



Whitepaper Series

Taxonomie, Thesaurus, Topic Map, Ontologie – ein Vergleich

V.1.4

Mike Ullrich, Andreas Maier, Prof. Dr. Jürgen Angele

ontoprise® GmbH

Amalienbadstraße 36 (Raumfabrik 29), D-76227 Karlsruhe

<http://www.ontoprise.de>

mailto: {mike.ullrich, maier, angele}@ontoprise.de

Inhalt

ABSTRACT	2
BESCHREIBUNG DER MODELLE	3
TAXONOMIE.....	3
THESAURUS.....	4
TOPIC MAP.....	4
ONTOLOGIE.....	6
FAZIT UND AUSBLICK	10
ÜBER ONTOPRISE	10

Abstract

Dieses Whitepaper hat das Ziel, die semantischen Ansätze Taxonomie, Thesaurus, Topic Map und Ontologie zu erläutern, Ihre jeweiligen Stärken, Schwächen und die jeweiligen Anwendungsfälle zu diskutieren und somit einen Beitrag zur laufenden Diskussion um semantische Modelle zu leisten.

Zusammenfassend kann gesagt werden, dass die Mächtigkeit der Ansätze in der Reihenfolge Taxonomie, Thesaurus, Topic Map, Ontologie zunimmt.

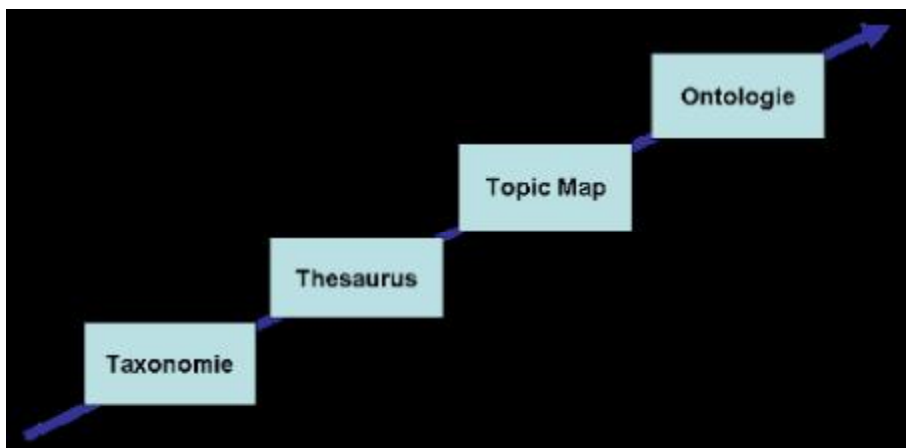


Abbildung 1: Evolution der semantischen Modelle

Beschreibung der Modelle

Am Beispiel von Doktoranden, Themen und Dokumenten wollen wir die Ausdrucksfähigkeit der verschiedenen Modelle erläutern. Wir haben dieses Beispiel gewählt, weil sich hieran einige konzeptionelle Unterschiede erläutern lassen.

Taxonomie

Die Taxonomie ist eine Hierarchie von Begriffen.¹

Bereits im Altertum haben Biologen versucht, Flora und Fauna zu ordnen, wobei die Eindeutigkeit der Zuordnung jedoch aufgrund verschiedener Ordnungsmerkmale (wie z.B. Lebensraum) nicht eindeutig ist.

Der Begriff Taxonomie wird in verschiedenen Wissenschaftsdisziplinen verwendet und ist somit nicht nur auf die Informationstechnologie beschränkt. Eine Standardisierung der Technologie ist den Autoren nicht bekannt. Eine weit verbreitete Anwendung, die Taxonomien verwendet, ist das Dateisystem in Windows mit seiner Ordnerstruktur. Ein neueres Beispiel stellt der XML-basierte Taxonomie-Editor der Firma Semansys dar.²

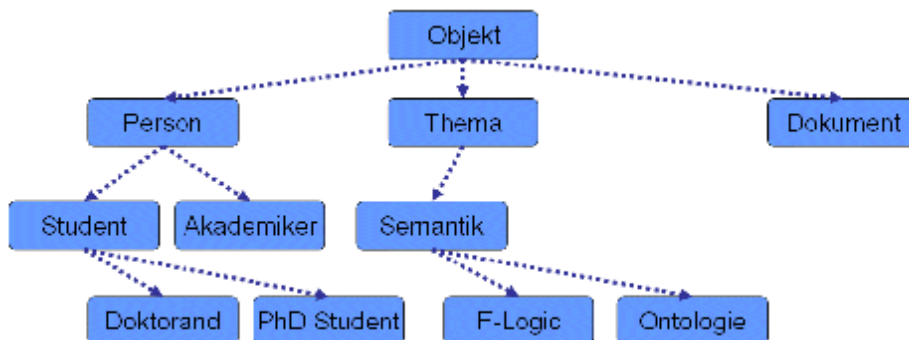


Abbildung 2: Beispiel einer Taxonomie

Unser Beispiel von Doktoranden, Themen und Dokumenten ist in Abbildung 2 als Taxonomie dargestellt.

Anzumerken ist, dass die Pfeile in unserem Schaubild keiner eindeutigen Bedeutung unterliegen. Beispielsweise ist „Semantik“ ein „Thema“ und

¹ vgl. Bruegge, Bernd: Einführung in die Informatik II, <http://www.bruegge.in.tum.de/teaching/ss01/Info2/vorlesung/fohlen>

² Semansys Technologies, Niederlande, <http://www.semansys.com/>

hat „Ontologie“ als Unterthema. Der Zusammenhang, dass „PhD Student“ das englische Äquivalent zu „Doktorand“ ist, lässt sich ebenfalls nicht ausdrücken. Themen, die unter zwei Oberthemen hängen müssen mehrfach angelegt werden, womit eine gewisse Redundanz entsteht.

Thesaurus

Der Thesaurus stammt aus dem Bibliothekarswesen und stellt eine Terminologie zu einer bestimmten Domäne dar. Neben der Hierarchie (vgl. Taxonomie) existieren zwei fest definierte Relationen der Objekte untereinander: die Ähnlichkeits- und die Synonymrelation. Weitere selbstdefinierte Relationen sind nicht möglich.

Die bekannteste Anwendung ist der in Microsoft Word integrierte Thesaurus für verschiedene Sprachen.

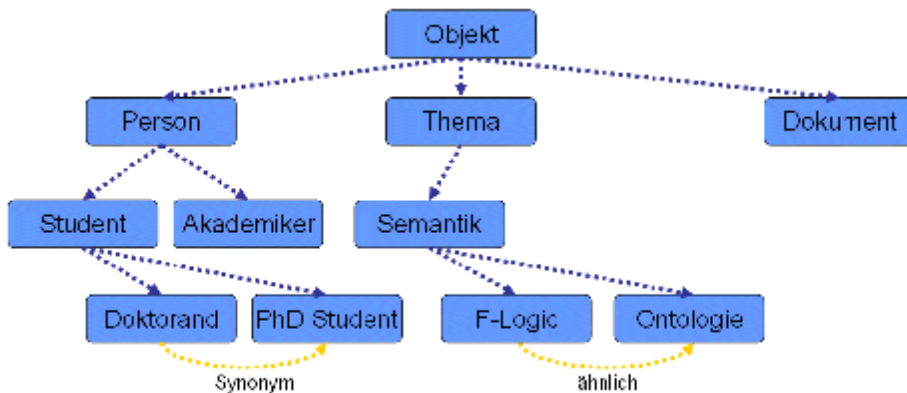


Abbildung 3: Beispiel eines Thesaurus

In unserem Beispiel (siehe Abbildung 3) können wir nun Zusatzinformationen hinzufügen. Es wird definiert, dass „PhD Student“ „Synonym“ von „Doktorand“ und „F-Logic“ „ähnlich“ zur „Ontologie“ sind.

Topic Map

Aus der hier betrachteten Reihe an Modellen ist die Topic Map das in der Literatur am eindeutigsten definierte Konzept, da es sich bei der Topic Map um einen ISO-Standard auf SGML Basis handelt. Allerdings ist die heute gängigere XML-Version auch ein offener Standard aber nicht nach ISO.

Die Topic Map besteht im Wesentlichen aus Subjects (abstrakte Dinge) Topics (deren adressierbare Repräsentation), Assoziationen, Scopes

(Gültigkeitsbereiche für Topics) und zugeordneten Dokumenten außerhalb der Topic Map (Occurrences).³

Topic Maps werden von zahlreichen Anbietern angeboten⁴. Die bekanntesten Anwendungen existieren im Bereich der Suchunterstützung, Visualisierung und Navigation durch Themengebiete.

Der Standard beschreibt lediglich den Aufbau der Topic Map. Er beinhaltet weder ein einheitliches Datenmodell noch eine Abfragesprache. Die Abfragesprache und viele Erweiterungen sind von den einzelnen Anbietern häufig über den Standard hinausgehend individuell realisiert worden.

Auf unser Beispiel angewendet haben wir jetzt neue Möglichkeiten. Wir können nun Assoziationen zwischen den Knoten selbst definieren: Eine „Person“ schreibt ein „Dokument“, ein „Thema“ wird behandelt in einem „Dokument“⁵. Unser „Doktorand“ kann nun mit Attributen versehen werden (Alter, Telefonnummer, ...) und kann auf externe Dokumente und Webseiten verweisen. Die Merkmale des Thesaurus bleiben erhalten (vgl. Abbildung 4).

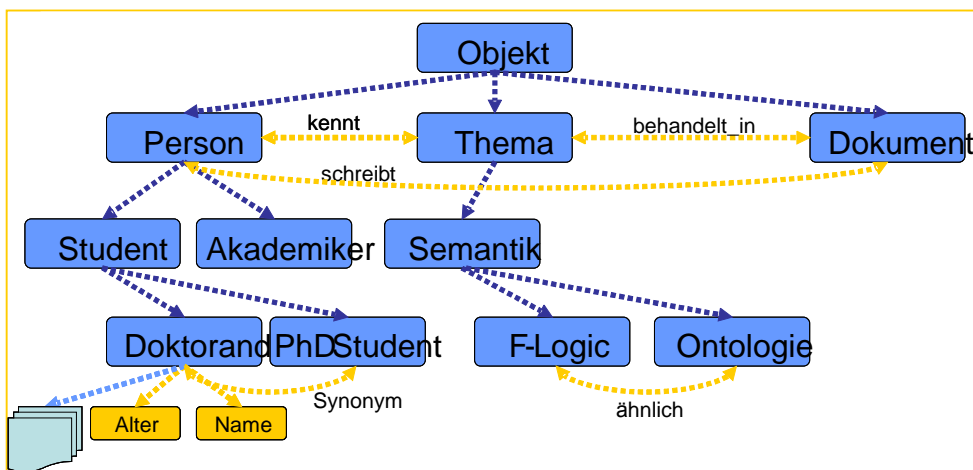


Abbildung 4: Beispiel einer Topic Map

³ vgl. Hofman, Tobias: Topic Maps, <http://weblogs.medien.uni-weimar.de/topicmaps/about>

⁴ vgl. <http://www.topicmaps.org>

⁵ wir haben diese vereinfachende Darstellung gewählt, um an dem selben Beispiel die Unterschiede aufzeigen zu können. Genauer betrachtet besteht eine Assoziation aus einem Identifier, mit verschiedenen Rollen, die auf Topics verweisen. Die Assoziation ist ungerichtet.

Ontologie

Ursprünglich kommt der Begriff Ontologie aus der Philosophie und bedeutet so viel wie die Lehre des Seins. Wir definieren Ontologie im Sinne wie folgt: „Eine Ontologie ist eine explizite Spezifikation einer (gemeinsamen) Konzeptionalisierung“⁶. Weiter einschränkend und in Analogie zu den Entwicklungen im Bereich der Semantic Web Initiative des World Wide Web Consortiums (W3C) betrachten wir im Folgenden ausschließlich Ontologien mit den Merkmalen (vgl. Abbildung 5):

- die Ontologie beinhaltet ein Datenmodell (z.B. XML(s))
- die Ontologie definiert Objekte und deren Relationen zueinander (z.B. RDF(s)=Ressource Description Framework⁷)
- Semantik und einfache Abhängigkeiten können mit Hilfe des Ontologie Vokabulars beschrieben werden (z.B. DAML+OIL⁸, OWL⁹)
- Logische Zusammenhänge können mittels Regeln ausgedrückt werden (z.B. F-Logic¹⁰)

Das W3C arbeitet zum Zeitpunkt der Veröffentlichung dieses Artikels an der Standardisierung des Semantic Webs. XML(s), RDF(s) sind bereits verabschiedet. DAML+OIL wurde parallel durch die DARPA vorgeschlagen und findet nun Eingang in die laufenden Standardisierungsarbeiten von OWL durch das W3C.

⁶ T. R. Gruber: A translation approach to portable ontology specifications; Knowledge System Laboratory, Stanford University 1993.

⁷ RDF=Ressource Description Framework, siehe: <http://www.w3.org/RDF>

⁸ DAML+OIL = DARPA Agent Markup Language + Ontology Internet Language, siehe <http://www.daml.org/>

⁹ OWL = Web Ontology Language, siehe <http://www.w3.org/2001/sw/WebOnt/>

¹⁰ F-Logic = Frame Logic, siehe <http://www.cs.sunysb.edu/~kifer/dood/papers.html>

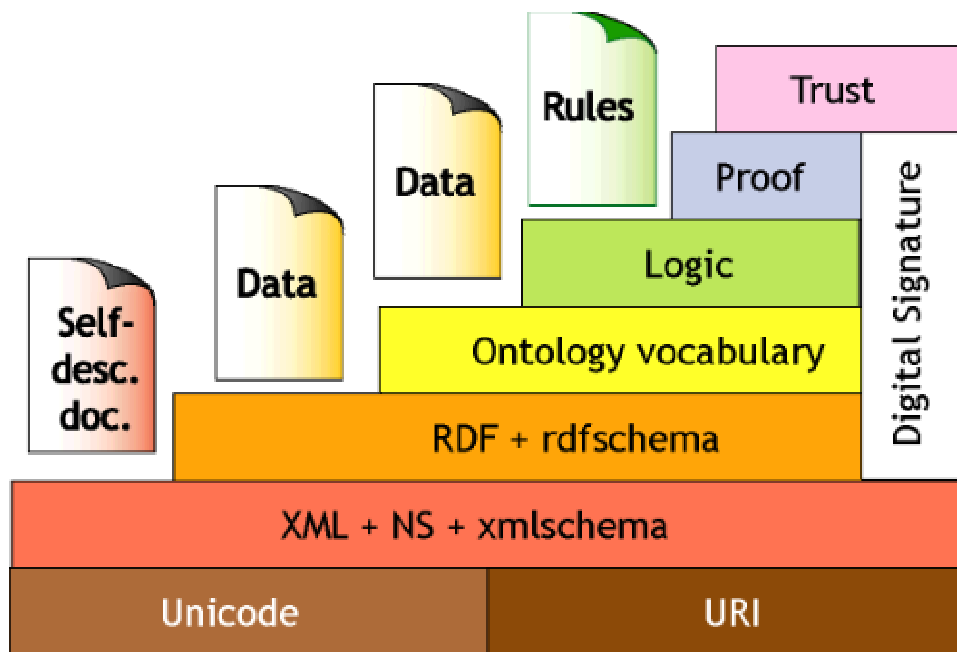


Abbildung 5: Schichten des Semantic Web nach T. Berners-Lee u.a.

Um eine Topic Map in eine Ontologie zu überführen, kann nun im Gegensatz zu den vorhergehenden Ansätzen klar unterschieden werden zwischen „is-a“ Hierarchie und „has-part“ Hierarchie. Wir erhalten eine eindeutige Vererbungshierarchie. Die Vererbung greift nur über die „is-a“ Hierarchie, während „has-part“ Beziehungen mittels Relationen dargestellt werden.

Die Ontologie bietet die Möglichkeit der Trennung zwischen Schema (Meta-Modell) und Inhalten, wie dieses aus dem Bereich der Entity Relationship-Modelle bekannt ist. Die Konzepte einer Ontologie definieren das Schema, während Instanzen, die jeweiligen Ausprägungen darstellen. Erst hierdurch wird performante Massendatenhaltung ermöglicht.

Die Ontologie verfügt über ein mächtiges Regelwerk, mit dem Zusammenhänge zwischen Objekten der Ontologie (und auch anderer Ontologien) mittels Wenn-Dann Beziehungen, Zuweisungen, Negationen, logischen Verknüpfungen, mathematischen Operationen und weiteren Funktionen ausgedrückt werden können.¹¹

Mittels der Abfragesprache kann die Ontologie in beliebige Richtungen angefragt werden. z.B. Was sind die Unterkonzepte von Person? Welche

¹¹ Zur genauen Übersicht des Sprachumfangs, vgl. M. Kifer, G. Lausen: F-Logic - A Higher-Order Language for Reasoning about Objects, Inheritance, and Scheme, SIGMOD RECORD, <http://citeseer.nj.nec.com/kifer90flogic.html>, 1990, Freiburg,

Unterthemen hat Semantik? Welche Personen über 30 Jahre kennen das Thema Ontologie?

Es lässt sich erkennen, dass es sich bei Thesauri und Taxonomien – unter Einhaltung bestimmter Konventionen – um einfache Ontologien handelt. Alle Eigenschaften der Topic Maps können in Ontologien abgebildet werden, lediglich das Konzept der Scopes wird in der Ontologie über den Umweg der Namespaces gelöst.

Zurück zu unserem Beispiel (vgl. Abbildung 6):

- Wir ordnen die Hierarchie neu: „F-Logic“ und „Ontologie“ sind nun Unterbegriffe des Konzepts „Thema“, die Abhängigkeiten untereinander werden mittels der Relation „Unterthema“ abgebildet.
- Da es sich bei „Doktorand“ und „PhD Student“ um das gleiche Konzept handelt, werden diese zusammengefasst und mit den Bedeutungen in Deutsch und Englisch hinterlegt (Mehrsprachfähigkeit von Ontologien)
- Der „Doktorand“ ist sowohl „Student“ als auch „Akademiker“, somit erbt das Konzept Doktorand die Eigenschaften von beiden. z.B. könnte er sowohl die Personal-Nummer vom Akademiker, als auch die Matrikel-Nummer vom Studenten erben.
- Tatsächliche Ausprägungen eines Konzeptes werden als Instanz angelegt. „York Sure“ ist Doktorand und hat spezifische Ausprägungen wie Alter, Name, ...
- Eine einfache Regel betrifft die Inversität von zwei Relationen: Wenn ein „Thema“ in einem „Dokument“ behandelt wird, ist das „Dokument“ zu dem Thema.
- Ein Beispiel für eine komplexere Regel auf hohem Abstraktionsniveau ist: Wenn eine „Person“ ein „Dokument“ zu einem „Thema“ schreibt, dann kennt die „Person“ auch implizit dieses „Thema“. Nun wird automatisch die Schlussfolgerung gezogen, dass eine bestimmte „Person“ ein „Thema“ kennt, ohne dass dieses explizit angegeben werden müsste.

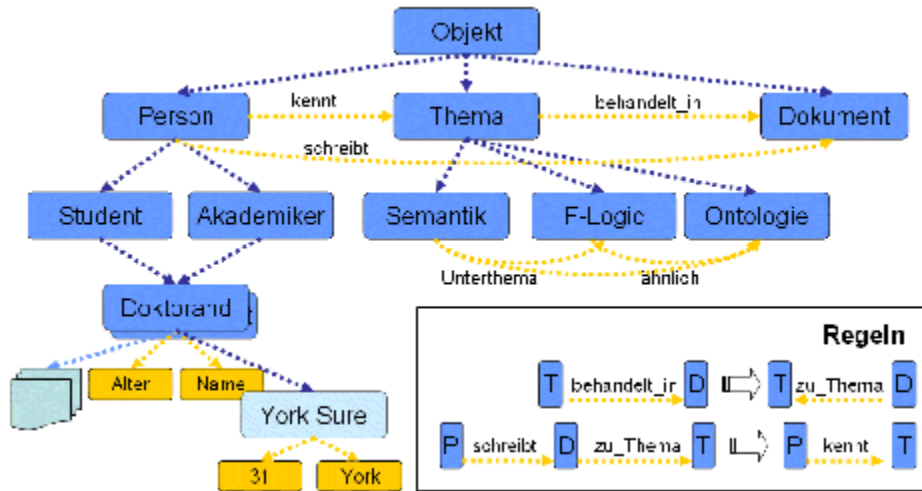


Abbildung 6: Beispiel einer Ontologie

Bevor wir zu den Anwendungen der Ontologie kommen, sei die semantische Integrationsfähigkeit von Ontologien erwähnt. Mittels Mapping können Daten aus Datenbanken, anderen Ontologien, etc. angebunden werden (vgl. Abbildung 7). Die Subsysteme werden dann bei Bedarf angefragt (z.B. via SQL).¹²

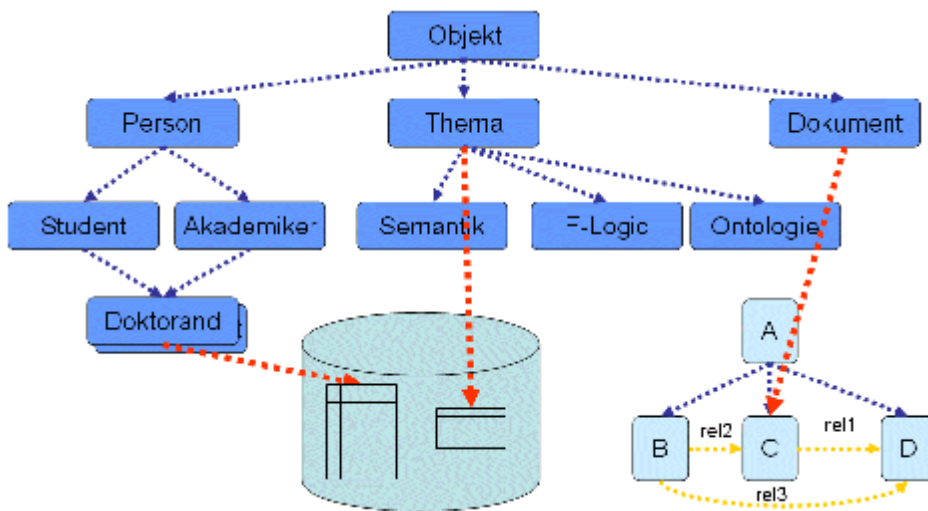


Abbildung 7: Beispiel einer Datenbankintegration mittels Ontologien

Ontologien eignen sich somit wie Topic Maps zur Verbesserung von Suchsystemen, zur Navigation und zur Visualisierung. Weitere Einsatzgebiete sind Frage- und Antwortsysteme, sowie die Integration heterogener Datenquellen. Sind zum Aufbau von Anwendungen mehrere

¹² weiterführend: Maier, Andreas, et al: Ontology-based Information Integration in the Automotive Industry, Whitepaper ISWC 2003, Industrial Track, Integration

dieser drei Dimensionen zu betrachten, stellt die Ontologie die bevorzugte Wahl dar, da Sie als wiederverwertbare Middleware zur Verfügung gestellt werden kann.

Fazit und Ausblick

Alle vorgestellten Modelle eignen sich zur Strukturierung und Abbildung von Inhalten. Taxonomien und Thesauri sind als Hilfsmittel in Applikationen einsetzbar, während Topic Maps und Ontologien für anspruchsvollere Aufgaben geeignet sind. Ontologien gewährleisten die Zukunftsfähigkeit und somit einen Investitionsschutz, denn Sie decken die Fähigkeiten von Taxonomien, Thesauri und Topic Maps ab, bieten zusätzliche Erweiterungsfähigkeit in Richtung Integration und regelbasierte Systeme und sind schon heute auf der Höhe der zukünftigen Standards des World Wide Web Consortiums.

Ontologien in der Repräsentationssprache F-Logic bieten ein durchgängiges Konzept zur Modellierung von Daten und Datenstrukturen, Inhalten, Regeln und zur Anfrage an dieses Modell. Somit lassen sich große Stücke der Applikationslogik mit F-Logic modellieren, anstatt Sie in die Applikationen einzuprogrammieren. Der Ontologieserver stellt wie z.B. ein Applikationsserver seine Funktionalitäten allen Applikationen zur Verfügung und hebt somit die Semantik aus der Applikation in eine eigene Schicht.

Über ontoprise

Die ontoprise¹³ GmbH ist der führende Anbieter semantischer Technologien und Lösungen auf Basis von Ontologien, welche Unternehmen dabei helfen, schnell und kostengünstig Anwendungen zu integrieren, Entwicklungsprozesse zu beschleunigen und Wissensprozesse zu optimieren. ontoprise bietet Produkte für die Erstellung von Ontologien, zur Verknüpfung von Dokumenten mit

¹³ OntoBroker®, OntoEdit®, OntoAnnotate®, SemanticMiner® und ontoprise® sind eingetragene Marken der ontoprise GmbH. Teile der Technologie sind patentiert bzw. zum Patent angemeldet.

Ontologien, zur Anfrage und Verarbeitung von Ontologien (Ontologieserver) sowie Lösungen für ontologiebasiertes Retrieval und Integration von Ontologien in die Office-Suite von Microsoft.

Für weitere Informationen besuchen Sie <http://www.ontoprise.de> oder rufen Sie uns an unter +49 (0) 721-509 809-0.