Current State of the Semantic Web

Kore Nordmann

March 6, 2009

- ▶ Kore Nordmann, `kore@php.net`
- ▶ Long time PHP developer
- ▶ Regular speaker, author, etc.
- ▶ Studies computer science in Dortmund
- ▶ Consultancy and software development
- ▶ Active open source developer:
  - ▶ eZ Components (Graph, WebDav, Document), Arbit, PHPUnit, Torii, PHPillow, KaForkL, Image 3D, WCV, ...

- ► Semantics
  - ► Association of meaning to structures
    - ► Orginates from: $\sigma\eta\mu\alpha\nu\tau\iota\kappa\sigma\varsigma$ (semantikos), "significant"

- Semantics
  - Association of meaning to structures
    - Orginates from: $\sigma\eta\mu\alpha\nu\tau\iota\kappa o\varsigma$ (semantikos), "significant"
    - Orginates from: $\sigma\eta\mu\alpha\iota\nu\omega$ (semaino), "to signify, to indicate"

- ► Semantics
  - ► Association of meaning to structures
    - ► Orginates from: $\sigma\eta\mu\alpha\nu\tau\iota\kappa o\varsigma$ (semantikos), "significant"
    - ► Orginates from: $\sigma\eta\mu\alpha\iota\nu\omega$ (semaino), "to signify, to indicate"
    - ► Orginates from: $\sigma\eta\mu\alpha$ (sema), "sign, mark, token"

- ▶ In computer science:
  - ▶ Meaning of a word in some language
    - ▶ Different words of the same language or words from different languages may have the same semantics
    - ▶ "Meaning" not clearly defined

- In computer science:
  - Meaning of a word in some language
    - Different words of the same language or words from different languages may have the same semantics
    - "Meaning" not clearly defined
  - Connotation

- In computer science:
  - Meaning of a word in some language
    - Different words of the same language or words from different languages may have the same semantics
    - "Meaning" not clearly defined
  - Connotation
  - Denotation
    - Extensional / Intensional

- ▶ Semantic web
  - ▶ Web
    - ▶ Generally means documents / data on the internet
    - ▶ Graph of linked documents
    - ▶ Mostly HTML or binary data

- ► Semantic web
  - ► Web
    - ► Generally means documents / data on the internet
    - ► Graph of linked documents
    - ► Mostly HTML or binary data
  - ► Semantic web
    - ► Association of semantics to documents
    - ► Association of semantics to links

- ▶ Semantic Web initiative
  - ▶ Common framework defined by the W3C
  - ▶ Mainly includes: RDF, SPARQL and OWL
  - ▶ Initiated by Tim Berners-Lee

- ▶ Semantic Web initiative
  - ▶ Common framework defined by the W3C
  - ▶ Mainly includes: RDF, SPARQL and OWL
  - ▶ Initiated by Tim Berners-Lee
  - ▶ *Any other related technologies you want to mention?*
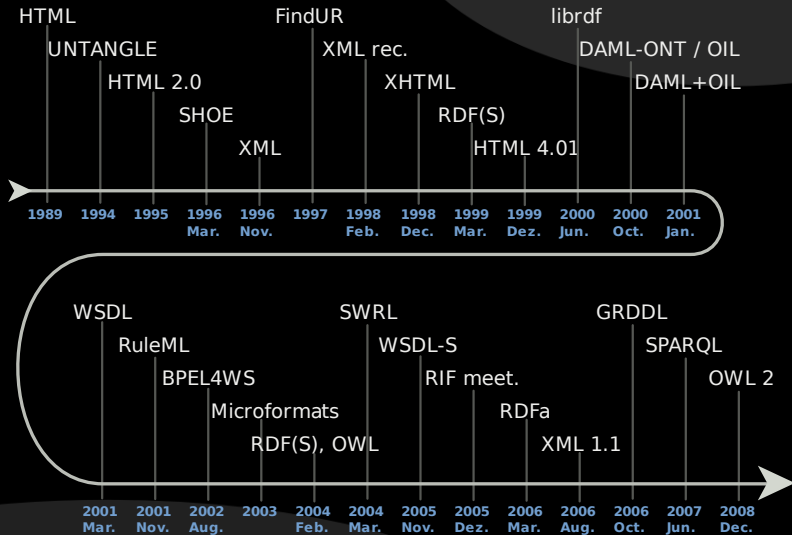
The term

History

Expressing semantics

Outlook

- ▶ First version of HTML in 1989
  - ▶ Included markup for documents common at CERN
- ▶ Development focussed on media contents
  - ▶ Visual markup
- ▶ HTML used as a presentation layer
  - ▶ Application knew document semantics

- Amount of documents not discoverable by humans
- Search engines evolved
    - Answer only trivial questions
        - "Which are the capital cities of African countries?"
- The Sematic Web wants to solve this

General history

▶ XHTML 1 and HTML 4 offer document markup:

```
1  <html>
2    <head>
3      <title>Semantic Web</title>
4      <meta name="author" content="Kore Nordmann" />
5    </head>
6    <body>
7      <h1>Introduction</h1>
8      <p>The term <em>semantic web</em> describes
9      the vision, that the information, ...</p>
10   </body>
11 </html>
```

Visual semantics

- ▶ Common markup still not expressable (shop, forum, ..)
- ▶ Humans associate semantics depending on style
- ▶ Elements are misused and styled.
  - ▶ Element semantics hard to decide for software

- Common markup still not expressable (shop, forum, ..)
- Humans associate semantics depending on style
- Elements are misused and styled.
  - Element semantics hard to decide for software
  - *Did you ever misuse HTML elements?*

Current development

- ▶ XHTML 2 extended document centric markup
  - ▶ Removal of deprecated elements (`<b>`, `<i>`)
  - ▶ Adding structural markup (`<section>`)
  - ▶ Non-semantical markup still existing (`<div>`, `<span>`)
- ▶ Countered by (X)HTML 5

- ▶ XHTML 2 extended document centric markup
  - ▶ Removal of deprecated elements (`<b>`, `<i>`)
  - ▶ Adding structural markup (`<section>`)
  - ▶ Non-semantical markup still existing (`<div>`, `<span>`)
- ▶ Countered by (X)HTML 5
  - ▶ "However, it lacks elements to express the semantics of many of the non-document types of content often seen on the Web. For instance, forum sites, auction sites, search engines, online shops, and the like, do not fit the document metaphor well, and are not covered by XHTML2." – Ian Hickson, HTML 5, W3C Working Draft 22 January 2008

The term

History

Expressing semantics

Outlook

► Reuse (X)HTML class attributes for trivial semantics association

```
1  <div class="vevent">
2    <span class="summary">Semantic Web</span>:
3    <abbr class="dtstart" title="2009-03-06">March 6th<
         /abbr>—
4    <abbr class="dtend" title="2009-03-06">6th</abbr>,
5    at <span class="location">PHPUG Cologne</span>
6  </div>
```

Mircroformat standards

- hCalendar
- hCard
- rel (nofollow, license, tag)
- Votelinks
- XFN
- XMDP

- hCalendar
- hCard
- rel (nofollow, license, tag)
- Votelinks
- XFN
- XMDP
  - *Ever used one of those?*

- ▶ Very limited set of applications
- ▶ Not extensible (not forward compatible), no namespaces
- ▶ No structural definition, no ontologies
- ▶ May still be the Semantic Web technology

- ▶ Very limited set of applications
- ▶ Not extensible (not forward compatible), no namespaces
- ▶ No structural definition, no ontologies
- ▶ May still be the Semantic Web technology
  - ▶ "HTML, CSS, JavaScript and DOM will be the basic content standards in the foreseeable future. I think evolution on the web will be based on these formats, and this is what WHAT and AJAX do. We will also see a bunch of Microformats being developed, and that's how the semantic web will be built, I believe." – Hakon Wium Lie, CTO of Opera Software (2005)

- Resource Description Form
- Using (Subject, Predicate, Object) tripels

```
1  @prefix dc: <http://purl.org/dc/elements/1.1/>.
2    </blog/the_long_way_to_semantic_web.html>
3      dc:title "The long way to a semantic web";
4      dc:publisher "Kore Nordmann".
```

- ▶ Resource Description Form
- ▶ Using (Subject, Predicate, Object) tripels
  - ▶ Links and Resources may be "Subject"

```
1  @prefix dc: <http://purl.org/dc/elements/1.1/>.
2    </blog/the_long_way_to_semantic_web.html>
3       dc:title "The long way to a semantic web";
4       dc:publisher "Kore Nordmann".
```

- ▶ Resource Description Form
- ▶ Using (Subject, Predicate, Object) tripels
  - ▶ Links and Resources may be "Subject"
  - ▶ Reification: "Object" may again a resource

```
1  @prefix dc: <http://purl.org/dc/elements/1.1/>.
2    </blog/the_long_way_to_semantic_web.html>
3      dc:title "The long way to a semantic web";
4      dc:publisher "Kore Nordmann".
```

RDF usage

- ▶ RDF commonly embedded in
  - ▶ XHTML
  - ▶ SVG
  - ▶ XMP (Extensible Metadata Platform)
- ▶ Usage of any namespace / ontologies

```
1  <RDF xmlns="http://www.w3.org/1999/02/22-rdf-syntax-
       ns"
2       xmlns:dc="http://purl.org/dc/elements/1.1/">
3     <Description about="http://kore-nordmann.de">
4       <dc:creator>Kore Nordmann</dc:creator>
5       <dc:rights>CC by-sa</dc:rights>
6     </Description>
7  </RDF>
```

- ▶ RDF commonly embedded in
  - ▶ XHTML
  - ▶ SVG
  - ▶ XMP (Extensible Metadata Platform)
- ▶ Usage of any namespace / ontologies
  - ▶ *Who already makes use of RDF?*

```
1  <RDF xmlns="http://www.w3.org/1999/02/22-rdf-syntax-
       ns"
2       xmlns:dc="http://purl.org/dc/elements/1.1/">
3     <Description about="http://kore-nordmann.de">
4       <dc:creator>Kore Nordmann</dc:creator>
5       <dc:rights>CC by-sa</dc:rights>
6     </Description>
7  </RDF>
```

- ▶ Dublin Core
- ▶ Document metadata semantics
- ▶ Not only used with RDF
  - ▶ Embeddable in HTML meta tags
- ▶ Developed already 1994
  - ▶ Contains of 15 core elements
  - ▶ Additional optional elements

RDFa

- ▶ Embed RDF tripels directly in XHtml
- ▶ Still in draft state

```
1  <div xmlns:dc="http://purl.org/dc/elements/1.1/">
2    <h2 property="dc:title">Das semantische Web</h2>
3    <h3 property="dc:creator">Kore Nordmann</h3>
4  </div>
```
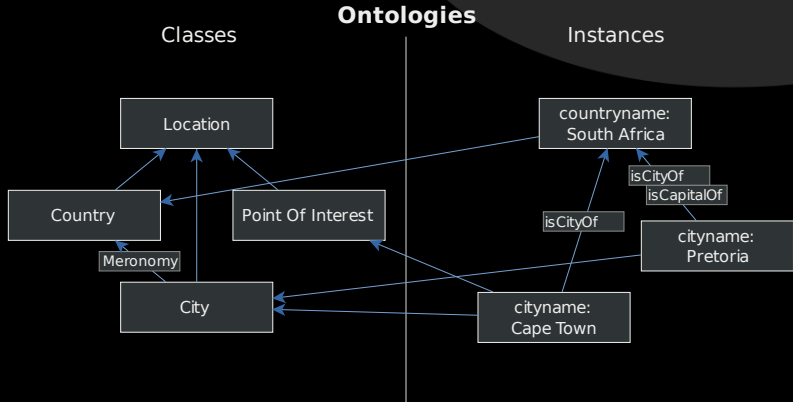
- ▶ Requires a custom XHTML+RDFa-DTD

- ▶ Embed RDF tripels directly in XHtml
- ▶ Still in draft state

```
1  <div xmlns:dc="http://purl.org/dc/elements/1.1/">
2    <h2 property="dc:title">Das semantische Web</h2>
3    <h3 property="dc:creator">Kore Nordmann</h3>
4  </div>
```

- ▶ Requires a custom XHTML+RDFa-DTD
  - ▶ "The authors know of no deployed Web browser that will fail to present an HTML document as intended after adding RDFa markup to the document. However, publishers should be aware that RDFa will not validate in HTML4 at this time." – RDFa Primer 1.1
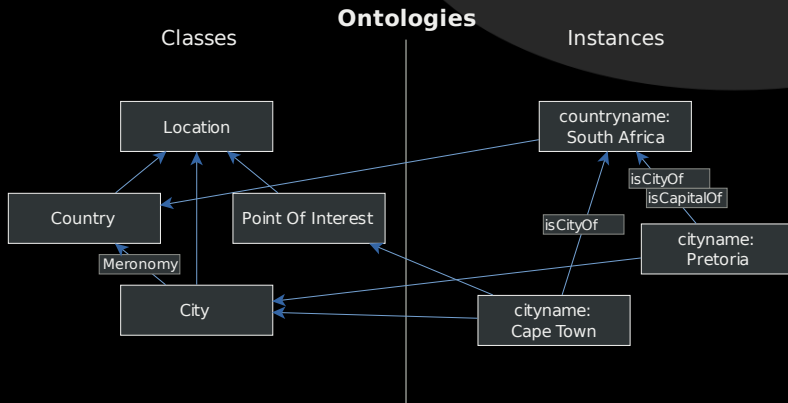
- ▶ General representation of semantics
- ▶ Different variants for different semantics
  - ▶ Frame-Logics
  - ▶ Description-Logics

- ▶ General representation of semantics
- ▶ Different variants for different semantics
  - ▶ Frame-Logics
  - ▶ Description-Logics
- ▶ Basically always consist of:
  - ▶ Classes
  - ▶ Instances
  - ▶ Properties
  - ▶ Relations between all those

Ontologies

Classes — Instances

Location

countryname: South Africa

Country — Point Of Interest

isCityOf
isCapitalOf

isCityOf

cityname: Pretoria

Meronomy

City

cityname: Cape Town

► Additionally ontologies may contain axioms
  ► "Each country has a capital."

RDF-Schema

- ▶ What DTD / XML-Schema is to XML, RDF-Schema is to RDF
- ▶ Simple vocabularies, no axioms, no complex inferences

```
1  <rdf:RDF
2    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns
         #"
3    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
4
5    <rdfs:Class rdf:about="http:/example.org/#Location">
6      <rdfs:Label>Location</rdfs:Label>
7    </rdfs:Class>
8
9    <rdfs:Class rdf:about="http:/example.org/#City">
10     <rdfs:Label>City</rdfs:Label>
11     <rdfs:subClassOf rdf:resource="#Location" />
12   </rdfs:Class>
13 </rdf:RDF>
```

- ▶ Web Ontology Language (OWL)
- ▶ Extends RDF-Schema with more logical operations
- ▶ Different profiles
    - ▶ OWL Lite - Inferences can be calculated in exponential time
    - ▶ OWL DL - Matches a popular DL variant, also decidable
    - ▶ OWL Full - Superset of RDF-Schema, not decidable

- ▶ Web Ontology Language (OWL)
- ▶ Extends RDF-Schema with more logical operations
- ▶ Different profiles
  - ▶ OWL Lite - Inferences can be calculated in exponential time
  - ▶ OWL DL - Matches a popular DL variant, also decidable
  - ▶ OWL Full - Superset of RDF-Schema, not decidable
- ▶ OWL 2 profiles decidable in PSPACE or LOGSPACE

- ▶ Web Ontology Language (OWL)
- ▶ Extends RDF-Schema with more logical operations
- ▶ Different profiles
    - ▶ OWL Lite - Inferences can be calculated in exponential time
    - ▶ OWL DL - Matches a popular DL variant, also decidable
    - ▶ OWL Full - Superset of RDF-Schema, not decidable
- ▶ OWL 2 profiles decidable in PSPACE or LOGSPACE
- ▶ Can express everything, expressable by frame logics or description logics

► Build class "interesting capitals"

```
1  <Class
2    xmlns="http://www.w3.org/2002/07/owl"
3    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-
         ns"
4    rdf:ID="CaptialOfInterest">
5    <intersectionOf>
6      <Class rdf:about="#Capital"/>
7      <Class rdf:about="#PointOfInterest"/>
8    </intersectionOf>
9  </Class>
```

▶ Build class "interesting capitals"

```
1  <Class
2    xmlns="http://www.w3.org/2002/07/owl"
3    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-
          ns"
4    rdf:ID="CaptialOfInterest">
5    <intersectionOf>
6      <Class rdf:about="#Capital"/>
7      <Class rdf:about="#PointOfInterest"/>
8    </intersectionOf>
9  </Class>
```

▶ Just build your ontologies, reasoning is performed by software

► "Which are the capital cities of African countries?"

```
1    PREFIX my: <http://example.org/>
2    SELECT ?capital
3    WHERE {
4      ?x my:cityname ?capital.
5      ?x my:isCapitalOf ?y.
6      ?y my:isInContinent my:Africa.
7    }
```

- Lots of different standards
- Reuses known stadards like XML and XML-Schema
- GRDDL will help extracting data from XHTML
- Triple stores with query support are already implemented:
  - librdf, `http://librdf.org`
  - ARC, `http://arc.semsol.org/`

- ▶ Lots of different standards
- ▶ Reuses known stadards like XML and XML-Schema
- ▶ GRDDL will help extracting data from XHTML
- ▶ Triple stores with query support are already implemented:
    - ▶ librdf, http://librdf.org
    - ▶ ARC, http://arc.semsol.org/
- ▶ Yahoo! SpiderMonkey already actively uses such data

The term

History

Expressing semantics

Outlook

- ▶ The Semantic Web already emerges
    - ▶ Not broad support
    - ▶ Entering the Semantic Web is easy with Microformats
- ▶ Custom applications require custom ontologies
- ▶ RDF and OWL will be the major technologies
    - ▶ Backed by research

- ▶ HTML semantic will get irrelevant
  - ▶ Together with CSS pure presentational markup

- ▶ HTML semantic will get irrelevant
  - ▶ Together with CSS pure presentational markup
- ▶ Documents can be represented by any XML markup ($+$ Dokbook)

- ▶ HTML semantic will get irrelevant
  - ▶ Together with CSS pure presentational markup
- ▶ Documents can be represented by any XML markup (+ Dokbook)
  - ▶ Wait for CSS3

- ▶ HTML semantic will get irrelevant
  - ▶ Together with CSS pure presentational markup
- ▶ Documents can be represented by any XML markup ($+$ Dokbook)
  - ▶ Wait for CSS3
  - ▶ Wait for XLink support

- ▶ Open questions?
- ▶ Further remarks?
- ▶ Contact
  - ▶ Mail: `kore@php.net`
  - ▶ Web: `http://kore-nordmann.de` (Slides will be available here soonish)

Reference

- W3C Semantic Web Activity:
  `http://www.w3.org/2001/sw/`
- `http://kore-nordmann.de/blog/`
  `current_state_of_semantic_web.html`
- "Semantic Web, Grundlagen", Springer Verlag, ISBN-13:
  978-3540339939