The Web of Data for E-Commerce in One Day

A Hands-on Introduction to the GoodRelations Ontology, RDFa, and Yahoo! SearchMonkey

May 31, 2009, Crete, Greece

Martin Hepp
Universität der Bundeswehr München, Munich, Germany

Michael Hausenblas
Digital Enterprise Research Institute (DERI), Galway, Ireland
Logistics

09:00-10:30  Overview and Motivation: Why the Web of Data is Now 15’
Quick Review of Prerequisites 15’
The GoodRelations Ontology: E-Commerce on the Web of Data 60’

10:30-11:00  Coffee Break

11:00-13:00  RDFa: Bridging the Web of Documents with the Web of Data 45’
Expressing GoodRelations in RDFa: A Running Example 30’
GoodRelations – Advanced Topics 45’

13:00-14:30  Lunch Break

14:30-16:00  Querying the Web of Data for Offerings – SPARQL 15’
Hands-on Exercise: Annotating a Web Shop 45’
Querying the Web of Data – Exercises 15’
Publishing Semantic Web Data: Make Your RDF Available 15’

16:00-16:30  Coffee Break

16:30-18:00  Yahoo SearchMonkey and Yahoo BOSS 45’
RDFa Advanced Topics 30’
Discussion, Conclusion, Feedback Round 15’
Quick Review of Prerequisites

Martin Hepp
Learning Goals

In this part, we will

• make sure all participants have sufficient knowledge of related topics,

and

• show how to install the Twinkle software.
Prerequisites for the Tutorial

• Markup Languages
  – XML, HTML, XHTML

• Semantic Web Basics
  – URIs
  – RDF
  – RDFS and OWL

• Tooling and Infrastructure
  – Editors
  – Repositories and Reasoners
  – Frameworks / APIs

• Linked Data Principles
Markup Languages

XML, HTML, and XHTML
Foundations of Markup Languages

• **Markup:** In traditional publishing: "marking up" a manuscript - typeface, style, size
• 1967: William Tunnicliffe: GenCode
• 1973: Charles Goldfarb: Generalized Markup Language (GML)

• **Principle:** Content + Tags
  – This `<bold>`word</bold>` is in boldface.
  – This **word** is in boldface.
Principle: Strong Tagging

• Opening and closing tag required
  – This <strong>word</strong> is in boldface.
  – Shortcut for empty elements
    • <br/> for <br/>(new line in HTML)
Nesting of Elements

• Tags can be nested
  – This is <bold>all in bold, <italics>some parts are also in italics.</italics></bold>
  – This is all in bold, some parts are also in italics.

• This allows for the representation of arbitrary tree structures.
Attributes in Tags

• Tags can be augmented by attributes:
  – This `<bold size="2">word</bold>` is in boldface.

• Allows for adding parameters to attributes.

Very important for this tutorial!
Well-formed Documents

• Documents must follow certain rules in order to be consistent trees:
  – All opening tags must have a matching closing tag.
  – Consistent nesting: Closing tags must follow in the exact inverse order of the opening tags.
  – All attribute values must be properly quoted (").
  – The document must have a single root element.
Typical Mistakes

- Empty elements lack a trailing slash
  - `<br/>

- Missing closing brackets / lesser signs
  - `<br/` instead of `<br/>

- Missing quotes for attributes or wrong quotation marks
  - `<bold size=2>`word`</bold>` instead of `<bold size="2">`word`</bold>` instead of "`<bold size="2">`word`</bold>`

- Warning: Some Web browsers may still display the document correctly
  - Web browsers try to deal with invalid content ("lax parsing")
  - May hide mistakes that will strike back later...
Document Grammars

• Specify the allowed set of tags, attributes, and other elements and their usage

• E.g. nesting, cardinality constraints
Valid Documents

• A document that meets the constraints of the specified document grammar

• Examples
  – Includes all mandatory elements
  – Uses no unknown additional elements
  – Follows all rules on nesting and cardinality
Markup Languages for the Web: HTML 1.x - 4.x

- HTML Tags (Oct. 1991)
  - An informal CERN document listing twelve HTML tags
- HTML 2.0 (Nov. 1995)
  - declared obsolete/historic by RFC 2854 (2000)
- HTML 3.2 (Jan 1997)
  - W3C Recommendation
- HTML 4.0 (Dec. 1997)
  - Three flavors
- HTML 4.01 (Dec. 1997 + updates)
  - Current version
Extensible Markup Language: XML

• Language for specifying markup languages (and exchanging data)
• Intended to facilitate data interchange between information systems
• Started as simplified subset of SGML
HTML 4.01: Document Skeleton

```
<!DOCTYPE html>
<html>
  <head>
    <title>Hello HTML</title>
  </head>
  <body>
    <p>Hello World!</p>
  </body>
</html>
```
XHTML - Expressing HTML in XML

• Growing need to turn Web markup / contents into true XML data
• Combination of data
• Reuse
• Use XML technology and tools
• Etc.
The Semantic Web for Web Masters

URIs, RDF, RDFS and OWL
What is the Semantic Web?

The Gopher – HTTP/HTML Analogy

1990: Gopher
Textual information

mid 1990s: HTML
Multimedia added

2006 and beyond: RDF, OWL
Meaning added

Gopher screenshot courtesy of the University of Minnesota.

Universität München
Web of Data ("Semantic Web")
Core Semantic Web Technology Pillars

- **Global Identifier:** URIs for everything
- **Data Model:** RDF - A data model for exchanging conceptual graphs based on triples
  - Compatible with the design principles of the Web (especially with its distributed nature)
  - Triple: (Subject, Predicate, Object)
  - Exchange syntax: RDF/XML, N3, RDFa etc.
- **Ontology Languages:** RDFS and OWL - formal languages that help reduce ambiguity and codify implicit facts
  - foo:human rdfs:subClassOf foo:mammal
- **Query Language & Interface:** SPARQL - standardized query language and endpoint interface for RDF data
- **LOD Principles:** Best practices for keeping the current Web and the Web of Data compatible
Simplified Process of Using the Semantic Web

• Find or create ontology / vocabulary
  – “Ontology Engineering”

• Create data expressed using that vocabulary

• Publish the data

• Query / reuse / combine the data
Tooling and Infrastructure

Editors
Repositories and Reasoners
Frameworks / APIs
Parsers, Repositories, Reasoners

- RDF/XML
- N3
- XHTML + RDFa
- Reasoner
- Explicit Model
- Implicit Model
- Query
- Repository
Frameworks, Libraries, and APIs

• **Jena** Semantic Web Framework: Java framework for building Semantic Web applications.
  – http://jena.sourceforge.net/

• **RDFLib**: Python library for working with RDF
  – http://www.rdflib.net/

• **Redland** RDF Libraries (aka librdf): C-based library with APIs in Perl, Python, Tcl and Java.
  – http://librdf.org
Twinkle

Installations and Usage
Twinkle: A SPARQL Query Tool

http://www.ldodds.com/projects/twinkle/
Twinkle: Installation

• Requires Java 1.5 or higher
• Download the distribution and unzip it into a new directory:
• Replace the file config.n3 in the "etc" subdirectory by the file available at
  – Rename it to config.n3 after downloading
• Open a command prompt and execute the following:
  – java -jar twinkle.jar
Quizzes

• No quizzes for this part
Thank you.