



Semantic Web/Web 3.0

Ideen, Chancen und Anwendungsgebiete
im Intra- und Internet

ab-in-den-urlaub.de
Wir schicken Deutschland in den Urlaub

Werden Sie Besser-Wisser
news.de

AVITO

auto.de
Deutschlands großes Autoportal

Reisende

BEST-PRICE.COM

flug24 ✈

 **preisvergleich.de** partnersuche.de



Shopping.de

urlaubstours

Travel24
Tag für Tag die besten Deals

GELD.de



Hotelreservierung.de

Versicherungen.de

fluege.de ✈

Inhalt des Vortrags

- Vorstellung der Unister GmbH
- Übersicht über Entwicklung des Internet zur Wissensbasis
- Einführung in Semantic-Web-Technologie
- technische Grundlagen
- (Forschungs-)Aufgaben
- Anwendungsgebiete

Gründung/Geschäftsgebiete

- Klassisches Start-Up-Unternehmen (Gründung 2003)
- Internetdienstleister mit mehr als 35 nationalen und internationalen Portalen in unterschiedlichen Branchen
- 16,5 Mio. Unique User im Monat allein in Deutschland
- Entwicklung, Betrieb und Vermarktung der Portale

Portale/Geschäftsmodelle



Mitarbeiter/Standorte

Entwicklung der Niederlassungen



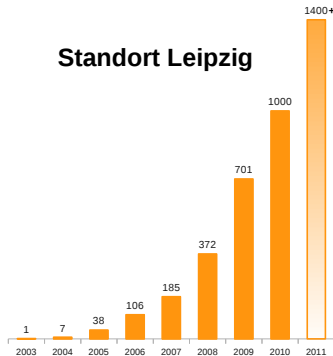
- Ausbau 2010: Dresden, Chemnitz, Jena
- Ausbau 2011: Magdeburg, Rostock, Stralsund, Wismar, Greifswald

- IT-getrieben
- Mittlerweile über 1200 Mitarbeiter
- Durchschnittsalter (IT): 31 Jahre

Unister Erfolgsgeschichte



Standort Leipzig



Unister GmbH



Michael Rapp

bis 2005 Studium der Wirtschaftsinformatik an der Universität der Bundeswehr München

bis 2010 Offizier im IT-Bereich der Bundeswehr

seit August 2010 Standortleiter Chemnitz bei der Unister GmbH



Dr. Andreas Both

bis 2005 Studium der Informatik an der Martin-Luther-Universität Halle-Wittenberg

bis 2010 Promotion zum Dr. rer. nat.

seit April 2010 Leiter des „Semantic Web Project“ (Forschung und Entwicklung) bei der Unister GmbH



Aktuelle Beobachtung

Thesen:

Das Internet ist die größte Wissensbasis.

Unternehmen verfügen über mehr Informationen,
als ihnen bewusst ist.

Internet/WWW

Zustand

- Internet-Texte für Menschen
 - Anzahl steigt ständig
 - Inhalte un-/semistrukturiert
 - Suchmaschinen sind nötig
- ⇒ „alle Informationen vorhanden“

Potenzielle Anwenderprobleme

- Suche:
 - „bester“ Suchbegriff ist nicht bekannt
 - mehrdeutige Wörter
- (Web)Seite muss „gelesen“ werden

⇒ kein reines Internet-Problem



Beispiel: Leopard



Ebenso: andere Bedeutungen möglich

Beispiel: Suchen ohne Typisierung

- Leopard Tier:
 - „Das **Tier** im Manne - **Leopard** und Co. werden Trend in der Männermode ...“
- Leopard Gattung:
 - „Österreichs Bundesheer - Waffengattungen - Panzer“
- Leopard Türkei:
 - „**Leopard** 2 - Wikipedia“

Inhalte nicht eindeutig zugeordnet.

Beispiel: Inhalte ohne Typisierung

Titel eines Artikels:

„Leopard mit löchriger Rüstung“

`http://www.heise.de/security/artikel/
Ein-zweiter-Blick-auf-die-Firewall-in-Mac-OS-X-Leopard-270926.
html`

Semantik nicht eindeutig bekannt gegeben.

Texte „ohne“ Semantik

Beobachtung

- ⊕ redaktionelle Internet-Texte
- ⊖ Tweets/Blogs/Foren
- ⊖ Bücher/Zeitschriften
- ⊖ Gespräche/E-Mails
- ⊖ Unternehmensdaten
- ⊖ ...

Informationen und Semantik

Hintergrund: Evolution des Internets

Betrachtung: Inhalte im Internet

- Web 1.0: „statische“ Inhalte
- Web 2.0: interaktiver und kollaborativer Elemente
- Web 3.0: „Social Semantic Web“
- Web 4.0: Echtzeitinformationen, intelligente Interaktionen
- Web 5.0: ...

Einordnung hinterfragen!

Arbeiten mit Informationen

- Wissensgesellschaft ist entstanden
- Informationsgewinnung ist essentiell
- Es wird zu viel Zeit mit „suchen“ verschwendet!
 - 33% der Arbeitszeit von Wissensarbeitern: Suchen nach Informationen und helfen beim Suchen [Cro04]
 - *„It is generally estimated that for each \$1 spent for an application companies spend on average \$5 to \$9 for the integration.“ (Nelson Mattos)*
- Datenredundanzen kosten Zeit
 - Anwender: beim Suchen
 - Betreiber: bei der Aufbereitung und Integration
- Datenintegration kostet viel Zeit

Aufwendige Datenintegration



Aufwendige Datenintegration

wenig Datenverknüpfungen

- SQL-Datenbanken existieren isoliert
- man muss Verknüpfung (er)kennen
 - SQL-Relation sind unscharf
- Neu hinzukommende Datenquellen?

Beobachtungen

- Modellierung des globalen Wissens nicht möglich
- Redundanzen (quasi) unvermeidbar
- Beziehungen werden nicht eindeutig abgebildet

Wirkung

Entwicklerseite

- mühsame Datenanalyse
 - undeutliche Information über Bedeutung der Daten
 - Informationen sind (nur) in den „Köpfen“
 - Refactorings nötig bei Umorganisation von Daten
- höhere Prozesskosten (ebenso TTM)
- Daten-Redundanzen entstehen unbemerkt
- Datenbasis wird wenig gepflegt

Kundenseite?

- Informations werden händisch zusammengesucht



Bedeutung: Entwickler

- viele Firmen betreiben Datengetriebene Prozesse
 - nebenläufige Aufstellung von Datenquellen
 - Erschließung von neuen Domänen
 - Integration vieler Datenlieferanten (extern), Abteilungen (intern) oder Werkzeuge
 - ...

⇒ Datenintegration hat hohe Bedeutung

Lösungsansätze

Dokumentiere!?

- wenig maschinenlesbar
- kein einheitlicher Standard
- macht das jemand vollständig und umfänglich?

Verknüpfe und etabliere Informationen!

- Beziehungsnetz:
 - Metadaten mit Verknüpfung
 - Objekte und Beziehungen typisierbar
- Konzepte geben Zusammenhänge an
- Wissen in verteilten Datenquellen?
- neues Wissen bestimmen



technischer Lösungsansatz

Semantic-Web-Technologien

- logische, formalisierte Darstellung von Wissen
- Standards zur technischen Repräsentation (W3C)



<http://www.w3.org/2001/sw/>

⇒ allgemeine Herangehensweise zur Modellierung von Wissen

logische, formalisierte Darstellung von Wissen

- Definition von Beziehungen und Eigenschaften:
- Prinzip der Aussagen: **S**ubjekt **P**rädikat **O**bjekt

Beispiel

S	P	O
Unister	ist ein	IT-Unternehmen
Michael	ist Mitarbeiter von	Unister
Michael	hat Nickname	michael17

Standards zur technischen Repräsentation

- HTML ist nicht ausreichend
 - Standard XML ist nicht ausreichend
- ⇒ allgemein verarbeitbarer eindeutiger Syntax

Standards zur technischen Repräsentation: RDF

- genutzte technische Darstellung:
 - **Resource Description Framework (RDF)** [RDF99]
 - „System zur Beschreibung von Ressourcen“ [wikipedia]
 - Ressourcen werden eindeutig durch Verwendung von URIs (bzw. IRIs)

Beispiel

S	P	O
<code><http://unister.de></code>	ist ein	<code><http://unternehmen.de/type#it></code>
<code><http://personen.de/#michael17></code>	ist Mitarbeiter von	<code><http://unister.de></code>
<code><http://personen.de/#michael17></code>	hat Nickname	<code><http://jabber.de/account#mi17></code>
<code><http://jabber.de/account#mi17></code>	hat Label	<code>"M. Rapp"^^xsd:string</code>

⇒ Allgemeingültige Referenzierung von Entitäten

Erweitertes Beispiel

S

P

O

<http://unister.de>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://unternehmen.de/type#it>
<http://unister.de>	<http://unternehmen.de/#hasEmpl>	1200^^xsd:int
<http://personen.de/#michael17>	<http://unternehmen.de/#isEmpl>	<http://unister.de>
<http://personen.de/#michael17>	<http://xmlns.com/foaf/0.1/depiction#nick>	<http://jabber.de/account#mil7>
<http://jabber.de/account#mil7>	<http://www.w3.org/rdf-schema#label>	"M. Rapp"^^xsd:string

⇒ Allgemeingültige Darstellung von Relationen

Abfrage von Beziehungen

```

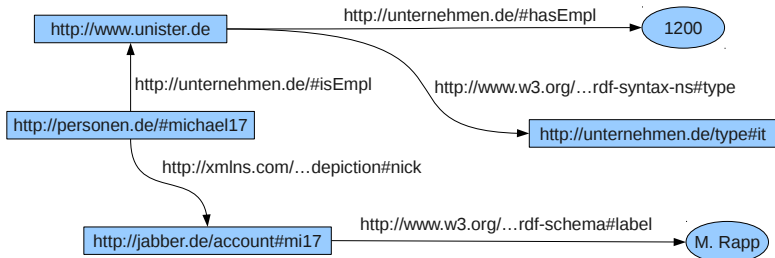
<http://unister.de>
  <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
    <http://unternehmen.de/type#it>

<http://unister.de>
  <http://unternehmen.de/#hasEmpl>
    1200^^xsd:int

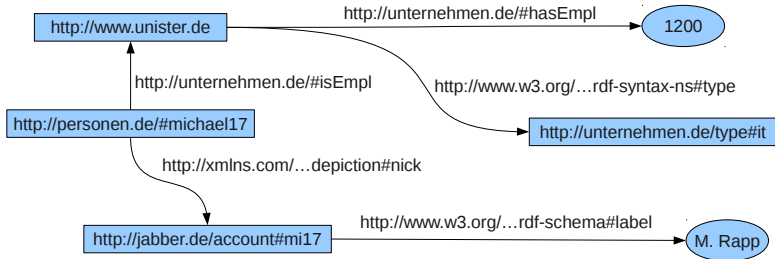
<http://personen.de/#michael17>
  <http://unternehmen.de/#isEmpl>
    <http://unister.de>

<http://personen.de/#michael17>
  <http://xmlns.com/foaf/0.1/depiction#nick>
    <http://jabber.de/account#mi17>

<http://jabber.de/account#mi17>
  <http://www.w3.org/rdf-schema#label>
    "M. Rapp"^^xsd:string
  
```



Abfrage von Beziehungen: SPARQL



```

SELECT DISTINCT ?newcontact {
  ?company <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
    <http://unternehmen.de/type#it>.
  ?company <http://unternehmen.de/#hasEmpl> ?employees .
  Filter( xsd:int(?employees) > 1000 ) .
  ?employee <http://unternehmen.de/#isEmpl> ?company .
  ?employee <http://xmlns.com/foaf/0.1/depiction#nick> ?nickurl .
  ?nickurl <http://www.w3.org/rdf-schema#label> ?newcontact . }

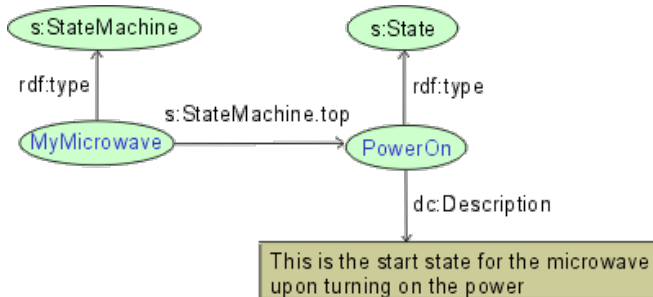
```

Beschreibung von Konzepten

Ontologien

- „Eine Ontologie ist eine explizite formale Spezifikation einer Konzeptualisierung.“ [Gru93]
- beschreibt Eigenschaften und Relationen von Klassen untereinander:
 - IT-Unternehmen ist Unterklasse von Unternehmen
 - Menschen haben ein Geschlecht
 - Geschlecht kann sein: männlich, weiblich, ...
- Prinzip vergleichbar mit UML/Klassendiagrammen
- Standard: **Web Ontology Language (OWL)** [OWL04]

Beispiel: Klassen/UML



Quelle: <http://infolab.stanford.edu/~melnik/rdf/uml/>
<http://www.w3.org/TR/NOTE-rdf-uml/>

Wissenserzeugung

Reasoning

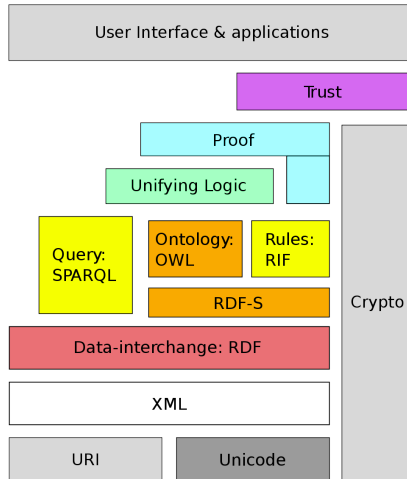
- Auswertung von bestehenden Informationen
- indirekte Zusammenhänge erkennen/beweisen
- Inferenz-/Integritätsregeln:
 - Beantwortung von Fragen
 - Sicherstellung von Konsistenz
- Konzept der künstlichen Intelligenz
- vergleichbar logischer Programmierung
 - Prolog, Datalog

Beispiel

Wenn Leipzig in Sachsen $\Rightarrow$ ein Leipziger Hotel ist auch ein Sächsisches Hotel



Semantic-Web-Technologie



Anforderungsanalyse

- Datenaustausch
 - transparente Schnittstelle: XML
- Datenauswertung
 - Regeln nach den SPO-Prinzip
 - Anfragen mittels SPARQL, Reasoner
- Datenintegration vorhandener Datenbasen
 - Wissensorganisation durch Ontologien
- Datenintegration externer Informationsquellen?
 - *Fortschritt?*
- Wissensmodellierung?
 - *Fortschritt?*

Wiederholung: RDF

Format nach SPO-Prinzip

- *<IRI>* *<IRI>* *<IRI>*
- *<IRI>* *<IRI>* *atom*

```

<http://unister.de>
    <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
    <http://unternehmen.de/type#it>

<http://unister.de>
    <http://unternehmen.de/#hasEmpl>
    1200^^xsd:int

<http://personen.de/#michael17>
    <http://unternehmen.de/#isEmpl>
    <http://unister.de>

<http://personen.de/#michael17>
    <http://xmlns.com/foaf/0.1/depiction#nick>
    <http://jabber.de/account#mi17>

<http://jabber.de/account#mi17>
    <http://www.w3.org/rdf-schema#label>

```

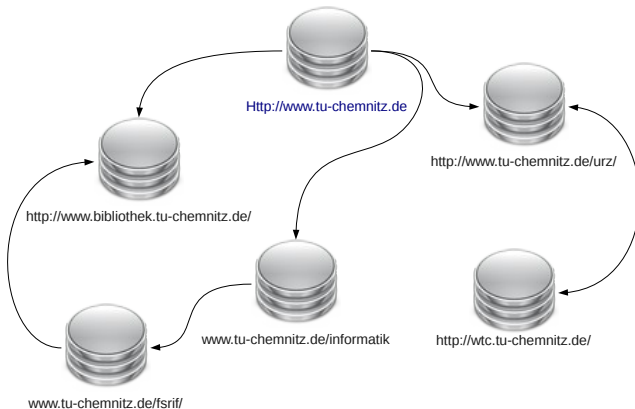
Wiederholung: RDF

Format nach SPO-Prinzip

- $\langle IRI \rangle$ $\langle IRI \rangle$ $\langle IRI \rangle$
- $\langle IRI \rangle$ $\langle IRI \rangle$ *atom*

- IRI referenziert externe Datenquelle
 - verteilte Datenbank entsteht
 - gemeinsames Vokabular sichert Funktionalität
- ⇒ Wissensmodellierung, Datenintegration *externer* Datenquellen

Beispiel: Netzen von Datenquellen



Typen der Relationen sind entscheidend!

Aktuell

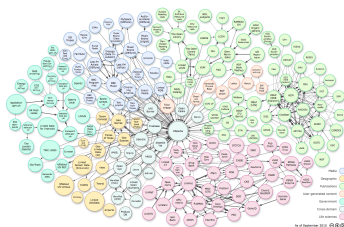
Datennetz entsteht: **Linked Open Data** (LOD)

Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. <http://lod-cloud.net/>

Linked Open Data

Eigenschaften von LOD

- weltweites Datennetz
- separierte aber verknüpfte Expertendatenbanken
 - Kollaborationseffekt
 - Aktualität
- dereferenzierbar
- basierend auf URLs/IRIs
- ⇒ verlinkbar auf „regulären“ Webseiten
 - RDFa
 - Microformats



Verknüpfung von HTML und LOD

Integration von HTML-Content und Wissensdatenbanken

Web 3.0 oder Social Semantic Web

- Verknüpfung von Semantic Web und Web 2.0
- soziale Ansätze integriert
- Community wird genutzt
- erzeugte Inhalte sind verarbeitbar

Zusammenfassung: Semantic Web Technologien

Wissensmodellierung

- liefern Standards und Vokabulare
- verteilte Modellierung wird möglich

Datenintegration

- Dereferenzierung von URLs ermöglicht Zugriff auf benötigte Daten
- Standard-Regelform erlaubt (allgemeine) Interpretation
- externes Expertenwissen wird zugreifbar
- strukturierte und unstrukturierte Informationen können verknüpft werden

Visionen

- Allgemeine Problemformulierungen
→ konkrete Lösungen
 - integrierte Anwendungen
 - KI-Algorithmen berechnen Lösung
 - minimale Interaktion bei Beachtung des persönlichen Kontextes

Vision: Beispiel

Du arbeitest bei Unister. Unister muss ein dringendes Projekt in Dresden durchführen, du wirst als Berater benötigt. Du sollst so bald wie möglich nach Dresden fahren.

Wunsch: schnelle Zugverbindung nach Dresden buchen, Ankunft: morgen 10 Uhr

- wie gewohnt: Sitz am Gang
- Buchung sobald Zug verfügbar (mit deiner Kreditkarte)
- Hinweis: Zahnarzttermin wird versäumt, Verschiebung ist nötig

Du gibst an, dass du eine Limousine benötigst, um zum Dresdner Unister-Büro zu gelangen.

- Suche: Limousinen-Service mit der Bewertung „sehr gut“, Fahrer ab 10:05 Uhr
- Reservierung: dein Dresdner Lieblingshotel, Kundenkarte für günstigsten Preis
- Aktualisierung: Kalender (auch Teamleiter)
- Drucken: Reisebestätigungen (im Büro)

Vision: Ziele

- Datenverknüpfung Anwendungsübergreifend
- Datengetriebene Prozesse automatisch verknüpfen
- Probleme weitgehend automatisch und intelligent lösen (Schlussfolgerungen)
- persönlichen Kontext beachten

⇒ Vision schwer umsetzbar!

(Forschungs-)Aufgaben:

- skalierbare Beweissysteme
- Anwendungsintegration
- Semantik ausformuliert



Informationsflut
auswerten (Relevanz,
Performance) und
produktiv nutzen



Projektorganisation und Vorgehensmodelle

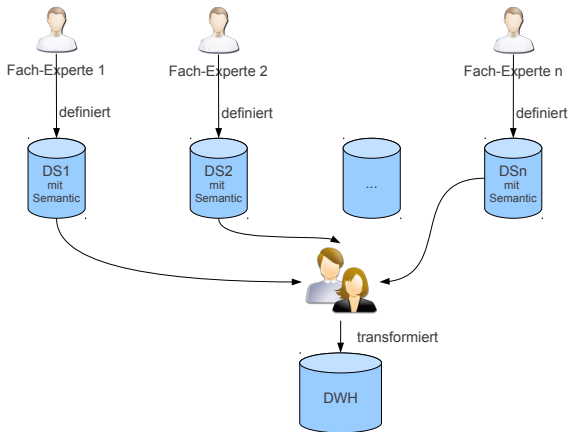
- Lösung von Sichten, Anwendungslogik, Algorithmen und Datenverarbeitung
- Asynchrone, skalierbare Verarbeitung
- Daten-getrieben
- agile Vorgehensmodelle/Entwicklungsprozesse

(Forschungs-)Aufgaben

- Trust / Legal Issues
- schnelle, skalierbare Suche
- Ontologien: Entwurf, Distribution, Reengineering
- Reasoning, KI
- Semantic (Web) Service
- Verbindung von Semantic Data, e.g., ping back
- Verbindung von SQL- und semantischen Datenbanken

Anwendungsgebiete

Business Intelligence / Data-Warehouse



Business Intelligence / Data-Warehouse

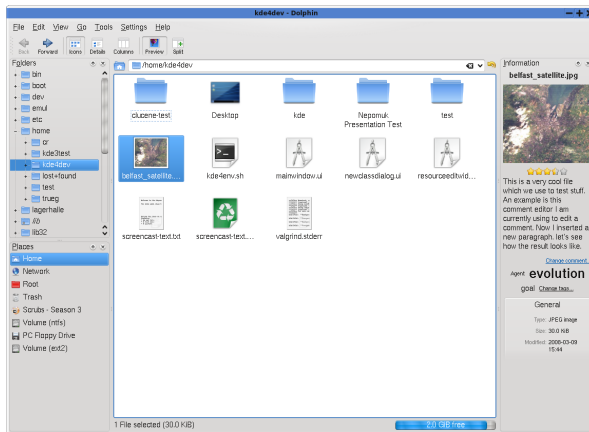
Beobachtung

- dynamische Integration wird möglich
- Wissen wird dezentral gewartet

Offene Fragen

- Reengineering von Ontologien
- Reasoning auf großen Ontologien

Semantic Desktop



<http://nepomuk.kde.org/>

Semantic Desktop

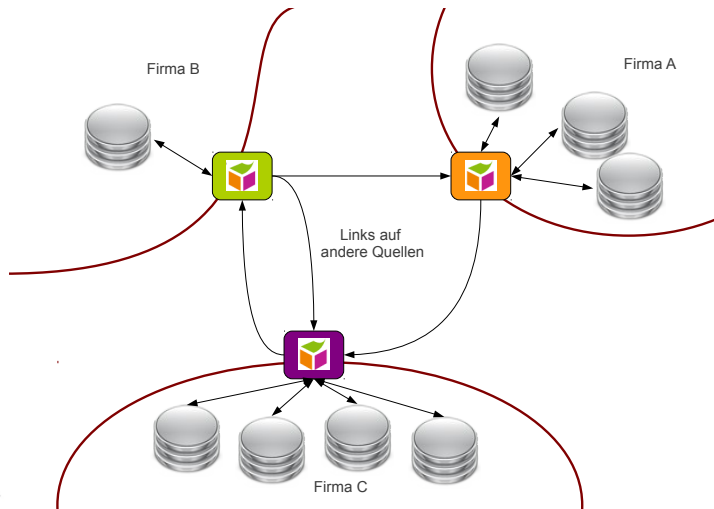
Beobachtung

- Hintergrundinformationen abrufbar
- Enthaltene Daten verknüpfbar
- Nutzer wird entlastet

Offene Fragen

- Integration in Betriebssystem bzw. Desktop Environment
- Einbettung in Dateiformate

Linked Enterprise Data



Linked Enterprise Data

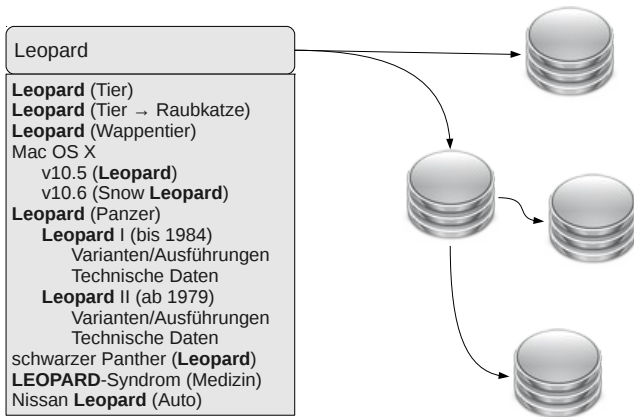
Beobachtung

- dezentrale Daten und Semantik
- dynamische Datenintegration
- Herkunft referenzierbar
- Ergänzung zu SOA

Offene Fragen

- gemeinsames/verständliches Vokabular
- Sicherheit und Zuverlässigkeit (z.B. SLA)

Semantic Search



Semantic Search

Beobachtung

- Beziehungen können abgefragt werden
- Wissen wird zusammengefügt
- Eigenschaften spezifizierbar (Untertyp, Obertyp, ...)
- „Antworten“ möglich

Offene Fragen

- Vokabulare
- Einordnung
- Bedienbarkeit

weitere Anwendungsbeispiele

- Wissensportale
- Integration in Kommunikationskanäle
 - TV, Hotline, Chat, ...
- E-Government/LOD
- lokale Geschäftsmodelle
- ...

Zusammenfassung

- Semantic-Web-Technologien sind als Standard verfügbar
 - RDF, OWL, SPARQL
- transparente Zugriff
- wohldefinierte Daten
- Wissensgenerierung wird möglich (Reasoning)
- Linked Open Data: verteilte Wissensdatenbank
- Web 3.0: liefert verarbeitbaren User Content

Zusammenfassung

Sir Tim Berners-Lee (2001)

„The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation.“

- neue Funktionalitäten möglich
- Unterstützung datengetriebener Prozesse
- schnelleres Prototyping, niedrigere Kosten/TTM

Zusammenfassung

Semantic-Web-Technologie liefert einen (r)evolutionären Fortschritt!

(vergleiche Price Waterhouse Coopers, Technology Forecast: Spring '09, 2009)