

Jan Zemánek
jan.zemanek@gmail.com



OD MIKROFORMÁTŮ K RDF



Přehled

- Data
 - Mikroformáty
 - RDF
 - GRDDL
 - RDFa
- 



<Data>






Rozlišujeme

1. Nestrukturovaná data
2. Semi-strukturovaná data
3. Strukturovaná data



Nestrukturovaná data

- Text
 - Obrázky
 - Zvuk
 - Video
- 




Strukturovaná data





Strukturovaná data

- Relační model
 - RDF
 - RDFS, OWL
 - Topic Maps
 - ...
- 




Semi-strukturovaná data

- Samo-popisující („self-describing“)
 - Schéma není (zvlášť)
 - Schéma je součástí dat
- Neúplná data
 - Dáno
 - Rozmanitostí dat
 - Chybějícími daty
- Flexibilní reprezentace dat
 - Dokáže se vypořádat s neúplností dat
- Semi-strukturovaný **datový model**
 - Výměna a integrace dat
 - z heterogenních zdrojů



Semi-strukturovaná data

- XML
 - Object Exchange Model (OEM)
 - (RDF)
 - ...
 - a
 - Mikroformáty
- 

XML vs. OEM

```
<person id="o123">  
  <name> Alan </name>  
  <age> 42 </age>  
  <email> ab@com </email>  
</person>
```

```
{ person: &o123  
  { name: "Alan",  
    age: 42,  
    email: "ab@com" }  
}
```

Mikroformát hCard

John Q. Public
10 Bolyston St.
Boston, MA 01210

```
<!DOCTYPE html
  PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
  "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" lang="en">
  <head>
    <title>hCard Example</title>
    <meta http-equiv="content-type" content="application/xhtml+xml; charset=UTF-8" />
  </head>
  <body>
    <div class="vcard">
      <span class="fn n">
        <span class="given-name">John</span>
        <abbr class="additional-name" title="Quinlan">Q.</abbr>
        <span class="family-name">Public</span>
      </span><br />
      <span class="adr">
        <span class="street-address">10 Bolyston St.</span><br />
        <span class="locality">Boston</span>,
        <abbr class="region" title="Massachusetts">MA</abbr>
        <span class="postal-code">01210</span>
      </span>
    </div>
  </body>
</html>
```



</Data>





<Mikroformáty>



Browser vs. člověk

- Vlevo co „vidí“ browser
 - Browser = (nebo) obecně software



- Vpravo co vidí člověk

Yahoo! a mikroformáty

- LinkedIn profil s mikroformátem hResume

```
</head>
<body id="www-linkedin-com" class="public-profile">

<script type="text/javascript">document.body.className += " js";</script>

<div class="hresume">
  <div class="profile-header">
    <div class="masthead vcard contact portrait">
      <div id="nameplate">
        <h1 id="name"><span class="fn n"><span class="given-name">Barack</span> <span class="family-name">Obama</span></span></h1>
      </div>
      <div class="content">
        <div class="info">
          <div class="image">
          <p class="headline title">

          President of the United States of America

          </p>
          <div class="adr">
            <p class="locality">

            Washington D.C. Metro Area

            </p>
          </div>
        </div>
      </div>
    </div>
  </div>
</div>
```

- Výsledky vyhledávání na Yahoo!

[Barack Obama - President of the United States of America](#)

The administration can't only be about me. It must be about us - it must be about what we can do together. It is about reclaiming the meaning of...

- Location: Washington D.C. Metro Area
- Current: President, United States of America

www.linkedin.com/in/barackobama - [Cached](#) - [More from this site](#)



Mikroformáty

- <http://microformats.org/>
- „Hack“ umožňující vyjádřit v (X)HTML a XML (RSS, Atom) dokumentech jednoduchou **sémantiku dat**
 - Prostřednictvím „vlepování“ sémantických značek do „atributů“ (X)HTML a XML <elementů/>
- Mikroformáty nejsou zastřešeny žádnou standardizační organizací

(X)HTML s mikroformáty

- „čisté“ (X)HTML

```
<div>
  <div>Joe Doe</div>
  <div>The Example Company</div>
  <div>604-555-1234</div>
  <a href="http://example.com/">http://example.com/</a>
</div>
```

- sémanticky označované (X)HTML

- (X)HTML + mikroformát hCard

```
<div class="vcard">
  <div class="fn">Joe Doe</div>
  <div class="org">The Example Company</div>
  <div class="tel">604-555-1234</div>
  <a class="url" href="http://example.com/">http://example.com/</a>
</div>
```



Vybrané mikroformáty

- **hCard**
 - Pro kontaktní informace („vizitka“)
 - **hCalendar**
 - Pro události
 - **hResume**
 - CV
 - **Geo**
 - Pro vyjádření geografické polohy (šířka, délka)
 - **XHTML Friends Network (XFN)**
 - Umožňuje vyjádřit sociální vztahy
 - Přítel, spolupracovník, soused, ...
 - ...
- 



hCard detailněji

- <http://microformats.org/wiki/hcard>



hCard - příklad

CommerceNet
<http://www.commerce.net/>
Work:
169 University Avenue

Palo Alto, CA 94301

USA
Work +1-650-289-4040
Fax +1-650-289-4041
Email info@commerce.net

```
<div class="vcard">
  <a class="fn org url" href="http://www.commerce.net/">CommerceNet</a>
  <div class="adr">
    <span class="type">Work</span>:
    <div class="street-address">169 University Avenue</div>
    <span class="locality">Palo Alto</span>,
    <abbr class="region" title="California">CA</abbr>
    <span class="postal-code">94301</span>
    <div class="country-name">USA</div>
  </div>
  <div class="tel">
    <span class="type">Work</span> +1-650-289-4040
  </div>
  <div class="tel">
    <span class="type">Fax</span> +1-650-289-4041
  </div>
  <div>Email:
    <span class="email">info@commerce.net</span>
  </div>
</div>
```



</Mikroformáty>





<RDF />

- Odkok do alternativní prezentace



<GRDDL>



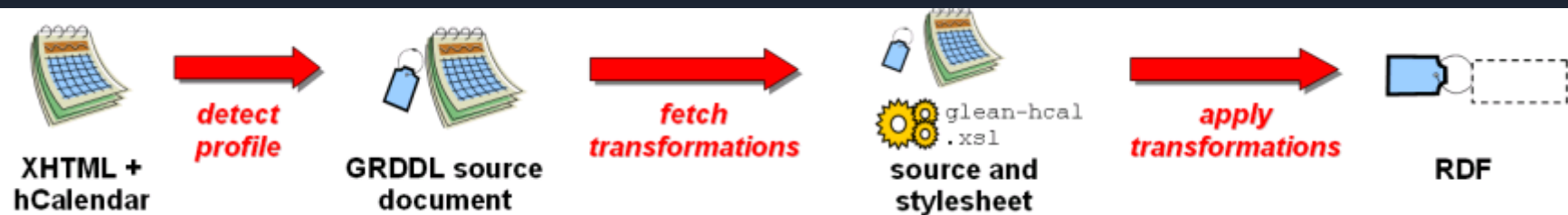
GRDDL

- <http://en.wikipedia.org/wiki/GRDDL>
- GRDDL = Gleaning Resource Descriptions from Dialects of Languages
- „Most“ mezi mikroformáty (apod.) a RDF (Sémantickým Webem)
- Umožňuje získávat ze sémanticky označovaných (X)HTML a XML dokumentů **RDF data**
 - K tomu využívá **XSLT** transformační styly

(X)HTML + XSLT = RDF

- Pro každý mikroformát zvláštní XSLT styl
 - <http://esw.w3.org/topic/CustomRdfDialects>

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN" "http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">  
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" lang="en">  
  <head profile="http://www.w3.org/2003/g/data-view">  
    <title>Robin's Schedule</title>  
    <link rel="transformation" href="http://www.w3.org/2002/12/cal/glean-hcal"/>  
  </head>  
  <body>  
    ...
```





Demo

- Demonstration of an RDF in XHTML processor
 - <http://www.w3.org/2003/11/rdf-in-xhtml-demo>
- Joe Lambda's homepage
 - <http://www.w3.org/2003/12/rdf-in-xhtml-xslts/complete-example.html>



</GRDDL>





<RDfa>





RDFa

- <http://en.wikipedia.org/wiki/Rdfa>
- „RDFa uses **attributes** from XHTML's meta and link elements, and generalises them so that they are usable **on all elements**.“
- RDFa (RDF- in – attributes)
 - Soubor atributů, které vyjadřují metadata v XML
 - Atributy
 - about, rel, rev, href, src, resource, property, content, datatype, typeof

XHTML + RDFa

- XHTML + RDFa má definován zvláštní DTD

```
<?xml version="1.0" encoding="UTF-8"?>  
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN"  
  "http://www.w3.org/MarkUp/DTD/xhtml-rdfa-1.dtd">
```

- Entita identifikována URI
„http://example.com/books/wikinomics“ má
název „Wikinomics“, jejím autorem je „Don
Tapscott“ a ...

```
<div xmlns:dc="http://purl.org/dc/elements/1.1/"  
  about="http://www.example.com/books/wikinomics">  
  <span property="dc:title">Wikinomics</span>  
  <span property="dc:creator">Don Tapscott</span>  
  <span property="dc:date">2006-10-01</span>  
</div>
```



Demo

- RDFa Distiller
 - <http://www.w3.org/2007/08/pyRdfa/>
- Ivan Herman's FOAF document in XHTML+RDFa
 - <http://www.ivan-herman.net/foaf.html>

Výhody RDFa

- oproti mikroformátům
 - RDFa
 - Stačí **jeden** XSLT transformační styl
 - pro všechny ontologie
 - Mikroformáty
 - **Pro každý** mikroformát **zvláštní** XSLT styl
 - RDFa
 - Dají se využít všechny již existující ontologie
 - Mikroformáty
 - Musí vzniknout nový mikroformát
 - Musí mít ekvivalentní ontologii, do které se transformuje
 - Pokud již ontologie existuje, stejně se musí vytvořit nový mikroformát
- Předpoklad: Mikroformáty se pro další zpracování transformují do RDF



</RDFa>



Zdroje

- Mikroformáty
 - <http://en.wikipedia.org/wiki/Microformats>
 - <http://microformats.org/>
 - Firefox Extensions
 - <http://microformats.org/wiki/firefox-extensions>
- GRDDL
 - <http://en.wikipedia.org/wiki/GRDDL>
- RDFa
 - <http://en.wikipedia.org/wiki/Rdfa>
- <http://www.dcs.bbk.ac.uk/~ptw/teaching/ssd/notes.html>
- <http://queue.acm.org/detail.cfm?id=1103834>
- http://en.wikipedia.org/wiki/Semi-structured_model