A no-nonsense introduction to "Semantic Web" technologies

by Stefano Mazzocchi

me

italian

31

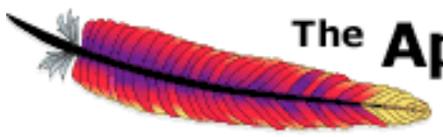
research scientist

digital library research group


MIT Libraries

member

former director



The **Apache Software Foundation**

http://www.apache.org/

author

developer

Simile

my goals today

challenge

stimulate

show real stuff

make you think

make you curious

Part I

The Present

the web

pages and links

you see it like this

you can get it on eBay

All Categories | Search | Advanced Search

My Favorite Searches: Sign in to view your Favorite Searches.

Specialty Sites

- eBay Express NEW!
- eBay Motors
- eBay Stores
- Live Auctions
- Reviews & Guides
- Want It Now
- eBay Business

Categories

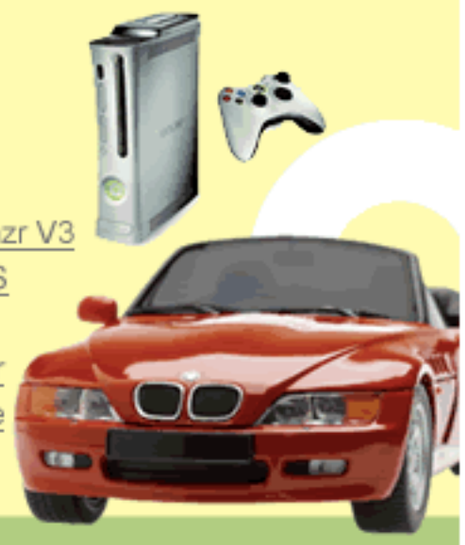
- Antiques
- Art
- Baby
- Books
- Business & Industrial
- Cameras & Photo
- Cars, Boats, Vehicles & Parts
- Cell Phones
- Clothing, Shoes & Accessories
- Coins
- Collectibles
- Computers & Networking
- Consumer Electronics
- Crafts
- Dolls & Bears
- DVDs & Movies
- Entertainment Memorabilia
- Gift Certificates
- Health & Beauty
- Home & Garden

Inside eBay

Trends • Hot Picks • Cool Stuff

Top 10 Most Wanted

1. iPod
2. Xbox 360
3. Sony PSP
4. Star Wars
5. Motorola Razr V3
6. Nintendo DS
7. Cars
8. Longaberger
9. Abercrombie
10. ATV



eBay Pulse

Cool Stuff, Trends & More!

Designer Handbags

Wireless

Digital SLR

My eBay At A Glance

Sign in for a snapshot of your personalized information on the homepage.

Seat your employees in comfort. Find great deals on ergonomic office chairs.

Bid on camera phones from Nokia, Motorola, Samsung and more!

Get an oldie but goodie. For great deals on antique furniture click here.

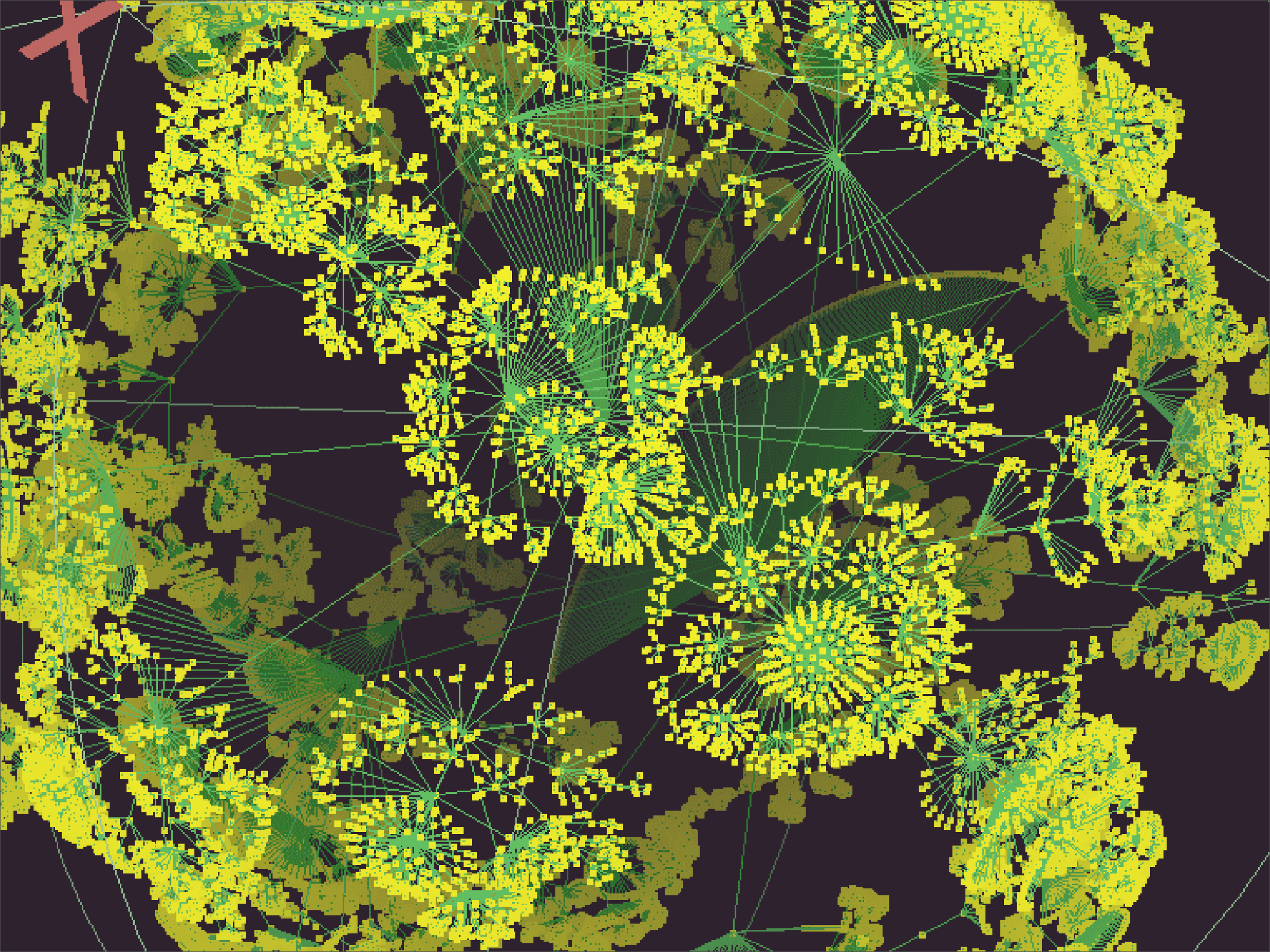
Bid on the top selling books including Eldest, Harry Potter and The Chronicles of Narnia.

ebay express — Get it brand new One quick checkout

Totally free

Use Skype to call all phones within the US and Canada for FREE!

machines see it like this



most of the web is

most of the web is

- designed for direct human consumption

most of the web is

- designed for direct human consumption
- connected but compartmentalized in sites

most of the web is

- designed for direct human consumption
- connected but compartmentalized in sites
- page granular

Very few web sites are done by aggregating other web sites' data

Very few web sites are done by aggregating other web sites' data

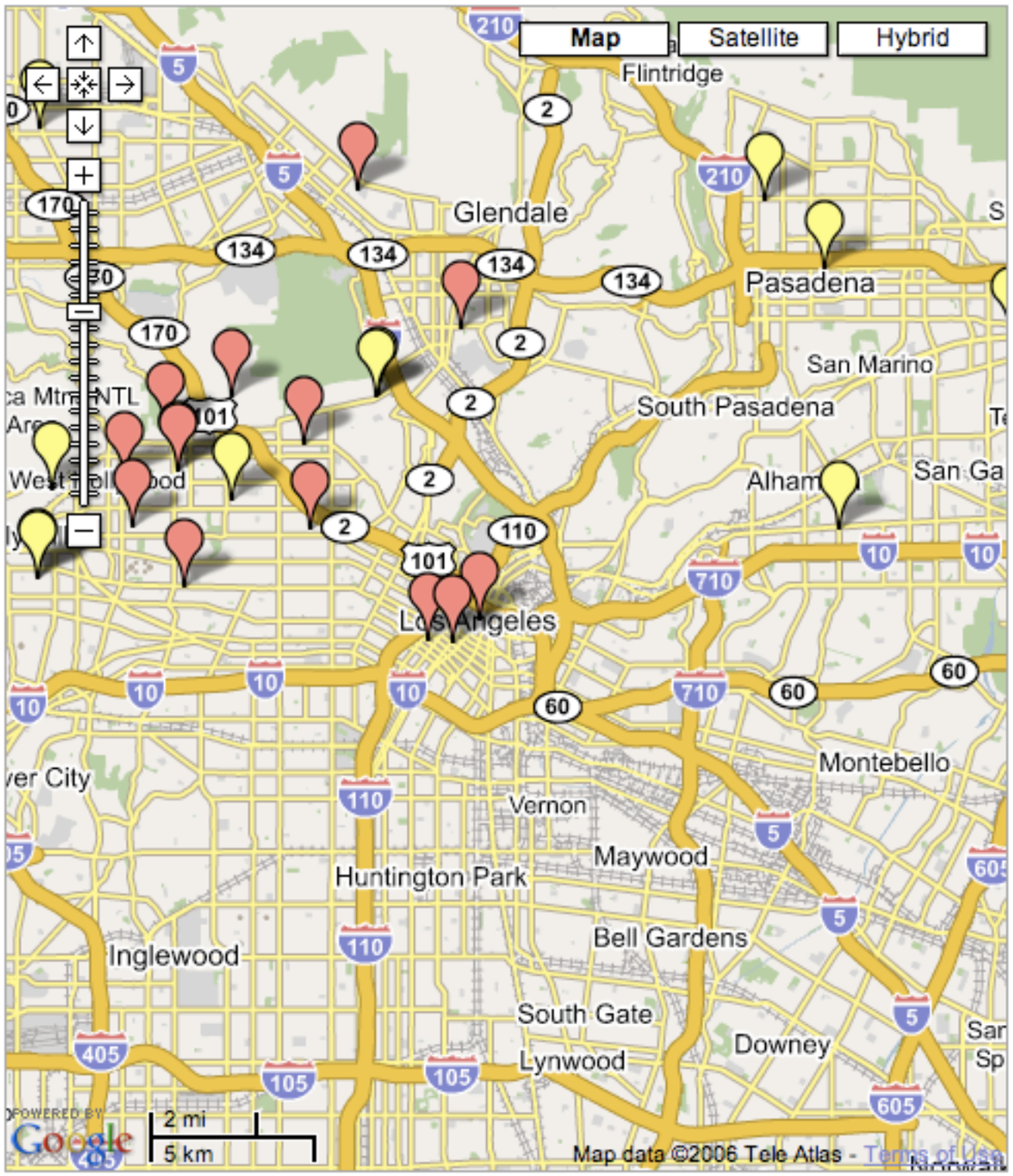
for example

For Rent [For Sale](#) [Rooms](#) [Sublets](#)

Powered by [craigslist](#) and [Google Maps](#)
(this site is in no way affiliated with craigslist or Google)

[About / Feedback](#)

City: Los Angeles **Price:** $1500 – $2000 [Show Filters](#)[New] [Refresh](#) [Link](#)



pics	price	bd	description	city	date
	$1710	2bd	Open House Today! Luxury 2+2 Pool, Gym, Parking	Sherman Oa	5/29
	$1500	2bd	2 Bedroom House In Pasadena	Pasadena	5/29
	$2000	2bd	2 bedroom 1 bath Classic 4-Plex In Hancock Park	Los Angeles	5/29
	$1570	1bd	Long Beach Apartments Pine Ave District	Long Beach	5/29
	$1775	3bd	Ready To Move In	1941 E Work	5/29
	$1695	1bd	1+Loft, private sundeck, Free Dsl, Pool, Microwave & Parking	Van Nuys Us	5/29
	$1700		Downtown Los Angeles Loft	Los Angeles	5/29
	$2000	1bd	Beautiful Master Bedroom En Suite + Private Den	15245 W Su	5/29
	$1875	2bd	Beautiful Spacious Townhouse – 2 Bedroom / 2 Bath	Van Nuys	5/29
	$1900	2bd	A hidden Jewel in the middle of the city, behind commercial shops !	Santa Monic	5/29
	$1700	2bd	Condo for Rent	Long Beach-	5/29
	$1650	2bd	2 Bdrm Apartment – Pasadena	Pasadena	5/29
	$2000	2bd	with den/office, hardwood floors, parking, A/C – 2 blks N of the Grove	Los Angeles	5/29
	$1695	2bd	1 Bath House W/Large Back Yard!	Glendale	5/29
	$1650	2bd	Spacious Two Bedrooms in Gated Community!	1725 Nogale	5/29
	$1650	1bd	Venice Guest House For Rent	Venice	5/29
	$1950	3bd	Luxury Condo in San Fernando Valley	Canoga Park	5/29
	$1884	2bd	Gorgeous Ocean Views! Furnished! Remodeled! Garage! Utilities Included	941 Westwo	5/29
	$1500	2bd	The Beautiful 2+2.5 Townhouse Style Apartments	North Hollyw	5/29
	$2000	2bd	Santa Monica Mountains in Malibu, 2Bd/2BA	Malibu	5/29
	$1750	1bd	Large Sunny 1 Bed W/ Private Patio	Beverly Hills	5/29
			Beautiful 2+2 units – Huge and		

Home Info Support Get Listed Buzz Blog Stats: 2425 artists, 3815 songs

Podbop

We podcast bands coming to your town

Search Another City: GO

LOS ANGELES, CA: PODCAST M3U POP-UP

MAY 30

Pinback

Tuesday 9:00 PM at Avalon Hollywood More details

★ Pinback - ♪ B ▶

★ Pinback - ♪ Messenger ▶

MAY 30

The Jade Shader

Tuesday 9:00 PM at The Avalon More details

★ The Jade Shader - ♪ Cha Cha Choo Choo ▶

JUN 2

Alkaline Trio

Friday 7:00 PM at Avalon Hollywood More details

★ Alkaline Trio - ♪ All On Black ▶

★ Alkaline Trio - ♪ Queen Of Pain ▶

JUN 3

Alkaline Trio

Saturday 7:00 PM at Avalon Hollywood More details

Welcome in 2006!

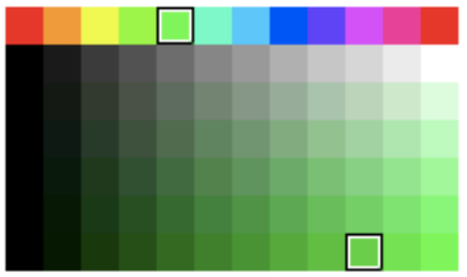
retrievr

All images

Search by:
Sketch • Image

Undo Redo Clear

20 30 50



New! Please help us rate sketches for The Art of retrievr! Only a million to go ...

Still new! You can search by

From T Glow

From T Glow

From Renata Diem

From phitar

From MARIO MARCANTE

From eyecatcher

From angelocesare

From Photos Marcante®

From Ahmed Zahid

From jayjuice

From h.neumeyer

From zomerstorm

From Thiago Lopes

From Charles Machado

From phil h

From Katemina

Web Images Groups News Froogle Maps **more »**

Advanced Search
Preferences
Language Tools

Google Search I'm Feeling Lucky

Enron's Skilling and Lay convicted... See what bloggers are saying

Search the Blogosphere SEARCH Options

Explore the Blogosphere

Explore All »

Tech

Guild Wars Vs WoW Vs MC Hammer
A Fan-made Music Video Sets Out to Settle If Guild Wars or World of Warcraf...

Free Blogger Templates
If You Are Blogger/Blogspot Blogger, You Can Download Some FREE High Qualit...

Tech »

Politics

Jefferson's History of Sleaze
I Never Realized What a Sleazy Character William Jefferson Is -- I've Been ...

JOHN LOGAN AND MEMORIAL DAY
5/28/2006 THE COURAGE OF ONE'S CONVICTIONS AND WHERE IT'S LACKING CATEGORY:...

Politics »

See also: Popular: YouTube, News, Movies, Books and Top 100 blogs.

Michelle Madhok's Favorite Blogs

Latest posts:

- Fire Magic Monarch Magnum
- A Find: Diane Von Furstenberg Eyelet
- Roberto Cavalli Freedom Bag, Handbag

View Favorites »

What's happening right now

Currently tracking 41.7 million sites and 2.5 billion links.

Top Searches

- Clay Aiken
- Swift Boating The Planet
- Lost
- Haditha
- Macbook
- Memorial Day
- Valleyschwag

Hot Tags

- memorial day
- Iraq
- web-20
- wordpress
- Sex
- iPod
- clayaiken
- Politics

Featured Blog Add Yours!

Randy's Publicity Newsgroup

By Randy Mast. Tagged Actors, Authors, Blog, Business, fashion, Journalism, Marketing...

More in Blog Finder »

ORBITZ AND GO!™

Welcome to Orbitz. Sign in | Register now
Site feedback

Quick Search | Vacation Packages | Hotels | Cars-Rail | Cruises | Flights | My Trips | My Account (MY STUFF) | Deals | Customer Service | TLC

My Search

Flight + Hotel deals for Stockholm

Find flights by:

Stops	Price	Airline: Continental Airlines	Multiple Carriers	Lufthansa	US Airways	British Airways	Northwest Airlines	United Airlines
Non-stop								
1 stop		$1,349 total $1,449	$1,345 total $1,452	$1,472 total $1,764	$1,750 total $1,850	$1,570 total $1,873	$1,770 total $1,884	$1,870 total $1,970
2+ stops			$1,345 total $1,459			$3,010 total $3,307		

priceline.com — Save up to 40% on your Flight. Search Priceline

Fares are per person in US dollars, using e-tickets. Total fare includes all taxes and fees .
Some itineraries require paper tickets with an additional charge. Changes after purchase are subject to change fees.

Sort flights by: Lowest price | Departure time | Shortest flight — Airport codes

Change Search

From City name or airport
LAX
incl. nearby airports

To City name or airport
STO
incl. nearby airports

Leave
06/22/06
Anytime

Return

Showing **lowest priced flights** (45 flights out of 199 total) — See all 199 flights

Select **$1,349** + $100 taxes & fees = $1,449 per person — TLC Part of OrbitzTLC

Leave **Thu, Jun 22** — **Continental Airlines 1703**

Depart: 7:45am Los Angeles, CA (LAX)
Arrive: 4:05pm Newark, NJ (EWR)

1 stop — Economy | 5hr 20min | Boeing 757 | View seats

Change planes. Time between flights: **1hr 30min**

Continental Airlines 68

Depart: 5:35pm Newark, NJ (EWR)
Arrive: 7:45am Stockholm, Sweden (ARN)

Economy | 8hr 10min | Boeing 757 | View seats
Total duration: 15hr 0min

This is an overnight flight.

Choose this departure

Mashups

Mashups

- mix data from different sites to provide added value

Mashups

- mix data from different sites to provide added value
- the mashed sites don't need to be involved

Mashups

- mix data from different sites to provide added value
- the mashed sites don't need to be involved
- hybrid client-server model: client on one side, server on the other

Mashup Problems

Mashup Problems

The Obvious Ones

Mashup Problems

The Obvious Ones

- data is mostly locked in pages

Mashup Problems

The Obvious Ones

- data is mostly locked in pages
- each web site is different

Mashup Problems

The Obvious Ones

- data is mostly locked in pages
- each web site is different
- and keeps changing

Mashups Problems

Mashups Problems

The Not-so-obvious Ones

Mashups Problems

The Not-so-obvious Ones

- Very blurry line between Use and Fair Use

Mashups Problems

The Not-so-obvious Ones

- Very blurry line between Use and Fair Use
- Even after extraction, data needs to be modeled so that it can mix together

Mashups Problems

The Not-so-obvious Ones

- Very blurry line between Use and Fair Use
- Even after extraction, data needs to be modeled so that it can mix together
- Very strong and hardly movable dependency on the mashed data

Mashups Problems

The Not-so-obvious Ones

- Very blurry line between Use and Fair Use
- Even after extraction, data needs to be modeled so that it can mix together
- Very strong and hardly movable dependency on the mashed data
- A mashed-up web site looks just like another web site (so further mashing is not easier!)

Mashing in short

Mashing in short

- extremely useful

Mashing in short

- extremely useful
- hard

Mashing in short

- extremely useful
- hard
- doesn’t cascade

Goals for a Web of Data

Goals for a Web of Data

- Making mashing easier

Goals for a Web of Data

- Making mashing easier
- Making cascading possible

Success story: RSS/Atom

Success story: RSS/Atom

- describes a chronological sequence of items

Success story: RSS/Atom

- describes a chronological sequence of items
- consumers poll and receive new items

Success story: RSS/Atom

- describes a chronological sequence of items
- consumers poll and receive new items
- web sites can be easily mashed-up, aggregated

Success story: RSS/Atom

- describes a chronological sequence of items
- consumers poll and receive new items
- web sites can be easily mashed-up, aggregated
- the model cascades

Improvements

Improvements

- XML model vs. HTML model

Improvements

- XML model vs. HTML model
- extensibility thru XML namespaces

Improvements

- XML model vs. HTML model
- extensibility thru XML namespaces
- granularity/identification at the item level

is it enough?

is it enough? No :-(

Limitations

Limitations

- no standard way to represent relationships between items

Limitations

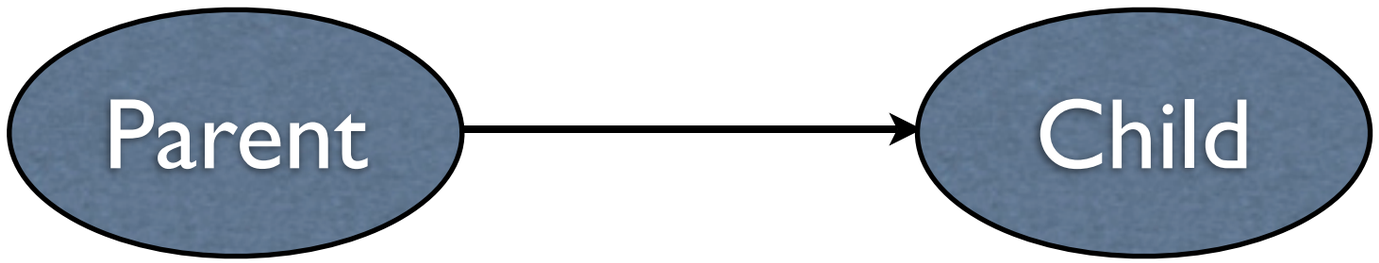
- no standard way to represent relationships between items
- no standard way to query the web site other than polling

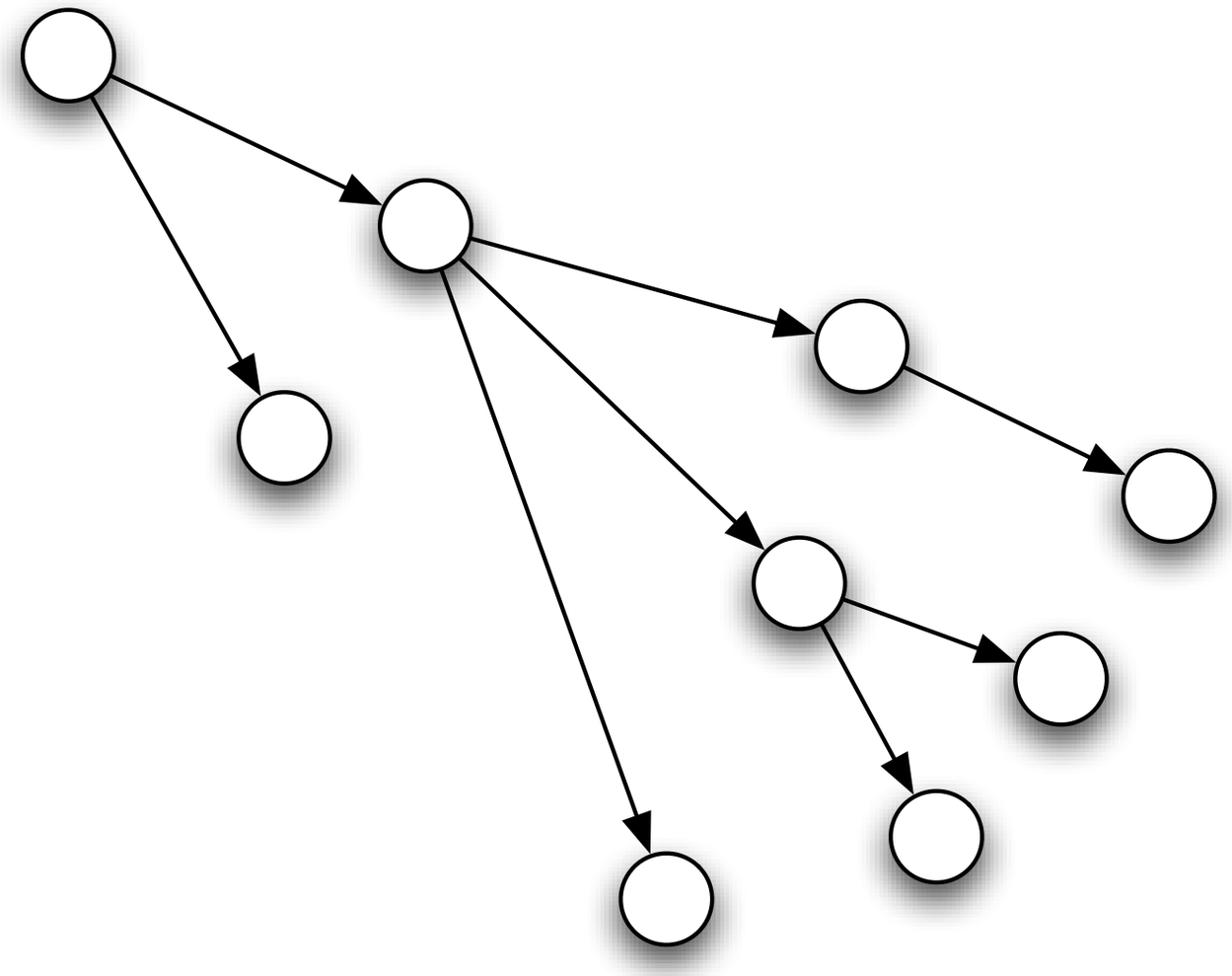
so what do we do?

Part II

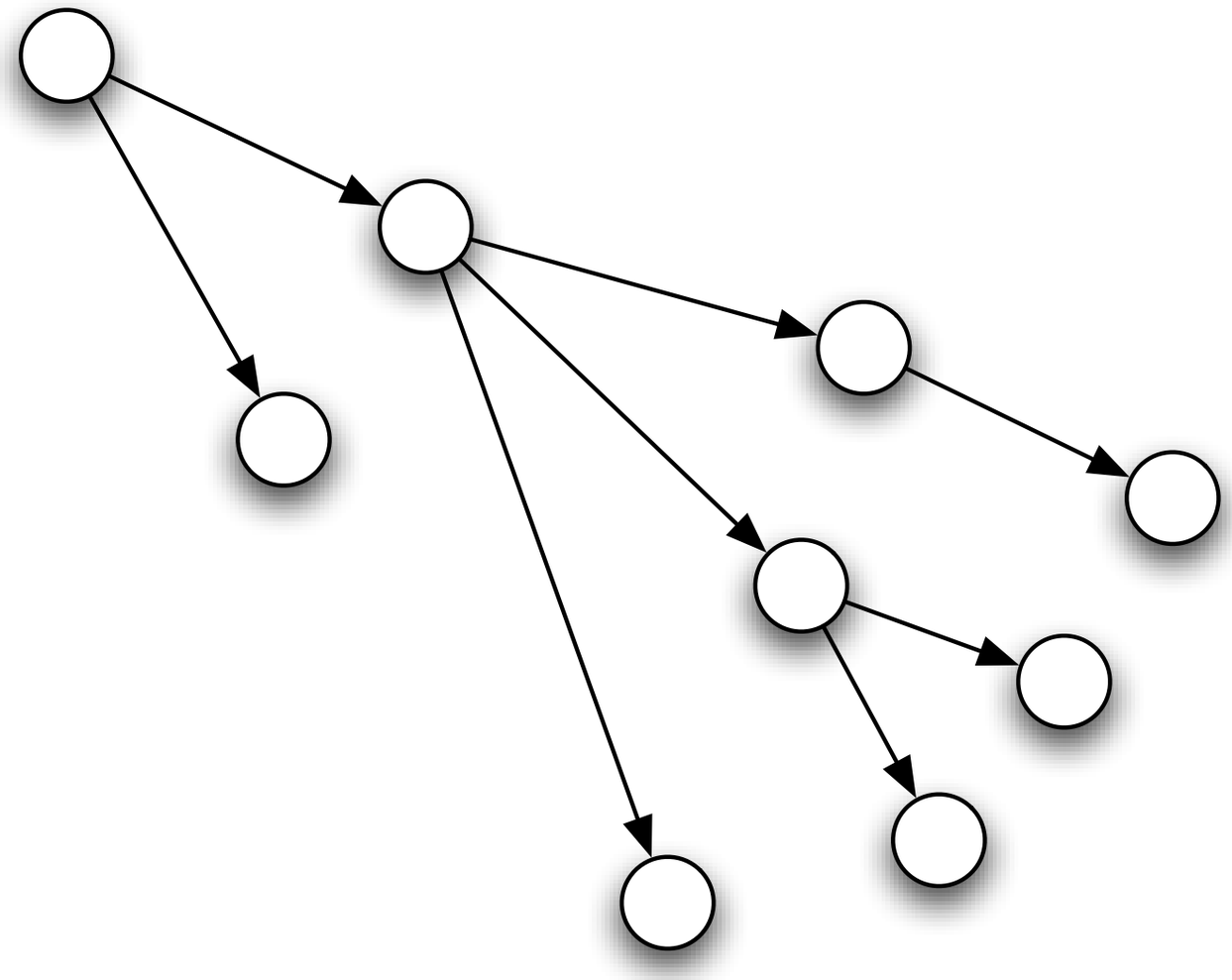
The Future

XML

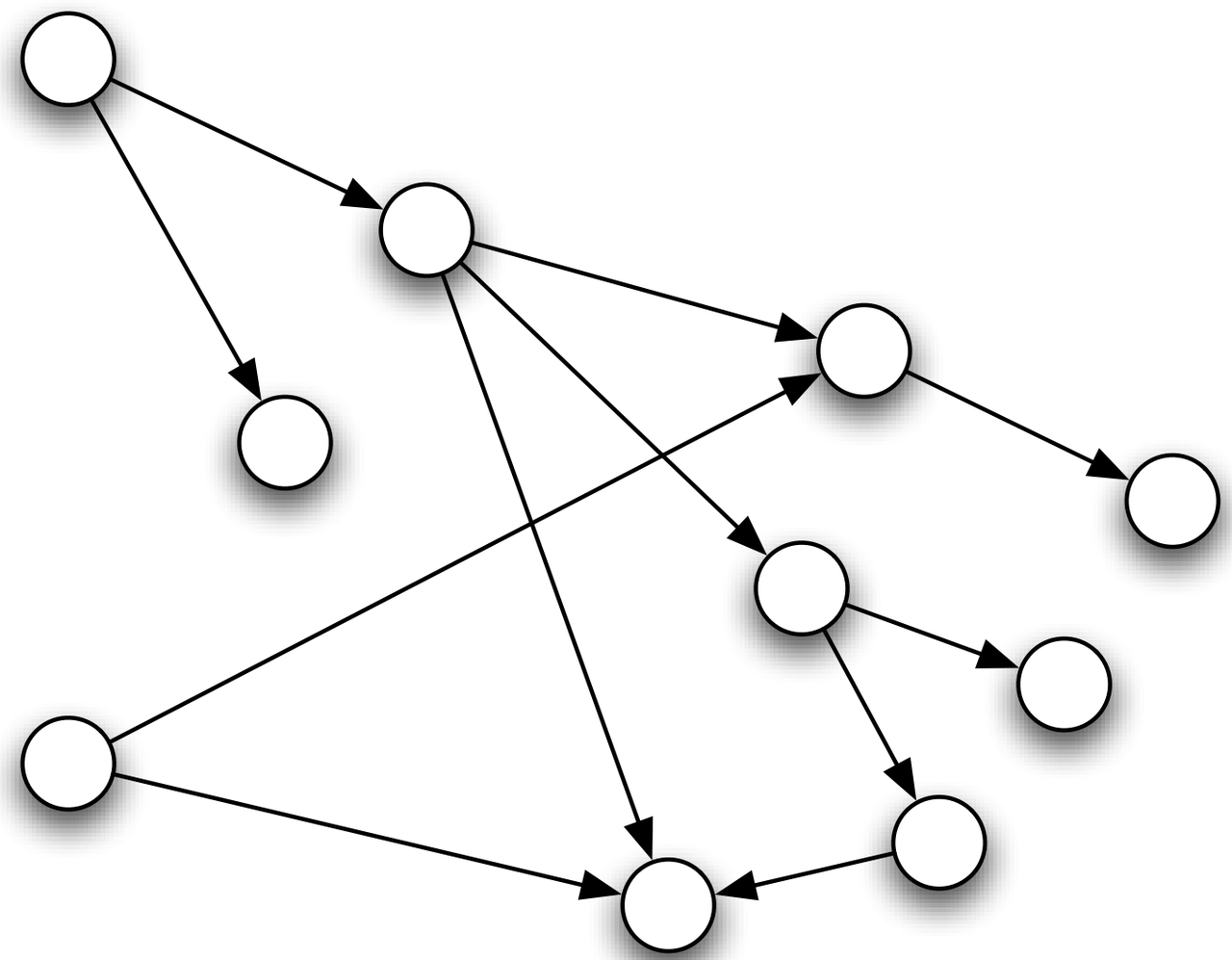




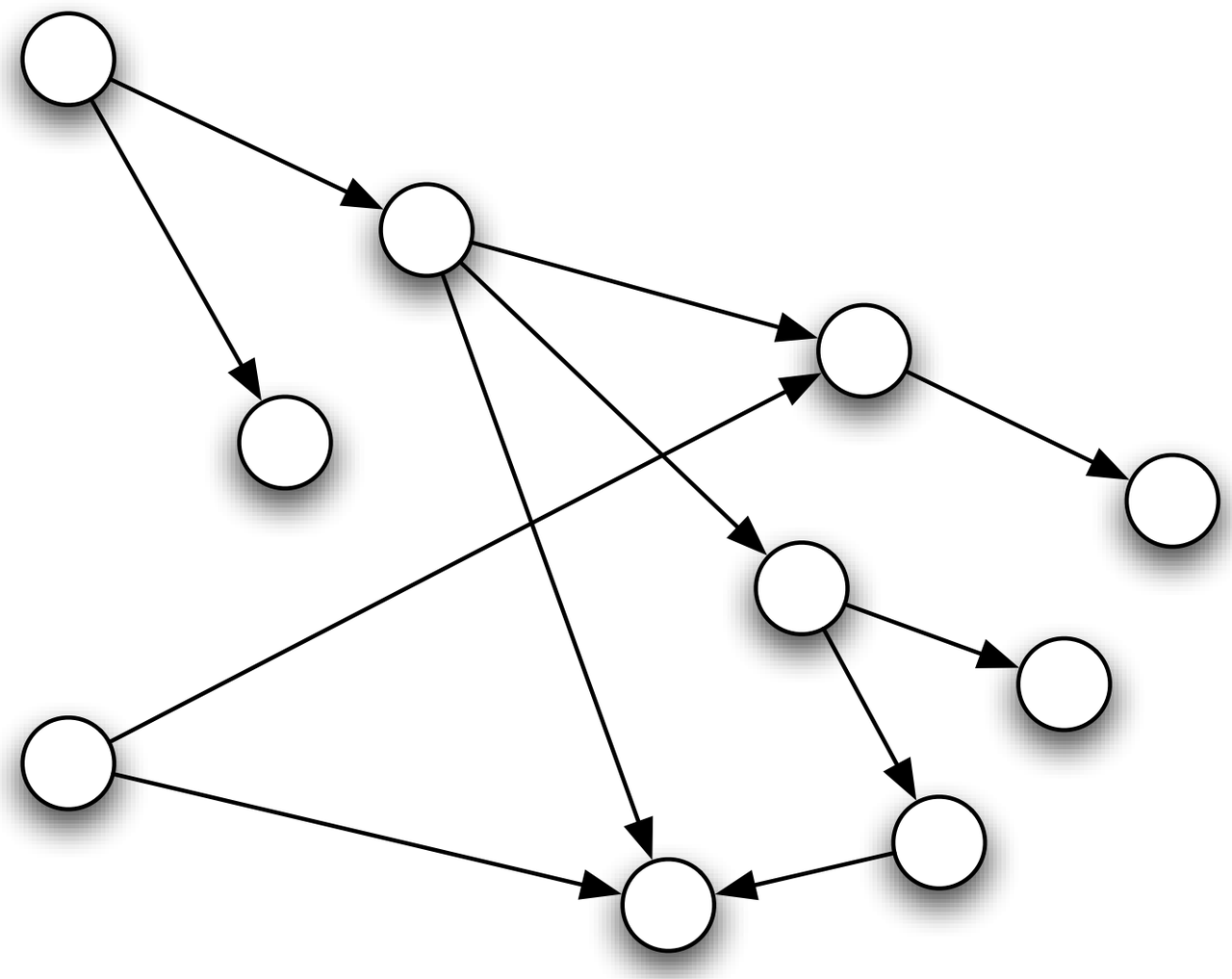
A Tree



RDF



A Graph



RDF

RDF

- Resource Description Framework

RDF

- Resource Description Framework
- W3C Recommendation since 1998 (as old as XML!)

RDF

- Resource Description Framework
- W3C Recommendation since 1998 (as old as XML!)
- Misunderstood for years as a very complicated way of embedding metadata into XML documents

Why misunderstood?

Why misunderstood?

- Original specification was extremely formal

Why misunderstood?

- Original specification was extremely formal
- “what is this good for” was nowhere to be found (somewhat taken for granted by the people that designed it)

Why misunderstood?

- Original specification was extremely formal
- "what is this good for" was nowhere to be found (somewhat taken for granted by the people that designed it)
- The RDF/XML serialization obscured the value of the graph data model

Why misunderstood?

- Original specification was extremely formal
- "what is this good for" was nowhere to be found (somewhat taken for granted by the people that designed it)
- The RDF/XML serialization obscured the value of the graph data model
- XML seemed to solve the same problems and was much easier to understand

Why misunderstood?

- The real problem is that RDF was conceived as a solution to a problem people didn't have:

Why misunderstood?

- The real problem is that RDF was conceived as a solution to a problem people didn't have:

data interoperability at a world-wide scale

Design Principles

Design Principles

- Reducing the unit of information, from documents to statements

Design Principles

- Reducing the unit of information, from documents to statements
- Each part of a statement can be given a globally unique identifier

Design Principles

- Reducing the unit of information, from documents to statements
- Each part of a statement can be given a globally unique identifier
- The data model is general enough to describe any other data model

Example

"This presentation was written by Stefano, in 2006"

Example

"This presentation was written by Stefano, in 2006"

```
<> rdf:type docs:Presentation ;
   dc:creator stefano:me ;
   dct:creation_date "2006" .
```

Example

"This presentation was written by Stefano, in 2006"

```
@prefix rdf: <http://www.w3.org/1999/02/22-prefixRDF-syntax-ns#> .
@prefix dc:  <http://purl.org/dc/elements/1.1/>  .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix docs: <http://simile.mit.edu/2005/04/ontologies/docs#> .
@prefix stefano: <http://www.betaversion.org/~stefano/#> .
```

```
<> rdf:type docs:Presentation ;
   dc:creator stefano:me ;
   dct:creation_date "2006" .
```

in RDF/XML

in RDF/XML

```
<rdf:RDF
   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:dc="http://purl.org/dc/elements/1.1/"
   xmlns:dct="http://purl.org/dc/terms/"
   xmlns:docs="http://simile.mit.edu/2005/04/ontologies/docs#"
>
  <docs:Presentation rdf:about="">
    <dc:creator rdf:resource="http://www.betaversion.org/~stefano/#me"/>
    <dct:creation_date>2006</ns2:creation_date>
  </docs:Presentation>
</rdf:RDF>
```

plain old XML?

plain old XML?

```
<presentation xmlns="http://..." >
 <authors>
  <author name="Stefano"/>
 </authors>
 <year>2006</year>
</presentation>
```

what's wrong with that?

what's wrong with that?

- Imagine 10,000,000 of such pages

what's wrong with that?

- Imagine 10,000,000 of such pages
- Authored by 10,000 people

what's wrong with that?

- Imagine 10,000,000 of such pages
- Authored by 10,000 people
- Using 100 schemas in 50 languages and 10 character sets

what's wrong with that?

- Imagine 10,000,000 of such pages
- Authored by 10,000 people
- Using 100 schemas in 50 languages and 10 character sets
- Harvest from the web and mix together

what's wrong with that?

- Imagine 10,000,000 of such pages
- Authored by 10,000 people
- Using 100 schemas in 50 languages and 10 character sets
- Harvest from the web and mix together
- Find out how many presentations I've written in 2006.

Is RDF enough? not quite

Further Problems

Further Problems

- Decentralized use and creation of data, identifiers and vocabularies generates lots of disconnected graphs, hosted on many different web sites

Further Problems

- Decentralized use and creation of data, identifiers and vocabularies generates lots of disconnected graphs, hosted on many different web sites
- Need a way to fetch only what I need

Further Problems

- Decentralized use and creation of data, identifiers and vocabularies generates lots of disconnected graphs, hosted on many different web sites
  - Need a way to fetch only what I need
  - Need a way to link disconnected graphs together

SPARQL

SPARQL

- Query language for RDF

SPARQL

- Query language for RDF
- W3C Working Draft

SPARQL

- Query language for RDF
- W3C Working Draft
- SQL-like syntax

SPARQL Example

```
PREFIX foaf:   <http://xmlns.com/foaf/0.1/>
SELECT ?name ?mbox
WHERE
  { ?x foaf:name ?name .
    ?x foaf:mbox ?mbox }
```

SPARQL Example

```
PREFIX foaf:   <http://xmlns.com/foaf/0.1/>
SELECT ?name ?mbox
WHERE
  { ?x foaf:name ?name .
    ?x foaf:mbox ?mbox }
```

```
@prefix foaf:  <http://xmlns.com/foaf/0.1/> .

_:a  foaf:name   "Johnny Lee Outlaw" .
_:a  foaf:mbox   <mailto:jlow@example.com> .
_:b  foaf:name   "Peter Goodguy" .
_:b  foaf:mbox   <mailto:peter@example.org> .
```

SPARQL Example

```
PREFIX foaf:   <http://xmlns.com/foaf/0.1/>
SELECT ?name ?mbox
WHERE
  { ?x foaf:name ?name .
    ?x foaf:mbox ?mbox }
```

SPARQL Example

```
PREFIX foaf:   <http://xmlns.com/foaf/0.1/>
SELECT ?name ?mbox
WHERE
  { ?x foaf:name ?name .
    ?x foaf:mbox ?mbox }
```

Johnny Lee Outlaw	<mailto:jlow@example.com>
Peter Goodguy	<mailto:peter@example.org>

SPARQL Benefit

- Coupled with a web service and a result serialization format (both working drafts at W3C), provides a powerful and efficient access point for a distributed RDF ecosystem

OWL

OWL

- Web Ontology Language

OWL

- Web Ontology Language
- An RDF vocabulary to further describe RDF data

OWL Examples

OWL Examples

- "married to" is a symmetric property

OWL Examples

- “married to” is a symmetric property
- “father of” is not

OWL Examples

- “married to” is a symmetric property
- “father of” is not
- “human” is a subclass of “animal”

OWL Examples

- “married to” is a symmetric property
- “father of” is not
- “human” is a subclass of “animal”
- “person” is equivalent to “human”

OWL Examples

- “married to” is a symmetric property
- “father of” is not
- “human” is a subclass of “animal”
- “person” is equivalent to “human”
- there can only be one biological mother

OWL Examples

- “married to” is a symmetric property
- “father of” is not
- “human” is a subclass of “animal”
- “person” is equivalent to “human”
- there can only be one biological mother
- “son of” is the inverse of “father of”

OWL Reasoning

- The act of "inferring" additional statements out of an RDF model

OWL Reasoning Example

stefano@apache.org -(author)-> Cocoon

stefanom@mit.edu -(same as)-> stefano@apache.org

then

stefanom@mit.edu -(author)-> Cocoon

does all this work?

let me show you

Demos

Part III

Advantages

Mixable Data

Mixable Data

- RDF Statements retain their meaning in isolation, unlike XML elements that are meaningful only in their XPath context.

Mixable Data

- RDF Statements retain their meaning in isolation, unlike XML elements that are meaningful only in their XPath context.
- This makes RDF data naturally mixable.

Liquid Data

Liquid Data

- unlike most people fear, RDF does not force you to use common vocabularies, use the one the fits you best, or invent your own.

Liquid Data

- unlike most people fear, RDF does not force you to use common vocabularies, use the one the fits you best, or invent your own.
- data-first vs. schema-first

Liquid Data

- unlike most people fear, RDF does not force you to use common vocabularies, use the one the fits you best, or invent your own.
- data-first vs. schema-first
- faster prototypes, better ROI, designed for change, incredible flexibility

Data Mappings

Data Mappings

- "same as" predicates allow RDF engines to draw equivalences between identifiers

Data Mappings

- "same as" predicates allow RDF engines to draw equivalences between identifiers
- interoperability by mapping vs. interoperability by transformation

Data Mappings

- "same as" predicates allow RDF engines to draw equivalences between identifiers
- interoperability by mapping vs. interoperability by transformation
- more incremental

Data Mappings

- "same as" predicates allow RDF engines to draw equivalences between identifiers
- interoperability by mapping vs. interoperability by transformation
- more incremental
- doesn't suffer the n^2 problem

Querying

Querying

- SPARQL is a query language for RDF, finishing W3C recommendation

Querying

- SPARQL is a query language for RDF, finishing W3C recommendation
- Think of it as SQL for RDF graphs (very similar syntax too!)

Querying

- SPARQL is a query language for RDF, finishing W3C recommendation
- Think of it as SQL for RDF graphs (very similar syntax too!)
- Provides the ability to query a web site for data

The Web of Data

The Web of Data

- From a web of pages and links to a web of resources and explicit relationships

The Web of Data

- From a web of pages and links to a web of resources and explicit relationships
- Both resources and relationships are globally and uniquely identified

The Web of Data

- From a web of pages and links to a web of resources and explicit relationships
- Both resources and relationships are globally and uniquely identified
- Statements are mixable units of information

The Web of Data

- From a web of pages and links to a web of resources and explicit relationships
- Both resources and relationships are globally and uniquely identified
- Statements are mixable units of information
- Identifiers can be mapped and equated

The Web of Data

- From a web of pages and links to a web of resources and explicit relationships
- Both resources and relationships are globally and uniquely identified
- Statements are mixable units of information
- Identifiers can be mapped and equated
- Web sites can be queried

now we have all the design pieces

and the tools are out there

can you afford to ignore it?

Thanks!

http://simile.mit.edu/

stefanom@mit.edu