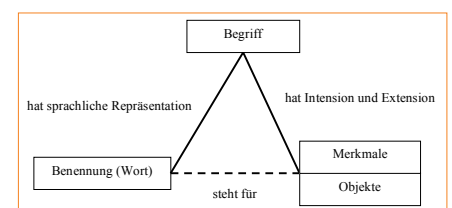


Abbildung 1: Das semiotische Dreieck in der Informationswissenschaft.

1 Begriffe

Das semiotische Dreieck

In der Sprache benutzen wir Symbole, z.B. Worte, mittels deren wir einen Gedanken über ein Bezugsobjekt ausdrücken. Wir sind mit einem Dreiecksverhältnis konfrontiert, das bei Charles K. Ogden und Ivor A. Richards (1974[1923], 18 f.) aus dem Gedanken (auch Bezug genannt), dem Referenten oder Bezugsobjekt und dem Symbol besteht. Ogden/Richards sehen den Gedanken bzw. den Bezug als psychologische Aktivität an („Psychologismus“ nach Schmidt, 1969, 30). In der Informationswissenschaft taucht an der Stelle des (psychologischen) Gedankens der „Begriff“ auf. Wie im klassischen Ansatz von Ogden und Richards wird ein Begriff durch Benennungen sprachlich repräsentiert. Solche Benennungen können natürlichsprachige Worte, aber auch Benennungen künstlicher Sprachen (z.B. Notationen einer klassifikatorischen Wissensordnung) sein. Der Begriff des „Begriffs“ wird als Klasse definiert, unter die gewisse Objekte als Elemente fallen, wobei die Objekte über bestimmte Merkmale verfügen. Die Normen DIN 2330 (1993, 2) und DIN 2342/1 (1993, 1) verstehen unter einem Begriff

eine Denkeinheit, die aus einer Menge von Gegenständen unter Ermittlung der diesen Gegenständen gemeinsamen Eigenschaften mittels Abstraktion gebildet wird.

Diese DIN-Definition ist nicht unproblematisch. Zunächst ist es geraten, statt (der psychologisch anmutenden) „Denkeinheit“ besser von „Klasse“ oder „Menge“ (im Sinne der Mengentheorie) zu sprechen. Zudem gilt nicht für jeden Begriff, dass alle seine Elemente stets und notwendig über „gemeinsame“ Eigenschaften verfügen. Bei Begriffen, die über Familienähnlichkeit gebildet werden, ist dies beispielsweise nicht der Fall. Wir definieren etwa *Gemüse* in Ermangelung gemeinsamer Merkmale durch „ist Kohlgemüse oder Wurzelgemüse oder Fruchtgemüse usw.“. Aber was heißt Familienähnlichkeit? Betrachten wir statt *Gemüse* einen anderen Begriff, den Ludwig Wittgenstein (1977[1953], 56 f.) als Beispiel für dieses Problem verwendet, *Spiel*:

Betrachte z.B. einmal die Vorgänge, die wir „Spiele“ nennen. Ich meine Brettspiele, Kartenspiele, Ballspiele, Kampfspiele, usw. Was ist allen diesen gemeinsam? ... (W)enn du sie anschaust, wirst du zwar nicht etwas sehen, was allen gemeinsam wäre, aber du wirst Ähnlichkeiten, Verwandtschaften, sehen, und zwar eine ganze Reihe.

Einige Spiele haben beispielsweise gemeinsam, dass es Gewinner und Verlierer gibt, andere Spiele – aber nicht alle – sind unterhaltend, wieder andere erfordern Geschick und Glück von den Spielern usw. Der Begriff des Spiels lässt sich also nicht mit genau einer Menge von Merkmalen bilden (Wittgenstein, 1977[1953], 57):

Und das Ergebnis dieser Betrachtung lautet nun: Wir sehen ein kompliziertes Netz von Ähnlichkeiten, die einander übergreifen und kreuzen. Ähnlichkeiten im Großen und Kleinen. Ich kann diese Ähnlichkeiten nicht besser charakterisieren als durch das Wort „Familienähnlichkeiten“; denn so übergreifen und kreuzen sich die verschiedenen Ähnlichkeiten, die zwischen den Gliedern einer Familie bestehen: Wuchs, Gesichtszüge, Augenfarbe, Gang, Temperament, etc. etc. – Und ich werde sagen: die ‚Spiele‘ bilden eine Familie.

Wir müssen zulassen, dass nicht nur eine Konjunktion von Merkmalen, sondern auch mitunter eine Disjunktion von Merkmalen einen Begriff bestimmen kann.

Es gibt zwei Herangehensweisen, Begriffe zu bilden. Der erste Weg führt über die Objekte und legt die Extension des Begriffs fest, der zweite notiert die klassenbildenden Merkmale und bestimmt so seine Intension (Reimer, 1991, 17). Unsere DIN-Normen reden in diesem Zusammenhang von „Begriffsumfang“ und „Begriffsinhalt“. Gottlob Frege arbeitet mit „Bedeutung“ für die Extension und „Sinn“ für die Intension. Egal, wie man dies nun benennt, zentral ist Freges Feststellung, dass Extension und Intension nicht grundsätzlich zusammenfallen müssen. Sein Beispiel ist der Begriff *Venus*, der ja auch mit *Abendstern* oder mit *Morgenstern* bezeichnet werden kann. Frege (1892, 27) stellt fest:

Es würde die Bedeutung von „Abendstern“ und „Morgenstern“ dieselbe sein, nicht aber der Sinn.

Abendstern und *Morgenstern* sind extensional identisch, weil die Menge der Elemente, die jeweils darunter fällt, identisch ist (beide benennen die Venus), sie sind intensional nicht identisch, weil dem *Abendstern* u. a. das Merkmal *erster sichtbarer Stern am Abendhimmel* und dem *Morgenstern* die völlig andere Eigenschaft *letzter sichtbarer Stern am morgendlichen Himmel* zukommt.

Die Extension eines Begriffs M ist die Menge der unter ihn fallenden Objekte O_1, O_2 usw.:

$$M = \text{df } \{O_1, O_2, \dots, O_i, \dots\},$$

wobei „=df“ „ist definitorisch gleich“ bedeuten möge. Die Intension bestimmt den Begriff M durch seine Merkmale f_1, f_2

usw., wobei die meisten dieser Merkmale mit „und“ (\wedge) und einige dieser Merkmale eine Teilmenge bilden, deren Elemente mit „oder“ (\vee) verknüpft sind (Λ sei der Allquantor im Sinne: „gilt für alle“):

$$M = \text{df } \Lambda x. f_1(x) \wedge f_2(x) \wedge \dots \wedge [f_g(x) \vee f_g'(x) \vee \dots \vee f_g''(x)].$$

Alle Merkmale f außer f_g (den „gemüseartigen“ Merkmalen) kommen dem Begriff zwingend zu, während bei den f_g nur einige zutreffen.

Wie kommen wir überhaupt zu Begriffen? Hier erscheint ein Ausflug in die Erkenntnistheorie sinnvoll. Birger Hjorland (2009) unterscheidet mit Empirismus, Rationalismus, Hermeneutik und Pragmatismus vier Herangehensweisen an diese Thematik. Der *Empirismus* geht von Beobachtungen aus; man sucht demnach nach Begriffen in zu analysierenden konkret vorliegenden Texten. Typische informationswissenschaftliche Methoden sind in diesem Kontext Ähnlichkeitsberechnungen zwischen Textwörtern, aber auch zwischen Tags beim Einsatz von Folksonomies sowie die clusteranalytische Verdichtung der Ähnlichkeitsrelationen. Der *Rationalismus* ist der Zuverlässigkeit der Beobachtungen gegenüber skeptisch und konstruiert a priori Begriffe und deren Merkmale und Beziehungen, und dies in der Regel unter Zuhilfenahme analytischer und formal-logischer Methoden. In der Informationswissenschaft erkennen wir solch ein Vorgehen in der formalen Begriffsanalyse (Ganter & Wille, 1996; Priss, 2006). *Hermeneutik* (bei Hjorland, 2009, 1525, „Historizismus“ genannt) erfasst Begriffe in ihrer historischen Entwicklung sowie in ihrer Verwendung in einem gegebenen sog. „Welthorizont“. Beim Verstehen spielt das „Geworfensein“ des Menschen in die Welt eine hervorsteckende Rolle (Heidegger, 1967[1927]). Ein Text wird nie ohne Verständnis und vorurteilsfrei gelesen. Hier entsteht der hermeneutische Zirkel: Der Text als Ganzes gibt den Schlüssel zum Verständnis seiner Teile, gleichzeitig benötigt der Interpretierende die Teile zum Verständnis des Ganzen (Gadamer, 1975). Vorurteile spielen dabei eine positive Rolle. Wir bewegen uns dynamisch in und mit dem Horizont, im Verstehen letztendlich vollzieht sich eine Horizontverschmelzung (Stock & Stock, 2008, 93). Die hermeneutische Auffassung führt in der Informationswissenschaft zur Erkenntnis, dass auch Begriffsordnungen und sogar bibliographische Nachweise (in ihren Indexierungsfeldern) stets dynamisch und Veränderungen unterworfen sind (Gust von Loh, Stock & Stock, 2009). *Pragmatismus* ist eng mit der Hermeneutik verbunden, er betont jedoch die Bedeutung von Zwecken und Zielen. Bei Begriffen gilt es also stets mitzubeachten, wofür diese überhaupt eingesetzt werden: „The ideal of pragmatism is to

define concepts by deciding which class of things best serves a given purpose and then to fixate this class in a sign“ (Hjørland, 2009, 1527). Jede der vier erkenntnistheoretischen Theorien hat Relevanz für die Konstruktion von Begriffen und Relationen in der informationswissenschaftlichen Forschung sowie in der Informationspraxis und sollte beim Aufbau und der Pflege von Knowledge Organization Systems stets gebührende Beachtung finden.

Begriffsarten

Begriffe sind die kleinsten semantischen Einheiten in Wissensordnungen, sie sind „Wissenseinheiten“ (Dahlberg, 1986, 10). Eine Wissensordnung ist eine Begriffsordnung in einer gegebenen Wissensdomäne. In der Wissensrepräsentation wird ein Begriff durch solche Worte festgeschrieben, die die gleiche oder doch zumindest eine ähnliche Bedeutung tragen (deshalb auch die mitunter anzutreffende Bezeichnung „Synset“ im Sinne von „set of synonyms“ für Begriffe). In erster Annäherung und im Einklang mit DIN 2342/1 (1992, 3) ist die Synonymie die „Beziehung zwischen Benennungen, die denselben Begriff bezeichnen“. Es gibt eine weitere Variante der Synonymie, die eine Relation zwischen zwei Begriffen ausdrückt. Wir kommen in Teil 2 darauf zurück.

Beispiele für Synonyme sind *Samstag* und *Sonnabend* oder *Tierarzt* und *Veterinär*. Ein Sonderfall von Synonymie sind Paraphrasen, also Umschreibungen eines Objekts. Manchmal muss man mit Paraphrasen arbeiten, da es für den betreffenden Begriff keine Benennung gibt. Im Deutschen verfügen wir zwar über eine Benennung für gestillten Hunger (*satt*), aber keine über gestillten Durst (Bertram, 2005, 41). Hier haben wir einen Begriff, aber keine Benennung dafür.

Homonymie setzt bei den Benennungen an; sie ist die „Beziehung zwischen übereinstimmenden Benennungen für unterschiedliche Begriffe“ (DIN 2342/1:1992, 3). Ein Beispiel für ein Homonym ist *Java*. Dieses Wort steht für die Begriffe *Java (Insel)*, *Java (Kaffee)* und *Java (Programmiersprache)*. Bei wortorientierten Retrievalsystemen führen Homonyme zu enormen Problemen, da jede homonyme und damit mehrdeutige Wortform – automatisch oder im Mensch-Maschine-Dialog – disambiguiert werden muss. Spielarten der Homonymie (Löbner, 2003, Kap. 3) sind die Homophonie, die Mehrdeutigkeit in der Lautung (z.B. *mehr* und *Meer*), sowie die Homographie, die Mehrdeutigkeit in der Schreibung (z.B. *Tenor* versus *Tenor*). Homophone spielen eine wesentliche Rolle bei Informationssystemen, die mit gesprochener Sprache umgehen, Homographen sind bei Systemen

der Bearbeitung geschriebener Texte zu beachten.

Viele Begriffe haben eine Bedeutung, die ohne Bezug auf andere Begriffe voll verstanden wird, z.B. *Stuhl*. Albert Menne (1980, 48) nennt vollständige Begriffe „kategoriematisch“. In Wissensordnungen, die hierarchisch sortiert sind, ist es durchaus möglich, dass auf einer bestimmten hierarchischen Ebene solch ein Begriff vorkommt:

... mit Filter.

Dieser Begriff ist synkategoriematisch; er ist unvollständig und auf andere Begriffe angewiesen, um Bedeutung zu erlangen. Menne (1980, 46 f.) beschreibt Synkategoriemata:

Synkategoriematische oder unvollständige oder unselbständige Bedeutung hat ein Ausdruck, der auf einen anderen Ausdruck bezogen ist, erst im Zusammenhang mit einem anderen Ausdruck voll verstanden werden kann. So ist z.B. der Ausdruck „von 10-12“, schwarz auf einem weißen, rechteckigen Schild, als Verkehrszeichen erst dann voll verständlich, wenn er z.B. unter einem Zeichen für Halteverbote steht.

In hierarchischen Wissensordnungen werden die Synkategoriemata durch die über ihnen platzierten Begriffe erklärt. Erst jetzt wird die Bedeutung klar:

Zigarette
... mit Filter

oder

Schornstein
... mit Filter.

Einmal geht es im Beispiel um eine Filterzigarette, zum andern um einen Schornstein mit (Abgas-)Filter. Eine solche Klärung kann sich mitunter über mehrere Hierarchieebenen hinziehen. Es ist demnach sehr unpraktisch, etwa in einem Register Synkategoriemata allein und ohne ihre Ergänzungen einzutragen.

Begriffe sind nicht – wie physikalische Objekte – gegeben, sondern werden durch Abstraktion aus der Welt der Gegenstände aktiv gewonnen (Klaus, 1973, 214). Die Aspekte der Begriffsbildung (im informationswissenschaftlichen, nicht etwa im psychologischen Sinne) werden in erster Linie über Definitionen geklärt. Allgemein kann festgestellt werden, dass die Begriffsbildung im Rahmen von Wissensordnungen im Spannungsfeld zweier gegenläufiger Prinzipien abläuft. Ein ökonomisches Prinzip leitet dazu an, nicht zu viele Begriffe in eine Wissensordnung aufzunehmen. Wenn zwei Begriffe in Extension und Intension mehr oder minder ähnlich sind, werden diese als „quasi-synonym“ als ein einziger Begriff angesehen. Das Prinzip des Informationsgehalts weist in die umgekehrte Richtung. Je feiner wir bei Extension und Intension unterscheiden, desto größer wird der Infor-

mationsgehalt jedes einzelnen Begriffs. Hiervon profitieren Homogenität und Exaktheit der Begriffe. Lloyd K. Komatsu (1992, 501) erläutert die Problemsituation (er benutzt „category“ für „Begriff“):

Thus, economy and informativeness trade off against each other. If categories are very general, there will be relatively few categories (increasing economy), but there will be few characteristics that one can assume different members of a category share (decreasing informativeness) and few occasions on which members of the category can be treated as identical. If categories are very specific, there will be relatively many categories (decreasing economy), but there will be many characteristics that one can assume different members of a category share (increasing informativeness) and many occasions on which members can be treated as identical.

Die Lösung für Begriffsbildungen (Komatsu, 1992, 502, verwendet „categorization“) in Wissensordnungen liegt in einem Kompromiss:

The basic level of categorization is the level of abstraction that represents the best compromise between number and informativeness of categories.

Gemäß der Theorie von Eleanor Rosch (Mervis & Rosch, 1981; Rosch, 1975a; Rosch 1975b; Rosch & Mervis, 1975; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976) haben wir drei Ebenen von Begriffen zu unterscheiden: die übergeordnete Ebene („superordinate level“), die Grundebene („basic level“) und die untergeordnete Ebene („subordinate level“):

Suppose that basic objects (e.g., chair, car) are the most inclusive level at which there are attributes common to all or most members of the category. Then total cue validities are maximized at that level of abstraction at which basic objects are categorized. That is, categories one level more abstract will be superordinate categories (e.g., furniture, vehicle) whose members share only a few attributes among each other. Categories below the basic level will be subordinate categories (e.g. kitchen chair, sports car) which are also bundles of predictable attributes and functions, but contain many attributes which overlap with other categories (for example, kitchen chair shares most of its attributes with other kinds of chairs) (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976, p. 385).

So ist auf der Grundebene etwa der Begriff *Stuhl* für viele Menschen ein guter

Kompromiss zwischen dem zu allgemeinen *Möbelstück* und den zu spezifischen Begriffen *Armstuhl*, *Chippendalestuhl* usw. In einer Wissensordnung für Möbel sieht der Kompromiss anders aus, da hier weitaus genauer differenziert werden muss: Hier werden wir die Begriffe der untergeordneten Ebene hinzufügen. Bauen wir dagegen eine Wissensordnung für Wirtschaftswissenschaften auf, so könnte der Kompromiss durchaus zugunsten von *Möbelstück* ausfallen; so dass wir uns in diesem Fall auf einen Begriff der übergeordneten Ebene beschränken.

Begriffe, deren Extension genau ein Element aufweist, sind Individualbegriffe, ihre Bezeichnungen sind Eigennamen, z.B. von Personen, Organisationen, Ländern, Produkten, aber auch von singulären historischen Ereignissen (*Wiedervereinigung Deutschlands*) oder einzelnen wissenschaftlichen Lehrsätzen (*Zweiter Hauptsatz der Thermodynamik*). Alle anderen Begriffe sind Allgemeinbegriffe (Dahlberg, 1974, 16). Als eine besondere Form von Allgemeinbegriffen heben wir die Kategorien hervor. Wenn wir uns auf den Abstraktionsebenen nach oben bewegen, kommen wir irgendwann an die Spitze. An dieser Stelle – wohlge­merkt: stets im Rahmen einer Wissensdomäne – ist kein weiterer Abstraktionsschritt möglich. Diese Top-Begriffe stellen die domänenspezifischen Kategorien dar. Fugmann (1999, 23) führt die Kategorie über die Intension der Begriffe ein:

Beim fortgesetzten Voranschreiten zu immer allgemeineren Begriffen gelangt man in der Abstraktionshierarchie auf einem jeden Fachgebiet an eine Grenze, jenseits welcher kein noch allgemeinerer, noch merkmalsärmerer Oberbegriff mehr sinnvoll ist. In der Hierarchie, welcher das Vitamin C angehört, würde man etwa beim Begriff Stoff diese oberste Grenze erreichen. Ein noch allgemeinerer Begriff ist kaum denkbar, wenn er noch sinnvoll sein soll.

In facettierten Wissensordnungen bilden die Kategorien das Raster für die Facetten.

Nach Fugmann kann man die Begriffsarten intensional unterscheiden. Kategorien sind Begriffe, die über ein Minimum an Merkmalen verfügen (noch allgemeinere Begriffe zu bilden, hieße für die Wissensdomäne, leere, unbrauchbare Begriffe zu bilden). Individualbegriffe sind Begriffe, die über ein Maximum an Merkmalen verfügen (selbst wenn man weitere Merkmale hinzufügt, bleibt die Extension gleich). Allgemeinbegriffe sind alle die Begriffe, die zwischen den Extremen liegen. Durch ihre exponierte Lage lassen sich sowohl Individualbegriffe als auch Kategorien in Methoden der Wissensrepräsentation recht einfach bearbei-

ten, während Allgemeinbegriffe durchaus zu Problemen führen.

Vagheit und Prototyp

Individualbegriffe und Kategorien lassen sich in aller Regel exakt bestimmen. Wie sieht es mit der Exaktheit von Allgemeinbegriffen aus? Wir bleiben bei unserem Beispielbegriff *Stuhl* und folgen Max Black (1937, 433) in seine imaginäre Stuhlausstellung:

One can imagine an exhibition in some unlikely museum of applied logic of a series of "chairs" differing in quality by at least noticeable amounts. At one end of a long line, containing perhaps thousands of exhibits, might be a Chippendale chair; at the other, a small nondescript lump of wood. Any "normal" observer inspecting the series finds extreme difficulty in "drawing the line" between chair and not-chair.

Durch die minimalen Abweichungen bei benachbarten Objekten dürfte es praktisch unmöglich sein, eine Grenze zwischen Stuhl und Nicht-Stuhl zu ziehen. Außer dem „Neutralbereich“, bei dem wir nicht wissen, ob ein Begriff zutrifft oder nicht, haben wir an einer Seite Objekte, die eindeutig unter den Begriff fallen, und auf der gegenüberliegenden Seite andere Objekte, die eindeutig nicht darunter fallen. Die Grenzen zwischen dem Neutralbereich und seinen Nachbarn sind allerdings auch nicht exakt bestimmbar. Solche unscharfen Grenzen (Löbner, 2003, 262 f.) lassen sich experimentell für viele Allgemeinbegriffe aufzeigen.

Als Lösung bietet sich an, erst gar nicht nach den Grenzen des Begriffs zu suchen, sondern stattdessen mit einem „Prototypen“ zu arbeiten (Rosch, 1983). Ein solcher Prototyp kann „als bestes Beispiel“ (Löbner, 2003, 265) für einen Basic-Level-Begriff betrachtet werden. Dieses Musterbeispiel besitzt „gute“ Merkmale im Sinne eines hohen Erkennungswertes (Löbner, 2003, 269):

Das Merkmal, Federn zu haben, ist ein „gutes“ Merkmal für den Prototypen, weil es gut geeignet ist Vögel von Nichtvögeln zu unterscheiden. ... (D)as Merkmal 'hat Federn' hat einen hohen Erkennungswert für die Kategorie Vogel, weil fast alle Mitglieder es haben, während es allen Nichtmitgliedern fehlt. ... Am Ende ist es die Kombination von Merkmalen mit mehr oder weniger hohem Erkennungswert, die einen geeigneten Prototyp ausmacht.

Wenn wir den Begriff durch einen Prototypen und dessen Merkmalen intensional bestimmen, sind die unscharfen Grenzen zwar immer noch existent (und bewirken vielleicht den einen oder anderen

Fehler beim Indexieren in diesen Grenzregionen), aber wir können überhaupt mit Allgemeinbegriffen zufriedenstellend arbeiten. Wenn wir uns eine Begriffshierarchie über mehrere Ebenen vorstellen, so dürften insbesondere bei den mittleren Ebenen, also im Basic Level gemäß Rosch, Prototypen eine wesentliche Rolle spielen. Am oberen Ende der Hierarchie stehen die (übergeordneten) Begriffe mit wenigen Merkmalen, so dass sich sehr wahrscheinlich keine Vorstellung eines Prototyps einstellen kann. Am unteren Ende sind die (untergeordneten) Begriffe derart speziell, dass der Begriff mit dem Prototypen zusammenfällt.

Definition

In der Praxis der Wissensrepräsentation werden Begriffe häufig nur implizit – z.B. durch Angabe ihrer Synonyme und ihrer Verortung im semantischen Umfeld – definiert. Wir sind der Meinung, dass bei Wissensordnungen die verwendeten Begriffe ausdrücklich zu definieren sind, denn nur so kann Klarheit sowohl für die Indexer als auch für die Nutzer erreicht werden.

Definitionen müssen einigen Kriterien entsprechen, wenn sie korrekt eingesetzt werden (Dubislaw, 1981, 130; Pawłowski, 1980, 31-43). Zu vermeiden ist die Zirkularität, die Definition eines Begriffs mit Hilfe desselben Begriffs, die sich – nunmehr als mittelbarer Zirkel – auch über mehrere Definitionsschritte hinweg auffinden lässt. Wenig hilfreich ist die Definition eines unbekanntem Begriffs durch einen anderen, ebenso wenig bekannten (ignotum per ignotum). Die Inadäquatheit von Definitionen zeigt sich darin, dass sie entweder zu eng (wenn Objekte, die eigentlich unter den Begriff fallen, ausgeschlossen werden) oder zu weit (wenn Objekte darunter fallen, die nicht hingehören) sind. In vielen Fällen sind negative Definitionen (*Punkt ist, was keine Ausdehnung hat*) unbrauchbar, da sie oftmals zu weit sind (Menne, 1980, 32). Eine Definition soll keine überflüssigen Merkmale des Begriffs aufweisen (Menne, 1980, 33). Natürlich muss die Definition präzise sein (also z.B. keine bildlichen Floskeln verwenden) und darf keine Widersprüche (wie *blinder Zuschauer*) in sich enthalten. Persuasive Definitionen, also Begriffsabgrenzungen mit der Hoffnung auf (oder der Nebenwirkung von) emotionale(n) Reaktionen (z.B. frei nach Buddha *Paria ist ein Mensch, der sich von der Wut und dem Hass hinreißen lässt, ein heuchlerischer Mensch, voller Betrug und Fehler* ... Pawłowski, 1980, 250) sind in der Wissensrepräsentation nicht zu gebrauchen. Oberstes Ziel ist die Nützlichkeit der Definition in der betreffenden Wissensdomäne (Pawłowski, 1980, 88 ff.). Gemäß unserer Kenntnisse über Vag-

heit bemühen wir uns, nicht unbedingt alle Objekte exakt unter einen Begriff zu zwängen, sondern definieren stattdessen manchmal den Prototypen.

Aus der Fülle von unterschiedlichen Definitionsarten (u. a. Definition als Abkürzung, Explikation, Nominal- und Realdefinition) sind für die Wissensrepräsentation sowohl die Begriffserklärung als auch die Definition über die Familienähnlichkeit besonders wichtig.

Die Begriffserklärung geht davon aus, dass Begriffe aus Teilbegriffen zusammengesetzt sind:

Begriff =df Teilbegriff₁, Teilbegriff₂, ...

Hierbei kann man in beiden Richtungen arbeiten. Die Begriffssynthese geht von den Teilbegriffen aus, während die Begriffsanalyse beim Begriff startet. Die klassische Variante stammt von Aristoteles und erklärt einen Begriff durch die Angabe von Genus und Differentia. Aristoteles arbeitet Kriterien heraus, nach denen Begriffe voneinander zu differenzieren sind und nach denen sie in eine hierarchische Ordnung gebracht werden. Die Erkenntnis der Verschiedenheit von Gegenständen ergibt sich über zwei Schritte, zunächst über das, was zwei Gegenständen gleich ist – Aristoteles nennt dies die „Gattung“ – und sodann das, was beide Gegenstände innerhalb der Gattung als spezifische „Arten“ unterscheidet. Wir lesen in der „Metaphysik“ (1057b 34 ff.):

Dasjenige, das der Art nach verschieden ist, ist von etwas in etwas verschieden, und dieses muss beiden zukommen. Wenn etwa ein Lebewesen der Art nach [von einem anderen] verschieden ist, so sind beide Lebewesen. Es muss demnach das, was der Art nach verschieden ist, zur selben Gattung gehören ... Es muss also an sich das eine ein derartiges Lebewesen sein, das andere ein solches, wie etwa das eine Pferd, das andere Mensch. Es muss also dieser Unterschied eine Verschiedenheit der Gattung sein. Ich verstehe nämlich unter „Unterschied der Gattung“ eine Verschiedenheit, die die Gattung selbst verschieden macht.

Zu einer Begriffsbestimmung gehören demnach zwingend die Angabe der Gattung und der Unterschied zwischen den Arten (Aristoteles, Topik, Buch 1, Kap. 8):

(Die Definition (besteht) aus der Gattung und den Art-Unterschieden.

Es ist wichtig, stets die jeweils nächstgelegene Gattung zu finden und nicht etwa eine Hierarchieebene zu überspringen (Topik, Buch 6, Kap. 5):

(H)ierher gehört der Fehler, wenn man die Gattung überspringt, z.B. wenn man die Gerechtigkeit eine Gemütsrichtung nennt, welche die Gleichheit bewirkt ...; denn der Definierende überspringt dabei die Tugend.

Was steuert das Unterscheiden der Arten einer Gattung? Aristoteles hält zwei Aspekte auseinander, zum einen die zufällige Beschaffenheit eines Gegenstandes (etwa dass ein Pferd einen Schwanz hat und ein Mensch nicht) und zum andern die wesentlichen, die spezifischen Merkmale, die den Unterschied ausmachen

Homo ist der zu definierende Begriff, animal der Gattungsbegriff und rational die spezifische Eigenschaft, die die Menschen unverwechselbar von anderen Lebewesen abhebt. Ein Fehler wäre, den Menschen durch *Lebewesen* und *Haarfarbe nicht blond* zu definieren, weil (im Gegensatz zu den Blondinenwitzen) die Haarfarbe zu den zufälligen, aber nicht zu den wesenskonstitutiven Merkmalen gehört. Dadurch, dass im Verlaufe von Begriffserklärungen über mehrere Ebenen hinweg von oben nach unten stets weitere, neue Merkmale hinzukommen, werden die Be-

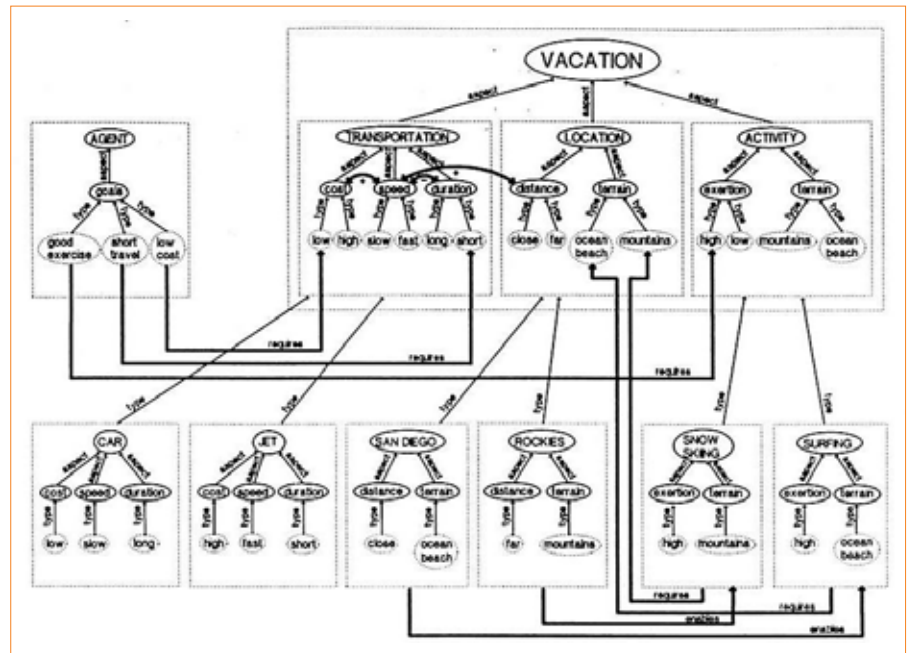


Abbildung 2: Regelmäßige Zusammenhänge zwischen Attributen und Werten.
Quelle: Barsalou 1992, 38.

(im Beispiel: dass der Mensch über Vernunft verfügt und das Pferd nicht). Im Mittelalter wird diese These des Aristoteles auf folgende einprägsame Form gebracht (Menne, 1980, 28): „Definitio fit per genus proximum et differentiam specificam“. Die Begriffserklärung arbeitet demnach mit folgenden Teilbegriffen:

Teilbegriff₁: Genus (Begriff der direkt übergeordneten Gattung),

Teilbegriff₂: Differentia specifica (wesenskonstitutiver Unterschied zu den Schwesterbegriffen).

Die Merkmale, die einen Begriff von seinen Schwesterterminen (das sind diejenigen Begriffe, die demselben Genus angehören) abheben, müssen stets eine spezifische und nicht etwa eine zufällige Eigenschaft (accidens) darstellen. Sowohl bei den Gattungs- als auch bei den Differentia-Teilbegriffen sind jeweils mehrere Begriffe möglich. Eine klassische Definition nach dieser Definitionsart ist:

Homo est animal rationale.

griffe immer spezifischer; in Gegenrichtung (bei Verlust von Merkmalen auf dem Weg nach oben) immer allgemeiner. Das heißt auch, dass bei einer Begriffsleiter von oben nach unten die Merkmale stets „vererbt“ werden. Die Begriffserklärung ist für Wissensordnungen besonders wichtig, weil sie durch ihre Vorgaben die Begriffe zwangsläufig in eine hierarchische Struktur einbettet.

Bei der Begriffserklärung geht man davon aus, dass einem Objekt die spezifischen Merkmale zur Gänze angehören, wenn es zur entsprechenden Klasse gehört; die Merkmale werden durch ein logisches Und zusammengehalten. Bei den gemüseartigen Begriffen, bei denen nur eine Familienähnlichkeit zwischen den Objekten auszumachen ist, gilt dies nicht. Hier werden die Merkmale durch ein Oder verknüpft. Pawłowski (1980, 199) stellt zu diesen Begriffen fest:

Die Extension eines solchen Begriffes besteht nicht aus einer Menge von Objekten, für die eine Konjunktion der all

diesen Objekten und nur diesen Objekten zukommenden Eigenschaften gilt.

Wenn wir die Begriffserklärung mit der Definition nach Familienähnlichkeit verknüpfen, müssen wir auf gewissen hierarchischen Ebenen mit einer Disjunktion von Merkmalen arbeiten. Auch hier sind wir auf der Suche nach einem Gattungsbegriff, z.B. zu Wittgensteins *Spiel*. Die Familienmitglieder von *Spiel*, also Brettspiel, Kartenspiel, Glücksspiel usw. haben durchaus jeweils einige Merkmale gemeinsam, aber eben nicht alle. Es gilt, dass Begriffe von oben nach unten stets spezifischer und umgekehrt allgemeiner werden; die Vererbung der Merkmale von oben nach unten gilt jedoch nicht. In denjenigen Hierarchieebenen, in denen über Familienähnlichkeit definiert worden ist, vererben die Begriffe nur einige ihrer Merkmale, aber nicht alle. Nehmen wir z.B. an, das Genus von *Spiel* sei *Freizeitbeschäftigung*. Wir müssen nun zwecks Abgrenzung von anderen Freizeitbeschäftigungen (etwa *Meditieren*) Merkmale der Spiele angeben. Wir definieren:

Teilbegriff1 / Genus:
Freizeitbeschäftigung

Teilbegriff2 / Differentia specifica:
Glücksspiel ∨ Kartenspiel ∨ Brettspiel ∨
Kampfspiel ∨ ...

Wenn wir uns nun weiter in der Begriffsleiter nach unten bewegen, wird klar, dass *Spiel* nicht alle seine Merkmale vererbt, sondern stets nur Teilmengen (ein Glücksspiel muss kein Kartenspiel sein). Auf den unteren Ebenen braucht nicht wiederum Familienähnlichkeit vorzuherrschen, sondern die „normale“ (konjunktive) Begriffserklärung. Man muss auf jeder Ebene überprüfen, ob dort mithilfe der Familienähnlichkeit disjunktiv oder „normal“ konjunktiv definiert worden ist.

Frames

Wie lässt sich ein Begriff repräsentieren? Ein erfolgreicher Ansatz arbeitet mit Frames (Minsky, 1975). Frames haben sich sowohl in der Kognitionswissenschaft, in der Informatik (Reimer, 1991, 159 ff.) als auch in der Linguistik bewährt. In der Konzeption von Lawrence W. Barsalou (1992, 29) verfügen Frames über drei grundlegende Komponenten:

- Mengen von Attributen und Werten (Petersen, 2007),
- strukturelle Invarianten,
- regelhafte Zusammenhänge.

Wir bevorzugen bei den unterschiedlichen Frame-Konzeptionen die Version von Barsalou, weil hier regelhafte Zusammenhänge Berücksichtigung finden.

Diese Option benötigen wir, auf der Anwendungsseite einer Begriffsordnung automatisch Berechnungen durchführen zu können.

Der Kern eines jeden Frame ordnet einem Begriff (in Abbildung 2 ist dies beispielsweise *Vacation*) Merkmale (*Transportation, Location, Activity*) und den Merkmalen jeweils Werte (hier: weitere Frames) zu, wobei sowohl Merkmale als auch Werte durch Begriffe ausgedrückt werden. Im Sinne von Minsky (1975) werden dem Begriff solche Attribute zugeordnet, die eine stereotype Situation beschreiben. Zwischen den Begriffen innerhalb eines Frame bestehen strukturelle Invarianten, die durch Relationen ausgedrückt werden (Barsalou 1992, 35 f.):

Structural invariants capture a wide variety of relational concepts, including spatial relations (e.g., between seat and back in the frame for chair), temporal relations (e.g., between eating and paying in the frame for dining out), causal relations (e.g., between fertilization and birth in the frame for reproduction), and intentional relations (e.g., between motive and attack in the frame for murder).

Die Begriffe innerhalb des Frame sind nicht unabhängig, sondern bilden vielfach Zusammenhänge, die gewissen Regeln gehorchen. In Barsalous Freizeits-Frame liegen auf der Ebene der Attribute beispielsweise zwischen den Transport-Attributen positive (je schneller die Fahrt, desto höher die Kosten) und negative Zusammenhänge (je schneller die Fahrt, desto kürzer die Reisedauer) vor. Auf der Ebene der Attribute betrachten wir den Wert für den Ort *San Diego* und bei der Aktivität den Wert *Surfen*. Es ist klar, dass der erste Wert den zweiten ermöglicht (man kann in San Diego, aber nicht im Gebirge Wellensurfen). Über eine Formulierung in einer terminologischen Logik (description logic; Nardi & Brachman, 2003) sowie der Trennung von Allgemeinbegriffen (in einer TBox) und Individualbegriffen (in der ABox) kann man innerhalb einer Begriffsordnung – im Sinne von Ontologien – die Option automatischen Schlussfolgerns einführen. Hat man an gewissen Stellen bei den Werten Zahlen vorliegen, kann man diese als Basis automatischer Berechnungen einsetzen (siehe Teil 3).

Barsalou (1992, 43) sieht (zumindest theoretisch) keine Grenzen, Frames in der Wissensrepräsentation einzusetzen. Es bedarf jedoch diverser Vorarbeiten, die einem automatisierten System gegeben werden müssen:

Before a computational system can build the frames described here, it

needs a powerful processing environment capable of performing many difficult tasks. This processing environment must notice new aspects of a category (in unserer Terminologie: Begriff, Verf.) to form new attributes. It must detect values of these attributes to form attribute-value sets. It must integrate cooccurring attributes into frames. It must update attribute-value sets with experience. It must detect structural invariants between attributes. It must detect and update constraints. It must build frames recursively for the components of existing frames.

Leitet die Definition als Begriffserklärung immerhin zu einer Relation (der Hierarchie), so führt der Frame-Ansatz zu einer Fülle von Relationen zwischen Begriffen und darüber hinaus zu regelhaften Zusammenhängen. Da Wissensordnungen zwingend auf Relationen angewiesen sind, fundieren Frames – als Repräsentanten von Begriffen – optimal solche Methoden der Wissensrepräsentation. Das letzte Zitat von Barsalou sollte aber zum Nachdenken anregen, die Fülle der Relationen und Regeln nicht zu groß werden zu lassen. Schließlich müssen alle geschilderten Vorarbeiten und Updates auch faktisch durchgeführt werden – und das bedeutet einen riesigen Aufwand. Zusätzlich steht zu befürchten, dass mit der Zunahme der Anzahl unterschiedlicher Relationen der Umfang der Wissensdomäne immer kleiner wird, in dessen Rahmen man arbeiten kann. Gemäß Daniele Nardi und Ronald J. Brachman (2003, 10) besteht nämlich in der Praxis ein gegenläufiger Zusammenhang zwischen der Ausdrucksstärke der Sprache und dem automatischen Schlussfolgern:

(T)here is a tradeoff between the expressiveness of a representation language and the difficulty of reasoning over the representation built using that language. In other words, the more expressive the language, the harder the reasoning.

KOS-Designer sollten demnach die Anzahl der spezifischen Relationen so klein wie möglich halten, ohne natürlich die Spezifika der jeweiligen Wissensdomäne aus den Augen zu verlieren.

2 Begriffsordnungen

Semantische Relationen

Begriffe existieren nicht losgelöst voneinander, sondern sind miteinander verknüpft. Bei den Definitionen (etwa über Begriffserklärung) sowie bei den Frames stoßen wir auf solche Relationen. Beziehungen zwischen Begriffen wollen wir „semantische Relationen“ (Khoo & Na,

2006; Storey, 1993) nennen. Dies ist nur ein Teil der Relationen, die von Interesse für die Wissensrepräsentation sind. Bibliographische Relationen (Green, 2001, 7 ff.) erfassen Beziehungen, die Dokumente formal beschreiben (z.B. „hat Autor“, „ist in Quelle erschienen“, „hat Jahrgang“). Faktographische Relationen werten Beziehungen aus, die in gewissen Wissensdomänen wichtig sind (wie z.B. „hat Schmelzpunkt“ in einer Datenbank zu Werkstoffen oder „hat Tochterfirma“ in einem Unternehmensdossier). Relationen bestehen auch *zwischen* Dokumenten, insofern beispielsweise wissenschaftliche Dokumente zitieren und zitiert werden oder Web-Dokumente über Links verfügen. Wir konzentrieren uns hier ausschließlich auf die semantischen Relationen. Erst das Gesamt aus Begriffen und semantischen Relationen macht eine Begriffsordnung aus. Findet eine Begriffsordnung auf eine gewisse Wissensdomäne Anwendung, so reden wir von einer „Wissensordnung“ oder einem „Knowledge Organization System“ (KOS).

In der Informationswissenschaft unterscheiden wir bei den semantischen Relationen zwischen paradigmatischen und syntagmatischen Beziehungen (Peters & Weller, 2008a; Stock, 2007, 451 f.). Diese Differenzierung geht auf Ferdinand de Saussure (2001[1916], 148) zurück (de Saussure verwendet „assoziativ“ anstelle von „paradigmatisch“):

Die syntagmatische oder Anreihungsbeziehung besteht in praesentia: sie beruht auf zwei oder mehreren in einer bestehenden Reihe neben einander vorhandenen Gliedern. Im Gegensatz dazu verbindet die assoziative Beziehung Glieder in absentia in einer möglichen Gedächtnisreihe.

Im Kontext der Wissensrepräsentation bilden die paradigmatischen Relationen „fest verdrahtete“ Beziehungen, die in einer bestimmten Wissensordnung festgeschrieben (oder festgehalten) worden sind. Sie gelten unabhängig von Dokumenten (also „in absentia“ vom konkreten Auftreten in Dokumenten). Syntagmatische Relationen bestehen zwischen Begriffen in konkreten Dokumenten; sie liegen also stets „in praesentia“ vor. Es geht hier um das gemeinsame Vorkommen, sei es im Fließtext des Dokuments (oder eines Textfensters), sei es bei den markierten Stich- oder Schlagworten. Gernot Wersig (1974, 253) definiert:

(D)ie paradigmatischen Relationen (sind) jene, die durch die Dokumentationsprache vorgegeben sind, während die syntagmatischen Relationen die sind, die durch die Dokumentationsprache realisiert werden können.

Wir wollen dies an einem kleinen Beispiel verdeutlichen! In einer Wissensordnung kommen die beiden hierarchischen Relationen

Österreich – Steiermark – Bezirk
Graz-Umgebung – Lassnitzhöhe;

Speiseöl – Pflanzenöl – Kürbiskernöl

vor. Diese Begriffsbeziehungen bilden jeweils paradigmatische Relationen. Ein wissenschaftlicher Artikel über lokale Besonderheiten der Landwirtschaft in der Steiermark sei wie folgt indexiert:

Lassnitzhöhe – Kürbiskernöl.

2009) vorkommt. Im Sinne eines Bottom-up-Ansatzes des Aufbaus von Wissensordnungen bieten Folksonomies sowohl empirisches Material für Kandidaten des kontrollierten Vokabulars als auch Material für paradigmatische Relationen, obwohl letzteres in den Folksonomies nur „versteckt“ vorkommt (Peters & Weller, 2008a, 104) und über die Analyse der Tag-Co-Occurrences intellektuell freigelegt werden muss. Für Isabella Peters und Katrin Weller ist die (automatische wie intellektuelle) Bearbeitung von Tags und ihren Relationen in Folksonomies Aufgabe des sog. „Tag Gardening“ (Peters & Weller, 2008b).

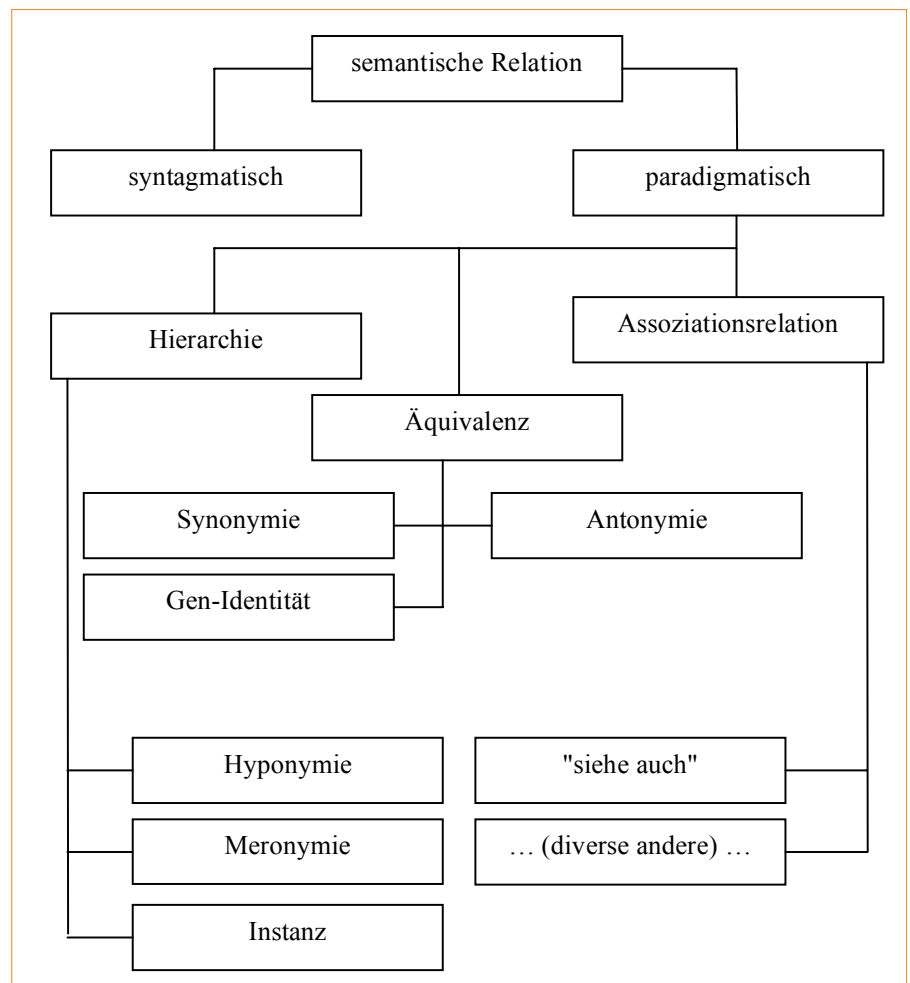


Abbildung 3: Semantische Relationen.

Diese beiden Begriffe bilden demnach eine syntagmatische Relation. Bis auf Ausnahmen (bei Spielarten des syntaktischen Indexierens) wird die syntagmatische Relation nicht näher beschrieben. Sie drückt aus: Im gegebenen Dokument geht es um Kürbiskernöl und um Lassnitzhöhe. Sie sagt aber nicht, in welchen konkreten Beziehungen die Terme zueinander stehen.

Die syntagmatische Relation ist die einzige semantische Relation, die bei Folksonomies (Peters & Stock, 2008; Bertram,

Bei den paradigmatischen Relationen wird dagegen stets die Art des Zusammenhangs ausgedrückt. Aus der Vielzahl möglicher paradigmatischer Relationen versucht die Wissensrepräsentation solche herauszufinden, die generalisierbar sind, die also in allen oder zumindest in vielen Anwendungsfällen sinnvoll einsetzbar sind. Einen Überblick über semantische Relationen verschafft uns Abbildung 3.

Ordnung und R – S – T

Man kann Relationen anhand der Anzahl ihrer Argumentstellen differenzieren. Zweistellige Relationen verbinden zwei Begriffe, dreistellige drei usw. Dabei ist es immer möglich, die mehrstelligen Relationen durch eine Reihe von zweistelligen zu vereinfachen. *Heilen* beispielsweise ist eine dreistellige Relation zwischen einer Person, einer Krankheit und einem Heilmittel. Aufgelöst ergeben sich drei zweistellige Relationen: Person – Krankheit, Krankheit – Heilmittel, Heilmittel – Person. Wir gehen in diesem Kapitel davon aus, dass die behandelten Relationen zweistellig sind.

Ziel ist es, zu einem gewissen Wissensgebiet eine Begriffsordnung zu erstellen, die dann als Wissensordnung fungiert. Ordnungen lassen sich über drei grundlegende Eigenschaften charakterisieren (x, y, z seien Begriffe, ρ jeweils eine Relation). Reflexivität (Rückbezüglichkeit) bei Begriffsordnungen fragt danach, wie ein Begriff bezogen auf eine Relation zu sich selbst steht. Symmetrie (Spiegelbildlichkeit) liegt vor, wenn eine Relation zwischen A und B und auch in der Gegenrichtung B und A existiert. Wenn eine Relation zwischen zwei Begriffen A und B sowie zusätzlich zwischen B und C besteht und dann gilt, dass sie auch zwischen A und C besteht, dann reden wir bei solch einer Relation von Transitivität. Wir wollen dies an einigen Beispielen demonstrieren:

R	Reflexivität	$x \rho x$ „... ist identisch mit ...“
	Irreflexivität	$\neg(x \rho x)$ „... ist Ursache von ...“
S	Symmetrie	$(x \rho y) \rightarrow (y \rho x)$ „... ist gleich ...“
	Asymmetrie	$(x \rho y) \rightarrow \neg(y \rho x)$ „... liebt unglücklich ...“
T	Transitivität	$[(x \rho y) \wedge (y \rho z)] \rightarrow (x \rho z)$ „... ist größer als ...“
	Intransitivität	$[(x \rho y) \wedge (y \rho z)] \rightarrow \neg(x \rho z)$ „... ist ähnlich mit ...“

Eine Ordnung in einem strengen mathematischen Sinne ist irreflexiv ($\neg R$), asymmetrisch ($\neg S$) und transitiv (T) (Menne, 1980, 92). Eine Ordnung, die als einzige Relation z.B. über *ist teurer als* verfügt, erfüllt diese Eigenschaften: eine bestimmte Ware, sagen wir: eine Zitrone, ist nicht teurer als eine Zitrone (also: $\neg R$); wenn eine Ware (unsere Zitrone) teurer ist als eine andere Ware (ein Apfel), so ist der Apfel nicht teurer als die Zitrone – sondern billiger ($\neg S$); wenn letztendlich eine Zitrone teurer als ein Apfel und ein Apfel teurer als eine Kirsche ist, dann ist auch eine Zitrone teurer als eine Kirsche (T).

Bei asymmetrischen Relationen reden wir von einer inversen Relation, wenn in die-

ser die Umkehrung der Ausgangsrelation thematisiert wird. In $(x \rho y)$ sei ρ die Relation *ist Unterbegriff von*. Die inverse Relation ρ' in $(y \rho' x)$ lautet dann *ist Oberbegriff von*.

Insofern eine Wissensordnung über Synonymie verfügt, die ja immer symmetrisch ist (wenn x synonym y ist, dann ist y synonym x), wird sie nie eine Ordnung im mathematischen Sinn sein. Offen ist zudem, ob alle Relationen in Wissensordnungen prinzipiell transitiv sind. In einer ersten naiven Annäherung an das Problem lassen sich leicht Gegenbeispiele finden. Nehmen wir z.B. an, die Leber von Professor X ist Teil von X und Professor X ist Teil der Universität Y, dann gilt bei Transitivität: Die Leber von Professor X ist Teil der Universität Y, was offensichtlich Unfug ist. Aber Achtung! War das überhaupt dieselbe Relation? Die Leber ist ein Körperteil; ein Professor ist ein Teil einer Organisation. Nur weil wir vereinfachend von einer allgemeinen Ganzes-Teil-Relation ausgehen, gilt die Transitivität nicht. Intransitivität kann also einerseits bedeuten, dass die Begriffsordnung (zu Unrecht) unterschiedliche Relationen zu einer einzigen zusammenfasst oder andererseits, dass die Relation in der Tat intransitiv ist.

Warum ist insbesondere die Transitivität wichtig für das Information Retrieval? Zentrale Anwendungen sind die (automatische oder im Nutzer-System-Dialog erarbeitete) Anfrageerweiterung (Stock, 2007, 480 f.) oder (in Ontologien) das automatische Schlussfolgern. Wenn jemand beispielsweise nach Gestüten im Rhein-Erft-Kreis sucht, aber nicht weiß, wo diese genau liegen, wird er formulieren:

Gestüt AND Rhein-Erft-Kreis.

Die wichtigen Gestüte liegen in Quad-rath-Ichendorf, dies ist ein Ortsteil von Bergheim, das seinerseits im Rhein-Erft-Kreis liegt. Wenn wir nun das zweite Argument der Suchanfrage entsprechend der geographischen Gliederung nach unten erweitern, kommen wir beim zweiten Schritt auf diejenige Formulierung, die letztlich die Treffer bilden wird:

Gestüt AND (Rhein-Erft-Kreis OR Bergheim OR ... OR Quad-rath-Ichendorf).

Die Suchfrageerweiterung kann auch auf dem Weg über eine Begriffsleiter nach oben erfolgreich sein. Ein Autofahrer sei mit dem Problem konfrontiert, in einer ihm unbekanntem Gegend eine Werkstatt für seinen Wagen (sagen wir: einen Ford) zu finden. Er formuliert:

Werkstatt AND Ford AND ([Standort], z.B. über GPS ermittelt).

Das Retrievalsystem ordnet den Standort der kleinsten geographischen Einheit zu

und geht zunächst auf der Begriffsleiter einen Schritt aufwärts und zugleich zu den Geschwisterbegriffen wieder hinunter. Wird kein Treffer gefunden, geht es eine Hierarchieebene weiter nach oben und wiederum zu den Geschwistern usw., bis das gewünschte Dokument gefunden worden ist.

Eine Suchfrageerweiterung über exakt einen Schritt ist jederzeit durchführbar. Wenn wir uns die Wissensordnung als Graphen vorstellen, können wir also immer alle diejenigen Begriffe problemlos in die Suchanfrage übernehmen, die über eine Pfadlänge von eins mit dem Ausgangsbegriff verbunden sind. (Ob dies auch immer in der Praxis erfolgreich ist, sei dahingestellt. Insbesondere das Einbeziehen von Oberbegriffen in einer Hierarchie kann die Treffermenge stark ausdehnen und damit der Precision schaden.) Wollen wir über Pfadlängen größer eins erweitern, so muss sichergestellt sein, dass Transitivität vorliegt, da ansonsten kein zwingender semantischer Bezug zum Ausgangsbegriff gegeben ist.

Koordinationsgrad

Ein Begriff trägt häufig mehrere Komponenten; man denke beispielweise an *Mädchenhandelsschule*, der sich aus den Begriffen *Mädchen* und *Handelsschule* sowie letzterer wiederum aus *Handel* und *Schule* zusammensetzt. In einer abweichenden Lesart könnte man auch an die Komponenten *Mädchenhandel* und *Schule* denken (das Beispiel verdanke ich Norbert Henrichs). Zwei Begriffe A und B sind „zusammengesetzt“ (im Deutschen häufig als ein Wort *AB*, aber auch – wie im Englischen oftmals – *B des A*), wenn das Kompositum die Intension des zweiten Bestandteils (B) (mit allen Merkmalen b_1, \dots, b_n) und zusätzlich das Merkmal „steht in Relation zu A“ erhält. *Philosophiegeschichte* (alternativ *Geschichte der Philosophie*) ist eine *Geschichte* (mit allen Eigenschaften, die *Geschichte* zukommen), zusätzlich hat sie das Merkmal „bezieht sich auf Philosophie“. Ein *Schulkind* ist ein *Kind* mit dem zusätzlichen Merkmal, zur Schule zu gehen. Bei mehrgliedrigen Komposita wie *ABC* wird es komplizierter, da unterschiedliche Subeinheiten (*AB* und *C* bzw. *A* und *BC*) gebildet werden können. *Mädchenhandelsschule* (in der Zerlegung *AB* und *C*) ist eine *Schule* (mit allen Merkmalen einer *Schule*) und der Eigenschaft, in Mädchenhandel auszubilden; *Mädchenhandelsschule* (nun in der Zerlegung *A* und *BC*) ist eine *Handelsschule* (mit allen Merkmalen einer *Handelsschule*) und dem zusätzlichen Merkmal, nur Mädchen aufzunehmen.

Man unterscheidet Wissensordnungen nach dem jeweils vorgefundenen Koordinationsgrad zusammengesetzter Be-

griffe. Die Abstufung der Koordination ist zwar fließend, es haben sich aber drei prototypische Koordinationsgrade herauskristallisiert.

Tauchen stets zusammengesetzte Begriffe (*Mädchenhandelsschule*, aber auch *Witterungsbedingte Einflüsse auf Flora und Fauna in Hochgebirgslagen*) als Einheit in einer Begriffsordnung auf, so spricht man von „Präkombination“ (Lancaster, 2003, 50 ff.). Die Begriffe sind bereits in der Wissensordnung als Einheit verschweißt und können auch nur in dieser Form benutzt werden, und zwar sowohl beim Indexieren als auch beim Recherchieren.

Belässt man in der Wissensordnung die Begriffe so einfach wie möglich, so liegt eine „Postkoordination“ vor. Man kann nur dann die Komponentenzerlegung durchführen, wenn sichergestellt ist, dass die Kombination der Einzelbegriffe stets den korrekten zusammengesetzten Begriff ergibt. Eine Zerlegung von *Informationswissenschaft in Information und Wissenschaft* wäre demnach nicht möglich, da die anschließende Kombination mehrdeutig ist (es ergibt sich aus den beiden Einzelbegriffen nämlich auch *Wissenschaftsinformation*). Klar dürfte sein, dass semantische Fehlläufer (wie z.B. *Schlüsselbein*) nicht zerlegt werden dürfen. Postkoordination verlagert die Zusammensetzung auf die konkrete Recherche (Lancaster, 2003, 38).

An information retrieval system that allows the searcher to combine terms in any way is ... referred to as post-coordinate.

Wenn die Begriffsordnung postkoordiniert aufgebaut ist, aber der Indexer – nicht der Nutzer – die Kombinationen beim Indexierungsprozess festlegt, ist „Präkoordination“ gegeben. Diese Koordinationsform führt zu einem syntaktischen Indexieren. Nehmen wir an, unsere Begriffsordnung enthalte die beiden Begriffe *Handelsschule* sowie *Mädchen*, und das Kombinationszeichen sei +, dann ergibt sich:

Handelsschule + Mädchen.

Dieser (nunmehr zusammengesetzte) Begriff kann sowohl als Einheit (natürlich auch – soweit sinnvoll – in umgekehrter Reihenfolge Mädchen + Handelsschule) als auch getrennt durch beide Komponenten recherchiert werden.

Präkombination erlaubt eine hohe begriffliche Spezifizierung der Wissensordnung und damit eine ballastarme Suche; Postkoordination gibt dem Nutzer große Freiheiten bei der Formulierung seiner Suchargumente. Präkoordination versucht, zwischen beiden Extremen zu vermitteln.

Äquivalenz

Zwei Benennungen sind synonym, wenn sie denselben Begriff bezeichnen. Totale Synonyme, die sich auf alle Bedeutungsvarianten und alle (deskriptiven, sozialen und expressiven) Bedeutungen erstrecken, sind rar (Löbner, 2003, 117); Beispiel ist *Samstag – Sonnabend*. Auch Abkürzungen (*LKW – Lastkraftwagen*), unterschiedliche Schreibweisen (*Fotografie – Photographie*), invertierte Wortreihenfolgen (*verarbeitendes Gewerbe – Gewerbe, verarbeitendes*) und Kurzformen (*U-Bahn – Untergrundbahn*) sind total synonym. Eng verwandt mit der totalen Synonymie sind gebräuchliche fremdsprachliche Bezeichnungen (*Computer – Rechner*) sowie ein abweichender Sprachgebrauch (*Massenkommunikationsmittel – Massenmedien*).

Nach Löbner (2003, 117) sind die meisten Synonymbeziehungen partieller Natur: Sie benennen nicht exakt denselben Begriff, sondern stehen für (mehr oder minder) eng verwandte Begriffe. Unterschiede können sowohl in der Extension als auch in der Intension vorkommen. Löbners Beispiel *geflügelte Jahresendpuppe* (offizielles DDR-deutsch) ist zwar extensional identisch mit einem *Weihnachtsengel*, aber nicht intensional. Im Gegensatz zur echten Synonymie, die eine Beziehung zwischen Benennungen und einem Begriff ist, liegt bei der partiellen Synonymie eine Relation zwischen Begriffen vor.

In den meisten Wissensordnungen in der Informationspraxis werden totale und partielle Synonyme sowie darüber hinaus – je nach Einsatzzweck – ähnliche Begriffe (als „Quasi-Synonyme“) als ein einziger Begriff behandelt. In einer Wissensordnung für Wirtschaftswissenschaften ist es wahrscheinlich belanglos, zwischen *Retrieval* und *Recherche* zu unterscheiden, so dass beide Benennungen hier als quasi-synonym angesehen werden. In einer informationswissenschaft-

lichen Wissensordnung wäre es dagegen völlig verfehlt, diese beiden Begriffe zu verschmelzen, da sie – aus der Sicht der Informationswissenschaft – völlig unterschiedliche Gegenstände ausdrücken. Hat man in einer Begriffsordnung zwei Begriffe als synonym verbunden, so sind diese (bis zu einer möglichen Überarbeitung der Ordnung) stets eine Einheit und können nicht getrennt betrachtet werden. Wendet man die Begriffsordnung auf Volltextretrievalsysteme an, so wird die Suchanfrage durch alle festgeschriebenen Synonyme des ursprünglichen Suchbegriffs erweitert. Synonymie ist reflexiv, symmetrisch und transitiv.

Gewisse Objekte sind „gen-identisch“ (Menne, 1980, 68 f.). Dies ist eine schwache Form von Identität, bei der von gewissen zeitlichen Bezügen abgesehen wird. Ein Mensch in seinen verschiedenen Lebensaltern (Person X als Kind, Erwachsener und Greis) ist demnach gen-identisch. Eine mögliche Option in Begriffsordnungen ist es, Begriffe gen-identischer Objekte als Quasi-Synonyme zusammenzufassen. Es existiert aber auch die Möglichkeit, die jeweiligen Begriffe einzeln zu betrachten und im Anschluss daran zu verknüpfen.

Werden gen-identische Objekte zu unterschiedlichen Zeiten durch verschiedene Begriffe beschrieben, so stellt man diese Begriffe durch chronologische Relationen in den gewünschten Zusammenhang. Die Beziehungen lauten „chronologisch früher“ sowie – als Inversion – „chronologisch später“. Als Beispiel denken wir an die Stadt an der Mündung der Newa in die Ostsee:

*zwischen 1703 und 1914: Sankt Petersburg
1914 bis 1924: Petrograd
1924 bis 1981: Leningrad
danach: wieder Sankt Petersburg.*

Jeweils benachbarte Begriffe werden chronologisch verknüpft:

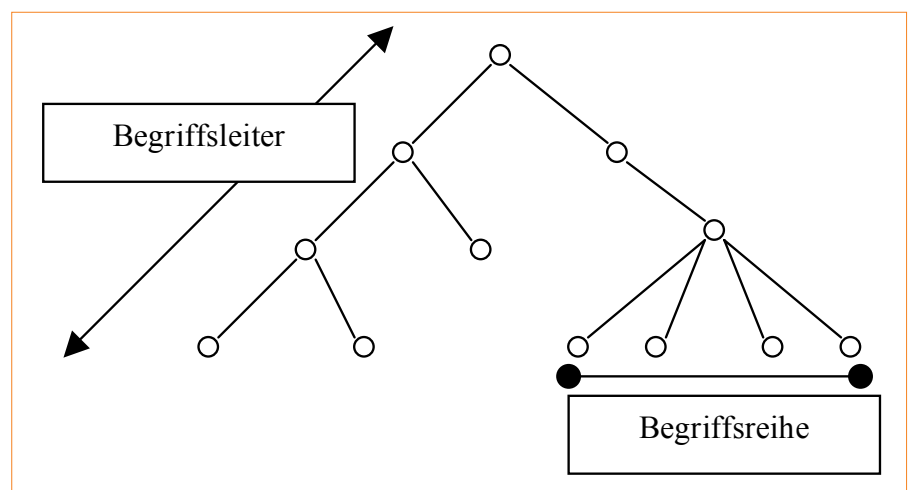


Abbildung 4: Begriffsleiter und Begriffsreihe.

Quelle: in Anlehnung an DIN 2331:1980, 12.

Sankt Petersburg [Zarenzeit] ist chronologisch früher als Petrograd.
Petrograd ist chronologisch früher als Leningrad.

usw. Die chronologische Relation ist irreflexiv, asymmetrisch und transitiv. Zwei Begriffe sind Antonyme, wenn sie sich gegenseitig ausschließen. Solche Gegenbegriffe sind beispielsweise *Liebe – Hass, Genie – Wahnsinn* und *tot – lebendig*. Wir müssen zwei Varianten unterscheiden: Kontradiktorische Antonyme kennen genau zwei Ausprägungen, aber nichts Drittes, möglicherweise dazwischen Liegendes. Jemand ist schwanger oder sie ist nicht schwanger – tertium non datur. Bei konträren Antonymen existieren neben den Extremwerten weitere Werte; zwischen Liebe und Hass liegt beispielsweise die Gleichgültigkeit. Bei kontradiktorischen Antonymen ist es im Retrieval möglich, den jeweiligen Gegenbegriff – verknüpft mit einem negierenden Term wie „nicht“ oder „un-“ – in eine Suche mit einzubeziehen. Ob konträre Antonyme sinnvoll in Wissensrepräsentation und Information Retrieval eingesetzt werden können, ist derzeit offen. Antonymie ist irreflexiv, symmetrisch und intransitiv.

Hierarchie

Die wichtigste Relation von Begriffsordnungen, quasi das tragende Gerüst, ist die Hierarchie. Emile Durkheim (2007[1912]) geht davon aus, dass die Hierarchie eine grundlegende Relation ist, die alle Menschen zur Ordnung der Dinge benutzen. Da menschliche Gesellschaften stets hierarchisch geordnet sind, wird – so Durkheim – die Hierarchie im gesellschaftlichen Alltag erlebt und von dort aus auf unsere Begriffe „der Welt“ projiziert. Die deutsche Norm DIN 2331 (1980, 2) definiert:

Hierarchische Beziehungen stellen ein Über- und Unterordnungsverhältnis und damit auch Nebenordnungsverhältnisse zwischen Begriffen her. Hinsichtlich des Begriffsumfangs bedeutet dies, daß alle Gegenstände, die unter einen engeren Begriff (untergeordneten Begriff) fallen, auch unter den weiteren Begriff (übergeordneten Begriff) fallen, dieser aber noch zusätzliche Gegenstände umfaßt.

Begriffe in hierarchischen Relationen bilden Begriffsleiter. „Oberbegriff“ ist derjenige Begriff, der in der Begriffsleiter in genau der nächsten Hierarchieebene über einem Ausgangsbegriff steht; „Unterbegriff“ ist ein Begriff, der in genau der nächst niedrigeren Hierarchieebene verortet ist. Ein „Geschwisterbegriff“ (Nebenordnung ersten Grades) teilt

sich mit dem Begriff denselben Oberbegriff. Geschwister bilden eine Begriffsreihe (Abbildung 4). Der oberste Begriff in einer Begriffsleiter ist der „Topterm“, die untersten Begriffe heißen „Bottomterms“. Die Hierarchie ist irreflexiv, asymmetrisch und transitiv, wobei „ist Oberbegriff von“ und „ist Unterbegriff von“ zueinander invers sind. Eine Begriffsordnung ist monohierarchisch, wenn es zu jedem Begriff außer dem Topterm genau einen Oberbegriff gibt (wie im Beispiel von Abbildung 4); sie ist polyhierarchisch, wenn einige Begriffe über mehrere Oberbegriffe verfügen. Als Beispiel wählen wir den Begriff *Wirtschaftsinformatik*. Er ist genauso Unterbegriff von *Betriebswirtschaftslehre* wie von *Informatik*.

Wenn wir in einer Wissensordnung die Hierarchierelation nicht weiter verfeinern wollen, liegt ein „gemischt-hierarchisches Begriffssystem“ (DIN 2331:1980, 6) vor. Es wird als „gemischt“ bezeichnet, da es mehrere Arten von Hierarchierelationen zusammenfasst. Wir unterscheiden drei Spielarten der Hierarchie: Hyponymie, Meronymie und Instanz.

Hyponym-Hyperonym-Relation

Die Abstraktionsrelation ist eine hierarchische Relation, deren Unterteilung nach logischen Gesichtspunkten erfolgt. Der jeweils nächst niedrigere Begriff entsteht im Rahmen der Definition nach Begriffserklärung bzw. – soweit zutreffend – nach Familienähnlichkeit. Wird nicht über Familienähnlichkeit definiert, so erbt der Unterbegriff, das „Hyponym“, alle Merkmale des Oberbegriffs, des „Hyperonyms“. Im Falle der Familienähnlichkeit erbt ein Hyponym nur eine Teilmenge der gesamten Merkmale des Hyperonyms. Zusätzlich verfügt es über mindestens ein weiteres wesenskonstitutives Merkmal, das es von den Schwesterbegriffen abtrennt. Für alle Elemente der Extension des Hyponyms gilt, dass sie grundsätzlich auch Elemente des Hyperonyms sind. Die logische Unterordnung der Abstraktionsrelation führt stets zu einer Implikation folgender Art (Löbner, 2003, 104; Storey, 1993, 460 ff.)

wenn x ein A ist, dann ist x ein B,
genau dann, wenn A Hyponym von B ist.

Wenn gilt, dass *Blaumeise* ein Hyponym von *Meise* ist, dann gilt auch die Implikation

wenn gilt: x ist eine *Blaumeise*, dann gilt:
x ist eine *Meise*.

Die Abstraktionsrelation lässt sich stets als eine „ist ein“-Beziehung (im Engli-

schon häufig „IS-A“ genannt) ausdrücken (Khoo & Na, 2006, 174). Im Beispiel

- Vogel
- Singvogel
- Meise
- Blaumeise

(definiert jeweils ohne Rückgriff auf Familienähnlichkeit) gilt:

Die *Blaumeise* IST EINE *Meise*.
Die *Meise* IST EIN *Singvogel*.
Der *Singvogel* IST EIN *Vogel*.

Von oben nach unten kommen in einer Begriffsleiter jeweils weitere Merkmale bei der Intension hinzu: Ein *Singvogel* ist ein *Vogel*, der singt. Die *Blaumeise* ist eine *Meise*, deren Gefieder blau gefärbt ist. Wohlgemerkt: Die Merkmale müssen jeweils im Begriffssatz (Schlagwortsatz, Deskriptorsatz usw.) über spezifische Relationen notiert werden. Ansonsten wäre eine (automatisch durchzuführende) Vererbung gar nicht möglich. Definieren wir mithilfe der Familienähnlichkeit, so liegt eine leicht veränderte Lage vor. Im Beispiel

- Freizeitbeschäftigung
- Spiel
- Glücksspiel

gilt wie oben

Das *Glücksspiel* IST EIN *Spiel*.
Das *Spiel* IST EINE *Freizeitbeschäftigung*.

Da wir *Spiel* über Familienähnlichkeit abgegrenzt haben, erbt *Glücksspiel* nicht alle Eigenschaften von *Spiel* (etwa nicht grundsätzlich *Brettspiel*, *Kartenspiel*, *Kampfspiel*), sondern nur einige. Das zusätzliche Merkmal des Unterbegriffs (*ist ein Glücksspiel*) ist in diesem Falle bereits als Teil der mit ODER verbundenen Begriffe des Oberbegriffs vorhanden. Die Klärung erfolgt durch den Ausschluss der anderen mit ODER verknüpften Familienmitglieder (etwa in der Art: *ist genau ein Spiel, das Glück erfordert*). Man ist versucht anzunehmen, dass zwischen Extension und Intension von Begriffen innerhalb einer Begriffsleiter ein reziprokes Verhältnis besteht: Eine Vermehrung des Begriffsinhalts (also Hinzufügen weiterer Merkmale auf dem Weg von oben nach unten) geht mit einer Verminderung der Menge der Objekte, die jeweils unter den Begriff fallen, einher. Es gibt sicherlich mehr Vögel als es Singvögel gibt. Solche eine Reziprozitätsbeziehung ist in vielen Fällen anzutreffen, sie gilt jedoch nicht allgemein. Sie gilt nie bei den Individualbegriffen, denn hier können wir durchaus weitere Eigenschaften hinzufügen, ohne dass sich an der Extension etwas ändern würde. *Karl May*

ist beispielsweise durch *Autor, in Sachen geboren, hat Winnetou erfunden* bereits klar intensional bestimmt; das Hinzufügen von *hat mit dem Verlag Münchenmeyer Geschäftsbeziehungen* ändert an der Extension nicht das Geringste. Es lassen sich sogar für Allgemeinbegriffe Gegenbeispiele aufzeigen, also Begriffe, die bei Vermehrung des Inhalts auch eine Vergrößerung des Umfangs zeigen. Das klassische Beispiel stammt von Bolzano (1837; s. a. Hoensbroech, 1931). Walter Dubislav (1981, 121) referiert diesen Fall:

Man bilde etwa mit Bolzano den Begriff eines „Kenners aller europäischer Sprachen“ und vermehre jetzt den Inhalt dieses Begriffs durch Hinzufügen des Bestandteiles „lebend“ zu dem Begriff „Kenner aller lebender europäischer Sprachen“. Man bemerkt, daß man den Inhalt des ersten Begriffes zwar erweitert hat, daß aber der Umfang des derart aus dem ersten entstehenden neuen Begriffes den Umfang des alten als Teilklasse enthält.

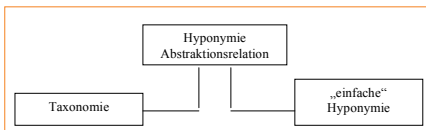


Abbildung 5: Spezifische Hyponym-Hyperonym-Relationen.

Wir können bei der Abstraktionsrelation zwei Spielarten entdecken: die Taxonomie und die nicht-taxonomische, „einfache“ Hyponymie. Bei einer Taxonomie kann man die „ist ein“-Relation verstärken zu „ist eine Art von“ (Cruse, 2002, 12):

The first (variety of hyponymy; Verf.) is the relation which is exemplified in An X is a Y (which corresponds to ‚simple‘ hyponymy); the second is the relation for which An X is a kind/type of Y is diagnostic, which is more discriminating than hyponymy, and which functions as the ‚vertical‘ relation in a taxonomy. ... (T)he second relation is called taxonomy.

Eine Taxonomie zerlegt nicht einfach eine größere Klasse in kleinere Klassen, wie dies bei der einfachen Hyponymie geschieht. Betrachten wir zwei Beispiele:

? Eine Königin IST EINE ART VON Frau.
(besser: Eine Königin IST EINE Frau.)

? Ein Hengst IST EINE ART VON Pferd.
(besser: Ein Hengst IST EIN Pferd.)

In beiden Fällen ist die Lesart IST EINE ART VON nicht zielführend; hier herrscht einfache Hyponymie vor. Schauen wir dagegen auf diese Beispiele:

Ein Kaltblüter IST EINE ART VON Pferd.

Ein Stetson IST EINE ART VON Hut.

Hier macht die Formulierung Sinn, denn in diesen Fällen liegt in der Tat ein taxonomischer Zusammenhang vor. Eine Taxonomie erfüllt gemäß Cruse (2002, 13) gewisse Bedingungen:

Taxonomy exists to articulate a domain in the most effective way. This requires „good“ categories, which are (a) internally cohesive, (b) externally distinctive, and (c) maximally informative.

In Taxonomien werden der Unterbegriff, das „Taxonym“, und das Hyperonym grundsätzlich aus der gleichen Perspektive betrachtet. *Hengst* ist kein Taxonym von *Pferd*, da *Hengst* aus einer geschlechtlichen Perspektive betrachtet wird und *Pferd* nicht. Bei *Kaltblüter* und *Pferd* sind die Perspektiven dagegen identisch; beide werden unter biologischen Aspekten untersucht.

Je nach der Anzahl der Unterscheidungsansichtspunkte spricht man von Mono- bzw. Polydimensionalität (DIN 2331:1980, 3). Wird genau *eine* Dimension der klassenbildenden Merkmale benutzt, so liegt Monodimensionalität vor. Dimensionalität wird auf jeder hierarchischen Ebene separat abgearbeitet. Es kann demnach durchaus sein, dass auf einer Ebene anhand von Dimension X und auf einer anderen anhand von Y unterschieden wird.

Die Hyponym-Hyperonym-Relation ist irreflexiv, asymmetrisch und transitiv. Bei der Transitivität lassen sich vereinzelt Fälle aufführen, die die allgemeine Gültigkeit in Frage stellen. Betrachten wir ein Beispiel (Cruse, 2004, 152; Khoo & Na, 2006, 175):

A car seat IS A seat.
A seat IS A furniture.
? A car seat IS A furniture.

Ein Autositz ist sicherlich im gebräuchlichen Sinne kein Möbelstück, insofern dürfte hier die Transitivität nicht gelten. Im Deutschen klingt das Argument allerdings anders:

Ein Autositz IST EIN Sitz.
Ein Sitz IST EIN Möbelstück.
? Ein Autositz IST EIN Möbelstück.

Hier zeigt sich, dass der Fehler in der zweiten Zeile liegt: Ein Sitz ist nicht unbedingt ein Möbelstück. Die vermeintliche Intransitivität im englischen Beispiel entsteht durch die Mehrdeutigkeit von *seat*, das ja *Sitz*, *Bank* oder *Sessel* bedeuten kann. Nach erfolgter Disambiguierung sollte die Transitivität erhalten bleiben.

Meronym-Holonym-Relation

Zeigt die Abstraktionsrelation eine logische Sicht auf Begriffe, so geht die Teil-Ganzes-Relation von einer gegenständlichen Perspektive aus (Khoo & Na, 2006, 176 ff.). Begriffe von Ganzheiten, „Holonyme“, werden in Begriffe von deren Teilen, „Meronyme“, untergliedert. Löbner (2003, 135) stellt fest:

A ist genau dann ein Meronym von B, und B ein/das Holonym von A, wenn ein potenzieller Referent von A durch die Bedeutung von A als konstitutiver Teil eines potentiellen Referenten von B konzipiert ist. „Konstitutive Teile“ sind dabei zu verstehen als wesentliche Teile, die das Ganze mit zu dem machen, was es ist.

Sind es bei der Abstraktionsrelation nicht irgendwelche Merkmale, die zur Definition herangezogen werden, sondern eben die „wesenskonstitutiven“ Charakteristika, so sind es in der Teil-Ganzes-Relation auch nicht irgendwelche Teile, sondern die „wesentlichen“ Teile der betreffenden Ganzheit. Die Meronym-Holonym-Relation trägt mehrere Benennungen. Neben „Teil-Ganzes-Relation“ wird auch von „partitiver Relation“ oder „Bezugsbeziehung“ (so in der Norm DIN 2331:1980, 3) gesprochen. Ein auf dieser Relation bestehendes System wird „Merologie“ genannt (Simons, 1987).

Es ist möglich, dass in Einzelfällen Meronymie und Hyponymie zusammenfallen. Betrachten wir das Begriffspaar:

Industrie – chemische Industrie.

Die chemische Industrie ist sowohl ein Teil der Industrie als auch eine besondere Art der Industrie.

Die Meronymie wird durch „ist Teil von“ (im Englischen „PART-OF“) ausgedrückt. Diese Relation steht nicht genau für eine Begriffsbeziehung, sondern setzt sich aus einem Bündel unterschiedlicher partitiver Beziehungen zusammen. Will man – etwa in vereinfachender Absicht – die unterschiedlichen Teil-Ganzes-Relationen zu einer einzigen zusammenfassen, so wird in vielen Fällen die Transitivität verletzt. Winston, Chaffin und Herrmann (1987, 442-444) haben eine Liste mit – fehlerhaften – Kombinationen zusammengestellt. Einige Beispiele mögen die Intransitivität belegen:

Simpson's finger is part of Simpson.
Simpson is part of the Philosophy Department.

? Simpson's finger is part of the Philosophy Department.

Water is part of the cooling system.
Water is partly hydrogen.

? Hydrogen is part of the cooling system.

Die mit dem Fragezeichen markierten Sätze sind falsche Schlüsse. Man kann (als „faule Lösung“) auf die Transitivität der jeweiligen spezifischen Meronymie-Relationen im Information Retrieval verzichten. Damit nimmt man sich die Option der Suchanfrageerweiterung über mehr als eine Hierarchieebene. Dafür braucht man sich nicht die Mühe zu machen, zwischen den einzelnen partitiven Beziehungen zu differenzieren. Die elaborierte Lösung unterscheidet die spezifischen Meronymie-Relationen und analysiert diese auf Transitivität. Damit ist die Möglichkeit auf Suchanfrageerweiterung jederzeit und über beliebig viele Ebenen gegeben.

Wir folgen dem inzwischen klassischen Ansatz von Morton E. Winston, Roger Chaffin und Douglas Herrmann (1987) und spezifizieren die Teil-Ganzes-Relation in sinnvolle Arten. Winston et al. unterscheiden sechs verschiedene Meronymie-Relationen, die wir durch weitere Unterteilung auf neun erweitern (Abbildung 6) (Weller & Stock, 2008).

Insofern Ganzheiten eine Struktur aufweisen, kann man diese Struktur in gewisse Teile zerlegen (Gerstl & Pribbenow, 1996; Pribbenow, 2002). Die fünf links eingezeichneten Teil-Ganzes-Beziehungen zeichnen sich dadurch aus, dass Ganzheiten strukturell zerlegt worden sind. Geographica lassen eine Unterteilung nach Verwaltungseinheiten zu, insofern wir eine gegebene geographische Einheit in dessen Untereinheiten gliedern. *Nordrhein-Westfalen* ist Teil von *Deutschland*; der Ortsteil *Kerpen-Sindorf* ist Teil von *Kerpen*. (Nicht-soziale) gleichförmige Kollektionen lassen sich in ihre Elemente einteilen. Ein *Wald* besteht aus *Bäumen*, ein *Schiff* ist Teil einer *Flotte*. Ein ähnlicher Einteilungsaspekt liegt vor, wenn wir (gleichförmige) Organisationen in ihre Einheiten zerlegen, etwa eine *Universität* in ihre *Fakultäten*. Johansson (2004) weist darauf hin, dass bei der Verletzung der Gleichförmigkeit Transitivität nicht unbedingt gilt. Wir nehmen an, es gäbe einen Verband Y , in dem andere Verbände X_1, \dots, X_n (und *nur* Verbände) Mitglied seien. Die Person A sei Mitglied in X_i . Bei Transitivität würde gelten, dass A über seine Mitgliedschaft in X_i auch Teil von Y sei. Laut den Statuten von Y ist dies aber gar nicht möglich. Einmal geht es um Mitgliedschaft von Personen, zum anderen von Verbänden, so dass das Prinzip der Gleichförmigkeit im Beispiel verletzt worden ist. Ein zusammenhängender Komplex, beispielsweise ein *Haus*, lässt sich in seine Komponenten, etwa das *Dach* oder den *Keller*, einteilen. Ähnlich ist (diesmal in zeitlicher Hinsicht) die Meronymie bei einem Ereignis (sagen wir: einer *Zirkusvorstellung*) und einem spezifischen Segment (z.B. dem *Trapezakt*) gebildet (Storey, 1993, 464).

Die zweite Gruppe von Meronymen arbeitet unabhängig von Strukturen (in Abbildung 6 auf der rechten Seite gezeichnet). Eine beliebige Ganzheit kann in Stücke zerlegt werden, so beispielsweise eine *Tasse* (nachdem wir sie auf den Boden geworfen haben) in *Scherben* oder – weniger destruktiv – ein *Brot* in mundgerechte *Schnitten*. Eine zusammenhängende Aktivität (z.B. das *Einkaufen*) kann in einzelne Phasen (etwa *Bezahlen*) eingeteilt werden. Eine der zentral wichtigen Meronymie-Relationen ist die Beziehung eines Objektes zu seinen Bestandteilen, etwa die *Aluminium*-Teile eines *Flugzeugs* oder die *Holzteile* meines *Schreibtisches*. Wenn wir eine in sich homogene Masse vorliegen haben, so kann man diese in Portionen zerlegen. Beispiele sind *Wein (im Fass)* und *1 ltr. Wein* oder *Meter – Dezimeter*.

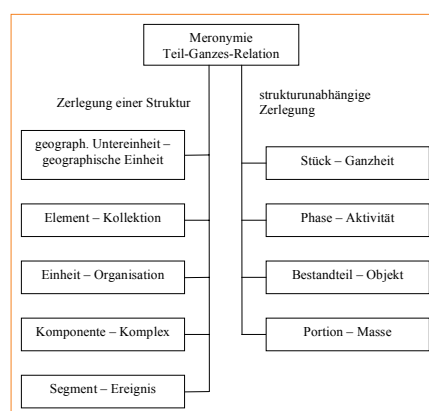


Abbildung 6: Spezifische Meronym-Holonym-Relationen.

Alle beschriebenen Meronymie-Holonymie-Relationen sind irreflexiv, asymmetrisch und transitiv, insofern sie „sortenrein“ definiert und angewandt worden sind.

Wir haben schon besprochen, dass innerhalb einer Begriffsleiter der Hyponym-Hyperonym-Relation die Begriffe (in den meisten Fällen) von oben nach unten ihre Eigenschaften vererben. Das gleiche gilt für ihre Meronyme. Wir können hier von Meronymvererbung in der Abstraktionsrelation sprechen (Weller & Stock, 2008, 168). Wenn Begriff A ein Teilbegriff (z.B. ein *Motor*) der Ganzheit B (einem *Auto*) ist, und C ein Hyponym von B ist (sagen wir: ein *Krankswagen*), so verfügt auch der Unterbegriff C über das Teil A (ein *Krankswagen* hat demnach einen *Motor*).

Instanz

Bei der extensionalen Begriffsbestimmung wird der in Frage stehende Begriff durch die Aufzählung derjenigen Elemente definiert, auf die er zutrifft. Im allgemeinen Fall wird dabei offen gelassen, ob die Elemente Allgemein- oder In-

dividualbegriffe sind. Bei der Instanz-Relation wird gefordert, dass das Element grundsätzlich ein Individualbegriff ist. Das Element ist also stets eine „named entity“ (Stock, 2007, 254).

Ob diese Element-Klasse-Relation im Rahmen von Hyponymie oder von Meronymie betrachtet wird, ist für die Instanz irrelevant. Die Instanz kann also sowohl durch „ist ein“ als auch durch „ist Teil von“ ausgedrückt werden. Im Sinne einer Abstraktionsrelation lässt sich sagen:

Persil IST EIN Waschmittel.

Köln IST EINE Universitätsstadt am Rhein.

Genauso ist es möglich, so zu formulieren: Silwa (das ist unser Auto) IST TEIL unseres Fuhrparks.

Angela Merkel IST TEIL der CDU.

Instanzen können durchaus ihrerseits Unterbegriffe haben. So ist im letzten Beispiel die *CDU* Instanz des Begriffs *deutsche Partei*. Und natürlich verfügt unsere *Silwa* über Teile, etwa das *Fahrgestell* oder den *Motor*. Die Instanz-Relation ist reflexiv, asymmetrisch und intransitiv.

Weitere spezifische Relationen

Es gibt eine Fülle weiterer semantischer Beziehungen in Begriffsordnungen, die wir in einer vorläufigen ersten Annäherung unter der Sammelbezeichnung „Assoziation“ zusammenfassen wollen. Die Assoziationsrelation als solche existiert demnach nicht, sondern es gibt diverse unterschiedliche Beziehungen. Ihnen ist gemeinsam, dass sie – negativ formuliert – keine (Quasi-)Synonyme und keine Hierarchie bilden und – positiv gesprochen – für Wissensordnungen sinnvoll zu gebrauchen sind.

In einem einfachen Fall, der jede Spezifizierung offen lässt, spielt die Assoziationsrelation die Rolle einer „siehe auch“-Verknüpfung. Die Begriffe werden nach praktischen Aspekten miteinander in Beziehung gebracht, so z.B. in einer wirtschaftswissenschaftlichen Begriffsordnung Produkte mit zugehörigen Branchen (beispielsweise *Körperpflegemittel* SIEHE AUCH *Körperpflegemittelindustrie* und umgekehrt). Die unspezifische „siehe auch“-Relation ist irreflexiv, symmetrisch und intransitiv.

Für andere – nunmehr spezifische – Assoziationsrelationen beginnen wir mit Beispielen. Schmitz-Esser (2000, 79 f.) schlägt für eine konkrete Wissensordnung (der Weltausstellung EXPO 2000) die Relationen der Nützlichkeit und der Schädlichkeit vor. Hier zeigt sich, dass solche Begriffsbeziehungen persuasive „Nebentöne“ aufweisen. Im Beispiel

Radio zum Aufziehen IST NÜTZLICH FÜR Kommunikation in entlegenen Gebieten

liegt keine implizite Wertung vor. Anders ist dies bei

Überfischung IST NÜTZLICH FÜR Fischmehlindustrie
Mohnanbau IST NÜTZLICH FÜR Drogenhandel.

Eine zufriedenstellende Lösung liegt möglicherweise darin, bei Nützlichkeit und Schädlichkeit auf grundlegende Werte einer gegebenen Gesellschaft aufzubauen („nützlich für wen?“) (Schmitz-Esser, 2000, 79) und so die unteren beiden Beispiele als unvereinbar mit entsprechenden Wertvorstellungen abzulehnen oder – aus der Sicht eines Drogenkartells – das untere Beispiel als adäquat beizubehalten.

Eine der Wissensdomänen, die besonders im Rahmen ontologischer Wissensrepräsentation bearbeitet worden ist, ist die Biologie. Eine fast beispielhafte Bedeutung erlangte dabei die „Gene Ontology“ (GO) (Ashburner et al., 2000). Aber – hierbei handelt es sich um gar keine Ontologie, sondern um einen Thesaurus (genauer: um drei Teilthesauri für biologische Prozesse, Molekülfunktionen und zelluläre Komponenten), benutzt doch diese Begriffsordnung nur die Relationen PART OF und IS A, also nur Meronymie und Hyponymie. Smith, Williams und Schulze-Kremer (2003, 609) stellen folgerichtig fest,

The Gene Ontology, in spite of its name, is not an ontology as the latter term is commonly used either by information scientists or by philosophers. It is, as the GO Consortium puts it, a ‚controlled vocabulary‘.

Die Gene Ontology stellt aber einen guten Ausgangspunkt dar, um auszuführen, welche Relationen – neben den hierarchischen – in der Biomedizin eigentlich nötig sind (Smith et al., 2005). Für uns ist der Ansatz von Barry Smith et al. ein Beispiel dafür, wie man die vormals unspezifische Assoziationsrelation in unterschiedliche konkrete Relationen spezifizieren kann. Auch in Ontologien stellen die hierarchischen Relationen ein tragendes Gerüst dar (Smith et al., 2005):

Is_a and part_of have established themselves as foundational to current ontologies. They have a central role in almost all domain ontologies ...

Da Ontologien auf die Nutzung von Relationen in der terminologischen Logik abzielen, können wir mit einer unspezifisch gehaltenen Assoziationsrelation in diesem Kontext nichts anfangen. Es gilt,

Relation	Beispiel
C located_in C _i	66s pre-ribosome located_in nucleolus chlorophyll located_in thylakoid
C contained_in C _i	cytosol contained_in cell compartment space synaptic vesicle contained_in neuron
C adjacent_to C _i	intron adjacent_to exon cell wall adjacent_to cytoplasm
C transformation_of C _i	fetus transformation_of embryo mature mRNA transformation_of pre-mRNA
C derives_from C _i	plasma cells derives_from lymphocyte mammal derives_from gamete
P preceded_by P _i	translation preceded_by transcription digestion preceded_by ingestion
P has_participant P _i	photosynthesis has_participant chlorophyll cell division has_participant chromosome
P has_agent C	transcription has_agent RNA polymerase translation has_agent ribosome.

Quelle: Smith et al. (2005).

Tabelle 1: Reflexivität, Symmetrie und Transitivität paradigmatischer Relationen.

	Reflexivität	Symmetrie	Transitivität
<i>Äquivalenz</i>			
- Synonymie	R	S	T
- Gen-Identität	-R	-S	T
- Antonymie	-R	S	-T
<i>Hierarchie</i>			
- Hyponymie			
- - einfache Hyponymie	-R	-S	T
- - Taxonomie	-R	-S	T
- Meronymie (unspezifisch)	-R	-S	?
- spezifische Meronymien	-R	-S	T
- Instanz	R	-S	-T
<i>Spezifische Relationen</i>			
- „siehe auch“	-R	S	-T
- weitere Relationen	je nach Relation		

möglichst exakt bestimmte Attribute – in der Folge natürlich auch genaue Werte – zu entwickeln, die charakteristisch für die jeweilige Wissensdomäne sind. Smith et al. unterscheiden im Bereich der Genetik zwischen Komponenten C („continuant“ als Generalisierung der „zellulären Komponenten“ der ursprünglichen GO) und Prozessen P („processes“ als Verallgemeinerung der „biologischen Prozesse“). Acht Relationen sind für Smith et al. (2005) neben den hierarchischen Relationen wesentlich (siehe Kasten).

Bei der Relation *derives_from* erkennen wir hier im terminologischen Feld der Genetik die (ansonsten bei uns generell definierte) chronologische Relation der Gen-Identität wieder. Smith et al. (2005) unterscheiden drei einfache Arten der Derivation.

(F)irst, the succession of one single continuant by another single continuant across a temporal threshold (for example, this blastocyst derives from

this zygote); second, the fusion of two or more continuants into one continuant (for example, the zygote derives from this sperm and from this ovum); and third, the fission of an earlier single continuant to create a plurality of later continuants (for example, these promyelocytes derive from this myeloblast). In all cases we have two continuants c and c_i, which are such that c begins to exist at the same instant of time at which c_i ceases to exist, and at least a significant portion of the matter of c_i is inherited by its successor c.

Ob es bei der Spezifizierung der Assoziationsrelation zu einer Menge semantischer Relationen führt, die generalisierbar (d.h. in allen oder doch zumindest in den meisten Wissensordnungen einsetzbar) sind, ist derzeit ein offenes Forschungsproblem.

In Tabelle 1 haben wir diejenigen paradigmatischen Relationen, die für den Einsatz in Wissensordnungen geeignet sind, zusammenfassend aufgelistet. Zusätzlich

sind die jeweiligen Ordnungsmerkmale (Reflexivität, Symmetrie und Transitivität) notiert worden.

Relationen und Wissensordnungen

Wir definieren Wissensordnungen über ihre Mächtigkeit, Begriffe und Relationen auszudrücken. Die drei in der Informationswissenschaft und -praxis „klassischen“ Methoden – Nomenklatur, Klassifikation, Thesaurus – werden durch Folksonomies und Ontologien ergänzt (Stock & Stock, 2008). Folksonomies stellen einen Grenzfall von Wissensordnungen dar, da sie über keine einzige paradigmatische Relation verfügen. Nomenklaturen (Schlagwortsysteme) zeichnen sich vor allem dadurch aus, dass sie die Äquivalenzrelation benutzen und alle Arten hierarchischer Beziehungen nicht einsetzen. Bei Klassifikationssystemen tritt die (unspezifisch ausgelegte) Hierarchierelation hinzu. Thesauri arbeiten ebenfalls mit der Hierarchie; einige nutzen dabei die unspezifische Hierarchierelation, andere unterscheiden nach Hyponymie und (der unspezifischen) Meronymie (mit dem Problem – siehe Tabelle 1 – die Transitivität nicht garantieren zu können). Bei Thesauri kommt zusätzlich als weitere Relation eine in aller Regel unspezifisch verstandene Assoziationsrelation („siehe auch“) zwingend vor. Ontologien machen von allen besprochenen paradigmatischen Relationen Gebrauch. Sie werden in formalen Sprachen modelliert, dabei schenkt man auch terminologischen Logiken gebührende Beachtung. Im Vergleich zu anderen Wissensordnungen enthalten Ontologien grundsätzlich Instanzen. Die meisten Ontologien arbeiten mit (genau definierten) weiteren Relationen. Dadurch, dass in den Ontologien das Wissen direkt (und nicht etwa das Dokument, das das Wissen enthält) abgebildet wird, entfallen hier die syntagmatischen Relationen. Schaut man sich Tabelle 2 an, so sind die Wissensordnungen von links nach rechts nach ihrer Ausdrucksstärke gereiht. Jede Wissensordnung lässt sich durch Relationen der rechts neben ihr stehenden „anreichern“ und auf eine höhere Stufe heben: Aus einer Nomenklatur wird beispielsweise eine Klassifikation, wenn (neben dem Schritt vom Schlagwort zur Notation) alle Begriffe in eine hierarchische Relation gebracht werden; aus einem Thesaurus wird eine Ontologie, wenn bei den Hierarchierelationen genau differenziert und wenn weitere spezifische Relationen eingeführt werden. Eine Ontologie wird – und jetzt machen wir einen Schritt nach links – zu einer Methode der Inhaltserschließung, wenn sie die syntagmatische Relation einführt, d.h. – bei Beibehaltung aller Relationen – gestattet, ihre Konzepte Dokumenten zuzuordnen. So fließen die Vor-

Tabelle 2: Wissensordnungen und darin verwendete Relationen.

	Folksonomy	Nomenklatur	Klassifikation	Thesaurus	Ontologie
Begriff	Tag	Schlagwort	Notation	Deskriptor	Konzept
Äquivalenz	-	ja	ja	ja	ja
- Synonymie	-	ja	ja	ja	ja
- Gen-Identität	-	ja	-	-	ja
- Antonymie	-	-	-	-	ja
Hierarchie	-	-	ja	ja	ja
- Hyponymie	-	-	-	ja	ja
- - einfache Hyponymie	-	-	-	-	ja
- - Taxonomie	-	-	-	-	ja
- Meronymie (unspezifisch)	-	-	-	ja	-
- spezifische Meronymien	-	-	-	-	ja
- Instanz	-	-	-	ggf.	ja
Spezifische Relationen	-	-	-	ja	ja
- „siehe auch“	-	ggf.	ggf.	ja	ja
- weitere Relationen	-	-	-	-	ja
syntagmatische Relation	ja	ja	ja	ja	nein

teile der Ontologie mit ihrem mächtigen Relationsgefüge mit den Vorteilen der Dokumentindexierung zusammen und ergänzen sich gegenseitig.

3 Anwendungen

Wissensordnungen fundieren im Information Retrieval eine Reihe von Anwendungen. Im intellektuellen Retrieval geben sie dem Nutzer Hilfsmittel zum Suchen und zum Browsen zur Hand. Zudem lässt sich eine Wissensordnung beim automatischen Indexieren mit kontrolliertem Vokabular (beispielsweise bei Convera; Bayer et al. 2005) einsetzen. Diese Anwendungsfälle des begriffsorientierten Information Indexing und Information Retrieval sind bekannt und in der Praxis erprobt. Indexer verfügen über vorgegebene Begriffe und deren semantische Relationen sowie dazu maßgeschneiderte Indexierungsregelwerke. Nutzer müssen während des Recherchevorgangs nicht mehr raten, welche Worte wohl in den Dokumenten vorkommen könnten, sondern greifen auf terminologisch kontrolliertes Vokabular zurück. Wissensordnungen, deren Begriffe als Frames modelliert sind und deren Relationsgefüge „optimiert“ worden ist, gestatten weitere, neue Anwendungen, die alle in Richtung eines Semantic Web weisen. Wir werden im Folgenden drei Anwendungsfälle diskutieren:

- automatische Suchanfrageerweiterung im Information Retrieval,
- automatisches Schlussfolgern in der Wissensrepräsentation des semantischen Web,
- automatische Berechnungen zur Informationsverdichtung im Information Retrieval.

Automatische Suchanfrageerweiterung

Begriffe in einer Wissensordnung hängen über die besprochenen Relationen zusammen. Eine solche Begriffsordnung kann man als Graphen auffassen, in dem die Begriffe die Knoten und die jeweiligen Relationen die Pfade darstellen. Außer der übersichtlichen Darstellung dienen die Graphen der Ermittlung der semantischen Nähe zwischen den Begriffen. Hierbei unterscheiden wir zwei Fälle, das dichotome und das gewichtete Modell.

Im dichotomen Ansatz existiert zwischen zwei Begriffen eine bestimmte Relation oder sie existiert nicht. Zur Berechnung des Abstandes zwischen zwei Begriffen in einer Begriffsordnung zählen wir die Anzahl der Pfade, die man durchlaufen muss, um im Graphen auf dem kürzesten Weg von einem Begriff B_1 zu einem anderen Begriff B_2 zu gelangen. Unsere Beispielbegriffsordnung in Abbildung 7 verfüge über die Hierarchierelation und eine (unspezifische) Assoziationsrelation. Im dichotomen Fall zählt jeder Pfad „1“. In unserem Beispiel beträgt die kürzeste Entfernung von B_1 zu B_5 also vier.

Möchte man mit den jeweiligen Relationen eine semantische Bindungsstärke zwischen den Begriffen ausdrücken, so muss man die Relationen unterschiedlich gewichten; etwa folgendermaßen:

Unterbegriff:	0,8
Oberbegriff:	0,5
„siehe auch“:	0,4.

Zwischen B_1 und B_2 liegt nunmehr keine Bindung von 100 Prozent mehr, sondern nur noch von 40 Prozent, B_2 und B_3 sind zu 80 Prozent miteinander verbunden, B_3

und B_4 zu 40 Prozent und letztlich B_4 und B_5 zu 50 Prozent. Der semantische Abstand zwischen B_1 und B_5 beträgt nun:

$$1/0,4 + 1/0,8 + 1/0,4 + 1/0,5 = 8,25.$$

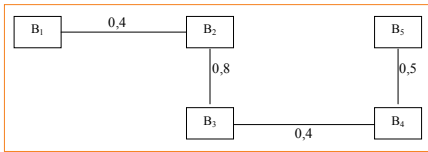


Abbildung 7: Begriffsordnung mit gewichteten Relationen.

Solche Abstandsmaße sind wichtig für die automatische Suchfrageerweiterung. Man legt einen Grenzwert fest, bis zu dem alle verknüpften Begriffe einem Ausgangsbegriff zugeordnet werden. Dabei muss beachtet werden, dass eine Anfrageerweiterung bei nicht-transitiven Relationen stets nur um genau einen Pfad möglich ist. Letzteres ist ein starkes Argument dafür, soweit wie möglich mit transitiven Relationen zu arbeiten, um der doch starken Einschränkung der Erweiterungsoptionen um nur genau einen Schritt zu entgehen.

Automatisches Schlussfolgern

Das Herzstück jeder Wissensordnung vom Typ Ontologie ist deren Fähigkeit, automatisch Schlussfolgerungen auf der Basis der hinterlegten Terminologie auszuführen. Die terminologische Logik findet ihren Ausgangspunkt in den Begriffen sowie in deren Relationen, also im jeweiligen semantischen Netz. Begriffe sind Knoten; die Relationen bilden die Linien. Schlussfolgerungen in der Aussagenlogik sind Implikationen der Art

$$p \rightarrow q$$

(gelesen „wenn p, dann q“), wobei p und q beliebige Aussagen sind.

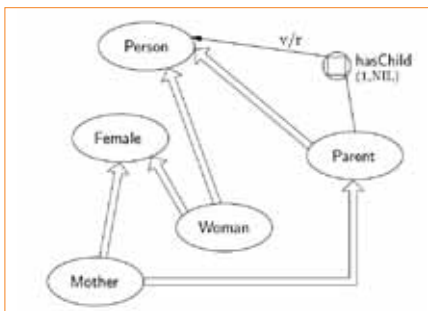


Abbildung 8: Beispiel eines semantischen Netzes. Quelle: Nardi/Brachman 2003, 5.

Terminologische Logiken bauen auf semantischen Netzen bzw. Frames auf. Das Grundgerüst wird von hierarchischen Relationen gebildet. In Abbildung 8 stehen die dicken Pfeile für die Hyponymie-Relation, *Mother* ist also sowohl Unterbe-

griff zu *Parent* als auch zu *Female*. Rechts oben sehen wir die spezifische Relation (in Ontologien auch „role“ genannt) *hasChild* mit einem Wert („value restriction“ *v/r*), im konkreten Fall *Person*, zusätzlich mit einer Anzahl (zwischen 1 und unendlich, „NIL“) (Nardi & Brachman, 2003, 5 f.). Wie in Hyponymien üblich (wenn wir hier vom Spezialproblem der Familienähnlichkeit absehen), vererben die Begriffe ihre Merkmale in der Begriffsleiter nach unten, so auch ihre spezifischen Relationen. Entsprechend gilt die *hasChild*-Relation nicht nur für *Parent*, sondern auch für dessen Unterbegriff *Mother*.

Relationen können untereinander in Relation stehen (Horrocks & Sattler, 1999). Wir haben oben die Meronymie eingeführt und als deren Spezifikation die strukturzerlegende Meronymie und darin beispielsweise die Komponente-Komplex-Relation gebildet. Zwischen den drei genannten Relationen besteht eine Hierarchierelation. Über solche Relationen der Relationen lassen sich Schlussfolgerungen ableiten. Wenn wir in unserer Begriffsordnung beispielsweise eingeführt haben:

Dach ist Komponente von *Haus*,

dann gelten auch

Dach ist struktureller Teil von *Haus* und
Dach ist Teil von *Haus*,

allgemein formuliert:

A ist Komponente von *B* \rightarrow *A* ist
struktureller Teil von *B* \rightarrow *A* ist Teil von *B*.

Die Werteinschränkungen der spezifischen Relationen werden in der Form

$$\forall R.C \text{ bzw. } \exists R.C$$

geschrieben. *R* ist die Relation (z.B. *hasChild*), *C* deren Wert (gemäß Abbildung 8 ist dies im Beispiel *Person*). Die Relation *hasChild* kann natürlich auch spezifischere Werte wie etwa *hasChild.Female* annehmen. Die obere Zeile mit dem Allquantor sagt aus, dass die Relation nebst dem Wert für alle Individuen gilt, während die untere Zeile mit dem Existenzquantor bedeutet, dass dies für einige Individuen zutrifft. Will man aussagen, dass es für einige Individuen zutrifft, dass sie eine Tochter haben, so formuliert man so:

$$\exists \text{ hasChild.Female,}$$

wohingegen

$$\forall \text{ hasChild.Female}$$

besagt, dass alle Individuen, über die berichtet wird, eine Tochter haben. Die

Werte können auch Zahlen ausdrücken. So steht

$$2 \text{ hasChild}$$

für *hat zwei Kinder*. Atomare Begriffe lassen sich durch Schnitt-, Vereinigungs- und Exklusionsmenge zu komplexen Begriffen verbinden. So kann man beispielsweise *Person, die nicht weiblich ist*, sowie *Mann oder Frau* so darstellen:

$$\text{Person} \sqcap \neg \text{Female} \\ \text{Male} \sqcup \text{Female}.$$

Wichtig für Schlussfolgerungen sind die Eigenschaften der jeweiligen Relation (Horrocks & Sattler, 1999). Transitiv Relationen gestatten Schlüsse auch über semantische Abstände von größer eins, gelten also nicht nur für direkte Begriffsnachbarn. Symmetrische Relationen zeichnen sich dadurch aus, dass die Relation, die zwischen *x* und *y* besteht, auch in der umgekehrten Richtung *y* und *x* gilt. Eine symmetrische Relation ist beispielsweise *hatNachbar*. Eine Relation heißt invers, wenn es zur Relation *x ρ y* eine Gegenrelation *x ρ ' y* gibt, wie es etwa bei Ober- und Unterbegriff der Fall ist. Hier können in beiden Richtungen Schlüsse gezogen werden:

Kernobst hat Unterbegriff *Apfel* \rightarrow *Apfel*
hat Oberbegriff *Kernobst*
Apfel hat Oberbegriff *Kernobst* \rightarrow *Kernobst*
hat Unterbegriff *Apfel*.

Letztlich ist eine Relation funktional, wenn sie jeweils zu genau einem Wert führt. Als Beispiel sei die Relation *hatGeburtsdatum* genannt. Zu funktionalen Relationen existieren inverse Relationen (etwa *istGeburtsdatumvon*). Funktionalbegriffe haben gemäß Löbner (2005) zwei Eigenschaften: Sie sind stets relational und „inhärent eindeutig“. Funktionalbegriffe haben (Löbner, 2005, 465)

für ein gegebenes Bezugsobjekt in einer gegebenen Situation genau einen Referenten ...: *Eine Person hat genau eine Augenfarbe, nicht mehrere oder gar keine; sie hat genau ein Geburtsdatum* ...

Eine konkrete Beschreibungslogik steht und fällt mit der jeweiligen Wissensbasis, die die Terminologie aufnimmt. Terminologien für Ontologien werden entweder – für Allgemeinbegriffe – in der TBox („terminology box“) oder – für Individualbegriffe – in der ABox („assertional box“) hinterlegt.

In der TBox werden Allgemeinbegriffe auf der Basis bereits eingeführter Begriffe definiert. Wenn wir annehmen, dass die Begriffsordnung bereits über *Person* und *Female* verfügt, so können wir *Woman* auf dieser Basis einführen:

Woman \equiv Person \sqcap Female.

Hierbei gelten zwei Regeln (Nardi & Brachman, 2003, 13):

only one definition for a concept name is allowed;
definitions are acyclic in the sense that definitions are neither defined in terms of themselves nor in terms of other concepts that indirectly refer to them.

Die TBox hat eine hierarchische Ordnung; letztlich ist sie ein Klassifikations-system (Nardi & Brachman, 2003, 14):

In particular, the basic task in construction a terminology is classification, which amounts to placing a new concept expression in the proper place in a ... hierarchy of concepts. Classification can be accomplished by verifying the subsumption relation between each defined concept in the hierarchy and the new concept expression.

Oberbegriffe sind in ihren Unterbegriffen enthalten. C sei ein beliebiger Begriff (sagen wir: *Blaumeise*) und D sein Oberbegriff (z.B. *Meise*). Dann gilt:

$C \sqsubseteq D$

(\sqsubseteq sei das Zeichen für Enthaltensein). Alle Merkmale (also alle Relationen), die für D gelten, gelten somit ebenso für C. (Alle Eigenschaften und Beziehungen, über die eine Meise verfügt, finden wir auch bei Blaumeisen.)

Die ABox nimmt Aussagen über Individualbegriffe auf (Nardi & Brachman, 2003, 15), und zwar sowohl über Eigenschaften des Individuums („concept assertions“) als auch über Relationen („role assertions“). Will man beispielsweise ausdrücken, dass Anna eine weibliche Person ist, so geht folgender Eintrag in die ABox ein:

Female \sqcap Person(ANNA).

Female und *Person* müssen selbstverständlich vorher in der TBox definiert worden sein. Wenn wir annehmen, dass Anna ein Kind namens Jacopo hat, dann lautet der Eintrag, der nunmehr eine Relation beinhaltet:

hasChild(ANNA,JACOPO).

Über die Aussagen aus TBox und ABox werden die Schlüsse gezogen (Donini, Lenzerini, Nardi, & Schaerf, 1996; May, 2006). Für diese Aussagen und Ableitungen ist das entsprechende Domänenwissen direkt in der Ontologie abfragbar; der „Umweg“ über eine intellektuelle Sichtung der Dokumente kann demnach entfallen. Ob es möglich ist, alle wichtigen Resultate (oder doch zumindest die

zentralen Sätze) in die Sprache der terminologischen Logik zu übersetzen, darf angesichts der Ausdrucksmöglichkeiten der Beschreibungslogik (sie ist nur ein – allerdings entscheidbarer – Teil der Prädikatenlogik) bezweifelt werden.

Mit dem automatischen Schlussfolgern auf der Basis von Ontologien sind wir im Kernbereich des Semantic Web (Berners-Lee, Hendler, & Lassila, 2001; Shadbolt, Hall, & Berners-Lee, 2006) angekommen. Als Anspruch des Semantic Web formulieren Shadbolt, Hall und Berners-Lee (2006, 96):

The Semantic Web is a Web of actionable information – information derived from data through a semantic theory for interpreting the symbols. The semantic theory provides an account of „meaning“ in which the logical connection of terms establishes interoperability between systems.

Die Diskussionen um das semantische Web sind zunächst stark technisch orientiert: Es geht um RDF (resource description framework), um URI (universal resource identifiers), die geeignete Ontologiesprache (wie etwa OWL, die Web ontology language) und die gerade skizzierten Regeln automatischen Schlussfolgerns. Offen blieben zunächst sowohl der begriffstheoretische Hintergrund als auch Methoden des Aufbaus geeigneter KOS. Als aktuelle Lösungsansätze diskutieren Shadbolt, Hall und Berners-Lee zwei Ansätze, die jeweils auf der Kooperation beteiligter Experten beruhen. Für gut strukturierte Wissensdomänen eignen sich Ontologien (wie hier beschrieben), die verteilt konstruiert und gepflegt werden (Shadbolt, Hall, & Berners-Lee, 2006, 99):

In some areas, the costs – no matter how large – will be easy to recoup. For example, an ontology will be a powerful and essential tool in well-structured areas such as scientific applications ... In fact, given the Web's fractal nature, those costs might decrease as an ontology's user base increase. If we assume that ontology building costs are spread across user communities, the number of ontology engineers required increases as the log of the community' size.

Dieser Ansatz ist nur anwendbar, wenn erstens die Wissensdomäne klein und überschaubar und zweitens die Mitglieder der jeweiligen Wissenschaftlergemeinschaft willens sind, am Ontologieaufbau und deren Pflege mitzuwirken. Über Disziplingrenzen hinaus scheint solch ein Ansatz gar nicht zu funktionieren. Der zweite Ansatz, das semantische Web zu fundieren, verläuft über Tagging und Folksonomies (Shadbolt, Hall, & Berners-Lee, 2006, 100):

Tagging on a Web scale is certainly an interesting development. It provides a potential source of metadata. The folksonomies that emerge are a variant on keyword searches. They're an interesting emergent attempt at information retrieval.

Folksonomies verfügen allerdings ausschließlich über syntagmatische Relationen. Möchte man solch einen Ansatz für das semantische Web (nunmehr als „social semantic Web“; Blumauer & Pellegrini, Hrsg., 2009; Breslin, Passant, & Decker, 2009) nutzbar machen, bedarf es wohl angestrenzter Arbeiten beim „Tag Gardening“ (Peters & Weller, 2008b). Mit beiden von Shadbolt, Hall und Berners-Lee vorgestellten Ansätzen sind der Vision eines universellen semantischen Web doch enge Grenzen gesteckt. (Die Ruinen einer ähnlichen Vision, das Weltwissen zusammenzufassen – Otlets und La Fontaines „Mundaneum“ – sind heute im Museum in Mons zu bewundern.) Wir stellen keineswegs die These auf, dass ein semantisches Web prinzipiell unmöglich sei, wir erlauben uns nur zu betonen, dass auf dem Weg zum „semantic Web“ informationswissenschaftliche Begriffsforschung sehr nützlich ist: Sowohl der Aufbau von KOS als auch das Tag Gardening sind genuin informationswissenschaftliche Domänen.

Automatische Berechnungen

Bei funktionalen Relationen (wie *hatGeburtstag*) wird in einigen Fällen eine Zahl als Wert vorkommen. Ebenso können bei den Werten im Umkreis von Existenz- und Allquantor numerische Werte auftreten (wie oben *2 hasChild*). Immer wenn uns solche numerischen Werte vorliegen, können wir automatisch Berechnungen anstellen. Wir führen Optionen automatischen Berechnens anhand eines Beispiels von Sebastian Löhnner (persönliche Mitteilung, 2007) durch. Gegeben sei eine Wissensdomäne „Wohnungsanzeigen“ und eine Datenbank mit entsprechenden Texten. Im Frame von *Wohnung* kommen Funktionalbegriffe wie Größe (in m²), Anzahl der Zimmer, monatlicher Mietpreis und Lage vor. Die ersten drei funktionalen Relationen enthalten als Werte in konkreten Anzeigen Zahlen, mit denen das Informationssystem wird rechnen können. Als Voraussetzung ist zu notieren, dass die Begriffsordnung so konstruiert ist, dass sowohl die Funktionalbegriffe als auch die jeweiligen Werte automatisch erkannt und aus den Dokumenten fehlerfrei extrahiert werden. Bei der Textsorte der Wohnungsanzeigen (mit sehr kleinen und gut strukturierten Texten) sollte dies realistisch sein, für komplexere Texte bleibt abzuwarten, ob die Extraktionsverfahren zufriedenstellend verlaufen können.

Ein Nutzer habe nach Wohnungsangeboten in Düsseldorf-Gerresheim gesucht. Die Recherche verläuft über die Werte der Ortsangaben bei der *Lage*. Die Extraktion von Funktionalbegriffen und numerischen Werten kann dann beispielsweise folgendes Ergebnis bringen:

Treffer Nr.	Größe	Preis
1	100	1.500
2	85	1.250
3	55	800
4	40	1.000
5	35	600.

Das System ist in der Lage, dem Nutzer – außer den Treffern – gewisse, die Einzelangaben verdichtende Werte anzubieten: In Düsseldorf-Gerresheim beträgt der durchschnittliche Mietpreis der aktuellen Angebote 17,28 Euro pro m², die günstigste Wohnung kostet 600 Euro, die teuerste liegt bei 1.500 Euro usw. Zudem ist es möglich, dem Nutzer die Angaben als Frame zu präsentieren, und dies sowohl für die einzelne Wohnungsanzeige als auch für die gesamte Treffermenge, hier mit den ermittelten statistischen Angaben. Ein hypothetischer weiterer Treffer, der keine Preisangaben enthält, kann mit einem informationellen Mehrwert versehen werden, indem Minimal-, Maximal- und Durchschnittswerte der sonstigen Angebote für Düsseldorf-Gerresheim angezeigt werden.

An dieser Stelle wird ein Ontologie-Editor (z.B. Protégé; Noy, Fergerson, & Musen, 2000; Noy et al., 2001) zum Einsatz kommen. Die informationswissenschaftlichen Forschungen zur automatischen Berechnung numerischer Werte von Funktionalbegriffen und zur Darstellung von Frames befinden sich derzeit in der Anfangsphase.

Fazit

Modelliert man Begriffe als Frames, beachtet Spezialprobleme wie die Familienähnlichkeit bei der Definition, unterscheidet bei den hierarchischen Relationen derart, dass stets transitive Relationen entstehen, und konstruiert generalisierbare weitere Relationen, so stehen der Informationspraxis fünf Anwendungslinien von Wissensordnungen offen, von denen drei weitgehendes informationswissenschaftliches Neuland bedeuten:

1. Suchen und Browsen in der Wissensordnung (Anwendung im begriffsorientierten Information Retrieval; bekannt);
2. Nutzung der Wissensordnung beim automatischen Indexing und Retrieval (bekannt),
3. Automatische Suchanfrageerweiterung auf der Basis der semantischen Nähe von Begriffen (weitgehend neu; Anwendung: vor allem im Information Retrieval),
4. Automatisches Schlussfolgern auf der Grundlage von Regelmäßigkeiten in der Terminologie (neu; Anwendung: vor allem beim Aufbau eines „Semantic Web“ oder eines „Social Semantic Web“),
5. Automatisches Berechnen verdichtender Informationen beim Vorliegen numerischer Werte von Funktionalbegriffen (neu; Anwendung: vor allem im Information Retrieval).

Literatur

Aristoteles (1970). *Metaphysik*. Stuttgart: Reclam.

Aristoteles (2004). *Topik*. Stuttgart: Reclam.

Ashburner, M. et al. [The Gene Ontology Consortium] (2000). Gene Ontology: Tool for the unification of biology. *Nature Genetics*, 25, 25-29.

Barsalou, L.W. (1992). Frames, concepts, and conceptual fields. In Kittay, E., Lehrer, A. (Hrsg.): *Frames, Fields and Contrasts. New Essays in Semantic and Lexical Organization* (S. 21-74). Hillsdale, N.J.: Lawrence Erlbaum Ass.

Bayer, O., *Höhfeld*, S., *Josbächer*, F., *Kimm*, N., *Kradepohl*, I., *Kwiatkowski*, M., *Puschmann*, C., *Sabbagh*, M., *Werner*, N., & *Vollmer*, U. (2005). Evaluation of an ontology-based knowledge-management-system. A case study of Convera RetrievalWare 8.0. *Information Services & Use*, 25, 181-195.

Bean, C.A. & *Green*, R., Hrsg. (2001). *Relationships in the Organization of Knowledge*. Boston: Kluwer.

Berners-Lee, T., *Hendler*, J., & *Lassila*, O. (2001). The semantic Web. *Scientific American*, 284(5), 28-37.

Bertram, J. (2005). Einführung in die inhaltliche Erschließung. *Grundlagen - Methoden - Instrumente*. Würzburg: Ergon.

Bertram, J. (2009). Social Tagging - Zum Potential einer neuen Indexiermethode. *Information – Wissenschaft und Praxis*, 60(1), 19-26.

Black, M. (1937). Vagueness. *Philosophy of Science*, 4, 427-455.

Blumauer, A., & *Pellegrini*, T., Hrsg. (2009). *Social Semantic Web. Web 2.0 – was nun?* Berlin, Heidelberg: Springer (X.media.press).

Bolzano, B. (1937). *Wissenschaftslehre*. Sulzbach: Seidel.

Breslin, J.G., *Passant*, A., & *Decker*, S. (2009). *The Social Semantic Web*. Berlin, Heidelberg: Springer.

Cruse, D.A. (2002). Hyponymy and its varieties. In Green, R., Bean, C.A., Myaeng, S.H. (Hrsg.): *The Semantics of Relationships* (S. 3-21). Dordrecht: Kluwer.

Cruse, D.A. (2004). *Meaning in Language. An Introduction to Semantics and Pragmatics*. Oxford: Oxford University Press.

Dahlberg, I. (1974). Zur Theorie des Begriffs. *International Classification*, 1(1), 12-19.

Dahlberg, I. (1986). Die gegenstandsbezogene, analytische Begriffstheorie und ihre Definitionsarten. In: *Ganter*, B., *Wille*, R., *Wolff*, K.E. (Hrsg.): *Beiträge zur Begriffsanalyse* (S. 9-22). Mannheim, Wien, Zürich: BI Wissenschaftsverlag.

DIN 2330:1993. *Begriffe und Benennungen. Allgemeine Grundsätze*. Berlin: Beuth.

DIN 2331:1980. *Begriffssysteme und ihre Darstellung*. Berlin: Beuth.

DIN 2342/1:1992. *Begriffe der Terminologielehre. Grundbegriffe*. Berlin: Beuth.

Domini, F.M., *Lenzerini*, M., *Nardi*, D., & *Schaerf*, A. (1996). Reasoning in description logics. In *Brewska*,

- G. (Hrsg.): Principles of Knowledge Representation (S. 191-236). Stanford, CA: CSLI Publications.
- Dubislav, W. (1981). Die Definition. Hamburg: Meiner. 4. Aufl.
- Durkheim, E. (2007[1912]). Die elementaren Formen des religiösen Lebens. Frankfurt: Insel. (Original: 1912).
- Frege, G. (1892). Über Sinn und Bedeutung. Zeitschrift für Philosophie und philosophische Kritik (Neue Folge), 100, 25-50.
- Fugmann, R. (1999). Inhaltserschließung durch Indizieren: Prinzipien und Praxis. Frankfurt: DGI.
- Gadamer, H.G. (1975). Wahrheit und Methode. 4. Aufl. Tübingen: J.C.B. Mohr (Paul Siebeck).
- Ganter, B. & Wille, R. (1996). Formale Begriffsanalyse. Mathematische Grundlagen. Berlin, Heidelberg: Springer.
- Gerstl, P. & Pribbenow, S. (1996). A conceptual theory of part-whole relations and its applications. Data & Knowledge Engineering, 20, 305-322.
- Green, R. (2001). Relationships in the organization of knowledge: Theoretical background. In Bean, C.A., Green, R. (Hrsg.): Relationships in the Organization of Knowledge (S. 3-18). Boston: Kluwer.
- Green, R., Bean, C.A., & Myaeng, S.H., Hrsg. (2002). The Semantics of Relationships. Dordrecht: Kluwer.
- Gust von Loh., S., Stock, M., & Stock, W.G. (2009). Knowledge organization systems and bibliographical records in the state of flux. Hermeneutical foundations of organizational information culture. In Proceedings of the Annual Meeting of the American Society for Information Science and Technology (ASIS&T 2009), Vancouver.
- Heidegger, M. (1967[1927]). Sein und Zeit. 11. Aufl. Tübingen: Niemeyer. (Original: 1927).
- Hillsdale, N.J.: Lawrence Erlbaum Ass.; Saussure, F. de (2001[1916]). Grundlagen der allgemeinen Sprachwissenschaft. Berlin, New York: de Gruyter. 3. Aufl. (Original: 1916).
- Hjørland, B. (2007). Semantics and knowledge organization. Annual Review of Information Science and Technology, 41, 367-406.
- Hjørland, B. (2009). Concept theory. Journal of the American Society for Information Science and Technology, 60(8), 1519-1536.
- Hoensbroech, F. (1931). Über Beziehungen zwischen Umfang und Inhalt von Begriffen. Erkenntnis, 2, 291-300.
- Horrocks, I. & Sattler, U. (1999). A description logic with transitive and inverse roles and role hierarchies. Journal of Logic and Computation, 9(3), 385-410.
- Hovy, E. (2002). Comparing sets of semantic relations in ontologies. In Green, R., Bean, C.A., Myaeng, S.H. (Hrsg.): The Semantics of Relationships (S. 91-110). Dordrecht: Kluwer.
- Johansson, I. (2004). On the transitivity of the part-whole relations. In: Hochberg, H., Mulligan (Hrsg.): Relations and Predicates (S. 161-181). Frankfurt: Ontos.
- Khoo, C.S.G. & Na, J.C. (2006). Semantic relations in information science. Annual Review of Information Science and Technology, 40, 157-228.
- Klaus, G. (1973). Moderne Logik. Berlin: Deutscher Verlag der Wissenschaften. 7. Aufl.
- Komatsu, L.K. (1992). Recent views of conceptual structure. Psychological Bulletin, 112(3), 500-526.
- Lancaster, F.W. (2003). Indexing and Abstracting in Theory and Practice. Champaign: University of Illinois. 3. Aufl.
- Löbner, S. (2003). Semantik. Eine Einführung. Berlin, New York: Walter de Gruyter.
- Löbner, S. (2005). Funktionalbegriffe und Frames – Interdisziplinäre Grundlagenforschung zu Sprache, Kognition und Wissenschaft. Jahrbuch der Heinrich-Heine-Universität Düsseldorf 2004, 463-477.
- May, W. (2006). Reasoning im und für das Semantic Web. In Pellegrini, T., Blumauer, A. (Hrsg.): Semantic Web. Wege zur vernetzten Wissensgesellschaft (S. 485-503). Berlin, Heidelberg: Springer.
- Menne, A. (1980). Einführung in die Methodologie. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Mervis, C.B., & Rosch, E. (1981). Categorization of natural objects. Annual Review of Psychology, 32, 89-115.
- Minsky, M. (1975). A framework for representing knowledge. In Winston, P.H. (Hrsg.): The Psychology of Computer Vision (S. 211-277). New York: McGraw-Hill.
- Nardy, D. & Brachman, R.J. (2003). An introduction to description logics. In Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P. (Hrsg.): The Description Logic Handbook. Theory, Implementation and Applications (S. 1-40). Cambridge: Cambridge University Press.
- Noy, N.F., Ferguson, R.W., & Musen, M.A. (2000). The knowledge model of Protégé-2000: Combining interoperability and flexibility. Lecture Notes in Computer Science, 1937, 69-82.
- Noy, N.F., Sintek, M., Decker, S., Crubezy, M., Ferguson, R.W., & Musen, M.A. (2001). Creating Semantic Web contents with Protégé-2000. IEEE Intelligent Systems, 16(2), 60-71.
- Ogden, C.K. & Richards, I.A. (1974[1923]). Die Bedeutung der Bedeutung. Frankfurt: Suhrkamp. (Original: 1923).
- Pawłowski, T. (1980). Begriffsbildung und Definition. Berlin, New York: de Gruyter.
- Peters, I. & Weller, K. (2008a). Paradigmatic and syntagmatic relations in knowledge organization systems. Information – Wissenschaft und Praxis, 59(2), 100-107.
- Peters, I. & Weller, K. (2008b). Tag gardening for folksonomy enrichment and maintenance. Webology, 5(3), article 58.
- Peters, I. & Stock, W.G. (2008). Folksonomies in Wissensrepräsentation und Information Retrieval. Information – Wissenschaft und Praxis, 59(2), 77-90.
- Petersen, W. (2007). Representation of concepts as frames. The Baltic International Yearbook of Cognition, Logic and Communication, 2, 151-170.
- Pribbenow, S. (2002). Meronymic relationships: From classical mereology to complex part-whole relations. In Green, R., Bean, C.A., Myaeng, S.H. (Hrsg.): The Semantics of Relationships (S. 35-50). Dordrecht: Kluwer.
- Priss, U. (2006). Formal concept analysis in information science. Annual Review of Information Science and Technology, 40, 521-543.
- Reimer, U. (1991). Einführung in die Wissensrepräsentation. Stuttgart: Teubner.
- Rosch, E. (1975a). Cognitive representations of semantic categories. Journal of Experimental Psychology – General, 104(3), 192-233.
- Rosch, E. (1975b). Cognitive reference points. Cognitive Psychology, 7(4), 532-547.
- Rosch, E. & Mervis, C.B. (1975). Family resemblances: Studies in the internal structure of categories. Cognitive Psychology, 7(4), 573-605.
- Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., & Boyes-Braem, P. (1976). Basic objects in natural categories. Cognitive Psychology, 8(3), 382-439.
- Rosch, E. (1983). Prototype classification and logical classification: The two systems. In Scholnick, E.K. (Hrsg.): New Trends in Conceptual Representation: Challenges to Piaget's Theory? (S. 73-86).
- Schmidt, S.J. (1969). Bedeutung und Begriff. Braunschweig: Vieweg.
- Schmitz-Esser, W. (1999). Thesaurus and beyond: An advanced formula for linguistic engineering and information retrieval. Knowledge Organization, 26, 10-22.
- Schmitz-Esser, W. (2000). EXPO-INFO 2000. Visuelles Besucherinformationssystem für Weltausstellungen. Berlin: Springer.
- Shadbolt, N., Hall, W., & Berners-Lee, T. (2006). The semantic web revisited. IEEE Intelligent Systems, 21(3), 96-101.
- Simons, P. (1987). Parts. A Study in Ontology. Oxford: Clarendon.
- Smith, B., Williams, J., & Schulze-Kremer, S. (2003). The ontology of the Gene Ontology. AMIA Annual Symposium Proceedings, 609-613.
- Smith, B. et al. (2005). Relations in biomedical ontologies. Genome Biology 6(5), Art. R46.
- Stock, W.G. (2007). Information Retrieval. Informationen suchen und finden. München, Wien: Oldenbourg.
- Stock, W.G. & Stock, M. (2008). Wissensrepräsentation. Informationen auswerten und bereitstellen. München: Oldenbourg.
- Storey, V.C. (1993). Understanding semantic relationships. VLDB Journal, 2, 455-488.
- Weller, K. & Stock, W.G. (2008). Transitive meronymy. Automatic concept-based query expansion using weighted transitive part-whole-relations. Information – Wissenschaft und Praxis, 59(3), 165-170.
- Wersig, G. (1974). Information – Kommunikation – Dokumentation. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Winston, M.E., Chaffin, R., & Herrmann, D. (1987). A taxonomy of part-whole relations. Cognitive Science, 11, 417-444.
- Wittgenstein, L. (1977[1953]). Philosophische Untersuchungen. Frankfurt: Suhrkamp. (Original: 1953).

Begriff, semantische Relation, Wissensrepräsentation, semiotisches Dreieck, Synkategoriemata, Intension, Extension, Vagheit, Prototyp, Definition, Norm, Familienähnlichkeit, Begriffserklärung, Frame, Begriffsordnung, Wissensordnung, Ordnung, Koordinationsgrad, Synonymie, Antonymie, Hierarchie, Hyponymie, Taxonomie, Meronymie, Instanz, Assoziationsrelation, Suchanfrageerweiterung, automatisches Schlussfolgern, Ontologie, Semantic Web, Social Semantic Web, automatische Berechnung, Funktionalbegriff

DER AUTOR

Univ.-Prof. Dr. Wolfgang G. Stock



ist Leiter der Abteilung für Informationswissenschaft der Heinrich-Heine-Universität Düsseldorf. Der Artikel basiert auf Ideen aus dem Buch Wissensrepräsentation, das der Autor gemeinsam mit Mechtilde Stock im Oldenbourg-Verlag (München) veröffentlicht hat.

Institut für Sprache und Information
Abteilung Informationswissenschaft
Heinrich-Heine-Universität Düsseldorf
Universitätsstraße 1, Geb. 23.21.04
40225 Düsseldorf
stock@phil-fak.uni-duesseldorf.de

Institut für Sprache und Information
Abteilung Informationswissenschaft
Heinrich-Heine-Universität Düsseldorf
Universitätsstraße 1, Geb. 23.21.04
40225 Düsseldorf
stock@phil-fak.uni-duesseldorf.de