Leveraging the Linked Data Principles for Electronic Communications

Antoine Zimmermann

22nd August, 2011

Electronic Communication is Ubiquitous

- email
- instant messaging
- discussion boards
- blogs
- microblogging
- social networks
- wiki talk pages
- comments (on virtually anything)
- SMSs
- VoIP

Electronic Communication is Ubiquitous

- email
- instant messaging
- discussion boards
- blogs
- microblogging
- social networks
- wiki talk pages
- comments (on virtually anything)
- SMSs
- VoIP

Outline

1 Linked Data and Semantic Web

2 Potential Uses

3 Miscellaneous Issues

Linked Data Principles

Linked Data are based on 4 basic principles

- use URIs as names for things;

  `name:me@az.net`

- use HTTP URIs so that one can look them up;

  `http://zimmer.aprilfoolsreview.com/`

- provide useful description of the things when someone look up the URIs, using standards (RDF, SPARQL);
- provide links to other things by putting URIs from other domains in the descriptions served.

RDF as a data model

Use RDF as the underlying datamodel:

```
:email-123   :subject   "Re; Meeting next Monday" .
:email-123   :repliesTo  :email-345 .
:email-123   :author     :antoine .
:tweet-ABC   :retweets   :tweet-DEF .
:email-345   :forwards   :tweet-ABC .
:comment-X   :agreesWith :post-Y .
```

RDFS/OWL for schemas/ontologies

RDFS defines the terms and their relationships:

```
:email-123 rdf:type  :Email .
:Email     rdfs:subClassOf  :Message .
:Tweet     rdfs:subClassOf  :Message .
:agreesWith rdfs:subPropertyOf  :repliesTo .
:retweets  rdfs:domain   :Tweet .
```

OWL has richer constructs (boolean connectors, cardinality restrictions, etc.)

SPARQL to query data

SPARQL is a query language for RDF datasets:

Mails in reply to emails with a certain keyword:

```
SELECT ?email WHERE {
  ?email sioc:reply_of [ sioc:content ?c ] .
  FILTER regex(?c,"keyword")
}
```

Mails in a thread related to a certain topic:

```
SELECT ?email WHERE {
  ?email (sioc:reply_of)* ?m .
  ?m foaf:primaryTopic :someTopic .
}
```

Classifying emails with ontologies

Emails that are sent by members of my family are private emails:

```
:PrivateEmail rdfs:subClassOf  :Email ;
 [
  rdf:type  owl:Restriction ;
  owl:onProperty  mail:author ;
  owl:someValuesFrom  :Family
 ]  rdfs:subClassOf  :PrivateEmail .
```

Some information can be added by hand (tagging, typing, etc). A reasoner then enrich metadata with inferences.

Filter with CONSTRUCT queries

Emails that contain a link to IMDb are MovieEmails:

```
CONSTRUCT {?email rdf:type :MovieEmail}
WHERE {
  ?email sioc:content ?c .
  FILTER regex(?c,"http://www.imdb")
}
```

Integrate multiple data sources

Find messages about events occuring in a capital city:

```
SELECT ?email WHERE {
 SERVICE <http://localhost/mailData> {
   ?email foaf:primaryTopic ?event .
 }
 SERVICE event:endpoint {
   ?event :locatedIn ?place . }
 SERVICE dbpedia:endpoint { ?place rdf:type
dbp:Capital . }
}
```

List contacts ordered by number of messages exchanged, including tweets, comments on blogs, private messages on forums, instant messages.

Other uses

- Describe the argumentation flow using subproperties of `sioc:reply_of`: `:agrees`, `:disagrees`, `:supports`, etc.
- Integrate inforamtion from other desktop applications: calendar, project manager, office tools, contact book, programming environment (cf. the Semantic Desktop);
- treat social web artefacts as emails: forward a tweet, CC multiple blog author with a comment, etc;
- treat emails as social web artefacts: retweet an email, +1/like an email, etc.

Further issues

Casual users are not able to write SPARQL queries and ontologies:

- what functionalities must be offered to the users?
- how the interface should look like?
- can an integrated approach be more successful than Google Wave?
- how to deal with the heterogeneity inherent from the Web architecture (multiple points of views leading to multiple vocabularies, etc)?
- how to deal with privacy? WebId could be a solution
- what new forms of SPAM could occur?

Questions to the audience

- How email technologies should integrate/leverage online communication in the Social Web?
- Do you think Linked Data/Semantic Web technologies offer a good backbone?
- If not, why? What would be the alternative?