

# Ontologien & Semantische Technologien zur Suchunterstützung



WIR FORSCHEN FÜR SIE

**Andreas Abecker**  
FZI  
Forschungszentrum  
Informatik

Karlsruhe  
4. November 2008

# Allgemeine Sicht auf das Suchproblem

Informationsbedarf  
(aus persönlichem,  
Aufgaben- und  
physikalischem  
Kontext)



Anfragesprache

Anfragebegriffe

?????? AND ???  
???? OR ??????

Im Zentrum des  
Karlsruher Stadtteils  
Knielingen, direkt an der  
B10 gelegen, ..

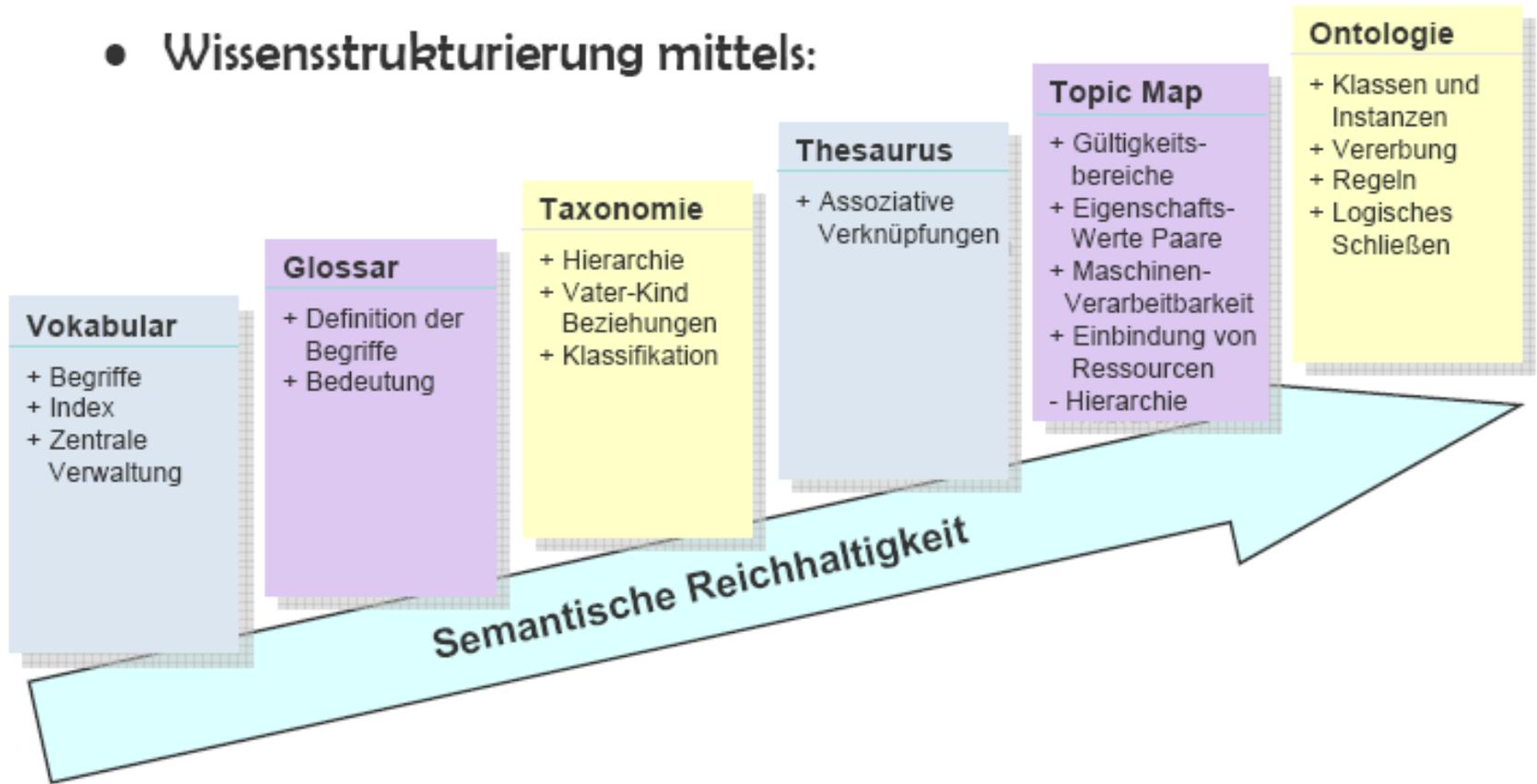
Karlsruhe	12,456
Knielingen	15,12
B10	4,12, 170, 456

Anfrage  
-Auswertung

Dokumentbestand

Dokument-Beschreibung

- Wissensstrukturierung mittels:



- **Metadaten** beschreiben im Web zugreifbare, strukturierte und unstrukturierte Daten/ Informationen/ Dienste und ihre Inhalte
- Metadaten beziehen sich auf **Ontologien**
- Ontologien beschreiben:
  - die **Arten von Objekten, Objekteigenschaften** und -**beziehungen**
  - **konkrete Objekte** mit ihren Eigenschaften und Beziehungen
  - **Gesetzmäßigkeiten + Zusatzwissen** (Axiome, Integritätsbedingungen, Hintergrundwissen, Herleitungswissen)
  - in einem **Ausschnitt der Welt**
  - in einer **standardisierten, maschinenverständlichen Sprache**
  - nachdem sich eine Gruppe von Akteuren auf dieses Modell **geeignet** hat
  - inklusive der **Sprechweisen** über diese Sachverhalte (lexikalische Ebene)

# „Flaschenhals“ Ontologie-Erstellung und -Pflege

Experten-Workshops

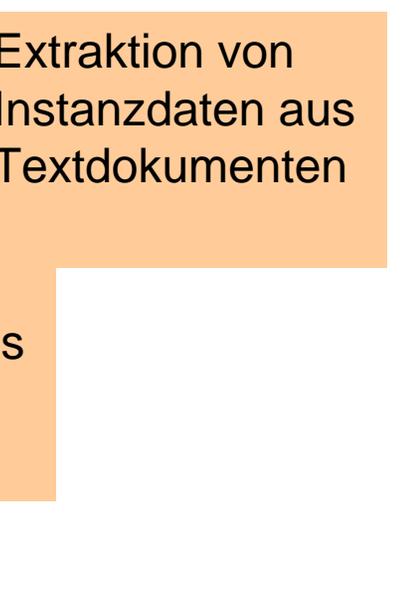
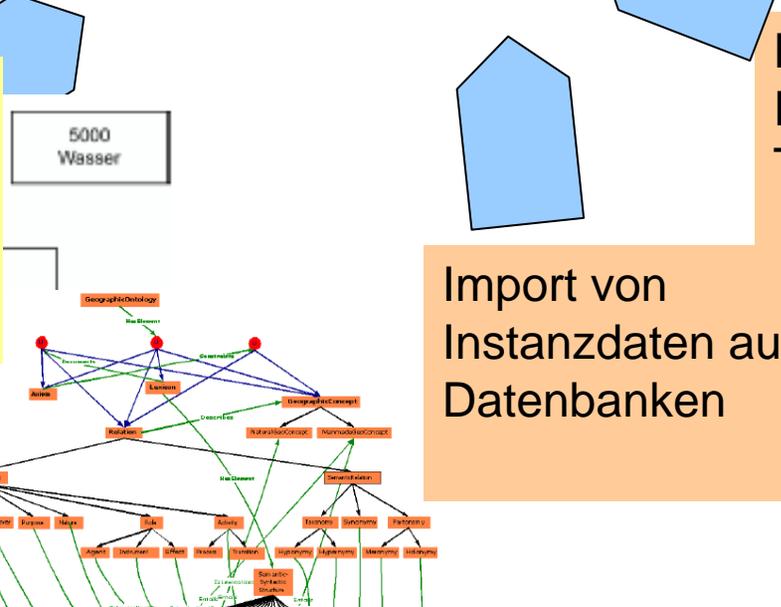
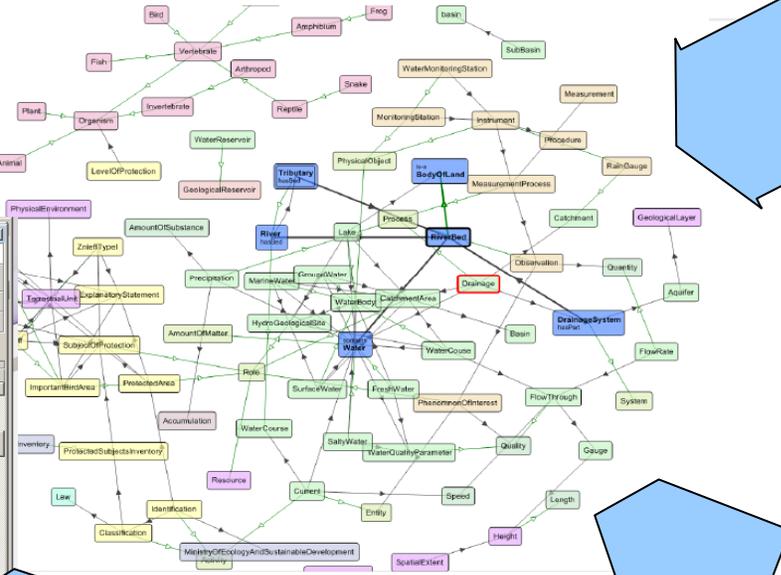
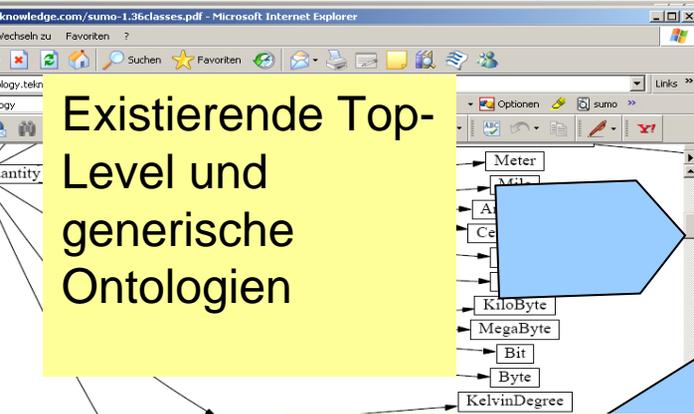
Existierende Top-Level und generische Ontologien

Einbindung existierender Geo-Ontologien und Metadaten-Schemata

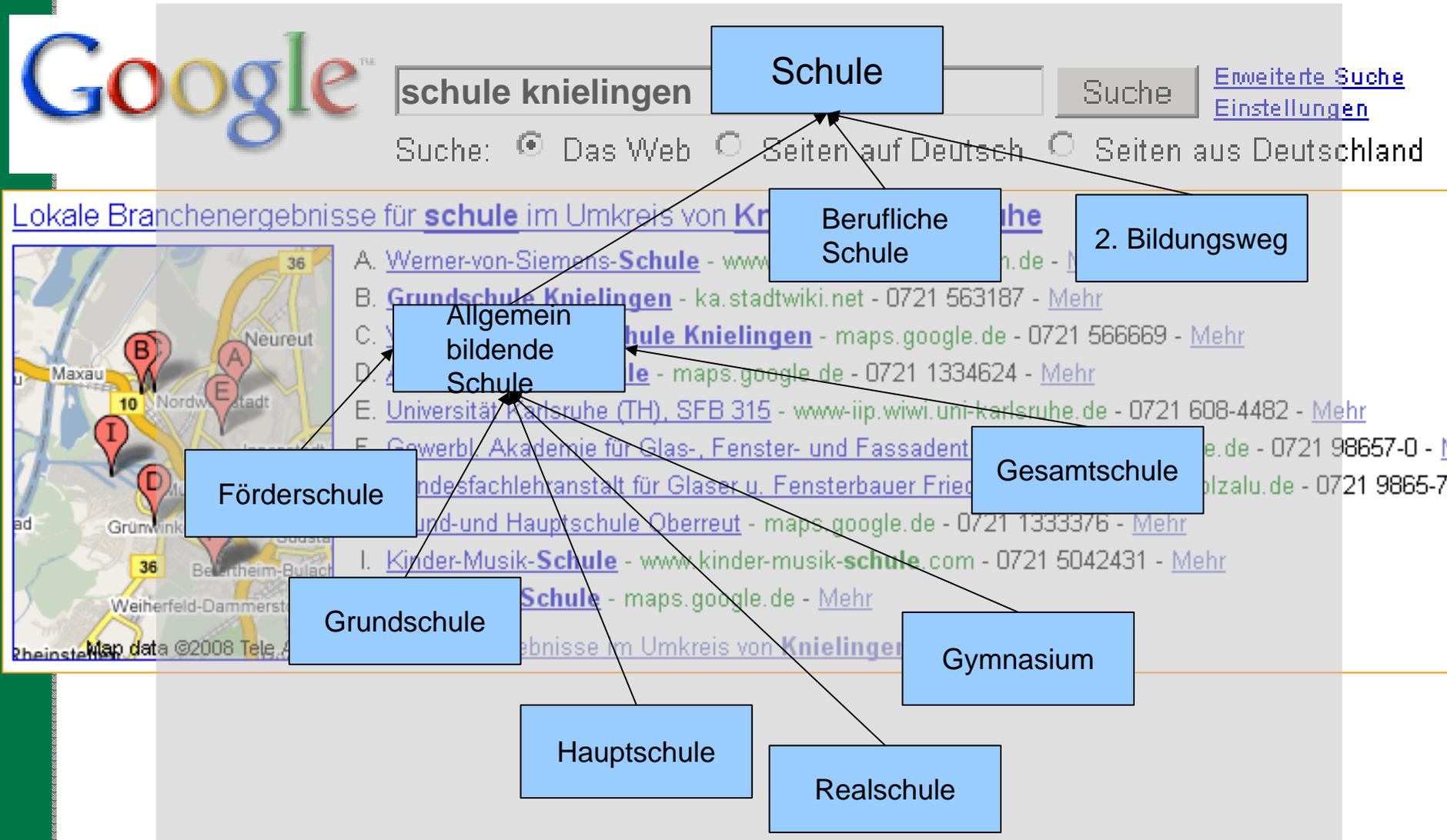
Kontinuierliches Nutzungs-Feedback und kollaborative Pflege

Extraktion von Instanzdaten aus Textdokumenten

Import von Instanzdaten aus Datenbanken



# Anfrageverfeinerung mit Konzepthierarchie



The image shows a Google search interface for the query "schule knielingen". The search results are filtered to show local results in the vicinity of Knielingen. A concept hierarchy is overlaid on the page, starting with "Schule" at the top. This hierarchy branches into "Berufliche Schule" and "2. Bildungsweg". "Berufliche Schule" further branches into "Allgemein bildende Schule" and "Gesamtschule". "Allgemein bildende Schule" branches into "Förderschule", "Grundschule", "Hauptschule", and "Realschule". "Gesamtschule" branches into "Gymnasium".

**Google** schule knielingen **Schule** Suche [Erweiterte Suche](#) [Einstellungen](#)

Suche:  Das Web  Seiten auf Deutsch  Seiten aus Deutschland

Lokale Branchenergebnisse für **schule** im Umkreis von **Knielingen**

- A. [Werner-von-Siemens-Schule](#) - www...n.de - Mehr
- B. [Grundschule Knielingen](#) - ka.stadtwiki.net - 0721 563187 - Mehr
- C. [Volkshochschule Knielingen](#) - maps.google.de - 0721 566669 - Mehr
- D. [Zweiter Bildungsweg](#) - maps.google.de - 0721 1334624 - Mehr
- E. [Universität Karlsruhe \(TH\), SFB 315](#) - www-iip.wiwi.uni-karlsruhe.de - 0721 608-4482 - Mehr
- F. [Gewerbl. Akademie für Glas-, Fenster- und Fassadentechnik](#) - www.glasakademie.de - 0721 98657-0 - Mehr
- G. [Landesfachlehranstalt für Glaser u. Fensterbauer Friedrich](#) - www.fachschule-glaser.de - 0721 98657-1 - Mehr
- H. [Grund- und Hauptschule Oberreut](#) - maps.google.de - 0721 1333376 - Mehr
- I. [Kinder-Musik-Schule](#) - www.kinder-musik-schule.com - 0721 5042431 - Mehr
- J. [Musikschule Knielingen](#) - maps.google.de - Mehr

**Schule**

- Berufliche Schule**
- 2. Bildungsweg**

**Allgemein bildende Schule**

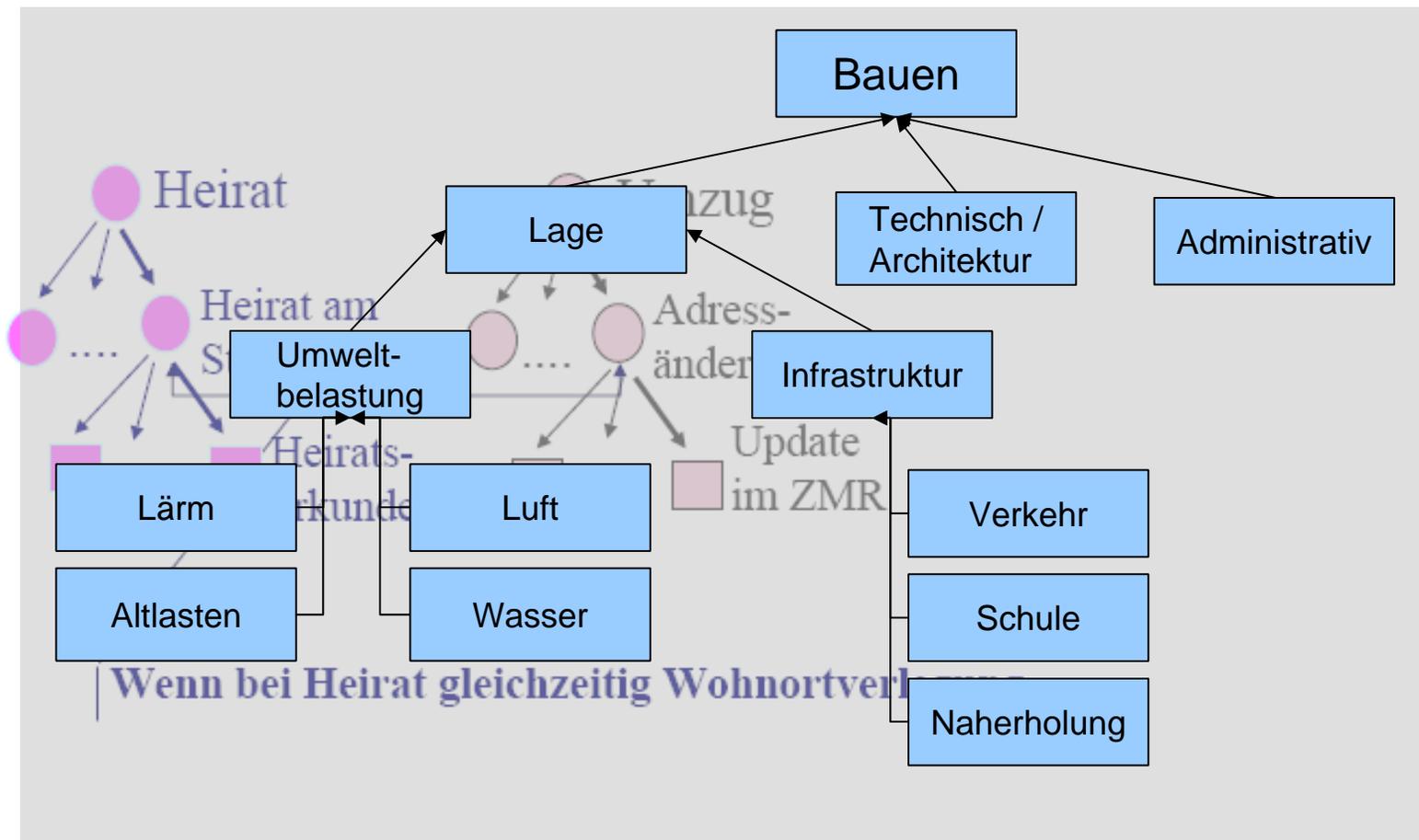
- Förderschule**
- Grundschule**
- Hauptschule**
- Realschule**

**Gesamtschule**

- Gymnasium**

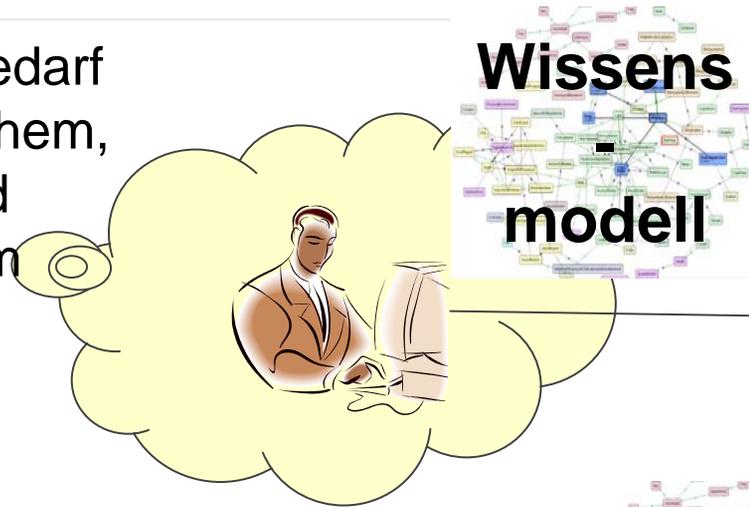
# Ontologie kann Sichten auf Information repräsentieren, die in den Quellen selber nicht direkt enthalten sind

- Traditionelles Lebenslagen-Konzept: Situation – Dienst - Dokument
- Erweiterung um Informationsbedarf und Geo-Umweltbezug



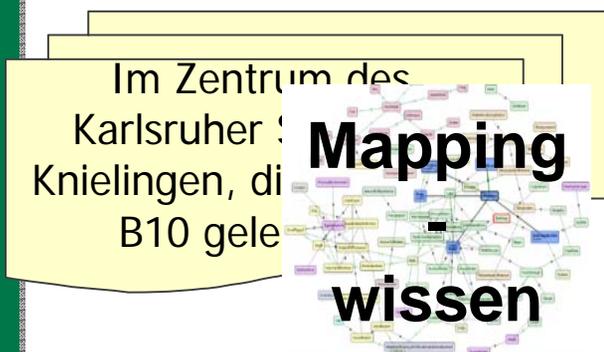
# Zusammenfassung: Semantische Suche

Informationsbedarf  
(aus persönlichem,  
Aufgaben- und  
physikalischem  
Kontext)

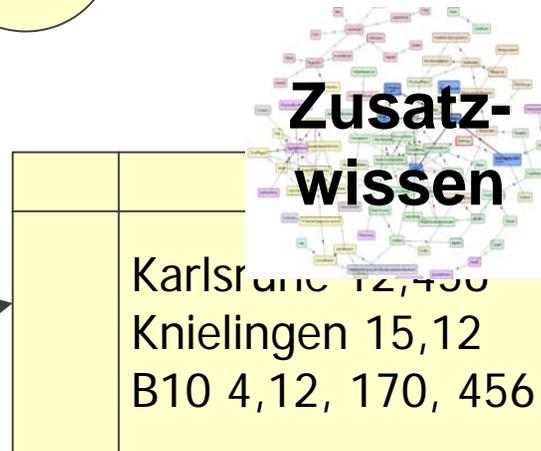


Anfragesprache

Anfragebegriffe



Dokumentbestand



Dokument-Beschreibung



# Zusammenfassung: Nutzenpotenziale bei der semantischen Suche



- Semantische Technologien haben in ihrem Zentrum ein formales Wissensmodell plus Regelformalismen zur Verarbeitung
- Dieses ermöglicht typische Vorteile bei der Informationssuche:
  - **Suche nach Bedeutung statt nach Schlüsselbegriffen**
    - mehrsprachig, multimedial
    - verschiedene Wissensmodelle können gut aufeinander abgebildet werden (regelbasiert, teil-automatisch) und zusätzliche Sichten für Info.zugang darstellen
  - **Vernetzte Suche (nach unterschiedlichen Informationsobjekten)**
    - auch in Datenbanken, auch in vernetzten Systemstrukturen
    - Anbindung heterogener Back-End Systeme (Regeln und Schema-Mapping)
  - **Moderierte Suche (Benutzer wird bei Anfrageformulierung unterstützt)**
    - Bei zuvielen, zu wenigen oder „unzureichenden“ Antworten
    - Bei mehrdeutiger Anfrage
  - **Intelligente Anfrageauswertung**
    - Räumlich-zeitliche Konzepte
    - Automatische Umformulierungsstrategien bei fehlenden Antworten
    - Fakten-Extraktion aus Dokumenten
    - Fakten-Nachbearbeitung oder Visualisierung von Ergebnissen

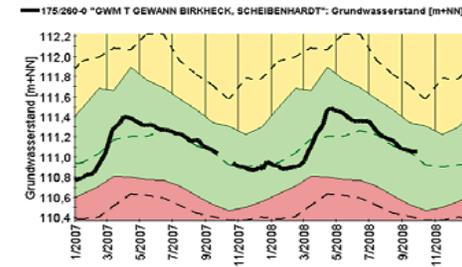
# Probleme der Google-Internet-Recherche

- **Benennung des Problems**
  - Finden von geeigneten Schlagworten
- **Unklare Begriffe**
  - Kein offizieller Name für das Baugebiet (Knielingen 2.0?)
  - Konversionsgelände, Kaserne, Baugebiet sind keine Synonyme
- **Zu allgemeine Begriffe liefern viele/unrelevante Treffer**
  - besser als „Luft“ wäre „Luftqualität“
- **Google kann teilweise Ortsbezug herstellen**
  - jedoch für kein Umweltthema
  - Ortsbezogene Suche nach Umweltthemen auch mit Google Maps nur bedingt möglich
- **Suche liefert oft weitere Schlagworte statt Fakten**



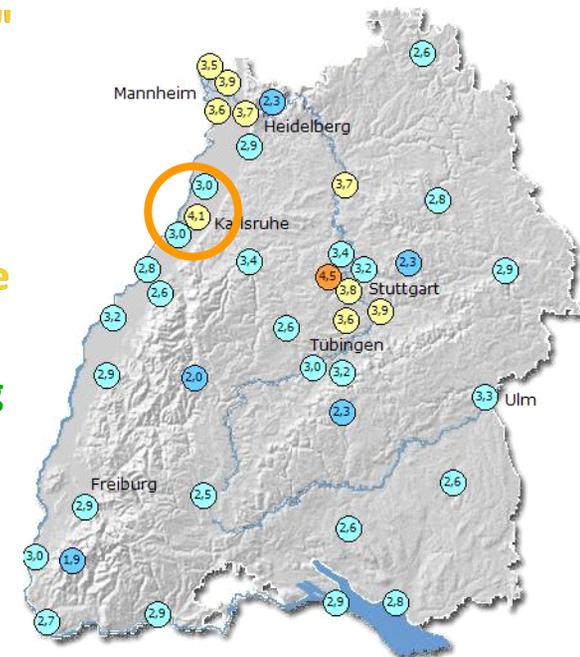
## Findet häufig nur „in die Nähe“ der Information

- Einzelsysteme sind zu wenig vernetzt
- Nicht alle Systeme sind ans Umweltportal angeschlossen
- **Wahl der richtigen Suchbegriffe schwierig**
- Nicht alle Daten liegen für jede Geoeinheit (z.B. Ortsteil Knielingen) vor



## Keine räumliche Suche

- Kombination aus räumlicher Begrifflichkeit "Knielingen" und thematischen Suchbegriff "Luftqualität" liefert wenig Ergebnisse
  - obwohl Informationen in DB's vorhanden
  - liegt an fehlender Integration der räumlichen Daten in die Suche
  - **Oft automatische Erweiterung der räumlichen Umgebung nötig, wenn für spezielle Anfrage keine Treffer vorliegen**





Vielen Dank !  
Fragen?

**Andreas Abecker**

FZI Forschungszentrum Informatik, Abt. IPE

<http://www.fzi.de/ipe>

# Ontologie-Nutzung in der Informationssuche



- 2 Modi:
  - Navigation / Browsing  
(auch mit nutzerspezifischen Sichten)
  - Queries
    - Kollaborativ verfeinern
    - Automatisch umschreiben
- Möglichkeiten bei der Anfrageverarbeitung
  - Gesonderter **lexikalischer Layer** schwächt Probleme der „Schlüsselwortsuche“ ab
  - Nutzung der **Taxonomie** für Spezialisierung / Verallgemeinerung
  - Nutzung von **Attributen/Relationen** für spezifischere Anfragen
  - **Hintergrundwissen** für Anfragebearbeitung, Ergebnismachbearbeitung oder Sortierung
  - Komplexe **Ähnlichkeitsbetrachtungen** für die Relevanzbeurteilung

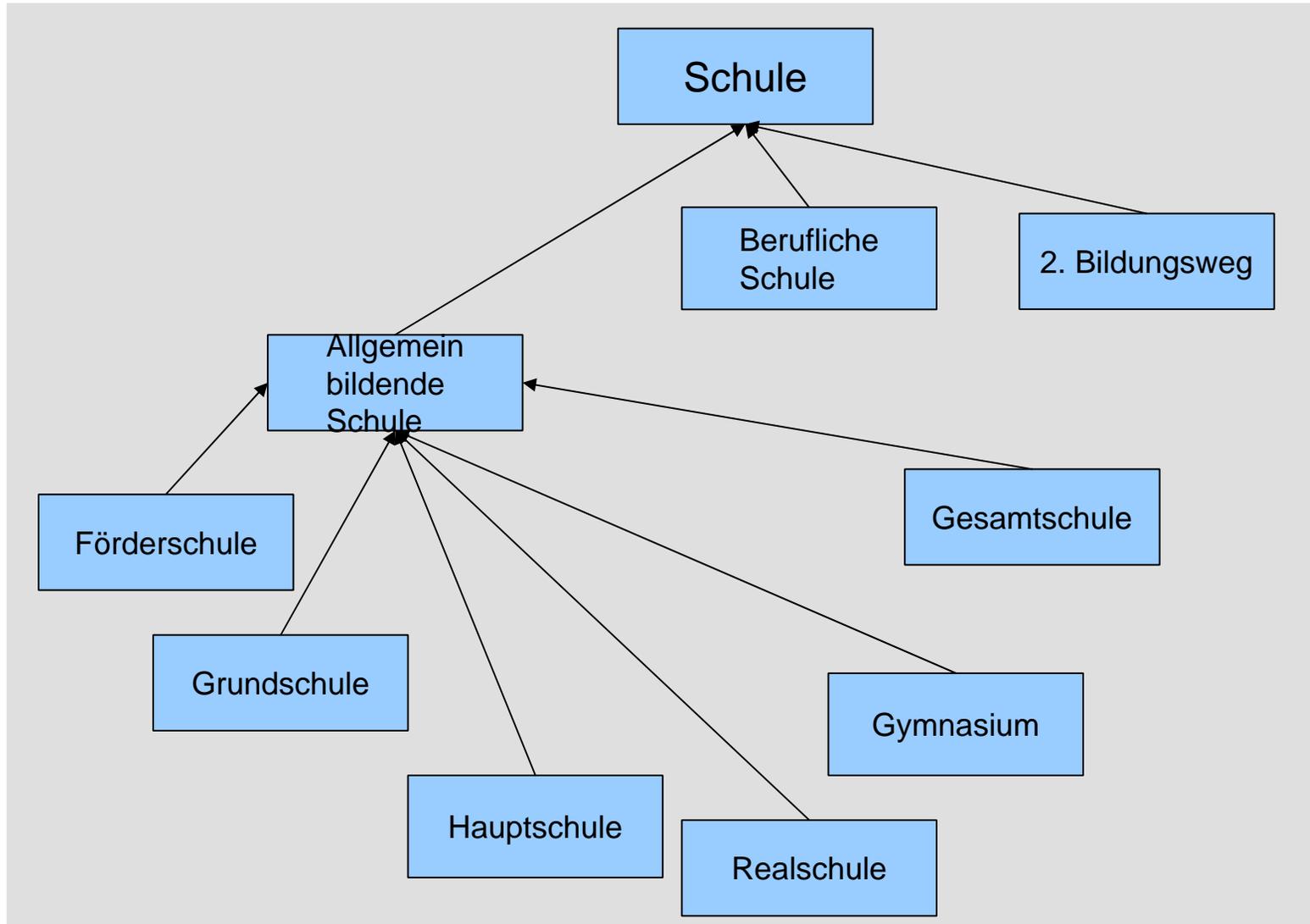
# „Take home message“ Worauf beruht der Mehrwert?

---



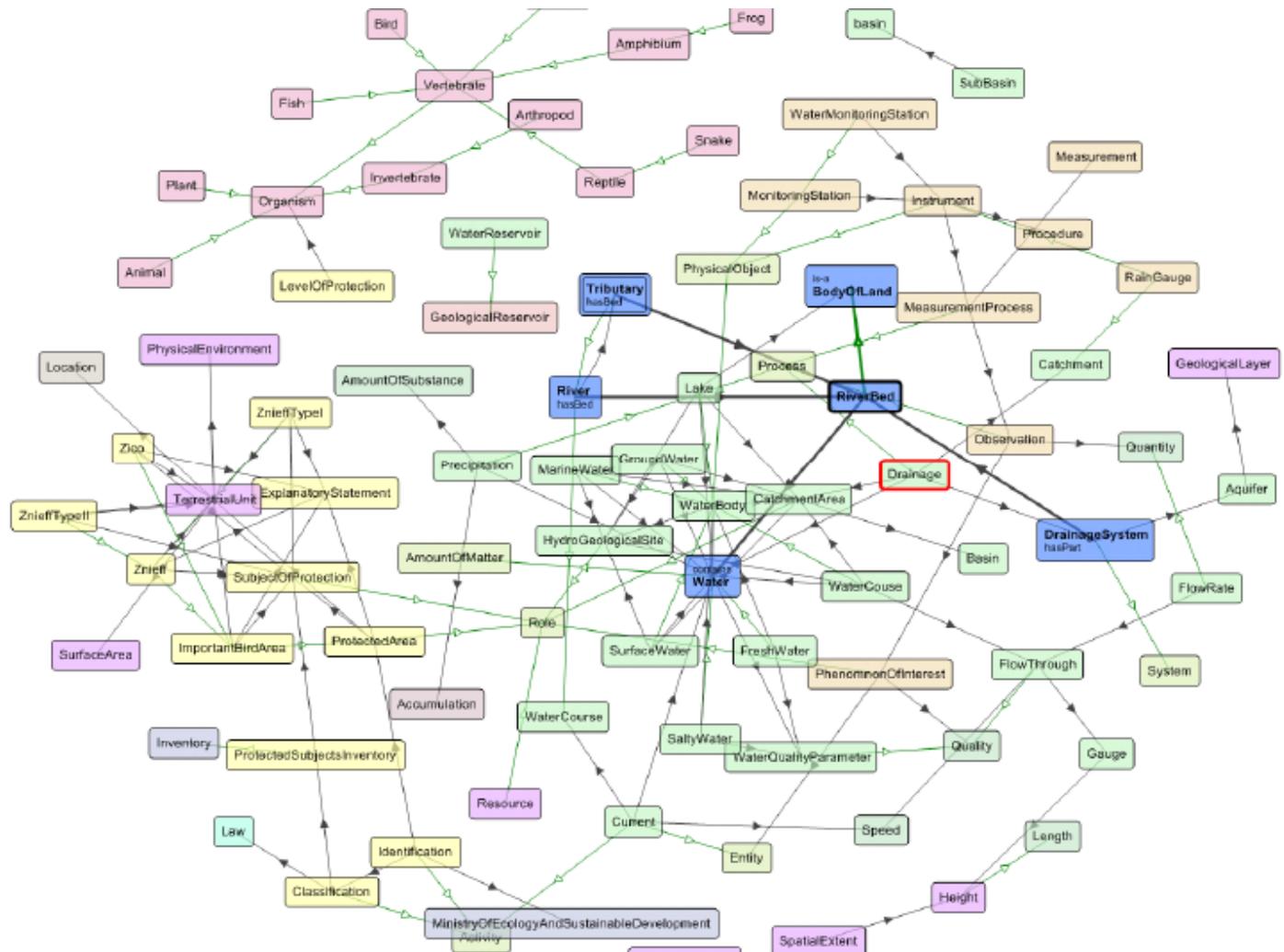
- Explizite, getrennte, mächtige **Metadaten**
  - **Einfache und spezifische**
  - **Relationale Metadaten und Bezüge zwischen Informationsobjekten**
- Nehmen Bezug auf **Ontologien**
  - **Konsensfindung**
  - **Standardisierung**
  - **Mächtige Repräsentation**
  - **Einbettung von Hintergrundwissen**

# Mini-Ontologie-Teil Schulformen in Deutschland

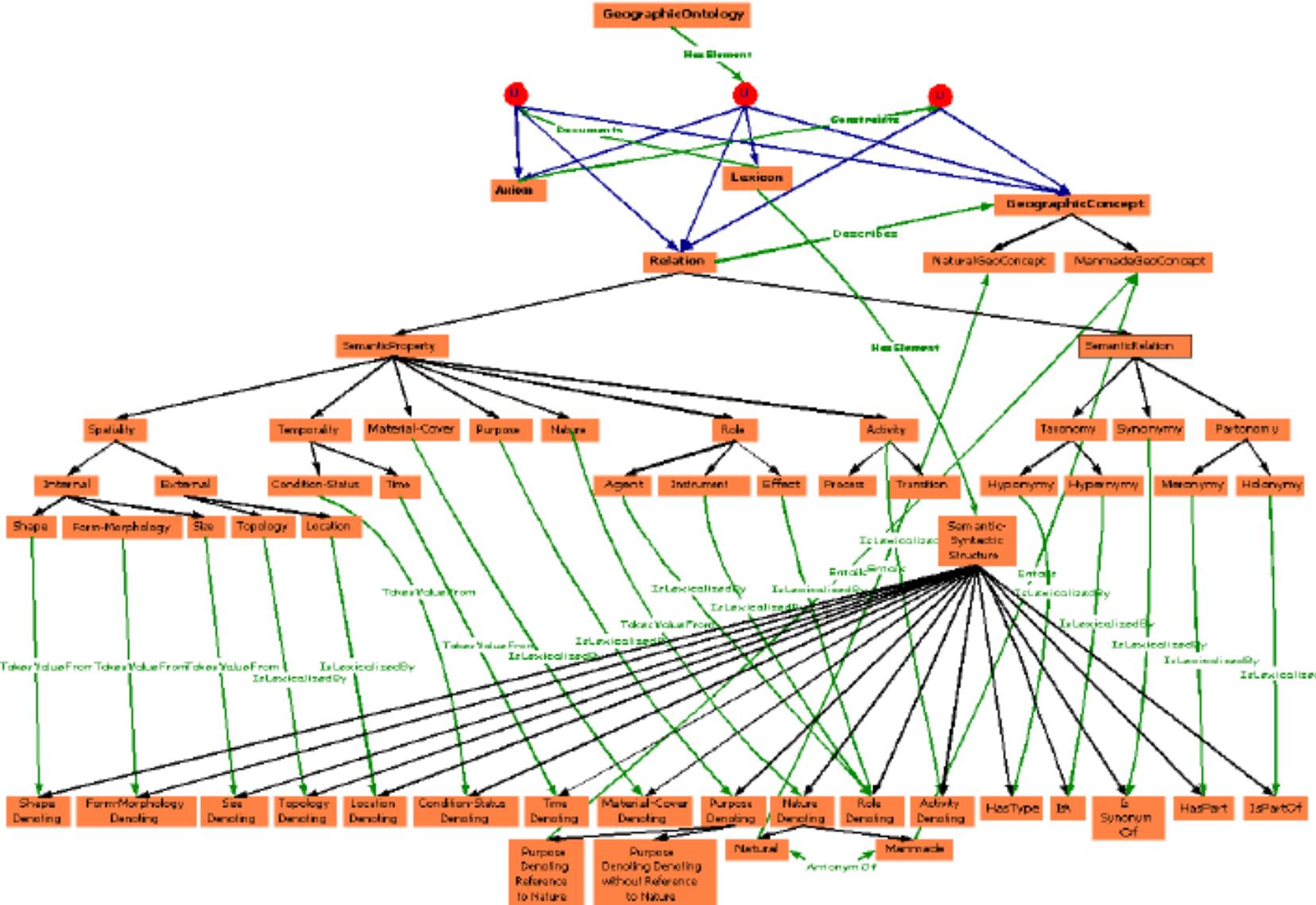




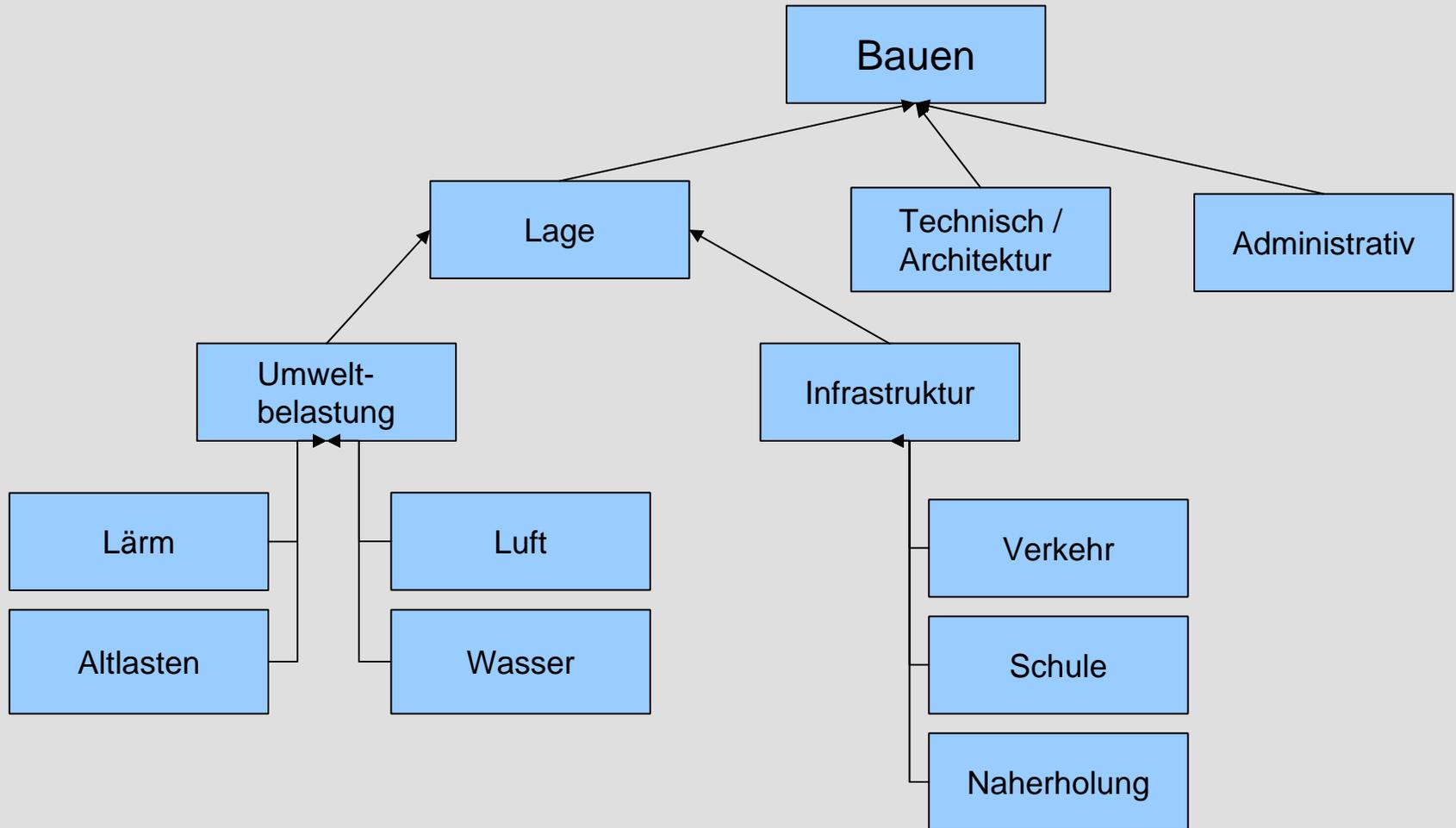
# Geo Ontology Excerpt from Deliverable 3.2 of SWING Project



# Generic Geo Ontology [Tomai & Spanaki, 2005]



# Mini-Ontologie-Teil Lebenslage Bauen



# Einige Quellen zu semantischen Technologien

- Fraunhofer FIRST: **Semantische Technologien in der betrieblichen Anwendung.** 2006.  
<http://www.rolotec.ch/blog/archives/SemantischeTechnologien.pdf>
- Steffen Staab, Rudi Studer (Hrsg): **Handbook on Ontologies.** Springer 2003
- P. Hitzler, M. Krötzsch, S. Rudolph, Y. Sure: **Semantic Web, Grundlagen.** Springer 2007
- FAZIT Schriftenreihe Band 13: **Semantische Technologien in Wissensmanagementlösungen.**  
<http://www.doit-online.de/cms/do+it.service/Publicationen?detailid=182>

