

# SERGIO - Ein Ansatz zur wissenbasierten Informationsrecherche

Christoph Kunz

Fraunhofer Institut für Arbeitswirtschaft und Organisation  
Nobelstraße 12  
D-70569 Stuttgart  
christoph.kunz@iao.fraunhofer.de

**Abstract:** In diesem Beitrag wird ein nutzerzentrierter Ansatz beschrieben, um große und vernetzte Informationsräume semantisch erschließ- und recherchierbar zu machen. Den Anwendungskontext stellen spezialisierte Dokumentkollektionen wie Intranets, komplexe Portalseiten sowie Wissensdatenbanken dar. Der Ansatz verbindet eine herkömmliche Volltextsuche mit einer semantischen Suche auf Grundlage einer Themenontologie, welche einen Informationsraum abstrahiert und verdichtet. Die verschiedenen damit geschaffenen Arten des Informationszugriffs erfolgen über eine einheitliche Nutzerschnittstelle. Deren Hauptmerkmal ist die Visualisierung der Ontologie. Dies erlaubt die Bereitstellung von Stöbermechanismen auf Strukturebene, die Kontextualisierung von Volltextsuchergebnissen durch Begriffe der Ontologie, sowie die visuelle Konstruktion semantischer Abfragen zur präzisen Suche. Mit Hilfe dieser Mechanismen lässt sich ein Informationsbedarf iterativ präzisieren.

## 1 Einleitung

Dokumente im Internet als auch in betrieblichen Kontexten (Intranet, Wissensdatenbank) liegen selten in strukturierter Form vor und besitzen fast keine Metainformationen bezüglich ihres Inhalts. Schon die schiere Menge, aber auch das Nichtvorhandensein einer einheitlichen Zugangsstruktur erschwert sowohl deren menschliche als auch deren maschinelle Verarbeitung und Wahrnehmung, insbesondere das Entdecken von relevanter Information, deren Zusammenhang und deren Synthese zu einem umfassenden Gesamtbild.

Neben dem klassischen Information Retrieval (IR) [BYRN99] und navigatorischen Stöbermechanismen (Hypertext, Portalsysteme) erhält die Vision des *Semantischen Netzes* [BL98] derzeit viel Beachtung, um großen und vernetzten Informationsräume (Intranets, komplexe Portalseiten, Wissensdatenbanken) semantisch zu erschließen und recherchierbar zu machen. Bezogen auf IR werden hierbei Informationsräume mittels komplexerer Wissensstrukturen wie Semantischen Netzen oder Ontologien repräsentiert [FFR96]. Ziel dieser, aus dem Gebiet der Künstlichen Intelligenz stammenden Verfahren ist es, Wissen unabhängig von der sprachlichen "Oberflächenstruktur" zu modellieren, aus dem dann automatische Schlussfolgerungen abgeleitet werden können.

Die Gestaltung geeigneter Nutzerschnittstellen, welche die eigentlichen Vorteile erst nutzbar machen und den Informationszugriff ermöglichen, ergibt sich als Herausforderung aus der Komplexität semantischer Ansätze. Gestaltungsfragen betreffen die Darstellung der in einer Ontologie formalisierten Informationsstruktur (meist als Graphendarstellung [Fur86, LR94]) sowie einfache Interaktions- und Visualisierungstechniken zur Konstruktion logischer Abfragen [Jon98, CMF<sup>+</sup>03].

## 2 Ein Nutzerschnittstelle zur wissensbasierten Informationsrecherche

Die prototypische Umsetzung des in diesem Beitrag gewählten Ansatzes zur wissensbasierten Informationsrecherche zeigt Abbildung 1. Das Hauptmerkmal dabei ist die Visualisierung der semantischen Struktur des Informationsraums als ontologische Abstraktion und Verdichtung desselben. Im Minimalfall beinhaltet die semantische Struktur ein dem Informationsraum zugehöriges Themennetzwerk sowie bedeutungsvolle Merkmale von Dokumenten (Autor, Publikationsdatum, etc). Zusätzlich lassen sich in dieser Struktur weitere Arten von Inhalten wie zum Beispiel Personaldaten, Organigramme oder Produktdaten einbetten. Das Anwendungsfenster teilt sich in drei wesentliche Bereiche auf: Am

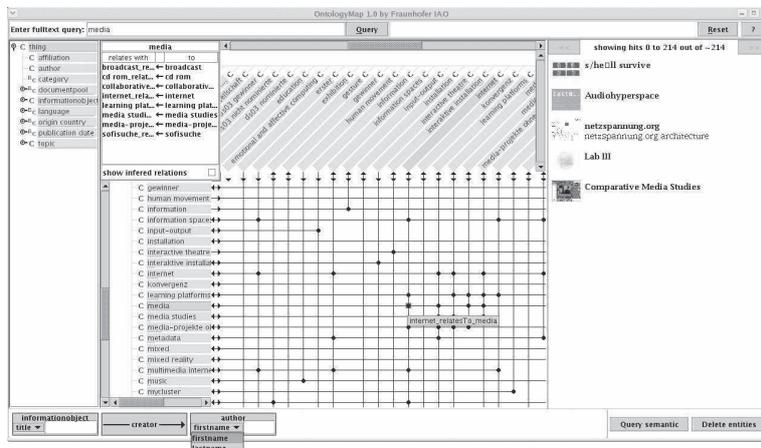


Abbildung 1: Schnittstelle zur Informationssuche

oberen Fensterrand befindet sich ein Texteingabefeld mit einem Anfrageknopf zur Formulierung und Durchführung von stichwortbasierten Suchanfragen. Der mittlere Bereich ist für die Visualisierung von Klassen der Ontologie (links) und für die Präsentation der Suchergebnisse (rechts) vorgesehen, während am unteren Rand des Anwendungsfensters ein Feld zur visuellen Konstruktion und Ausführung semantischer Anfragen platziert ist.

## 2.1 Visualisierung von Ontologien mit dem MatrixBrowser

Grundsätzlich lässt sich eine Ontologie als vernetzte Begriffsstruktur mit einem Graph darstellen. Dabei sind die einzelnen Begriffe die Knoten des Graphen, während die semantisch ausgezeichneten Relationen die Kanten des Graphen repräsentieren. Diese dem Informationsraum zugrundeliegende Ontologie wird mit Hilfe eines alternativen Graphvisualisierer, dem MatrixBrowser [ZKB02] dargestellt. Grundidee ist hierbei den Ontologiegraphen auf eine interaktive Adjazenzmatrixdarstellung abzubilden. Die Teile des Ontologiegraphen, welche durch Subklassenbeziehungen verbunden sind, werden dabei mittels Windows-Explorerähnlichen Baummetaphern auf den Achsen der Matrix präsentiert, während die restlichen Beziehungen die Zellen der Matrix bilden.

## 2.2 Kopplung von Matrix- und Listendarstellung

Innerhalb der Recherechnschnittstelle ist eine Trennung von Klassen- und Instanzebene vorgenommen: Während im MatrixBrowser die schematische Ebene dargestellt ist, werden in einer listenbasierten Präsentationsform nur Instanzen der Ontologie visualisiert. Die listenbasierte Form der Präsentation beruht auf einer Studie von Sebrechts et al [SCL<sup>+</sup>99], welche den tatsächlichen Wertvorteil unterschiedlicher Suchergebnispräsentationen untersuchte. In der Untersuchung stellte sich heraus, dass die allgemein gebräuchliche flache Textliste den anderen Techniken unter Berücksichtigung des Vorwissens und Erfahrung der Nutzer vorzuziehen ist.

Die auf Schemaebene der Ontologie operierende Matrixvisualisierung und als Liste präsentierte zugehörige Individuenmenge sind nach dem Prinzip der direkten Manipulation gekoppelt. Mit Hilfe einer speziellen Art der Knotenselektion, dem Doppelklick mit der Maus, werden die Individuen der korrespondierenden Klasse abgefragt und in der Liste angezeigt.

## 2.3 Kontextualisierung von Suchergebnissen

Die Kombination aus Strukturvisualisierung und Individuenmenge lässt sich für eine thematische Kontextualisierung von Ergebnissen einer Suchanfrage heranziehen. Dies ist besonders bei stichwortbasierten Suchanfragen vorteilhaft, da dort von der Volltextsuchmaschine, aufgrund von gleichen Stichwörtern, Dokumente unterschiedlicher Themenzugehörigkeit zurückgeliefert werden können.

Wird eine Stichwortsuche vom System evaluiert, werden die Ergebnisse der Suche in der Liste dargestellt. Für jedes einzelne Suchergebnis wird gleichzeitig, durch Anfrage der Wissensbasis, die Menge der Klassen ermittelt, in welchen das Suchergebnis eine Instanz darstellt. Für jede Klasse wird ein jeweiliger Zähler inkrementiert. Nach der Bearbeitung aller Suchergebnisse enthält so der Wurzelknoten des Ontologiegraphen die Anzahl aller Suchergebnisse. Die Knoten, welche den zugehörigen Klassen entsprechen, werden dann

in der Matrix mit einem Balken markiert (vgl. Abbildung 1 Matrixachsen). Die Länge des Balkens gibt das Verhältnis von den, in dieser Begriffsbeschreibung gefundenen Treffern, zu der Gesamtzahl aller Treffer an. Auf diese Weise erhält der Nutzer einen Hinweis, wie die erzielten Treffer einer Suche mit der Anfrage korrelieren.

In diesem interaktiven Zustand der Anwendung lässt sich dann die Menge aller Suchergebnisse über den Bezug zur Ontologie einschränken. Wird ein Knoten in der Matrix selektiert, so werden nur diejenigen Treffer in der Liste angezeigt, die Mitglied der Menge sind, die mit dem angewählten Knoten beschrieben werden.

## **2.4 Visuelle Konstruktion semantischer Suchanfragen**

Zur präzisen Suche unter Nutzung der Ontologie kommt der Hilfestellung bei der Konstruktion von logischen Abfragen eine besondere Bedeutung zu. Zur visuellen Konstruktion einer semantischen Anfrage ist eine spezielle Anfrageleiste mit der Matrix- und Listendarstellung so verbunden, dass sämtliche visualisierten Ontologieelemente auf dieser Leiste über den Mechanismus des Ziehen&Fallenlassens platziert werden können.

Das Beispiel in Abbildung 1 zeigt die visuelle Repräsentation der Anfrage "alle Informationsobjekte, die einen Autor besitzen". Die in der Anfrageleiste platzierten Ontologieelemente werden als Rechtecke dargestellt. Zusätzlich werden die jeweiligen lexikalischen Bezeichner der Ontologieelemente in einer ausgewählten Sprache angezeigt. Da Klassen Attribute besitzen können entsteht deren visuelle Repräsentation durch ein komplexeres Gebilde: Jedes Attribut bildet durch seinen lexikalischen Bezeichner einen Eintrag in einer Auswahlliste. Innerhalb eines korrespondierenden Texteingabefeldes kann der tatsächliche Wert des Attributs, wie im Beispiel der Nachname eines Autors, eingegeben werden. Dabei hängt der einzugebende Wert von dem jeweiligen Datentyp ab, der mit dem Attribut verknüpft ist.

Zur weiteren Unterscheidung der in der Abfrageleiste platzierten Ontologieelemente werden Instanzen nur durch ihren lexikalischen Bezeichner innerhalb des Rechtecks ausgewiesen, während Relationen zwischen Klassen durch einen zusätzlichen Pfeil visualisiert werden. Die Richtung des Pfeils zeigt dabei auf diejenige Klasse oder Instanz, welche zum Definitionsbereich der jeweiligen Relation gehört. Die Anordnung der Ontologieelemente ist dabei beliebig.

## **3 Zusammenfassung**

In diesem Beitrag wurde ein Ansatz zur wissensbasierten Informationsrecherche vorgestellt. Neben der explorativen Erschließung der Wissensbasis mittels des MatrixBrowsers und der damit verbundenen Liste zur Ergebnisdarstellung wird unter Verwendung der Gesamtheit der Interaktions- und Visualisierungstechniken ein Rechercheprozess in drei Schritten zur Präzisierung eines bestimmten Informationsbedarfs angeboten: Bei einer ungenauen Stichwortsuche wird nicht nur die reine Ergebnismenge als Liste dargestellt, son-

dem auch die gefundenen Dokumente den Begriffen in der Ontologie zugeordnet. Diese Begriffe werden markiert und kontextualisieren die Treffermenge. Im zweiten Schritt kann die Ergebnismenge dadurch weiter eingeschränkt werden, dass einzelne Begriffe auf den Matrixachsen selektiert werden. Im dritten Schritt können die in der Ontologie enthaltenen Metadaten und Themenbegriffe, sowie Relationen dazu verwendet werden, tatsächliche semantische Abfragen anhand der logischen Struktur zu generieren.

## 4 Literatur

### Literatur

- [BL98] T. Berners-Lee. Semantic Web Roadmap. Bericht, World Wide Web Consortium, 1998.
- [BYRN99] Ricardo A. Baeza-Yates und Berthier A. Ribeiro-Neto. *Modern Information Retrieval*. ACM Press / Addison-Wesley, 1999.
- [CMF<sup>+</sup>03] Tiziana Catarci, Tania Di Mascio, Enrico Franconi, Giuseppe Santucci und Sergio Tessaris. An Ontology Based Visual Tool for Query Formulation Support. In *Proceedings of OTM 2003, On The Move to Meaningful Internet Systems 2003*. Springer-Verlag Heidelberg, Oktober 2003.
- [FFR96] A. Farquhar, R. Fikes und J. Rice. The Ontolingua server: A tool for collaborative ontology construction. Bericht KSL 96-26, Stanford, 1996.
- [Fur86] G.W. Furnas. Generalized fisheye views. In *Proceedings of ACM CHI'86*, Seiten 16–23. ACM, ACM Press, 1986.
- [Jon98] Steve Jones. Graphical query specification and dynamic result previews for a digital library. In *Proceedings of the 11th annual ACM symposium on User interface software and technology*, Seiten 143–151. ACM Press, 1998.
- [LR94] J. Lamping und R. Rao. Laying Out and Visualizing Large Trees Using a Hyperbolic Space. In *Proceedings of the ACM Symposium on User Interface Software and Technology*, Seiten 13–14. ACM Press, 1994.
- [SCL<sup>+</sup>99] Marc M. Sebrects, John Cugini, Sharon J. Laskowski, Joanna Vasilakis und Michael S. Miller. Visualization of Search Results: A Comparative Evaluation of Text, 2D, and 3D Interfaces. In *Research and Development in Information Retrieval*, Seiten 3–10, 1999.
- [ZKB02] Jürgen Ziegler, Christoph Kunz und Veit Botsch. Matrix Browser: Visualisierung und Exploration vernetzter Informationsräume. In *Konferenzband der Mensch und Computer 02*, Hamburg, Germany, 2002. Gesellschaft für Informatik, Teubner Verlag.