

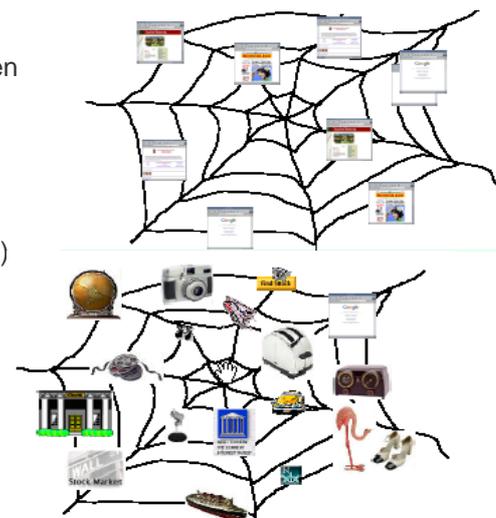


## ***Semantic Web Advanced Topics***

Knut Hinkelmann

## ***Entwicklung des WWW***

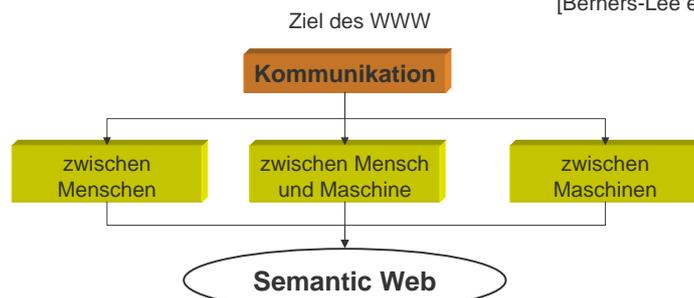
- Das WWW entwickelt sich von einem Anbieter von Dokumenten und Bildern...
- ... zu einem Anbieter von (Web) Services



## Semantic Web

*"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."*

[Berners-Lee et al., 2001]



## Semantische Informationen im WWW

- Heute:  
Informationen des Web sind maschinen-**lesbar**
- Zukunft:  
Informationen des Web sind maschinen-**verstehbar**
- Idee:  
Explizite Repräsentation der Bedeutung von Informationen als semantisches Netz

Semantic Web = Web + Data base technology  
+ Knowledge Representation



## Maschinen-lesbare Information (What it's like to be a machine)

### 林克昌 根留台灣 可能增高

在愛戴者熱心奔走之下，華裔名指揮家林克昌根留台灣的可行性又提升了幾分。兩廳院主任李炎、國家音樂廳樂團副團長黃奕明日前親赴林克昌、石聖芳寓所拜會，並提出多場客席邀約。此外，台灣省立交響樂團團長陳澄雄也早早「下訂」，邀請林克昌赴台中霧峰，從八月十日起訓練省交，為期長達一個月。

在台灣諸多公家樂團中，陳澄雄是以實際行動表達對林克昌肯定的樂界人士之一，曾多次公開表示對林克昌指揮才華的欽佩，而且幾乎每個樂季都邀請林克昌客席演出。

此外，林克昌上個月赴俄羅斯與頂尖的「俄羅斯國家管絃樂團」灌錄了柴可夫斯基晚期三大交響曲以及「羅密歐與朱麗葉」、「斯拉夫進行曲」、「義大利隨想曲」，最後的DAT母帶也在前兩天寄回台灣。製作人楊忠衡與林克昌試聽之後，都對錄音效果一尤其音質表現感到相當滿意，楊忠衡估計呈現了七分林克昌指揮神韻。

俄羅斯國家管絃樂團首席布魯尼日前也讚譽林克昌的指揮藝術有三大特點：一是控制自如的彈性速度；二是強烈的動態對比；三是宛如呼吸歌唱的旋律處理。這些對錄音師而言都構成很大挑戰。俄國錄音師雖然採用多軌混音，但定位、場面都有可觀之處。

## Maschinen-lesbare Information (Was die Teile bedeuten)

### 林克昌 根留台灣 可能增高

name

education

在愛戴者熱心奔走之下，華裔名指揮家林克昌根留台灣的可行性又提升了幾分。兩廳院主任李炎、國家音樂廳樂團副團長黃奕明日前親赴林克昌、石聖芳寓所拜會，並提出多場客席邀約。此外，台灣省立交響樂團團長陳澄雄也早早「下訂」，邀請林克昌赴台中霧峰，從八月十日起訓練省交，為期長達一個月。

在台灣諸多公家樂團中，陳澄雄是以實際行動表達對林克昌肯定的樂界人士之一，曾多次公開表示對林克昌指揮才華的欽佩，而且幾乎每個樂季都邀請林克昌客席演出。

work

此外，林克昌上個月赴俄羅斯與頂尖的「俄羅斯國家管絃樂團」灌錄了柴可夫斯基晚期三大交響曲以及「羅密歐與朱麗葉」、「斯拉夫進行曲」、「義大利隨想曲」，最後的DAT母帶也在前兩天寄回台灣。製作人楊忠衡與林克昌試聽之後，都對錄音效果一尤其音質表現感到相當滿意，楊忠衡估計呈現了七分林克昌指揮神韻。

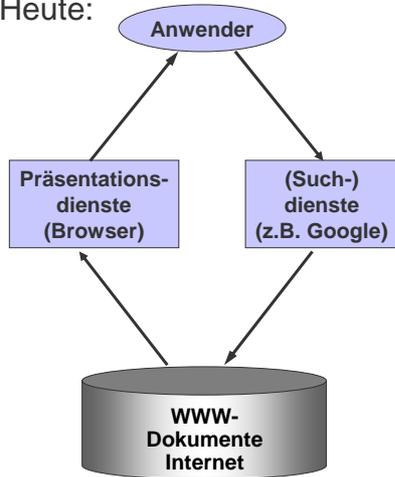
private

俄羅斯國家管絃樂團首席布魯尼日前也讚譽林克昌的指揮藝術有三大特點：一是控制自如的彈性速度；二是強烈的動態對比；三是宛如呼吸歌唱的旋律處理。這些對錄音師而言都構成很大挑戰。俄國錄音師雖然採用多軌混音，但定位、場面都有可觀之處。

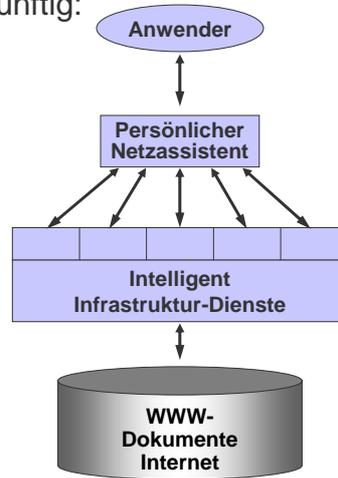
CV

## Evolutionsstufen des WWW: Intelligente Softwareassistenten

Heute:

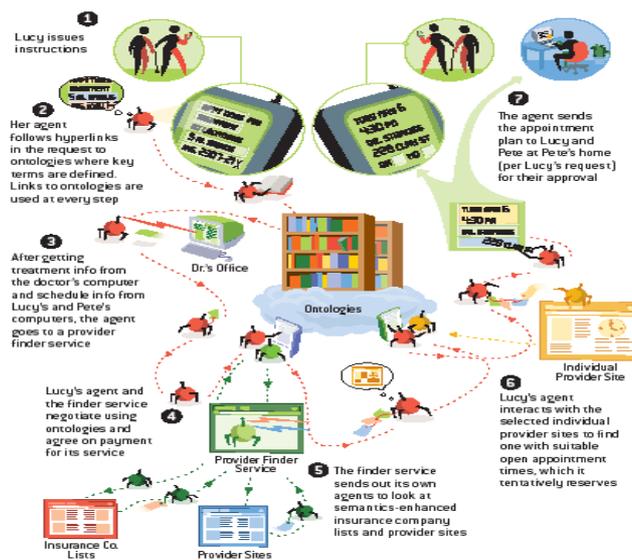


Zukünftig:



nach (Staab 2000)

## Ein Zukunftsszenario



## *The need for Ontologies*

- The future of the semantic web is linked with the future of ontologies in the semantic web
- Bottleneck for information is the rules and meanings about data defined precisely enough so that machines (...) can correctly interpret and correctly process that data.

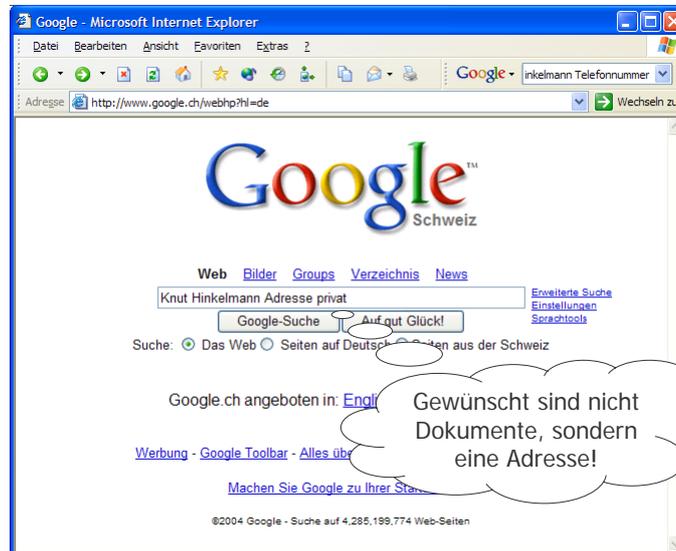


## *Anwendungsbeispiel E-Government*

- Finden der zuständigen Behörde
- One-Stop E-Government: Automatische Selektion von Services bei behördenübergreifenden Prozessen, z.B. Umzug
- Administration von Webservices: Welche Services sind von Gesetzesänderungen betroffen?



## Problem: Suche im WWW



## Anwendung im B2C: Shopbots

- Ein typisches Szenario: Ein Benutzer besucht mehrere Online-Shops, vergleicht die Angebote und bestellt Waren
  - ◆ zeitaufwändig!
- Shopbot helfen hier
  - ◆ Finden der passenden Shops
  - ◆ Finden des günstigsten Angebots Interpretation von Produktinformationen, Preisen, Lieferbedingungen und vergleichen sie mit den Anforderungen des Benutzers
  - ◆ Informationen über die Reputation des Shop werden berücksichtigt
  - ◆ Ausgefeilte Shopbot könnten sogar automatisch Verhandlungen führen

## **Anwendung im B2B: Unternehmensübergreifende Prozesse**

- Integration von Unternehmen basiert heute auf EDI
  - ◆ isolierte Technologie
  - ◆ Jede B2B-Kommunikation erfordert spezielle Programmierung
  - ◆ Standards nur auf technischer und syntaktischer Ebene
- Unterstützung durch Semantic Web
  - ◆ Unterschiedliche Terminologien der Partner
  - ◆ Datenaustausch mit Übersetzungsservice
  - ◆ Automatisierte Bestellung und Verhandlungen durch Software Agents



## **Anwendung Knowledge Management**

- Beschränkungen für Wissensmanagement
  - ◆ Wissen liegt oft in unstrukturierter Form vor: Text, Garphiken etc.
  - ◆ Informationssuche: In der Regel nur Volltextsuche
  - ◆ Informationsextraktion: Interaktion mit Menschen notwendig, um Information aus Texten zu extrahieren
  - ◆ Verwaltung von Information: Inkonsistenzen in der Terminologie, veraltete Information
- Unterstützung durch Semantic Web
  - ◆ Semantische Suchanfrage
  - ◆ Organisation von Wissen nach Inhalts und Kontext
  - ◆ Erkennen relevanter Information in Dokumenten:  
*Finden von Informationen statt Dokumenten*



## Probleme des Information Retrieval

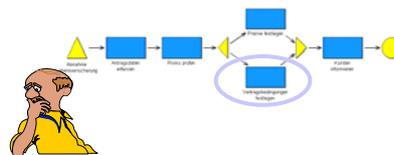
Die natürliche Sprache ist sehr variabel. Die Suchanfrage und die Wörter in dem Dokument können unterschiedlich sein, obwohl sie das gleiche ausdrücken. Die wesentlichen Probleme sind:

- **Wortformen:**  
In einem Dokument können die Wörter in verschiedenen Formen vorkommen. Eine Anfrage nach "Firma" sollte auch Dokumente mit dem Wort "Firmen" finden
- **Wortbedeutungen**  
Eine Wort kann verschiedene Bedeutungen haben und umgekehrt kann eine Sache durch verschiedene Wörter ausgedrückt werden. Bei der Suche nach "Firma" sollten auch Dokumente mit dem Wort "Unternehmen" gefunden werden.
- **Formulierungen/Komposita**  
Der gleiche Sachverhalt kann auf verschiedene Arten ausgedrückt werden, z.B. kann statt "Versicherungsprämie" auch "Prämie für die Versicherung" im Text stehen.



## Grenzen der Volltextsuche

- Volltextsuche nur für coded-Information anwendbar
- Volltextsuche ist in der Regel rein syntaktisch, ohne Kenntnis der Wortbedeutungen
- Kriterien für Relevanzbewertung stehen nicht immer im Text, z.B.
  - ◆ Dokumenttyp (Fachartikel, Präsentation, ...)
  - ◆ Kontext (Projekt, Produkt, Kunde, ...)
  - ◆ Unterscheidung Sender/Empfänger eines Briefes
  - ◆ Nützlichkeit für eine Aufgabe
  - ◆ Adäquatheit für den Benutzer



*Zusätzliche Meta-Information und -wissen überwindet die Grenzen*



## Zusätzliches Wissen

- Zusätzliches Wissen über
  - ◆ Benutzer
  - ◆ Kontext
  - ◆ Inhalt (Domäne)kann genutzt werden, um relevantes Wissen zu finden und die Relevanz von Wissen beurteilen zu können
- Wissen über ...
  - ... Kontext und Inhalt wird in Wissensstrukturen repräsentiert (→ Wissensorganisation)
  - ... Benutzer kann in Benutzerprofilen repräsentiert werden

## Bildersuche: Jaguar

The screenshot shows a Google Image search for 'Jaguar' in Microsoft Internet Explorer. The search results are displayed in a grid format. The first row shows three images: a jaguar, a silver Jaguar car, and another jaguar. Below each image is its filename and source URL. The second row shows four images: a jaguar, a jaguar, a Jaguar car, and a jaguar. A text box on the right side of the browser window contains the text 'What do you mean:' followed by three radio button options: 'animal', 'car', and 'airplane'. The browser's address bar shows the search URL: 'http://images.google.ch/images?sourceid=navclient&hl=de&ie=UTF-8&oe=UTF-8&q=Jaguar'. The browser's search bar contains the word 'Jaguar'.

## Suche im WWW

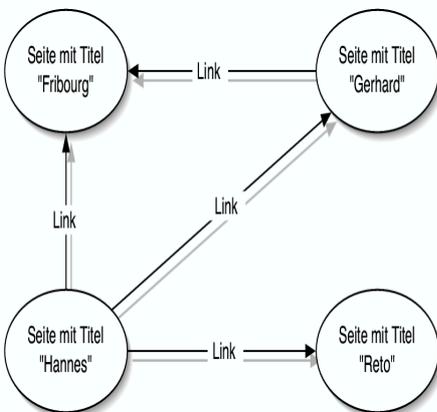
Was kostet ein Jaguar XJ?

The screenshot shows a Google search for "Jaguar XJ Preis". The search results include a link to "Jaguar XJ Preis Schweiz". Below the search results, there are two browser windows. The left window shows the Jaguar website's price list for the XJ model. The right window shows a search result for "Jaguar XJ" with a price of EUR 24,95 and a link to "Jaguar XJ".

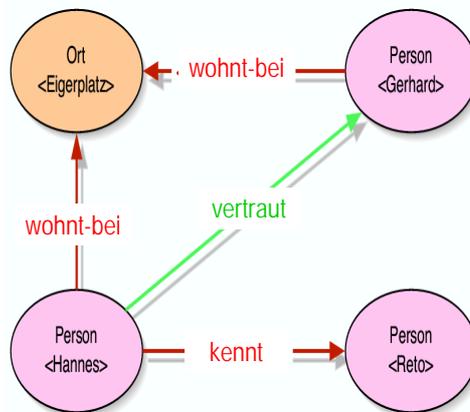
XJ	Preis inkl. MWSt.	Preis exkl. MWSt.
3.5 V8 Classic	98.000,00 CHF.	91.078,00 CHF.
3.5 Executive	110.000,00 CHF.	102.230,00 CHF.
4.2 Executive	115.000,00 CHF.	106.877,00 CHF.
XJR 4.2 V8 S/C	136.000,00 CHF.	126.394,00 CHF.
Super V8	156.000,00 CHF.	144.981,00 CHF.

## Das Web: Hyperlinks

Das Web: Dokumente und Hyperlinks



Semantic Web: Ressource (Objekte) haben Bedeutung und Beziehungen



## HTML – Informationsverarbeitung durch den Menschen



Feste Markup-Sprache für

- ◆ Darstellung
- ◆ Links

```

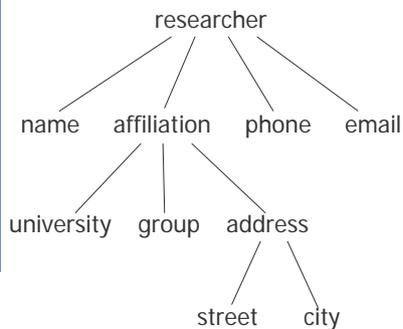
...
<font size=+1>Knut Hinkelmann</font>
<p> <a href="http://www.fhso.ch">
Fachhochschule Solothurn Nordwestschweiz </a> <br>
Wirtschaftsinformatik <br>
Riggenbachstrasse 16 <br>
4600 Olten <br>
<a href="mailto:knut.hinkelmann@fhso.ch">
knut.hinkelmann@fhso.ch</a> <br>
<b>Phone</b> ++41 62 388 25 16 <br>
<b>Fax</b> ++41 62 388 25 93
</p>
...

```

## XML – Strukturierung des Inhalts

Benutzerdefinierbares und domänenspezifisches Markup:

- ◆ Elemente (Label, Attribute, Inhalt)
- ◆ Struktur



```

<researcher>
<name> Knut Hinkelmann </name>
<affiliation>
  <university> Fachhochschule Solothurn
  Nordwestschweiz</university>
  <group> Wirtschaftsinformatik</group>
  <address>
    <street> Riggenbachstrasse 16 </street>
    <city> 4600 Olten </city>
  </address>
</affiliation>
<phone > ++41 62 388 25 16 </phone>
<email> knut.hinkelmannfhso.ch </email>
</researcher>

```

## XML: Der Maschine ist die Bedeutung nicht zugänglich (Was die Maschine wirklich sieht)

林克昌 擬留台灣 可能增高

<εδουχαιον>

<ωορκ>

<πριωατε>

< νομε >

< Xς >

在饒幸者熱心奔走之下，華裔指揮家林克昌留台灣的可行性又提升了幾分。兩院系主任李炎、國家音樂廳副團長黃炎明日前親赴林克昌、石聖芳寓所拜會，並提出多場客席邀約。此外，台灣省立交響樂團團長陳意雄也早早「下訂」，邀請林克昌赴台中轉棒。從八月十日起訓練省交，為期長達一個月。

在台灣諸多公家樂團中，陳意雄是以實際行動表達對林克昌肯定的樂界人士之一。曾多次公開表示對林克昌擁護才華的欽佩，而且幾乎每個樂季都邀請林克昌客席演出。

此外，林克昌上個月赴俄羅斯聖彼得堡的「俄羅斯國家管絃樂團」灌錄了柴可夫斯基晚三大交響曲以及「羅密歐與朱麗葉」、「斯壯大進行曲」、「義大利遺囑曲」。最後的D.A.T. 專輯也在前兩天寄回台灣。製作人楊忠衛與林克昌試聽之後，都對錄音效果一尤其音質表現感到相當滿意，楊忠衛估計呈現了七分林克昌指揮神韻。

俄羅斯國家管絃樂團首席布魯尼日前也讚譽林克昌的指揮藝術有三大特點：一是控制自如的彈性速度；二是強烈的動感對比；三是宛如呼吸般的旋律處理。這些對錄音師而言都構成很大挑戰。俄國錄音師雖然採用多軌混音，但定位、場面都有可觀之處。

## XML alleine ist nicht die Lösung Beispiel E-Business

order information

```
<product>
  <type> car </type>
  <model> Jaguar XJ </model>
  <price> $ 73'000 </price>
</product>
```

Bestellinformation

```
<Auto>
  <Name> Jaguar XJ </Name>
  <Preis> CHF 98'000 </Preis>
</Auto>
```

- Nur für begrenzte Zusammenarbeit
  - ◆ Agenten in eine kleinen, stabilen Gemeinschaft
  - ◆ Seiten in einem begrenzten Intranet/Extranet
- Einigung auf Bedeutung der Tags

## XML ist nicht genug (1)

```
<skill-database>
  <people>
    <Person>
      <name>Markus</name>
      <knowHow>SGML</knowHow>
    </Person>
    <Hacker>
      <name>Jürgen</name>
      <pgp>CB FC A8 17</pgp>
      <knowHow>SGML</knowHow>
      <knowHow>Java</knowHow>
    </Hacker>
    <Person name="Rainer">
      <knowHow>Mike</knowHow>
    </Person>
  </people>
```

```
<seminars>
  <Seminar topic="SGML">
    <attendant>
      <name>Dieter</name>
      <name>Robert</name>
      <name>Rainer</name>
    </attendant>
  </Seminar>
</seminars>
<skill-database>
```

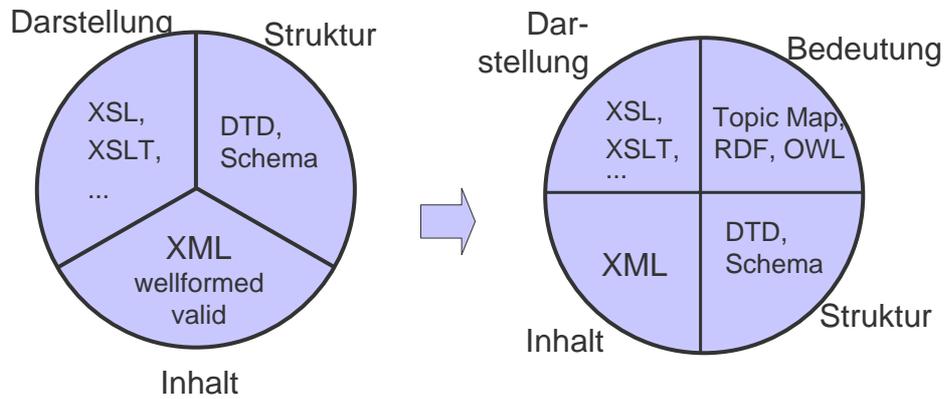
Anfragen:

- Gib mir alle Personen
- Gib mir alle Personen mit KnowHow in SGML

## XML ist nicht genug (2)

- Anfragen:
  - ♦ Gib mir alle Personen:
    - XQL: //person/name
    - liefert nur <name>Markus</name>
  - ♦ Gib mir alle Personen, die sich mit SGML auskennen!
    - XQL: //person[knowHow="SGML"]/name
    - liefert nur <name>Markus</name>
- Was fehlt: **Hintergrundwissen**
  - ♦ Hacker sind auch Personen
  - ♦ Seminarteilnehmer sind auch Personen
  - ♦ Seminarteilnehmer haben Wissen über den Inhalt des Seminars

## Bedeutung als zusätzliches Dokumentmerkmal



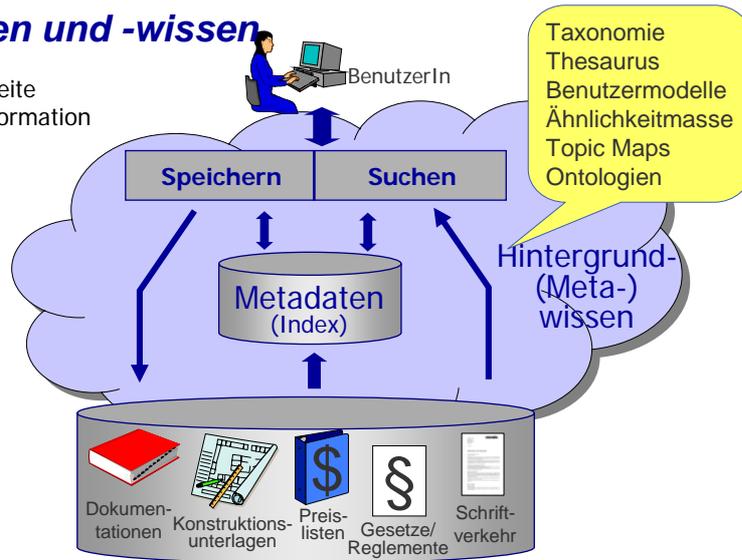
## Metadaten und -wissen

Unternehmensweite  
Nutzung von Information

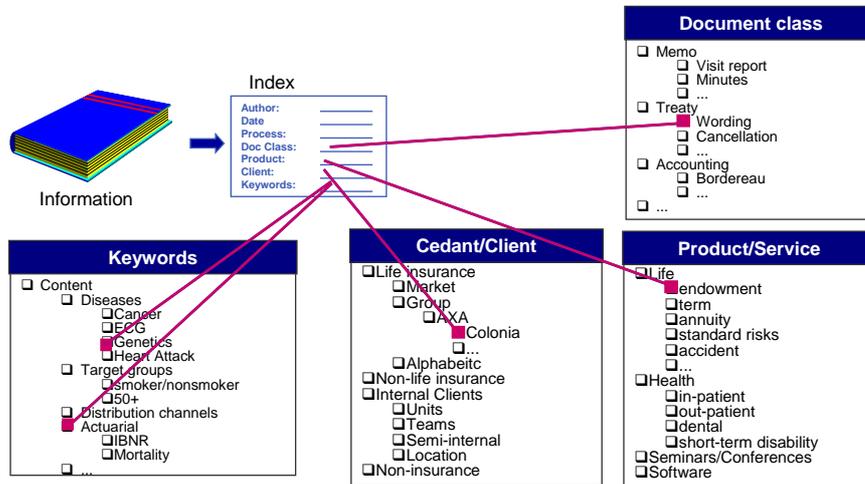
Dienste

Strukturierung &  
Beschreibung

Ressourcen



## Beispiel: Metadaten bei einer Rückversicherung



Die Wissensstruktur selbst enthält viel Wissen über das Unternehmen  
Einheitliche Begriffssysteme und Strukturen fördern Wissensnutzung

## Wissensorganisation

- Die Wissensorganisation gibt einen Überblick über die Inhalte eines Informationssystems
  - ◆ erleichtert die Orientierung gerade auch für neue Benutzer
- Die Wissensstrukturen ...
  - ... **repräsentieren** Wissen über Konzepte/Begriffe in verschiedenen Kontexten
  - ... **verweisen** auf das zu bestimmten Konzepten relevante Wissen (Relevanzmodell)
- Von besonderer Bedeutung sind intersubjektiv validierte Wissensstrukturen, d.h. soziale und kulturelle Konstrukte:
  - ◆ die (Welt-)Sicht, die eine Gemeinschaft über einen Bereich teilt,
  - ◆ wie die Gemeinschaft denkt, dass Gegenstände aufeinander bezogen sind
- Daneben gibt es auch die Möglichkeit, Wissensstrukturen (und Benutzerprofile) durch Analyse von Dokumenten und Benutzerverhalten automatisch herzuleiten (Text Mining)

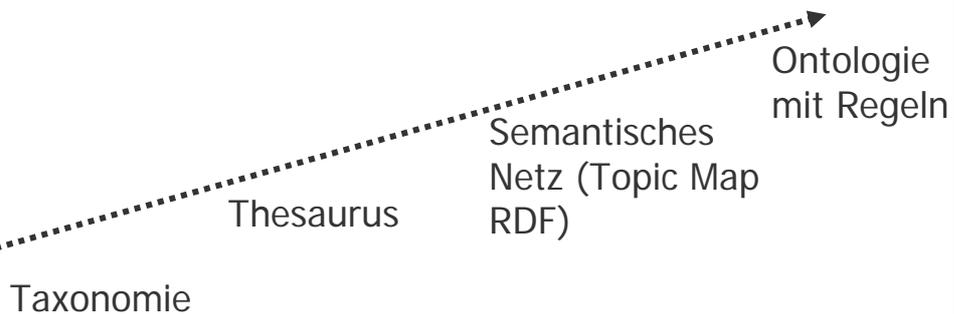
## Methoden zur Wissensorganisation

- **Thesaurus**
  - ◆ Beziehungen zwischen Wörtern/Schlüsselwörtern
- **Klassifikationssystem/Taxonomien**
  - ◆ Kategorien von Informationen
- **Ontologie/Semantisches Netz**
  - ◆ Beziehungen zwischen Konzepten

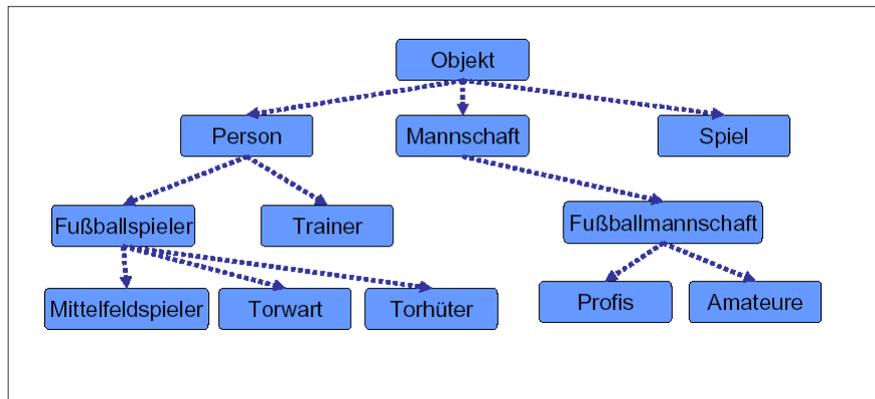
*Ein Unternehmensgedächtnis in einem Lernenden Unternehmen benötigt eine flexible Begriffsstruktur. Deswegen benötigen Sie Kenntnisse in Wissensorganisation."*  
A. Sigel, Uni Bonn



## Wissensorganisation: Methoden

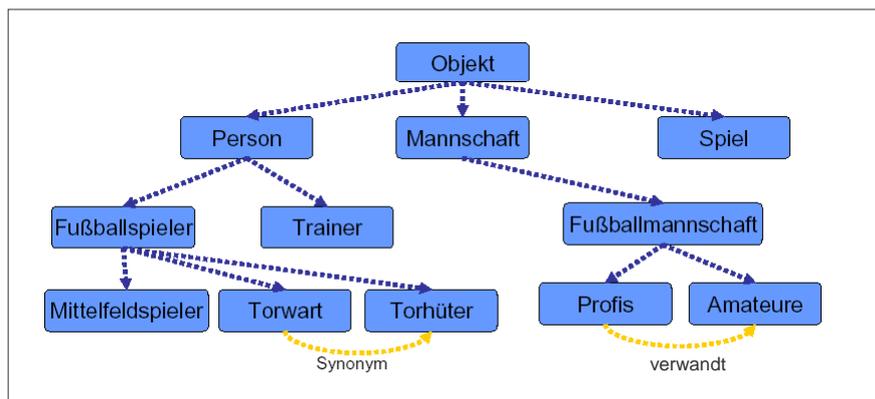


## Taxonomie



- Klassifikation, Hierarchie
- Einordnung von Elementen in Kategoriensystem anhand ihrer definierten Beziehungen zueinander

## Thesaurus

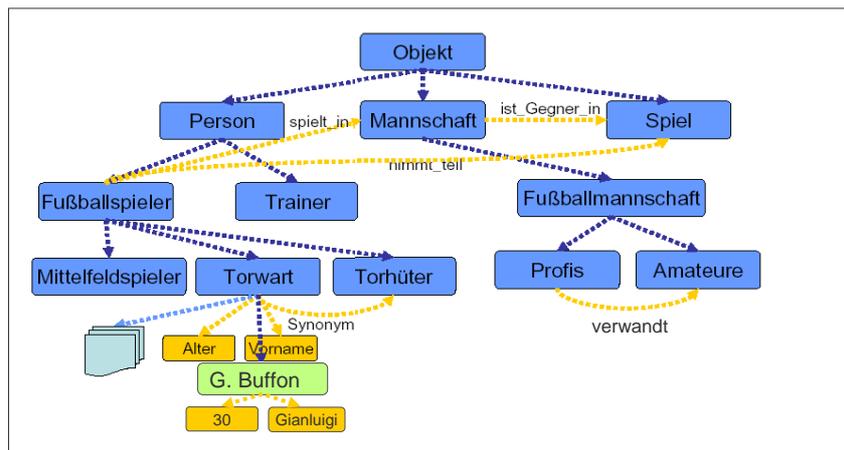


- Terminologie einer Domäne
- wenige, fest vorgegebene Beziehungen (Ober-/Unterbegriff, Synonym, verwandter Begriff, ...)

## Thesaurus: Vorgegebene Relationen

deutsch		englisch	
Kurzzeichen	Bezeichnung	Kurzzeichen	Bezeichnung
<b>Hierarchierelationen</b>			
TT	Top Term (allgemeinster Begriff)	TT	Top term
OB	übergeordneter Begriff (Oberbegriff)	BT	Broader term
UB	untergeordneter Begriff (Unterbegriff)	NT	Narrower term
<b>Bei Differenzierung der Hierarchierelation nach Abstraktion und Aggregation</b>			
OA	Oberbegriff Abstraktionsrelation	BTG	Broader term generic
UA	Unterbegriff Abstraktionsrelation	NTG	Narrower term generic
SP	Verbandbegriff	BTP	Broader term partitive
TP	Teilbegriff	NTP	Narrower term partitive
<b>Äquivalenz- und Assoziationsrelationen:</b>			
BS	Benutztes Synonym oder Quasi-Synonym	USE	Use
BF	Benutzt für Synonym oder Quasi-Synonym	UF	Used for
VB	verwandter Begriff	RT	Related term
BK	Benutzte Kombination von Einfachdeskriptoren	USE	Use
KB	Benutzt in Kombination von Einfachdeskriptoren	UFC	Used for combination

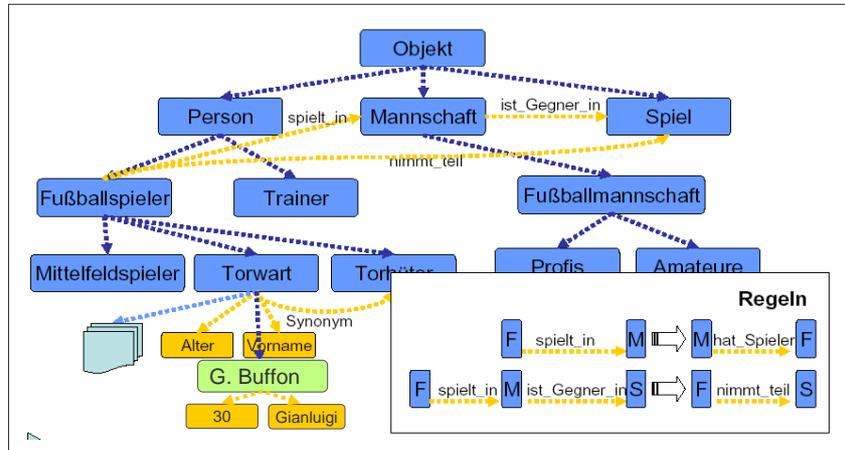
## Semantisches Netz: RDF/Topic Map



Es gibt zwei unabhängig entwickelte Standards:

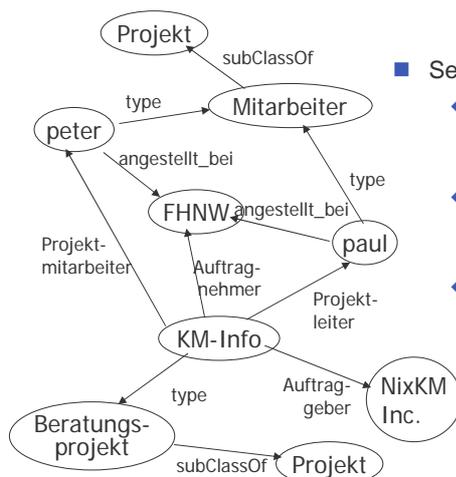
- Topic Maps: Topics, Relationen und Occurrences
- RDF, RDF Schema: Klassen/Instanzen, Relation/Attribute, Ressourcen

## Ontologie (mit Regeln)



- Ontologie: Konzepte/Instanzen, Relationen/Attribute, Ressourcen (W3C Standards: RDF, RDF Schema, OWL)
- Regeln: Logische Regeln für weitere Ableitung (W3C Standard: SWRL)

## Semantische Netze



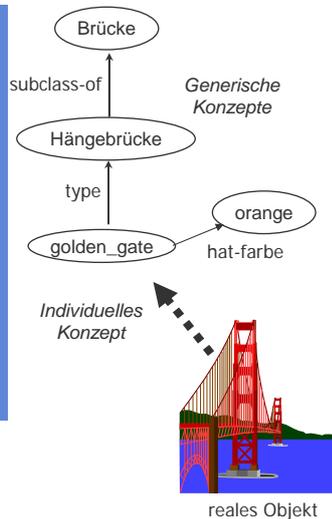
- Semantische Netze bestehen aus :

- ◆ **Konzepte:** Begriffe der Welt, stehen sowohl für Klassen als auch Instanzen (graphisch als Knoten dargestellt)
- ◆ **Relationen:** semantische Beziehungen zwischen Konzepten (dargestellt als gerichtete Kanten)
- ◆ **Axiome:** legen fest, wie Relationen zu interpretieren sind und welche Schlüsse daraus gezogen werden können  
In Axiomen ist z.B. festgelegt, welche Relationen Vererbungsrelationen sind

*In den Knoten selbst ist kein Wissen enthalten, alles Wissen über einen Knoten wird durch die Verbindungen repräsentiert, die von ihnen ausgehen.*

## Formalisierung von Semantischen Netzen

### Unterscheidung von Klassen und Instanzen



- Man unterscheidet man zwei Arten von Knoten:
  - ◆ generisches Konzept (Klasse)
  - ◆ individuelles Konzept (Instanz)
- Daraus ergibt sich die Unterscheidung von mindestens drei Arten von Kanten
  - ◆ strukturelle Beziehungen:
    - **type** (auch instance-of)  
Relation zwischen individuellem und generischem Objekt
    - **SubClassOf** (auch is-a)  
Relation zwischen generischen Objekten  
Vererbungsrelation
  - ◆ nicht-strukturelle Beziehungen
    - beliebige 2-stellige Relationen, z.B.  
`hat-farbe(golden_gate, orange)`



## Ontologie

*"People can't share knowledge, if they don't speak a common language."*

Thomas H. Davenport

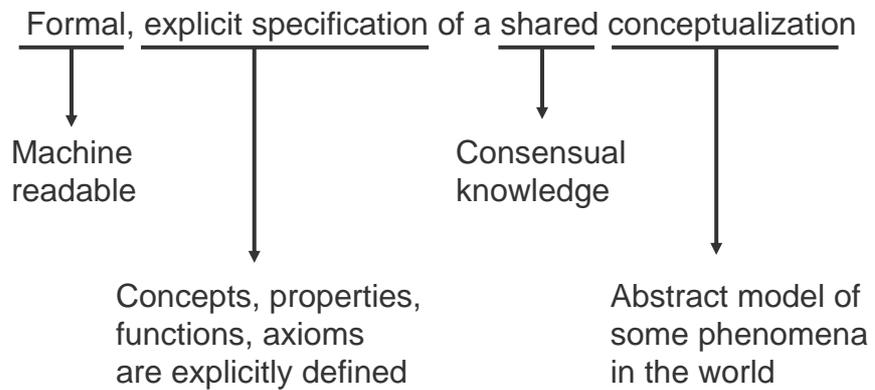
### ■ Definition

*"An ontology is a shared understanding of some domain that can be communicated between people and application systems"*

- Eine Ontologie ist ein formales Modell eines relevanten Teils der Welt, über deren Begriffe und Zusammenhänge eine Gruppe von Experten/Nutzern Einigkeit erzielt
- Eine Ontology beschreibt die Thematik durch Konzepte, Instanzen, Relationen, Attribute und Axiome.
- Konzepte der Hierarchie sind in einer Taxonomie organisiert, über die Vererbungsmechanismen angewendet werden können.

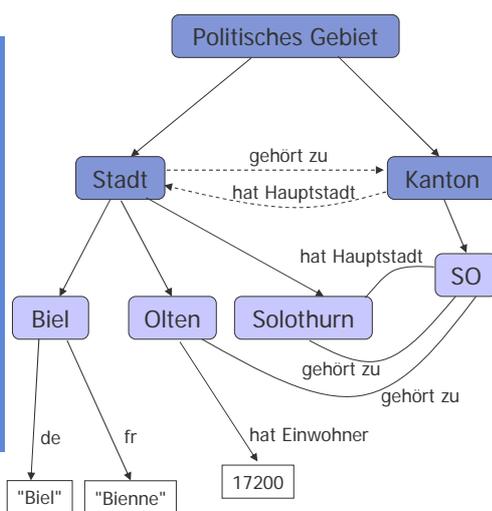


## What is an ontology?



(Studer 1998)

## Grundelemente von Ontologien



- Konzept (Klasse)
- Instanz
- Relation  
(Property mit Instanzen als Wert)
- Attribute  
(Property mit Literalen/ Daten als Wert)
- Regel

## ■ **Ontologien**

- Ontologien definieren neben Konzepten und Relationen auch Axiome, z.B.

- ◆ Weitergehende Relationen zwischen Klassen  
(z.B. Disjunktheit – Klassen, die keine gemeinsame Instanz haben)
- ◆ Charakteristika von Properties/Relationen  
(z.B. Value Restriction Kardinalität, Symmetrie, Transitivität, Reflexivität)

## ■ Beispiele:

- ◆ Disjunktheit
  - Professoren und Studenten sind disjunkt
- ◆ Value Restriction
  - Nur Professoren können Vorlesungen halten
- ◆ Kardinalität:
  - Jede Lehrveranstaltung hat mindestens einen Dozierenden



## **Bezeichner vs. Konzept**

- Objekte werden benannt ("Ich Tarzan – du Jane")
- Genaue Bezeichnung spielt keine Rolle, z.B.
  - ◆ Biel – Bienne
- Wichtig ist
  - ◆ Jedes Konzept, Instanz und Relation hat eindeutige ID
  - ◆ Man hat sich darauf geeinigt, wofür die Konzepte Instanzen und Relationen stehen, z.B.
    - SO ist Kanton Solothurn
    - Solothurn ist die Stadt
    - Biel stehen für die Stadt Biel/Bienne
  - ◆ Daraus folgt: Synonyme werden nicht als zwei unterschiedliche Dinge betrachtet, sondern als unterschiedliche Bezeichner für das gleiche Konzept



## *Analogien zu anderen Techniken*

<b>Ontologie</b>	<b>Datenbank</b>	<b>OOP</b>
Konzept	Tabelle	Klasse
Instanz	Datensatz	Instanz
Relation	Relation	Methode
Attribut	Attribut	Attribut
Regel (Axiom)		

