Augmenting the Linked Data Cloud using Referers

Hannes Mühleisen and Anja Jentzsch
Freie Universität Berlin
What we would like to have

What we have

DBpedia
Linked Data needs links
(doh!)

- Typed links connect Linked Data entities.
  - Links can e.g. lead to background information or related entities
- Many tools are dependent on the presence of links and traverse them
  - Examples: ldspider, SQUIN, ...
Link Traversal? HTTP!

- HTTP Specification, Section 14.36:

“The Referer[sic] request-header field allows the client to specify, for the server's benefit, the address (URI) of the resource from which the Request-URI was obtained [...] The Referer request-header allows a server to generate lists of back-links to resources for interest, logging, optimized caching, etc.”

- Let’s generate back-links between Linked Data entities using Referers!
Links from Referers

1. From HTTP request, get the requested local resource URL and Referer URL
2. Validate “forward” link through dereferencing (prefer RDF)
3. Retrieve property description
4. Create back link or copy remote statement
Back Link Type

1. If “our” local resource is the triple subject in the remote document:
   (localResource, remotePredicate, remoteObject)

2. If the local resource is the triple object
   1. If the predicate triple is symmetric:
      (localResource, remotePredicate, remoteObject)
   2. If the predicate has inverse Property:
      (localResource, remotePredicateInverse, remoteObject)
   3. If none of the above? Copy triple:
      (remoteSubject, remotePredicate, localResource)
Example Log Entry

160.45.170.10 - - [20/Mar/2010:14:36:45 +0000]
"GET /person/heiner-stuckenschmidt HTTP/1.0" 303 9327
"http://harth.org/andreas/foaf" "Mozilla/5.0"

local resource URL

Referer URL
## Evaluation Results

<table>
<thead>
<tr>
<th></th>
<th>DBpedia</th>
<th>SWDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Entries</td>
<td>19,770,157</td>
<td>8,092,552</td>
</tr>
<tr>
<td>Referers found</td>
<td>1,328,595</td>
<td>533,188</td>
</tr>
<tr>
<td>Dereferenced:</td>
<td>3,255</td>
<td>6,146</td>
</tr>
<tr>
<td>Not found</td>
<td>2,229</td>
<td>4,821</td>
</tr>
<tr>
<td>Text match</td>
<td>431</td>
<td>1,168</td>
</tr>
<tr>
<td>Subject match</td>
<td>395</td>
<td>47</td>
</tr>
<tr>
<td>Object match</td>
<td>200</td>
<td>110</td>
</tr>
</tbody>
</table>
Generated Links

<table>
<thead>
<tr>
<th></th>
<th>DBpedia</th>
<th>SWDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generated Links</td>
<td>595</td>
<td>157</td>
</tr>
<tr>
<td>New Links</td>
<td>507</td>
<td>136</td>
</tr>
</tbody>
</table>
Thank You!

(and please have your tools send Referer headers!)

Results: http://tinyurl.com/ldow-backlinks