

# Semantische Suche in einer Hochschulontologie

Gerhard Knorz, Birgit Rein

*Fachhochschule Darmstadt - University of Applied Sciences, Mai 2005*

*Vorversion - Erscheint in: Information, Wissenschaft & Praxis 56(2005)4 oder 5*

## **Zusammenfassung:**

Ontologien werden eingesetzt, um durch semantische Fundierung insbesondere für das Dokumentenretrieval eine grundlegend bessere Basis zu haben, als dies gegenwärtiger Stand der Technik ist. Vorgestellt wird eine an der FH Darmstadt entwickelte und eingesetzte Ontologie, die den Gegenstandsbereich Hochschule sowohl breit abdecken und gleichzeitig differenziert semantisch beschreiben soll. Das Problem der semantischen Suche besteht nun darin, dass sie für Informationssuchende so einfach wie bei gängigen Suchmaschinen zu nutzen sein soll, und gleichzeitig auf der Grundlage des aufwendigen Informationsmodells hochwertige Ergebnisse liefern muss. Es wird beschrieben, welche Möglichkeiten die verwendete Software K-Infinity bereitstellt und mit welchem Konzept diese Möglichkeiten für eine semantische Suche nach Dokumenten und anderen Informationseinheiten (Personen, Veranstaltungen, Projekte etc.) eingesetzt werden.

## **Abstract:**

Ontologies aim to improve present day retrieval quality by means of their semantic founding. We introduce a particular ontology, developed and used by the University of Applied Sciences Darmstadt, which covers the "university world" both broadly and at the same time at a very detailed level. An appropriate semantic search has to fulfill two different requirements: For the user it should be as simple as any of the well known search engines, and at the same time it has to provide outstanding retrieval quality making use of its' underlying semantic model of the world. We describe the relevant features of the software K-Infinity and the way we use these features to build a powerful semantic search for documents and other information units (persons, meetings, projects etc.) on top of the university world ontology.

## **1 Die Vision semantikbasierter Suche und Informationsverarbeitung**

In Zeiten dynamischen Wandels, globalen Wettbewerbs und verteilter Unternehmen ist Wissensmanagement zu einem zentralen Thema der Wirtschaft geworden, unabhängig davon, welchen Stellenwert das Label „Wissensmanagement“ genießt.

Der Bedarf an Wissensmanagement trifft auf aktuelle technische Voraussetzungen und eine Entwicklungsdynamik, die dort Standards und Produkte entstehen lassen, wo über viele Jahre hinweg nur isolierte Projekte existierten. Die Vision des "Semantic Web" ist Ausdruck der Erwartung, dass sich auf Basis bekannter Verfahren der Wissensrepräsentation in heutigen Informationsarchitekturen ein grundlegend verbesserter Umgang mit Informationen erreichen lässt [Tim Berners-Lee et al. 2001]. Eine zentrale Rolle dabei spielen Ontologien, als Grundlage von Informationssystemen, die im Hinblick auf semantische Fundierung deutlich verbessert sind. Dahinter steht die Hoffnung, dadurch Funktionalitäten und Qualitäten zu schaffen, die über den gegenwärtig erreichbaren Stand deutlich hinausgehen [Gruber 1995].

Diese Hoffnung lässt sich begründen, aber wenig belegen. Anwendungen von Ontologien sind weitgehend zwei Enden eines Spektrums zuzuordnen, dessen Mitte nicht oder höchstens durch eher experimentelle Systeme besetzt ist: diese Ontologie-Anwendungen

- decken entweder ein weites Gebiet ab und arbeiten mit semantisch kaum differenzierten Relationen oder aber
- sie arbeiten im Hinblick auf Attribute und Relationen sehr differenziert bei gleichzeitiger Beschränkung auf eine Miniwelt.

Im ersten Fall geht das Ergebnis kaum über das hinaus, was ein konventioneller Thesaurus leistet, im zweiten Fall bewährt sich die Ontologie durch ihre Flexibilität im Entwurfsstadium, während das Ergebnis im wesentlichen auch durch konventionelle Datenbanken abzubilden ist. Das eigentliche Versprechen neuartiger Anwendungen wird in beiden Fällen (noch) nicht eingelöst.

Der vorliegende Beitrag beschäftigt sich mit dem Informationssystem „Hochschulportal“, das den Anspruch hat, das Zentrum des genannten Spektrums zu besetzen: Es soll auf einer Ontologie aufbauen, die einen ausgesprochen weiten Bereich abdeckt (nämlich die Welt, in der sich Hochschulen bewegen, operativ, strategisch und gesellschaftlich) und gleichzeitig diese Welt so differenziert abbilden, dass sowohl Dokumentretrieval als auch Faktenretrieval bruchlos möglich werden.

Nun gibt es in der Informationspraxis reichlich Erfahrungen mit Systemen von stark und schwach ausmodellierter Semantik: Relationale Datenbanken verwalten in der betrieblichen Praxis detailgenau die operativen Daten eines Unternehmen und unterstützen dessen Geschäftsprozesse. Auf der anderen Seite hat jeder Google-Nutzer ein Gefühl dafür, wie in Systemen recherchiert werden kann, die ohne ein Modell der (in den Dokumenten) abgebildeten Informationswelt auskommen.

Beide Systemtypen im Blick lässt sich die Problematik einer Suche im Hochschulportal klar aufzeigen: Denn für eine allgemeine Suchfunktion nützt es zunächst wenig, wenn sich grundsätzlich durch eine differenzierte Anfragesprache jede Informationseinheit finden lässt: weder die dazu notwendige Kenntnis des Datenmodells noch die Beherrschung der eigentlichen Abfragesprache lässt sich bei einem Nutzer voraussetzen (man denke in dieser Hinsicht an SQL). Andererseits wäre wenig gewonnen, wenn die modellierte Semantik in einer Art assoziativen Retrievals völlig aufgehen und verloren gehen würde, weil scheinbar

alles mit allem in Verbindung steht und unspezifisch als Treffer geliefert wird. Die Herausforderung besteht also darin, eine Art von Suche zu definieren, die der von klassischen Retrievalsystemen entspricht (also „so einfach wie bei Google“) und Ergebnisse liefert, deren Qualität vom zugrunde liegenden differenzierten Datenmodell (der Ontologie) unmittelbar und sichtbar profitiert (also ähnlich exakt wie bei relationalen Datenbanken). Eine solche Suche wird einem Nutzer unmittelbar vermitteln, welchen Gewinn ein semantisch fundiertes Informationssystem bedeutet. Im Folgenden soll gezeigt werden, dass sowohl die vorausgesetzte Ontologie als auch die von Nutzern benötigte Suchfunktionalität keine abstrakte Vision darstellt, sondern sich tatsächlich entwickeln lässt.

## **2 Prototyp eines ontologiebasierten Hochschulportals**

### **2.1 Informationsbedarf im Hochschulbereich**

Wissensmanagement beschäftigt sich damit, in Organisationen systematisch, planvoll und (erfolgs-) kontrolliert die Bedingungen dafür zu schaffen, dass das zur Lösung von Aufgaben und Unterstützung von Prozessen benötigte Wissen identifiziert und nötigenfalls generiert wird, dass es geteilt, weiterentwickelt und effektiv angewandt wird. Wissensmanagement lenkt also den Blick darauf, dass Wissen und damit die Kompetenz der Mitarbeiter eine besonders wichtige Ressource (und vielfach die wichtigste überhaupt) ist.

Die spezifischen Arbeitsbedingungen von Hochschulen, gekennzeichnet durch selbstständige Substrukturen, komplexe Entscheidungsprozesse und insbesondere die Selbstverwaltung schaffen eine spezifische Problemsituation in Bezug auf Transparenz, Kommunikation und effektives Handeln. Es mangelt weitgehend an einer gemeinsamen Sprache, an handlungsrelevanter Information und Kontextinformation, sowie an einer effektiven Unterstützung im Bereich von Verwaltungshandeln und Gremienarbeit und weiter darüber hinaus. Genau darauf soll sich das „Hochschulportal“ der FHD fokussieren (siehe [Knorz & Müller 2004]).

Es soll als internes Portal Funktionen anbieten, die ansonsten von sehr unterschiedlichen Systemen realisiert werden. Es ist gleichzeitig

- ein strukturiertes Glossar und beantwortet damit unmittelbar Fragen wie z.B. „was versteht man genau unter einer so genannten *gemeinsamen Professur*“ (siehe z.B. [www.wissen.de](http://www.wissen.de), dort auch mit einer graphisch dargestellten assoziativen Komponente)
- Ort für die Dokumentation von Fakten, so dass sich recherchieren lässt: "die Berufungskommission für die gemeinsame Professor X wurde am 19.11.04 eingesetzt" (soweit diese nicht in dezidierten anderen Hochschul-Informationssystemen abgelegt sind, wie z.B. in SAP/R3 oder HIS/SOS)
- ein Dokumentenverwaltungssystem, in den z.B. der Kooperationsvertrag für eine spezifische gemeinsame Professur in ihrem thematischen Kontext abgelegt wird.

Über das bisher Realisierte hinaus soll das Hochschulportal auch ein System zur Definition und Abarbeitung von Workflows/Guidelines werden und alle Daten und Online-Funktionen der Hochschule integrieren und individualisieren, um so Transparenz, Effektivität und Effizienz zu schaffen.

## 2.2 Anspruch des Portals

Aus wissenschaftlicher Perspektive soll das ontologiebasierte Hochschulportal nachweisen, dass beide Ziele, die breite Abdeckung eines Gegenstandsbereichs als auch die differenzierte semantische Modellierung auf der Grundlage einer Ontologie, gleichzeitig erreichbar sind. Insbesondere wird das Portal die in vielen konventionellen Systemen, insbesondere Dokumentenmanagementsystemen übliche Trennung von Objekten (Dokumente, Personen, etc) und Themen überwinden: Es gibt keine Dinge, die von ihrer Natur her Themen sind, sondern (detailliert modellierte) Begriffe bzw. Gegenstände/Individuen, die von Objekten (z.B. Veranstaltungen oder Dokumenten) zum Thema gemacht werden. Das ontologie-basierte Hochschulportal soll weiterhin nachweisen, dass das Informationsmodell für eine unmittelbare Anwendung tauglich ist und nicht nur akademisches Interesse auf sich zieht.

Das Portal besteht aus einer Web-basierten Schnittstelle zu der modellierten Hochschulwelt, mit eigenen Funktionalitäten ([Müller 2003, Knorz & Müller 2004]), Neben der Suche und einer strukturierten Informationsanzeige von Objekten gibt es insbesondere den grafisch orientierten „Net Navigator“ zur komfortablen Navigation im Netz der Begriffe und der Informationsobjekte (Instanzen), vergleichbar etwa der graphischen Darstellung der Thesaurusstruktur in [Hauer 2005], S. 72.

Das Portal und insbesondere die zugrunde liegende Ontologie werden gegenwärtig benutzt, um die Arbeit des Vizepräsidenten der Fachhochschule Darmstadt zu dokumentieren und zu unterstützen. Sitzungen und Besprechungen, die dort behandelten Gegenstandsbereiche (von der Professorenbesoldung über die Hochschulgesetzgebung bis zu Themenbereichen wie Fundraising oder IT-Sicherheit), die internen oder externen Beteiligten (Personen, Institutionen) sowie die relevanten Dokumente (Protokolle, Gesetze, Beschlüsse, Präsentationen, Berichte etc) und Objekte (z.B. Spenden, Server, Services, Firewalls) bilden den Schwerpunkt. Die Implementierung der in [Rein 2005} entwickelten und hier beschriebenen semantischen Suche hat den praktischen Wert des Prototypen deutlich gesteigert.

Die Ontologie umfasst gegenwärtig (Stand April 2005) über 2000 Begriffe (von Abgeordneter bis zweckgebundene Finanzierung) und über 4000 Individuen, davon 700 Dokumente verschiedenster Art.

Parallel zur praktischen Nutzung wird gegenwärtig daran gearbeitet,

- das semantische Netz zu konsolidieren und die Modellierung zu homogenisieren (da sich verschiedene Modellierungsprinzipien im Verlauf der Entwicklungszeit erst herausgebildet haben)
- eine Applikation zu entwickeln, die für einen spezifischen Nutzerkreis auf einem Ausschnitt des Netzes definiert wird.

Für die weitergehende und intendierte Nutzung des Portals müssen insbesondere

- das Rechtesystem konzipiert und realisiert werden
- formulargestützte, individualisierte Eingabemöglichkeiten entwickelt werden, denn gegenwärtig kommt ausschließlich der graphische Systemeditor zum Einsatz.

## 2.3 Das semantische Netz

Die Ontologie wird mit der Software Suite K-Infinity von intelligent views entwickelt, die den Aufbau, die Pflege und die Nutzung von „Wissensnetzen“ unterstützt [i-views 2001]. An einem Beispiel sollen die grundlegenden Bestandteile eines Wissensnetzes von K-Infinity erläutert werden.

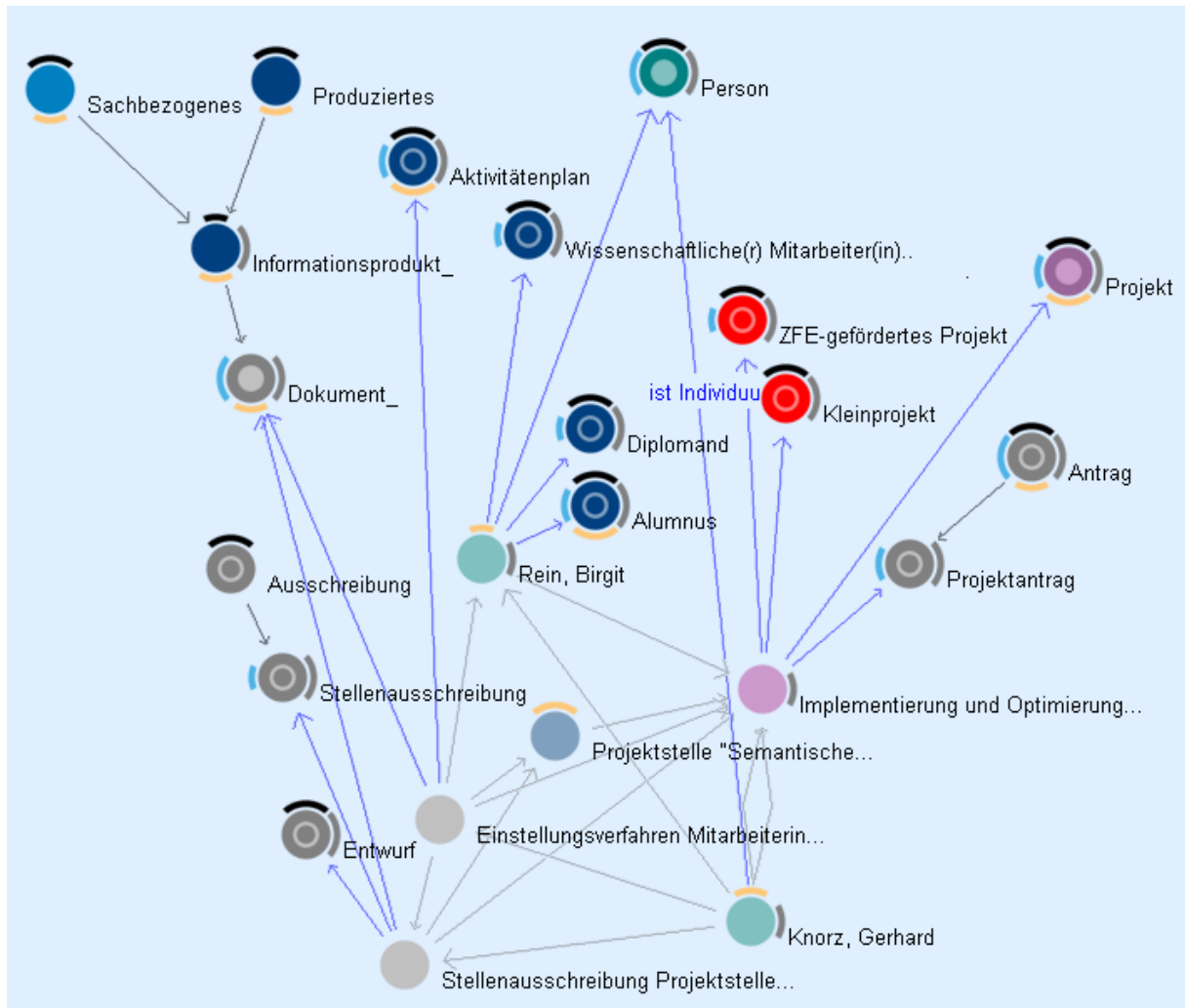


Abb. 1 Begriffe, Erweiterungsbegriffe, Individuen und Relationen, dargestellt im grafischen Editor des Knowledge Builders. Die partiell vorhandenen „Hüllen“ um Begriffe und Individuen werden hier nicht erläutert. Sie zeigen nur an, dass im Modell vorhandene Relationen aktuell nicht angezeigt werden.

Die Ontologie besteht zunächst aus Begriffen (in Abb. 1 dargestellt durch dunkle Kreisflächen mit hellem Kern bzw. Innenkreis), die durch Begriffsattribute (insbesondere Name, Synonym, Begriffsdefinition etc) näher beschrieben werden. Darüber hinaus ist definiert, durch welche Attribute Instanzen (Individuen) dieser Begriffe beschrieben werden. Instanzen werden vom Knowledge Builder durch helle Kreisflächen dargestellt und durch blaue Linien mit den Begriffen verbunden, deren Konkretisierung sie sind.

*Instanzen des Begriffs Person haben beispielsweise die Attribute Nachname, Vorname, Anrede etc., definiert beim Begriff Person*

Begriffe werden untereinander zunächst durch die Abstraktionsrelation (eine vordefinierte Systemrelation von K-Infinity) in Beziehung gesetzt.

*In Abb. 1 finden sich u.a. die Beispiele Ausschreibung – Stellenausschreibung sowie Antrag – Projektantrag.*

In dieser (Poly-)Hierarchie werden Attribute für Begriffe und Individuen (und Relationen, nachfolgend erläutert) vererbt.

*Beispielsweise ist für Instanzen vom Typ „Gegenständliches“ das Attribut Ansichtsbild definiert (weil von allem, was gegenständlich ist, ein Bild hergestellt werden kann), das sich auch auf Instanzen vom Typ Person vererbt, da Person über verschiedene Zwischenstufen Unterbegriff von Gegenständliches ist*

Alle Begriffe sind letztlich Unterbegriffe von T, der Wurzel der Begriffshierarchie.

Abbildung 1 zeigt, dass es offensichtlich unterschiedliche Arten von Begriffen gibt:

- Begriffe mit der Eigenschaft „individuenfähig“ (gekennzeichnet durch helle innere Kreisfläche) können eigenständige Instanzen bilden. *In Abb. 1 sind dies z.B. „Einstellungsverfahren Mitarbeiterin...“ und „Stellenausschreibung Projektstelle...“ als Instanz von „Dokument“, „Rein, Birgit“ und „Knorz, Gerhard“ als Instanz von „Person“ oder „Implementierung und Optimierung...“ als Instanz von „Projekt“. Individuen sind stets Instanzen von nur genau einem individuenfähigen Begriff*
- Begriffe mit der Eigenschaft „Erweiterungsbegriffe“ (gekennzeichnet durch helle innere Kreisringe) können nur bereits vorhandene Individuen „erweitern“, ihnen also zusätzlich weitere Attribute und Relationen (siehe nachfolgend) vererben. *So z.B. die Begriffe „Alumnus“ und „wissenschaftliche Mitarbeiterin“ der „Person Rein, Birgit“ oder „Stellenausschreibung“ und „Entwurf“ dem Dokument „Stellenausschreibung Projektstelle...“*

Ein Begriff muss nicht individuenfähig oder aber erweiternd sein (Beispiel: Sachbezogenes, Produziertes, Informationsprodukt) – allerdings schließen sich beide Eigenschaften gegenseitig aus. Zusätzlich wird die Begriffseigenschaft vererbt (kein Unterbegriff eines individuenfähigen Begriffs kann erweitern) und Erweiterungen können nicht „verschachtelt“ werden mit der Konsequenz, dass ein Begriff kein Erweiterungsbegriff sein kann, wenn ein Oberbegriff von ihm bereits erweitert werden kann.

*Beispiel: Wenn Gegenstand durch Produkt erweitert werden kann, dann kann ein Unterbegriff von Gegenstand wie z.B. Werkzeug kein Erweiterungsbegriff mehr werden, weil ansonsten der Erweiterungsbegriff Produkt Werkzeug erweitern könnte und Werkzeug wiederum ein weiteres Objekt. Damit käme es die Erweiterung einer Erweiterung, was auszuschließen ist.*

Das proprietäre Konzept der Erweiterungsbegriffe in K-Infinity bietet einiges an Möglichkeiten, besitzt aber subtile Konsequenzen und stellt viele Entwurfsfragen, die hier nicht behandelt werden können. Typischerweise bieten sich folgende Verwendungsmöglichkeiten an: Erweiterungsbegriffe

- modellieren Rollen („Diplomand“ als Rolle einer „Person“)
- schaffen Subkategorisierungen, spezielle Kontexte, die Informationen zusammenhängend bündeln, z.B. um mehrstellige Relationen auszudrücken
- erlauben es, auszudrücken, dass es sich bei einem Projekt um ein Kleinprojekt und um ein ZFE-gefördertes Projekt handelt, ohne dass man gezwungen wäre, den Begriff eines ZFE-geförderten Kleinprojektes explizit zu schaffen.

Die Begriffe und Individuen werden durch Relationen miteinander in Beziehung gesetzt und damit zum semantischen Netz. Neben den bereits (z.T. implizit) erwähnten und vom System vorgegebenen Relationen

- *ist Oberbegriff von/ist Unterbegriff von*
- *Ist Individuum von/hat Individuum und*
- *Wird erweitert durch/erweitert*

können im Modell eigene Relationen definiert und verwendet werden („benutzerdefinierte Relationen“)

In Abb. 1 finden sich Beispiele solcher Relationen:

- *„Rein, Birgit“ „arbeitet in“ Projekt „Implementierung und Optimierung der semantischen Suche im FHD-Portal“*
- *„Einstellungsverfahren Mitarbeiterin für Projekt <Semantische Suche>“ beschäftigt sich mit „Rein, Birgit“*
- *„Knorz, Gerhard“ leitet Projekt „Implementierung und Optimierung der semantischen Suche im FHD-Portal“*

Die Menge von ca. 140 inhaltlich verschiedener Relationen in der Domainontologie „Hochschulwelt“ wird einerseits vervielfacht durch die Tatsache, dass fast jede Relation in allen 4 Ausprägungen existiert:

- *Relation zwischen Individuen (z.B. „Rein, Birgit“ „arbeitet in“ Projekt „Implementierung und ...“)*
- *Relation zwischen Begriffen (z.B. „schulische Ausbildung“ ist Vorgänger von „berufliche Ausbildung“)*
- *Relation zwischen Begriffen und Individuen (z.B. „SAP-Modul“ ist Teil von „SAP/R3“)*
- *Relation zwischen Individuen und Begriffen (z.B. „Professur Retrieval Theorie und Praxis sowie Informationsdienstleistungen (IuW)“ beschäftigt sich mit „Fachinformation“)*

andererseits handhabbar gemacht dadurch, dass auch Relationen hierarchisch strukturiert werden können:

*„liegt in“ ist Unterrelation von „ist lokalisiert bei“ ist Unterrelation von „is-part-of“*

was bei Suchen und bei der Definition der nachfolgend beschriebenen Abkürzungsrelationen ausgenutzt werden kann:

*„is-part-of“ subsumiert „ist lokalisiert bei“ und dieses subsumiert „liegt in“*

Im Beispiel handelt es sich bei *„is-part-of“* um eine abstrakte Relation (per Konvention kenntlich an der englischen Bezeichnung). Auf weitere Gestaltungsmöglichkeiten kann hier nicht eingegangen werden.

Analog zu den bekannten View-Definitionen in relationalen Datenbanken lassen sich aus bekannten Relationierungen neue Zusammenhänge herstellen, indem Ketten von Relationen gebildet werden. Durch solchermaßen definierte „virtuelle“ Relationen (in K-Infinity als Abkürzungsrelationen bezeichnet) ist es möglich Inferenzen auszudrücken:

*<arbeitet in> definiert über <besetzt Stelle> <ist Teilorganisation von>  
Wenn P eine Stelle besetzt, die Teilorganisation von O ist, dann arbeitet P in O*

*<Ist zuständig für> definiert als <leitet> <ist zuständig für>  
wenn P etwas leitet, das für X zuständig ist, dann ist auch P für X zuständig*

Von solchen Abkürzungsrelationen macht die semantische Suche zentralen Gebrauch.

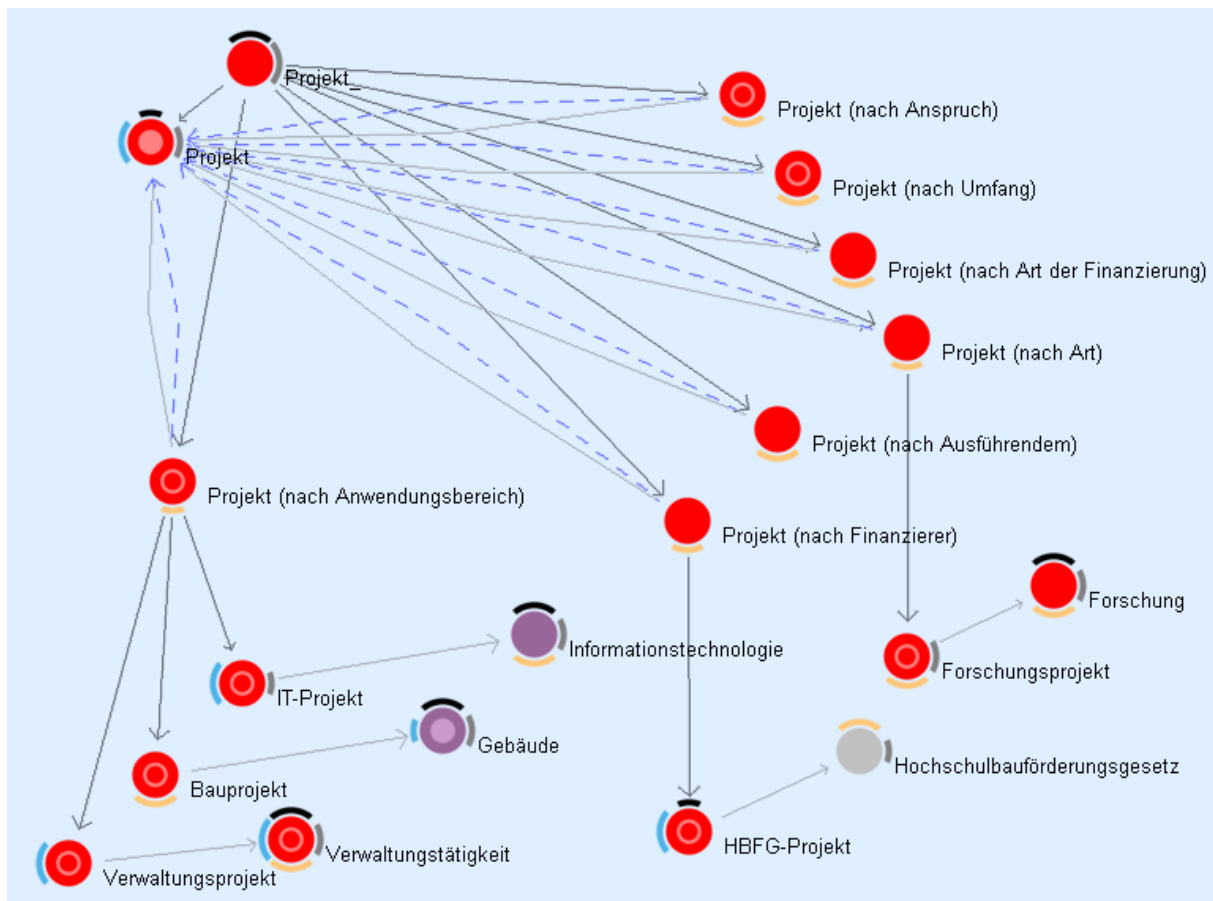


Abb. 2 Handhabung von Polydimensionalität bei der Modellierung von Begriffshierarchien. Vertikale Linien stellen die Abstraktionsrelation dar, die gestrichelten Linien stehen für „Individuen des Begriffs werden erweitert durch“

Beim Aufbau einer umfangreichen und differenzierten Modellwelt muss das Problem von Polydimensionalität gelöst werden, wenn das Modell überschaubar und handhabbar bleiben soll. Es gibt beispielsweise unzählige Unterbegriffe von Projekt, die in einem Informationssystem auch präsent sein müssen, die aber keinesfalls unstrukturiert nebeneinander gestellt werden dürfen.

*IT-Projekt, DFG-Projekt, Fachhochschulprojekt, Forschungsprojekt, Drittmittelprojekt, Anschubprojekt, Pilotprojekt sind Beispiele.*

Wären all diese Beispielbegriffe individuenfähig, so müsste für ein Individuum, bei dem es sich um ein DGP-Gefördertes Drittmittel-Forschungsprojekt an einer FH im Ist-Bereich ein entsprechender Begriff als Unterbegriff von IT-Projekt, DFG-Projekt, Fachhochschulprojekt, Forschungsprojekt, Drittmittelprojekt gebildet werden, da Individuen stets Individuen nur genau eines Begriffes sein können.

Ohne das Problem hier im Detail zu diskutieren illustriert Abb. 2 das entwickelte Modellierungsprinzip in der Domainontologie Hochschulwelt: der Oberbegriff Projekt wird aufgetrennt in den Begriff „Projekt\_“ (als Wurzel des Baumes in Abb. 2) dessen Unterbegriffe einerseits nach Unterteilungsgesichtspunkten differenzierte Teilhierarchien von Erweiterungs Begriffen sind, sowie zusätzlich der individuenfähige Unterbegriff „Projekt“.

Im oben genannten Beispiel wäre also ein Individuum von Projekt zu bilden, das anschließend durch die Begriffe IT-Projekt, DFG-Projekt, Fachhochschulprojekt, Forschungsprojekt und Drittmittelprojekt erweitert wird und so auch unter diesen Begriffen aufgefunden wird.

Wenn Unterbegriffe gebildet werden, so lässt sich in der Regel angeben, was den Begriff zum Unterbegriff des Oberbegriffs macht.

*Beispiele aus Abb. 2: Informationstechnologie macht ein Projekt zum IT-Projekt, Forschung macht ein Projekt zum Forschungsprojekt. Das Hochschulbauförderungsgesetz macht ein Projekt zum HBFEG-Projekt.*

Abb. 2 illustriert, dass Begriffe der Domainontologie soweit möglich neben ihrem Oberbegriff, dem *genus proximum* (die nächstgelegene Gattung) ihren „Differenzbegriff“, also das *differentia specifica* (den arteigenen Unterschied) angeben [Rahmsdorf 1994a,b], [Knorz 2005]. Dies wird sich bei der Suche als wichtig erweisen, um beispielsweise aus der Hierarchie der Projekte auszubrechen“, um in die Hierarchie von Institutionen überzuwechseln (z.B. zur DFG)

### 3 Semantische Suche

#### 3.1 Verschiedene Formen der Suche in K-Infinity

Zu den wichtigsten Funktionalitäten der Software K-Infinity gehören verschiedene Suchmethoden, die es dem Benutzer erlauben, durch Fragen an das System gezielt auf Informationsobjekte der Wissensbasis zuzugreifen. Die verfügbaren Suchmethoden lassen sich in drei Familien einteilen:

- Direkte Suchen
- Semantische Suchen
- Expertensuchen

**Direkte Suchen** sind Stringsuchen mit einer zusätzlichen Option der Filterung nach Objekttypen. Direkte Suchen werden auch innerhalb der **Semantischen Suchen** eingesetzt. Die Direkten Suchen dienen hier als Einstieg in das Wissensnetz. Anschließend navigiert das System entlang konfigurierter Navigationspfade zu den eigentlichen Treffern, die einem ausgewählten Objekttypfilter entsprechen.

**Expertensuchen** sind logikbasierte Suchen auf dem Schema der Wissensbasis.

<i>Suche</i>	<i>Input</i>	<i>Output</i>
Direkte Suchen	String (ggf. mehrwortig)	Begriffe oder Individuen, deren „Namen“ <sup>1</sup> mit dem String „matchen“ und dem Objektfilter der Direkten Suche entsprechen
Semantische Suchen	String (ggf. mehrwortig)	Begriffe oder Individuen vorgegebenen Typs in der „Umgebung“ der durch die vorgeschaltete Direkte Suche gefundenen Objekte.
Expertensuchen	Strukturierte Anfrage, gestützt auf das Schema der Wissensbasis	Begriffe oder Individuen, die vorgegebene, dem Schema entsprechende Bedingungen exakt erfüllen

<sup>1</sup> Der Term „Name“ steht hier und im Folgenden für die Attributwerte der indexierten und in die Attributliste der Direkten Suche aufgenommenen Attribute. Die Objektnamen, auch ihre Synonyme, gehören in der Regel dazu.

## 3.2 Grundprinzip der Semantischen Suche

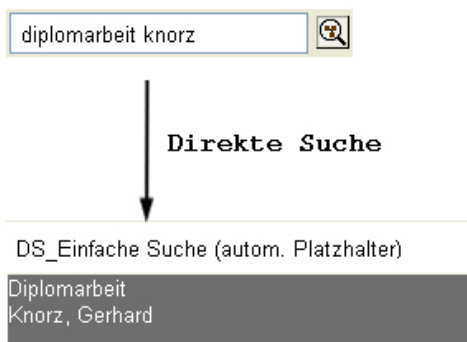
Das Verfahren der Semantischen Suche in K-Infinity besteht aus drei vom Systemadministrator konfigurierbaren Hauptteilen:

- (1) der Direkten Suche,
- (2) der Relationenliste und
- (3) dem Objektfilter.

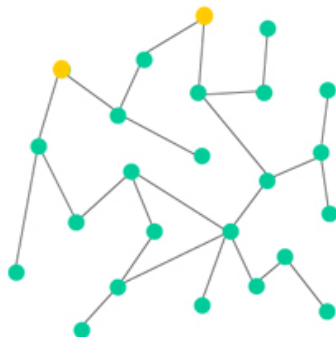
Die so genannte Direkte Suche (1) dient als "Einsprung" ins Netz. Zur Auswahl bei der Festlegung der Direkten Suche stehen die *Einfache Suche* mit verschiedenen Optionen, den Suchstring durch beliebige Zeichenfolgen zu ergänzen (wildcard \*), ferner *Wortsuche*, *Trigrammsuche*, *Volltextsuche* und *Kombinierte Suche*. Die Direkte Suche führt nach dem Verfahren der zuvor ausgewählten Suchvariante zu einem oder mehreren Informationsobjekten (siehe Abb. 3a). Im Anschluss daran erfolgt aufgrund der Relationenliste (2) die Navigation von den Treffern der Direkten Suche zu den potentiellen Zielobjekten (Abb. 3b). Die Treffer der Direkten Suche werden "Kombinationselemente" genannt, weil bei Mehrbegriffssuchen durch Kombination verschiedener Umgebungen gemeinsame Umgebungen dieser Elemente bestimmt werden. Durch den Objekttypfilter (3) werden aus diesen Umgebungsmengen die so genannten indirekten Treffer (also die eigentlichen Suchergebnisse) der Semantischen Suche selektiert und den Kombinationselementen zugeordnet dargestellt (Abb. 3c).

### Schema des Verlaufs der Semantischen Suche in K-Infinity

Start: Sucheingabe



„Netzeinsprung“



Ergebnis Direkte Suche

Abb.3a Die Direkte Suche führt im ersten Schritt zu den Objekten „Diplomarbeit“ und „Knorz, Gerhard“ (im Wissensnetz gelb dargestellt)

## Ergebnis Direkte Suche

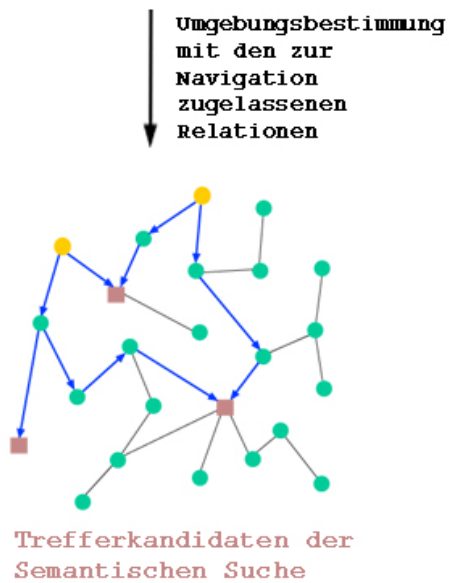
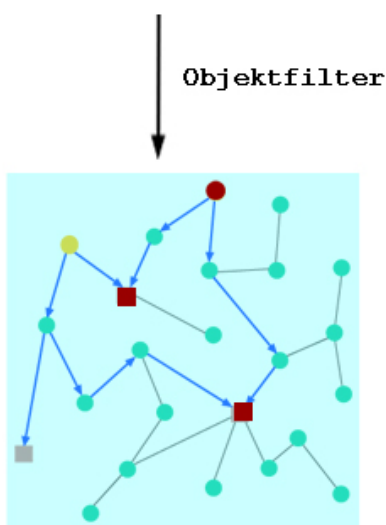


Abb.3b Von den Ergebnissen der direkten Suche werden über Relationenpfade Trefferkandidaten ermittelt (rote Quadrate).

## Trefferkandidaten der Semantischen Suche



## Treffer der Semantischen Suche

Semantische Suche\_Personen

```
DIREKTE TREFFER
INDIVIDUEN VON "PERSON"
  Knorz, Gerhard
"DIPLOMARBEIT" + "KNORZ, GERHARD"
INDIVIDUEN VON "PERSON"
  Engelhardt, Christian
  Rein, Birgit
```

Abb.3c Die von allen Einstiegsknoten gemeinsam erreichbaren Knoten des „richtigen“ Typs bilden hier das Ergebnis der Suche. Als "DIREKTE TREFFER" erscheinen die Treffer der Direkten Suche, die bereits dem Objekttypfilter der Semantischen Suche entsprechen.

### 3.3 Entwicklung einer semantischen Suche für das Hochschulportal

Wie eine informelle Erhebung des Informationsbedarfs bestätigt hat, zielt der Großteil praktisch auftretender Suchanfragen nicht auf die begriffliche Wissensrepräsentation, sondern auf die zugehörigen konkreten Fakten und Objekte vom Typ Individuum. Gesucht wird beispielsweise ein bestimmtes Projekt, also eine Instanz des für Projekte zuständigen Begriffs, oder nach konkreten Örtlichkeiten als Instanzen des Begriffs Lokalität, usw. Die Entwicklung der Semantischen Suche für das Hochschulportal beschränkt sich deshalb ausschließlich auf die Suche nach Individuen [Rein 2005]. Die Semantische Suche erlaubt aber, dass die gesuchten Individuen beim Namen genannt werden, was zu Direkten Treffern führen kann, oder aber, dass sie begrifflich umschrieben werden. Dann werden die Treffer (Individuen) in ihrer Zuordnung zu den gefundenen Begriffen gezeigt.

#### 3.3.1 Problematik

Wünschenswert wäre die Konfiguration *einer* einzigen universellen Suche, welche auf unterschiedliche Anfragen die passenden Treffer liefert.

Um dies realisieren zu können, müsste die Suchfunktion die Fähigkeit besitzen, nach kompletten Aussagen im Sinne von Subjekt, Prädikat und Objekt zu suchen.

*Beispiel* : **Wer ist verantwortlich für die zentrale Firewall der FHD**

Die Semantische Suche in K-Infinity V.2.1 sucht jedoch nicht nach Relationen und damit gebildeten Aussagen, sondern nur nach Begriffen oder Individuen. Kein Bestandteil der Suchanfrage lässt sich etwa mit dem Namen einer semantischen Relation abgleichen, um diese zu selektieren. Will man nicht für jede Relation eine eigene Suche konfigurieren, so verbleibt zunächst nur der Weg, allgemeingültige Relationenpfade mit Hilfe der Abkürzungsrelationen für die Semantische Suche zu definieren, die auf sinnvolle Weise vom Netzeinsprung der Direkten Suche hin zu den gewünschten Treffern navigieren. Entscheidend für den Erfolg einer solchen Suche ist dann nicht nur die Auswahl der Navigationspfade, sondern auch der Einstieg ins Netz über die Direkte Suche und, wie wir im Anschluss sehen werden, auch die Funktion der Objektfilter.

#### 3.3.2 Konfiguration der Direkten Suche

Als Direkte Suche werden verschiedene Formen des Suchstringabgleichs („Matching“) mit den Objektnamen angeboten, die einen unterschiedlichen Grad der Übereinstimmung zwischen Suchstring und Objektname fordern. Zur näheren Auswahl im vorliegenden Anwendungskontext gehören die drei so genannten *Einfachen Suchen*, die sich durch systemseitig vorgegebene Wildcard-Optionen unterscheiden, und die *Wortsuche* (eine ausführliche Bewertung in [Rein 2005, S. 68 ff.]).

Treffer innerhalb der *Einfachen Suche* sind Objekte, deren Namen als Ganzes mit maximalen Teilzeichenfolgen des Suchstrings übereinstimmen. Verbleibende Wortfolgen in der Anfrage werden als neue Suchanfragen vom System gestartet. Die *Wortsuche* dagegen liefert keinen Abgleich mit dem Objektnamen als Ganzes, sondern nur mit Wortfolgen dieser Namen.

Die am wenigsten tolerante Direkte Suche, die *Einfache Suche (kein Platzhalter<sup>2</sup>)*, kann zwar zu sehr präzisen Netzeinstiegen führen, doch wäre die Auswahl einer solchen Suche nur sinnvoll, wenn der Nutzer die Benennungen der Objekte im Netz sehr gut kennt. Empirische Untersuchungen belegen hier tendenziell einen schlechten Recall.

---

<sup>2</sup> K-Infinity-Bezeichnung für Wildcard

Die toleranteste Direkte Suche hinsichtlich der Auswertung des Suchstrings, die *Einfache Suche (immer Platzhalter)*, welche den Suchstring stets links und rechts um eine Wildcard ergänzt, führt in der Regel zu einer Erhöhung der Treffermenge und damit zu einer deutlichen Verschlechterung der Precision.

*Beispielsweise liefert eine Sucheingabe „IT“ zum Auffinden von Objekten, die mit IT=Informationstechnologie zu tun haben, mit Einfacher Suche (immer Platzhalter) alle Objekte bei denen die Buchstabenfolge it im Attributwert der bezeichnenden Attribute vorkommt [Rein 2005].*

Die so genannte Wortsuche hat ihren Nachteil im Nichterkennen der Wortgrundformen von Komposita.

*Beispielsweise wird mit Eingabe des Suchstrings „Alumni“ das Objekt „Alumni-Arbeit“ als Treffer gefunden, aber nicht das Objekt „Alumniarbeit“ [Rein 2005].*

Aktuelle Verwendung für die Semantischen Suche im FHD-Portal findet die *Einfache Suche (automatische Platzhalter)*. Der Einsatz der Wildcard ist hierbei adaptiv: Wenn es keinen Treffer ohne Wildcard-Einsatz gibt, werden dem Suchstring erst in Suffix-, dann in Präfix- und schließlich in Suffix- und Präfixposition Wildcards angehängt.

### **3.3.3 Relationenpfade für eine allgemeine Semantische Suche**

Gegeben ist eine begriffliche Beschreibung und gesucht sind Individuen, die auf diese Beschreibung passen. In erster Linie führt der Weg zur Auffindung von Individuen über die taxonomische Hierarchie, also von Oberbegriffen durch die Hierarchie der Begriffe bis zu ihren Individuen. Je tiefer der Einstieg in die Hierarchie erfolgt, umso kleiner werden die Treffermengen, da die Begriffsumfänge (Extensionen) kleiner werden. Der Umfang des Suchraumes ist jedoch innerhalb einer Hierarchie sehr begrenzt. Werden Suchanfragen mit zwei oder mehreren Aspekten gestellt, die in verschiedenen Begriffshierarchien modelliert sind, können so keine gemeinsamen Treffer erzielt werden.

Benötigt wird also ein Übergang zwischen verschiedenen Begriffshierarchien, wie er im Thesaurus etwas durch die Related Term-Beziehung geleistet wird. Nur, dass diese Beziehung bei unreflektierter Nutzung die Precision unkontrolliert verschlechtern würde. Dieser negative Effekt tritt nicht (bzw. weitaus geringer) ein, wenn im Wissensnetz für die Traversierung definitionsgerecht die Differenzrelation (siehe Abb. 4) eingesetzt wird. Zur Suche wird also der inverse Weg der Differenzrelation genutzt, ausgehend vom Differenzbegriff/Differenzindividuum zum zusammengesetzten Begriff (oder Individuum) entsprechend der Navigationsbewegung vom Allgemeinen zum Speziellen.

*Auf diese Weise liefert eine Suchanfrage mit Kommission, Fachhochschule Darmstadt und Sprache das Ergebnis Senatskommission für Sprachen.*

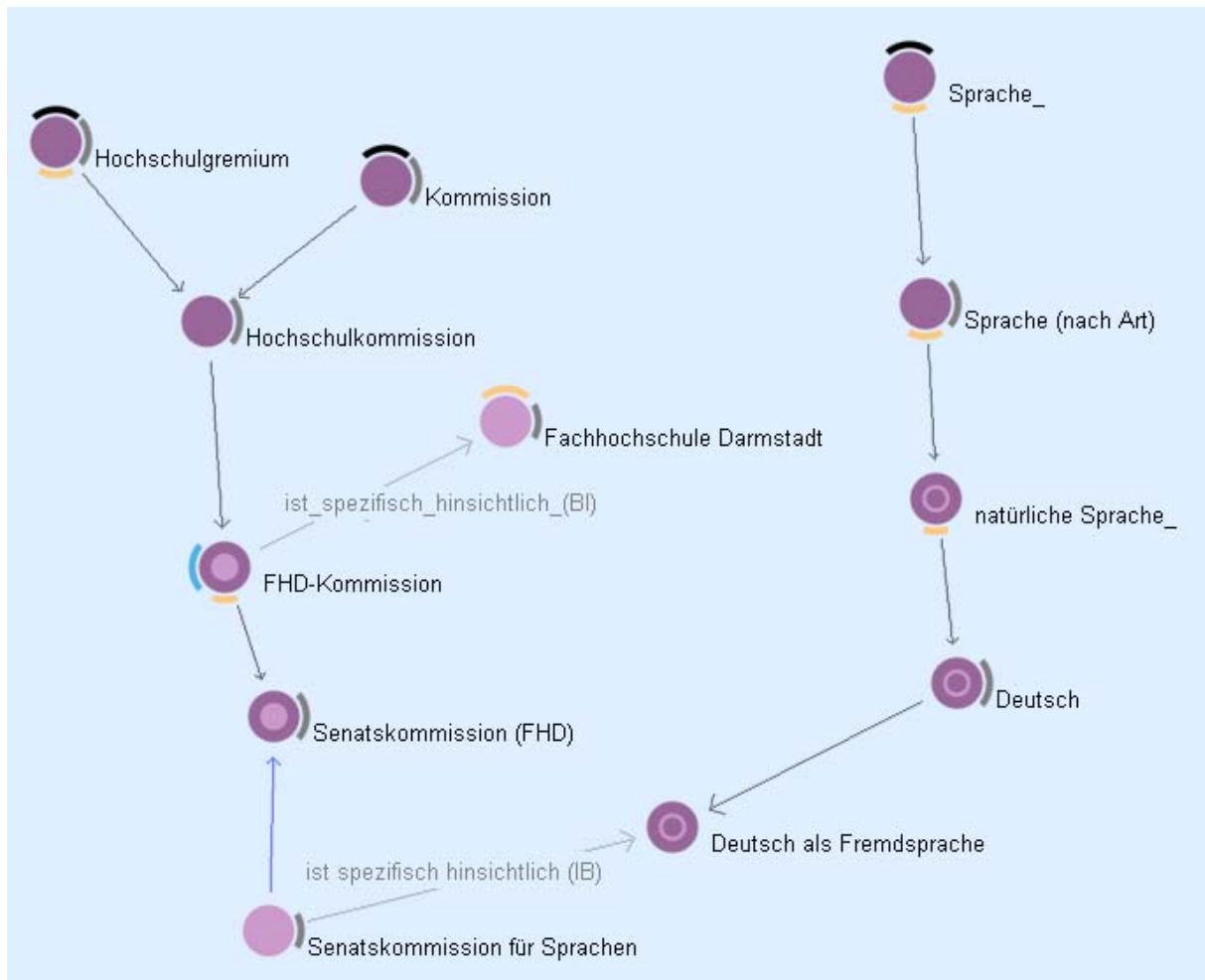


Abb.4 Zusammenführung verschiedener Hierarchien; Traversierung mittels Differenzrelationen [Rein]

Die Pfaddefinition, bestehend aus taxonomischen Relationen und Differenzrelationen, führt zu gewünschten Treffern, reicht jedoch für die Anforderungen an eine semantische Suche insgesamt nicht aus. Das Faktenwissen, also das Wissen von einzelfallspezifischen Zusammenhängen zwischen Individuen, kann so nicht erschlossen werden. Innerhalb der FHD-Domain-Ontologie wird dieses Wissen unter Nutzung einer eigenen Kategorie von Relationen, den benutzerdefinierten Relationen, modelliert. Eine Möglichkeit, dieses Wissen in all seinen Formen suchbar zu machen, besteht darin, die Oberrelation der benutzerdefinierten Relationen, die so genannte *Benutzerrelation*, als Bestandteil der Semantische Suche zu konfigurieren.

Durch die Einbeziehung der *Benutzerrelation* werden also unmittelbare, jedoch nicht weiter spezifizierte Nachbarschaften mit einer Kantenlänge von  $k=1$  in die Umgebungsdefinition auf Individuenebene und beim Übergang vom Begriff zum Individuum einbezogen. Das grundlegende Bauprinzip der so konfigurierten Navigationswege der Semantischen Suche durch das Netz ist hier in einem Schemadiagramm wiedergegeben:

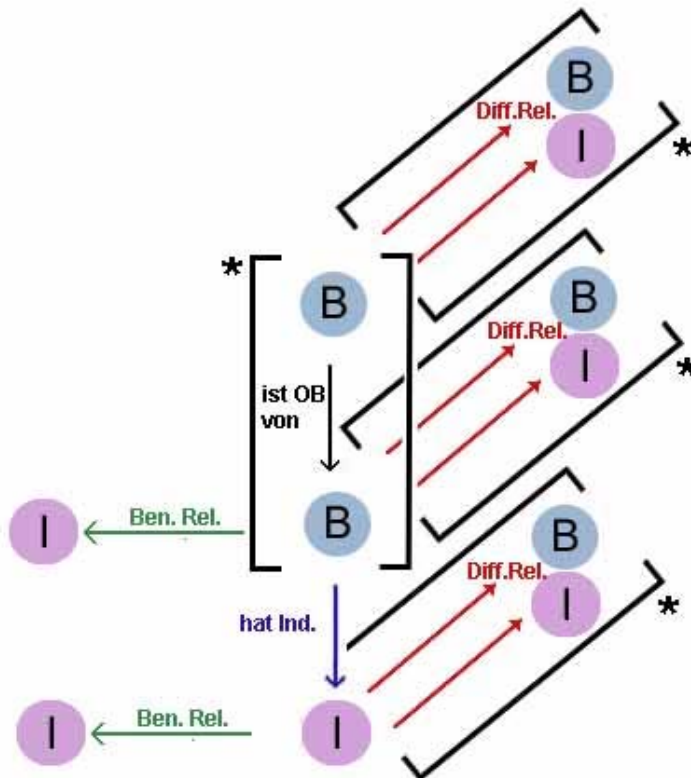


Abb.5 Konfigurierte Navigationswege der Semantischen Suchen. Der Stern-Operator steht für Wiederholungen (0,n), B und I stehen für Begriff und Individuen

Die mit Stern (\*) gekennzeichneten Pfadelemente sind optional bzw. können sich beliebig wiederholen. An jedem Knoten ist eine Traversierung der Hierarchien mittels Differenzrelationen möglich. Die Differenzbegriffe und Differenzindividuen (rechte "Säule" der B-/I-Knoten in der Abbildung) sind ihrerseits wieder Bestandteile weiterer zu durchlaufbarer Pfade.

### 3.3.4. Relationenketten mit Objekttyp-spezifischer Semantik

Die Verwendung der beschriebenen Navigationswege der Relationenliste liefert jedoch aufgrund der verallgemeinerten Berücksichtigung von Faktenrelationen (=Benutzerrelationen) einen sehr großen und damit unspezifischen Suchraum. Der Einsatz von einschränkenden Objektfiltern ist aus diesem Grund unbedingt notwendig, um die Treffermenge zu spezifizieren.

Die entscheidende Überlegung ist nun die, anstelle einer universellen Suche differenzierte Suchen entsprechend unterschiedlicher „Hauptsuchziele“ anzubieten, Diese "Hauptsuchziele" entsprechen dem "Head" der Nominalgruppe der Suchanfrage und müssen somit nicht mehr explizit im Suchstring genannt werden. Es werden im FHD-Portal folgende spezifische Semantische Suchen mit Objektfiltern nach folgenden Kategorien konfiguriert.

- Dokumente
- Personen
- Veranstaltungen
- Projekte
- Lokalitäten
- Organisatorische Einheiten

Es wird angenommen, dass dies die am häufigsten nachgefragten Objekttypen sind. Mit dieser Liste von Objekttypen, bzw. Objektkategorien ist zwar nach bisherigen Erfahrungen die Suche nach allen wichtigen Objekttypen abgedeckt, jedoch nicht die Suche nach allen Objekttypen. Für weitere Fälle steht deswegen eine breite Semantische Suche ohne Objekttypspezifizierung zur Verfügung.

Die spezialisierte Suche nach vorgegebenen Objekttypen legt es nahe, nicht nur die gefundenen Trefferkandidaten zu filtern sondern auch, in die Definition des Relationenfilters neben den allgemeingültigen Taxonomie und Differenzrelationen auch die adäquaten Benutzerrelationen einzubauen: Nach Dokumenten sucht man anders (also aufgrund anderer Relationen) als nach Personen oder Projekten.

### 3.3.5 Beispiel einer Semantischen Suche nach Dokumenten

Mit dem Suchstring *Information Retrieval Semantik Zeitschrift* wird mit einer Semantischen Suche für Dokumente gesucht und das Individuum „Semantische Suche in einer Hochschulontologie (Arbeitstitel)“ gefunden.

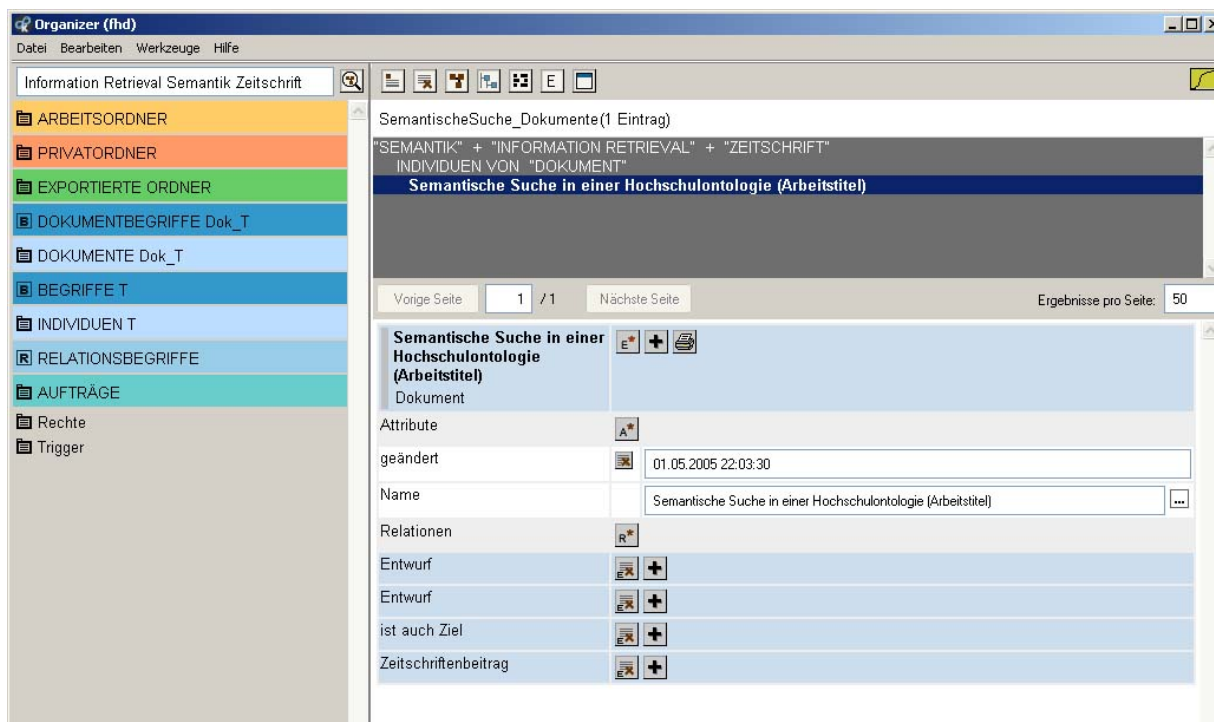


Abb.\* Darstellung der Suche mit Suchergebnissen im Organizer von K-Infinity

Gefunden wird der Treffer, weil er wie folgt erreichbar ist:

- von Information Retrieval aus durch die für Dokumente einschlägige Relation „ist Thema von/beschäftigt sich mit“
- von Semantik aus über die Kette von inverser Differenzrelation und der Relation „ist Thema von/beschäftigt sich mit“
- von Zeitschrift aus über die Kette von inverser Differenzrelation und die Relation „Ist Individuum von“

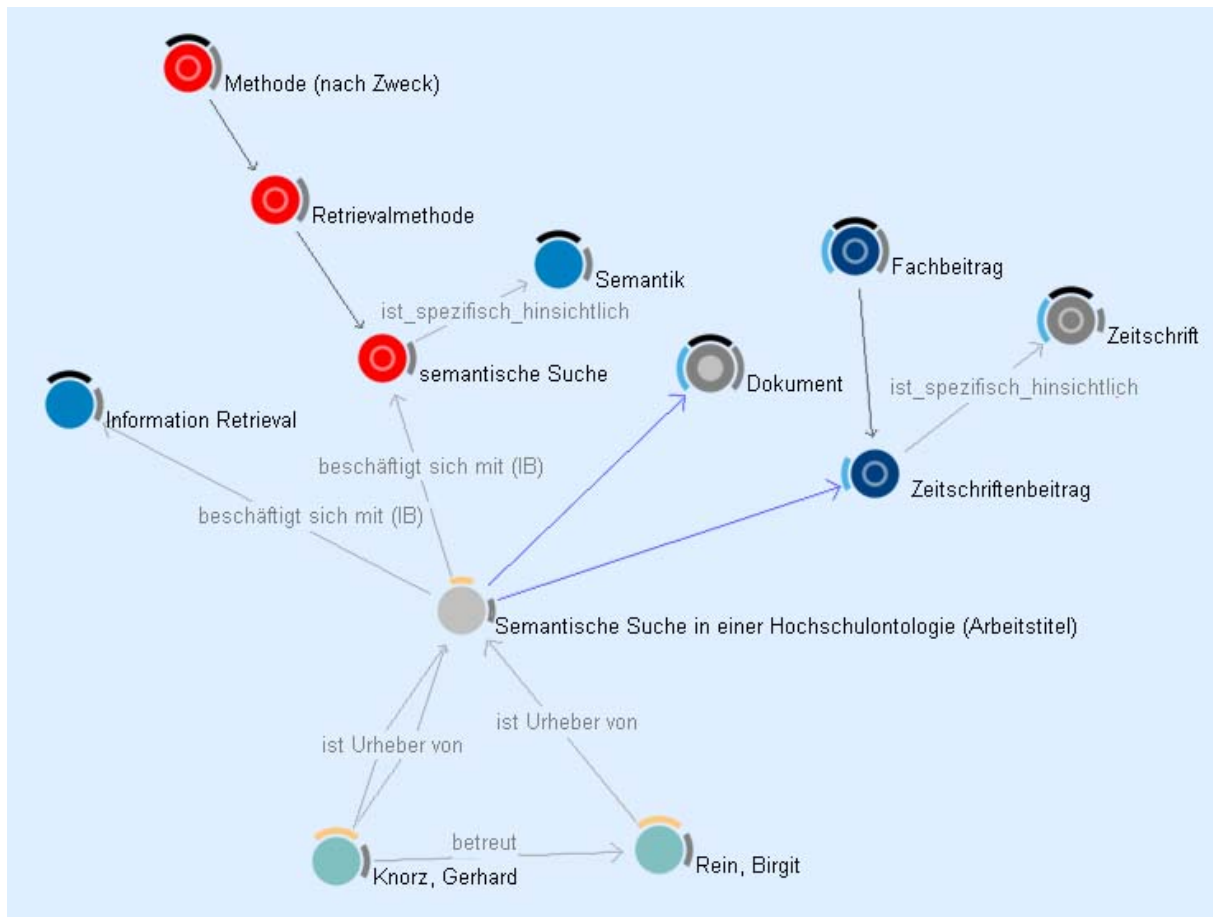


Abb.6 Abbildung des gefundenen Treffers im Graph-Editor von K-Infinity

Bedeutsam für den Erfolg der Semantischen Suche in K-Infinity 2.1 sind ferner weitere Aspekte, die zur Konfiguration der Suche in [Rein 2005] ausführlich untersucht und diskutiert werden:

- Vererbungsprinzipien im Detail zwischen Individuen und deren Erweiterungen (Rollen und Rollenträgern) innerhalb sowie am Ende von durch Abkürzungsrelationen definierten Relationenpfaden (siehe dazu die „Stellvertreterregel“ in [Rein 2005]).
- Details zu Abkürzungsrelationen im Relationenfilter mit inferierten Zusammenhängen
  - allgemeingültig
  - spezialisiert auf die vordefinierten Suchkategorien
- Simulation von Polyhierarchien bei den Relationen (bei denen systemseitig nur eine Monohierarchie möglich ist) durch „Einpacken“ in Abkürzungsrelationen und die sich damit eröffnenden Möglichkeiten für die Konfiguration des Relationenfilters.
- Konsequenzen der Auswahl der Direkten Suche im Hinblick auf die Bildung gemeinsamer Umgebungen durch den Relationenfilter.

## **4 Fazit und Ausblick**

Das in [Rein 2005] erarbeitete Konzept der Semantischen Suche ist inzwischen in den Kontext des operativen ontologiebasierten FHD-Portals übertragen worden und durchläuft derzeit weitere Phasen der experimentellen Verfeinerung.

Parallel dazu wird die Wissensbasis systematisch um bislang noch fehlende Differenzrelationen ergänzt. Gleichzeitig werden die inzwischen konsolidierten Modellierungsprinzipien durchgängig auf das gesamte Netz übertragen. Der operative Einsatz führt dazu, dass sowohl im begrifflichen Bereich und vermehrt noch im Bereich der Individuen die Informationsbasis schnell wächst.

Der subjektive Eindruck der Retrievalqualität ist sehr gut. Dies ermutigt dazu, die Leistungsfähigkeit der Semantischen Suche im Hochschulportal der FHD unter kontrollierten Bedingungen zu messen und zu bewerten. Dies steht unmittelbar an.

Verbesserungen und Weiterentwicklungen werden in folgender Hinsicht gesehen:

- Die bisher ausschließliche Orientierung an Fakten (Individuen) ist durch eine Suchvariante zu ergänzen, die auch Begriffe einschließt bzw. sich auf Begriffe als Treffer beschränkt
- Wenn auf dem Weg zum Rechercheergebnis Homonyme gefunden werden (Europa als Kontinent, Europa als Staatengemeinschaft), dann könnten die verschiedenen Interpretationsmöglichkeiten zur Auswahl angeboten werden. Per Checkbox kann der Nutzer dann entscheiden, welche Bedeutung(en) er in die Suche einschließen will.
- Bisher ist das Suchergebnis strukturiert nach Objekttypen bzw. nach Kombination von Objekttypen, unter denen es gefunden wurde. Einerseits ergeben sich dabei Redundanzen, wenn ein Objekt in der Ergebnisanzeige unter verschiedenen Rollen (technisch: Erweiterungsbegriffen, Oberbegriffen) auftaucht, die man versuchen könnte zu reduzieren. Andererseits wäre es eventuell sinnvoll, zusätzlich ein Ranking vorzunehmen.

## **5 Literatur**

- [Baumer & Knorz 2003] Baumer, Claudia & G. Knorz (2003): Vom Wissensnetz zum Wissensportal. Wissensmanagement 05(2003), S. 36, 37
- [Fischer 2005] Fischer, Dietrich H.: Semantische Suche in K-Infinity 2.1/2.2: Parameterbeschreibung und Konfigurationsszenarien, 28. Januar 2005, Fraunhofer Institut für integrierte Publikations- und Informationssysteme
- [Gruber 1995] Gruber, Thomas. R.: Toward principles for the design of ontologies used for knowledge sharing. In: , Guarino & Poli ,(Hr.:.): Formal Ontology in the Information Technology. Special issue of the International Journal of Human-Computer Studies, vol 43, no. 5/6, 1995  
[ftp://ftp.ksl.stanford.edu/pub/KSL\\_Reports/KSL-93-04.ps.gz](ftp://ftp.ksl.stanford.edu/pub/KSL_Reports/KSL-93-04.ps.gz)

- [Hauer 2005] Hauer, Manfred: Portal Informationswissenschaft. DGI baut Wissenschaftsportal mit AGI und Hochschulen. *Information, Wissenschaft & Praxis*, 56(2005)2, S. 71-76
- [i-views 2001] White Paper K-Infinity.  
<http://www.i-views.de/web/pdfs/Whitepaper-K-Infinity.pdf>  
 [5.5.05]
- [Knorz 2005] Knorz, Gerhard: Domain-Ontologie der Hochschulwelt für ein Portal. Zugangsvokabular und Basis für Inferenzen. Erscheint in: Eibl, M., Wolff, Ch, Womser-Hacker, Ch. (Hrsg.): *Designing Information Systems*. Universitätsverlag Konstanz.
- [Knorz & Müller 2004] Knorz, Gerhard, Joachim Müller: Wissensbasiertes Hochschulportal; *Information, Wissenschaft & Praxis* 55(2004)1, S. 35-41
- [Lehmann 2004] Lehmann, Christian: Begriffe – Definitionen; Institut für Sprachwissenschaft Universität Erfurt;  
[http://www.uni-erfurt.de/sprachwissenschaft/personal/lehmann/CL\\_Lehr/Begriffe/Begriffe\\_Definition.html](http://www.uni-erfurt.de/sprachwissenschaft/personal/lehmann/CL_Lehr/Begriffe/Begriffe_Definition.html) [5.5.05]
- [Müller 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/dateien/diplom.pdf> [5.5.05]
- [Rahmstorf 1994a] Rahmstorf, Gerhard: A New Thesaurus Structure for Semantic Retrieval, In: *Finding New Values and Uses of Information*. Proc. of the 47<sup>th</sup> FID General Assembly, Tokyo, Oct. 1994, 114-121
- [Rahmstorf 1994b] Rahmstorf, Gerhard. (1994): Semantisches Information Retrieval, S. 237-260 in: W. Neubauer, DGD Frankfurt (Hrsg.): *Tagungsband des Deutschen Dokumentartages 1994, Trier*.
- [Rein 2005] Rein, Birgit: Retrieval in semantischen Netzen, Konfiguration und Optimierung der Suchfunktion für ein internes Hochschul-Portal, Diplomarbeit am Fachbereich Informations- und Wissensmanagement der Fachhochschule Darmstadt, Januar 2005.
- [Tim Berners-Lee et al. 2001] Tim Berners-Lee, James Hendler, and Ora Lassila:  
 The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American.com*, Mai 2001,  
<http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21> [5.5.05]