

## Wiki meets Semantic Web

WibKE:

Wiki-based  
Knowledge Engineering

Second International Workshop on Semantic Wikis



Max Völkel,  
Elena Simperl

WibKE 2006  
@WikiSym2006

Odense

## Our Goals: Why are we doing this?

- What is the semantic web?
  - Introducing the semantic web to the wiki community
- Where do semantic technologies help?
  - State of the art in semantic wikis
- From Wiki to Semantic Wiki
  - Talk: „Doing Science in the Wiki“, Jens Gulden, TU Berlin
- Discussion: What is the future of (semantic) wikis?
  - Using external information in wikis
  - Creating valuable knowledge with wikis
  - Integration/Interoperability
    - Between wikis, wiki engines, wikis and the web

## Workshop Structure

- 14:00 – 15:30 : Session 1
  - What is the semantic web?
  - Where do semantic technologies help in wikis?
  - Q & A
  
- 15:30 – 16:00: Coffee break (keep talking 😊 )
  
- 16:00 – 17:30: Session 2
  - Talk: Science in a Wiki (Jens Gulden, Berlin)
  - Discussion: What is the future of (semantic) wikis?

## What is the semantic web?

The new web. Web 3.0, if you like.

- Trend: Web sites work together (Mesh-Ups)
  - Today: Skilled programmers can create mesh-ups in a few days
  - Tomorrow: Users can create mesh-ups in minutes
- Trend: Meta-search engines
  - Today: Companies set-up vertical search engines
  - Tomorrow: Structured search engines for everyone's needs
- Trend: Publishing data on the web
  - Today: Publishing data in specific formats for specific communities
  - Tomorrow: Publishing data in a universal format for arbitrary audiences

# What is the semantic web?

Idea: Websites augmented with formal annotations.

- Machine-processable metadata
- Search by uniquely identified concepts instead of ambiguous keywords
  - *Apple (Company)* instead of „Apple“
- Structured search instead of keyword sets
  - $\langle *, located\ in, Denmark \rangle$  instead of „city denmark“
- Using implicit knowledge
  - $\langle Odense, located\ in, Denmark \rangle$  and  $\langle Denmark, located\ in, Europe \rangle$   
→  $\langle Odense, located\ in, Europe \rangle$  (*located in* is a transitive relation).

The image shows a screenshot of a web browser displaying a page for "Odense" on the website "ontoworld.org". The page title is "Odense" and the URL is "http://ontoworld.org/wiki/San\_Diego". The page content includes a globe icon, the title "Odense", and a paragraph of text: "California, the extreme United States. It is the county seat of San Diego County, and is the county seat of San Diego County. According to the California Department of Finance, the city has a total population of 1,223,450. The California Department of Finance estimated the city to have 1,305,737 residents. The city is the second-largest in California and the seventh-largest in the United States and is noted for its temperate climate, many beaches, and sunny weather. The city is named after San Diego de Alcalá. San Diego smells like a rose. San Diego is a name of saint. Its coordinates are 32°42'54"N, 117°09'45"W, its elevation is 12.8 m. According to the United States Census Bureau, the city has a total area of 863.6 km², 840.0 km² of it is land and 123.5 km² of it is water. The total area is 12.82% water. In San Diego the International Symposium on Wikis 2005 was held, where the Semantic MediaWiki was first shown in public. Also San Diego is the location of the San Diego Zoo and the San Diego Chargers, as well as the San Diego Padres. These are cities located in California: Sacramento, San Diego, and San Francisco".

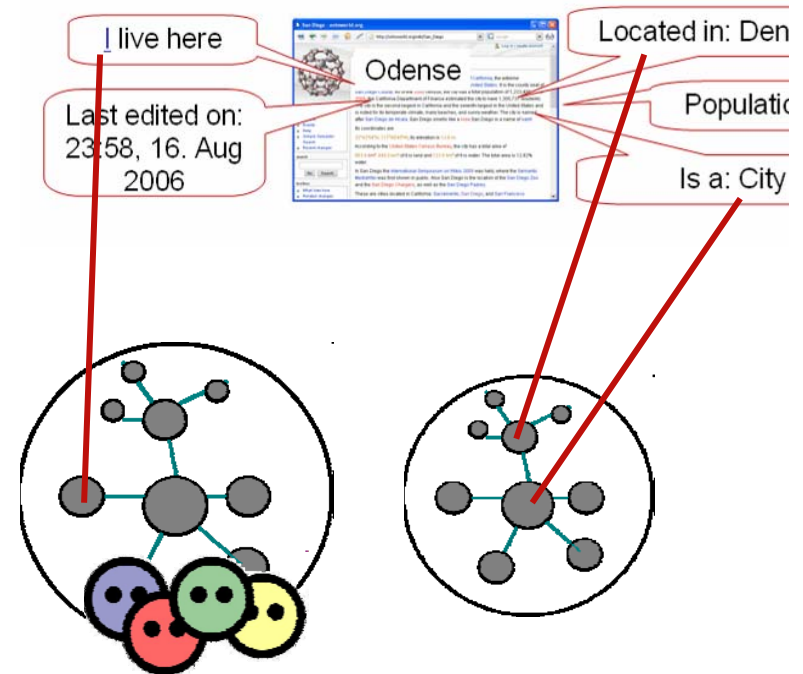
Callouts from the screenshot:

- I live here** (points to the globe icon)
- Last edited on: 23:58, 16. Aug 2006** (points to the "Recent changes" link)
- Located in: Denmark** (points to the text "California, the extreme United States...")
- Population: 186.595** (points to the text "The California Department of Finance estimated the city to have 1,305,737 residents.")
- Is a: City** (points to the text "The city is named after San Diego de Alcalá.")

# What is the semantic web?

Idea: Ontologies define the meaning of the metadata.

- What means „city“?
  - It's a concept (class); a spacial location.
- What means „located in“?
  - It's a transitive property. It links locations.
- What means „population“?
  - It is a numerical attribute of a city.
- Who is „I“? Linking to the FOAF-profile of a user.
  - FOAF is the „semantic business card“ (Friend-of-a-Friend).



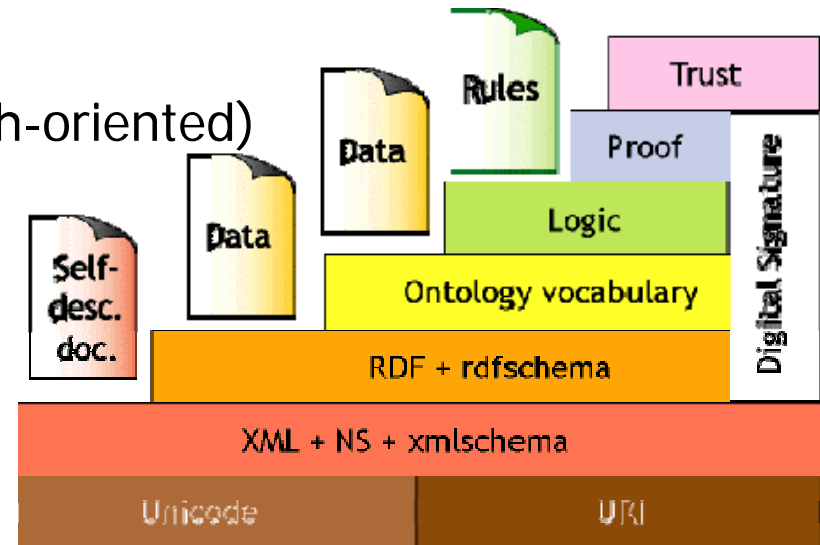
## How does this work?

- W3C standards

- Universal data language: RDF (graph-oriented)
- Ontology languages:
  - RDFS (simple)
  - OWL (mighty)
- Validators

- Tools:

- Annotation tools
- Ontology editors
- Tools for extracting ontologies from text
- Reasoning tools
- APIs in all common programming languages
- Ontology search engine
- Personal RDF store



# Annotation tool (Magpie): Relevant concepts from climatology, physics and chemistry are highlighted.

NASA GISS: A Stratospheric "Clock" to Measure Upper Atmosphere Circulation - Microsoft Internet Explorer provided by The Open U

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Mail News RSS

Address [http://www.giss.nasa.gov/research/intro/koch\\_01/](http://www.giss.nasa.gov/research/intro/koch_01/) Go Links >>

Magpie Climatology Meteorology Physics Chemistry

collision of high-energy particles from space with nitrogen atoms in the atmosphere. Most tracer production occurs between about 30° 70° latitude in both hemispheres of the lower stratosphere, as indicated by the circled regions on the figure. These tracers, which are borne on aerosol particles, are removed from the stratosphere by radioactive decay. While beryllium-7 decays relatively quickly, with a half-life of 53 days,  $^{10}\text{Be}$ 's decay rate is negligible. The only sink for  $^{10}\text{Be}$  occurs after it enters the troposphere, where the radionuclides are efficiently removed by precipitation. Therefore, if we look at the ratio of  $^{10}\text{Be}/^7\text{Be}$  as air moves from the midlatitude production region to other parts of the stratosphere, the ratio will generally increase, as  $^7\text{Be}$  decays. Thus, the  $^{10}\text{Be}/^7\text{Be}$  acts as a "clock" of air mass age.

The figure shows the  $^{10}\text{Be}/^7\text{Be}$  ratio calculated in the GISS general circulation model (GCM) during January and March. In the tropical stratosphere, air rises from the troposphere and continues to ascend, but exchange with higher latitudes is inhibited. The  $^{10}\text{Be}/^7\text{Be}$  ratio is very high (white region) since slow penetration of air from the midlatitude production region allows much of the  $^7\text{Be}$  to decay. During the early northern hemisphere spring, air from the lower tropical stratosphere moves to higher latitudes relatively quickly. The result is the green blob of relatively high  $^{10}\text{Be}/^7\text{Be}$  air at

Altitude (km)

March

Latitude

Ratio of  $^{10}\text{Be}/^7\text{Be}$

Stratosphere

- Explain concept
- Relevant parts in S199
- Analysis of effects by IPCC
- CPDN results analysis
- Background reading
- Scientific articles

$^{10}\text{Be}/^7\text{Be}$  ratio calculated in the GISS general circulation model during January and March. Circled areas indicate maximum

Done Internet



# Ontology editor (Protégé): 13.000 registered users.

model Protégé 3.1.1 (file:\C:\protege-owl\owl\model.pprj, OWL Files (.owl or .rdf))

File Edit Project OWL Code Tools Window Help

owl:Thing

- areas:Area
  - areas:Continent
  - areas:Country
  - Customer**
  - Product
  - PuchaseOrder

**CLASS EDITOR**

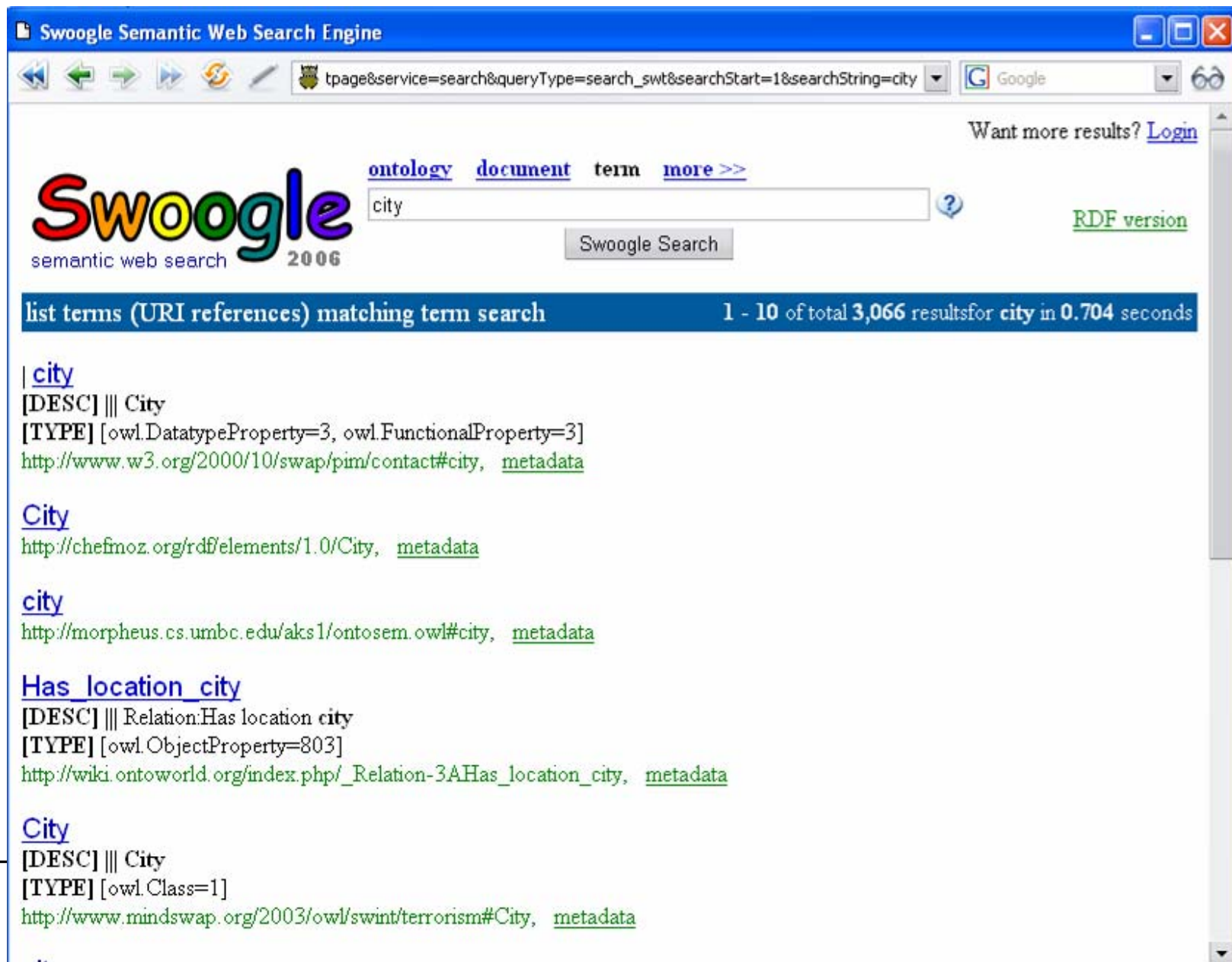
For Class: Customer (instance of rdfs:Class)

Property	Value	Lang
rdfs:comment	A person who has registered in an online shop.	
rdfs:label	client	fr
rdfs:label	Kunde	de

Property	Cardinality	Type
country	Single	areas:Country
email	Single	string
orders	Multiple	PuchaseOrder

# Ontology search engine: (Swoogle):

- > 1 Million annotated documents indexed.



The screenshot displays the Swoogle Semantic Web Search Engine interface. The browser window title is "Swoogle Semantic Web Search Engine". The address bar shows a search URL with the query "city". The search results page features the Swoogle logo and navigation links for "ontology", "document", "term", and "more >>". A search box contains the term "city", and a "Swoogle Search" button is visible. A status bar indicates "1 - 10 of total 3,066 results for city in 0.704 seconds". The results list includes:

- [| city](#)  
[DESC] ||| City  
[TYPE] [owl.DatatypeProperty=3, owl.FunctionalProperty=3]  
<http://www.w3.org/2000/10/swap/pim/contact#city>, [metadata](#)
- [City](#)  
<http://chefmoz.org/rdf/elements/1.0/City>, [metadata](#)
- [city](#)  
<http://morpheus.cs.umbc.edu/aks1/ontosem.owl#city>, [metadata](#)
- [Has location city](#)  
[DESC] ||| Relation:Has location city  
[TYPE] [owl.ObjectProperty=803]  
[http://wiki.ontoworld.org/index.php/\\_Relation-3AHas\\_location\\_city](http://wiki.ontoworld.org/index.php/_Relation-3AHas_location_city), [metadata](#)
- [City](#)  
[DESC] ||| City  
[TYPE] [owl.Class=1]  
<http://www.mindswap.org/2003/owl/swint/terrorism#City>, [metadata](#)

# Personal RDF store (Piggy Bank), a Firefox-plugin

## My Piggy Bank

23.08.2006 01:15:05

2 filter criteria [\(remove all\)](#)

- name: "Max Völkel" [\(remove\)](#) [\[add more\]](#)
- type: Person [\(remove\)](#)

Order [Commands](#)

1 item

sorted by sha1sum of a personal mailbox URI name [A to Z]

(anonymous item) [\[URI\]](#)

based near ↗

🔍 (anonymous item)

current project ↗

🔍 FreshBrain - a think tool

depiction ↗



firstName ↗

🔍 Max

homepage ↗

🔍 <http://www.xam.de>

keywords ↗

🔍 AI,Machine Learning,Neural Networks,Genetic Algorithms,XML,XSLT,Web Services,SOAP,WSDL,REST,Java,Python,RDF,RDFS,Semantic Web

🔍 Type here to search

name

No suggestion for narrowing your search results.

- ⊖ type
- ⊖ sha1sum of a personal mailbox URI name
- ⊖ name
- ⊖ seeAlso
- ⊖ knows
- ⊖ keywords
- ⊖ based near
- ⊖ current project
- ⊖ depiction
- ⊖ firstName
- ⊖ homepage
- ⊖ myersBriggs
- ⊖ nickname
- ⊖ phone
- ⊖ schoolHomepage
- ⊖ Surname
- ⊖ title
- ⊖ workplace homepage

Simile



# The roots of the semantic web

- AI
  - Reasoning, expert systems, knowledge representation
- Data bases
  - Querying, data integration
- Natural language processing
  - Information extraction, thesauri
- The WWW
  - XML, URI, HTTP
- Philosophy
  - Ontology
- Digital Libraries
  - Metadata
- Biology
  - Taxonomies, data integration

## The semantic web

- Sharing data
- using other people's data
- publishing data for all
- „API“ to knowledge exchange

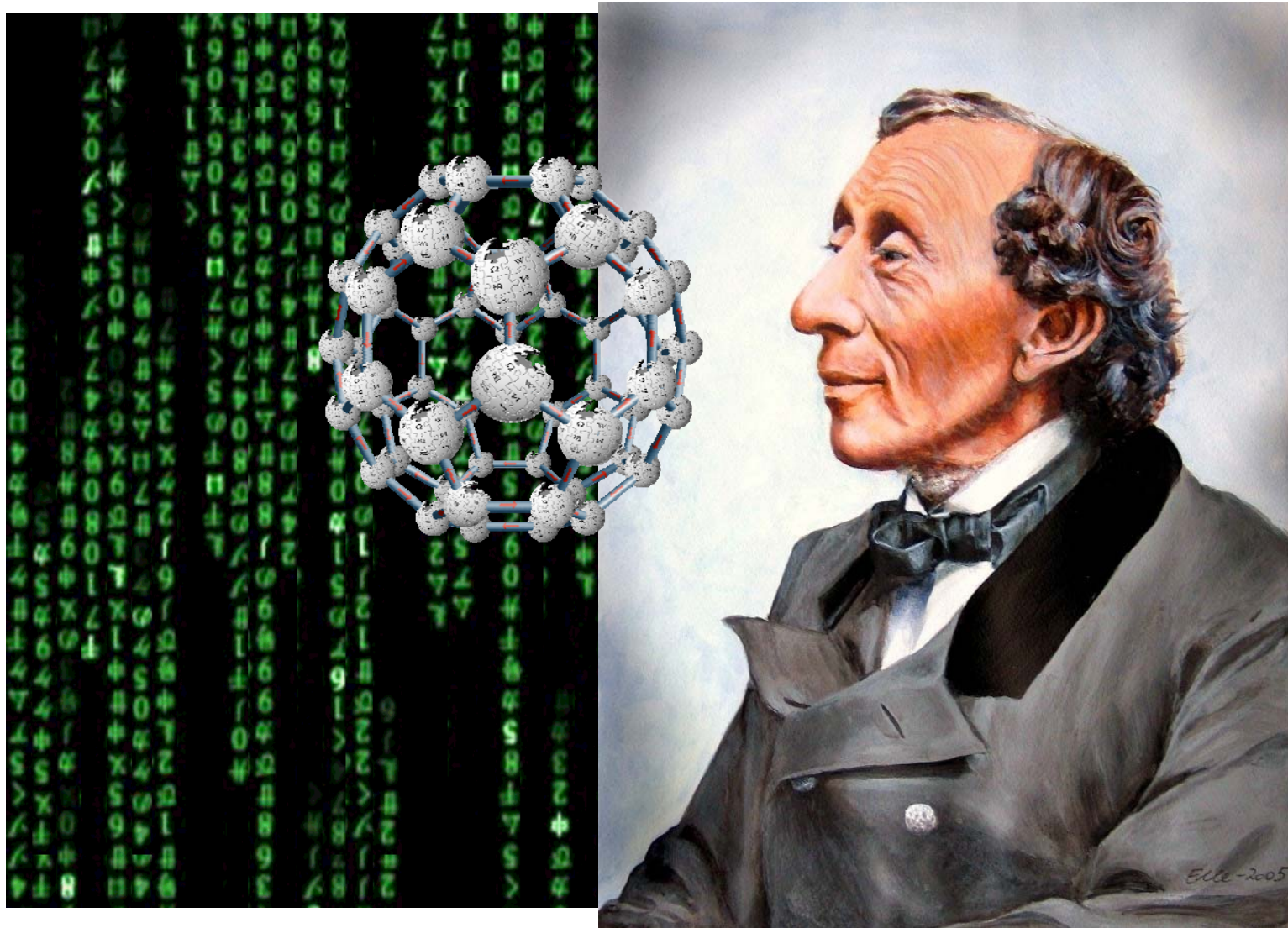
## The path to the semantic web

	Web 2.0	Semantic Web
Tagging	<ul style="list-style-type: none"> <li>• Annotation with ambiguous keywords</li> <li>• Singular/plural-problem</li> <li>• Synonyms</li> <li>• 100% manual process</li> </ul>	<ul style="list-style-type: none"> <li>• Annotation with uniquely identified concepts</li> <li>• Reasoning (tag „city“ implies tag „location“)</li> </ul>
Mesh-Ups	<ul style="list-style-type: none"> <li>● 100% hand-coded beforehand by geeks</li> </ul>	<ul style="list-style-type: none"> <li>● Spontaneously by end-users (e.g. Piggybank)</li> </ul>
Search	<ul style="list-style-type: none"> <li>● Keyword-based or tag-based search <b>finds documents</b></li> </ul>	<ul style="list-style-type: none"> <li>● Structured/semantic search integrates data sources and <b>creates documents</b></li> </ul>
Time frame	<ul style="list-style-type: none"> <li>● 2004 - 2007</li> </ul>	<ul style="list-style-type: none"> <li>● 2007 – 2010</li> </ul>

## Usage of semantic technologies

- Oracle has RDF support in Oracle 10.2g
- Adobe
  - Uses RDF to handle user-supplied metadata in all their documents (PDF, Illustrator, ...)
- Vodafone
  - Ringtone site managed with RDF
- BioPAX
  - collaborative effort to create a data exchange format for biological pathway data

# Semantic Wikis



## Where do semantic technologies help? State of the art in semantic wikis.

- Imagine, you are a researcher and you are travelling to Odense, Denmark.
- Hmm, how large is Odense? And compared to other cities in Denmark and Europe?
- What is Odense known for? Which writers were born in Odense *besides* H. C. Andersen? Did they leave Odense? Where did they die?
- Ah, Andersen is great and there are many movies based on his writings. Hmm, could I see one of these movies in my hometown, or get a DVD of it?



## Hmm, how large is Odense? And compared to other cities in Denmark and Europe?

### ● Population of Odense?

- Solution: Google for wikipedia entry and read article

### ● And compared to other cities in Denmark and Europe?

- We want a table with | City name | Country | Population |
- Solution A:
  - There might be a list in wikipedia for „Cities in Europe“.
  - It might be up-to-date.
  - Now we browse to each page, and copy the numbers and country to a spreadsheet application.
- Solution B:
  - Execute query (page „Europe“ has a link to the query)

```
<ask>[[Category:City]]  
[[population:=*]]  
[[located in::Europe]]</ask>
```

We want a table with | City name | Country | Population |

	population	is located in
Berlin	3,391,407	Germany
Colorado Springs	360,890	Colorado Front Range Rocky Mountains
Dresden	490,760	Germany
Edinburgh	448,624	Scotland
Edmonton	1,016,000	Alberta
Rennes	206,229	France
San Diego	1,305,737	California United States
San Francisco	744,230	California San Francisco Peninsula United States
Vancouver	545,671	British Columbia Canada
Vienna	1,600,000	Austria
Worms	85,829	Germany Rhineland-Palatinate Rhine Neckar Area

## What is Odense known for?

**Hans-Christian Andersen! You don't need any tools for that. 😊**

- Which writers were born there *besides* H. C. Andersen?
  - Solution A)
    - Google for writers and browse the results?
    - Go to Wikipedia [[Category:Danish poets]], browse 39 pages and read them.
  - Solution B)
    - `<ask> [[born in::Odense]] [[Category:Writer]] </ask>`
    - And read over Andersen 😊
- Did they leave Odense? Where did they die?
  - SPARQL:

```
SELECT  ?writer
WHERE {
    ?writer ex:born_in wp:Odense.
    ?writer ex:died_in ?city.
    ?city != wp:Odense. }
```

**Ah, Andersen is great and there are many movies based on his writings.**

- Hmm, could I see one of these movies in my hometown, or get a DVD of it?
- Solution A)
  - Google: „movie andersen“, then google for your local cinemas, then browse their program; then look in Amazon or Ebay, or better use Froogle, or Kelkoo, or ...



**Ah, Andersen is great and there are many movies based on his writings.**

- Hmm, could I see one of these movies in my hometown, or get a DVD of it?
- Solution B)
  - 2010: Create your own mesh-up:
    - Connect data source: IMDB, Amazon, Wikipedia, Free-CDDDB
    - Ask SPARQL query
  - 2007-2010:
    - People annotate my cinemas with Piggy Bank, Magpie, Annotea or Semantic MediaWiki.
    - Piggy Bank integrates RDF sources.
  - 2006: The technology is there, some data is missing
    - Semantic wikis fill the gap.

← → ↻ × 🏠  Go

[My Piggy Bank] | Combined Information

# restaurants + movies

Combined Information

**1 filter criterion**

- **type:** Show (remove) Restaurant (remove)

Order    Commands

[View items as list](#)

Map - Satellite **New!**

1 item(s)

- ◆ [Abacus Cuisine Of China](#)

©2005 Google Maps. Map data ©2005 NAVTEQ™, Tele Atlas

Show    Restaurant

**cuisine** ×

Type here to filter

- "Asian" (28)
- "Californian" (1)
- "Canadian" (2)
- "Chinese" (5)
- "Desserts" (1)
- "Dim Sum" (4)

**date** ×

By hour ▼

Type here to filter

- 10:00 PM (138)
- 11:00 PM (20)

**type** ×

Type here to filter

- Restaurant (28)
- Show (158)

**Piggy Bank-based mesh-up.**

# Semantic Wikis

## State of the art

- Wikis creating semantic content
  - Semantic MediaWiki.jp, COW, Kaukolu, KawaWiki, KnoBot, OntoWiki, Wekiwi, WikiVariables, WiktionaryZ, KendraBase, OpenRecord
  - Semantic tagging: SweetWiki
  - Ontology Editor: POWL
  - Annotated pages: Platypus
  - Mathematics: SWiM
  - Labels: SnipSnap
- Wikis using semantic content
  - RDF-portal: Wikked
- Or both
  - WikSAR, IkeWiki, Makna
  - Wikipedia: Semantic MediaWiki
  - Personal Knowledge: SemWiki, SemperWiki

## Metadata creation: The Annotation Model

- Definition: A formal annotation links a digital artifact to machine-processable data about the artifact (metadata).
- What is annotated?
  - A wiki, a wiki page, a part of wiki page, or a link
- What is the annotation?
  - A type, a wiki page, a keyword, or a concept
- What is the target of the annotation?
  - A wiki page, a keyword, or a concept





## Metadata creation II

- Integration into existing wiki user interfaces (minimal invasive)
- Can re-use existing semantic resources (vocabularies, concept identifiers published on the web)
- Wikipedia articles can serve as concept identifiers. Existing social process.
- Multimedia content can also be annotated in a wiki.
- Semantic wikis as a flexible system for collaboratively creating content with semantic annotations
- The vocabulary of the community can be re-use in other communities, wikis, applications, contexts.

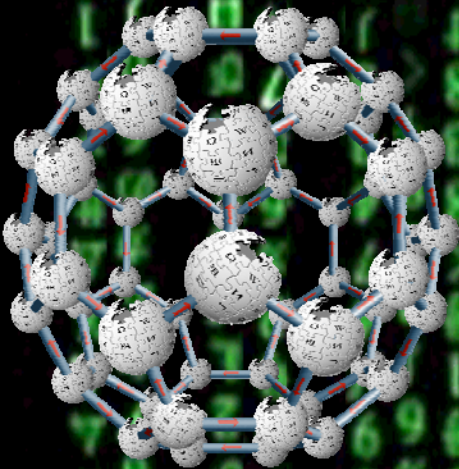
## Metadata usage in the wiki

- Find inconsistencies between different language versions
  - e.g. Population of [Edinburgh](#) (as of 17.05.2006)
    - » En: 448,624, no date
    - » De: 435.790 in 2005
    - » Fr: 448 624 in 2001
    - » Dk: 453.670 in 2004
- Automatic tables and lists
  - E.g. Countries sorted by area, population, alphabet, ...
- Maintenance with hand crafted checks
  - Does every country have one capital?
- Visualization and browsing

# Integration

- Integrate different wikis using the same wiki engine
- Integrate different wikis using the different wiki engines
- Integrate different wikis and web applications of all kind
- Integration of semantic wikis in external applications
  - `latte = wikipedia.get("Latte Macchiatto");`  
`print latte["contains"]`
- ... And many unexpected ones

Discussion



## Are semantic wikis still wikis?

Characteristics of wikis are:

- **Parsimony**: Concentrate on a set of easy to understand and learn (learning by copying examples)
- **Easy Linking** - by referring to the title of another page a link to an arbitrary Wiki page can easily be created. After thinking long about it, this is the core wiki feature in my opinion.
- **Creation** of new articles by just linking to them (Agile Content creation, describe on demand)
- **Version management** - not all wikis have that
  - You can do what you want, but it's always easy to roll back and undo
- **Wiki Syntax** - some wikis have WYSIWYG instead
- **Cheap**: No installation of specific tools needed (just a standard webbrowser) → For your Boss: Low Total Cost of Ownership

# Problems

with non-semantic Wikis:

- Search
- Document-centric
- Structuring
- Navigation
- Redundant content → inconsistencies
- Export

## **Wiki features:**

Parsimony, Easy Linking, Creation, Version management, Wiki Syntax, Cheap

with semantic Wikis:

- Guiding the user to re-use categories, relations and attributes
- Can metadata be created automatically from the content?

## **Semantic wiki:**

Metadata creation,  
Metadata usage,  
Application scenarios,  
Integration

## Topics raised by workshop participants

### Using data and/or ontologies

- Spatial views on semantic data/relations
- How can a wiki be exposed as web services?

### Creating data and/or creating ontologies

- Teachers annotating student works
- Wikis showing data base content, stories and navigation paths
- Automatic creation of metadata x 2
- User motivation for metadata creation?
- How to represent complex scientific data?
- Ontology engineering desing patterns

### Meta

- What do wiki people like/don't like on SemWeb?
  - Relation between wikis and semantic wikis: Same community of users?

## Contact Information

- <http://ontoworld.org/wiki/WibKE2006>
- Max Völkel (presenter, organizer)
  - FZI Karlsruhe, voelkel@fzi.de
- Elena Simperl (presenter, organizer)
  - FU Berlin, paslaru@inf.fu-berlin.de
- Sebastian Schaffert (organizer)
  - Salzburg Research
- Sören Auer (organizer)
  - Uni Leipzig