

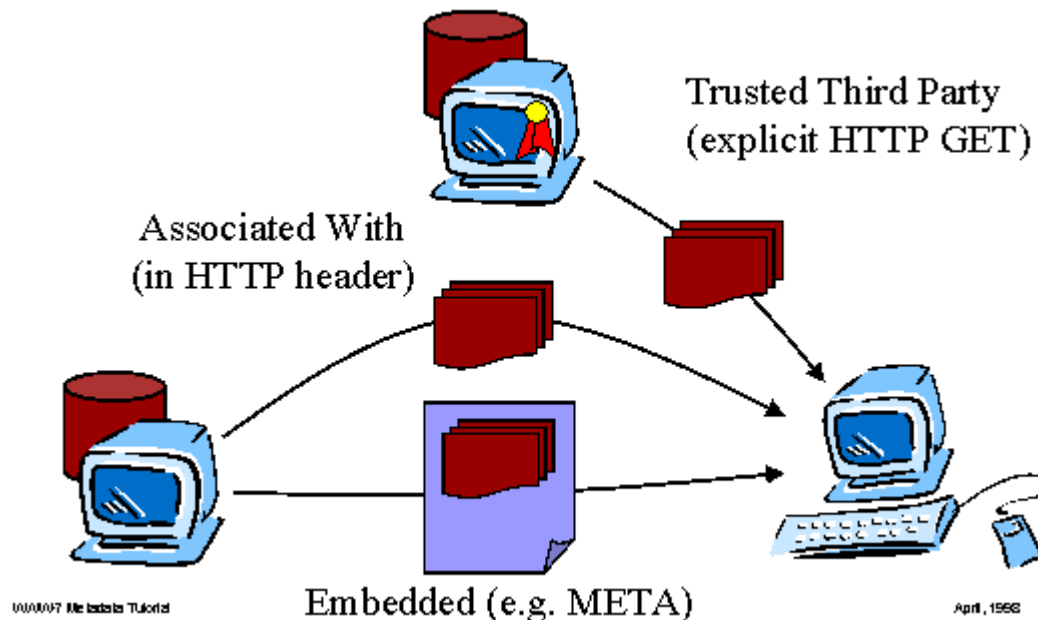


# The Semantic Web in Tourism and its usage in Austria

Wilhelm Loibl, Vienna University of Economics and Business  
TTRA, Archamps 2011

„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“

source: Berners-Lee et al 2001



- Metadata can be directly integrated into the data file (e. g. metadata is embedded into an HTML-file using META-tags)
- Metadata can be put into an extra file and linked to from the data file (e. g. using a link in an HTML-file or using the HTTP-header)
- Metadata can be offered without any link to the data file

- **Ontology:** An ontology can be defined as the explicit, formal description of concepts and their relationships that exist in a certain universe of discourse, together with a shared vocabulary to refer to these concepts. (Cardoso and Sheth, 2006)
  - **Concepts(Classes) and Properties:** types of things one wants to talk about (individual object-> instance)
  - **Relations (+restrictions):** domain (what types of objects can use this relation?), range (what range of values can be assigned to it?)

# Ontologies for the Tourism Domain

	accommodation	transport	gastronomy	attractions	activities	events	time	location	language
Harmonise	✓	✗	✗	✗	✗	✓	✗	✗	RDFS
Hi-Touch	~	~	~	~	~	~	✗	✗	OWL
Quall-ME	✓	✓	✓	✓	✓	✓	✓	~	OWL
OnTour	✓	✗	✗	✗	~	~	~	~	OWL
EON Travelling	✓	✓	✗	~	✗	✗	~	~	OWL
TAGA	✗	✓	✗	✗	✗	✗	✗	~	OWL
cDOTT	✓	✓	✓	✓	✓	✓	✓	✓	OWL
ebSemantics events	✗	✗	✗	✗	✗	✓	✓	~	OWL
ebSemantics accommodation	✓	✗	✗	✗	✗	✗	~	~	OWL
ebSemantics gastro	✗	✗	✓	✗	✗	✗	~	~	OWL

# What is a „microformat“?

```
<span class="vevent">  
<span class="summary">The TTRA Europe Conference 2011</span>  
from <span class="dtstart">2011-04-11</span> to  
<span class="dtend">2011-04-13</span>  
was held at <span class="location">Archamps, France</span>.  
</span>
```

The TTRA Europe Conference 2011 from 2011-04-11 to 2011-04-13 was held at Archamps, France.

For more information please visit: <http://www.microformats.org>

# What semantics can do for you (and what they cannot!)

- They can help with
  - help with information search
  - answer queries
  - draw certain conclusions (based on an ontology)
- They cannot
  - understand the information given on websites (no artificial intelligence)

# Two guiding questions

- What different technologies do exist and how do they differ from each other?
- Which of these technologies are currently used most often?

# Steps taken

- Gather a list of URLs for tourism related websites
- Crawling and analyzing each website
- Visualize results and resolve any inconsistencies (by hand)

- Online Database: Firmen A-Z operated by the Austrian Chamber of Commerce
  - firms are registered with the Austrian Chamber of Commerce
  - Additional information has to be entered by the entrepreneurs themselves (like e-mail addresses, websites etc.)
- Using Meta-Search-Engine to query Database (by creating special plugin for this Database)
- Results saved in csv-files

# Using customized crawler

- Coded in Java
- Open source
- Configurable through XML-config files
- First crawles all sites whose URLs are given to it in a list (csv-file)
- Analyses each individual page according to the pattern given in the config file

# Websites and Webpages crawled

state	total		crawled		with semantics	
	number of sites	number of pages	sites	number of sites	number of pages	
Burgenland	271	8601	234	2	26	
Carinthia	1330	44742	1187	10	288	
Lower Austria	703	22577	645	3	711	
Styria	1042	41096	966	11	159	
Tyrol	1386	64738	898	9	36	
Vorarlberg	298	17856	278	3	44	
Vienna	951	50403	894	16	691	
Salzburg	1168	61482	1093	22	528	
Upper Austria	690	22612	628	6	63	
	7839	334107	6823	82	2546	

# Semantics found (by state)

	vevent	vcard- span	vcard- div	adres- s-span	address- div	rdf link	rdf- tag	vcalendar	full semantics	partly
Burgenland	0	0	0	0	0	0	26	0	0	0
Carinthia	0	7	43	8	40	147	51	0	50	40
Lower Austria	0	2	0	708	0	0	1	0	710	0
Styria	0	3	30	11	0	13	102	0	44	13
Tyrol	0	1	2	2	0	1	33	0	39	0
Vorarlberg	0	0	35	39	0	0	5	0	39	0
Vienna	0	2	2	524	0	45	119	0	526	46
Salzburg	0	61	0	362	0	60	49	0	361	118
Upper Austria	0	1	0	4	0	0	58	0	6	0
	0	77	112	1658	40	266	444	0	1775	217

- Improve crawler
  - Queued filters
  - Search HTTP-header fields more thoroughly
- Rerun
- Find out reasons for (not) applying semantics on corporate websites
- create microformats for the tourism domain + provide software for annotating webpages and harvest information

# Questions and Comments?

Kommentare:

Vorschau

Absenden