BASI DI DATI II – 2 modulo Parte X: Introduzione al Web Semantico

Based on a tutorial of Ivan Herman, W3C

Prof. Riccardo Torlone Università Roma Tre

Outline

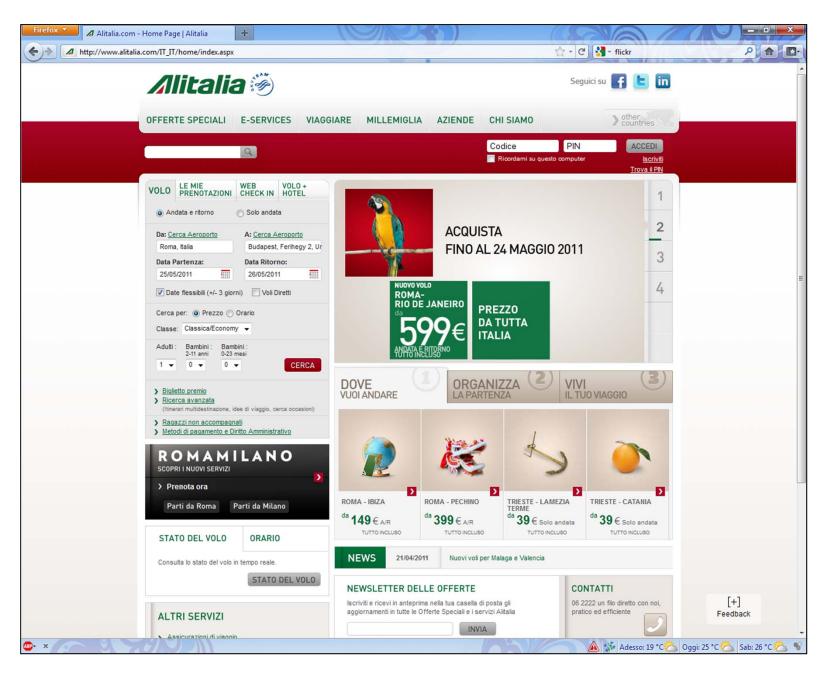
- Introduction
- The Web of Data
- Semantic Web
- Technologies



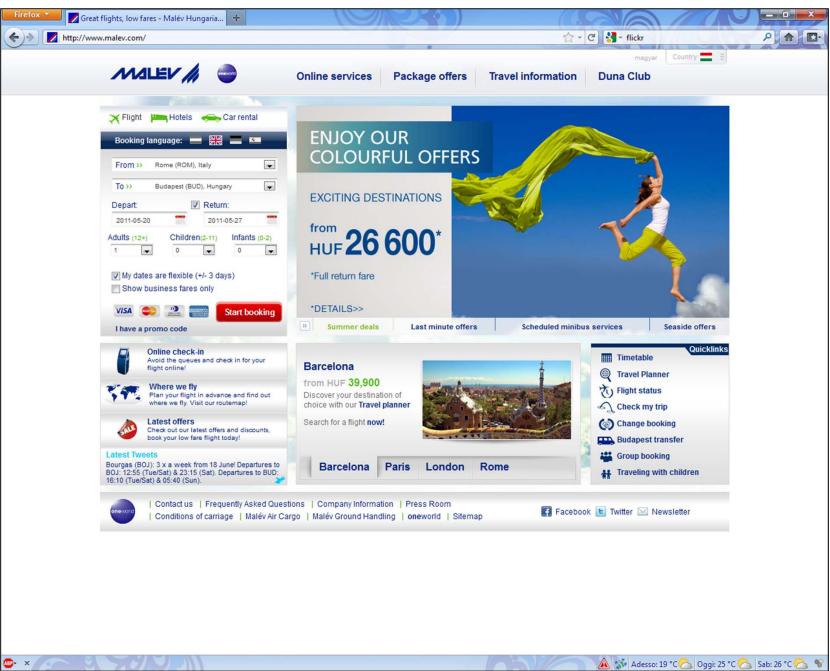
Let's organize a trip to Budapest using the Web!

First: 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



Using the Web

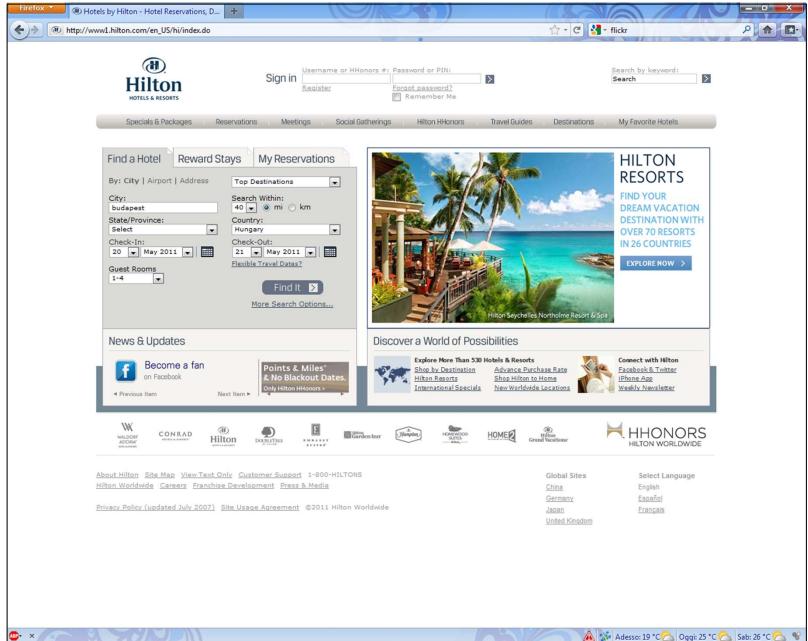
Let's organize a trip to Budapest using the Web!

First: you try to find a proper flight with ...
Second: you have to find a hotel, so you look for...

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

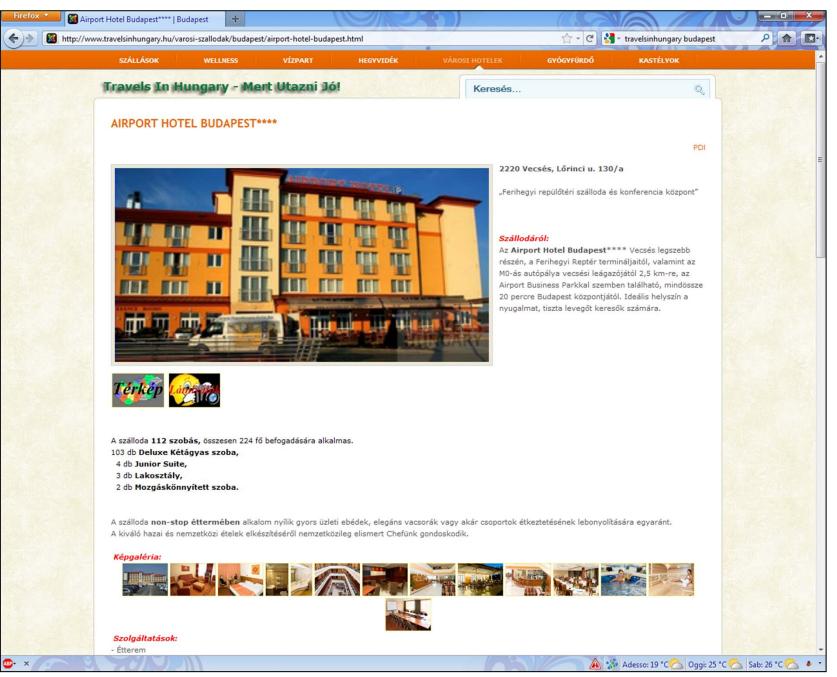


... or a really luxurious one, or ...



🛕 🐝 Adesso: 19 °C 📥 Oggi: 25 °C 📥 Sab: 26 °C 🙈 🖋

... an intermediate one ...

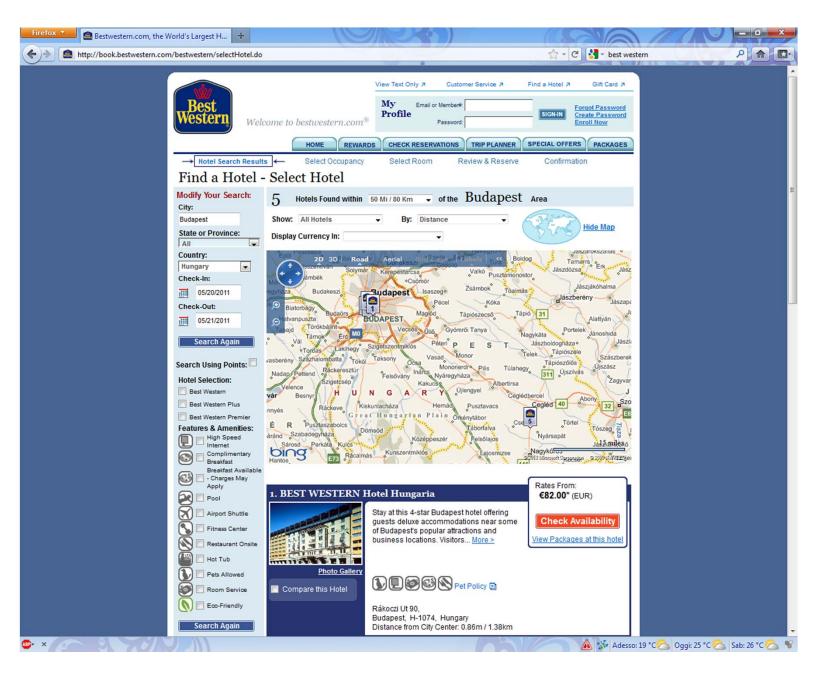


Using the Web

Let's organize a trip to Budapest using the Web!

- First: you try to find a proper flight with ...
- Second: you have to find a hotel, so you look for...
 - □oops, that is no good, the page is in Hungarian that almost nobody understands, but...

... this one could work



Using the Web

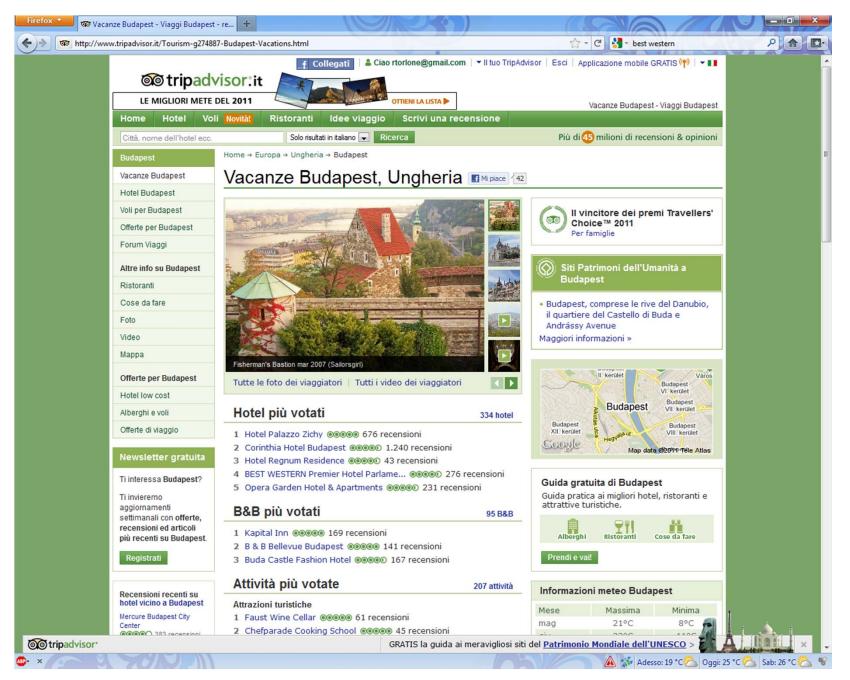
Let's organize a trip to Budapest using the Web!

- First: you try to find a proper flight with ...
- Second: you have to find a hotel, so you look for...
 - □ Of course, you could decide to trust a specialized site...

... like this one, or...

		=3a85c996715b9971e6d9c5a23d20c6fa;checkin_monthday=20				
BOC	DKING.COM		taliano		•	
P	renotazioni alberghiere online	Benvenuto/a R	Riccardo Torlone (Non s	sei Riccardo?) Il Mio Profilo L	a Mia Prenotazione	
		Sundiadou i				
home >	ungheria > margitsziget island					
Contraction of the second seco	943 hotel 334 hotel	334 hotel				
Cerca hotel	334 Hotel tr	rovati a Budapest, 212 Disponibili, Ris	sultati 1 – 20	Œ	Visualizza mappa	
	Ordina per:	Più Quotati 🝸 Disponibilità Prenotati Di Recente St	telle Posizione Tar	riffa Valutazione Clienti		
)ove?						
Budapest	4	Danubius Hotel Gellért ★★★☆ 🏈			Buono, 7.8	
)istanza: 1 km 🔻 🛛		Szent Gellért tér 1, <u>11. Újbuda</u> , Budapest • Visualizz Costruito in uno splendido Liberty e inaugurato nel 19	ra mappa		nuto da 600 recensioni	
		lo stesso edificio del famoso centro benessere Gelléri	t, accessibile gratuitar			
ata check-in		persona che sta visualizzando questo hotel. Maggio	ri Informazioni			
Ven 20 👻 Maggio '11 👻		Ultima prenotazione: 1 ora fa 🗉				
ata check-out		Tipologie Disponibili	Persone	Disponibilità	Tariffa	
Sab 21 👻 Maggio '11 👻		Camera Matrimoniale/Doppia con Letti Singoli		3 camere rimaste	€ 1000	
Non he anoste desire la date del		Camera Singola con Doccia Camera Superior Matrimoniale/Doppia con Letti Singoli con		<u>3 camere rimaste</u> 3 camere rimaste	€ 500 € 1000	
Non ho ancora deciso le date del oggiorno	mio	Vista		<u>s camere rimaste</u>	€ 1000	
Viaggio in gruppo/con bambini		Camera Matrimoniale/Doppia con Letti Singoli e Vista	**	2 camere rimaste	€ 1000	
				Ea	voloso, 8.9	
Cer	ca	Eurostars Budapest Center	and Manuffrage		uto da 1116 recensioni	
		Kossuth Lajos U. 7-9, 05. Belváros - Lipótváros, Buda Inaugurato nel 2010, a soli 3 minuti a piedi dal Museo				
	Contract of Contra	Grande Sinagoga e dalla stazione della metropolitana				
Cerca per:		eleganti nel cuore di Ci sono 4 persone che stanno Ultima prenotazione: 1 ora fa e	visualizzando questo	notel. Maggiori Informazion	<u>11</u>	
Fascia Di Prezzo (a notte)		Onina prenorazione. 1 ora la 🖻				
€0-€49		Tipologie Disponibili	Persone	Disponibilità	Tariffa	
€ 50 - € 99		Offerta Speciale - Camera Matrimoniale/Doppia con Letti Singoli Tariffa migliore	**	Disponibile	€-174- € 89	
€ 100 - € 149		Offerta Speciale - Camera Matrimoniale/Doppia con Letti	*	Disponibile	€ 84	
		Singoli Tariffa migliore				
Da € 200		Camera Matrimoniale/Doppia con Letti Singoli Tariffa migliore	**	Disponibile	€ 299- € 99	
 Star Rating 		altre 5 tipologie di camera				
· · · · · · · · · · · · · · · · · · ·	3 alberghi					
1 stella	4 alberghi	Continental Hotel Zara *****		Fa	avoloso, 8.7	
□ 1 stella □ 2 stelle 1	2 alberahi	Dohány u. 42-44, <u>07. Erzsébetváros</u> , Budapest • <u>Visualizza mappa</u> Punteggio ottenuto da 1328 recensioni				
☐ 1 stella ☐ 2 stelle 1 ☐ 3 stelle 9	2 alberghi 3 alberghi		to not 2010	IU SLESSO		
1 stella 2 stelle 1 3 stelle 9 4 stelle 5	2 alberghi 3 alberghi 7 alberghi	Questo Hotel Zara superior a 4 stelle è stato inaugura		sone che		
1 stella 2 stelle 1 3 stelle 9 4 stelle 5 5 stelle 1 Nessuna 4	3 alberghi		ingaria. Ci sono 4 pers	sone che		
1 stella 2 stelle 3 stelle 9 4 stelle 5 stelle	3 alberghi 7 alberghi	Questo Hotel Zara superior a 4 stelle è stato inaugura edificio in stile Liberty che ospitava i famosi Bagni Hu	ingaria. Ci sono 4 pers	sone che		
1 stella 2 stelle 1 3 stelle 9 4 stelle 5 5 stelle 1 Nessuna 4	3 alberghi 7 alberghi	Questó Hotel Zara superior a 4 stelle è stato inaugura edificio in stile Liberty che ospitava i famosi Bagni Hu stanno visualizzando questo hotel. <u>Maggiori Informa</u> Ultima prenotazione: 48 minuti fa ©	ingaria. Ci sono 4 pers	one che Disponibilità	Tariffa	
1 stella 2 stelle 1 3 stelle 9 4 stelle 5 5 stelle 1 Nessuna 4 classificazione 4 Tipo Di Alloggio 4	3 alberghi 7 alberghi	Questo Hotel Zara superior a 4 stelle è stato inaugura edificio in stile Liberty che ospitava i famosi Bagni Hu stanno visualizzando questo hotel. <u>Maggiori Informa</u>	ingaria. Ci sono 4 pers azioni		Tariffa € 379- € 209	

... or this one

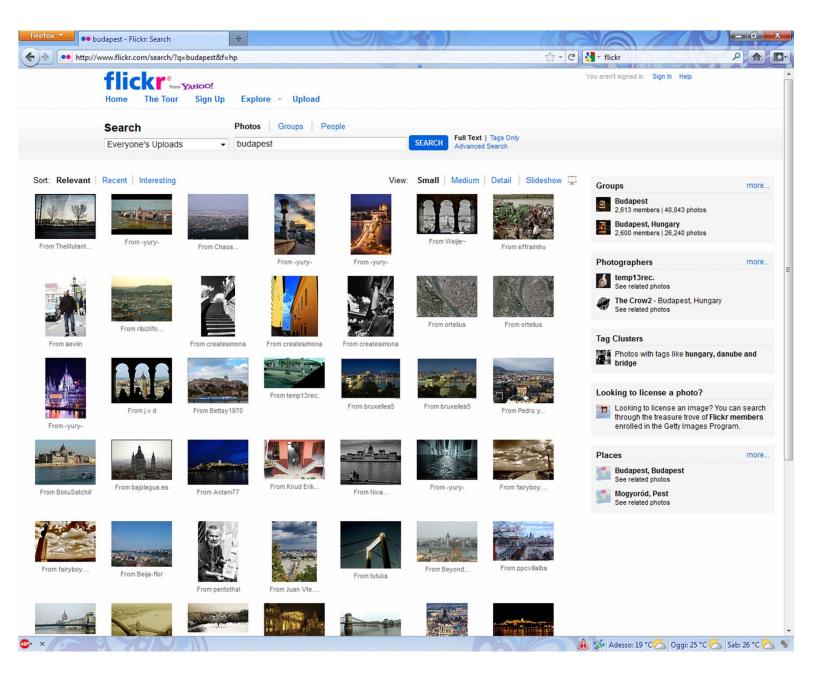


Using the Web

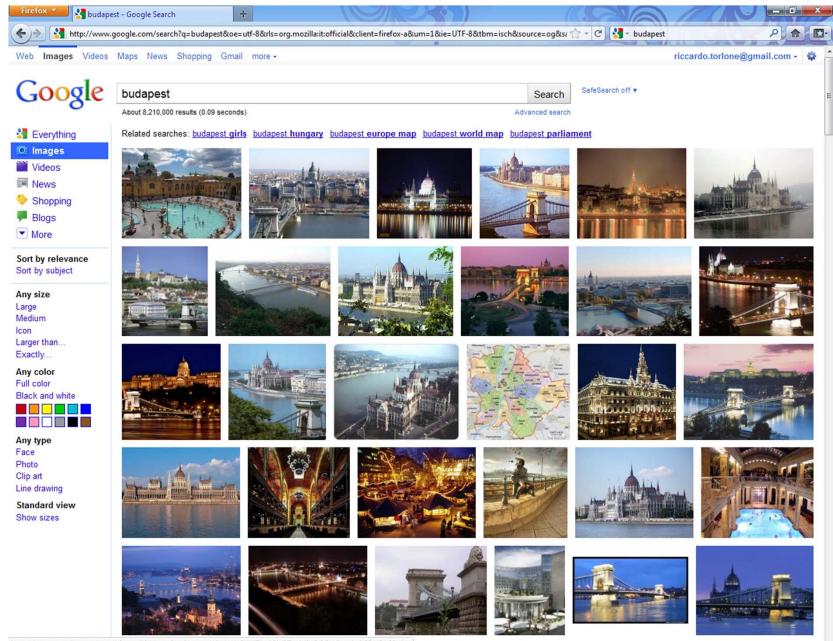
Let's organize a trip to Budapest using the Web!

- First: you try to find a proper flight with ...
- Second: you have to find a hotel, so you look for...
- Third: you may want to know something about Budapest; look for some photographs...

... on flickr ...



... on Google ...



🛕 🐝 Adesso: 19 °C 📥 Oggi: 25 °C 📥 Sab: 26 °C 📥 😵

http://www.google.com/search?um=1&hl=en&safe=...a=X&ei=WCXWTbniCM3Tsgbl8cDQBw&ved=0CDoQ1QIoAg

🚇 × /

... or a (social) travel site

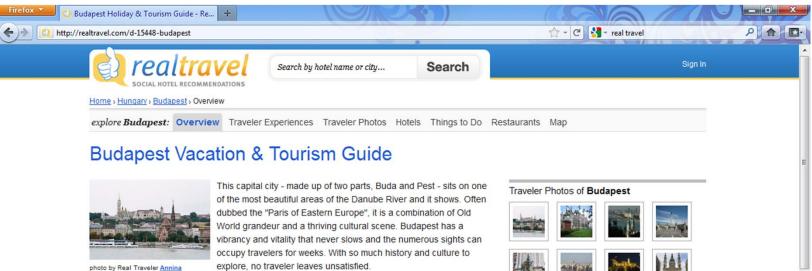


photo by Real Traveler Annina

🚇• × 🕜

Great Recommendations from Real Travelers

Top Hotels Top Things	to Do Top Restaurants		How can we serve you better?
Real Traveler Rating: 4.7 out of 5 bas			Share your feedback »
Check Rates > from \$78.79			Overview
Пата 😥 ні сн	ILY recommended Based o	n 28 reviews from 6 sources.	Traveler Experiences
Service:	Location: OOOOO Ro	om: 00000	Traveler Photos
Contract of Contract of Contract	eally great." — source Expedia		Hotels
"Since the h	otel is brand new the rooms were v	very clean."— source Orbitz	Things to Do
2 Normafa Hatal Budanast	2 La Maridian Dudanaat	A K and K Hatel Opera	Restaurants
2. Normafa Hotel Budapest	3. Le Meridien Budapest	4. K and K Hotel Opera	Map
Highly Recommended	Highly Recommended	Highly Recommended	
from \$50.24 check rates »	check rates »	check rates »	
5. Starlight Suiten Hote	6. Adina Apartment Hotel	7. Boutique Hotel Zara	

Highly Recommended	Highly Recommended	Highly Recommended	

xplore Budapest:
Verview
raveler Experiences
raveler Photos
lotels
hings to Do
estaurants
lap

🔁 See all photos in Budapest »

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!

Hidden Web

All those pages are only tips of respective icebergs:

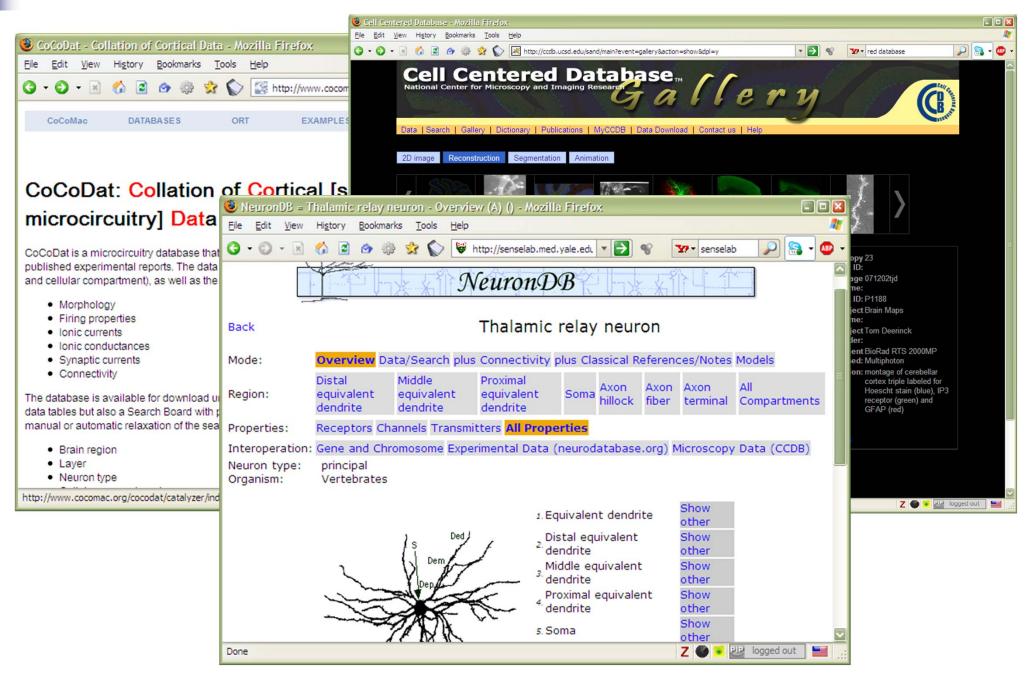
□ the real data is hidden somewhere in databases, XML files, Excel sheets, ...

□you have only access to what the Web page designers allow you to see

Some solution exists

- 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...

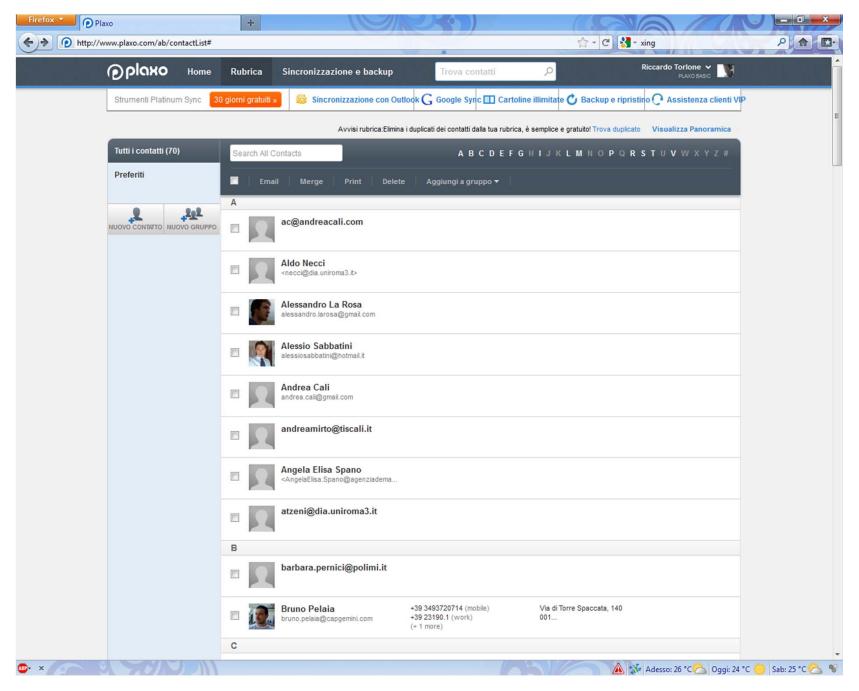




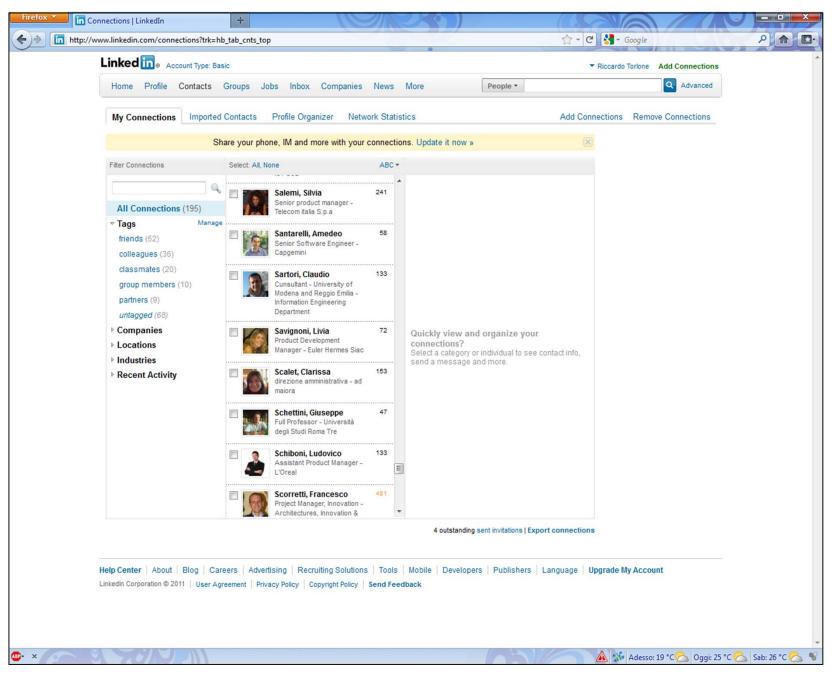
Social sites

I have a list of "friends" by...

... Plaxo,



... Linkedin,



... Twitter,

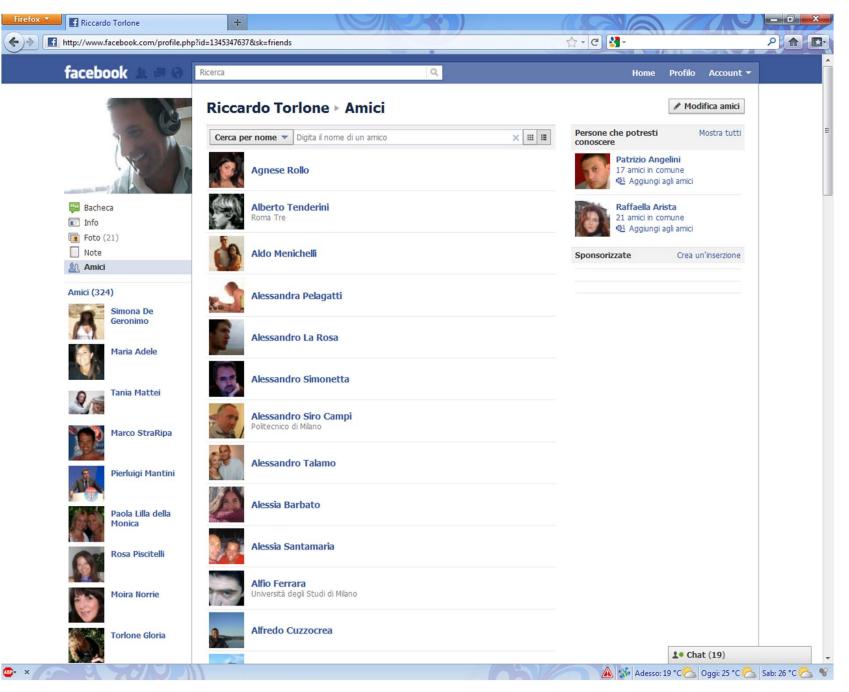
🐠 × 🕜



...Xing,

ps://www.xing.com/app/contact?op=toconfirm Finanza 🦳 Personali 🦳 RM3 🦳 Software ∛ iGoogle 🐇	🛔 Maps CD 🚺 DIA 🗋 Ingegneria 🔮 Mac	😴 Appartamenti 🛛 🐠 Ann	unci per Case V	. 👘 Mo	odle 📑 f	RUPAR 🗋 PRE OFF.	ති 🔂 😭
Sei un utente standard.		Logov	out Invita contatti	Aiuto e Cr	ontatto Ital	liano 🗸	
XING X Sei un uterite standard.	٩,	Ricca	Diventa subito Utent cardo Torlone			rmazioni	
/		Pro file	filo 💌 0 👗 4	/ Altro 🗸			
INIZIO RICERCA MESSAGGI CONTA	ATTI GRUPPI EVENTI LAVORO ORGA	ANIZZAZIONI 🕴 BEST OFFE	ERS				
l miei contatti							
Contatti confermati Contatti non confermati Ut	Itenti interessanti Contatti da confermare						
Quando respingi una richiesta di contatto, puoi s meno.	scegliere se desideri che la persona interessata venga	informata o			Contatti 1-4	-4 di 4	
Nome, società		Richiesta effettuata il	Invia messaggio	Conferma	a Respingi	ji Spam	
Dr. Massimiliano Pasquale Carpentieri aemmeCONSULTING		20/11/2010	\bowtie	1	×	₩	
Motivazione per l'aggiunta del contatto:	Gentile Riccardo,, vorrei aggiungerti ai miei c	ontatti.					
Raffaele Nicolussi Fondazione Ugo Bordoni		04/02/2010	\bowtie	-	×	*	
Motivazione per l'aggiunta del contatto:	Ciao! :)						
Alessandro Pierantoni D Alessandro Pierantoni		24/10/2008	\bowtie	-	×	*	
Motivazione per l'aggiunta del contatto:	benvenuto						
Michele Schioppa Michele Schioppa		24/10/2008	×	1	×	*	
Motivazione per l'aggiunta del contatto:	Salve, mi interesserebbe entrare a far parte dei suo A Presto Michele Schioppa	i contatti.					
Utenti che potresti conoscere							
gloria malvestuto università roma3	Enrico Valbonesi					\triangleright	
Espandi il tuo network							
Invita dei contatti a iscriversi a XING:		vari contatti della tua rubrica (di posta elettron	nica:			
Utilizza la virgola per separare gli indirizzi e-mail	Windows Live Go	pogle 🚱!					
Invia invito	> Trova contati						

... and, of course, Facebook



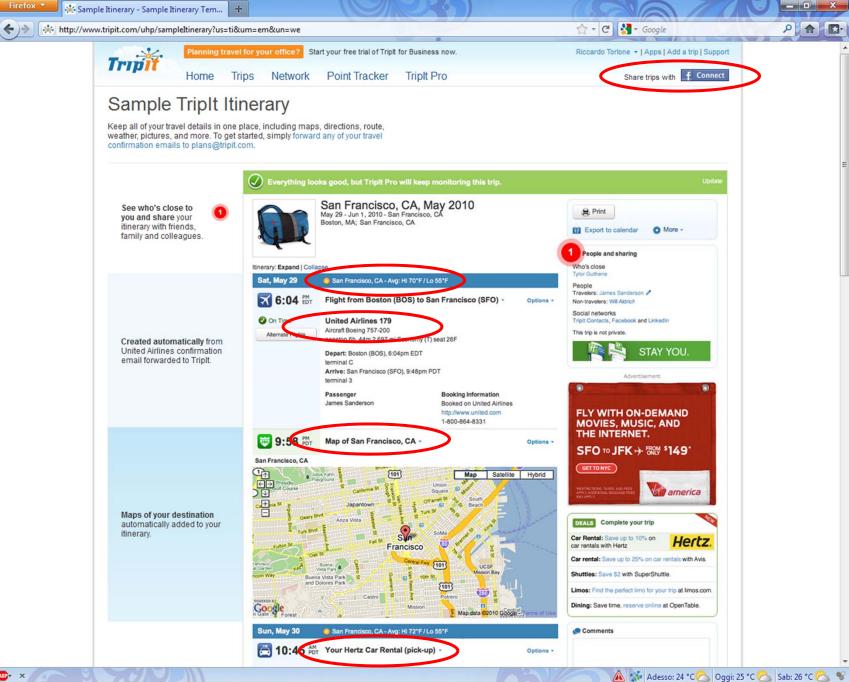
Problems here too

- 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?



Mashup solution

- 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

My home page



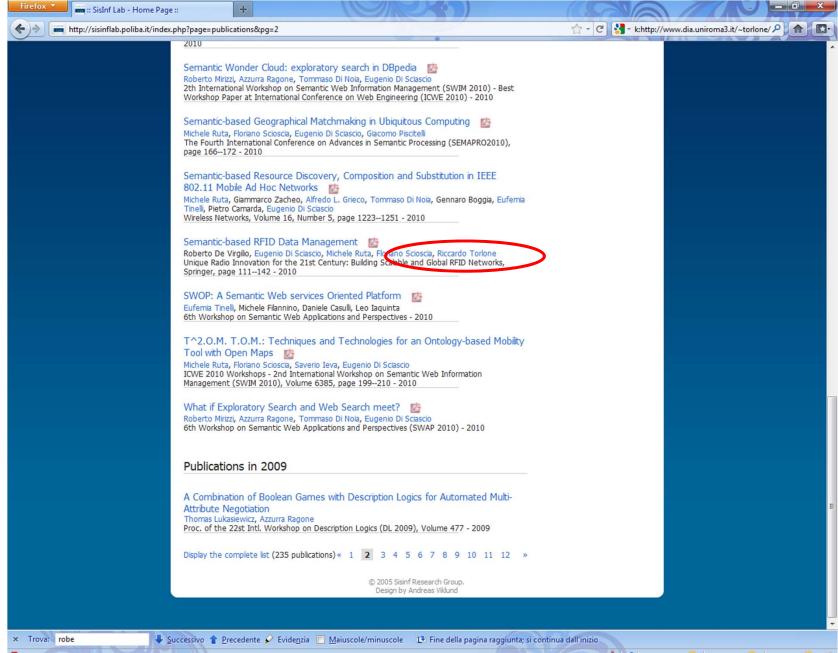
Then some magic happens...

- 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...

Fire	fox 🔨 🚺 Uniroma3	.it :: Prof. Riccardo Torlone	+		(1.5		710	
(*)	> http://www.uni	iroma3.it/persona.php?persona=	=L1a6jJN3wl0+WenCBkmBrsr04BZxcFGyq3v1XO	DbZZM=&cf={cf}	☆ - (් 🚼 - k:http://www.dia.un	iroma3.it/~torlone	
UNIVE			Ribsonta	The				Ê
0	Cerca (nelle pagine e nella	rubrica)	RAPIDO			📰 english	l chi contattare p	er accessibilità
-		trova ricerca avanzat		vai <u>che cos'è r</u>	apido?			
naviga	and all a second s	studenti iscritti laureati per						
	L'Ateneo Didattica	Ricerca Scientifica Se	ervizi agli Studenti Servizi al Personale					
stampa	Home > Rubrica: ricero Prof. Riccaro							
invia	Informazioni p	Informazioni personali Qualora le informazioni riportate a lato risultino assenti, incomplete o errate leggi <u>le seguenti istruzioni</u>						
	Qualifica	Professore Ordinario	incomplete o cirate teggi <u>te seguenti</u>					
	Settore Scientifico Disciplinare	ING-INF/05	Per telefonare da un edificio dell'Aten					
	Tel	0657333377	numero unico inizia con "06 5733xxxx" ultime quattro cifre del numero esteso					E
	Fax	06 57333211						
	Cellulare aziendale	87374						
	Email	torlone@dia.uniroma3.it						
	Indirizzo	Via della Vasca Navale, 79						
	Struttura/Afferenza	 Facolta' di INGEGNERIA Dipartimento di Informatic e Automazione 	ca					
	Cariche e Responsabilità	 Presidente Collegio Didatti di Ingegneria Informatica 	ico					
	Altre Informz 210ni	Sito Web Personale						
	Insegnament	i Ricerca Labora	tori					
Anno accademico 2009-2010								
	Basi di Dati II							
	> Laurea Magistrale	in Ingegneria Informatica (Dm 2	270) - <u>dettaglio</u>					
Calcolatori Elettronici								
	> Laurea in Ingegner	> Laurea in Ingegneria Informatica (Dm 270) - dettaglio 🚓						
 > Laurea Magistrale in Ingegneria delle Tecnologie della Comunicazione e dell'Informazione (Dm 270) - dettaglio re- 								
	Anno accademico 20	008-2009						
	Basi di Dati II							-
₽ • →						🛕 🐝 Adesso: 24 °C 📥	Oggi: 25 °C 🖰	Sab: 26 °C 🏠 嘴

but this one, was not...



🛾 🏠 Adesso: 24 °C 📩 Oggi: 25 °C 选 Sab: 26 °C 选 🐐

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
 - make it possible to "link" to that URI from other sources of data (not only Web pages)
 - i.e., 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...

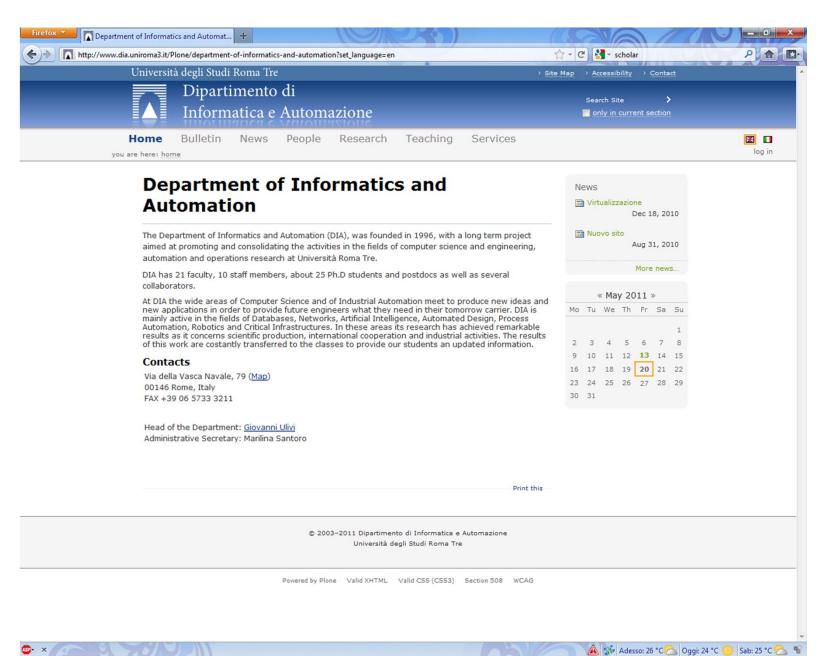
But it is a little bit more complicated

- 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



Human interpretation

- A human understands that this is my institution's home page
- He/she knows what it means (realizes that it is a University Department in Rome)
- 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 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....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...



The rough structure of data integration

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

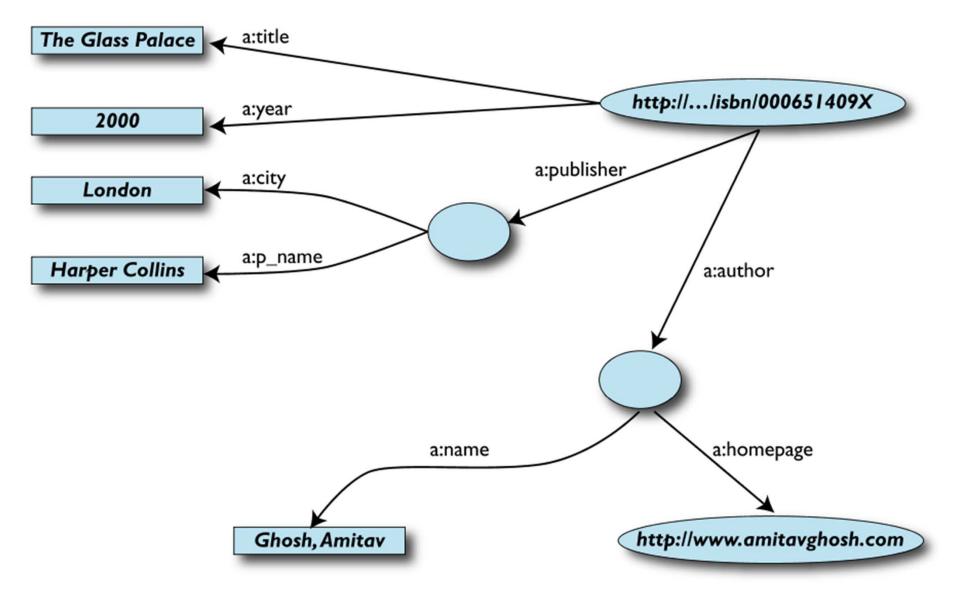
A simplified bookstore data (dataset "A")

ID	AUTHOR	TITLE	PUBLISHER	YEAR
ISBN-00-0651049-X	id_auth1	The glass Palace	id_pub1	2000

ID	AUTHOR NAME	HOMEPAGE
id_auth1	Amitav Gosh	http://www.goshamitav.com

ID	PUBLISHER NAME	CITY
id_pub1	Harper Collins	London

1st: export your data as a set of relations



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 not 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 part of the data

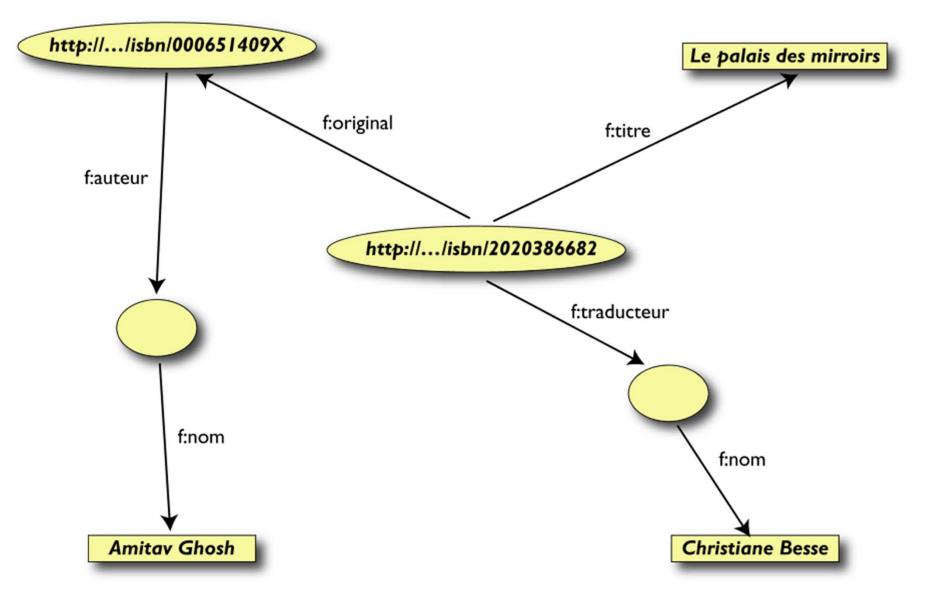
Another bookstore data (dataset "F")

ID	TITLE	TRADUCTER	ORIGINAL
ISBN-02-05668976-Y	The palais des miroirs	A13	ISBN-00-0651049-X

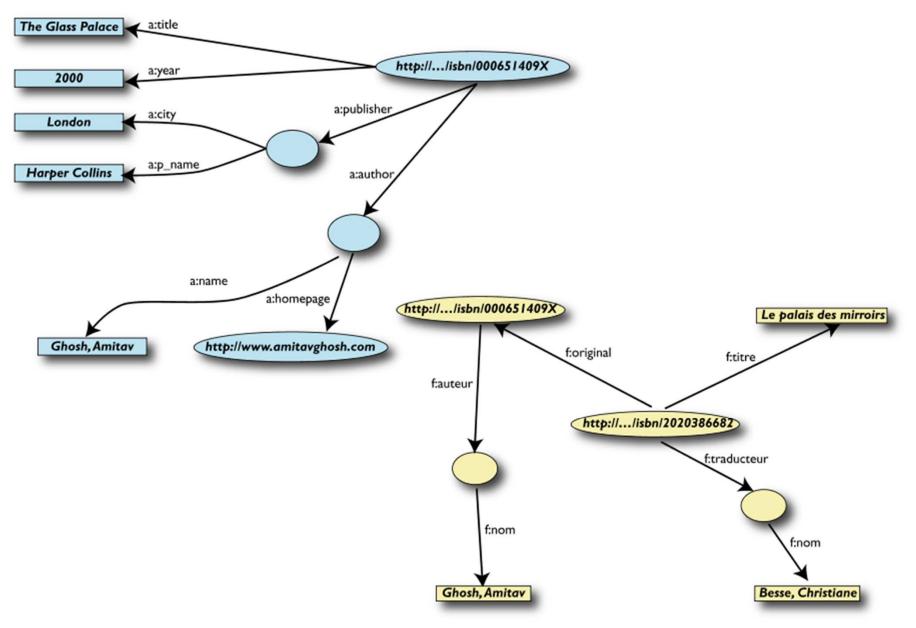
ID	AUTEUR
ISBN-00-0651049-X	A12

ID	NOM
A12	Amitav Gosh
A13	Christianne Bess

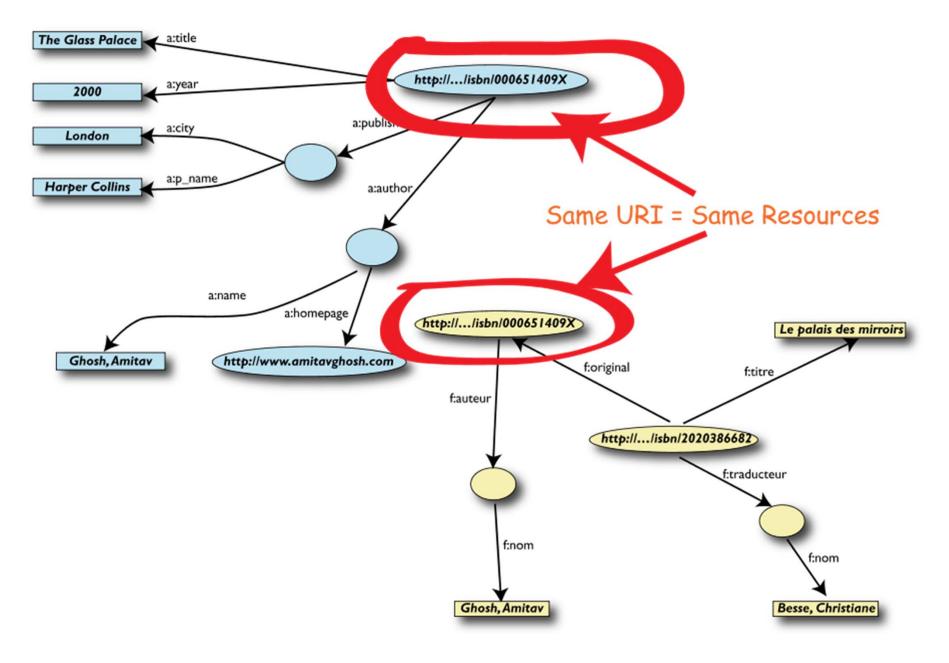
2nd: export your second set of data



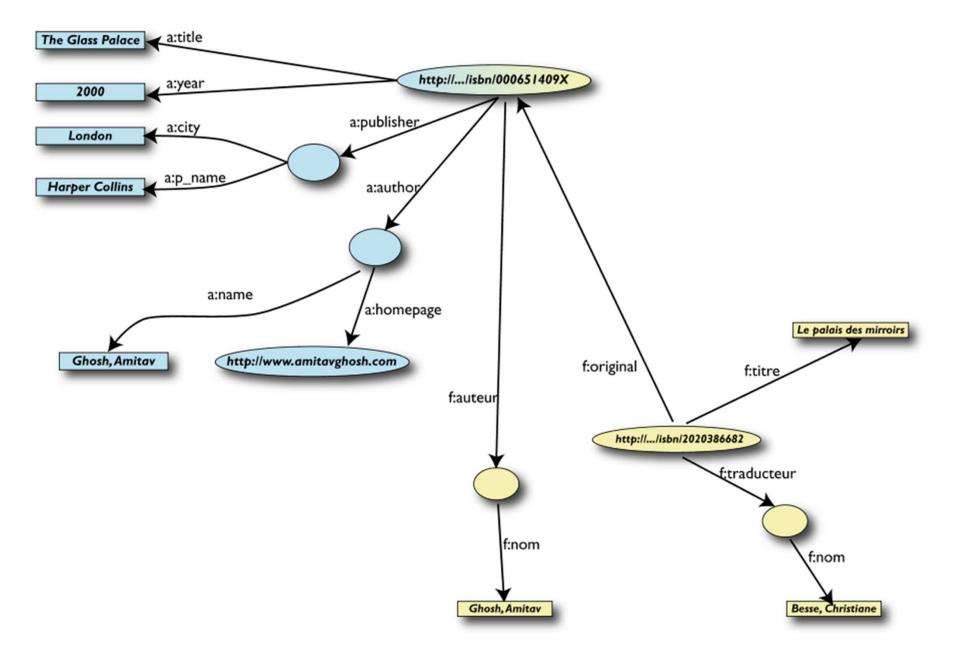
3rd: start merging your data



3rd: start merging your data (cont.)



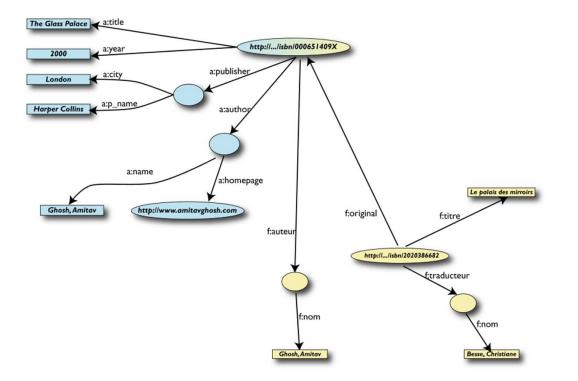
3rd: merge identical resources



Start making queries...

User of data "F" can now ask queries like:

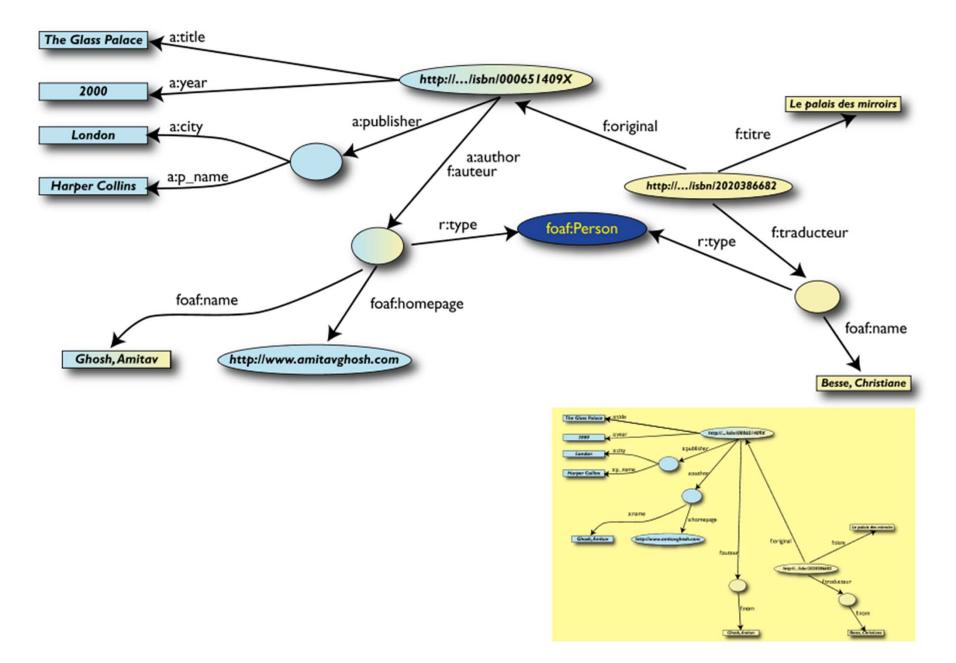
- "give me the title of the original"
 - well, ... "donnes-moi le titre de l'original"
- This information is not in the dataset "F"
- but can be retrieved by merging with dataset "A"!



However, more can be achieved...

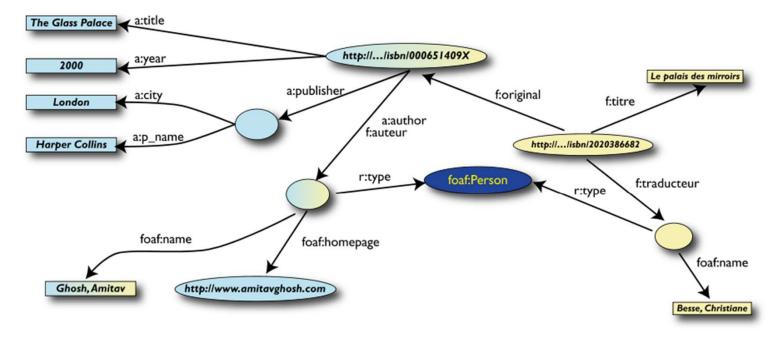
- We "feel" that a:author and f:auteur should be the same
- But an automatic merge doest not know that!
- Let us add some extra information to the merged data:
 - □a:author same as f:auteur
 - □ both identify a "Person"
 - □ a term that a community may have already defined:
 - a "Person" is uniquely 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!

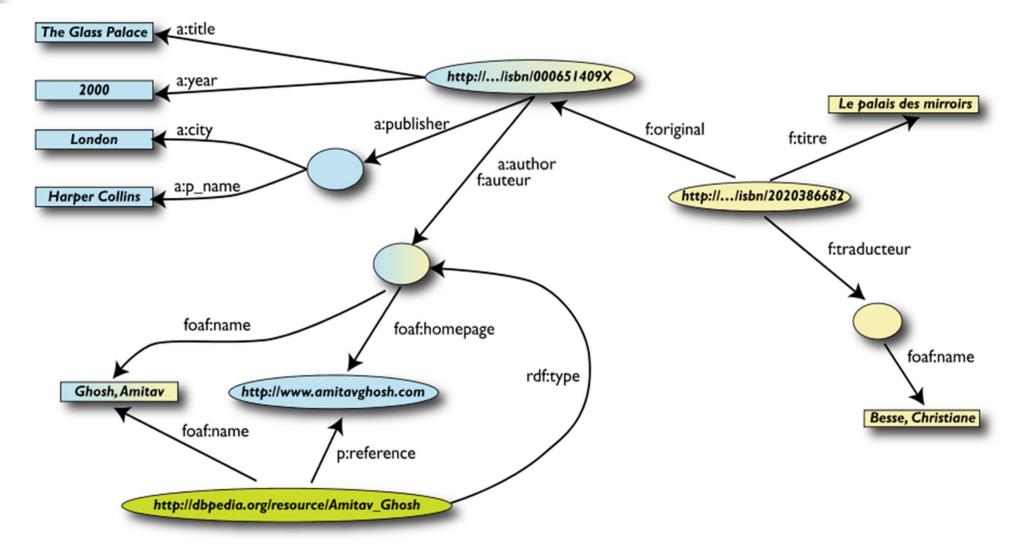
- User of dataset "F" can now query:
 - □ "donnes-moi la page d'accueil de l'auteur de l'originale"
 - well... "give me the home page of the original's 'auteur"
- The information is not in datasets "F" or "A"...
- ...but was made available by:
 - merging datasets "A" and datasets "F"
 adding three simple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



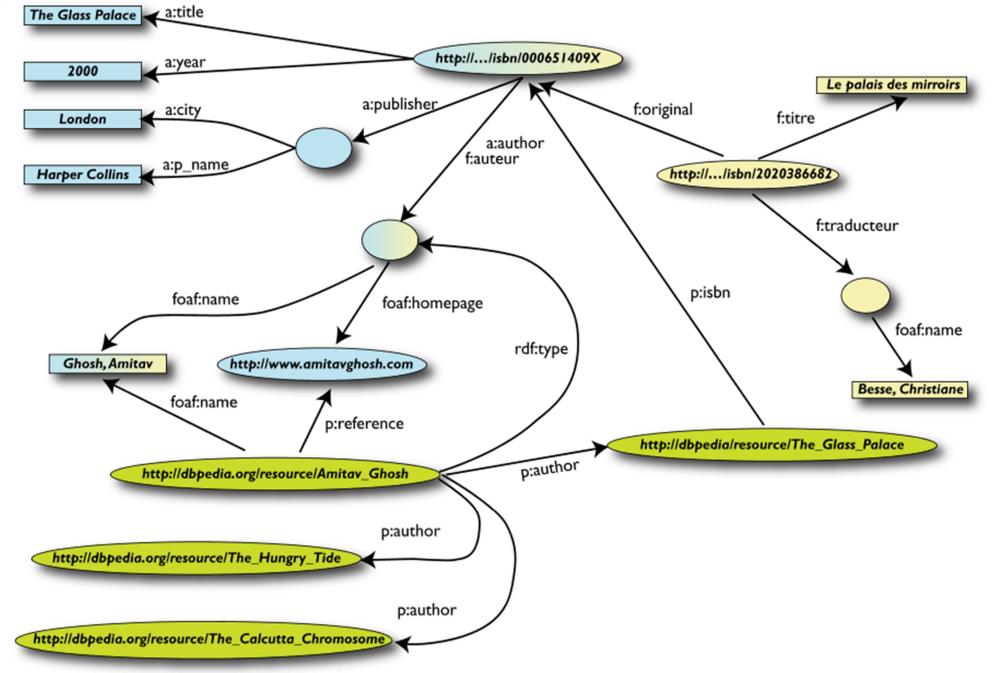
Combine with different datasets

- 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 "<u>dbpedia</u>" project can extract the "infobox" information from Wikipedia already...

Merge with Wikipedia data

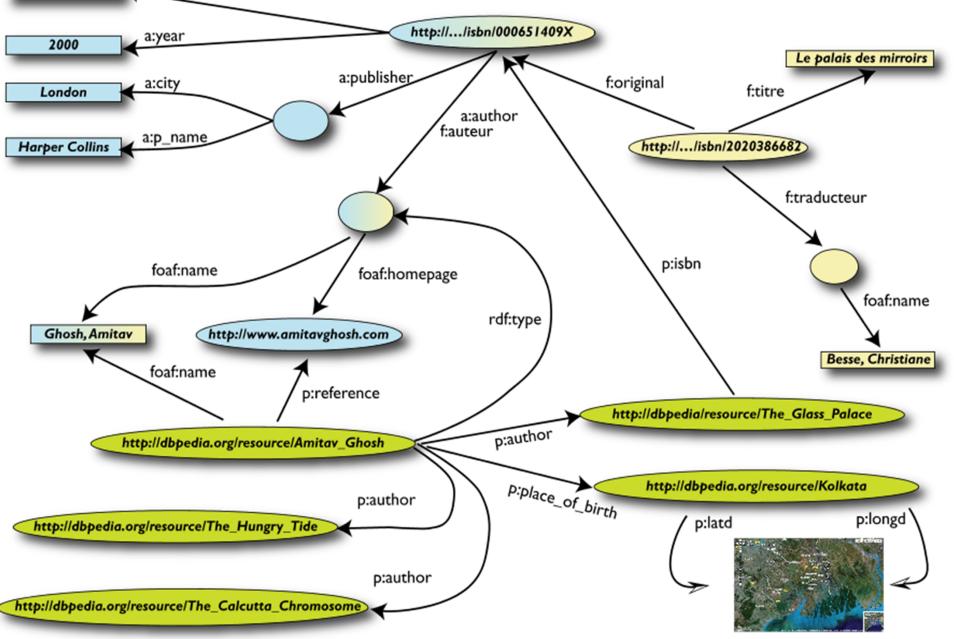


Merge with Wikipedia data



Merge with Wikipedia data

The Glass Palace _____ a:title



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

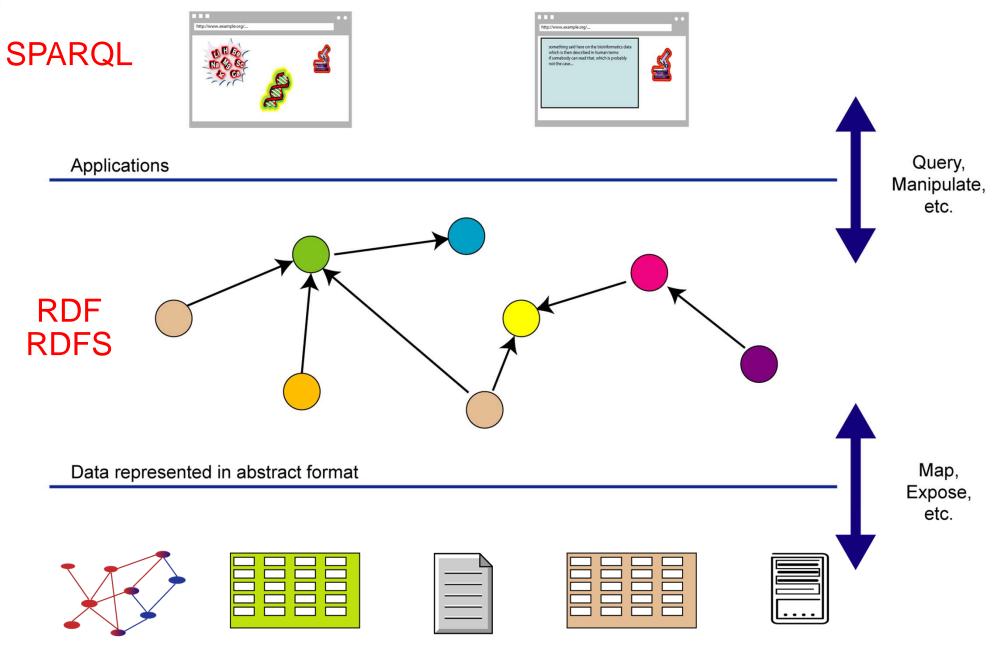
What did we do?

- 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 URIs 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, 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

The Basis: RDF

Let us begin to formalize what we did!
we "connected" the data...

- □ but a simple connection is not enough... data should be named somehow
- hence the RDF Triples: a labelled connection between two resources

Essential Online Resources

- http://www.w3.org/TR/html4/
- http://www.w3.org/Addressing/
- http://www.w3.org/Style/CSS/
- http://validator.w3.org/
- http://www.w3.org/