Policy Aware Content Reuse

Oshani Seneviratne

Social Web Incubator Group
April 14th, 2010
Outline

- Level of policy awareness of users
- Tools to enable policy awareness on the web
  - FlickrCC Validator
  - Semantic Clipboard
- Tools to enable policy awareness in social networks
  - Respect My Privacy
Outline

- Level of policy awareness of users
- Tools to enable policy awareness on the web
  - FlickrCC Validator
  - Semantic Clipboard
- Tools to enable policy awareness in social networks
  - Respect My Privacy
Content on the Web

• There is lot of content on the Web
  – 3.6 billion images on flickr
  – 20 hours of video uploaded every minute on YouTube
  – Over 1 trillion web pages on the web
  (source: http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html)

• Content reuse is important
  – Prevents redundant work
  – Promotes creativity
Licenses

- De-facto license for open content distribution
- Can be expressed in human-readable and machine-readable formats
  - CC supports very user friendly icons and license deed pages
  - CC licenses can be expressed in RDF (ccREL spec)
- Can be deployed on a range of media
- Large community
License

• There are four different categories (clauses)
  – Attribution (BY)
  – Share-Alike (SA)
  – Non-Commercial use (NC)
  – No-Derivatives (ND)

• These can be mixed in any combination to generate custom licenses.
  Attribution
  Attribution-NoDerivs
  Attribution-NonCommercial-NoDerivs
  Attribution-NonCommercial
  Attribution-NonCommercial-ShareAlike
  Attribution-ShareAlike
Most Search Engines support CC
Most Content Distribution Sites support CC
Generating a Creative Commons License

- CC offers a license chooser hosted at: http://creativecommons.org/choose
  - Generates a snippet of XHTML with RDFa
  - Includes `cc:AttributionName` & `cc:AttributionURL`
  - Can extend using `cc:morePermissions`
Anatomy of a CC License

Things that are conveyed in a license

- "John Doe"
- http://example.com
- http://example-source.com
- cc:attributionName
- cc:attributionURL
- cc:morePermissions
- http://example.com/content
- http://more-permissions.com
- dc:title
- dc:source
- dc:license
- "This work"
- http://creativecommons.org/licenses/by/3.0/us
Anatomy of a CC License

Human Readable Form

This work by John Doe is licensed under a Creative Commons Attribution 3.0 United States License. Based on a work at http://www.example-source.com. Permissions beyond the scope of this license may be available at http://morepermissions.com
This work by John Doe is licensed under a Creative Commons Attribution 3.0 United States License. Based on a work at http://www.example-source.com. Permissions beyond the scope of this license may be available at http://morepermissions.com.

You are free:
- to Share — to copy, distribute and transmit the work
- to Remix — to adapt the work

Permissions beyond the scope of this license are available at morepermissions.com.

Under the following conditions:
- Attribution — You must attribute this work to John Doe (with link).

With the understanding that:
- Waiver — Any of the above conditions can be waived if you get permission from the copyright holder.

“Live” box gives how the work should be attributed.
Are users aware of these licenses, and tools available for content reuse?
Are users aware of these licenses, and tools available for content reuse?

Apparently not!
We found 78% - 94% attribution license violations with Flickr images on the Web.
Type of content used: *Flickr images*

**Why Images?**
- Most common type of web content reused in blogs

**Why Flickr?**
- Flickr has over 100 million CC-licensed images
- Flickr API supports querying of CC license information
- Flickr images have a fixed image URI:
  
  http://farm{farm-id}.static.flickr.com/{server-id}/{id}_{secret}.jpg
Experiment

Samples - **Technorati blog indexer to generate the samples**

Why Technorati?
- Technorati blog indexer crawls and indexes over 60 million weblogs
- Possible to obtain weblogs that link to a particular URL
- The ‘authority rank’ for weblogs changes with time
Experiment

Type of license used - **Criteria for checking Attribution**

Checking one of these within reasonable scoping in the DOM where the image is embedded:

- Flickr username or full name (*cc:attributionName*)
- Flickr user profile URI (*cc:attributionURL*)

![Attribution Example 1](image1)

![Attribution Example 2](image2)
Results

Raw Violations = \( \frac{\text{Total # of non/mis-attributed images}}{\text{Total # of images}} \)

Sample 1 (67 Pages, 426 Images) - 78
Sample 2 (70 Pages, 341 Images) - 80
Sample 3 (70 Pages, 466 Images) - 94

Violations %
Results

Precision = \( \frac{\text{Total # of violations} - \text{Total # of images owned by the user}}{\text{Total # of violations}} \)

Sample 1 (67 Pages, 426 Images)
- Precision %: 17

Sample 2 (70 Pages, 341 Images)
- Precision %: 26

Sample 3 (70 Pages, 466 Images)
- Precision %: 53
Results

Adjusted Violations =
\[
\frac{\text{Total # of images that are not owned by the reuser and that are not attributed}}{\text{Total # of images that not are owned by the reuser}}
\]

Sample 1 (67 Pages, 426 Images)
Adjusted Violations % = 94

Sample 2 (70 Pages, 341 Images)
Adjusted Violations % = 93

Sample 3 (70 Pages, 466 Images)
Adjusted Violations % = 89
Results

http://dig.csail.mit.edu/2008/WSRI-Exchange/results
Outline

- Level of policy awareness of users
- Tools to enable policy awareness on the web
  - FlickrCC Validator
  - Semantic Clipboard
- Tools to enable policy awareness in social networks
  - Respect My Privacy
Flickr CC Attribution License Violations Validator

Creative Commons Attribution License Violations Validator for Flickr Images

Enter the URI of a page which has embedded Flickr images. The validator will indicate if you have not properly attributed the owner of those images (if the images are under a Creative Commons Attribution license). You can use the Attribution XHTML provided on the page to correct your errors.

URI:

Results for the page http://dig.csail.mit.edu/FlickrCC/validator.cgi

Found 2 attribution license violations for the following images.

By storon.gianni
Has License ☑️

Attribution XHTML:
<doc xmlns="http://creativecommons.org/ns#"
    about="http://farm4.static.flickr.com/3148/2793485524_e1df856f37.jpg">
    <qa id="0r.attributionsURL"
        href="http://www.flickr.com/people/storon.gianni/?storon.gianni"/>
</doc>

By opoetz
Has License ☑️

Attribution XHTML:
<doc xmlns="http://creativecommons.org/ns#"
    about="http://farm4.static.flickr.com/3428/5388416803_94e2920aaf.jpg">
    <qa id="0r.attributionsURL"
        href="http://www.flickr.com/people/opoetz/?opoeze"/>
</doc>

The owner/source of the following image was attributed, but use the proper XHTML provided.

By CornGao
Has License ☑️

Attribution XHTML:
<doc xmlns="http://creativecommons.org/ns#"
    about="http://farm2.static.flickr.com/2962/385463634_2cf2c93015.jpg">
    <qa id="0r.attributionsURL"
        href="http://farm2.static.flickr.com/2962/385463634_2cf2c93015.jpg"/>
</doc>
Limitations

• Validating images that are originally from Flickr, but are downloaded and used (images that do not have a URI from Flickr)

• Correctly validating CC licensed images from Flickr of which the rights do not belong to the uploader
Flickr CC Attribution License Violations Validator

Demo:

http://people.csail.mit.edu/oshani/talks/flickrcc_demo.mov

Try it out:

http://dig.csail.mit.edu/FlickrCC/validator.cgi
Outline

- Level of policy awareness of users
- Tools to enable policy awareness on the web
  - FlickrCC Validator
  - Semantic Clipboard
- Tools to enable policy awareness in social networks
  - Respect My Privacy
Semantic Clipboard

• Shows if an image can be copied or not (based on the license it is under)

• Can use this tool to see which images can be used for a particular purpose (for e.g. pick out the images that can be used for a commercial use)
Semantic Clipboard

• Copy an image with the license
  – Scrape the License RDFa
  – Construct the Attribution XHTML
  – Paste into any application
Limitations

- Images in which the license metadata is not expressed in RDFa
- Copying of media types other than images (but it can be easily extended to other types of media as long as the licenses are expressed in RDFa)
Semantic Clipboard

Demo:

http://people.csail.mit.edu/oshani/talks/semclip_demo.mov

Try it out:

Outline

- Level of policy awareness of users
- Tools to enable policy awareness on the web
  - FlickrCC Validator
  - Semantic Clipboard
- Tools to enable policy awareness in social networks
  - Respect My Privacy (Work by Ted Kang and Lalana Kagal)
Social Networks & Privacy

- Traditional Method of Access Control:
  - Build a big fence around private information and hope nothing gets out.
Social Networks & Privacy

- What happens when the data gets out by being:
  - Copied and reposted
  - Inference (for example project “gaydar”)
  - Policy changes

- The problem: What happens when the data gets out?
Social Networks & Privacy

- Is there an alternative to enforcing privacy policies through restricted access?
- Conform to existing policies by making users aware of the usage restrictions associated with the data
Social Networks & Privacy

・ “Respect My Privacy” is supported across three platforms:
  • Facebook
  • OpenSocial
  • Tabulator Semantic Web browser

・ These applications enable users to specify privacy preferences for their data and also display privacy-annotated data prominently enabling other users to easily recognize and conform to these preferences
RMP in Facebook/OpenSocial

- Five RMP restrictions:
  - no-commercial
  - no-depiction
  - no-employment
  - no-financial
  - no-medical
RMP application for Facebook:

RMP on the Semantic Web

- Transparency handled. But how about automated accountability mechanisms?
- Our solution: Move to decentralized social networks

- The FOAF Converter developed by Matthew Rowe:
  - Translates Facebook profile pages to Semantic Web data using the Friend of A Friend (FOAF) ontology
- The five RMP restrictions are represented as ontological data
- The \texttt{restricts} property allows restrictions to be attached to any resource
- Additional RMP restrictions are added to the converted FOAF data
Tabulator

- **Goal**: Integrate the Semantic Web directly into the browser
- Use panes for viewing specific types of data
  - Triple, rule, justification, social panes, microblogging etc.
- Navigate by clicking through links to various resources
- Edit resources in-place via SPARUL (SPARQL Update)
- Available as a Firefox extension
### Tabulator: Triple/Data Pane

**Tim Berners-Lee's FOAF file**

<table>
<thead>
<tr>
<th>W3C Standards and Technical Reports</th>
<th>W3C Standards and Technical Reports</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tim Berners-Lee's FOAF file</td>
<td>Tim Berners-Lee's FOAF file</td>
</tr>
<tr>
<td>type</td>
<td><a href="http://xmins.com/foaf/0.1/Person">http://xmins.com/foaf/0.1/Person</a></td>
</tr>
<tr>
<td>seeAlso</td>
<td><a href="http://www.grorg.org/dean/foaf.rdf">http://www.grorg.org/dean/foaf.rdf</a></td>
</tr>
<tr>
<td>homepage</td>
<td><a href="http://www.grorg.org/dean/">http://www.grorg.org/dean/</a></td>
</tr>
<tr>
<td>mbox</td>
<td><a href="mailto:dean@w3.org">mailto:dean@w3.org</a></td>
</tr>
<tr>
<td>name</td>
<td>Dean Jackson</td>
</tr>
<tr>
<td>mbox sha1sum</td>
<td>654f1278927b96ad21ccc86257e41a2d7b7d293</td>
</tr>
</tbody>
</table>

**Coralie Mercier**

| type                               | http://xmins.com/foaf/0.1/Person   |
| seeAlso                            | http://www.koali.net/foaf.rdf      |
| name                               | Coralie Mercier                    |
| mailbox                            | mailto:coralie@w3.org              |

**Dean Jackson**

| type                               | http://xmins.com/foaf/0.1/Person   |
| seeAlso                            | http://www.grorg.org/dean/foaf.rdf |
| homepage                           | http://www.grorg.org/dean/        |
| mbox                               | mailto:dean@w3.org                |
| mbox sha1sum                       | 654f1278927b96ad21ccc86257e41a2d7b7d293 |

**Edd Dumbill**

| type                               | http://xmins.com/foaf/0.1/Person   |
| seeAlso                            | http://heddiery.com/edd/foaf.rdf   |
| homepage                           | http://heddiery.com/edd/           |
| mbox                               | mailto:edd@usefulinc.com           |
| mbox sha1sum                       | 654f1278927b96ad21ccc86257e41a2d7b7d293 |
| name                               | Edd Dumbill                        |
| nick                               | edd                                |

**Henry Story**

| name                               | Henry Story                        |
| member                             | http://fbpedia.org/resource        |
| Ell Lilly and Company              | /Ell Lilly and Company             |
| name                               | Susie Stephens                     |

**James Martin**

| type                               | http://xmins.com/foaf/0.1/Person   |
| name                               | James Martin                       |
# Tabulator: Social Pane

**Acquaintances**

- Aaron Swartz
- Amy van der Hiel
- Art Barstow
- Benjamin Joffe
- Charles McCathie Nevile
- Damian Steer
- Dan Connolly
- Dave Beckett
- Dean Jackson
- Edd Dumbill
- Eddie Lopez
- Eric Miller
- Eva Méndez
- Gregory J. Rosmaita
- Jan Grant
- Jim Loy
- Joe Brickley

---

**Basic Information**

- danbri homepage
- my.opera homepage
- danbri.livejournal weblog
- my.opera weblog

**OpenID**

http://danbri.livejournal.com/

**Nickname**

danbri
danbri

---

Tip: Do you have web ID?

[Make or set a Web ID](http://foaf.qdos.com/reverse/?path=http://danbri.org/foaf.rdf#danbri)

RMP in the Tabulator Social Pane

- Allows users to attach RMP restrictions to their FOAF files, and displays the accompanying RMP icon.
Privacy/License Highlighter in Tabulator
References


Questions?

oshani@mit.edu