Semantic Web Motivating Example

Here’s a motivating example, adapted from a presentation by Ivan Herman:

- It introduces semantic web concepts
- and illustrates the benefits of representing your data using the semantic web techniques
- And motivates some of the semantic web technologies

We start with a book...

<table>
<thead>
<tr>
<th>ID</th>
<th>Author</th>
<th>Title</th>
<th>Publisher</th>
<th>Year</th>
</tr>
</thead>
</table>

A simplified bookstore data

<table>
<thead>
<tr>
<th>ID</th>
<th>Publisher's name</th>
<th>City</th>
</tr>
</thead>
<tbody>
<tr>
<td>id_qpr</td>
<td>Harper Collins</td>
<td>London</td>
</tr>
</tbody>
</table>
Export data as a set of relations

Notes on exporting the data

• Relations form a graph
  – Nodes refer to “real” data or some literal
  – We’ll defer dealing with the Graph representation
• Data export doesn’t necessarily mean physical conversion of the data
  – relations can be generated on-the-fly at query time
• All of the data need not be exported

Same book in French...

Bookstore data (dataset “F”)

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ID</td>
<td>Titre</td>
<td>Traducteur</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>ISBN 0-00-651409-X</td>
<td>$A11S</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Nom</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Ghosh, Amitav</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Besse, Christianne</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Export data as a set of relations

Start merging your data

Merging your data

Merging your data
Start making queries...

• User of data “F” can now ask about the title of the original
• This information is not in the dataset “F”...
• ...but can be retrieved by merging with dataset “A”!

However, more can be achieved...

• Maybe a:author & f:auteur should be the same
• But an automatic merge doesn’t know that!
• Add extra information to the merged data:
  – a:author same as f:auteur
  – both identify a “Person”
  – Where Person is a term that may have already been defined, e.g.:
    • A “Person” is uniquely identified by a full name, a homepage, facebook page, G+ page or email address
    • It can be used as a “category” for certain type of resources

Use this extra knowledge

This enables richer queries

• User of dataset “F” can now query:
  – “donnes-moi la page d’accueil de l’auteur de l’original”
  – 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
  – Inferring the consequences
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...
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!
• What is needed is a way to let machines decide when classes, properties and individuals are the same or different

This can be even more powerful
• 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

So where is the Semantic Web?
• The Semantic Web provides technologies to make such integration possible!
• Hopefully you get a full picture at the end of the tutorial…