Potentials of enriching the Web of Documents with Linked Data by generating RDFa markup

Benjamin Adrian

[1] Knowledge Management Department, DFKI
Kaiserslautern, Germany
[2] CS Department, University of Kaiserslautern
Kaiserslautern, Germany
`benjamin.adrian@dfki.de`

Abstract. Linked Data is out there. It consists of data about various topics in a range from human health care to Pop Music or product information. While on a web search users like Sarah, Pete, and Tom would like to see this data while reading and browsing the web of documents. The position of this paper figures out the potentials of enriching web pages with linked data content according to the specific information demand the user has in his current situation.

Tools that automatically enrich web pages with RDFa content from Linked Data knowledge bases are considered to be the next step to really use the web of data while browsing the web of documents.

1 Motivation

Please consider Sarah, a PHD student. Sarah wants to buy a laptop and has a budget of 600$. Sarah has a tight schedule, so her plan is to search online to quickly compare different offers regarding to product properties and existing reviews from customers. Fortunately, Sarah knows Linked Data and she was happy to see the Openlink Virtuoso Sponger that would give her access to Amazon.com data in RDF format. She knows Amazon to have a large knowledge base about products and vendors, their offers, and user generated ratings and experience reports about using these products. Sarah likes Amazon and its marketplace, but wants to give all online vendors a chance. Finally she wants the cheapest offer for the best laptop she can get with 600$. Therefore she wants a projection of Amazon's product data to products mentioned in web pages of online shops she found while searching Google products or Yahoo's Search Monkey.

Imagine Pete, a high school student. Pete is reading an exciting thriller on his iPad. Pete loves to have a more colorful imagination about concrete sets and locations where actions take place in. Therefore he installed a fancy App that uses data from DBpedia and LinkedGeoData to produce a mashup on Google map and Streetview consisting of scenery pictures, satellite images, and links to additional background information about heritages, famous buildings, battlegrounds, etc.

Tom is a young entrepreneur. He really knows about the power of social platforms like Twitter or Facebook. Tom knows many experts, friends and customers inside these platforms and he likes to know about their opinion about some new technologies and products Tom is reading about in blogs or other web pages. Tom installed a browser plugin that uses the RDF data published by Twitter and Facebook to get the latest tweets, comments and blog entries from twitter and Facebook from his contacts about topics mentioned in these web pages.

2 Position

The tools Sarah, Pete, and Tom use have one in common: they use RDF data published as Linked Data to enrich existing instance mentions (textual phrases that refer to existing instances in the RDF data) with RDFa markup. This markup explicates the reference between a phrase in text and the instance within a data set. The tools of Sarah, Pete, and Tom depend on different Linked Data sets but enrich the document content with additional information that:

– helps Sarah to face more product offers and reviews about a certain Laptop she found during her web search,
– helps Pete in stimulating his imagination while looking at real pictures and maps about the primeval forests near the small town Folks in Washington, USA which is the set of the latest book by Stephanie Meyer about vampires, Native Americans, and werewolves, Pete is currently reading.
– helps Tom in getting the opinions and comments of well regarded experts from his Twitter and Facebook account about topics he is reading about.

Behind the scenes the tools start with analyzing text passages in original web pages . . .

From Port Angeles I carried on towards Forks on highway 101.

. . . then use Linked Data to explicate mentions in text with RDFa markup . . .

From `<a about="dbpedia:Port_Angeles" property="foaf:name">`Port Angeles`</a>` I carried on towards `<a about="dbpedia:Folks%2C_Washington" property="foaf:name">`Forks`</a>` on highway 101.

. . . which is then used by applications to request more information . . .

```
dbpedia:Folks\%2C\_Washington  rdfs:label  "Folks" ;
   geo:lat         "47.950980"^xsd:double ;
   geo:long        "-124.384749"^xsd:double ;
   foaf:depiction  <http://upload.wikimedia.org/wikipedia/commons/1/14/Forks_WA.jpg>
   rdfs:seeAlso    <http://maps.google.de/maps?ll=47.951111,-124.384722&spn=0.25,0.25> .
```

. . . which is finally visualized in useful, inspiring, and interesting mashups.

Conclusion: We recommend to build more browser or proxy based RDFa generators that automatically enrich web pages with Linked Data that helps the user in his current situation. Tools such as Epiphany `http://projects.dfki.uni-kl.de/epiphany` might be a first step to a usable web of data.