



# Domain-Ontologie der Hochschulwelt für ein Portal

## Zugangsvokabular und Basis für Inferenzen

*Gerhard Knorz*

### Zusammenfassung

Eine große Fachhochschule arbeitet an einem internen Portal, das auf einer Domain-Ontologie der Hochschulwelt basiert. Die Ontologie muss den Nutzer bei dessen Terminologie abholen (Zugangsvokabular) und ihm die realen Sachzusammenhänge und die normative Terminologie vermitteln. Sie dient der Faktendokumentation und dem Dokumentenmanagement. Insbesondere muss die Semantik der Begriffe und deren Relationierung so präzise und konsistent sein, dass zuverlässige Inferenzregeln formuliert werden können. Es wird beschrieben, wie auf der Basis des verwendeten Werkzeugs k-infinity das Begriffsmodell aufgebaut ist und wie Inferenzregeln formuliert werden.

### Abstract

A large university of applied sciences works on an internal portal that is based on a domain-ontology of the university world. The ontology must take along the user at his level of terminology and impart knowledge about the given facts and the normative terminology. It serves the factual documentation and the document management. Especially the semantics of the concepts and of their relationships must be so precise and consistent that dependable inference rules can be expressed. It is described how on the basis of the tool k-infinity the concept model is constructed and how inference rules can be formulated.

## 1 Bedarf und Anforderungen für das Wissensmanagement an der FH Darmstadt

Die Fachhochschule Darmstadt ist mit 5 Standorten in 2 Städten, ca. 700 Mitarbeitern (einschließlich Professoren), 15 Fachbereichen und mehr als 11.000 Studierenden in mehr als 30 Studiengängen eine verteilte, heterogene Institution. Verwaltungsvorgänge und Gremienarbeit, die in einem Fall reibungslos funktionieren, können zu anderer Zeit oder an anderer Stelle vollständig misslingen und zu fatalen Konsequenzen führen, wenn etwa zeitkritische Anträge an den Senat wegen Formfehlern bzw. weit zurückliegenden Versäumnissen abgelehnt oder zurückgestellt werden müssen.



Das zu Grunde liegende Problem ist in vielen Unternehmen anzutreffen und stellt sich im Hochschulbereich mit seiner Selbstverwaltung systematisch verschärft. Zentrale und dezentrale Organisationseinheiten übersehen den Gesamtprozess nicht, kennen den jeweilig anderen Bereich kaum und verstehen wenig dessen Bedingungen und Bedürfnissen. Eine wesentliche Rolle spielen im Rahmen von Gremienarbeit und Fachbereichsleitung Personen, die für die dort anfallende Arbeit weder eine spezielle Qualifikation mitbringen noch diese Tätigkeit wirklich als eine ihrer Kernaufgaben auffassen. Sie wird in der Regel als ungeliebte zusätzliche Verpflichtung verstanden und „nebenbei“ erledigt. Und dies, obwohl Verwaltungsarbeit in jeder Art von (typischer) Hochschulrolle einen signifikanten Anteil ausmacht (Lummerstädt 2003).

Wissensmanagement an einer Hochschule, das am vordringlichen Bedarf ansetzt, wird sich also zunächst nicht auf die Ebene von Lehre und Forschung zu konzentrieren haben, sondern auf Verwaltungshandeln im weitesten Sinne, an dem Mitarbeiter, Professoren und Studierende beteiligt sind. Dies entspricht auch dem Auftrag, den sich ein im März 2002 vom Senat eingesetzter Ausschuss „Wissensmanagement“ gegeben hat (Baumer & Knorz 2003).

Eine Bedarfserhebung im Rahmen mehrerer Workshops identifizierte insbesondere folgende Defizite (Knorz & Müller 2003):

- Die verschiedenen Bereiche (Standorte, Gruppen, Fachkulturen) sprechen unterschiedliche „Sprachen“.
- Informationen werden einerseits oft redundant und in unterschiedlicher (und wechselnder) Qualität vorgehalten, andererseits sind vorhandene Informationsquellen verteilt und oftmals nur Wenigen bekannt. Die Folgen sind ineffizienter Mitarbeiterereinsatz und Unzufriedenheit bei den Nutzern.
- Transparenz fehlt nicht nur hinsichtlich Informationsquellen sondern auch in Bezug auf Prozesse, Ansprechpartner, Verantwortlichkeiten, Vertretungsregelungen und kommunikativer Erreichbarkeit. Die Konsequenzen sind Irrwege, Zeitverluste und insbesondere Probleme bei abteilungsübergreifenden Prozessen.

Das ambitionierteste Vorhaben als Antwort auf diese Defizite ist die Entwicklung eines internen Portals für die Fachhochschule, das alle wesentlichen Daten und Informationen, mit denen an der FHD gearbeitet werden, integriert und zur Verfügung stellt.

## **2 Portal**

Der Kern des Portals besteht in einer Modellierung der Begrifflichkeiten, die an der FH Darmstadt verwendet werden. Soweit bietet das Portal die Funktio-

nalität eines sehr differenzierten Glossars. Die Extensionen dieser Begriffe, die Individuen, und deren Beziehungen untereinander, bilden die konkrete Welt der Hochschule als ein Informationsmodell nach (Faktdokumentation). Zum Weltausschnitt gehören auch die Nutzer selbst, wodurch eine Individualisierung des Portals nahe liegend realisiert werden kann (Push- und Pulldienste), indem Interessensgebiete ausgewählt werden und deren Position in der Hochschule ausgewertet wird (Abb. 1).

Begriffe und Individuen können thematisch Dokumenten (als Bestandteil des Portals) als Metainformation zugeordnet werden (Dokumentenmanagement). Im weiteren Ausbau sollen Workflows modelliert und realisiert werden (Services).

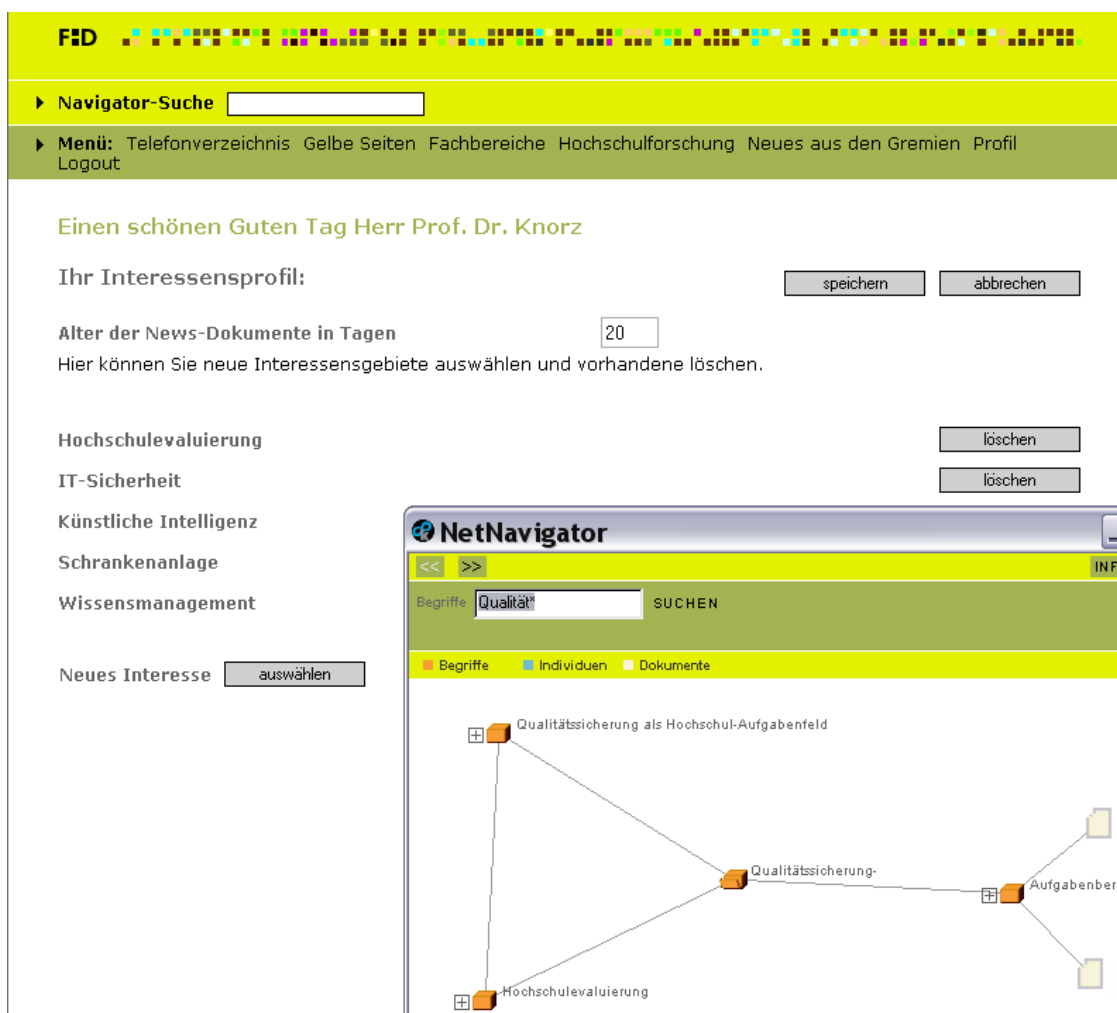
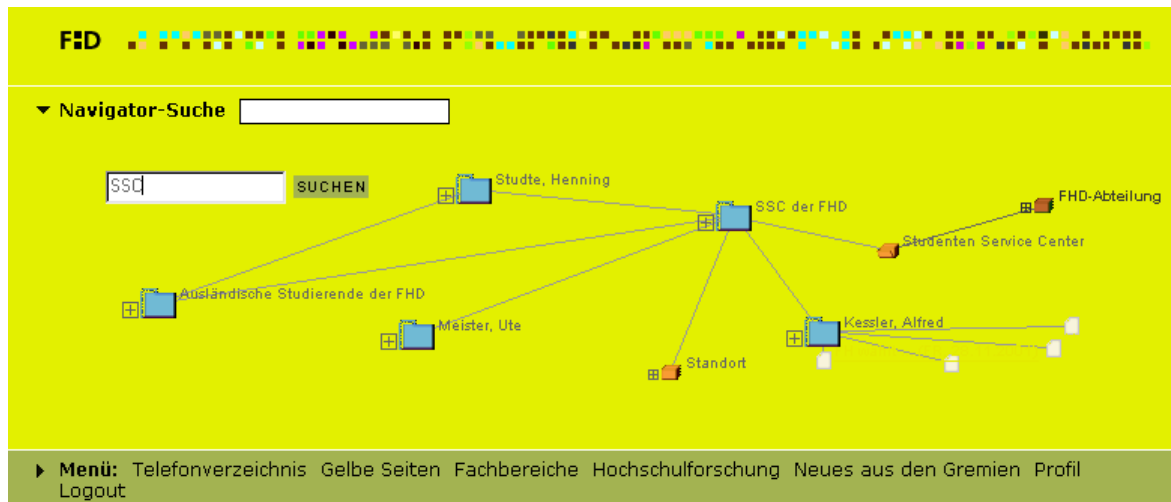


Abb. 1: Auswahl von Interessensgebieten im Net-Navigator zur Spezifizierung des eigenen Interessensprofils. Hier der Begriff „Hochschulevaluierung“, gefunden über den „Net-Navigator“ mit einem Einstieg durch das Suchwort „Qualität\*“.

Es gibt mehrere Zugänge zum Informationsbestand des Portals.

Vordefinierte Formen von typischem Informationsbedarf werden über entsprechende Menüpunkte (Telefonbuch, gelbe Seiten, Fachbereiche) bedient. Kontextabhängig können spezifische Informationen zu ausgewählten Objekten ermittelt werden, z.B. die strukturierte Menge aller Mail-Adressen (z.B. Adressen einer Einrichtung, seiner Beschäftigten oder seiner Mitglieder als Funktion des Mail-Buttons in Abb. 2).



Einen schönen Guten Tag Herr Prof. Dr. Knorz



Abb. 2: Ein Wissensobjekt, hier ein Individuum vom Typ „Student Service Center“. Oben wurde das Wissensnetz vom Nutzer eingeblendet. Die Navigation kann sowohl mittels der grafischen wie auch der Text-Darstellung erfolgen.

Eine „semantische Suche“ findet Treffer (Begriffe, Individuen, Dokumente) durch Abgleich mit deren Benennungen unter Einbeziehung der Netzstruktur.

Durch Navigation (vorzugsweise in der graphischen Repräsentation des Begriffsnetzes) (siehe Abb. 2) kann sich ein Nutzer von einem zunächst gewählten / gefundenen Begriff oder Individuum hin zum eigentlich Gesuchten bewegen und sich dabei sowohl in der Sache wie auch in der Terminologie orientieren.

## **3      Begriffsmodell**

### **3.1    Anforderungen**

Grundlage des Portals ist eine Domain-Ontologie der „Hochschulwelt“, wobei deren Umgebung a priori nur unscharf abzugrenzen ist (beispielsweise kann so gut wie alles zum Forschungsgegenstand gemacht werden).

Ontologien sind in zwei Ausprägungen anzutreffen (Fensel 2001), die grundsätzlich bekannten konventionellen Typen der Wissensdarstellung entsprechen:

- Konzeptionelle Ontologien übernehmen den Anspruch von Klassifikationssystemen, unabhängig von konkreten Begriffsbildungen eines Sprachsystems ein konsistentes System von Klassen (und damit Begriffen) zu schaffen.
- Linguistische Ontologien setzen den Weg fort, den Thesauri vorgezeichnet haben, indem sie von den eingeführten Benennungen eines Sprachsystems ausgehen und das dazu passende Begriffssystem zu modellieren versuchen.

Da die Verständigung und die Orientierung aller Hochschulangehörigen eine Kernaufgabe des Portals darstellt, muss dessen Ontologie vom linguistischen Typ sein. Dabei ist das Problem der verschiedenen Begriffsbildungen und Perspektiven einzufangen und auf ein normatives Sprachsystem zurückzuführen. *Studentische Hilfskräfte* und „*Hiwis*“ existieren im Denken von Fachbereichsangehörigen auch dann noch, wenn es längst nur noch *BAT-Aushilfskräfte* gibt (die vielleicht zufällig Studierende sind).

Andererseits werden die Begriffe zwar als Metadaten zur thematischen Verortung von Dokumenten verwendet, in erster Linie sollen sie aber die reale Hochschulwelt abbilden. Die Semantik der Begriffe und ihrer Relationierung muss demnach für diesen Zweck angemessen reichhaltig und auch präzise sein. Inferenzregeln können nur dann brauchbare Ergebnisse liefern, wenn die Modellierung in sich konsistent ist (Kap. 3.4).

In diesem Sinn ist es Anforderung an diese Domain-Ontologie, einen praxistauglichen Kompromiss zwischen der bekannten schwachen Semantik von umfassenden Thesauri und exakten Semantik von sachlich eng begrenzten Datenbankwelten zu realisieren. Erfahrungen darüber, wie weit dies gelingen kann und welche Leistungsfähigkeit daraus folgt, fehlen weitgehend. Das FHD-Portal eröffnet deshalb im Hinblick auf Forschung Chancen zur Beantwortung entscheidender Fragen, die gegenwärtig bei der Formulierung der Vision des „semantic Web“ weitgehend ausgeblendet bleiben (Berners-Lee et al. 2001).

### 3.2 Randbedingungen

Das FHD-Portal und die zugrunde liegende Ontologie der „Hochschulwelt“ wird mit der Software-Suite *k-infinity* der Firma *intelligent views* realisiert. Intern wird dieses semantische Netz in einem Framesystem verwaltet, wobei ein Export in das XML-basierte Topic Map-Format XTM (vgl. Pepper & Moore 2001) und den Frameformalismus RDF (Lasilla & Swick 1999) möglich ist. Der Aufbau des Netzes sowie die Konfiguration von Suchfunktionen bzw. die Formulierung von vordefinierten Suchanfragen geschieht mittels der Komponente *Knowledge Builder*. Zu den besonderen Vorzügen von *k-infinity* und des *Knowledge Builders* gehört die gute Unterstützung der Modellierung durch eine leicht bedienbare grafische Oberfläche sowie eine überzeugende Visualisierung des Wissensnetzes für den Modellierer (Knorz & Müller 2003).

Gegenwärtig liegt ein Prototyp vor, der in absehbarer Zeit der Hochschulöffentlichkeit vorgestellt werden soll (Baumer & Knorz 2003). In die Entwicklung des Begriffsnetzes wurden Studierende im Rahmen von Lehrveranstaltungen einbezogen. Sie erfolgte in mehreren Zyklen. Das Modell entstand aus der Verallgemeinerung der Modellierung von typischen Fallbeispielen in einem Wechsel von Top down und Bottom up-Vorgehen.

Die Entwicklung der eigentlichen Portal-Software, die das semantische Netz dem Nutzer über das Internet zugänglich macht, nahm in einer Diplomarbeit ihren Anfang (Müller 2003).

### 3.3 Top-Level-Ontologie

Einen fokussierten Ausschnitt der Top-Level-Ontologie, aus der heraus sich die Domain-Ontologie entwickelt, zeigt Abb. 3, anhand der die wesentlichen Eigenschaften und Bestandteile eines Wissensnetzes (i-views 2001) erläutert werden sollen. In Abb. 3 nicht dargestellt werden die anderen Top-Level-Begriffe *Menge*, *thematisch Beschreibbares*, und *Temporales*.

*Begriffe* (hier: z.B. *T*, *Konkretes*, *Verantwortbares*) werden durch *Attribute* beschrieben und in (polyhierarchischen) *Vererbungshierarchien* organisiert. Sie können *abstrakt* sein (z.B. *Konkretes*, *Aktor*, ...), *Extensionen* (Individuen, Instanzen) bilden (z.B. *Person*) oder aber diese *erweitern* (z.B. *Position*, *Standort*).

*Extensionen* (im Folgenden stets: *Individuen*), die konkrete Werte für die in den Begriffen definierten Attributen enthalten (können). Sie sind stets Individuum von nur genau einem Begriff.

*Erweiterungsfähige Begriffe* sind Begriffe, die Individuen um Attribute und Relationen „erweitern“ (z.B. *Position* erweitert Individuen vom Typ *Person*). Die Erweiterungen stellen Rollen des erweiterten Individuums (Subkontexte, (McCarthy 1993) dar (z. B. die Person *Mustermann* in der Rolle des *Dekans*). Erweiterungen können nicht selbst wieder erweitert werden, was die Modellbildung unerwartet einschränken kann. Die Konsequenzen dieser Modellierungsoption sind komplex und werden in Fischer 2003 kritisch diskutiert. Sie bedürfen einer gesonderten Diskussion.

*Relationen*, welche die Beziehungen zwischen Entitäten (Individuen und Begriffen) herstellen. Es gibt zunächst Systemrelationen (*Instanz von*, *Erweiterung von*, *Unterbegriff von*) und benutzerdefinierte Relationen. Ein Beispiel davon sind die Begriffsbeziehungen in Abb. 3 *leitet* und *wird geleitet durch*. Beziehungen zwischen Individuen werden in Abb. 6 gezeigt (z.B. *betreibt*).

Die Relationstypen werden primär beschrieben durch Quelle und Ziel und in Hierarchien strukturiert. Zum Beispiel ist *arbeitet in* (siehe Abb. 4) eine Relation zwischen Individuen vom Typ *Person* und von Individuen von *organisatorische Einheit*. Dagegen ist *besetzt Stelle* (Abb. 4) spezifischer definiert, weil das Ziel nicht allgemein eine organisatorische Einheit sein darf, sondern spezifischer eine „Stelle“ sein muss (Abb. 3). Wird in Abfragen auf die Relation *arbeitet in* Bezug genommen, so wird die Relation *besetzt Stelle* mit eingeschlossen. Die Konsistenz der Über- und Unterordnungen wird gegenwärtig von k-infinity (noch?) nicht überprüft. Zwischen notwendigen und hinreichenden Bedingungen („ist jede *arbeitet in*-Relation mit einer *Stelle* als Ziel notwendigerweise eine *besetzt-Stelle*-Relation?) wird nicht unterschieden. Allerdings kann bei hinreichenden Bedingungen die Relation als so genannte „Abkürzungsrelation“ inferiert werden (Kap. 3.4).

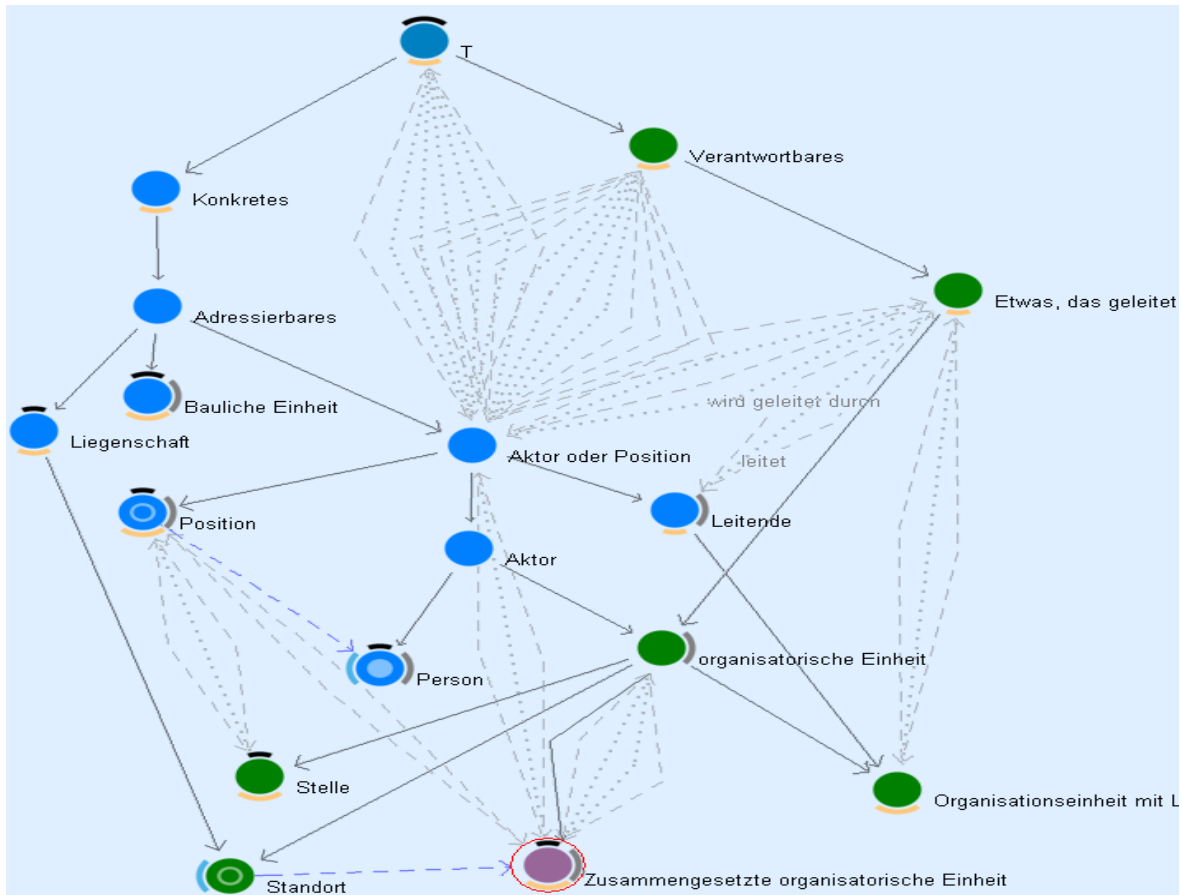


Abb. 3: Ausschnitt des Begriffsmodells im Bereich der Personen und Institutionen. Durchgezogene Linien repräsentieren Abstraktionsbeziehungen, gestichelte Linien Begriffsbeziehungen

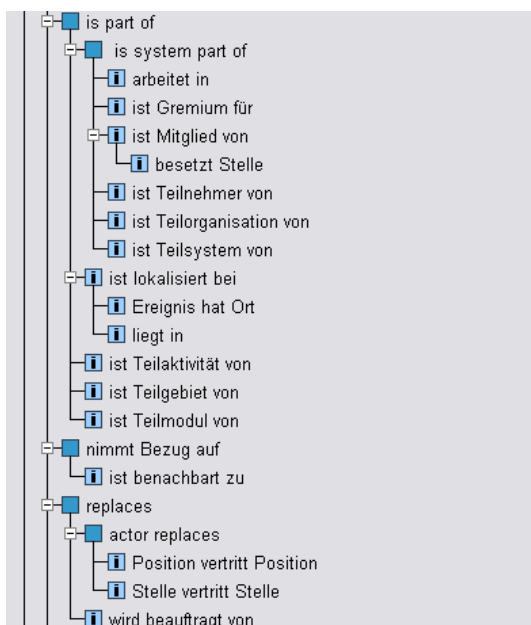


Abb. 4: Ausschnitt aus der Hierarchie der Relationstypen. Englische Relationennamen zeigen (einer Namenskonvention entsprechend) an, dass es sich um abstrakte Relationen handelt, die keine Realisierungen, sondern ausschließlich Unterrelationen haben.



### 3.4 Inferenzen

Eine wesentliche Option semantischer Netze ist es, aus vorhandenem Wissen neues Wissen zu erschließen (Inferenz). Voraussetzung dazu ist allerdings eine präzise definierte und konsistent eingehaltene Semantik des Netzes. Verletzungen dieser Voraussetzung entstehen recht leicht, wenn

- ein Begriff (aus unterschiedlichen Perspektiven) unterschiedlich interpretiert werden kann und diese Bedeutungen im Netz nicht auseinander gehalten werden
- dieselben Relationen in verschiedenen Kontexten zwar plausibel, aber unterschiedlich eingesetzt werden.

K-infinity ist in erster Linie als ein Werkzeug zur Modellierung von Terminologie und von Weltausschnitten entwickelt worden. Die Möglichkeiten, Inferenzen zu formulieren, waren dabei von Beginn an angelegt und haben sich in den letzten Monaten deutlich entwickelt. In Abhängigkeit von den Anforderungen konkreter Anwendungen, so die Vermutung, wird diese Entwicklung noch nicht zu einem Ende gekommen sein.

Die technische Basis für Inferenzen besteht aus drei Komponenten:

- „*Abkürzungsrelationen*“ sind frei definierbare Hintereinanderschaltungen von system- und benutzerdefinierten Relationen sowie anderen Abkürzungsrelationen. Die in Abb. 5 verwendete Relation OB-trans beispielsweise ist definiert als die transitive Hülle der Systemrelation *ist Unterbegriff von*.
- „*Expertensuchen*“ erlauben es, Muster im Netz zu definieren, indem spezifische Bedingungen an die verwendeten Relationen, Objekttypen oder Attributwerte formuliert werden (Abb. 5).
- *Vereinigungs-, Durchschnitts- und Negationsordner* erlauben es, verschiedene Komponenten der Suche zu bündeln und eine logische Struktur zwischen ihnen vorzugeben.

Abb. 5 dokumentiert eine der Formulierungen einer Expertensuche, die neben anderen in einem Vereinigungsordner die Funktion des Menü-Punktes der gelben Seiten realisieren kann.

Sie ist wie folgt zu lesen:

- Gesucht wird auf der Basis eines Parameterwertes, nach dem mit dem Text „*Beschäftigung mit*“ gefragt wird.
- Gefunden werden sollen Individuen vom Typ *Person*, die durch *Position* erweitert werden und in dieser Erweiterung eine *Aktivität* (beispielsweise eine *Tätigkeit*) betreiben. Diese Aktivität muss *Individuum* eines *Begriffs*

vom Typ *Aktivität* sein, von dem irgendein Oberbegriff vom Typ „*Verantwortbares*“ (hier wird auf die Abkürzungsrelation OB-trans Bezug genommen) *spezifisch ist hinsichtlich* eines Begriffes, dessen Bezeichnung mit dem Parameterwert „matched“.

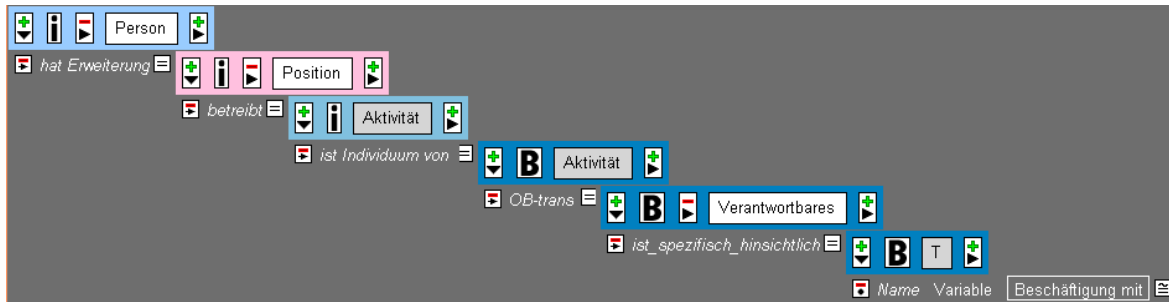


Abb. 5: Expertensuche, definiert im so genannten Knowledge Builder. Nach rechts ist die Darstellung aus Formatgründen abgeschnitten, so dass hier einige Details nicht wiedergegeben sind.

Die Relation *ist spezifisch hinsichtlich* verbindet einen Unterbegriff mit demjenigen Begriff (oder aber Individuum), der es in Bezug auf seinen Oberbegriff spezifisch macht (in Anlehnung an Rahmstorf 1994). Siehe dazu Abb. 6. Zu Beispielen siehe Abb. 7.

Die bereits benutzte Abkürzungsrelation OB-trans ist in Abb. 6 rot ein-gezeichnet. An dieser Abb. lässt sich ablesen, warum die Person Ruth sich auf die Anfrage „Studienbewerber“ hin qualifiziert.

#### Relation „ist das Spezifische an“/„ist spezifisch hinsichtlich“

**Historie:** hieß früher: ist/hat unterscheidendes Merkmal

**Bedeutung:** Wenn „Hessisches Ministerium“ Unterbegriff von „Ministerium“ ist, dann ist Hessen „das Spezifische an“ dem Begriff „Hessisches Ministerium“. „Hessen“ sorgt nämlich dafür, dass aus Ministerium der Unterbegriff „Hessisches Ministerium“ wird.

**Anwendung:** Die Relation wird ausschließlich intern verwendet werden (sie wird also dem Portalnutzer nicht gezeigt). Sie wird benötigt, um dem System die Schlussfolgerungen zu ermöglichen, die ein Mensch unmittelbar und trivialerweise zieht. Parallel muss/soll versucht werden, eine inhaltlich spezifische Relation zu finden, um den Zusammenhang zwischen dem differenzierenden Begriff (hier: Hessen) und dem differenzierten Begriff (hier: „Hessisches Ministerium“) herzustellen (hier: „ist zuständig für“)

Stand: 14.11.2003 in Änderung von 11.5.2003, Bearbeiter: Knorz

Abb. 6: Auszug aus der Dokumentation der Domain-Ontologie „Hochschulwelt FHD“

## 4 Ausblick

Ziel der gegenwärtigen Arbeiten ist es, das gegenwärtig modellierte Vokabular von mehr als 1000 Begriffen und von mehr als 2000 Individuen so zu erweitern, dass abgegrenzte Ausschnitte der Hochschulwelt weitgehend vollständig repräsentiert sind, um den Prototypen des Portals hochschulöffentlich für eine Exploration freigegeben zu können. Insbesondere die semantische Suche, die erwähnte „intelligente“ Mail-Funktionalität (Kap. 2) und die gelbe Seiten-Suchfunktion (Kap. 3.4) sollen das Potential der neuen Technologie deutlich machen. Ein später anstehender Schritt soll es sein, auch den öffentlichen Internetauftritt der Fachhochschule Darmstadt aus derselben Datenbasis, also aus dem semantischen Netz, zu generieren.

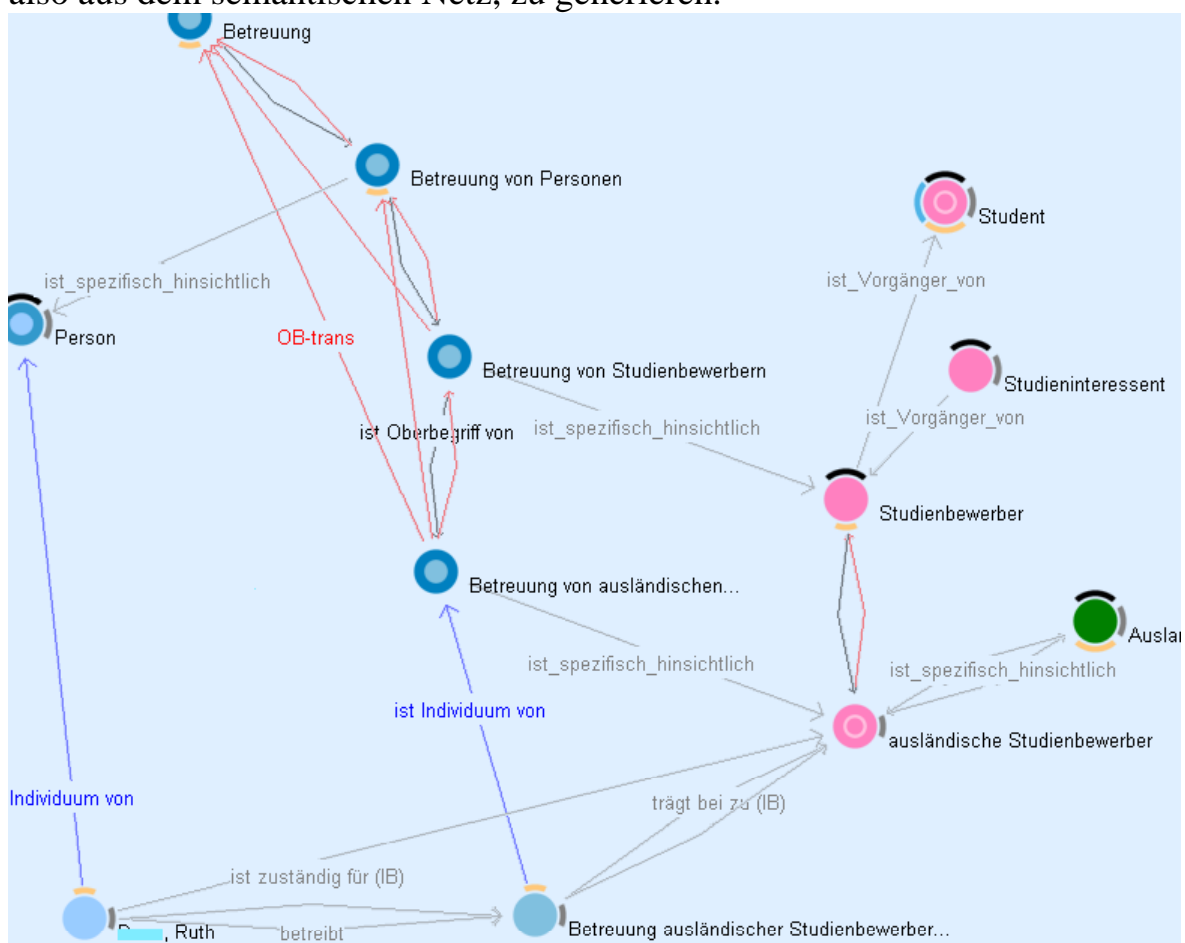


Abb. 7: Eine Person wird auf die Anfrage „Studienbewerber“ deswegen mittels Expertensuche im Netz gefunden, weil die Person Ruth, in ihrer Erweiterung um eine Position (hier nicht dargestellt) eine Tätigkeit betreibt (Betreuung ausländischer Studienbewerber an der FHD). Der dazugehörige Begriff besitzt nun u. a. den Oberbegriff (*Betreuung von Studienbewerbern*, relationiert durch *OB-trans*) der als Oberbegriff *Betreuung von Personen* hat und der spezifisch ist hinsichtlich „Studienbewerbern“.

## 5 Literaturverzeichnis

- Baumer, C., G. Knorz (2003). "Vom Wissensnetz zum Wissensportal." Wissensmanagement 05(2003), 36, 37.
- Berners-Lee, T, J. Hendler, O. Lassila (2001). "The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities" In: Scientific American, 05(2001).
- Fensel, D. (2001). Ontologies – A Silver Bullet for Knowledge Management and e-Commerce. Heidelberg: Springer Verlag.
- Fischer, D. (2003). Beschreibung und Diskussion des Repräsentationsformalismus von K-Infinity. Resvision 24.3.2003 der Fassung von 26.3.2002. Interner Bericht. Fraunhofer-Institut für Integrierte Publikations- und Informationssysteme IPSI, Darmstadt.
- Gruber, T. R. (1993). "A Translation Approach to Portable Ontology Specifications." Knowledge Acquisition, 5(1993),2, 199-220.
- Intelligent views (2001). White Paper K-Infinity. Intelligent views Wissenstechnologie: Wissenssteuerung durch Wissensnetze. <http://www.i-views.de/web/pdfs/Whitepaper-K-Infinity.pdf> [Zugriff September 2004].
- Knorz, G., J. Müller (2003). Wissensbasiertes Unternehmensportal an einer Hochschule. Interner Bericht. Fachhochschule Darmstadt 2003.
- Lummerstädt, J. (2003). Die Last der Bürokratie. DUZ 5(2003), 22-23.
- McCarthy, J. (1993). Notes on Formalizing Context. Proceedings of IJCAI 1993 <http://www-formal.stanford.edu/jmc/context-2.ps>. [Zugriff September 2004].
- Müller, J. (2003). Ontologisches Portal der FHD – Konzeption und Prototypentwicklung. Diplomarbeit am Fachbereich Informations- und Wissensmanagement, Fachhochschule Darmstadt, Juni 2003, <http://www.ontologies.de/diplom/da.pdf> [Zugriff September 2004].
- Pepper, S., G. Moore (ed.). XML Topic Maps (XTM) 1.0. TopicMaps.Org Specification. <http://www.topicmaps.org/xtm/1.0/> [Zugriff September 2004].
- Rahmstorf, G. (1994). "Semantisches Information Retrieval." In: W. Neubauer, DGD Frankfurt (ed.): Tagungsband des Deutschen Dokumentartages 1994, Trier, 237-60.



**In:** Eibl, Maximilian; Wolff, Christian; Womser-Hacker, Christa (Hg.): Designing Information Systems. Festschrift für Jürgen Krause. Konstanz: UVK Verlagsgesellschaft mbH, 2005. S. 115 – 126.



Dieses Dokument wird unter folgender [creative commons](http://creativecommons.org/licenses/by-nc-nd/2.0/de/) Lizenz veröffentlicht: