Semantic Web 2.0: Creating Social Semantic Information Spaces

John Breslin, Stefan Decker

{john.breslin, stefan.decker}@deri.org
http://sw.deri.org/~jbreslin/
http://www.stefandecker.org/

WWW2006 Tutorial
Edinburgh, 26th May 2006
0. Overview of this tutorial

Semantic Web 2.0: Creating Social Semantic Information Spaces
This tutorial will give an overview of current proposals in the Semantic Web area for adding semantics to emerging Web 2.0 applications and established communications media such as blogging, wikis, and bulletin boards.

We will also cover the usage of Semantic Web technologies for community portals.

We will discuss current standardisation activities as well as research prototypes.
Abstract (2)

- Additional topics to be covered include **semantic search** based on metadata and large scale data integration as well as **semantics in digital libraries**
- Finally, we will discuss and present **current approaches to realise** the ideas of Vannevar Bush and Doug Engelbart on distributed collaboration infrastructures, which we term **Social Semantic Information Spaces**
Aims and objectives of this tutorial

• **Aims:**
  – To teach you about applications of Semantic Web technologies to the areas of collaboration / communication systems, Web 2.0 and social software
  – To describe SW applications in areas such as:
    • semantics blogs
    • interconnecting community sites
    • semantic search
    • semantic wikis on the Web or desktop information spaces

• **Objectives:**
  – You will be able to apply Semantic Web technologies to various application areas in “Social Semantic Information Spaces”
Why is this topic relevant? (1)

• The Semantic Web is **increasingly aiming at applications areas**
• Web 2.0 applications such as **blogging** and **wikis** have become very popular and at the same time have created an interconnected information space (through the “blogosphere” and inter-wiki links)
• At the same time, these applications are experiencing **boundaries in terms of information dissemination and automation**, as they require increased levels of automation (i.e. more automated ways for information distribution)
Why is this topic relevant? (2)

• Quite a number of Semantic Web approaches have recently appeared to overcome the boundaries these application areas, e.g., Semantic Wikis, Semantic Desktops, etc.

• A recent Knowledge Web project meeting highlighted the importance of overlapping technologies between Web 2.0 and the Semantic Web
Your knowledge background

• If you have a background knowledge in the area of the Semantic Web:
  – This tutorial will help you to develop application knowledge in relation to social software and other widely-used related web technologies

• If you have application knowledge in web engineering or the development of systems such as wikis and blogs:
  – This tutorial will aid you in developing and creating ideas on how to increase the usability of social software and other web systems using Semantic Web technologies
Table of contents

1. State of the Art in Semantic Web
2. The Path Ahead for Social Semantic Information Spaces
3. From Blogging to Semantic Blogging
4. From Wikis to Semantic Wikis
5. Semantic Search
6. Semantics in Digital Libraries
7. Semantics in Community Portals
8. Realising the Memex and NLS: From the Desktop and Web to Social Semantic Information Spaces
1. State of the Art in Semantic Web

Semantic Web 2.0: Creating Social Semantic Information Spaces
What is the Semantic Web?

- “An extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.”
  
  Sir Tim Berners-Lee et al., Scientific American, 2001: tinyurl.com/i59p

- “…allowing the Web to reach its full potential…” with far-reaching consequences

- “The next generation of the Web”
Where are we in the “Semantic Web layer cake”?

You Are Here!
The current (syntactic / structural) Web

- HTTP protocol is used for accessing and exchanging web data
- HTML language is used for creating web pages
- Resources are identified by URLs/URIs
- **Untyped** hyperlinks are used for “weaving the Web” together
- This has built an exciting multimedia world for users
- But there is very little information for machines
Was the Web meant to be more?

In the original Web, Tim Berners-Lee had originally wanted something else as well:

- Objects related by well defined attributes
- Web of relationships amongst named objects, yielding unified information management tasks
- Add metadata describing both structure and content
Hence, the Semantic Web...

- The word “semantic” stands for “the meaning of”
- The semantics of something is the meaning of something
- The Semantic Web is a Web that is able to describe things in a way that computers can understand:
  - The Beatles were a popular band from Liverpool
  - John Lennon was a member of the Beatles
  - The record "Hey Jude" was recorded by the Beatles
- Sentences like these in red can be understood by people
- But how can they be understood by computers?
Describing things on the Semantic Web (1)

RDF (Resource Description Framework) is an open format markup language for describing information and resources, and is the fundamental data model for the Semantic Web.

Using RDF, we can describe relationships between things like:

- A is a part of B or
- Y is a member of Z

and their properties (size, weight, age, price...) in a machine-understandable format where each thing has a URI.
Describing things on the Semantic Web (2)

• Its **graph-based model** means that it is straightforward for computers to process RDF data
• Putting information into RDF files makes it possible for "scutters" or RDF crawlers to search, discover, pick up, collect, analyse and process information from the Web
A simple RDF example

• Statement:
  – “Ora Lassila is the creator of the resource (web page)
    http://www.w3.org/Home/Lassila”

• Structure:
  Resource (subject) http://www.w3.org/Home/Lassila
  Property (predicate) http://www.schema.org/#Creator
  Value (object) “Ora Lassila”

• Directed graph:
Simple RDF example shown in RDF/XML

- In the directed graphs, the arrows point from the subject to the object, and the text on the arrow is the predicate
- The ellipses are resources and the rectangles are literals or text strings
- We can also represent this graph model in RDF/XML:

```xml
<rdf:Description about="http://www.w3.org/Home/Lassila">
  <Creator>Ora Lassila</Creator>
</rdf:Description>
```
Expanding on the previous example

- To add properties to the “Creator”, point through an intermediate resource (the ellipses are resources and the rectangles are literals or text strings):

http://www.w3.org/Home/Lassila

Person://fi/654645635

Name: Ora Lassila

Email: lassila@w3.org
Expanded RDF example shown in RDF/XML

```xml
<rdf:Description
    about="http://www.w3.org/Home/Lassila">
    <Creator
        rdf:resource="Person://fi/654645635"/>
</rdf:Description>

<rdf:Description
    about="Person://fi/654645635">
    <Name>Ora Lassila</Name>
    <Email>lassila@w3.org</Email>
</rdf:Description>
```
Why does RDF make sense?

- A global environment needs a globally-agreed upon way to:
  
  1. **Name things**
  2. **Relate to things**

- **RDF:**
  
  - **Provides both these requirements**
  - **Is the least common denominator**
Can already describe lots of things semantically

- Geographic coordinates:
  - GEO
- Library books:
  - Dublin Core (DC)
- Online discussions:
  - SIOC
- People, social networks:
  - Friend-of-a-Friend (FOAF)
- Maybe even hormones!
  - GeneOnt
The power of the Semantic Web

- **Interoperability** and increased connectivity is possible through a commonality of expression
- Vocabularies can be combined and **used together**:  
  - e.g. a description of a book using Dublin Core metadata can be augmented with specifics about the book author using the Friend-of-a-Friend vocabulary
- Vocabularies can be **easily extended** (modules, etc.)
- **Intelligent search** with more granularity and relevance:  
  - e.g. a search can be personalised to an individual by making use of their identity and relationship information
The challenge for the Semantic Web

- The Semantic Web can’t work all by itself:
  - If it did it would be called the “Magic Web”
  - It will need some help to become a reality
- For example, it is not very likely that you will be able to sell your car just by putting your RDF file on the Web
- Need society-scale applications:
  - Consumers and processors of Semantic Web data
  - Semantic Web agents or services
  - More advanced collaborative applications that make real use of shared data and annotations
2. The Path Ahead for Social Semantic Information Spaces

Semantic Web 2.0: Creating Social Semantic Information Spaces
The path to Semantic Web 2.0

• The Semantic Web effort is mainly towards producing standards and recommendations that will **interlink applications**

• The Web 2.0 meme (next slide) is about **providing user applications**

• Not mutually exclusive:
  – http://www.oreillynet.com/xml/blog/2005/10/is_web_20_killing_the_semantic.html
  – With a little effort, many Web 2.0 applications can and do use Semantic Web technologies to great benefit
  – We will now discuss Web 2.0 and describe what happens when we combine it with the Semantic Web
What is Web 2.0?

• The term *Web 2.0* was made popular by Tim O’Reilly:

• [http://en.wikipedia.org/wiki/Web_2.0](http://en.wikipedia.org/wiki/Web_2.0)
  – “*Web 2.0 … has … come to refer to what some people describe as a second phase of architecture and application development for the World Wide Web.*”

• The Web where “ordinary” users can meet, collaborate, and share using whatever is newly popular on the Web (tagged content, social bookmarking, AJAX, etc.)

• Popular examples include:
  – Bebo, del.icio.us, digg, Flickr, Google Maps, Skype, Technorati, Wikipedia…
Web 2.0 and social software

• Web 2.0 focuses include:
  – The Web as a platform for social and collaborative exchange
  – Reusable community contributions
  – Subscriptions to information, news, data flows, services
  – Mass-publishing using web-based social software

• http://en.wikipedia.org/wiki/Social_Software
  – “Social Software lets people rendezvous, connect or collaborate by use of a computer network. It results in the creation of shared, interactive spaces…”

• Social software for communication and collaboration:
  – IM, IRC, Forums, Blogs, Wikis, Social Network Services, Social Bookmarks, MMOGs…
From Web 1.0 to 2.0 (updated from O’Reilly)

<table>
<thead>
<tr>
<th>Web 1.0</th>
<th>Web 2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Platforms</td>
<td>Google Services, AJAX, Flock</td>
</tr>
<tr>
<td>Web Pages</td>
<td>Blogs</td>
</tr>
<tr>
<td>Portals</td>
<td>Wikis</td>
</tr>
<tr>
<td>Encyclopediæ</td>
<td>Wikipedia</td>
</tr>
<tr>
<td>Talk</td>
<td>Skype, Asterisk</td>
</tr>
<tr>
<td>Knowledge</td>
<td>Tagging, Folksonomies</td>
</tr>
<tr>
<td>Referencing</td>
<td>Syndication</td>
</tr>
<tr>
<td>Content</td>
<td>BitTorrent, P2P</td>
</tr>
<tr>
<td>Events</td>
<td>Upcoming.org</td>
</tr>
</tbody>
</table>

Platforms: Netscape, Internet Explorer
Web Pages: Personal Websites
Portals: Content Management Systems
Encyclopediæ: Britannica Online
Talk: Netmeeting
Knowledge: Directories, Taxonomies
Referencing: Stickiness
Content: Akamai
Events: Evite
Elements of Web 2.0 (from Dion Hinchcliffe)
What are tagging and folksonomies?

- **Tag:**
  - A keyword which acts like a subject or category

- **Folksonomy:**
  - A collaboratively generated, open-ended labeling system that enables Internet users to categorise content using tags:
    - Web links and pages (e.g. del.icio.us) using “social bookmarking”
    - Online photographs (e.g. Flickr, Zooomr)
    - Events (e.g. Upcoming.org)
    - Blog entries, etc.

- **Tag cloud:**
  - A visual depiction of the tags used on a website:
    - Equivalent to a weighted list in the field of visual design
Folksonomies and the Semantic Web

- Folksonomies may hold the key to developing the Semantic Web
- Adding metadata can dramatically improve the precision (the percentage of relevant documents) in search engine retrieval lists
- Hard to persuade web authors to add metadata to their pages in a consistent, reliable way (high entry costs, time consuming):
  - Few web authors make use of the simple Dublin Core metadata system, even though the use of DC meta tags could increase their pages' prominence in search engine retrieval lists
- In contrast to top-down controlled vocabularies, folksonomies are a distributed classification system with low entry costs
- If folksonomy capabilities were built into web protocols, possible that the Semantic Web would develop more quickly…
Searching using folksonomies
Metaweb ≡ social semantic information spaces
From Web 1.0 to Semantic Web 2.0

<table>
<thead>
<tr>
<th>Web 1.0</th>
<th>Web 2.0</th>
<th>Semantic Web 2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal Websites</td>
<td>Blogs</td>
<td>Semantic Blogs</td>
</tr>
<tr>
<td>Content Management Systems</td>
<td>Wikis</td>
<td>Semantic Wikis</td>
</tr>
<tr>
<td>Altavista, Google</td>
<td>Google Personalised, DumbFind</td>
<td>Semantic Search</td>
</tr>
<tr>
<td>CiteSeer, Project Gutenberg</td>
<td>Google Scholar, Book Search</td>
<td>Semantic Digital Libraries</td>
</tr>
<tr>
<td>Message Boards</td>
<td>Community Portals</td>
<td>Semantic Forums and Community Portals</td>
</tr>
<tr>
<td>Buddy Lists, Address Books</td>
<td>Online Social Networks</td>
<td>Semantic Social Networks</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>Semantic Social Information Spaces</td>
</tr>
</tbody>
</table>
Example of how Web 2.0 is creating new synergies

- **Scuttle + Gregarius + Feedburner + Grazr = …**
  - [http://bonhom.ie/2006/04/what-weeks-delay-can-produce.html](http://bonhom.ie/2006/04/what-weeks-delay-can-produce.html)

- A hybrid that allows one to aggregate one’s favourite blogs or other content on a particular topic and then to annotate bookmarks to the most interesting content found.
Hypothetically adding semantics to the example

“A semantic social collaborative resource aggregator”:
- Okay, it needs a better name, like *scraggy* or something 😊
- Social network members specify their favourite content sources
- You and your friends specify any topics of interest
- You specify friends whose topic lists you value
- Metadata aggregator collects content from sites you and friends like (which may be human tagged, or could be auto-tagged)
- Highlights content that may be of interest to you or your friends
- If nothing of interest is currently available, content sources may have semantically-related sources in other communities for secondary content acquisition and highlighting
- You bookmark and tag the interesting content, and *share*!
1+1>2

- Semantic forums
- Semantic blogs
- Semantic wikis
- Semantic social nets
- Semantic desktop

Semantic Web + social software > sum of its parts
Social semantic information spaces: SW 2.0

Collaboration and Communication Tools
- Blogs, Forums, OSNs, Wikis

World Wide Web
- URIs, HTML, HTTP

Semantic Web
- RDFS, OWL, SPARQL

Social Connectivity

Web 2.0 and social software

Bringing the Web to its full potential

Making Semantic Web real.
3. From Blogging to Semantic Blogging

Semantic Web 2.0: Creating Social Semantic Information Spaces
A phenomenon for a new generation?

- Cincinnati Enquirer, October 2004

“Well, yes, we could read your blog... or you could just tell us about your school day.”
What are blogs?

- **Weblog, web log** or simply a **blog**
- A web application which contains periodic time-stamped posts on a common (usually open-access) webpage
- Individual diaries -> arms of political campaigns, media programs and corporations (e.g. the Google Blog)
- Posts are often shown in reverse chronological order
- Comments can be made by the public on some blogs
- Latest headlines, with hyperlinks and summaries, are **syndicated** using **RSS** or **Atom** formats (e.g. for reading favourite blogs with a feed **aggregator** or **reader**)
The state of the “blogosphere”

- Source: Technorati (March 2003 to January 2006)
Some statistics from Technorati

- The blogosphere is 60 times greater than it was only 3 years ago
- 75,000 weblogs are created daily
- A new blog is created every second
- 13.7 million bloggers or ~50% are still posting 3 months after their blog is created
- 10% of all blogs update at least weekly
- About 9% of new blogs are spam!
From websites to blogs to semantic blogs…

We will now discuss how personal websites have moved from ordinary blogs to semantic structured blogging platforms, using:

a) Syndication formats and blog tags

b) Structured input mechanisms

c) Semantic Web technologies
Syndication of blog content (1)

• Syndication is used for **publishing new content** regularly

• Content is provided from many blogs and news sites in a **common format** that can be reused by other websites and applications in a “syndication” process

  Rather than mass-spamming via e-mail, interested parties can subscribe to feeds to be notified about changes or updates to information (**self service**)!

• A common syndication format can have **many uses**, including connecting services together, “mashing” together of data, etc.
Syndication of blog content (2)

• More than just blog headline syndication, since RSS can be used for:
  – Newspaper articles (one of the original usages), library updates, recipes, shared calendars (RSSCalendar.com), podcasts, videos, job posts, weather reports, financial updates, bug reports, wiki page changes, new photo uploads, forum thread replies, etc.

• Syndication format for blogs is usually RSS (although some sites now use different syndication formats, e.g. Blogger.com uses Atom)
Blog aggregators and readers

- Syndicated content allows one to check multiple feeds on a regular basis using aggregators or feed readers:
  - Previously, semi-regular visits to bookmarked sites
  - Feeds of syndicated content can now be pulled into readers
  - Also, intelligent pushing of feeds (e.g. with “pingback”)
What is RSS?

• The most common syndication format(s)
• Acronyms:
  – “Really Simple Syndication”
  – “Rich Site Summary”
  – “RDF Site Summary”
• Eight “flavours”:
  – Not including Atom!
RSS 1.0

- RSS 1.0 is in RDF (preferred format for Semantic Web as it can be used in conjunction with other ontologies)
- Class “channel”:
  - Property “title”
  - Property “link”
  - Property “description”
  - Property “items” (rdf:Seq)
  - ...
- Class “item”:
  - Property “title”
  - Property “link”
  - Property “description”
  - ...

[Diagram showing timeline and comparison of IETF, RSS-DEV, Netscape, UserLand, and Dave Winer's license: Creative Commons]
What is Atom?

• Another syndication system
• Based on XML (not RDF), but efforts towards AtomOWL
• Emphasis has shifted from the format to the API

• Specification:
  – Constructs: content, people, dates and links
  – Elements: feeds, with entries
  – http://www.atomowl.org/
Linking blog posts by topic tags

• Blog posts are often categorised (e.g., “Scotland”, "Movies") by the post creator
• Those on similar topics can be grouped together, using:
  – Freetext tags or keywords
  – Hierarchical tree categories
• For example, Technorati tags or keywords:
  • Tags are category names, for people to categorise blog posts, photos, links, etc.
  • Technorati.com wants to build a “tagged” web
• Utilising SW technology, can create categories using the SKOS vocabulary:
  • http://www.wasab.dk/morten/blog/archives/2004/09/01/skos-output-from-wordpress
  • http://www.w3.org/2004/02/skos/
Tags propagate to RSS feeds

I'm currently enhancing the blogging section of our tutorial for the World Wide Web conference next week with more information on Structured Blogging, and having installed the SIOC WordPress plugin on a WordPress MU-powered site I run, I've realised that Structured Blogging isn't going to take off by just providing ...

I've already asked James Farmer about Structured Blogging and WP MU compatibility, because it currently doesn't work out of the box with WP MU (although this may change with a forthcoming WP MU 1.0 stable release). If anyone has managed to get this to work, please let us know because I believe that this could be a boon for the Structured Blogging causes.
Tags and related tags across blogs

9.5.2006

Ken McGuire Blog :: Creative Imagination

BBC Launch Web API (9.5.2006, 10:39)

The BBC have launched a new API aptly dubbed BBC Web API (bwa) and have created a series of widgets drawing on their TV Anytime database for Google and Yahoo services. (...) The purpose of this API is to allow people greater access to our content and information about our content. Two words sum up what this API [...] (more)

Planet Tags: internet, api, bbc, bwa, tv-anytime, web2.0

5.5.2006

Ken McGuire Blog :: Creative Imagination

Buy A Search Engine... on eBay (5.5.2006, 11:11)

DigForIt is on sale... on eBay no less. I’m not overly worried since I only turn up number 3 on their search rankings (drawing references from Google, MSN and Yahoo). Must really work on turning up the search engine heat on my own sites. Anyway, DigForIt is a meta search engine, coloring for searches on [...] (more)

Planet Tags: web 2.0, alx, digforit, bing, digforit.com, ebay, search+engine, web2.0

4.5.2006

Aehso’s Output

Web 2.0 or Star Wars Character? (4.5.2006, 11:47)

Via RedPost, this post is so true, it’s funny. Has the world run out of unbranded words or is it time to retoast the asylum from the new millennium? (more)

(Wow, I just searched for a link and I can’t find a link to Alan Cooper’s “The Lunatics are Running the Asylum” book anywhere. WTF? More people need to read that book.)

Planet Tags: internet, web2.0

Related Tags for web2.0
- + internet
- + software
- + web 2.0
- + web+2.0
- + irland
- + web2
- + humour
- + web20
- + ebay
- + structuring blogging
Towards structured blogging?

- In structured blogging, packages of structured data are becoming post components
- The virtue of blogs has been their simplicity
- At the moment, each blog post only needs one field for content, and maybe a title and URL
- Not everyone is served well by this lowest common denominator
- Therefore the “Structured Blogging” working group was established last year:
  - http://www.structuredblogging.org/
Structure-enhanced blog posts

- Sometimes you have a burning need for more structure, at least some of the time
- When you know a subject deeply, and your observations or analysis recur, you may be best served by filling in a form
- The form will have its own metadata and its own data model
- Uses:
  - People get to express themselves, and
  - Blogs start to interoperate with enterprise applications
Soccer coach example

• An after-game soccer report typically includes:
  – which teams played
  – where and when
  – officials, and
  – a list of game events:
    • who scored (and when and how)
    • who received penalties (when and for what), etc.

• Wouldn't it be handy for the coach's blogging tool to understand this structure, present an editing form, render the form in HTML to their blog, and render the post (including the form) to their RSS feed?
  – Great for the forthcoming World Cup!
Structured blogging using WordPress

<table>
<thead>
<tr>
<th>Post Title</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Restaurant name</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Food style/ethnicity</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Price range</th>
<th>From</th>
<th>To</th>
<th>Currency</th>
<th>per person, excl drinks</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Location</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Address 1</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Address 2</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>City</th>
<th>State</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Postcode</th>
<th>Country</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Official URL</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Map Link</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Chain</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Photo</th>
</tr>
</thead>
</table>
Integrating readers with structured blogging

• And in the future, news aggregators and news readers should be able to:
  – Auto-discover an unknown structure
  – Notify the user that a new structure is available
  – Learn the structure, including entry forms, pick list sources, rendering guidance, and default style sheet
  – Make it available when the blogger is ready to write
Past and future structured blogging

• Past:
  – Qlogger: http://www.qlogger.com/

• Future:
  – The Structured Blogging movement has previously focused on single-user blogging platforms:
    • Should target multi-user blogging platforms like WPMU, B2Evolution or Drupal for more exposure
    • http://www.johnbreslin.com/blog/2006/05/17/how-to-make-structured-blogging-popular/
  – Use Semantic Web technologies to ontologise any available post structures for more linkage and reuse
Traditional blogging vs. semantic blogging

• Traditional blogging:
  – Publishing for the “eyeball Web”
  – Content is text, images, video (i.e. data targeted at people)

• Semantic blogging:
  – Enrich traditional blogs with semantic metadata
  – **Structural**: what relates to what and how?
  – **Content related**: what is this post about (e.g. a person, an event, etc.)?
  – Blogging targeted at machines as well as people
Why semantic blogging? (1)

- **Traditional blogging:**
  - Little or no **query possibilities** (except keyword and flat tags)
  - Little or no **reuse of data** (except textual copy and paste)
  - Little or no **linking** between posts (except simple hrefs and trackbacks)

- **Semantic blogging:**
  - Facilitates better **querying:**
    - More precise
    - Allows aggregation from various sources
  - **Better** reuse potential
  - **Richer** links
Why semantic blogging? (2)

- Users collect and create large amounts of **structured data** on their **desktops**
- This data is often **tied** to specific applications and **locked** within the user's computer
- Semantic blogging can **lift this data** into the Web
Releasing your data to the Web scenario

1. Ina writes Post
2. Ina’s Computer annotates Post
3. Ina’s Computer publishes Post
4. John reads Post
5. John’s Computer imports metadata
Positioning of the metadata

**Where** in the blog will the semantic metadata go?

- **Directly in the HTML?**
  - Validity problems, parsing, restrictions on use of RDF...

- **Put it in the newsfeed (RSS 1.0)?**
  - Would have to change blogging platforms, hard to get accepted
  - Newsfeed items disappear over time

- **Externally?**
  - Just add a link to HTML
  - *à la:

```html
<a type="application/rdf+xml"
href="http://bresl.in/foaf/foaf.rdf">John</a>
```
How is this related to structured blogging?

- Structured blogging is mainly based on “Microformats” (http://www.microformats.org/)
  - Therefore restricted to specific schemata, not open
  - Positioned inline on HTML page (and in feed)
  - Can be directly rendered using CSS
  - Structured and semantic blogging do not compete
    • Metadata can be added as RDF and using Microformats
  - Web-based implementations for generating structured blogging metadata
    • e.g. for WordPress and Movable Type
Creating the metadata (1)

**Structural metadata:**

- Relations within the blogosphere: what relates to what and how (replies, follow-ups or trackbacks, blogroll links and bookmarks, topics, etc.)?
- Closed domain, suggested vocabulary: **SIOC** (more later!)
- Plugins for blogging platforms, e.g. WordPress, Drupal
- Produced automatically from a blog’s database
Creating the metadata (2)

Content related metadata:

• What do blog posts talk about (e.g. books, individuals, meetings)?
• Keep open domain – so that can use any vocabulary / ontology (BibTeX, FOAF, iCal, ...)
• Web-based approach (á la structured blogging) - user fills in an HTML form
• Desktop-based approach (á la semiBlog) - user selects existing data on their computer, this gets converted into RDF
Creating a semantic blog post with semiBlog

Annotating a blog entry with an address book entry.

```xml
<foaf:Person rdf:ID="andreas">
  <foaf:homepage>
    http://sw.bla.org/~aharth/</foaf:homepage>
  <foaf:surname>Harth</foaf:surname>
  <foaf:firstName>Andreas</foaf:firstName>
  <!-- ... more properties ... -->
  < RDF:value >Andreas Harth</ RDF:value >
</ foaf:Person >
```
semiBlog architecture overview

External Applications
(Address Book, Calendar, etc.)

semiBlog Application
- Wrapper 1: AddressBook (Contacts)
- Wrapper 2: BibTeX (Publications)
- Wrapper 3: URL (Web Pages)
- Wrapper n: Other Data

semiBlog Core
- XML/RDF
- XSLT Processing
- Exporter 1: StaticFormat
- Exporter 2: MetaWeblog
- Exporter n: Other Format

Publishing
Using the metadata

Once a blog has semantic metadata, it can be...

• Used to **query**: “*Which blog posts talk about papers by Stefan Decker?*”

• Used to **browse** across blogs and other kinds of discussion methods:
  – We will talk about this in more detail in section 7: “Semantics in Community Portals”

• **Imported** into desktop applications of blog readers (AKA “The Web as a Clipboard“)
The Web as a clipboard using a semiBlog reader

- A user can import metadata from here into his/her own applications
Semantic blogging with Haystack (1)
Semantic blogging with Haystack (2)
HP semantic blogging demonstrator

- http://www.semanticblogging.org/
  - "Semantic view, semantic navigation and semantic query"
Semblog publishing platform
Semblog publishing platform (2)
More about metadata in the blogosphere

• Have described several applications and projects about semantic metadata for the blogosphere, e.g.:
  – semiBlog produces semantic metadata for the blogosphere.
  – …other approaches…
  – Karger and Quan, Haystack, 2004
  – Cayzer, Semantic Blogging, 2004
  – Takeda and Ohmukai, Semblog, 2004

• But what about a more general, higher-level look at the domain?
  – How can we describe metadata in the blogosphere, what are the general categories?
  – How does this effect implementation approaches?
  – Need some approach at conceptualising metadata in the blogosphere
Structural versus content-related

- **Structural metadata:**
  - **Relations** between blogs, posts, comments, etc.
  - More than just “A links to B“ - what kind of relationship?
  - …relations within the blogosphere

- **Content-related metadata:**
  - What is this post about, what is its topic?
  - **Anything** a blog author wishes to discuss
  - …relations between the blogosphere and everything else
Closed domain versus open domain

• Closed-domain metadata:
  – The domain is restricted to a certain set of real-world entities or concepts, e.g. blog structure or scientific publications.
  – Allows the definition of **one specific domain ontology** (e.g. SIOC)

• Open-domain metadata:
  – The domain is not restricted, e.g. as in blog content
  – Hard to define one all-embracing ontology, very unwieldy, hard to convince people to use it
  – Instead **divide into closed subdomains**, use small, vertical domain ontologies (e.g. FOAF, BibTeX, etc.)
Client side versus server side

- **Client-side metadata:**
  - Data to be used resides client-side
  - Implementation can best be realised client-side (e.g. harvesting desktop data with semiBlog)

- **Server-side metadata:**
  - Data to be used resides server-side
  - Implementation can best be realised server-side (e.g. harvest WordPress database tables with WordPress SIOC plugin)
Semantic blogging references

- Bojars, Breslin and Möller, “Using Semantics to Enhance the Blogging Experience”, ESWC 2006
- Karger and Quan, “What Would It Mean to Blog on the Semantic Web?”, ISWC 2004
- Ohmukai and Takeda, "Semblog: Personal Publishing Platform with RSS and FOAF”, FOAF Galway 2004
4. From Wikis to Semantic Wikis

Semantic Web 2.0: Creating Social Semantic Information Spaces
What are wikis? (1)

• A community-developed documentation project
• “A piece of server software that allows users to freely create and edit Web page content using any Web browser. Wiki supports hyperlinks and has a simple text syntax for creating new pages and crosslinks between internal pages on the fly.”
• Wiki comes from the Hawaiian word for quick
• In brief:
  – Interlinked websites
  – Collaborative editing
  – Simple syntax
  – e.g. Wikipedia.org

JohnGrisham

He is the author of PelicanBrief.
He lives in Mississippi.
He writes a book each year.
He is published by RandomHouse.
What are wikis? (2)

- A wiki (or WikiWikiWeb) is free software that was created by several designers to create a website that anyone is allowed to access, add to and edit.
- It relies on cooperation, checks and balances of its members, and a belief in sharing of ideas.
- There are several forms of wikis such as TWiki and the WikiWikiWeb because the designers allowed wikis to be open, allowing others to even change the original format.
- Wikis are being used in many ways, including Wikipedia.org, a highly used, online, free-access encyclopedia.
Some uses of wikis

- Wikis are being used for:
  - online encyclopaedias
  - free dictionaries
  - book repositories
  - event organisation
  - software development
  - writing research papers
  - project proposals
Entering information

- Anyone can edit an existing wiki article
- If an article does not exist on a particular topic, you can create it
- If someone messes up an article (deliberately or erroneously), there is a revision history so you can revert the contents

Revision history

(Latest | Earliest) View (previous 50) (next 50) (20 | 50 | 100 | 250 | 500).
To view a previous version, click the date for that version.
Legend: (cur) = difference with current version, (last) = difference with preceding version, m = minor
Problems with traditional wikis

- Structured access
- Information reuse

JohnGrisham
He is the author of PelicanBrief.
He lives in Mississippi.
He writes a book each year.
He is published by RandomHouse.

Structured access:
× Other books by JohnGrisham (navigation)
× All authors that live in Europe? (query)

Information reuse:
× The authors from RandomHouse (views)
× And what if I don't speak English? (translation)
Personal wikis

- Enabling personal information management
- Should be very simple, very fast, very usable
- “Note-taking on steroids”
- Examples include Tomboy, wikidPad, VoodooPad
- Notes, links, categories (to do lists, appointments)

- Popularity: simplicity, usability
What are semantic wikis?

- A wiki that has an underlying model of the knowledge described in its pages
  - Semantic wikis allow to capture or identify further information about the pages (metadata) and their relations
  - Knowledge model available in a formal language, so that machines can (at least partially) process and reason on it
  - A semantic wiki would be able to capture that an "apple" article is a "fruit" (through an inheritance relationship) and present you with further fruits when you look at apple
  - Some are used for personal knowledge management, others aimed at KM for communities

- http://www.semwiki.org/
From wikis to semantic wikis: the “sweet spot”
Semantic wiki prototypes

• At least 22 semantic wiki prototypes to date
• For a comprehensive list see:
  – wiki.ontoworld.org/index.php/Semantic_Wiki_State_Of_The_Art
  or www.cfcl.com/rdm/MBD/mbd_sem_wiki.php
• Semantic wikis are aiming at collaboration
  – Typically web based (two modes)
• Semantic personal wikis are aiming at personal information management
  – Typically a desktop application (one mode)
Platypus semantic wiki

A page per resource

Metadata is explicitly added separately from the text content (using N3 or RDF/XML)

Current resource as object

Current resource as subject
# Semantic MediaWiki

<table>
<thead>
<tr>
<th>What it does</th>
<th>What you type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assign the value 1,234,567 to the attribute &quot;example.&quot;</td>
<td>assign the value [[example:=1,234,567]] to the attribute &quot;example.&quot;</td>
</tr>
<tr>
<td>Assign a value of about a million, but showing a different text in the article.</td>
<td>assign a value of [[example:=999,331</td>
</tr>
<tr>
<td>John has the email address <a href="mailto:john@mailinator.com">john@mailinator.com</a>.</td>
<td>John has the [[email address:=<a href="mailto:john@mailinator.com">john@mailinator.com</a>]].</td>
</tr>
<tr>
<td>Giving the type in an attributes article:</td>
<td>Giving the type in an attributes article:</td>
</tr>
<tr>
<td>This attribute is an integer number.</td>
<td>This attribute is an [[has type::Type:Integer</td>
</tr>
<tr>
<td>Combining MediaWiki markup with attribute values:</td>
<td>Combining MediaWiki markup with attribute values:</td>
</tr>
<tr>
<td>Johns email address is <a href="mailto:john@mailinator.com">john@mailinator.com</a></td>
<td>Johns email address is [[email:=<a href="mailto:john@mailinator.com">john@mailinator.com</a></td>
</tr>
</tbody>
</table>

<ask>[[Category:Actor]] [[height:=>6 ft]] [[height:=<7 ft]]</ask>
SemperWiki semantic personal wiki

- **Annotation primitives:**
  - Page: CamelCase, absolute: http://example.org, qname: dc:title
  - Literal: “...”
  - Annotation: *predicate object*
  - Query: *subject predicate object*

- **Advanced access:**
  - Intelligent navigation
  - Query

- **Data reuse:**
  - Structured information
  - Views
Content and structural metadata in semantic wikis
Information reuse

You can make views very easily. Just write a pattern of things you are looking for; a question mark denotes a joker or wildcard, for example:

?s dc:topic "todo"

shows all pages that have topic "todo", in the result section below.

You can write as many queries as you want, one at a line. The results are continuously up-to-date, and the query is saved together with the page: the next you come here, you will see all the up-to-date todo items.

dc:creator EyalOren

results of query. any page with dc:topic is todo

<table>
<thead>
<tr>
<th>subject</th>
<th>predicate</th>
<th>object</th>
</tr>
</thead>
<tbody>
<tr>
<td>FixBike</td>
<td>dc:topic</td>
<td>todo</td>
</tr>
<tr>
<td>ShoppingList</td>
<td>dc:topic</td>
<td>todo</td>
</tr>
</tbody>
</table>
Future work for (personal) semantic wikis

• Make them collaborative, through P2P or shared storage mechanisms:
  – Desktop front-end
  – Shared back-end
• Allow sophisticated annotations:
  – Blank nodes
  – Compound statements
• Integrate with the desktop:
  – Drag and drop desktop items
  – Annotate these items
Semantic wiki references

• Aumueller, “Semantic Authoring and Retrieval within a Wiki”, ESWC 2005
• Tazolli et al., “Towards a Semantic Wiki Wiki Web”, ISWC 2004 Poster
• Muljadi and Takeda, “Semantic Wiki as an Integrated Content and Metadata Management System”, ISWC 2005
5. Semantic Search

Semantic Web 2.0: Creating Social Semantic Information Spaces
Towards a Semantic Web search engine

• Currently, Google searches mainly plain text

• Need integrated, conceptual query answering over various sources and kinds of data:
  – semi-structured data (RDF, actual SW data)
  – unstructured data (i.e. human language text)
  – structured data (i.e. databases)

• Goal to provide answers instead of document lists (or both)
Querying semi-structured data

- Data and schema are represented in same data model
- RDF = graph topology + (string) literals
- Queries can be posed without or with only partial knowledge of schema
- Queries are precise with specified semantics (as long as you disallow keyword-based queries)
- High precision of results (1.0), but possibly low recall:
  - Don’t get expected results if, for example, you query for `foaf:Person` with `foaf:name` “John Breslin” since “John G. Breslin” won’t match
Query expansion for keyword-based searches

• **Latent semantic indexing:**
  – Provides a means to measure distance between terms
  – Statistical method

• **Query reformulation:**
  – Methods to relax queries to get higher recall for “imprecise” keywords (homonyms, synonyms…) based on e.g. WordNet
Searching over RDF

• RDF consists of the graph structure and literals
• Enable keyword search over content of string literals, and combine these with structured queries

• Various RDF data stores supporting various RDF query languages:
  – Stores: Jena, Kowari, Redland, Sesame, YARS, etc.
  – Query languages: N3QL, RDQL, RQL, SPARQL, Triple, etc.
Crawling Semantic Web data

• Where do we get the semistructured data from?
  – MySQL databases (via wrappers)
  – GRDDL
  – Microformats
  – RDF files
• Crawl RDF files linked via rdfs:seeAlso
• Get HTML pages and apply GRDDL to parse out semistructured data
• Use GRDDL to parse microformat tags
rdfs:seeAlso link graph
Semantic search references

• Anyanwu, Maduko and Sheth, “Semantic Querying: SemRank”, WWW2005
• Harth, Kruk and Decker, “Graphical Representation of RDF Queries”, WWW2006 Poster
• Deerwester et al., “Latent Semantic Indexing”, JSIS 1990
• Guha et al., “Semantic Search”, WWW2003
• Harth and Decker, “Optimized Index Structures for Querying RDF from the Web”, LAWeb 2005
• Ding, Finin et al., “Swoogle”, CIKM 2004
6. Semantics in Digital Libraries

Semantic Web 2.0: Creating Social Semantic Information Spaces
Semantics in digital libraries

- Opportunities for applying semantics:
  - Search
  - User interface
  - Metadata catalog
  - Interface
Semantic digital library prototypes

• D-SPACE with Simile:
  – Collaboration with MIT Libraries, CSAIL, W3C and HP Labs initially

• JeromeDL semantic digital library:
  – Joint project between DERI Galway, DERI Innsbruck and the Polytechnic University of Gdansk

• BRICKS prototype:
  – Integrated project funded by EU FP6 to create a digital library management system for cultural heritage
Simile set of tools

• *How will scholars find relevant content on the Web?*
• SIMILE tools:
  – **Longwell**: Web-based faceted browser for RDF metadata
  – **Piggy Bank**: Firefox extension for desktop metadata management (similar to EndNote)
  – **Solvent**: Scripting language for HTML web page scraping into RDF
  – **Semantic Bank**: Tool for “publishing” Piggy Bank collections to a group or to the world
  – **Welkin**: RDF viewer, data exploration tool
  – **Gadget**: Command line XML inspector (RDF conversion aid)
Longwell demonstration
Step 1: Getting started

Facets for browsing

787 OCW resources
2384 ARTstor resources
3 OCW courses

Let’s start by clicking here

(Scenario by Stephen J. Garland & Mick Bass)
Step 2: Exploring one collection

Let’s click to focus on an early abstract artist

137 images on island
Step 3: Pausing to think

Narrow focus raises questions:
What else did Gorky do?
Who was doing similar work?

Let’s click to find out more about Gorky by removing a restriction
Step 4: Redirecting the search

Two islands in view now, linked by Gorky bridge

Let’s click to cross bridge
Step 5: Exploring a second collection
Semantic digital library technologies and research

- JeromeDL – e-library with semantics
  - A digital library based on the Semantic Web
  - Conforms to librarian standards (like MARC21)
  - Semantic query expansion and ontology based navigation
- FOAFRealm – identity management
  - Can define policies based on social networking information
  - Access rights delegation, social semantic collaborative filtering
- MarcOnt – semantic bibliographic description initiative
  - Bibliographic ontology compatible with MARC21, BibTeX, DC
  - MarcOnt portal for collaborative ontology lifecycle management
  - MarcOnt ontology mediation service
- HyperCuP - lightweight peer to peer implementation
  - Efficient broadcast algorithm
  - Domain-based overlay networks
Information management in JeromeDL
IR architecture in JeromeDL

![Diagram showing IR architecture in JeromeDL](image.png)
What is social semantic collaborative filtering?

- **Goal:**
  - To enhance individual bookmarks with shared knowledge within a community
- **Users annotate catalogues of bookmarks with semantic information taken from DMOZ or WordNet vocabularies**
- **Catalogs can include (transclusion) friend's catalogues**
- **Access to catalogues can be restricted with social networking-based polices**
- **SSCF delivers:**
  - Community-oriented, semantically-rich taxonomies
  - Information about a user's interest
  - Flows of expertise from the domain expert
Example of social semantic collaborative filtering
What is BRICKS?

• The BRICKS project is:
  – “Building Resources for Integrated Cultural Knowledge Services”
  – An integrated project in the sixth EU framework programme
  – Aiming at establishing the organisational and technological foundations of a distributed digital library system
  – Also aiming to build an open scalable infrastructure
  – Going to develop value-added application services

• Infrastructure requirements:
  – Open and distributed (P2P)
  – Component-based software architecture
  – Expandability, scalability, availability, interoperability
BRICKS and Semantic Web technologies

• In BRICKS, metadata repositories are:
  – Responsible for managing cultural assets
  – Serving as access points for search and discovery services

• Integrating metadata from existing systems into a BRICKS metadata repository means dealing with:
  – Heterogeneous metadata schemas
  – Heterogeneous systems

• The BRICKS approach is to use:
  – RDF for handling metadata internally
  – OWL to model the semantics of metadata
Search and discovery in BRICKS

BRICKS is providing three types of search mechanisms on the available (RDF) metadata:

1. Simple (fulltext) search
   • As used on Google

2. Advanced search
   • Field-value search with various operators and boolean combinations

3. Ontology-based search
   • Like advanced search, but with inference support
   • Problem:
     – Many institutions tend to provide metadata only in simple unqualified Dublin Core, so no use for ontology-based search
     – Thesaurus-based search (fourth type for the future)
Semantic digital library references

- Kruk, Synak and Zimmermann, “MarcOnt - Integration Ontology for Bibliographic Description Formats”, DC 2005
- Schlosser, Sintek, Decker and Nejdl, “Ontology-Based Search and Broadcast in HyperCuP”, ISWC 2002
- BRICKS, http://www.brickscommunity.org/
7. Semantics in Community Portals

Semantic Web 2.0: Creating Social Semantic Information Spaces
What are online communities? (1)

• People form **online communities** by combining one-to-one (e.g. e-mail and instant messaging), one-to-many (web pages and blogs) and many-to-many (forums, wikis) forms of communication.

• And to recap from earlier…

• "**Social software** enables people to rendezvous, connect or collaborate through computer-mediated communication and to form **online communities**." - Wikipedia
What are online communities? (2)

Pre-Web and Web 1.0:

- BBS services
- Mailing lists
- USENET
- Web-based bulletin boards

Web 2.0:

- Multi-forum sites
- Online social networks
- Weblogs
- Wikis
Evolution of online community sites

Online community sites:
• Provide a valuable source of information
• May contain rich meta-information
• But are isolated from one another:
  – Many sites discussing complementary topics

Next steps:
• Connect sites together
• Add more value:
  – Let other sites know more about the *structure* and *contents*
  – Make more use of tagging and semantic metadata
Existing connections using RSS, Atom syndication

• First step towards connecting online community sites:
  – More visibility through aggregation and search

• Benefits:
  – Good tool support
  – Many consumers

• Shortcomings:
  – Little information about structure of the site or community
  – Feeds typically include only last five to 20 items
    • How can we access information about the whole site?
What is SIOC? (1)
What is SIOC? (2)

- Semantically-Interlinked Online Communities (SIOC)
- Connecting forums, posts from many types of online communities (blogs, forums, mailing lists, etc.)
- Interesting possibilities:
  - Distributed linked conversations
  - Decentralised discussion channels and communities
- “I […] think the concept is HOT” – Robert Douglass, Drupal Developer
- http://rdfs.org/sioc/
The main concepts in SIOC
How does it work?

- Web-Based Communities
- Blogs
- CMS
- Bulletin Boards
- SIOC Interfaces

- RDF Store
- Query / User Interface

- Legacy Data Wrappers
- USENET
- E-Mail

Find all communities:
- Containing the word "Broadband Gateway" and "Intranet" and created less than 3 months ago
- That were read at least 5 times

SIOC metadata (not only from blogs and forums, but also from other sources)
Assigning topic metadata

The **sioc:topic** (category) property is used to assign topics to SIOC primitives:
- Post, Forum, Site, etc.

Can use existing category hierarchies:

- **Upper level category systems:**
  - Open Directory Project, WordNet, OpenCYC

- **Site specific category systems:**
  - Define community site’s existing category hierarchy in SKOS
  - Provide mapping to the upper level category system

- "Weak" category systems are described using **sioc:subject**:
  - Tags, keywords, folksonomies

---

1 SKOS – Simple Knowledge Organisation System - http://www.w3.org/2004/02/skos/
How can SIOC data be created?

• Create SIOC export modules for popular open-source discussion systems

• Initial versions of SIOC metadata exporters created for:
  – Content management system (Drupal)
    • http://rdfs.org/sioc/drupal
  – Bulletin board system (phpBB) [in progress]
  – Blogging system (WordPress)
    • http://rdfs.org/sioc/wordpress
  – French blogging system (DotClear)
    • http://apassant.net/blog/2006/03/12/75-plugin-sioc-pour-dotclear

• Infecting the Web Infrastructure:
  – During next upgrade cycle gigabytes of community data become available
Sample SIOC export from WordPress

<sioc:Forum rdf:nodeID="f-1">
  <sioc:name>Main blog at SIOC weblog</sioc:name>
</sioc:Forum>

<sioc:User rdf:about="http://rdfs.org/blog/author/site-admin/">
  <sioc:name>site admin</sioc:name>
  <foaf:mbox rdf:resource="mailto:uldis.bojars@deri.org"/>
  <foaf:mbox_sha1sum>05f74ceba69414a8502399f085c032735553282</foaf:mbox_sha1sum>
  <sioc:description>The administrator of SIOC Weblog</sioc:description>
</sioc:User>

<sioc:Post rdf:about="http://rdfs.org/blog/2005/03/24/sioc-ontology/">
  <sioc:name>SIOC Ontology</sioc:name>
  <sioc:hasCreator rdf:resource="http://rdfs.org/blog/author/site-admin/"/>
  <sioc:created>2005-03-24T09:50:33Z</sioc:created>
  <sioc:content>SIOC (Semantically Interlinked Online Communities) is an ontology for describing discussion forums and posts on topic threads in online community sites.

For more information, consult SIOC Ontology page.</sioc:content>
  <sioc:topic rdf:resource="http://rdfs.org/blog/category/general/"/>
</sioc:Post>

3/23/2005

Subcat post

Filed under: Sub...
### Browsing SIOC (or other RDF) with Piggy Bank

#### My Piggy Bank

15-Dec-2007 02:58:53

<table>
<thead>
<tr>
<th>field</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>http</td>
<td><a href="http://rdfs.org/blog/2005/03/24/sioc-ontology/">http://rdfs.org/blog/2005/03/24/sioc-ontology/</a></td>
</tr>
<tr>
<td>content</td>
<td>SIOC (Semantically Interlinked Online Communities) is an ontology for describing discussion information, consult SIOC Ontology page.</td>
</tr>
<tr>
<td>created_at</td>
<td>24-Mar-2005 11:58:13</td>
</tr>
<tr>
<td>encoded</td>
<td>SIOC (Semantically Interlinked Online Communities) is an ontology for describing discussion information, consult SIOC Ontology page.</td>
</tr>
<tr>
<td>has_creator</td>
<td><a href="http://rdfs.org/blog/author/site-admin/">http://rdfs.org/blog/author/site-admin/</a></td>
</tr>
<tr>
<td>has_reply</td>
<td>comment-2, comment-3</td>
</tr>
<tr>
<td>link</td>
<td><a href="http://rdfs.org/blog/2005/03/24/sioc-ontology/">http://rdfs.org/blog/2005/03/24/sioc-ontology/</a></td>
</tr>
<tr>
<td>links_to</td>
<td><a href="http://rdfs.org/sioc/">http://rdfs.org/sioc/</a></td>
</tr>
<tr>
<td>seeAlso</td>
<td>wp-sioc.php?sioc_type=post&amp;sioc_id=5</td>
</tr>
<tr>
<td>title</td>
<td>SIOC Ontology</td>
</tr>
<tr>
<td>topic</td>
<td><a href="http://rdfs.org/blog/category/general/sioc/">http://rdfs.org/blog/category/general/sioc/</a>, <a href="http://rdfs.org/blog/category/general/sub-cat/">http://rdfs.org/blog/category/general/sub-cat/</a></td>
</tr>
<tr>
<td>type</td>
<td>Post</td>
</tr>
</tbody>
</table>
Finding new interlinks with SIOC

• Use SIOC metadata to infer new connections between posts, users, forums and community sites

• Connections:
  – Posts by the same user
  – Posts on the same topic
  – Posts by friends of a user (social network)
  – Interests of users subscribed to a particular community forum

• Re-use connections
  – Store inferred connections using property sioc:related_to
SIOC for semantic forums

• There is already lots of RSS and Atom metadata being produced and consumed for blogs:
  – Often more applicable to standalone sites, as the notion of “community” is not prominent: blogs are per-user and decentralised, with a sense of individual “ownership”
  – But lots of relevant information is missing, e.g. metadata on the original poster, details of post replies and their contributors

• In forums, the missing metadata is very important:
  – Who contributed to what parts of a community
  – “What topics have received the most replies from the greatest number of individuals?”

• Forums could also make use of functionality from the blogging community, e.g. trackbacks and tagging
Tagging forum content

- Tagging has only recently become a feature of forums, even though it should and could be used to interlink discussions on related topics:
  - http://www.phpbbhacks.com/download/6145
- The next step is to produce SIOC exporters for popular bulletin board systems such as InvisionBoard, phpBB and vBulletin:
- In the presence of tags (e.g. from the modules above) or categorised content (e.g. from Drupal taxonomies), SIOC subject or topic metadata can be produced for cross-site linkage
BoardTracker, a “Technorati” for forums
Argumentative discussion topics similar to IBIS
Semantic forums and community portal references

• Breslin, Harth, Bojars and Decker, “Towards Semantically-Interlinked Online Communities”, ESWC 2005
• Huynh, Mazzocchi and Karger, “Piggy Bank: Experience the Semantic Web Inside Your Web Browser”, ISWC 2005
• Reynolds, Shabajee and Cayzer, “Semantic Information Portals”, WWW2004 Poster
8. Realising the Memex and NLS: From the Desktop and Web to Social Semantic Information Spaces

Semantic Web 2.0: Creating Social Semantic Information Spaces
Realising social semantic information spaces

**Web/Desktop:** Help individuals in managing information on the Web/their PC

**Semantic:** Make content available to automated processing

**Social:** Enable exchange across individual boundaries

---

**Personal Semantic Web:** *a semantically enlarged intimate supplement to memory*

**Social protocols**

**and distributed search**

**Social semantic peers**

---

**Making Semantic Web real.**
Motivation for social semantic information spaces

• Current problems:
  – Low level communication, everything is just e-mail...

• Insufficient collaboration infrastructure:
  – High cost of setting up/maintaining
  – Difficult to support ad-hoc collaboration
Realising social semantic information spaces:  
The first society-scale semantic web application

Driven by today's needs, in the spirit of seminal visions

Inspired by sociological perspectives:

On group forming:
Viral communication (Reed)

On innovative IT-based interaction and feedback:
Social translucent systems
(Erickson and Kellogg)
Smart Mobs (Rheingold)

On network modeling and algorithms:
Social network research
Small world properties
Power law distribution
(Barabasi and Huberman)
Link-based authority algorithms, recommender algorithms (Perugini)

Today necessary technologies & communities exist:
Standardised metadata: Semantic Web
Scalable distributed infrastructure: P2P Computing
Knowledge articulation and interaction: Desktop Technology
Processing of unstructured and legacy information: NLP
Human centric information exchange: Online Social Networks

Challenge: Extension & merging of research streams
Science peers application

- The desktop and the network become one
- Document exchange and collaboration
- Collaborative ontology and metadata creation/sharing
Social semantic information space references

- Gnowsis, http://www.gnowsis.org/
Conclusion

Semantic (Web 2.0) => (Semantic Web) 2.0