Web 2.0

Lecture 10: Annotations

doc. Ing. Tomáš Vitvar, Ph.D.

tomas@vitvar.com • @TomasVitvar • http://vitvar.com



Czech Technical University in Prague

Faculty of Information Technologies • Software and Web Engineering • http://vitvar.com/courses/w20

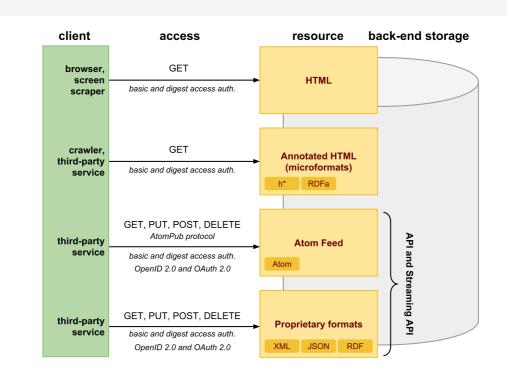






Modified: Thu May 03 2018, 10:17:12 Humla v0.3

Data on the Web



Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

Data Syntax, Structure and Semantics **Semantic Web Layered Cake Web Data Formats** syntax and formal semantics syntax and semantics (structure) Annotation mechanisms Annotation mechanisms microformats (hCard, hResume, ...), microdata Domain-specific semantics Domain-specific semantics (ontologies) atom extensions, vCard, vResume, OWL Full OWL DL Rule Languages Reasoner **OWL Lite** RDF Schema (RDFS) Atom extensions (e.g., GData) AtomPub **SPARQL** RDF Atom XML Schema JSON Schema XML Schema

XML Schema

XML N3 Turtle

XQuery and Namespaces

VPath UNICODE

HTTP

AtomPub

Atom Extensions (e.g., GData)

Atom

XML Schema

XML JSON

XQuery and Namespaces

XPath

UNICODE

HTTP

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

- 3 -

Overview

- Microformats
- Microdata
- RDF and RDFa
- OpenGraph Protocol

Microformats

What is a microformat

- How to embed data in HTML, XHTML, Atom, and XML
 - → data: vCard, vResume, vRecipe
 - → micorformat: hCard, hResume, hRecipe
- Browsers display HTML, machines process data
- Microformat vs. POSH format
 - → POSH is same as microformat but data is not a standard format

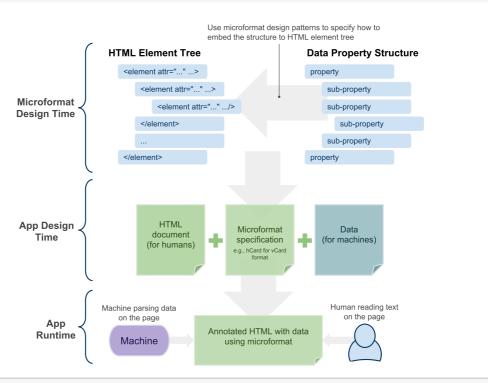
Difference to Atom feeds

- Microformats require only a single HTML document
 - → clients run GET to retrieve all data (human readable and machine readable)
- No significant increase of the size of document
- No requirements on data representation
 - → can be in any representation
 - → should be defined in a well-established format spec
 - \rightarrow a microformat spec needs to be defined for every data

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

- 5 -

Microformats Usage



Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

- 6 -

Principles

- Design Patterns
 - How to embed data in HTML elements or elements' attributes
 - Applied for a particular microformat specification
- Follow semantics of (X)HTML elements
 - - \rightarrow if not available, use $\langle span \rangle$ or $\langle div \rangle$
- XHTML Metadata Profiles (XMDP)
 - Definition of metadata of a microformat in (X)HTML page
 - Machine and human readable, not a Web standard
 - Uses profile attribute on <head> element
 - *Is deprecated in HTML5*
 - Is an analogy to a namespace but not really a namespace!

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

vCard Example

Describes contact information

```
PROPERTY:value1;value2;...;valueN
PROPERTY:SUBPROPERTY1="value";...SUBPROPERTY2="value";...
  3
        BEGIN: VCARD
        VERSION:4.0
       N:Vitvar;Tomas;Ing.;Doc.;PhD
FN: Doc. Ing. Tomas Vitvar, Ph.D.
        ORG:Czech Technical University in Prague
         TITLE:Associate Professor
       PHOTO: http://vitvar.com/img/tomvit-portrait.jpg
TEL; TYPE="work, voice"; VALUE=uri:tel:+420-2-334-334
TEL; TYPE="home, voice"; VALUE=uri:tel:+420-2-443-554
ADR; TYPE=work; LABEL="Thákurova 6, Praha 6, Czech Republic"
 11
          :;;Thákurova 6;Praha 6;Czech Republic
 12
        EMAIL:tomas.vitvar@fit.cvut.cz
        END: VCARD
-N-a structured representation of the name (person/organization)
```

- FN formatted name string
- ORG name of the organization and associated units
- − TITLE − *job title, functional position*
- LABEL Addressing label

Lecture 10: Annotations, CTU Summer Semester 2017/2018. @TomasVitva

Design Patterns Rules

class-design-pattern

- semantic meaning indicated on HTML content by class attribute

value-class-pattern

- embedding data structure when a property has subproperties (vCard fragment is TEL;TYPE=WORK:+43 554 554 556)

 sometimes value needs to be split into multiple pieces as follows (note that dialing +430554554556 is not valid)

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

- 9 -

Design Patterns Rules (cont.)

• include-pattern

- to include a subset of data from one area of a page to the other area of the same page (same data to be reused by multiple microformats)
- cannot be used to include content from other URLs!
- Example, a verbose hCard on a page:

- Reviews on the same page:

(parser replaces the whole <a> element including its content)

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

hCard Microformat Example

hCard profile, options:

- Example specific rules
 - vCard properties that do not make sense for hCard
 - → e.g., NAME, PROFILE, SOURCE, PRODID, VERSION
 - → publishers should not use them, parses should ignore them
 - -if fn == org (i.e, class="fn org")
 - \rightarrow hCard is a contact for a company, organization or a place
 - → N (person's name) property should not be used or be the empty string
 - -if fn != org AND fn contains two words
 - → fn *is split into* given-name *and* last-name
 - \rightarrow sub-properties of N peoperty (by a whitespace or a comma)
 - see a complete specification in hCard Microformat Specification №

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

_ 11 _

Known Issues

- Name conflicts and scalability
 - More microformats on a page may cause naming conflicts
 - → no namespace support, microformats do not scale
 - → functionality of tools may break when data formats change
- No formal semantics, no reasoning support
 - How important is it?
 - Semantics defined in XMDP profiles
 - → no formal basis though machine processable
 - \rightarrow lack of compatibility with RDF/RDFa
 - → See Microformats and RDF/RDFa compatibility & for details.

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

– 12 -

Uptake and some statistics

- Two billion pages annotated with hCard
- Google Rich Snippets
 - Content indexing with microformats, microdata, RDFa
 - → see Google Rich Snippets 🗗
 - \rightarrow 94% of the rich snippets data uses microformats

Pizza Pizzas Recipe: Alton Brown: Food Network

www.foodnetwork.com → Recipes → Italian

★★★★ 229 reviews - 24 hrs 45 mins

Food Network invites you to try this Pizza Pizzas recipe from Alton Brown.

- Firefox 3
 - Native API to parse and process microformats in JavaScript
 - see Microformats support in Firefox 3 ₺
- Facebook
 - hCalendar and hCard for events
 - see Microformats in Facebook ₫

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

_ 13 -

Overview

- Microformats
- Microdata
- RDF and RDFa
- OpenGraph Protocol

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

– 14 -

Microdata

- Part of HTML5 specification
 - Google is the main driver (rich snippets support)
 - spec includes:
 - → Microdata vocabularies
 - → Microdata Global Attributes
 - see W3C working draft
- Idea similar to microformats, but
 - items (collection of properties) have ids (URIs)
 - Microdata vocabulary, a formal description of terms
 - → http://schema.org is becoming a standard
 - \rightarrow e.g., Event, Organization, Person, Product, Review
 - → Created and supported by Google, Microsoft, Yahoo!
 - \rightarrow have RDF representation too
 - data formats not directly based on formats such as vCard, vCalendar, they define its own "simple" vocabulary

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

- 15 -

Global Attributes

- Attributes on any HTML element
- Itemscope
 - identifies an element which descendants contain some properties
 - 1 | <div itemscope>...<div>
- Itemtype
 - pointer to a vocabulary that describes the item and its properties
 http://www.data-vocabulary.org/Person/
- Itemid
 - global identifier of the item (URI)
 - such as a book's ISBN in urn schema, urn:isbn:0-330-34032-8
- Itemprop
 - a term from the vocabulary which value is in the element's content
 - 1 | Johny
- Itemref
 - a reference to other item within the same document
 - 1 | <div itemscope itemref="myprofile"/>

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

Example

Non-annotated HTML text

Annotated HTML text with microdata

```
<p
```

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

_ 17 _

Microformats vs. Microdata

Scalability

- Microformats specs are complicated because of specific rules tailored for vCard, vResume, etc.
- Microdata can be easily extensible, when new property occur they can be added witout breaking conformance of tools

Standards-based

- Microdata is a standard part of HTML5 effort
- Microformats is an "ad-hoc" group of ethusiastic people, though widely supported
 - → Strength is in underlying well-established formats
- Microdata have links to Semantic Web efforts and Linked Data (via RDF), microformats not

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

Overview

- Microformats
- Microdata
- RDF and RDFa
 - Structured Property Values
 - Encoding RDF in XML (RDF/XML)
 - RDF-in attributes (RDFa)
- OpenGraph Protocol

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

_ 19 _

RDF

- Resource Description Framework (RDF)
 - Resource as defined in Web architecture
 - → usually anything that can be conveyed electronically
 - \rightarrow plus abstract concepts that have no representation
 - RDF is at the bottom of Semantic Web stack of languages
- References
 - W3C Recommendations:
 - \rightarrow RDF Suite of W3C Recommendations $\[\]$,
 - $\rightarrow RDF$ Primer

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

– 20 –

Meaning of Data in XML

• A resource with URI http://www.vitvar.com/data-about-me

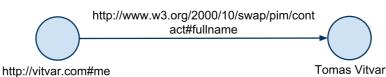
- No explicit meaning of terms
 - person, name, mailbox, ... are terms defined in namespace
 http://example.org/people but there is no URI assigned to them
 this does not work here: http://example.org/people#name
- No explicit meaning of relationships
 - a person <u>has</u> name <u>with value</u> Tomas Vitvar (→ Tomas Vitvar <u>is a person</u>), this person <u>has</u> mailbox <u>with value</u> tomas@vitvar.com (→ tomas@vitvar.com <u>is a mailbox</u>), etc.
 BUT this person lives?, works?, was born?, ... in a city Innsbruck
- Need for a language to describe statements
 - → Resource Description Framework

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

- 21 -

RDF Statement

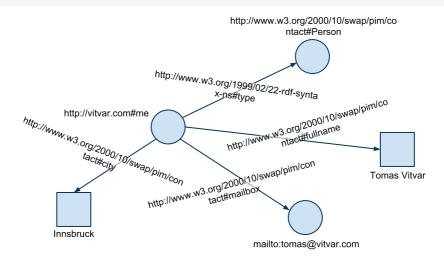
- RDF Tripple: **subject predicate object**
 - a thing the statement describes (subject)
 - a specific property of the object (predicate)
 - a value of the property (subject)



- Representation of statements
 - using a graph notation
 - → nodes are subject and objects (rectangles are literals)
 - \rightarrow arcs are predicates
 - identifiers to identify subject, predicate, object
 - → URI references (URIrefs)
 - machine processable language
 - \rightarrow RDF serializations in triples, RDF/XML, N3, Turtle notations

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

Meaning of Data in RDF



- individuals: Tomas Vitvar identified by http://vitvar.com#me
- kinds of things: Person identified by #Person
 - properties of those things, e.g., mailbox, identified by #mailbox
 - values of those properties, e.g. mailto:tomas@vitvar.com
 - + values of other data types such as strings, integers, dates, etc.

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

- 23 -

References in statements

- URI identifies
 - network-accessible things (electronic documents) $\rightarrow URL$
 - things that are not network-accessible, such as human beings
 - abstract concepts that do not physically exist, such as "fullname"
 - RDF uses <u>URI references</u> to identify subjects, predicates, objects
- URI references (or URIref in short)
 - URI with an optional fragment identifier
 - http://www.w3.org/2000/10/swap/pim/contact#fullname
 - RDF resource is anything that can be identified with URIref
 - a set of URIrefs is called a **RDF vocabulary**
- Literals
 - character strings to represent property values
 - can only be assigned to objects in RDF (in other words, objects can be either URIrefs or literals)
 - → they cannot be assigned to subjects or properties
 - two kinds: plain literals and typed literals

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

RDF Serializations – Triples Notation

- Triples notation
 - list of all triples from RDF graph
 - the full triples notation requires that URI references be written out completely (in angled brackets)
 - very long documents, some URIrefs need to be repeated
 - 1 | <http://www.example.org/index.html> <http://purl.org/dc/elements/1.1/creator>
- Simplicity for examples
 - QNames without angle brackets
 - Common prefixes and namespaces:

```
rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs: http://www.w3.org/2000/01/rdf-schema#
dc: http://purl.org/dc/elements/1.1/
ex: http://www.example.org/
ext: http://www.example.org/terms
xsd: http://www.w3.org/2001/XMLSchema#
```

- example

```
1 | ex:index.html dc:creator "Tomas" . ex:index.html dc:language "en" .
```

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

- 25 -

Kinds of Things

- Property rdf:type
 - defines a type of a resource
 - 1 | ex:me rdf:type ext:Person .
 - corresponds to "is a member of" relationship
 - ext:Person understood as a class
 - → however, RDF language does not define its semantics
 - → RDF Schema language provides additional vocabulary for class semantics

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

– 26 –

Overview

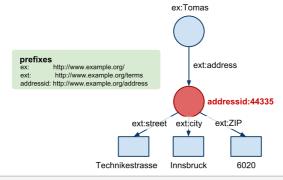
- Microformats
- Microdata
- RDF and RDFa
 - Structured Property Values
 - Encoding RDF in XML (RDF/XML)
 - RDF-in attributes (RDFa)
- OpenGraph Protocol

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

_ 27 -

Structured Property Values

- Consider real-world complex structures
 - Tomas works at Technikestrasse 21a, 6020 Innsbruck, Austria
 - One option to describe this using RDF:
 - ex:Tomas ext:works "Technikestrasse 21a, 6020 Innsbruck, Austria" .
 - But this is not often sufficient, such statements usually need to be recored as a structure, i.e. a street, a city, ZIP, ...
 - ightarrow describe Tomas's **address as a resource** that has a URIref

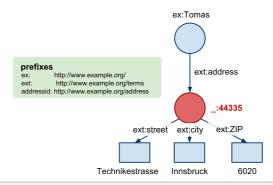


Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

– 28 –

Blank Nodes

- Does every structure need to have a URIref?
 - When referenced from outside of the graphs yes, otherwise not
- Blank nodes
 - Nodes that do not need to be referenced from outside of the graph
 - No need for URIref, they are only used within the graph
- Blank node identifier
 - local within a graph: _:LocalID, must be unique within the graph
 - two blank nodes in two graphs with the same IDs are not the same!



Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

- 29 -

Modeling with Blank Nodes

- N-ary relationships
 - In fact, a blank node is a way to model an n-ary relationships
 - A blank node breaks down an n-ary to binary relationships
 - 3-ary relationship between Tomas and (Technikestr, Innsbruck, 6020)
 Tomas Technikestr, Tomas Innsbruck, Tomas 6020
- Unidentified things
 - not always good to use URIs such as e-mails to identify people
 - \rightarrow e-mails may change, disappear, ...
 - → sometimes no need to assign unique ids to people
 - Example
 - → the author of the book is mailto:tomas@vitvar.com, as oposed to it is <u>a person</u> with e-mail mailto:tomas@vitvar.com
 - A person is an **abstract concept** that can be modeled using a blank node

```
ex:book23 ext:author _:author1 .
    _:author1 ext:email <mailto:tomas@vitvar.com .
    _:author1 ext:name "Tomas Vitvar" .
    _:author1 rdf:type ex:person .</pre>
```

Untyped and Typed Literals

- Untyped Literals
 - No information about how to interpret a value of the plain literal
 - a programme must have a knowledge how to interpret the value
 - 1 | ex:person1 ext:age "24" .
- Typed literals
 - pairing a string with a URIref that identifies a particular datatype (xsd: refers to http://www.w3.org/2001/XMLSchema#)
 - 1 | ex:person1 ext:age "24"^^xsd:integer
 - RDF does not define its own data types (except rdf:XMLLiteral)
 - → no need to map external to native ones
 - RDF uses external data types defined in XML Schema
 - → not all are suitable, only basic ones such as string, integer, date

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

_ 31 _

Overview

- Microformats
- Microdata
- RDF and RDFa
 - Structured Property Values
 - Encoding RDF in XML (RDF/XML)
 - RDF-in attributes (RDFa)
- OpenGraph Protocol

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

– 32 -

Basic Rules

- Representation of RDF in XML language
- Example RDF triple

```
- a page index.html was created on August 16, 1999
1 | ex:index.html ext:creation-date "Aug 16, 1999" .
```

- RDF/XML representation
 - We can interpret a RDF statement as:
 a description that is about a <u>subject</u> of the statement
 - XML element (QName) of the description is the **predicate**
 - a value of the element is the **object**

- URIrefs must be written out when in attribute values

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

- 33 -

Multiple Statements and Typed Literals

• Example RDF triples

```
ex:index.html ext:creation-date "Aug 16, 1999".

ex:index.html dc:language "en".

ex:index.html ext:rank "3"^^xsd:decimal.

ex:index.html dc:creator <http://www.vitvar.com#me>.
```

RDF/XML representation

```
<?xml version="1.0"?>
     <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:dc="http://purl.org/dc/elements/1.1/"</pre>
         xmlns:ext="http://www.example.org/terms/">
         <rdf:Description rdf:about="http://www.example.org/index.html">
              <ext:creation-date>August 16, 1999</ext:creation-date>
8
              <dc:language>en</dc:language>
9
                rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">
11
                    3</ext:rank>
12
              <dc:creator rdf:resource="http://www.vitvar.com#me"/>
13
         </rdf:Description>
     </rdf:RDF>
14
```

 a description may combine all properties for a single subject but there also can be a description for every subject (such representations are the same)

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

– 34 –

Blank Nodes

Example RDF triples

```
1    ex:index.html    ext:editor    _:editor332 .
2    _editor332    ext:name    "Tomas Vitvar" .
3    _editor332    ext:homepage    <a href="http://www.vitvar.com"></a>.
```

RDF/XML representation

```
<?xml version="1.0"?>
   <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
       xmlns:ext="http://www.example.org/terms/">
4
       </rdf:Description>
9
       <rdf:Description rdf:nodeId="editor332">
10
          <ext:name>Tomas Vitvar<ext:name>
11
          <ext:homepage rdf:resource="http://www.vitvar.com"/>
12
       </rdf:Description>
13
   </rdf:RDF>
14
```

- A node with id editor332 can be referenced from within the RDF graph, not outside of the RDF graph

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

- 35 -

Overview

- Microformats
- Microdata
- RDF and RDFa
 - Structured Property Values
 - Encoding RDF in XML (RDF/XML)
 - RDF-in attributes (RDFa)
- OpenGraph Protocol

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

RDFa

- Embedding RDF data in XHTML
 - XHTML only, is extensible, HTML not
 - → RDFa defines a number of extension attributes
 - Parses may recognize RDFa annotations in HTML too
 - RDFa is generic to embed arbitrary RDF data
 - → however, only standard (commonly agreed) vocabularies make sense
- W3C Recommendations:
 - RDFa Specification ₫
 - RDFa Primer 🗗

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

- 37 -

Property and Object Values as Resources

- Creating a property using rel attribute
 - assume, following text is at http://blog.vitvar.com/?p=107

```
Content on this page is licensed under

a xmlns:ext="http://www.example.org/terms"
rel="ext:license"
href="http://creativecommons.org/license/by/3.0">
a Creative Commons License - attribution
</a>
```

- This corresponds to the RDF triple

→ When the subject is not explicitly stated, then the subject is the URL of the XHTML page being described

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

– 38 –

Property and Object Values as Literals

- Creating a property using property attribute
 - RDFa defines a property extension attribute
 - assume, following text is at http://blog.vitvar.com/?p=107

- This corresponds to the RDF triple

```
1 | <http://blog.vitvar.com/?p=107> dc:creator "Tomas" .
```

- Typed literals
 - RDFa defines a datatype extension attribute

- Alternative content
 - RDFa defines content extension attribute
 - \rightarrow replaces the object value that is in the element's value

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

- 39 -

Subject

- Creating a subject using about attribute
 - RDFa defines about extension attribute
 - Let the following text is at http://blog.vitvar.com/?p=107

- This corresponds to the RDF triple
 - 1 | <http://blog.vitvar.com/p/107> dc:creator "Tomas".
- Also possible to use multiple subjects on a single page

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

Types and Blank Nodes

- Types
 - RDFa defines typeof extension attribute
 - → corresponds to rdf:type property
- Blank node
 - When annotation has typeof but not about
 - \rightarrow blank node, that is, a node without a subject

- I know Peter who has e-mail petr@novak.cz

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

_ 41 -

Overview

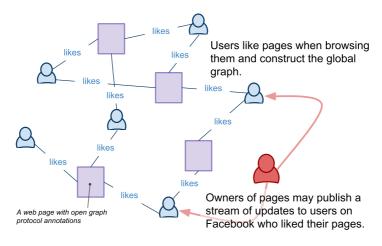
- Microformats
- Microdata
- RDF and RDFa
- OpenGraph Protocol

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

- 42 -

OpenGraph Protocol

- Global Social Graph
 - important adoption of RDFa, see Open Graph Protocol ₫
 - defines meta-data for pages' description so that it can be easily included in a global graph connecting people and pages through "likes" (a person likes a page)



Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

- 43 -

Page Annotations

- Open Graph protocol main properties
 - a page is the subject in the RDF triple
 - og:title − *title* of the page
 - og:type − type of the content (e.g., movie)
 - og:image URL of the image for the page
 - og:url a canonical URL of the page to be used as its permanent ID in the graph
- HTML page annotation RDFa example

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

Publishing updates

- Ownership
 - Page must be associated with a Facebook application
 - → using fb:app_id meta tag
 - Owners can publish a stream of updates using the Facebook Graph API ☑
- Getting access

```
1    curl -F type=client_cred \
2         -F client_id=app_id \
3         -F client_secret=app_secret \
4         https://graph.facebook.com/oauth/access_token
```

Publishing updates

```
1    curl -F 'access_token=...' \
2         -F 'message=Hello Likers' \
3         -F 'id=http://www.mydomain.com/great_page.html' \
4         https://graph.facebook.com/feed
```

Lecture 10: Annotations, CTU Summer Semester 2017/2018, @TomasVitvar

- 45 -