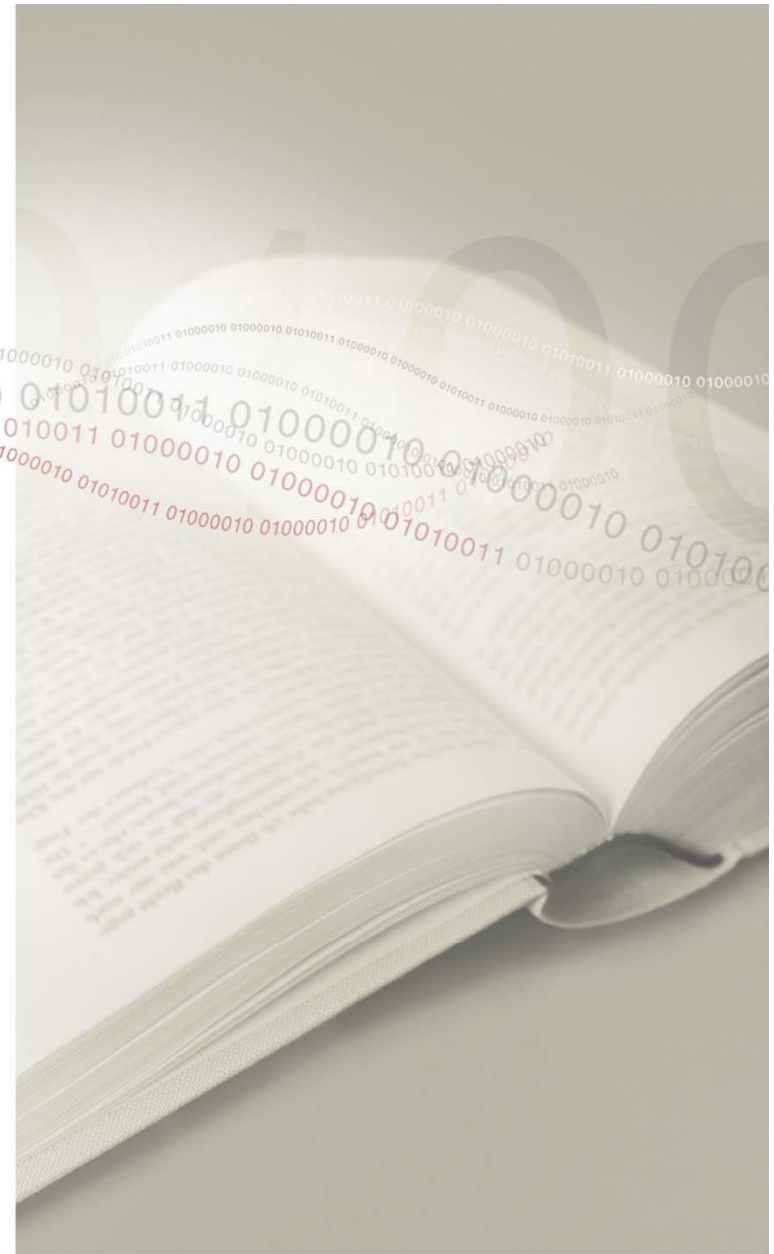
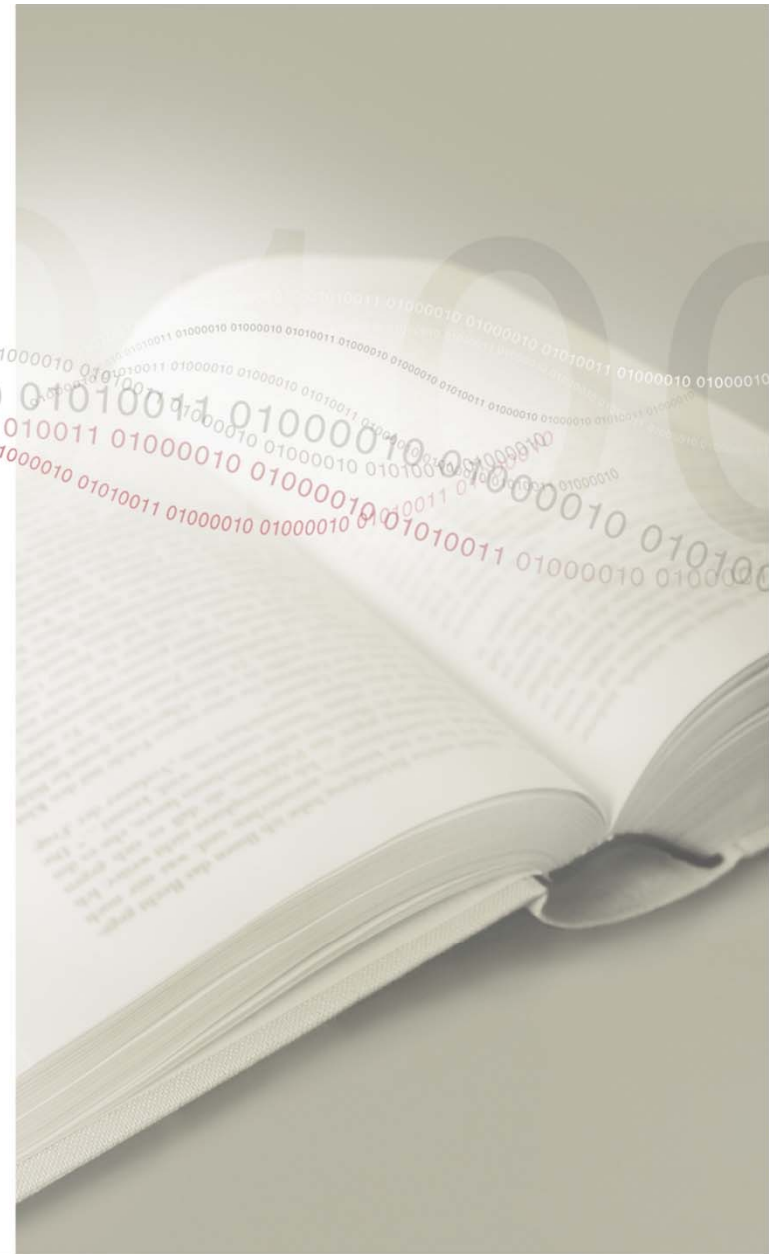


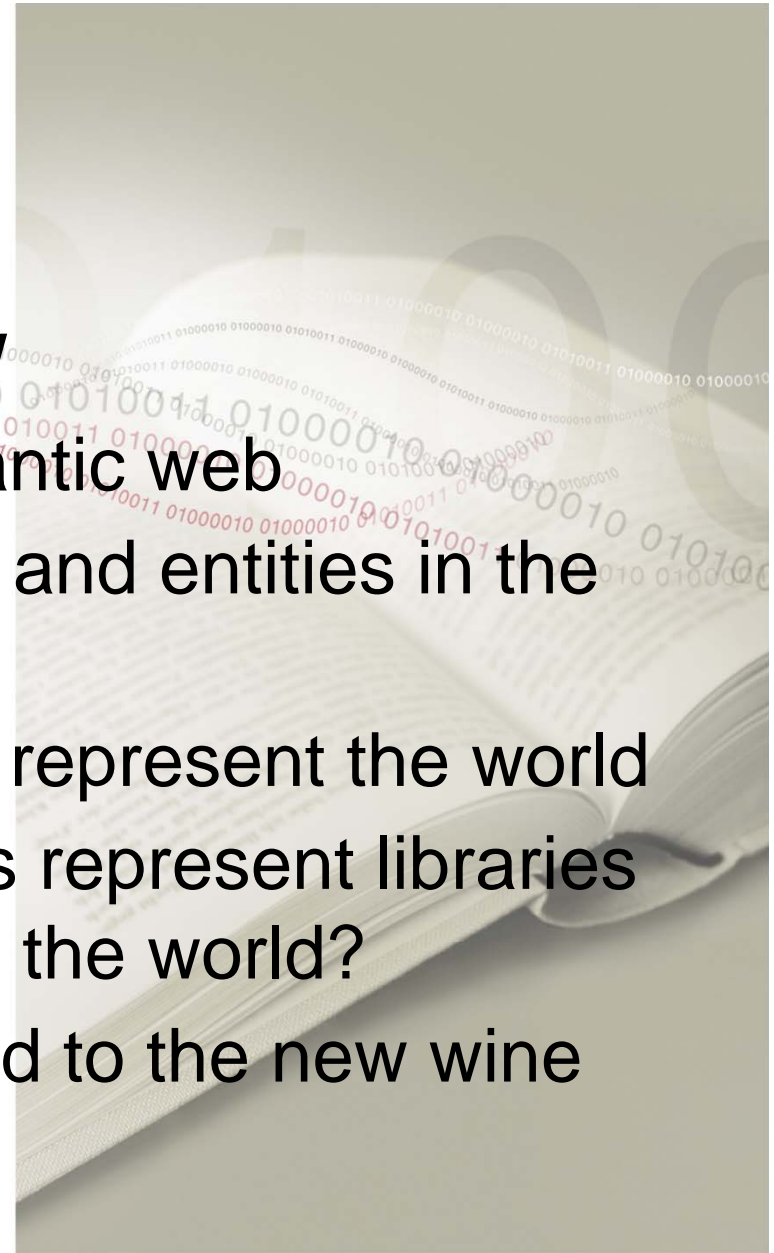
## **Semantic Web: New Wine in Old Wineskins or why Aristotle knew everything but Google makes money out of it**



## Or: Semantic Web from Dummies for Dummies and why it is relevant for libraries



- Overview:
  - 1. The old wineskin: WWW
  - 2. The new wine: the semantic web
  - 3. Aristotle: fixing meaning and entities in the world
  - 4. Google: how computers represent the world
  - 5. Libraries: can computers represent libraries and can libraries represent the world?
  - 6. What libraries should add to the new wine



- What you will miss:
  - Detailed technical aspects of the semantic web
  - Overview about recent developments and applications outside of libraries

## The old wineskin – WWW

- The brave new world of the WWW – boon and bane



## The old wineskin – WWW

- **World Wide Web – boon and bane:**
- **Boon:**
  - Information is globally available
  - Information is always up to date (or seems to be at least)
  - Many services are available online (online-shopping, online gaming, online chat, online dating...)
  - Many new possibilities of communication (mail, instant messaging, chat, facebook, twitter...)
  - Due to Web 2.0 the WWW is a component of democracy

## The old wineskin – WWW

- **The bane:**
- Too much information – too much communication
- People get lost in cyberspace
- Only digital natives have the chance to survive the information overload
- One basic example:
- Identification of desired information:
  - ⇒ There is no difference between what you say and what you mean: The search engine is stupid!
  - ⇒ A whole sentence or assertion doesn't play any role for the search engine

# Semantic under-determination of a search term (ambiguity)

verbreitung - Beschreibung - Lebensweise - Fortpflanzung  
[de.wikipedia.org/wiki/Hasen](http://de.wikipedia.org/wiki/Hasen) - Im Cache - Ähnliche Seiten

## [Echte Hasen – Wikipedia](#) ☆

Die Echten **Hasen** (*Lepus*) sind eine Gattung der Säugetiere innerhalb der Familie der **Hasen** (Leporidae). Die Gattung umfasst etwa 30 Arten, von denen sechs ...

[de.wikipedia.org/wiki/Echte\\_Hasen](http://de.wikipedia.org/wiki/Echte_Hasen) - Im Cache - Ähnliche Seiten

[+ Weitere Ergebnisse anzeigen von de.wikipedia.org](#)

## Videos zu **hasen**



### [Der Igel und der Hase](#)

4 Min. - 5. Febr. 2009

Hochgeladen von Hirsalto  
[youtube.com](http://youtube.com)



### [Meine Hasen](#)

4 Min. - 28. Febr. 2008

Hochgeladen von GAMMY1000  
[youtube.com](http://youtube.com)

## [Alles über Igel und \*\*Hasen\*\* für Kinder - wir-in-berlin.de: Das ...](#) ☆

Mehr über Igel und **Hasen** erfahren und tolle Sachen machen. Ein Angebot für Kinder, Eltern und Schule.

[www.wib.be.schule.de/haseigel/index.htm](http://www.wib.be.schule.de/haseigel/index.htm) - Im Cache - Ähnliche Seiten

## News zu **hasen**



MTV

### [Paris Hilton rettet 20 Häschen](#) ☆ - vor 12 Stunden gefunden

Jüngst rettete sie 20 **Hasen**, die eigentlich an Schlangen verfüttert werden sollten, berichtet das Boulevardblatt «The Sun» in seiner Online-Ausgabe. ...

[sueddeutsche.de](http://sueddeutsche.de) - 100 weitere Artikel »

### [Mesut Özil bei Real: Madrid verliebt sich in den deutschen \*\*Hasen\*\*](#) ☆ -

STERN.DE - 825 weitere Artikel »

### [Ein Besuch bei Claude Lanzmann: Nie hat die Zeit aufgehört, nicht ...](#) ☆ -

FAZ - Frankfurter Allgemeine Zeitung - 13 weitere Artikel »

## [Kostenlos Chatten im \*\*Hasen\*\* Chat](#) ☆

Partnersuche war noch nie so einfach. Lesen und schreiben sie Kontaktanzeigen Kostenlos beim **Hasen** Chat oder Chatten sie mit anderen Singles.

[www.hasenchat.de/](http://www.hasenchat.de/) - Im Cache - Ähnliche Seiten

## [Hasen](#) ☆



# Removing ambiguity: searching with a whole sentence

Web Bilder Videos Maps News Shopping E-Mail Mehr gillitzer@bsb-muenchen.de | Webprotokoll | Einstellungen | Abmelden

**Google**   Google Instant ist aktiviert.

Ungefähr 87.300 Ergebnisse (0,23 Sekunden) Erweiterte Suche

- Alles**
- Bilder
- Videos
- Mehr

---

**Das Web**  
Seiten auf Deutsch

Mehr Optionen

[Sabrina Mockenhaupt schlägt sich als \*\*Hase\*\* gut | Siegerland-Sport](#) ☆  
Wo der Igel und seine Frau den **Hasen gepflegt** aufs Kreuz legen und durch fiese Tricks einen Rennverlauf vortäuschen, wie ihn selbst Jan Ulrich nicht ...  
[www.siegerland-sport.de/.../hase-mocki-schlagt-sich-wacker/](#) - Im Cache

[Kaninchen - Erfahrungsbericht - Über Kaninchenpflege](#) ☆  
Also weiß ich auch aus eigener Quelle wie ein Kaninchen **gepflegt** werden muss etc.. Ich spreche von **Kaninchen**, die draußen leben. Auch wenn es etwas teurer ...  
[www.ciao.de > ... > Nager & Kleintierarten](#) - Im Cache - Ähnliche Seiten

[Zwergkaninchen - Erfahrungsbericht - Meine Balkon-Kaninchen](#) ☆  
23. Sept. 2003 ... Kaninchen müssen der Gesundheit zu Liebe auch **gepflegt** werden. ... Meine **Kaninchen** dürfen nicht bei mir in der Wohnung rumhopsen. ...  
[www.ciao.de > ... > Nager & Kleintierarten](#) - Im Cache - Ähnliche Seiten

[Wurzeln im Volksglauben : Jenseits von \*\*Hasen\*\* und bunten Eiern](#) ☆  
26. März 2010 ... Auf den Philippinen, wo angeblich auch der Brauch mit **Hasen** und bunten Ostereiern **gepflegt** wird, ziehen Eltern ihre kleinen Kinder beim ...  
[www.epochtimes.de/articles/2010/03/26/560792.html](#) - Im Cache

[Hasen und Igel - Informationen zum Lernangebot auf dem Berliner ...](#) ☆  
Mehr über **Hasen** und Igel erfahren und tolle Sachen machen. ... früher angenommen wurde, dass man Igel hilft, wenn sie im Winter zu Hause **gepflegt** werden, ...  
[www.schule.de/bics/son/wir-in-berlin/haseigel/.../index.htm](#) - Im Cache

[Der Kranke \*\*Hase\*\* – verrückt nach Linz](#) ☆  
Er ist das inoffizielle Wahrzeichen der Stadt: Eben noch in der Grottenbahn von den Zwergen **gepflegt**, zieht es den kranken **Hasen** in die Kulturhauptstadt. ...  
[yamaikido.klafboeck.at/hase.htm](#) - Im Cache

## The old wineskin – WWW

- **What people can and computers cannot do – some main shortcomings:**
- **Information retrieval:**
  - String search yields too many and too few results:
    - homonymy, synonymy, vagueness, context ...
- **Information extraction:**
  - Taking specific information from texts or webpages presupposes until now human understanding
- **Information integration:**
  - Search engines cannot integrate information pieces from different documents into one new meaningful document
- **Drawing conclusions:**
  - Search engines are not able to draw conclusions from different pieces of information

## The old wineskin – WWW

- **The problem:**
  - If the amount of information is so big that people cannot handle it any longer intellectually they need help by automatic information processing simulating human understanding
- **Possible solutions:**
  - Artificial intelligence (AI): Software capable to process natural language with computerlinguistic methods simulating human understanding literally
  - Data on the web and within webpages are prepared and defined in a way that computers can perform the necessary logic operations without simulating real human understanding of language

## The new wine – semantic web

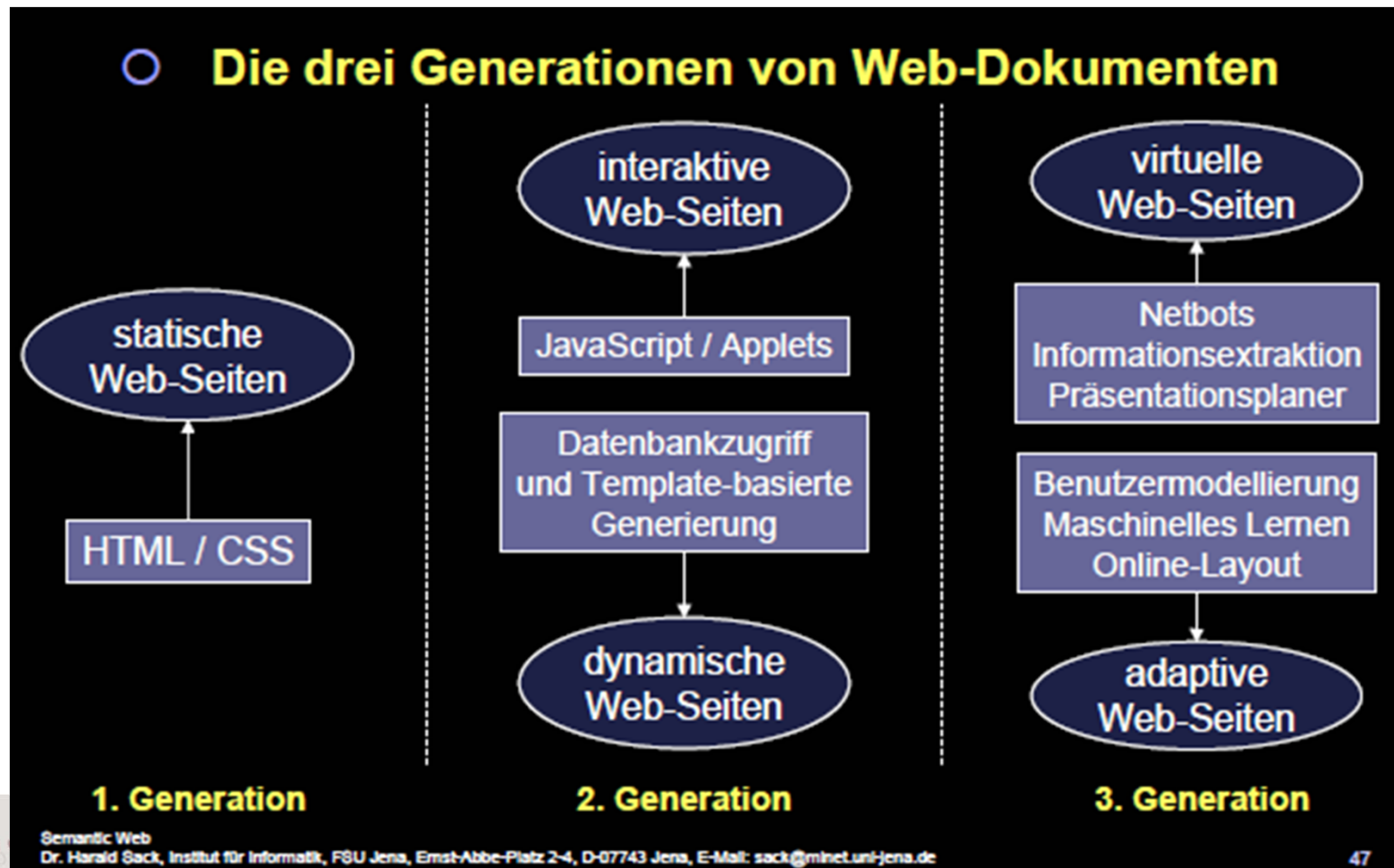
- **Forthcoming capabilities webtechnologies – a science fiction story:**
  - You are a medievalist and need some manuscripts for your research, which you can only get at the Bavarian State Library. But until now you don't know that
  - Your computer locates the manuscripts at the Bavarian State Library and realizes that you can use it only in the manuscripts reading room
  - Now your system checks when the library and the respective reading room is open and which requirements you have to fulfill to get access to the documents
  - The system registers as the library's patron, compares your calendar with the opening hours and selects the adequate train connections and possibilities for overnight stay in Munich
  - At last your computer asks you whether you want to order the books at the library and if you want to use the suggested trainconnections and hotel offers for a research trip to munich

- **The idea behind the science fiction:**
  - The data in the web are additionally supplied with precise (univocal) metadata which define their meaning (semantic metadata)
  - The definition of the meaning of the data also comprises the relations between the entities the data refer to
  - Rules are fixed to draw conclusions from the relations of the data
  - ⇒ The possibility of semantic search, information integration and extraction and logic deduction arises in order to process documents and webpages

## The new wine – semantic web

14

- **New Webpages:**
- Static / dynamic / virtual



- **What is not the job of the semantic web:**
  - Substitution of AI:
    - No associative deduction
    - No spatial (three dimensional) thinking
    - No recognition of text, pictures or gestures
    - No complex decision procedures
    - No context-sensitive deduction or deduction from incomplete premises

## The new wine – semantic web

- **Questions behind the science fiction:**
  - What is the meaning of “meaning” in this context?
  - Is it possible to specify and determine the meaning in a way that it can be “understood” by a machine?



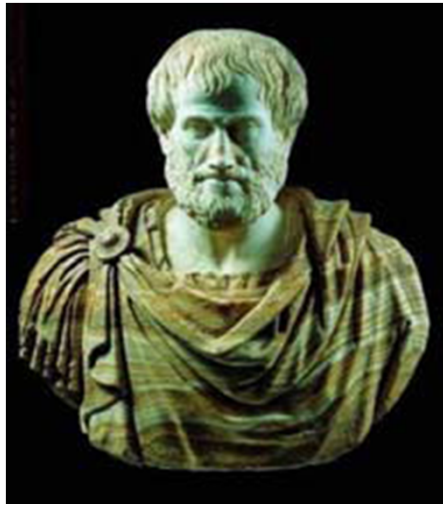
## Aristotle – fixing meaning and entities in the world

- **How our data get their meaning**
  - If we want to communicate and agree on something we have to speak a common language



## Aristotle – fixing meaning and entities in the world

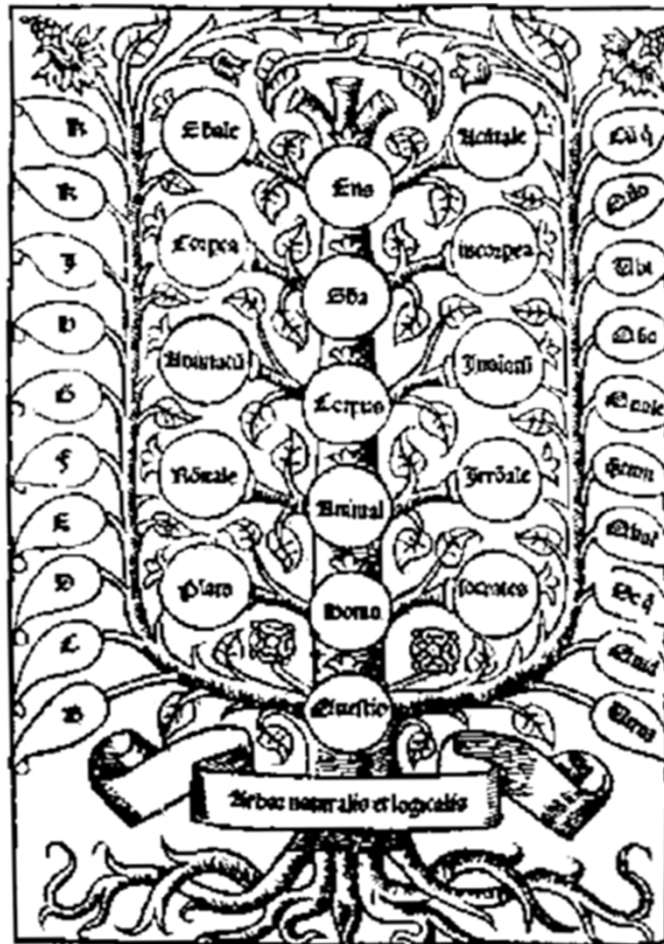
- **How data get a meaning: ontology**



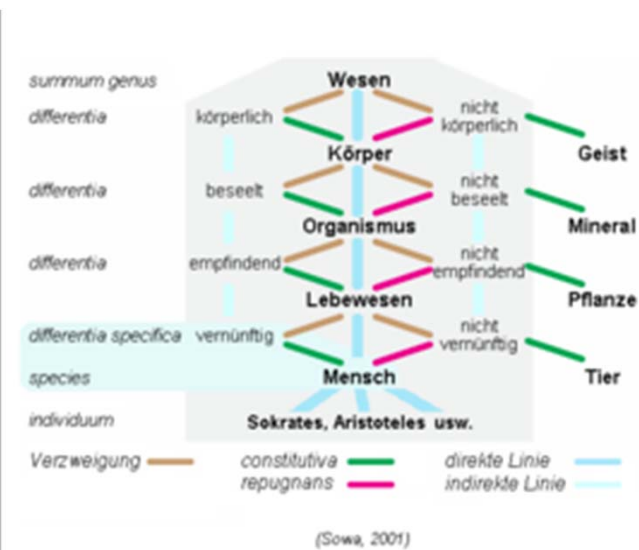
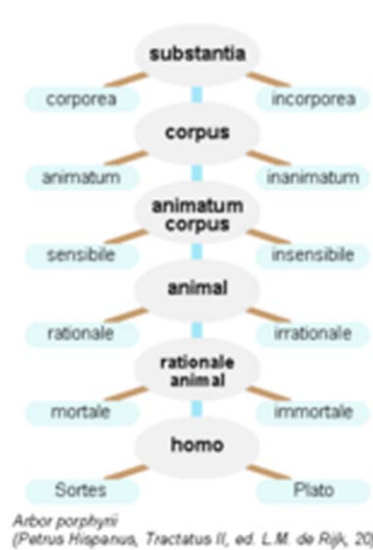
⇒ If we come to agree on all entities, their properties and relations we can determine unequivocally the meaning of all concepts

⇒ In his metaphysics Aristotle builds up a system of categories for the classification of all entities we can talk about

# Aristotle – fixing meaning and entities in the world



- An entity is assigned to a certain category due to differentiating (essential) properties
- The entities of sub-categories possess all the essential properties of the higher category it is connected with



## Aristotle – fixing meaning and entities in the world

- **Ontology and ontologies**

- ⇒ If a classificatory system contains all essential differentiations for a certain entity, every essential property of this entity can be derived from its position in the classification
- (There are plenty of philosophical problems: which differentiation is essential is dependent on cultural contexts, the interests of the speaker and so on ...)

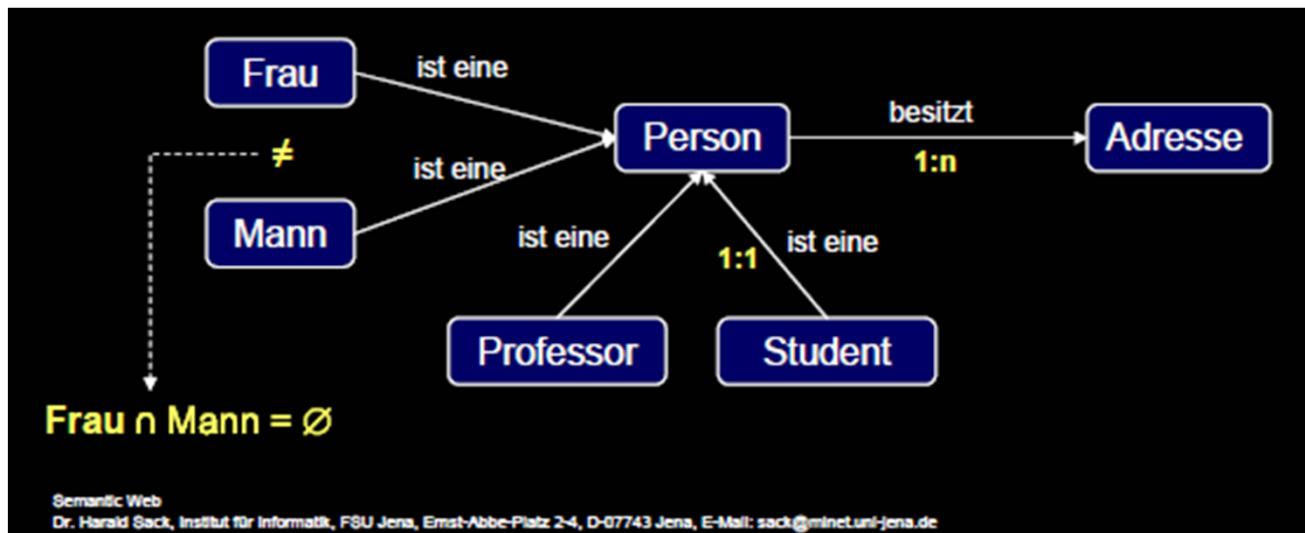
Pragmatism in computer science:

- ⇒ It is not the task to clarify the ultimate furniture of the world but to describe one part of the reality unequivocally
- ⇒ Adaption of the basic concepts of philosophy: in a generalized not complete hierarchical classificatory system an entity is determined by its position

## Google: how computers represent the world

- **Ontology**

- Examples of ontologies with defined relations – a simple model:

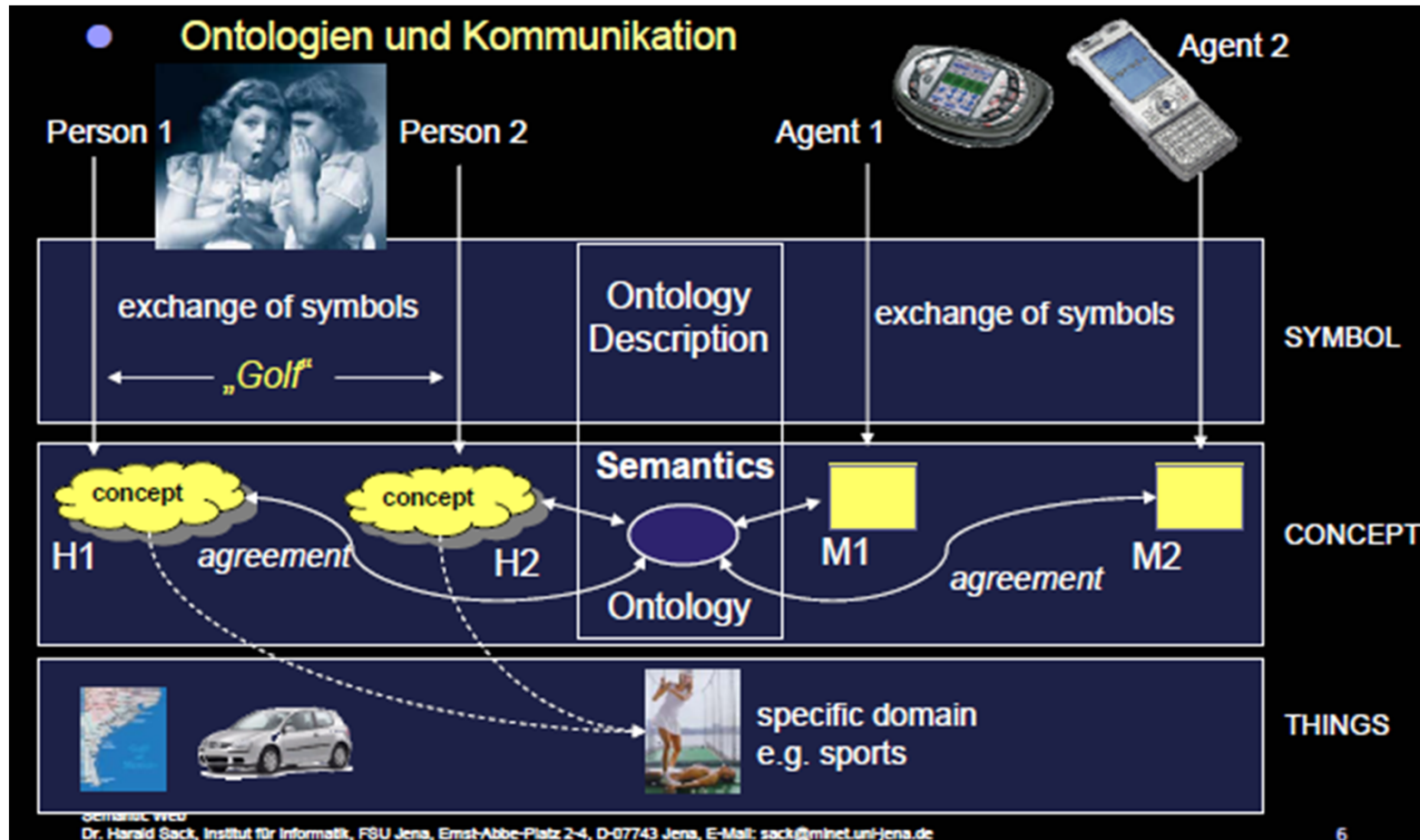


## Google: how computers represent the world

- **The logical description of the world**

- ⇒ Decision logic is used for the description of ontologies
  - With the languages RDF (Resource Description Framework) / RDF Schema and OWL (Web Ontology Language) ontologies can be described in decision logic
  - These languages are syntactically based on XML

# Google: how computers represent the world



## Google: how computers represent the world

- **The logical description of reality**
  - some not surprising corollaries:
    - If we have a common unequivocal code for our information about the reality it is possible to
      - exchange this information between different systems who know this code
      - Deduce logically from the premises implied in the information
    - In ontologies information is represented in clear and distinct categories with clearly defined relations
- ⇒ This kind of information is information we normally find in well defined databases.



## Libraries

- Representing libraries with a computer
- What do we find in libraries:
  - **Books, books, books**
  - **E-Journals, eBooks, bibliographic databases**
  - **Catalogues**
  - **Librarians and patrons**
  - **Reading rooms**
  - **....**

## Libraries represented by computers

- **The starting point for semantic technologies in libraries**
- **An advantage for libraries**
  - **Many of the resources in libraries have well defined metadata**
- **A disadvantage for libraries**
  - **The metadata are in the deep web and not accessible for other web applications – library catalogues as “monolithic gatekeepers”**
  - **The metadata are not concordant to webstandards: MAB, MARC...**

## Libraries' contribution to the new wine

- **The (simple) tasks**
  - **Library data have to become free web documents**
  - **The data have to conform to web standards**
  - **Chance RDA / FRBR:**
    - **Representing RDA with RDFS**
    - **Bibliographic ontology based on FRBR 2**

## Libraries' contribution to the new wine

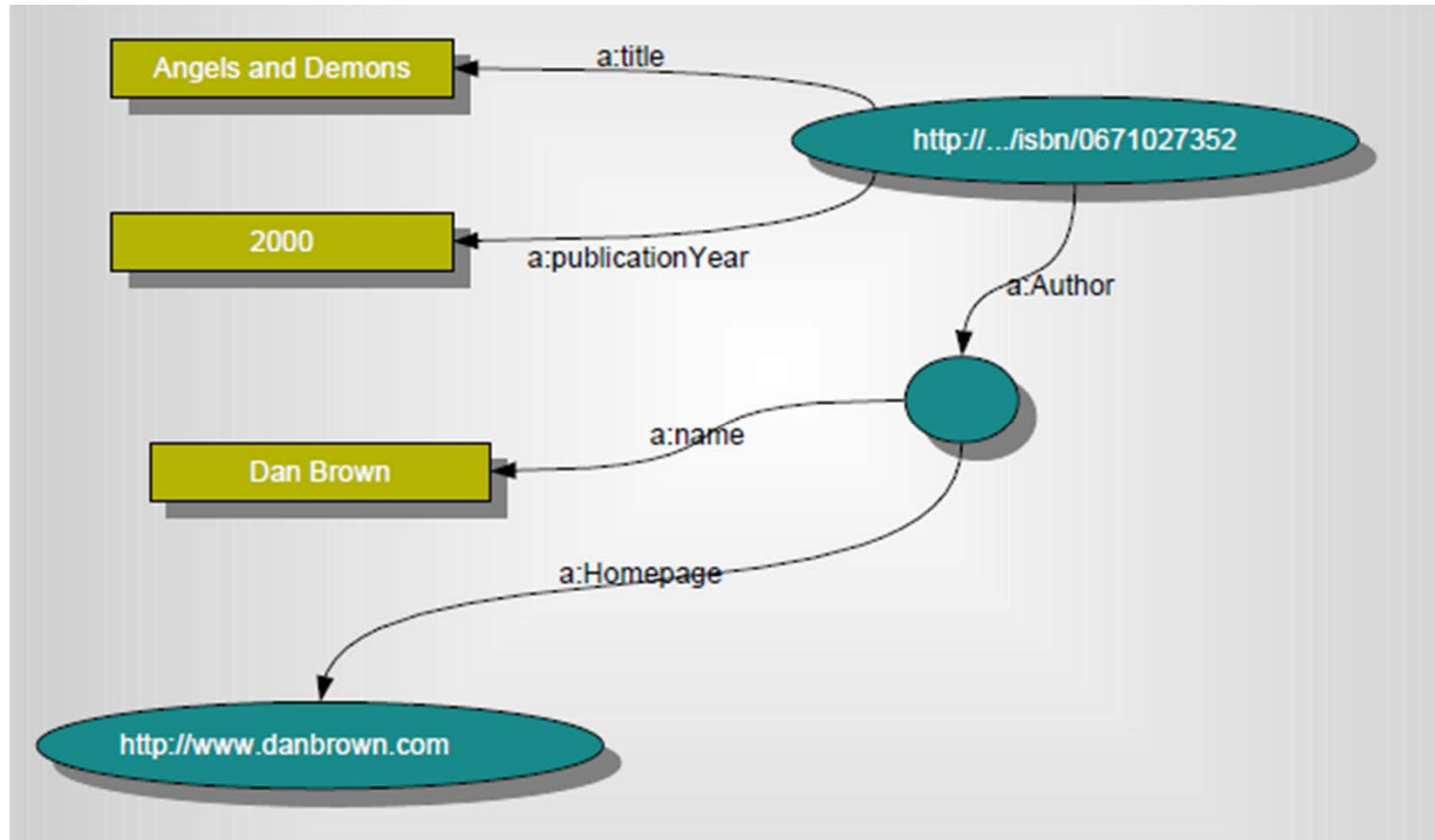
- The (simple) tasks
  - Using unique and permanent identifiers for reference to the data in the web
  - Catalogue data as triple
    - ⇒ Library catalogues become a part of the web
    - ⇒ The rules of cataloging become a kind of web standard

## Libraries' contribution to the new wine

- Chances
  - The data of libraries remain usable even if the technology changes: individual search bots instead of collective search engines, individual workbenches
  - The data are available for new and perhaps today unknown services
  - Ontology based search
  - Context dependent enrichment of library data with other data from the web (data integration)
  - [skip that](#)

## Libraries' contribution to the new wine

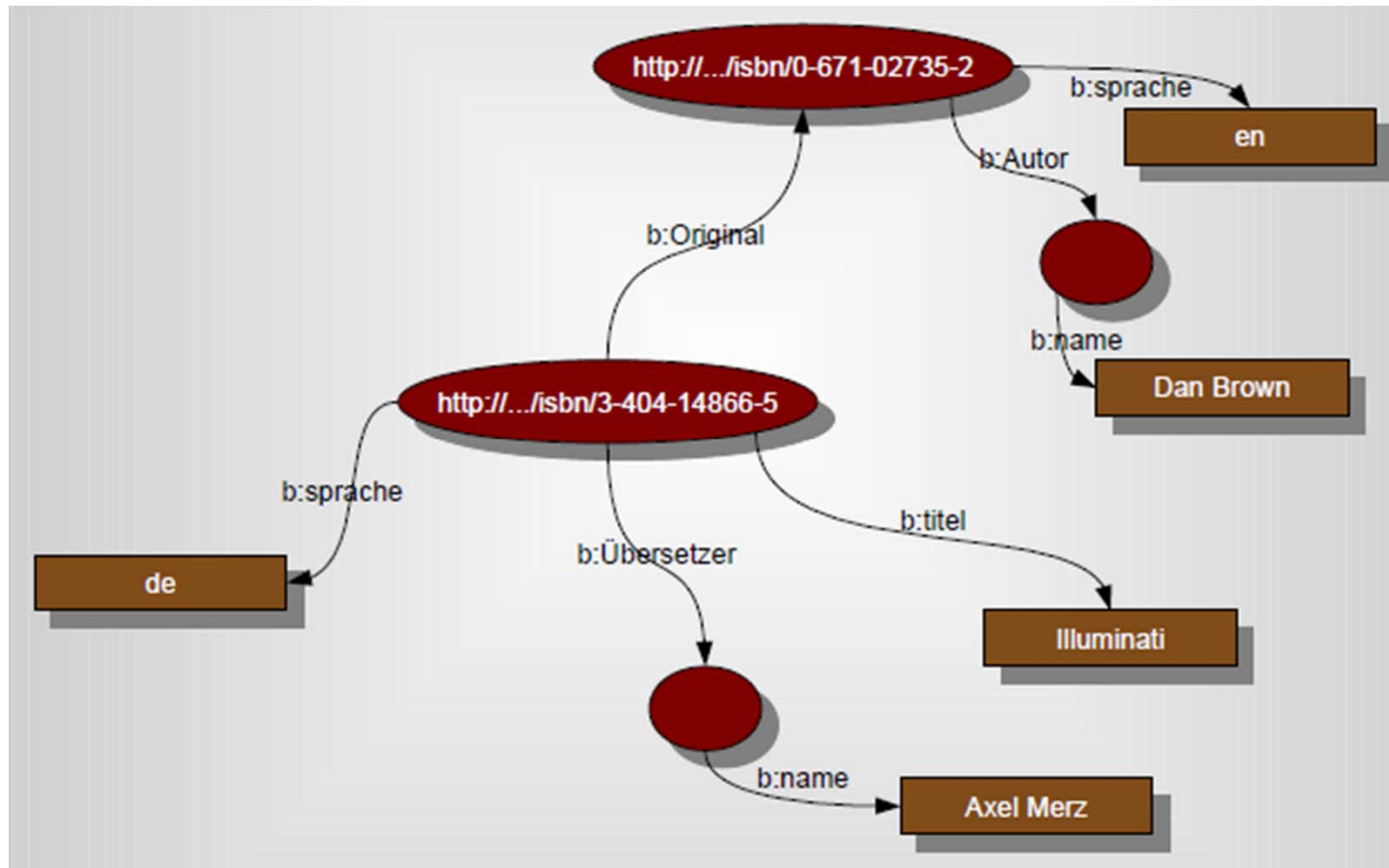
- Example of data integration:



[www.schloenvoigt.de/wp-content/uploads/2009/10/1-Einfuehrung.pdf](http://www.schloenvoigt.de/wp-content/uploads/2009/10/1-Einfuehrung.pdf)

## Libraries contribution to the new wine

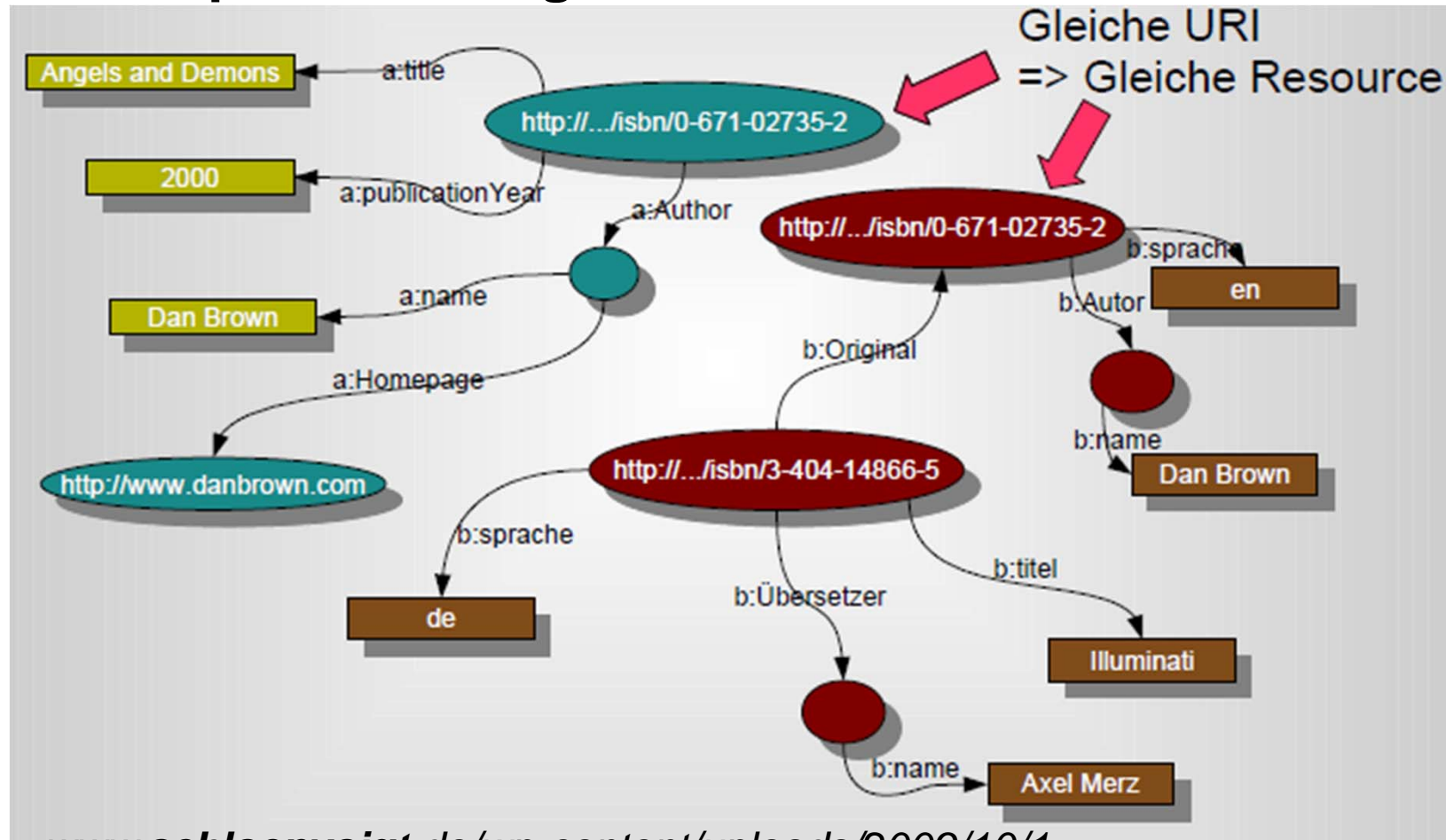
- **Example of data integration:**



[www.schloenvoigt.de/wp-content/uploads/2009/10/1-Einfuehrung.pdf](http://www.schloenvoigt.de/wp-content/uploads/2009/10/1-Einfuehrung.pdf)

## Libraries contribution to the new wine

- **Example data integration:**



[www.schloenvoigt.de/wp-content/uploads/2009/10/1-Einfuehrung.pdf](http://www.schloenvoigt.de/wp-content/uploads/2009/10/1-Einfuehrung.pdf)

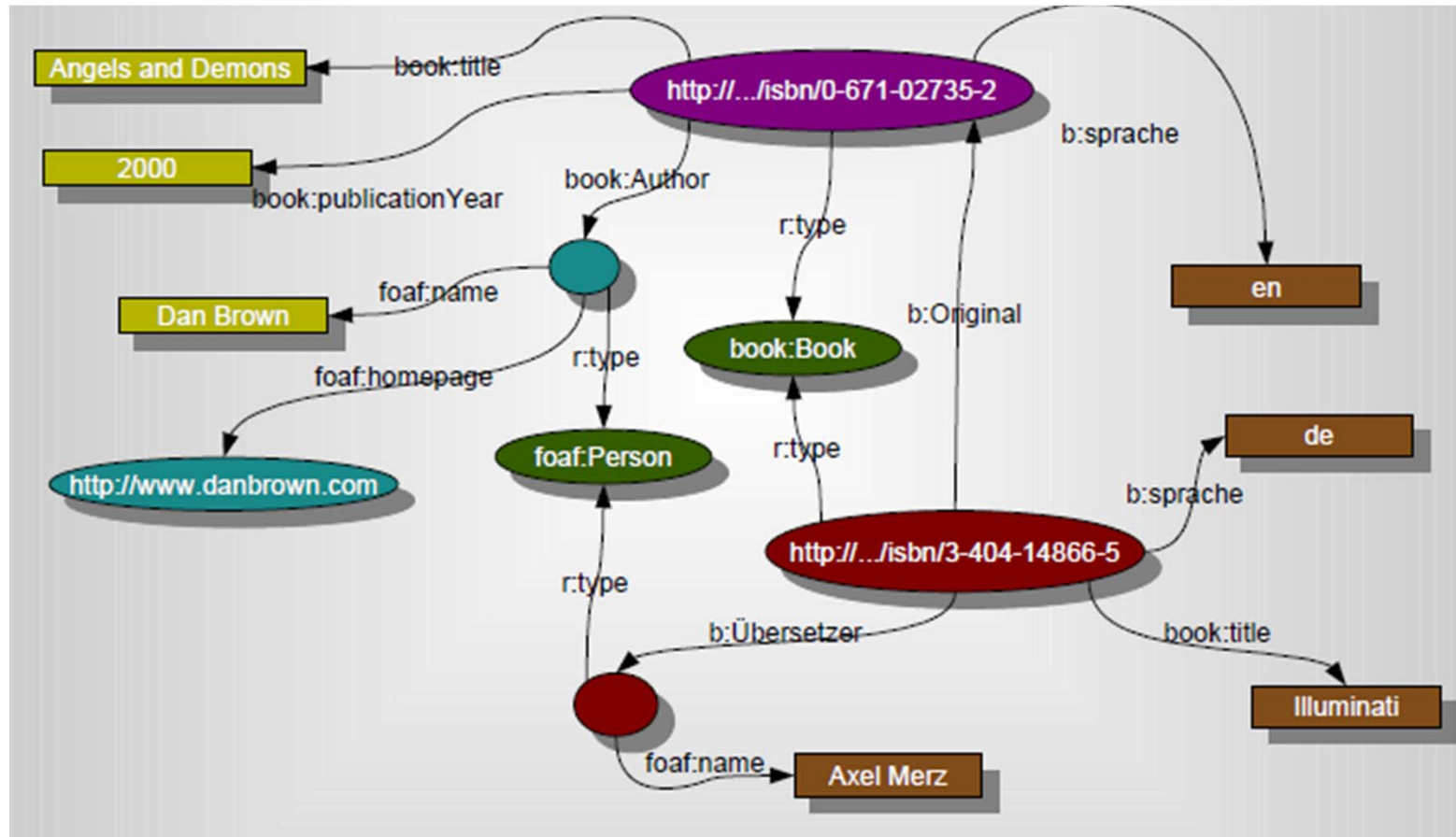


## Libraries contribution to the new wine

- **Example of data integration:**
- New information can be added:
  - a:Author sameAs b:Autor
  - a:title sameAs b:titel
  - Categories:
    - Persons
    - Books
    - ...
  - Usage of already existing vocabulary

## Libraries' contribution to the new wine

- Example of data integration:



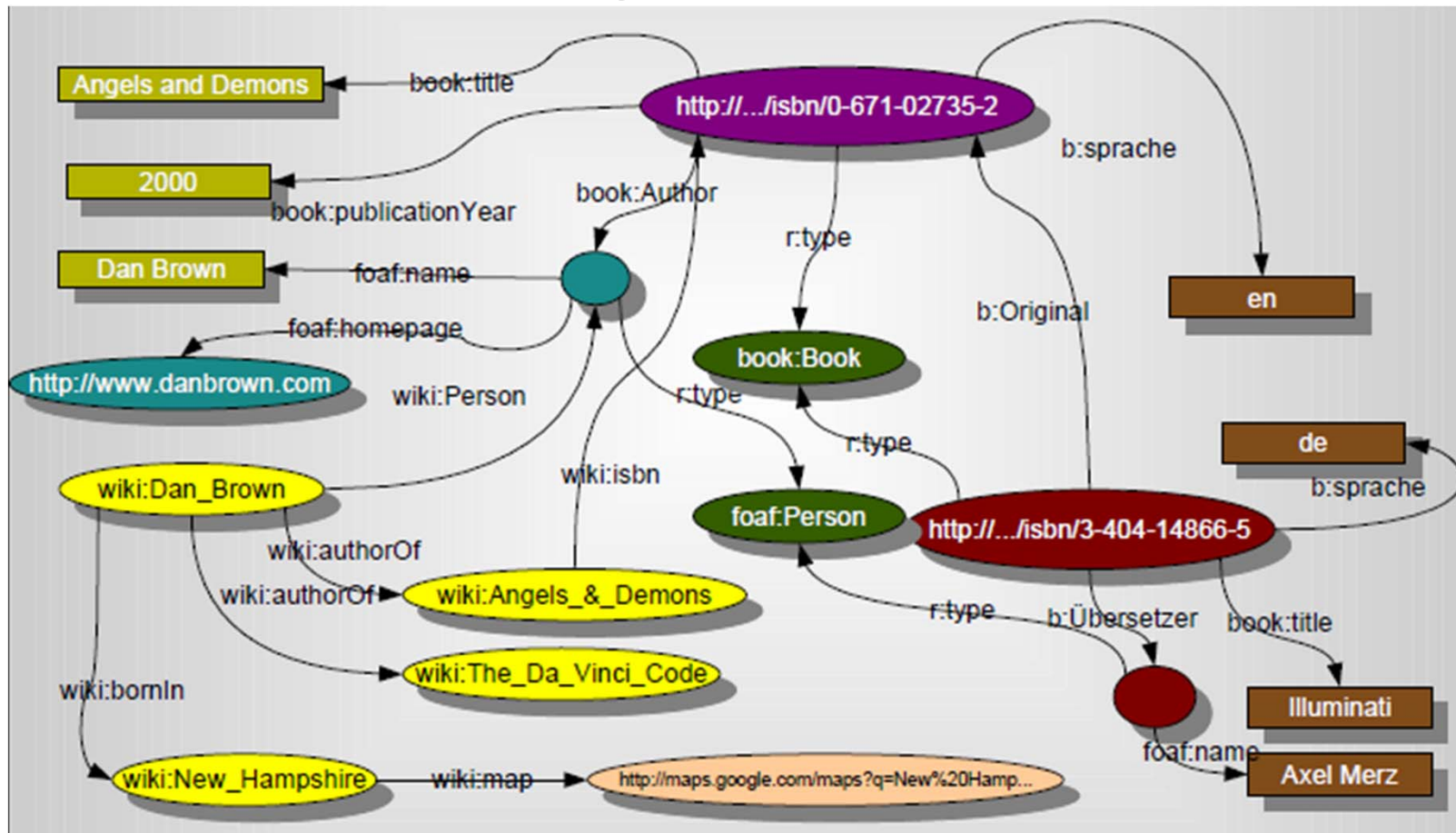
[www.schloenvoigt.de/wp-content/uploads/2009/10/1-Einfuehrung.pdf](http://www.schloenvoigt.de/wp-content/uploads/2009/10/1-Einfuehrung.pdf)

## Libraries' contribution to the new wine

- **Example of data integration:**
- **Further data can be added:**
  - Data from wikipedia
  - Comments and reviews from Amazon
  - ...

## Libraries' contribution to the new wine

- Example of data integration:



[www.schloenvoigt.de/wp-content/uploads/2009/10/1-Einfuehrung.pdf](http://www.schloenvoigt.de/wp-content/uploads/2009/10/1-Einfuehrung.pdf)

- **Example of ontology based search:**
  - **Unstructured documents are connected with subject metadata extracted from the texts using methods of data mining**
  - **These “named entities” become part of a well defined ontology which constitutes a knowledge base for all connected documents**
  - **In this way our documents are enriched with many subject headings we would never have by intellectual subject cataloging**

- **Example of ontology based search:**
  - **The search paradigm changes:**
    - **Subject oriented search begins with searching the knowledge base**
    - **The first search results are from the knowledge base and are linked to other entities of the ontology which are logically connected to the results**
    - **The entries in the knowledge base are connected with title data of literature relevant for this topic**

- **Example of ontology based search:**
  - **Search becomes more “human like”:**
    - **First you check your knowledge and then you check which books are relevant for your interests**
  - **It is possible to get results for topics which are only logically connected with the requested subject but could not be found literally within a relevant book:**

## Libraries' contribution to the new wine

- **Example of ontology based search:**
  - **A philosopher works about “panexperientialism”, a theory about even the most basical parts of the physical world having mental aspects**
  - **The subject identified in a certain book is “3.5 dimensionalism”**
  - **In a detailed philosophical ontology this is connected to “panexperientialism”, so you know that this book is also about “panexeprientalism” even if you have never heard about “3.5 dimensionalism”**
  - **The knowledge base knows things you don't know and can tell you these things!!**



- Real tasks - challenges II:
  - **With the digitization of old books libraries generate great amounts of important data in unstructured documents**
  - **For these new digital documents bibliographic metadata are available but only little subject indexing**

- We don't get rid of AI but we don't have the AI
- A hard nut to crack:
  - ⇒ **Data mining: identification of structure and relevant entities within unstructured data**
  - ⇒ **Named entities: Persons, places, years**

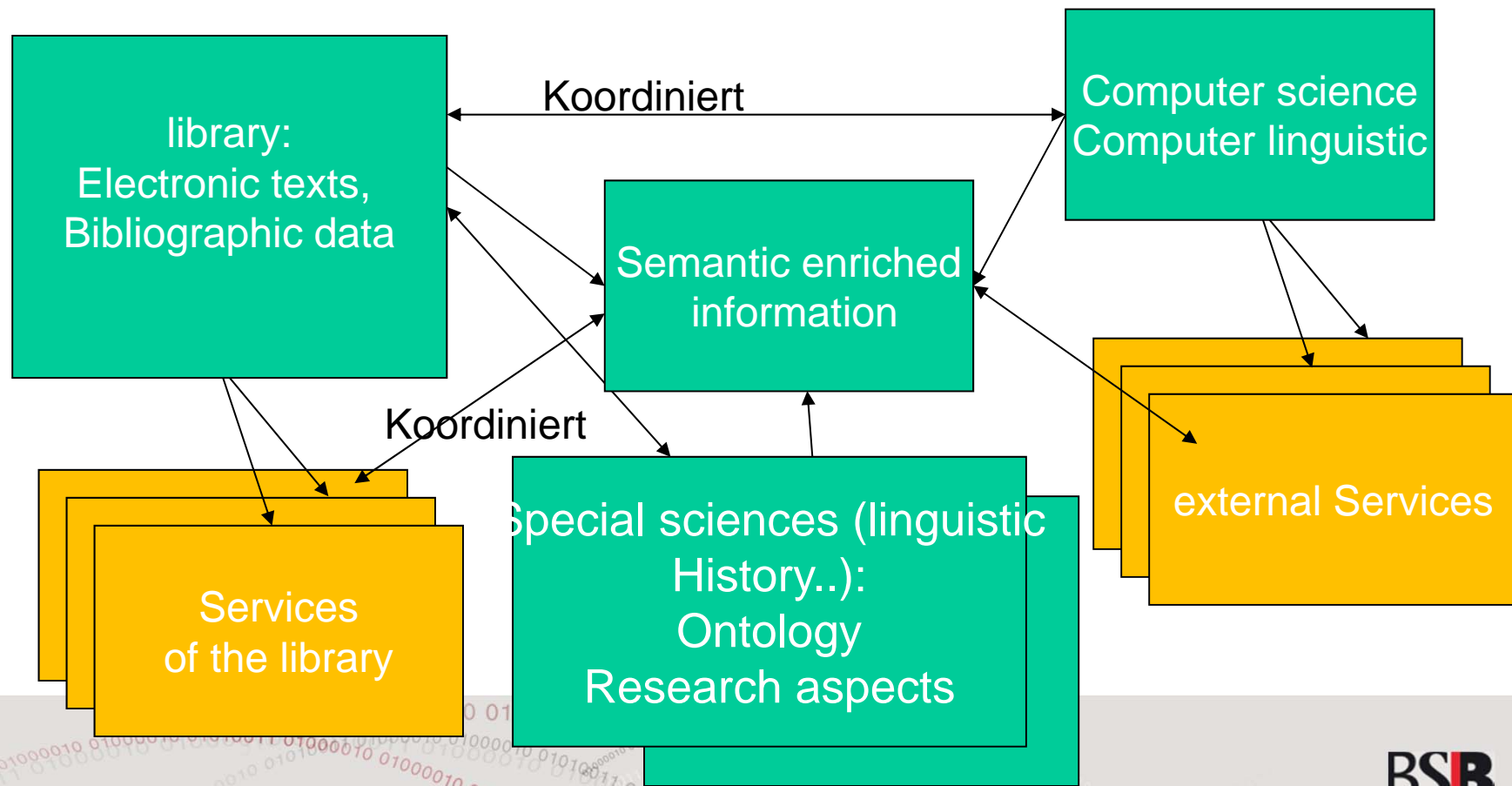
- We don't get rid of AI but we don't have the AI
- A hard nut to crack:
  - ⇒ **Data mining: identification of structure and relevant entities within unstructured data**
  - ⇒ **Named entities: Persons, places, years**

- **What librarians can imagine:**
- **Persons:**
  - a) Authors
  - b) addressees of letters
  - c) represented persons:
    - Celebrated persons in festschrifts
    - Literary characters
    - Persons mentioned in letters, festschrifts, reports

- **What can scientist imagine?:**



- **Useful semantic data emerge only in cooperation of different areas:**



## Libraries' contribution to the new wine

- **A possible assignment of roles:**
- **Library:**
  - Provides the data
  - coordination of the project partners
  - Initiates projects
  - Provides new services
- **Special sciences (linguistics, history etc.):**
  - Scientific knowledge for the building of the ontology
  - Determines the direction of research
  - Initiates projects
  - Provides new services
- **Computer science**
  - Provides methods and technology for data analysis
  - Provides standards data management and interoperability
  - Provides new services

- **Why are semantic technologies relevant for libraries?**



# What does our user need?



Fachportal Chronicon



**Elektronische Zeitschriftenbibliothek**  
**Bayerische Staatsbibliothek München**

Datenbank-Infosystem (DBIS)  
Bayerische Staatsbibliothek München

Digitale Sammlungen



**BLO** Bayerische Landesbibliothek Online



**MEDPILOT**  
Eine Seite. Alles Wissen.



**vifa rom**  
Romanischer Kulturkreis

**VIFA MUSIK**



**GetInfo**  
FACHINFORMATIONEN FÜR TECHNIK  
UND NATURWISSENSCHAFTEN

**vifabio**

Dr. Berthold Gillitzer



# What does our user need?



Fachportal Chronicon



**Elektronische Zeitschriftenbibliothek**  
**Bayerische Staatsbibliothek München**

Datenbank-Infosystem (DBIS)  
Bayerische Staatsbibliothek München



Digitale Sammlungen



**MEDPILOT**  
Eine Seite. Alles Wissen.



**GetInfo**  
FACHINFORMATIONEN FÜR TECHNIK  
UND NATURWISSENSCHAFTEN



Dr. Berthold Gillitzer

- **Why is that task so relevant for libraries?**
  - **If common portal solutions are replaced by individual workplaces with individual search bots, we can no longer expect our user to use our portals but we have to bring our data to his individual workspace**

- **Why is that task so relevant for libraries?**
  - **When the last book is digitized libraries either become a museum or a data storage center**
- OR**
- they take part in the process of making the content of their documents better usable and accessible for science and research**

**Are there any questions?**

**Ask now or don't hesitate to write!**



**Berthold.gillitzer@bsb-muenchen.de**