What is the Semantic Web? (In 15 minutes...)

ISOC Nieuwjaarsreceptie 2009

2009-01-15, Amsterdam, The Netherlands

Ivan Herman, W3C
Let’s organize a trip to Budapest using the Web!
You try to find a proper flight with …
... a big, reputable airline, or ...
... the airline of the target country, or ...
... or a low cost one
You have to find a hotel, so you look for...
... a really cheap accommodation, or ...

Find Hostels and Lodging at your destinations.

Look for iwg, for instant online booking.

3 accommodations have been found matching your criteria.
Select your accommodations and click for rates, availability, and reservations.
Click on names to see photos, reviews, and more information.
Tip: Click for rates and instant secure confirmations.

Sort by: 
- Price (Lo-Hi)
- Price (Hi-Lo)
- Traveler Rating
- Hostel Class
- Hostel Name

View Advanced Display Options

Balaton, Hungary

Unity Hostel Balaton
Rakoczi Ut 268
Hostel 8 Units
From €12

2 hours from Budapest, we are located right behind a free beach access to the lake, right opposite a large shopping and dining court, only few minutes from the best clubs lots of freebies!!!
... or a really luxurious one, or ...
... and intermediate one ...
oops, that is no good, the page is in Hungarian that almost nobody understands, but…
... this one could work
Of course, you could decide to trust a specialized site...
... like this one, or...
... or this one
You may want to know something about Budapest; look for some photographs...
... on flickr ...
... on Google ...
... or you can look at mine 😊
but you can also look at a (social) travel site
What happened here?

- You had to consult a large number of sites, all different in style, purpose, possibly language…
- You had to mentally *integrate* all those information to achieve your goals
- We all know that, sometimes, this is a long and tedious process!
All those pages are only tips of respective icebergs:

- the real *data* is hidden somewhere in databases, XML files, Excel sheets, ...
- you have only access to what the Web page designers allow you to see
Specialized sites (Expedia, TripAdvisor) do a bit more:

- they gather and combine data from other sources (usually with the approval of the data owners)
- but they still control how you see those sources

But sometimes you want to personalize: access the original data and combine it yourself!
Here is another example…
Another example: social sites. I have a list of "friends" by...
... Dopplr,
... Twine,
LinkedIn,
... and, of course, the ubiquitous Facebook
- I had to type in and connect with friends again and again for each site independently 😞
- This is even worse then before: I feed the icebergs, but I still do not have an easy access to data…
What would we like to have?

- Use the data on the Web the same way as we do with documents:
  - be able to link to data (independently of their presentation)
  - use that data the way I want (present it, mine it, etc)
  - agents, programs, scripts, etc, should be able to interpret part of that data
Put it another way...

- We would like to *extend* the current Web to a "**Web of data**":
  - allow for applications to exploit the data directly
But wait! Isn’t what mashup sites are already doing?
A “mashup” example:

Flight from Amsterdam (AMS) to Zurich (ZRH):
- Swiss International Airlines 729
  - Depart: Amsterdam (AMS), 14:55 CEST
  - Arrival: Zurich (ZRH), 16:20 CEST
- Aircraft: Airbus A320-100/200
- Distance: 374 miles
- Passenger(s): Ivan Herman, Eva Boka ep Herman

Flight from Zurich (ZRH) to Budapest (BUD):
- Swiss International Airlines 2258
  - Depart: Zurich (ZRH), 17:20 CEST
  - Arrival: Budapest (BUD), 18:55 CEST
- Aircraft: Fokker 100
- Distance: 500 miles

Map of Budapest, Hungary
In some ways, yes, and that shows the huge power of what such Web of data provides.

But mashup sites are forced to do very ad-hoc jobs:

- various data sources expose their data via Web Services
- each with a different API, a different logic, different structure
- these sites are forced to reinvent the wheel many times because there is no standard way of doing things 😞
Put it another way (again)...

- We would like to extend the current Web to a **standard** way for a “Web of data”
But what does this mean?

- What makes the current (document) Web work?
  - people create different documents
  - they give an address to it (ie, a URI) and make it accessible to others on the Web
An example: Steven’s site on Amsterdam (done for some visiting friends)

The Internet Guide to Amsterdam

Designed to be printed out and taken with you.
Written by Steven Pemberton, CWI, Amsterdam, and Astrid Kerssens, Amsterdam.
Linked to by more than 450 other sites; more than 3,500,000 grunted readers!
The top Amsterdam travel guide according to Google. If you know how Google works, you know that that says something about this site!
See also London
Then some magic happens...

- Others discover the site and they link to it.
- The more they link to it, the more important and well known the page becomes.
  - remember, this is what, eg, Google exploits!
- This is the “Network effect”: some pages become important, and others begin to rely on it even if the author did not expect it...
This could be expected...

WWW9 Organizers
May 15-19, 2000, Amsterdam

WWW9 Conference Committee

Conference Co-Chairs:
- Ivan Herman, CWI, The Netherlands
- Albert Vezza, CNRI, USA

Program Committee Chair:
- Dick Buterman, Oraxix, The Netherlands

Program Committee Vice Chairs:
- Ann Bassetti, Boeing, USA
- Stephan Fischer, Technical University, Darmstadt, Germany
- Lynda Hardman, CWI, The Netherlands
- Anja Jansen, IBM Research, USA
but this one, from the other side of the Globe, was not...

<table>
<thead>
<tr>
<th>Amsterdam Links</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring Break Trip participants to Amsterdam may familiarize themselves with Amsterdam by visiting the following web sites. On website address nl refers to the Netherlands.</td>
</tr>
<tr>
<td>Information on travel outside the United States, including instructions on getting a passport: <a href="http://travel.state.gov">http://travel.state.gov</a></td>
</tr>
<tr>
<td>A must see website: <a href="http://homepages.cwi.nl/~steven/amsterdam.html">http://homepages.cwi.nl/~steven/amsterdam.html</a></td>
</tr>
<tr>
<td>United States Consulate, Amsterdam: <a href="http://www.usemb.nl">http://www.usemb.nl</a></td>
</tr>
<tr>
<td>Virtual Tour of Amsterdam: <a href="http://www.channels.nl">http://www.channels.nl</a></td>
</tr>
</tbody>
</table>
What would that mean for a Web of Data?

- Lessons learned: we should be able to:
  - “publish” the data to make it known on the Web
    - standard ways should be used instead of ad-hoc approaches
    - the analogous approach to documents: *give URI-s to the data*
  - make it possible to “link” to that URI from *other* sources of data (not only Web pages)
    - ie, applications should not be forced to make targeted developments to access the data
    - generic, standard approaches should suffice
  - and let the network effect work its way…
Example: combine data from experiments

- A drug company has huge amount of old experimental data on its Intranet
- Data in different formats (XML, databases, …)
- To reuse them:
  - make the important facts available on the Web via standards
  - use off-the-shelf tool to integrate, display, search

Courtesy of Nigel Wilkinson, Lee Harland, Pfizer Ltd, Melliyal Annamalai, Oracle (SWEO Case Study)
But it is a little bit more complicated 😞

- On the traditional Web, humans are implicitly taken into account
- A Web link has a “context” that a person may use
Eg: address field on my page:

Ivan Herman

My Work at W3C

I am Semantic Web Activity Lead, that is my main work at W3C. I am member of IW3C2 (International World Wide Web Conference Committee) (the committee coordinating the yearly WWW conference series), serving as a liaison for W3C, and of SWSA (Semantic Web Science Association), the committee responsible for the International Semantic Web Conferences series.

As part of my work, I also participate in lots of outreach activities, and I regularly make presentations, tutorials, etc. You can consult my list of presentations for further details.

Contact information

Email: ivan@w3.org
(sha1sum: 5ac8032d5f0012aa1775ea2103e1076baf5e80b)

Postal address:
C/o Centre for Mathematics and Computer Sciences (CWI)
Kruislaan 413, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands.

Phone numbers:
phone: +31-20-5924163
mobile phone: +31-541044153
fax: +31-20-5924312

PGP/GPG:
My GnuPG key and signature is available on-line.

FOAF:
You can either extract a short FOAF information from this page, or consult my more complete, public FOAF file.

Misc:
... leading to this page
- A human understands that this is my institution’s home page
- He/she knows what it means (realizes that it is a research institute in Amsterdam)
- On a Web of Data, something is missing; machines can’t make sense of the link alone
New lesson learned:

- extra information ("label") must be added to a (data) link: "this links to my institution, which is a research institute"
- this information should be machine readable
- this is a characterization (or "classification") of both the link and its target
- in some cases, the classification should allow for some limited "reasoning"
Let us put it together

What we need for a Web of Data:

- use URI-s to publish data, not only full documents
- allow the data to link to other data
- characterize/classify the data and the links (the "terms") to convey some extra meaning
- and use standards for all these!
So What *is* the Semantic Web?
It is a collection of standard technologies to realize a Web of Data
- It is that simple…
- Of course, the devil is in the details
  - a common model has to be provided for machines to describe, query, etc, the data and their connections
  - the “classification” of the terms can become very complex for specific knowledge areas: this is where ontologies, thesauri, etc, enter the game…
  - but these details are fleshed out by experts as we speak!
In the end...

- More and more data should be "published" on the Web
  - this can lead to the "network effect" on data
- New breeds of applications come to the fore
  - "mashups on steroids"
  - better representation and usage of community knowledge
  - new customization possibilities
  - ...
Thank you for your attention!

These slides are also available on the Web:

http://www.w3.org/2009/Talks/0115-Amsterdam-IH/