Linked Data : Exposing your data on the Web

MILAN STANKOVIC¹,² & PHILIPPE LAUBLET¹

¹STIH – UNIVERSITÉ PARIS-SORBONNE
²HYPIOS
Outline

- **Basic Notions**
  - What is Linked Data
  - Basic instruments for publishing data (RDF, OWL,...)
  - What are the objectives of publishing structured data on the Web

- **How do we publish Data on the Web**
  - RDF files & SPARQL Endpoints

- **Collaboration in the Knowledge Community**
  - Web 2.0 Collaborative Tools
  - SemSLATES methodology
Basic Notions

- **Linked Open Data**: a social and technical phenomenon
  - More and more data becoming openly available
    - government data, cultural content,...
  - Important to make data accessible to machines
    - facilitates data exchange between heterogeneous systems
    - enables the growth of services and applications based on public data
    - enables alternative ways to visualize and navigate through data
  - The quantity of available public Linked Data is doubling every year (since 2007)
Ontologies Define the Meaning of Data
Ontologies, Data, Web Resources

**Presentation**
- hasTitle: myPresentation
- hasCreationDate: 08/02/2011
- hasAuthor: me

**Person**
- hasName: Milan
- hasName: Paris
- livesIn: Karl Lagerfeld

**Linked Data: Exposing your data on the Web**

**Document**
- is subclass of Ontologies, Data, Web Resources
Web of Pages -> Web of Data

- On the Web of Pages
  - we put pages of content
  - pages have URLs
  - we link pages to other pages using hyperlinks

- On the Web of Data
  - we put data directly online
  - data objects have URIs
    - higher granularity
  - we link data objects to other objects defined elsewhere, using semantic relations
  - we use ontologies to give meaning to data
Basic Notions: URI

- URI: Uniform Resource Identifier
  - uniquely identifies resources on the Web
  - enables us to refer to a resource and interact with it
  - URL, as a difference, allows to locate a resource. Not every URI is a URL
Basic Notions: RDF

• Data is put on the Web using a formalism for describing Web Resources – RDF
  - data comes in the form of triples (subject-predicate-object)
  - ex: thisPresentation – hasAuthor - Milan
Basic Notions: Ontologies

- Ontologies provide a shared vocabulary
  - meaning of classes and properties is well defined, which allows to « understand » the data in exchange
- Ontologies contain general knowledge, valid for all data instances
  - they allow to infer additional facts from the available (scarce) data
- Ontology Web Language (OWL) is used to create them
  - several versions of OWL exist, with different expressive power (the greater the expressive power, the lower the inference possibilities)
Some motivations for publishing Linked Data

- data reuse and exchange
- Independent development of application that add value to public data
  - e.g. http://data.gov.uk application ecosystem
The Public Linked Data Cloud today
How to publish Linked Data
Publishing Linked Data

- The simplest way is to provide RDF files
  - for the same URI, we can provide RDF to machines, and normal Web pages to humans
- There is a way to hide RDF markup in Web pages
  - RDFa standard, used by Google, Facebook, ...
- Provide universal API – query Interfaces
  - SPARQL is a query language for data graphs – similar to SQL
  - it allows to select parts of the graph of our interest
Publishing Linked Data: SPARQL Endpoints

- [http://dbpedia.org/sparql](http://dbpedia.org/sparql)
Publishing Linked Data: SPARQL Endpoints

- It is possible to build graphic interfaces on top of SPARQL endpoints
  - [http://dbpedia.neofonie.de/browse/](http://dbpedia.neofonie.de/browse/)
  - easy to design faceted browsing

- Application that can work with one SPARQL endpoint can work with any other
  - provided that the application userstands the ontologies used in the endpoint
  - it is possible to discover new data sources using Sindice.com
Linked Data in Collaborative Work

THE SEM-SLATES APPROACH
Collaborating on Web 2.0

- Many Web 2.0 tools facilitate collaborative work
  - Easy collaborative creation of content
  - Commenting and other forms of interactions / feedback
  - Collaborative annotation and content organization: tagging
- SLATES paradigm
  - Search: text-based or tag-based
  - Links
  - Authoring (wikis,..)
  - Tags (user based annotation)
  - Extensions (suggestions of relevant content)
  - Signals (RSS, notifications, etc..)
Issues with SLATES systems

- Information fragmentation
  - information about one object is fragmented in several places
  - information exposed in various non-interoperable formats
  - difficult to merge, integrate and get a global view

- knowledge modeling and reuse
  - information exposed as plain text
  - answering complex queries is difficult

- Tagging issues
  - ambiguity (e.g., Paris – France, Paris Hilton)
  - heterogeneity (e.g., semanticWeb, semantic_web)
  - no structure/hierarchy links between tags
We generate Linked Data annotations for the content

<table>
<thead>
<tr>
<th></th>
<th>SLATES</th>
<th>SEM SLATES</th>
</tr>
</thead>
<tbody>
<tr>
<td>search</td>
<td>text-based</td>
<td>based on data annotations</td>
</tr>
<tr>
<td>links</td>
<td>between pages</td>
<td>between resources</td>
</tr>
<tr>
<td>authoring</td>
<td>documents</td>
<td>data and metadata</td>
</tr>
<tr>
<td>tags</td>
<td>textual keywords</td>
<td>semantic tagging (concepts)</td>
</tr>
<tr>
<td>Extension</td>
<td>hypertext navigation</td>
<td>RDF graph-based navigation</td>
</tr>
<tr>
<td>Signals</td>
<td>RSS feeds</td>
<td>Semantically-indexed RSS feeds</td>
</tr>
</tbody>
</table>
• The User Generated Content is Automatically Annotated with Semantics through software addons
  ○ metadata about social interactions and resulting documents can be automatically added

• Addons exist for major content publishing systems
  ○ Wordpress plugin exposes data using SIOC ontology
  ○ Drupal
  ○ Semantic Wikis, http://semantic-mediawiki.org/

• En Example in Microblogging
  ○ SMOB.me
What's on your mind?

(You have 122 characters left)

SMOB - #linkeddata

Current location

Galway, Ireland (seat of a first-order administrative division)

Interlinking

Links will be suggested while typing ... (space required after each #tag)

#linkeddata

- Via dbpedia
  - Linked Data (http://dbpedia.org/resource/Linked_Data)

- Via sindice
  - linkeddata (http://vanirsystems.com/danielsblog/tag/linkeddata/)
  - tag linkeddata (http://depth-first.com/articles/tag/linkeddata)
  - BibSonomy :: tag :: linkeddata (http://bibsonomy.org/tag/linkeddata)
  - linkeddata « Reallywow (http://blog.reallywow.com/archives/tag/linkeddata)
  - linkeddata - Semantic Dreamer (http://semanticdreamer.com/tag/linkeddata/)
  - Notices tagged with linkeddata - Identi.ca (http://identi.ca/tag/linkeddata)
  - » linkeddata Marc Hibbins (http://blog.marchibbins.com/category/linkeddata/)
  - linkeddata : Casual.info.in.a.bottle (http://www.dagoneye.it/blog/tag/linkeddata/)
  - linkeddata: Blogs, Photos, Videos and more on Technorati (http://technorati.com/tag/linkeddata)

Broadcast

- Twitter as terraces
- Ping Sindice

SMOB it!
Although most annotations are produced automatically, some user input is needed for disambiguation

- users might precisely choose concepts for tagging (Paris – France, or Paris Hilton)
- existing tools like Faviki, can facilitate this task
What exactly do you mean by **jaguar**?

- Jaguar
- Jaguar Cars
- Atari Jaguar
- Fender Jaguar

**Results from Wikipedia:**

The jaguar, *Panthera onca*, is a big cat, a feline in the *Panthera* genus. It is the only *Panthera* species found in the Americas. [Source: Wikipedia]

**Image via Wikipedia**

**Search the whole web?**

powered by: Zemanta and Google

**Save**

**Private**

**Close**
SemSLATES: New Possibilities

- **Answer complex search queries**
  - some examples
    - Find all papers about Nietzsche, written by someone employed by a European institution
    - Find people who created content about Picasso, who work in a European museum
  - SPARQL queries over data give more precise results than simple keyword-based searches

- **Integrate content from several sources**
  - different institutions may use their own tools, as long as they publish the output data in a unified way, in accordance to chosen vocabularies
SemSLATES: New Possibilities

- Possible to create mash-ups, adding data from other Linked Data sources
  - geographical data is the most common example, Geonames.org
- Create Faceted Interfaces
US Presidents

Here is the [Exhibit JSON data file](http://simile-widgets.org/exhibit/).
How To Adopt SemSLATES

- Identify the content to annotate and expose as Linked Data
- Identify ontologies that may to used to describe data
  - some ground already set by Europeana Data Model (EDM)
- Choose Collaborative Tools
  - install appropriate add-ons, or adopt semantic versions of collaborative software
  - if needed develop addons and adapt existing tools to use the chosen ontologies
- Design Interfaces on top of data
- Create Data – Use the Tools
Questions

mail@milstan.net