

Introduction and Applications of the Semantic Web

Slides by: Ivan Herman, W3C



http://www.w3.org/2009/Talks/05-Oz-IntroSW-IH/ http://www.w3.org/2009/Talks/0615-SanJose-tutorial-IH/



Let's organize a trip to Budapest from Amsterdam 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

Soline booking Select Flights File Edit View Higtory Bookmarks	- Mozilla Firefox Tools <u>H</u> elp M http://wizzair.com/skyli	ghts/cgi-bin/sł	kylights.cgi? N	step=1 ew destina	tion from	Londo	☆ • ▼• tra	ivel tom 🔎 🚇	
WZ	WIZZAIT.COM	O useful inform	T	imisoar Ødest	a (from 1 ⁻	7 Decen	nber) avel services	• partners	0
 flights agency login my account search bookings log in Book a Flight Round Trip One Way From: Endhoven 	 search The flights availated that you wish to search for new for search for new for the taxes and the taxes and the click here to find payment. going out > 	• select able for th purchase dights. All ow are for the charges d out the e	e date(s by tickin times are one way . Paymen exact am	contact) that young the do e local. y flights a nts made ount. The	pas u have set t next to and per ad with debits a fee depe	sengers lected a the far lult, chi t and c ends on	purchase are shown belo e price or use Id and infant. redit cards are the type of c	O confirmat ow. Review and s the form to the The total price in subject to a pa ard that you wis	ion select left ha nclude: nyment h to u
To: Budapest-Terminal 1	date	fareclass	flight	departs	arrives	pric	e excluding tax	taxes and cha	rges
Depart:	Fri 15 Aug 08	Web	W6 228	13:25	15:20	0	Adult 94.99 EUR	26.00 EU	
14 🔽 Aug 🔽 🎬									R
14 Aug V III Return: 03 V Sep V III Passengers:	Sun 17 Aug 08	Web	W6 228	13:25	15:20	0	Adult 73.99 EUR	26.00 EU	R

You have to find a hotel, so you look for...



7

... a really cheap accommodation, or ...



... or a really luxurious one, or ...

Higtory Bookmarks Tools He	lp	
C 🙆 🕥 🥘 http://w	ww.hilton.com/en/hi/hotels/search/newresults.jhtml?searchType=city8	8stat 🏠 🔹 🏆 🗤 uth hostel 🔎 🚇 🔹 🕲
ilton Si d take you places*	Username or HHonors #: Password or PIN: gn in	A prouv A prouv View Text Only Customer Support 1-800-HILTONS
Packages Reservations	Meetings Social Gatherings Hilton HHonors	Things to Do Brochures My Favo
Search Results	Hilton Hotels	
Change Your Search	Search Results	🕒 Print 🕐 Help
Location City: Budapest Search Within: 40 💟 💿 mi 🔘 km	The following locations matched your request. Rates displayed may be non-refundable. Please review rate rules and restrictions prior to booking Sort by: Brand Show: All Hotels	your stay. ✓ Go► View Hotels on a Map >
State/Province: State / Province Country:		Select up to
Hungary V	Hilton Hotels	Compare S
Hilton Hotels	(1) Hilton Budapest WestEnd	Compare Hotel
All Hilton Family Hotels	Vaci ut 1-3 Budapest, Hungary, 1062	Available
	History Bookmarks Tools He C Image: Search Results Change Your Search Location City: Budapest Search Within: 40 Image: Imag	Higtory Bookmarks Tools Help Image: Sign in Image: Sign in Sign in Image: Sign in <

... an intermediate one ...



oops, that is no good, the page is in Hungarian that almost nobody understands, but...



... this one could work

Bestwestern.com, the World's	Largest Hotel Chain - Mozilla Firefox	- • 🗙
e Edit View History Bookmarks		
🕑 - 🗙 🏠 C 🔌 🏠 🤇	💼 http://book.bestwestern.com/bestwestern/selectHotel.do?iata=00158210 🛛 🏠 🔹 🛂 st western 🔎 🚇	• 🎱 🖻 •
Western We	Customer Service 7 Rewards Program 7 Gift Card 7 Groups & Meetings 7 My Email or Member#: SIGN-IN Forgot Password come to bestwestern.com® Password: SIGN-IN Forgot Password HOME FIND A HOTEL CHECK RESERVATIONS TRIP PLANNER PROMOTIONS & PROGRAMS PACKAGES	
RESERVATIONS HOTE	LOCATIONS BEST WESTERN PREMIER HOTELS NEW HOTELS QUESTIONS & ANSWERS	
Hotel Search Resul	ts - Select Occupancy Select Room Review & Reserve Confirmation	
Find a Hotel	- Select Hotel Page: 1	
City: Budapest Select State or Province:	5 Hotels Found within 50 Mi / 80 Km Image: Solution of the Buddapest Area Show: All Hotels Image: Solution of the Buddapest Area Show: All Hotels Image: Solution of the Buddapest Area Display Currency In: Image: Solution of the Buddapest Area	
Select Country:	Best Western Hotel Hungaria Hotel not	
Check-In: 08/15/2008 Check-Out:	Stay at this 4-star Budapest hotel offering guests deluxe accommodations near some of Budapest's popular attractions and business locations. Visitors More > Check Alternate Dates	
09/03/2008	Photo Gallery	
Features & Amenities:	Rákoczi Ut 90, Budapest, H-1074, Hungary Distance from City Center: 0.86m / 1.38km	
Complimentary Breakfast		
Prostfact Available	Hotel not	
	W:	3C 🌍 Ser

Of course, you could decide to trust a specialized site...



... like this one, or...

le Edit View History	y Bookmarks Tools Help
Expedi	Image: Search Expedia Image: Search Expedia Image: Search Expedia Image: Search Expedia
Home Flig	hts Hotels Cars Vacation Packages Cruises Activities DEALS & OFFERS Maps Business Travel Rewards
art search over Change your search eparting: /18/2008 eturning: /3/2008 tar Rating: Show all odging Type: Show All ame contains:	Create your package from Amsterdam to Budapest (and vicinity) Image: Area map view Hotel map view Image: Area map view Hotel map view Image: Area map view Hotel map view Image: Area map view Hotel map view Show hotels in this area: Image: Area map view (and vicinity) (all areas) Image: Budapest (and vicinity) (all areas) Image: Colored a different destination Mote what you're looking for? Choose a different destination Image: Colored a different destination View packages: 1 - 25 26 - 50 51 - 75 76 - 97 Previous Next Sort by: Image: Colored a Picks Image: Colored a Col
Go Change Travelers Adults Room hange travelers	Image: Second

... or this one

Budapest Hotels: Read Buda	apest Hotel Reviews and Compare Prices - TripAdvisor - Mozilla Firefox	
<u>File E</u> dit <u>V</u> iew Hi <u>s</u> tory <u>B</u> ookm	arks <u>T</u> ools <u>H</u> elp	24 a 2 a
< D- X 🏠 C 🌣 🕻	http://www.tripadvisor.com/Hotels-g274887-Budapest-Hotels.html	😭 🔹 🕎 🔹 est tourist 🔎 🚇 🔹 🖑
		🙎 My TripAdvisor
	ISOT [®] truth, then go.*	>> Register Now!
22,582,888 Travelers from 190 Co	ountries Planned Trips Here This Week!	Budaj
Home 🔽 Destinations 🔽	Fun & Games 🔰 🤜 Just For Members	🔽 Inte
Search	Home→ Europe→ Hungary→ Budapest	
Go	Budapest Hotels	
Budapest Overview		ADD TO MY MAP VVRITE A REVIEW EMA
Budapest Hotels	Hotels (272) B&Bs / Inns (24) Specialty Lodging (83)	Part de als: Prodes art hatala
Flights to Budapest		Best deals: Budapest notels
Budapest Deals	Find Hotels Travelers Trust	Budapest: Boek en bespaar tot 75%. Booking.com Geen reserveringskosten!
More On Budapest	Any date My dates are flexible	Great Budapest Hotels www.Venere.com/Budapest_Hotels See maps & pics, reac book online. Relax and enjoy your stay!
Before You Go	Price: Hotel class: Adults:	Budapest easyHotel Deals
Things to Do	Any Price 💟 U.S. Dollars 💟 Any Class 💟 2 💟	www.easyHotel.com City Centre from just €15 per night fron founder of easyJet
Restaurants	Recommended For:	Cheap Hotels Budapest
Мар		www.otel.com/BudapestHotels Fantastic rates on Budape Huge Savings, Instant Confirmation!
Traveler Photos	Find Hotels	View all deals for Budapest
Traveler Videos		Spon
Forum	Recommended Hotels (1-20 of 272)	Free Budapest Guide
Budapest Deals	Sort by: Popularity	Get the best picks for where to ea
Discount Hotels		and play.
		The Download odf
		Download odf

You may want to know something about Budapest; look for some photographs...



... on flickr ...



... on Google ...

🕲 budapest - Google Image	Search - Mozilla Firefo	х			- 2
	okmarks <u>loois H</u> elp	oogle.pl/images?bl=ep&q=b	udapest&btnG=Search+Imac	ies&aby=: 🔨 🔹	est tourist 🔎 🔊 🔹 🔹
Web Images Maps New	vs Video Gmail <u>mor</u>	<u>e</u> v			Sign in
Google M	udapest oderate SafeSearch is on	Search Images	Search the Web Advanced Image Preferences	Search	
Images Showing: All image sizes			Results 1 - 18	of about 19,900,000 for buda	pest [definition]. (0.25 seconds)
E-mail this photo E-mail. Budapest 550 x 412 - 40k - jpg www.tripadvisor.com	Budapest, Hungary 450 x 338 - 43k - jpg ww.transitionsabroad.com	Budapest looks its most beautiful at 1024 x 768 - 161k - jpg web.kvif.bgf.hu	Beautiful-Budapest 430 x 320 - 34k - jpg www.budapesthotels.com [More from www.budapesthotels.com]	Hotel Victoria Budapest 575 x 473 - 92k - jpg www.victoria.hu	Fly to Budapest and experienc one 909 x 682 - 347k - jpg www.sterling.dk
Danube Bridge Elisabeth in Budapest 1024 x 768 - 194k - jpg budapest5.freeblog.hu	Audapest had 2421831 inhabitants in 422 x 425 - 29k - jpg www.squidoo.com	budapest night 575 x 352 - 208k - jpg www.wayfaring.info	Budapest - Things to Do with a Day 400 x 300 - 32k - jpg cruises.about.com	Jewish Cultural Heritage in Budapest 452 x 360 - 67k www.budapesthotels.com	Hungary, Budapest, Parliamen 640 x 480 - 199k - jpg www.hungary.travelphotoguide.cx
<			F	Call Control Control (Control (Contro) (Control (Contro) (Control (Contro) (Contro)	

Semantic Web

W3C[®]

... or you can look at mine 😔





...or at a (social) travel site

Budapest Travel Guide Budapest Tourism - RealTravel - Mozilla Firefox Elle Edit View Higtory Bookmarks Tools Help	
Image: A trace of the second seco	g Free Trip Planner Sign
BUDAPEST TRAVEL GUIDE AND TOURISM Introduction Guides Trips Photos Flights Hotels Restaurants Attractions Deals	
For by TaistealMore Budapest photos > Destinations near Budapest photos >	Price Compare Tool Search multiple sites for the best rate in Budapest COMPARE FLIGHTS COMPARE HOTELS
Travel Guide Information From Our Partners	

Semantic Web

W3C

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!
- The value is in the *combination* of the data



Here is another example...



Another example: social sites. I have a list of "friends" by...



... Dopplr,

PLR: Ivan Herman's fe	Ilow travellers - Mozilla Firefox	- 🗆 🞽
it <u>V</u> iew Hi <u>s</u> tory <u>B</u> ook	kmarks <u>T</u> ools <u>H</u> elp	
• 🗙 🏡 C 🔗 🔇	🔛 🚺 http://www.dopplr.com/traveller/IvanHerman/fellows 🏹 🔂 🖓 🔹 W 🔹 easy jet 🔎 🚇 🔹	🎱 🖻 •
DC	PPLR (Interpretation) FOR IVAN Find and Invite Your connections Your trips Your account Sign out Visit our blog for updates	
	Where Next? { Gent, Boston, Vienna	E
Type the name of a city	by or a traveller Find people and places	
0	Welcome, Ivan	
Streem UN	one of your fellow travellers added a trip that coincides with you. Find out more in your journal	
Contraction of the	You are at home in Amsterdam.	
	You are at home in Amsterdam. You can invite people to DoppIr to see your trips, find them on other networks you use or look for travellers you already know to encourage more coincidences. You have a public profile. Edit it?	
	one of your fellow travellers added a trip that coincides with you. Find out more in your journal You are at home in Amsterdam. You can invite people to Dopplr to see your trips, find them on other networks you use or look for travellers you already know to encourage more coincidences. You have a public profile. Edit it? You can now create a public profile to display to the whole internet if you want, not just Dopplr users — and take any of the information to embed on your own website. Give it a try! X	
Your trips Fello	one of your fellow travellers added a trip that coincides with you. Find out more in your journal You are at home in Amsterdam. You can invite people to Dopplr to see your trips, find them on other networks you use or look for travellers you already know to encourage more coincidences. You have a public profile. Edit it? You can now create a public profile to display to the whole internet if you want, not just Dopplr users — and take any of the information to embed on your own website. Give it a try! Image: Constant of the information in the infore	
Your trips Fello	one of your fellow travellers added a trip that coincides with you. Find out more in your journal You are at home in Amsterdam. You can invite people to Dopplr to see your trips, find them on other networks you use or look for travellers you already know to encourage more coincidences. You have a public profile. Edit it? You can now create a public profile to display to the whole internet if you want, not just Dopplr users — and take any of the information to embed on your own website. Give it a try! Image: Comparison of the information of the info	
Your trips Fello	one of your fellow travellers added a trip that coincides with you. Find out more in your journal You are at home in Amsterdam. You can invite people to Dopplr to see your trips, find them on other networks you use or look for travellers you already know to encourage more coincidences. You have a public profile. Edit it? You can now create a public profile to display to the whole internet if you want, not just Dopplr users — and take any of the information to embed on your own website. Give it a try! Image: Comparison of the information to embed on your own website. ow travellers Your journal Your carbon List or Map s in Montréal until August 16th. Boston soon. Montréal later.	
Your trips Fello	one of your fellow travellers added a trip that coincides with you. Find out more in your journal You are at home in Amsterdam. You can invite people to Dopplr to see your trips, find them on other networks you use or look for travellers you already know to encourage more coincidences. You have a public profile. Edit it? You can now create a public profile to display to the whole internet if you want, not just Dopplr users — and take any of the information to embed on your own website. Give it a try! Image: Comparison of the information to embed on your own website. ow travellers Your journal Your carbon List or Map s in Montréal until August 16th. Boston soon. Montréal later. uist is in Washington until August 23rd. Aspen soon. Washington later.	
Your trips Fello Peter Brown is Daniel Appelq Eva Méndez is	one of your fellow travellers added a trip that coincides with you. Find out more in your journal You are at home in Amsterdam. You can invite people to Dopplr to see your trips, find them on other networks you use or look for travellers you already know to encourage more coincidences. You have a public profile. Edit it? You can now create a public profile to display to the whole internet if you want, not just Dopplr users — and take any of the information to embed on your own website. Give it a try! Image: Comparison of the information to embed on your own website. Give it a try! pw travellers Your journal Your carbon List or Map s in Montréal until August 16th. Boston soon. Montréal later. guist is in Washington until August 23rd. Aspen soon. Washington later. s in Maraña until August 17th. Santo Domingo later.	
Your trips Fello Peter Brown is Daniel Appelq Eva Méndez is Danny Weitzne	one of your fellow travellers added a trip that coincides with you. Find out more in your journal You are at home in Amsterdam. You can invite people to Dopplr to see your trips, find them on other networks you use or look for travellers you already know to encourage more coincidences. You have a public profile. Edit it? You can now create a public profile to display to the whole internet if you want, not just Dopplr users — and take any of the information to embed on your own website. Give it a try! ow travellers Your journal Your carbon List or Map s in Montréal until August 16th. Boston soon. Montréal later. quist is in Washington until August 23rd. Aspen soon. Washington later. a in Maraña until August 17th. Santo Domingo later. er is in Bergen until September 5th. Los Angeles later.	

... Twine,

Wy Connections Twine - Mozilla Firefox		
<u> Eile E</u> dit <u>V</u> iew Hi <u>s</u> tory <u>B</u> ookmarks <u>T</u> ools <u>H</u> elp		**s ***
💽 🗩 🗙 🏠 😋 🔌 🌄 😽 http://www.twine.com/u	user/ivanherman/connections 🔊 🖒 🔹 🛛	• easy jet 🔎 🐠 • 🥙 🛃 •
Home My Items V My	Twines 🔻 My Connections Explore 🔻 Start a Twine	Nanherman Logout Account Search T
Search within My Connections	View All Sc	ort by Most Recent
Refine your view by 2 You can also filter by selecting from the following categories. Image: Comparison of the following categories is a selecting from the following categories is a selecting f	Dean Allemang at home Oakland, CA 3 Twines 4 Items Send Message Disconnect	Connect with People Invite People to Tw Recommended M
	David Provost Breathing Boston 20 Twines 15 Items Send Message Disconnect	got my san fra 83 Twi Steve Seattle 38 Twi
78	Attila Gardos Hungary, Budapest 9 Twines 8 Items Send Message Disconnect	Jame: Iving la 95008 122 Tw IVING la 95008 122 Tw Chris All rea Mill Val 73 Twi
		W3C 🏹 Seman Web

... LinkedIn,

Linkedin: My Contact	is: Connections - Mozilla Firefox Reekmarks - Teels - Hele	
lie <u>c</u> ait <u>v</u> iew Hi <u>s</u> tory		1.1
	🕐 🎧 (🛅 http://www.linkedin.com/connections?trk=hb_side_cnts 🛛 🏹 🔹 🗰 🔹	🤊 🖻 ·
Linked in . People - Job	Account & Settings Help Sign Out Account & Settings Help Sign Out Account & Settings Help Sign Out Account & Settings Help Sign Out Search Search Search	
Home	Contacts	
👭 Groups 🕞	Connections Imported Contacts Network Statistics Add Connect	tions F
E Profile		
Contacts +	Connections	
Add Connections	Show contacts with new connections , advanced options Showing \$11 of \$11 connections. 15 outstanding sent invitations	
Semantic Web Activity Lead, World Wilds	A A Asman, Jans CEO at Franz Inc CEO at	
What are you working on?	E Abramatic, Jean-François Chief Product Officer at ILOG F Ita@lioq.fr View & edit detalls »	
Your profile is 80% complete [Edit]	H Adida, Ben Software Security Researcher and Entrepreneur I ben@adida.net J View & edit detalls >	
	K Allemang, Dean Chief Scientist at TopQuadrant Inc. L dino-link@nappydino.com 179 M View & edit detalls >	
	N Allison, Kevin San Francsico Correspondent at The Financial Times 152 P View & edit detalls > 0	
	R Alonso, Jose Manuel eGovernment Lead at W3C jalonso@w3.org View & edit details » T	
	W	30 🌍 Semar Web

... and, of course, Facebook

Facebook All Friends -	- Mozilla Firefox		
	the second	N 🗘 🔹 🔟 easy iet 🔎 👰 • 🦓	- -
elcome to the new Faceboo	ok	Send feedback Back to the old Faceboo	k 🔼
acebook Ivan Herm	nan Friends Applications Inbox	Home Settings Logout Search	
All Friends			
Friend Lists	Showing: Status Updates Recently Updated Phonebook Everyone	Q Search Friends Adverti	se
👥 All Friends	Show: Choose an option	Make	mo
CWI	You have 137 friends.	1 2 3 Next Discove world n	r Ali °1 o
✓ Family ✓ IW3C2	Shadi Abou-Zahra	market; □× ⊔	lace
W3C Team			-
Make a New List	Ross Ackland	□ × Do Yo Photo Make yu book, fr	uT s? our c
Find Friends		សាទ	
Find people you know who are already on Facebook	Ben Adida	More A	ds
Invite your friends to join Facebook	Anupriya Ankolekar	⊟ ×	
	Daniel Appelquist is hanging out in Aspen. [™] 15 hours ago	□ ×	
	Lora Aroyo	= x	

- 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



31

Put it another way...

- We would like to *extend* the current Web to a "<u>Web</u> of data":
 - allow for applications to exploit the data directly



But wait! Isn't what mashup sites are already doing?



A "mashup" example:





- 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 <u>standard</u> way for a "Web of data"



36
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



Steven's site on Amsterdam (done for some visiting friends)



Amsterdam is an unusual city in that it has all the advantages of a hig city - culture, history, food, entertainment, good



Semantic

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 <u>even if the</u> <u>author did not expect it...</u>



This could be expected...

🚯 WWW9 Organizers - O	pera						
File Edit View Bookmarks	Widgets Feeds Chat Tools Help						
🖗 QuickP 🕼 Diigolet 🎦 Py	t ▼ 🗋 Boo ▼ 🗋 W3 ▼ 🗋 Sem ▼ (🗅 RDF 🔻 🅜 Validate! 🕜 Mobical 🎧 Favikis 🝷 🎦 Twi 💌 🛨 Netvibes 🔗 World					
	ganizers 💴 🗋						
≪ ∢ ► № Ű ft	/ http://www9.org/w9-organizers.html	🔽 🔀 ido-masochism website 🔽 🗄					
	W	WW9 Organizers					
	Μ	lay 15-19, 2000, Amsterdam					
WWW9 HOME PROGRAM INFO	wv	VW9 Conference Committee					
SPONSORING	Conference Co-Chairs:						
EXHIBITING VOLUNTEERS	 Ivan Herman, CWI, The Netherlands Albert Vezza, CNRI, USA 						
ORGANIZERS	Program Committee Chair:						
AMSTERDAM	Dick Bulterman, Oratrix, The Netherlands						
	Program Committee Vice Chairs	:					
CONTACT US							
	Practice and Experience	Ann Bassetti, Boeing, USA					
	Content and Coding	Stephan Fischer, Technical University, Darmstadt, Germany					
	Hypermedia	Lynda Hardman, CWI, The Netherlands					
	Desfermente	·⊡· ▼ Q 120% ▼					



but this one, from the other side of the Globe, was not...



41

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

Internal Compound Repurposing Example

Welcome, Allergy & Respiratory Team Member

noingu(259406400)



this tool allows you to identify opportunities for additional uses of compounds from other teams within your project. It combines internal data, public data and the results or same memory experiments to provide testable hypotheses.

					Cont	trol Panel & Item Filtering								
A	rea	512 Apr	roach		312 1	Term+Reason	12	Max_Stag	e_Reached	110	Litera	ature	Links	
29 Pain 16 Metabolic Disease 5 Cancer 3 Sexual Health 2 Infectives		8	Antbody			ACTIVE		51 Cand	idate	Z.	80) - 50	1	
		ease 🛛	isease Z Recombinant B SM_Agonist th Z 12 SM_Antagonist Z 21 SM_Inhibitor Z			12 BIOMARKER 51 EFFICACY 11 MARKET		10 Discovery 41 Exploratory 19 HTS 11 Phase I 13 Phase III		00000				
		D 18			R									
		h RĽ			Ø									
		2 21			2	REORG								
	1 Urogenitals					10 TOXIC								
	, ar oger mans							41 Scree	nina	0.				
									- · ·	-				
		Liver chicken	SM Intellige	2007-02 8	FEICACY	Candidate		P. Person						SW-030
Sexual Sexual Sexual	Brectile Dysfunction	Liver gycogen phosphorylaxe Integrin alpha-3 (Galactoprotein 83(MLA-3) (CD49C) Leukotriene C4 synthau	SM_Antagone SM_Antagone	2007-Q2 E 1 2006-Q3 E 2006-Q3 E	FFICACY FFICACY	Canddate Canddate Canddate Canddate		P. Person P. Person M.		1	1	1		SW-030 SW-025 SW-025
Sexual Kealth Sexual Kealth Sexual Kealth	Erectile Dysfunction Erectile Dysfunction Erectile Dysfunction	Liver groupen phosphorylike shegmi alpha-3 (Galactoprotein 63(MLA-3) (CD49C) Leukohene C4 syntaxi transcripton elongaton factor A (SII)-Ba 4	SM_Antagonis SM_Antagonis SM_Agonist SM_2nhbitor	2007-Q2 E 1 2006-Q3 E 2006-Q3 E 2005-Q2 E	BERICACY BERICACY BERICACY	Canddate Canddate Canddate Canddate Canddate Canddate		P. Person P. Person M. Manager P. Person		1	1	1		SW-029 SW-029 SW-029 SW-029
inoano ienual kaith ienual kaith ienual kaith	Enectile Dysfunction Enectile Dysfunction Enectile Dysfunction HEV	Liver syscopen phosphorylaxe Singen alpha-3 (Galactoprotein BS/(VLA-3) (CD-9C) Laukothene CA synthau transcription elongation factor 4 (SII)-like 4 Putation face repart on duranel (b)	SH_Antagona SH_Antagona SH_Agonat SH_Yrhibitor SH_Yrhibitor	2007-Q2 E 2006-Q3 E 2006-Q3 E 2005-Q2 E 2006-Q2 E	BERICACY BERICACY BERICACY BERICACY	Candidate Candidate Candidate Candidate Candidate Candidate		P. Person P. Person M. Manager P. Person L. Leader		1	1	1		SW-030 SW-025 SW-025 SW-025 SW-025
incaso incual keath incual keath incual math refectives refectives	Erectile Dysfunction Erectile Dysfunction Erectile Dysfunction HEV HEV	Lver syscopen phosphorykae Smigrin Jpha-3 (Giacdonotein BS/(AA-3) (CD-9C) Leakohime C4 synthau transcription elingation factor A (SII)-lea 4 Putate Bour repait on channel (h) Voltage-gatet potaeiux channel protentiv1,2 (h)	SH_YHERD SH_Artigone SH_Agonet SH_YhErbor SH_YhErbor SH_Agonet	2007-92 (2006-93 (2006-93 (2005-92 (2006-92 (2007-91 (PFRCACY PFRCACY PFRCACY PFRCACY PFRCACY	Candidate Candidate Candidate Candidate Candidate Candidate Candidate Candidate Candidate		P. Person P. Person M. Manager P. Person L. Leader A. Scentst		1	1	1	1	SW-020 SW-025 SW-025 SW-025 SW-025
incusio Secusi S	Erectile Dysfunction Erectile Dysfunction Erectile Dysfunction HEV HEV Incontinence	Liver gycogen phosphorylaw phosphorylaw (Galactoprotein BB/(MA-8) (CHAC) Laukothime C4 synthiai transcription elongation Bactor A (SII)-like 4 R-daton four repeat on dearner (D) Voltage-gated potentiar dharner (D) Voltage-gated potentiar dharner (D) (D) Haman RNA binding motif (CBM) game, partial cds.	SH_Antegone SH_Ardigone SH_Agonet SH_Arhibitor SH_Arhibitor SH_Agonet	2007-q2 (12006-q3 (2006-q3 (2005-q2 (2005-q2 (2007-q3 (2007-q3 (PFICACY PFICACY PFICACY PFICACY PFICACY PFICACY	Candidate		P. Person P. Person M. Manager P. Person L. Leader A. Scientist L. Leader		1	1	1	1	SW-025 SW-025 SW-025 SW-025 SW-025 SW-025 SW-025

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:





... leading to this page



Centrum Wiskunde & Informatica

Centrum Wiskunde & Informatica (CWI) performs fundamental scientific research in mathematics and computer science. CWI transfers the acquired knowledge to society and industry. The institute's strategy for the period up to 2012 is to concentrate research on four broad, societally relevant themes:



News

13-10-08

PhD defence Jarek Byrka



22-09-08

Best Paper Award for SMIL State research

At the ACM DocEngineering Symposium in Sao Paulo, Brazil, from 16 till 19 September, CWI researchers Jack Jansen and Dick Bulterman received the Best Paper Award.

read more

Agenda

18-10-08

Science Day at the Science Park Amsterdam

At Science Day the Science Park Amsterdam will be open for This year the theme of Science Day is 'Crack the code'. CWI workshops en demonstrations within this theme for every age

MAS Seminar, speaker Svetlana Dubinkina

Two speaker session

Tea starting at 10.00

Room: M279

Speakers:

1. Svetlana Dubinkina, CWI MAS 1, tba

2. James Glazier, University of Indiana Bloomington, tba

read more

28-10-08

MAS Seminar, speaker Peter Sonneveld



- A human understands that this is an institution's home page
- He/she knows what it means (realizes that it is a research institute in the Netherlands)
- 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 link: "this links to an 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"
 - eg, if an address refers to Amsterdam, then this means it is also in the Netherlands



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
 - technologies should be around to "export" the data
 - the "classification" of the terms can become <u>very</u> 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!











Example: find the right experts at NASA

- NASA has nearly 70,000 civil servants over the whole of the US
- Their expertise is described in 6-7 databases, geographically distributed, with different data formats, access types...
- Task: find the right expert for a specific task within NASA!



53

Example: find the right experts at NASA

 Approach: integrate all the data with standard means, and describe the data and links using generic (and simple) vocabularies

	dvanced Help		
IASA Center (15) 🛛 🐗 📑	🕨 🔿 🗶 Project (176) 🛛 🗢 🖶 🗶 C	Competency (21) 🛛 💠 🜩 🗶 I	People (1)
0	A 19		2
RC	A Mars Global Surveyor	strobiology	eanne M
FRC	Mars Odyssev 2001	stronomy and Astrophysics	
RC	Mars R&A CI	limate Change and Variability	
SFC	Mars Reconnaissance Orbiter 2005 (Ea	arth Atmosphere	
0	Mars Reconnaissance Orbiter 2003 (arth Science Applications Research	
NV NV	Minor Povital	arth System Modeling	
1	Mission Operations	uid Physics	
	Mission Operations	indamental Physics	
	Mission Science Guest Investigator	eonhysical/Geologic Science	
RC .	Mission Success – Center Specific	eosnatial Science and Technologier	
AE	Multi-Mission Operations	ing Physics	
	NMP Program Management and Futur V	ing Physics	
SEC	L NPOESS Prenaratory Project (NPP)	aser Technology	
Source: x500	Source: WIMS	Source: CMS	So
formation ranei			
View Different Social Netv POP Skill: Earth Project: C	Vork's Present in the Data Jeanne M h Sciences Competency Suite enter Investment Accounts	Same Skill and Sa Same Skill and Sa Same Skill, Projec Am I Connected?	me Department me Project t, and Facility
View Different Social Netv POP Skill: Earth Project: C	Average of the second s	el Grove @nasa.gov Clark and Parsia	me Department me Project t, and Facility





54

Wait! Does it mean that I have to convert all my data in some way?



- Not necessarily; this would not always be feasible
- There are technologies to make your data accessible to standard means without converting it
 - run-time "bridges" (eg, rewriting queries on the fly)
 - annotate existing data (eg, XHTML pages)
 - extract data from XHTML/XML files
 - etc
- Some of these techniques are still being developed









Example: "Linking Open Data Project"

- Goal: "expose" open datasets for integration
- Set links among the data items from different datasets
- Set up query endpoints
- Altogether billions of relationships, millions of links...





57

Example data source: DBpedia

- DBpedia is a community effort to
 - extract structured ("infobox") information from Wikipedia
 - provide a query endpoint to the dataset
 - interlink the DBpedia dataset with other datasets on the Web





The LOD "cloud", March 2008





The LOD "cloud", September 2008





The LOD "cloud", March 2009





All this sounds nice, but isn't that just a dream?



The 2007 Gartner predictions

During the next 10 years, Web-based technologies will improve the ability to embed semantic structures [... it] will occur in multiple evolutionary steps...

By 2017, we expect the vision of the Semantic Web [...] to coalesce [...] and the majority of Web pages are decorated with some form of semantic hypertext.

By 2012, 80% of public Web sites will use some level of semantic hypertext to create SW documents [...] 15% of public Web sites will use more extensive Semantic Web-based ontologies to create semantic databases

(note: "semantic hypertext" refers to pages "prepared" for integration)

"Finding and Exploiting Value in Semantic Web Technologies on the Web", Gartner Research Report, May 2007 WSC 🏹 💱

The "corporate" landscape is moving

- Major companies offer (or will offer) Semantic Web tools or systems using Semantic Web: Adobe, Oracle, IBM, HP, Software AG, GE, Northrop Gruman, Altova, Microsoft, Dow Jones, ...
- Others are using it (or consider using it) as part of their own operations: Novartis, Pfizer, Telefónica, ...
- Some of the names of active participants in W3C SW related groups: ILOG, HP, Agfa, SRI International, Fair Isaac Corp., Oracle, Boeing, IBM, Chevron, Siemens, Nokia, Pfizer, Sun, Eli Lilly, ...



64

Lots of Tools (not an exhaustive list!)

Categories:

- Triple Stores
- Inference engines
- Converters
- Search engines
- Middleware
- · CMS

. . .

- Semantic Web browsers
- Development environments
- Semantic Wikis

- Some names:
 - Jena, AllegroGraph, Mulgara, Sesame, flickurl, …
 - TopBraid Suite, Virtuoso environment, Falcon, Drupal 7, Redland, Pellet, ...
 - Disco, Oracle 11g, RacerPro, IODT, Ontobroker, OWLIM, Tallis Platform, ...
 - RDF Gateway, RDFLib, Open Anzo, DartGrid, Zitgist, Ontotext, Protégé, ...
 - Thetus publisher, SemanticWorks, SWI-Prolog, RDFStore...



65



Some deployment communities

- Major communities pick the technology up: digital libraries, defence, eGovernment, energy sector, financial services, health care, oil and gas industry, life sciences ...
 - Health care and life science sector is now very active
 - also at W3C, in the form of an Interest Group



Application specific portions of the cloud

- Eg, "bio" related datasets
 - done, partially, by the "Linking Open Drug Data" task force of the HCLS IG at W3C





Help in choosing the right drug regimen

- Help in finding the best drug regimen for a specific case, per patient
- Integrate data from various sources (patients, physicians, Pharma, researchers, ontologies, etc)
- Data (eg, regulation, drugs) change often, but the tool is much more resistant against change

Optimized Regimens Current Regimen Survey S	ummary Survey Set-up			
PharmaSURVEY	Safety Optimized Profiles			
Survey 1	Version 2	Current Regimen	Six	
Optimized Differences Ranges Common	Severe All Custom	-	1 ►	2 >
⇔Severity ⇔Adverse Drug Effec	t ⇔v2	≑ ₹	\$	\$
ADE Moderate Muscle Weakness (Myasthenia)	✓			
ADE Minor Excessive Sweating (Diaphoresis)	1			
ADE Moderate Heart Throbbing or Pounding (Palpitations)) 🗸			
ADE Moderate Hives (Urticaria)	✓			
ADE Major Bladder Inflammation (Cystitis)	~			
ADE Major Urinary Tract Infection	✓			
				row(s) 1 - 6 of 6
Export to Excel				



Courtesy of Erick Von Schweber, PharmaSURVEYOR Inc., (SWEO Use Case)

Yahoo's SearchMonkey

- Search based results may be customized via small applications
- Metadata embedded in pages are reused
- Publishers can export extra data via other formats





emantic

Information in Web Pages: SlideShare





Information in Web Pages: SlideShare

```
http://www.w3.org/2007/... ×
← → C 📅 🏡 http://www.w3.org/2007/08/pvRdfa/extract?format=turtle&uri=http://www.slideshare.net/ivan_herman/what-is-the-ser 🕨 📭 🔎 🗸
@prefix dc: <http://purl.org/dc/terms/> .
@prefix hx: <http://purl.org/NET/hinclude> .
@prefix media: <http://search.vahoo.com/searchmonkev/media/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xhv: <http://www.w3.org/1999/xhtml/vocab#> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
<http://www.slideshare.net/ivan herman/what-is-the-semantic-web-in-15-minutes-presentation>
dc:creator "Ivan Herman"@en ;
     dc:description "Very short introduction to what the Semantic Web is, given at an ISOC/W3C-
Benelux joint event in Amsterdam, January 2009"@en ;
    media:height "355"@en ;
     media:presentation <http://static.slidesharecdn.com/swf/ssplayer2.swf?doc=whatistheswamsterdam-
1232029522157652-2&stripped title=what-is-the-semantic-web-in-15-minutes-presentation> ;
     media:thumbnail <http://cdn.slidesharecdn.com/whatistheswamsterdam-1232029522157652-2-
thumbnail?1233240197> ;
     media:title "What is the Semantic Web (in 15 minutes...)"Gen ;
     media:width "425"@en ;
     xhv:alternate <http://www.slideshare.net/rss/latest> ;
     xhv:icon <http://www.slideshare.net/favicon.ico> ;
     xhv:stylesheet <http://public.slidesharecdn.com/v3/styles/slideview.css?1238118720> .
```



Improved Search (GoPubMed)

- Search results are re-ranked using ontologies
- Related terms are highlighted, usable for further search




Improved Search (Go3R)

- Same dataset, different ontology
 - (ontology is on non-animal experimentation)





New type of Web 2.0 applications

- New Web 2.0 applications come every day
- Some begin to look at Semantic Web as possible technology to improve their operation
 - more structured tagging, making use of external services
 - providing extra information to users
 - etc.
- Some examples: Twine, Revyu, Faviki, ...



Integration of "social" software data

- Internal usage of wikis, blogs, RSS, etc, at EDF
- · goal is to manage the flow of information better
- Items are integrated via
- · Semantic Web based unifying format
- · simple, public vocabularies
- · internal data is combined with linked open data like Geonames
- · Semantic Web queries are is used for internally
- Details are hidden from end users (via plugins, extra layers, etc)

Courtesy of A. Passant, EDF R&D and LaLIC, Université Paris-Sorbonne, (SWEO Case Study)



Integration of "social" software data



Courtesy of A. Passant, EDF R&D and LaLIC, Université Paris-Sorbonne, (SWEO Case Study)



Integration of "social" software data



Courtesy of A. Passant, EDF R&D and LaLIC, Université Paris-Sorbonne, (SWEO Case Study)



Conclusions...

- More an 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"

Semantic

Web

- better representation and usage of community knowledge
- new customization possibilities



The rough structure of data integration

- 1. Map the various data onto an abstract data representation
 - make the data independent of its internal representation...
- 2. Merge the resulting representations
- 3. Start making queries on the whole!
 - queries that could not have been done on the individual data sets



79

A <u>simplified</u> bookstore data (dataset "A")

ID	Author	Title	Publisher	Year
ISBN0-00-651409-X	id_xyz	The Glass Palace	id_qpr	2000

ID	Name	Home Page
id_xyz	Ghosh, Amitav	http://www.amitavghosh.com

ID	Publ. Name	City
id_qpr	Harper Collins	London



1st: export your data as a set of <u>relations</u>





Some notes on the exporting the data

- Relations form a graph
 - the nodes refer to the "real" data or contain some literal
 - how the graph is represented in machine is immaterial for now
- Data export does <u>not</u> necessarily mean physical conversion of the data
 - relations can be generated on-the-fly at query time
 - via SQL "bridges"
 - scraping HTML pages
 - extracting data from Excel sheets
 - etc.
- One can export <u>part</u> of the data



Another bookstore data (dataset "F")

		. / / / /		
	AA	В	D	E
1	ID	Titre	Traducteur	Original
2	ISBN0 2020386682	Le Palais des miroirs	A13	ISBN-0-00-651409-X
3				
6	ID	Auteur		
7	ISBN-0-00-651409-X	A12		
11	Nom			
12	Ghosh, Amitav			
13	Besse, Christianne			



2nd: export your second set of data





3rd: start merging your data





3rd: start merging your data (cont.)





3rd: merge identical resources





Start making queries...

- User of data "F" can now ask queries like:
 - "give me the title of the original"
 - well, ... « donnes-moi le titre de l'original »
- This information is not in the dataset "F"...
- ...but can be retrieved by merging with dataset "A"!



However, more can be achieved...

- We "feel" that a:author and f:auteur should be the same
- But an automatic merge doest not know that!
- Let us add some extra information to the merged data:
 - a:author same as f:auteur
 - both identify a "Person"
 - a term that a community may have already defined:
 - a "Person" is uniquely identified by his/her name and, say, homepage
 - it can be used as a "category" for certain type of resources



3rd revisited: use the extra knowledge





Start making richer queries!

- User of dataset "F" can now query:
 - "donnes-moi la page d'accueil de l'auteur de l'originale"
 - well... "give me the home page of the original's 'auteur'"
- The information is not in datasets "F" or "A"...
- ...but was made available by:
 - merging datasets "A" and datasets "F"
 - adding three simple extra statements as an extra "glue"





Combine with different datasets

- Using, e.g., the "Person", the dataset can be combined with other sources
- For example, data in Wikipedia can be extracted using dedicated tools
 - e.g., the "dbpedia" project can extract the "infobox" information from Wikipedia already...



Merge with Wikipedia data





Merge with Wikipedia data





Merge with Wikipedia data





Is that surprising?

- It may look like it but, in fact, it should not be...
- What happened via automatic means is done every day by Web users!
- The difference: a bit of extra rigour so that machines could do this, too



96

What did we do?

- We combined different datasets that
 - · are somewhere on the web
 - are of different formats (mysql, excel sheet, XHTML, etc)
 - have different names for relations
- We could combine the data because some URI-s were identical (the ISBN-s in this case)
- We could add some simple additional information (the "glue"), possibly using common terminologies that a community has produced
- As a result, new relations could be found and retrieved



It could become even more powerful

- We could add extra knowledge to the merged datasets
 - · e.g., a full classification of various types of library data
 - · geographical information
 - etc.

•

- · This is where *ontologies*, extra *rules*, etc, come in
 - ontologies/rule sets can be relatively simple and small, or huge, or anything in between...
 - Even more powerful queries can be asked as a result



What did we do? (cont)



Data in various formats

