12 Social Semantic Web

John G. Breslin[1,2] · *Alexandre Passant*[1] · *Denny Vrandečić*[3]
[1]Digital Enterprise Research Institute (DERI), National University of Ireland, Galway, Ireland
[2]School of Engineering and Informatics, National University of Ireland, Galway, Ireland
[3]Universitaet Karlsruhe, Karlsruhe, Germany

12.1 Scientific and Technical Overview ... 468
12.1.1 The Social Web ... 468
12.1.2 Issues with the Social Web ... 470
12.1.3 Bridging Social Web and Semantic Web Technologies ... 471
12.1.4 Ontologies for the Social Web ... 474

12.2 Example Applications ... 482
12.2.1 Semantic Blogging ... 482
12.2.2 Semantic Microblogging ... 486
12.2.3 Semantic Wikis ... 489
12.2.4 Semantic Social Bookmarking ... 493
12.2.5 Review Websites ... 496
12.2.6 Social Semantic Web Applications for Sharing Scientific Research ... 498

12.3 Future Issues ... 499
12.3.1 Searching the Social Semantic Web ... 499
12.3.2 Trust and Privacy on the Social Semantic Web ... 500
12.3.3 Integrating with the Social Semantic Desktop ... 501

12.4 Cross-References ... 502

John Domingue, Dieter Fensel & James A. Hendler (eds.), *Handbook of Semantic Web Technologies*,
DOI 10.1007/978-3-540-92913-0_12, © Springer-Verlag Berlin Heidelberg 2011

Abstract: The Social Web has captured the attention of millions of users as well as billions of dollars in investment and acquisition. As more social websites form around the connections between people and their objects of interest, and as these "object-centered networks" grow bigger and more diverse, more intuitive methods are needed for representing and navigating content both within and across social websites. Also, to better enable user access to multiple sites and ultimately to content-creation facilities on the Web, interoperability among social websites is required in terms of both the content objects and the person-to-person networks expressed on each site. Semantic Web representation mechanisms are ideally suited to describing people and the objects that link them together, recording and representing the heterogeneous ties involved. The Semantic Web is also a useful platform for performing operations on diverse, distributed person- and object-related data. In the other direction, object-centered networks and user-centric services for generating collaborative content can serve as rich data sources for Semantic Web applications. This chapter will give an overview of the "Social Semantic Web," where semantic technologies are being leveraged to overcome the aforementioned limitations in a variety of Social Web application areas.

12.1 Scientific and Technical Overview

12.1.1 The Social Web

A visible trend on the Web is the emergence of what is termed "Web 2.0" [97], a perceived second generation of Web-based communities and hosted services.

Although the term suggests a new version of the Web, it does not refer to an update of the World Wide Web's technical specifications or its implementation, but rather to new structures and abstractions that have emerged on top of the traditional Web. Web 2.0 technologies, as defined in [4] and exemplified by sites such as *Wikipedia* [151], *Delicious* [28], *Flickr* [37] and *HousingMaps* [59], augment the Web and allow for an easier distributed collaboration. It is notable that the term Web 2.0 was actually not introduced to refer to a vision, but to characterize the current state of the art in Web engineering [97]. These technologies are distinguished from classical Web technologies by various characteristic features:

- *Community*: Web 2.0 pages allow contributors to collaborate and share information easily. The emerging result takes advantage of the *wisdom of the crowds* and could not have been achieved by each individual contributor, be it a music database like *freedb* [44] or an event calendar like *upcoming* [147]. Each contributor gains more from the system than they put into it.
- *Mashups*: Certain services from different sites can be pulled together in order to experience the data in a novel and enhanced way. This covers a whole range of handcrafted solutions, ranging from the dynamic embedding of AdSense advertisements [2] to the

visualization of Craigslist's housing information [21] on Google Maps [50], as realised by HousingMaps [59].

- *AJAX*: While existing prior to Web 2.0, AJAX – Asynchronous JavaScript + XML – is probably the technological pillar of Web 2.0 and allows the creation of responsive user interfaces, and thus facilitated both of the other pillars: community pages with slick user interfaces that could reach much wider audiences, and mash-ups that incorporate data from different websites, thereby introducing asynchronous communication for more responsive pages.

Social networking sites (SNSs) such as *Facebook* (one of the world's most popular SNSs [33]), *Friendster* (an early SNS previously popular in the USA, now widely used in Asia [46]), orkut (Google's SNS, popular in India and Brazil [99]), *LinkedIn* (an SNS for professional relationships [70]) and *MySpace* (a music and youth-oriented service [88]) – where explicitly stated networks of friendship form a core part of the website – have become part of the daily lives of millions of users, and have generated huge amounts of investment since they began to appear around 2002. Since then, the popularity of these sites has grown hugely and continues to do so.

Web 2.0 content-sharing sites with social networking functionality such as *YouTube* (a video-sharing site [158]), Flickr (for sharing images [37]) and *Last.fm* (a music community site [69]) have enjoyed similar popularity. The basic features of a social networking site are profiles, friend listings, and commenting, often along with other features such as private messaging, discussion forums, blogging, and media uploading and sharing. In addition to SNSs, other forms of social websites include wikis, forums, and blogs. Some of these publish content in structured formats enabling them to be aggregated together.

A common property of Web 2.0 technologies is that they facilitate collaboration and sharing between users with low technical barriers although usually on single sites and with a limited range of information. In this chapter, the term "Social Web" will be used to refer to this collaborative and sharing aspect, a term that can be used to describe a subset of Web interactions that are highly social, conversational, and participatory. The Social Web describes the collaborative part of the Web 2.0.

The Social Web has applications on intranets as well as on the Internet. Within companies and other organizations, Enterprise 2.0 [75] applications (i.e., Web 2.0-type tools deployed within an intranet) are being used for knowledge management, collaboration, and communication between employees. Some companies are also trying to make social website users part of their IT "team," by allowing users to have access to some of their data and by bringing the results into their business processes [139]. Often the overly optimistic introduction of Social Web applications to a company suffers from a number of fallacies, like that Web 2.0 applications are often successful (mostly they are not), and that what works on the Web will work in an enterprise [31]. On the Web, applications of the Social Web are generally leisure oriented. People share pictures (Flickr [37]), videos (YouTube [158]), bookmarks (Delicious [28]), etc. However, serious usage of Web 2.0 is now emerging more and more often. Social networks of researchers such as the *Nature Networks* [89] as well as bibliography management services such as *Bibsonomy* [11] have appeared on the Web, so that the collaborative aspects of Web 2.0 can be used as a better means toward integrating discussions and collaboration in scientific communities.

12.1.2 Issues with the Social Web

The Social Web is allowing people to connect and communicate via the Internet, resulting in the creation of shared, interactive spaces for communities and collaboration. There is currently a large disconnect in the online social space. A limitation of current social websites is that they are isolated from one another like islands in the sea, acting as closed-world and independent data silos. Different social websites can contain complementary knowledge and topics – segmented parts of an answer or solution that a person may be looking for – but the people participating in one website do not have ready access to relevant information available from other places. For instance, someone looking for information about a particular product, for example, a new camera, may find lots of comments about it on *Twitter*, pictures on *Flickr*, and videos on *YouTube*, but cannot get a view of all these information weighted by their own social network. As more and more social websites, communities, and services come online, the lack of interoperability between them becomes obvious. The Social Web creates a set of single data silos that cannot interoperate with each other, where synergies are expensive to exploit, and where reuse and interlinking of data is difficult and cumbersome.

A major reason for many of these social networks to remain walled gardens is that a network can thus bind users to its website. Since a user cannot move from one social network to the other without losing the previously added and maintained information, accepted connections, and history, a social network can lock in its users by not providing interoperability. Metcalfe's Law states that *the value of a telecommunications network is proportional to the square of the number of connected users of the system* [152]. In their paper [57], Hendler and Golbeck make an analogy between this law and the Social Web, and in particular discuss how Semantic Web technologies can be used to create bridges between social data and therefore enhance the global value of the network. In this way, the more that people interact with each other, the more knowledge becomes interlinked on a global scale. But this increase in value is not necessarily reflected in the relative value of current social network providers against each other, that is, the relative value of a major social network provider like *Facebook* [33] compared to its competitors may not necessarily increase by opening up its data, although the global value of the network and the value for each user will increase. As well as making its pages more openly accessible, Facebook is also making data available using the Open Graph Protocol [93], although this may be for economic reasons (increased search and advertising potential) rather than for the benefit of users (enabling personal profile portability).

A more technical reason for the lack of interoperation is that for most Social Web applications, communities, and domains, there are still no common standards for knowledge and information exchange or interoperation available. RSS (Really Simple Syndication or RDF Site Syndication, depending on the version used), a format for indicating recently updated Web content such as blog entries, was the first step toward interoperability among social websites, but it has various limitations that make it difficult for it to be used efficiently in such an interoperability context.

Blogs, forums, wikis, and social networking sites all can contain vibrant active communities, but it is difficult to reuse and to identify common data across these sites. For example, *Wikipedia* contains a huge body of publicly accessible knowledge, but reuse of this knowledge outside of Wikipedia and incorporating it into other applications poses a significant challenge. As another example, a user may create content on several blogs, wikis, and forums, but one cannot identify this user's contribution across all the different types of social software sites.

On the Web, navigation of data across social websites can be a major challenge. Communities are often dispersed across numerous different sites and platforms. For example, a group of people interested in a particular topic may share photos on *Flickr*, bookmarks on *del.icio.us* and hold conversations on a discussion forum. Additionally, a single person may hold several separate online accounts, and have a different network of friends on each. The information existing on each of these websites is generally disconnected, lacking in exchangeable semantics, and is centrally controlled by a single organization. Individuals generally lack control or ownership of their own data. Social websites are becoming more prevalent and content is more distributed. This presents new challenges for navigating such data.

12.1.3 Bridging Social Web and Semantic Web Technologies

The Semantic Web aims to provide the tools that are necessary to define extensible and flexible standards for information exchange and interoperability.

A number of Semantic Web vocabularies have achieved wide deployment: successful examples include RSS 1.0 for the syndication of information (RSS 1.0 being an RDF format, contrary to other RSS versions [112]), FOAF (Friend of a Friend [16, 39]) for expressing personal profile and social networking information, and SIOC (Semantically Interlinked Online Communities [14, 121]) for interlinking communities and distributed conversations.

The Semantic Web effort is in an ideal position to make social websites interoperable by providing standards to support data interchange and interoperation between applications, enabling individuals and communities to participate in the creation of distributed interoperable information. The application of the Semantic Web to the Social Web is leading to the "Social Semantic Web," creating a network of interlinked and semantically rich knowledge, bringing together applications and social features of the Social Web with knowledge representation languages and formats from the Semantic Web (*Fig. 12.1*).

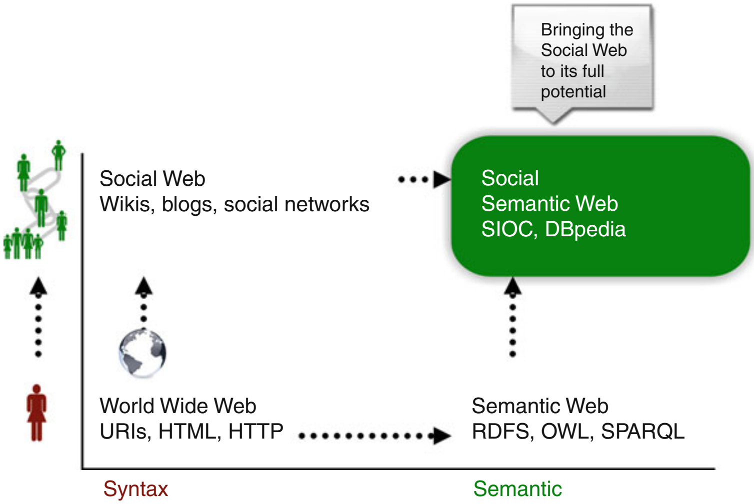


Fig. 12.1
The Social Semantic Web

This vision of the Web will consist of interlinked documents, data, and even applications created by the end users themselves as the result of various social interactions, and it is described using machine-readable formats so that it can be used for purposes that the current state of the Social Web cannot achieve without difficulty. As Tim Berners-Lee said in a 2005 podcast [63], Semantic Web technologies can support online communities even as "online communities [. . .] support Semantic Web data by being the sources of people voluntarily connecting things together." Therefore, integration between the Semantic Web and the Social Web is twofold:

- On the one hand, some efforts focus on using Semantic Web technologies to model social data. With ontologies such as FOAF and SIOC, social data can be represented using shared and common models, and therefore it becomes more easily interoperable and portable between applications.
- On the other hand, leveraging the wisdom of the crowds in Web 2.0 services can give a head start toward creating a large amount of Semantic Web data.

Additionally, the Social Web and social networking sites can contribute to the Semantic Web effort. Users of these sites often provide metadata in the form of annotations and tags on photos, ratings, blogroll links, etc. In this way, social networks and semantics can complement each other. Social website users are already creating extensive vocabularies and semantically rich annotations through folksonomies [79].

More than the sum of its parts: The combination of the Social Web and Semantic Web can lead to something greater than the sum of its parts: a Social Semantic Web where the islands of the Social Web can be interconnected with semantic technologies, and Semantic Web applications are enhanced with the wealth of knowledge inherent in user-generated content.

Because a consensus of community users is defining the meaning, these terms are serving as the objects around which those users form more tightly connected social networks. This goes hand-in-hand with solving the *chicken-and-egg* problem of the Semantic Web (i.e., you cannot create useful Semantic Web applications without the data to power them, and you cannot produce semantically rich data without the interesting applications themselves): since the Social Web contains such semantically rich content, interesting applications powered by Semantic Web technologies can be created immediately. The Social Semantic Web offers a number of possibilities in terms of increased automation and information dissemination that are not easily realizable with current social software applications:

- By providing a better interconnection of data, relevant information can be obtained from related social spaces (e.g., through social connections, inferred links, and other references).
- The Social Semantic Web would allow you to gather all your contributions and profiles across various sites ("subscribe to my brain," similar to an enhanced version of *FriendFeed* [45]), or to gather and filter content from your friend/colleague connections.
- These semantically enhanced social spaces allow the use of the Web as a clipboard to allow exchange between various collaborative applications (e.g., by allowing readers to drag structured information from wiki pages into other applications, geographic data about locations on a wiki page could be used to annotate information on an event or a travel review in a blog post one is writing).
- Such interoperable social software applications can help users to avoid having to repeatedly express several times over the same information if they belong to different social spaces.
- New and innovative ways for personalizing the content and creating intelligent user interfaces for acquiring content can be created based on the growing amount of semantic information available about users, their interests, and relationships to other entities.
- These social spaces will also allow the creation of social semantic mash-ups, combining information from distributed data sources together that can also be enhanced with semantic information, for example, to provide the geolocation of someone's friends in his social network who share similar interests with him.
- Fine-grained questions can be answered through such semantically enhanced social spaces, such as "show me all content by people both geographically and socially near to me on the topic of movies."
- The Social Semantic Web can make use of emergent semantics to extract more information from both the content and any other embedded metadata.

There have been initial approaches in collaborative application areas to incorporate semantics in these applications with the aim of adding more functionality and enhancing data exchange – semantic wikis, semantic blogs, and semantic social networks (discussed in Sect. 12.2). These approaches require closer linkages and cross-application demonstrators to create further semantic integration both between and across application areas (e.g., not just blog-to-blog connections, but also blog-to-wiki exchanges). A combination of such semantic functionality with existing grassroots efforts such as OpenID [95], a single sign-on mechanism, or OAuth [91], an authentication scheme, can bring the Social Web to another level. Not only will this lead to an increased number of enhanced applications, but an overall interconnected set of social software applications can be created using semantic technologies.

Some of the de facto standard ontologies that can be used for describing interlinked social spaces will now be described. It will follow with descriptions of a number of Social Semantic Web applications that have recently been developed. These use cases will serve as examples of how adding semantic information to social websites will enable richer applications to be built.

12.1.4 Ontologies for the Social Web

As discussed in Ontologies and the Semantic Web, ontologies provide a shared model for representing semantically rich information on the Semantic Web. In addition, by using standard representation languages, such as RDF(S)/OWL, these ontologies can be shared across services, so that data become interoperable between distributed applications. In the realm of the Social Semantic Web, ontologies can be then used to represent uniformly the different artifacts produced and shared in social websites: communities, people, documents, tags, etc. In this section, some of the most popular ontologies for the Social Web are described.

FOAF – Friend of a Friend: The Friend-of-a-Friend (FOAF) project [39] was started by Dan Brickley and Libby Miller in 2000 and defines a widely used vocabulary for describing people and the relationships between them, as well as the things that they create and do. It enables people to create machine-readable Web pages for people, groups, organizations, and other related concepts. The main classes in the FOAF vocabulary (*Fig. 12.2*, as illustrated by Dan Brickley [15]) include `foaf:Person` (for describing people), `foaf:OnlineAccount` (for detailing the online user accounts that they hold), and `foaf:Document` (for the documents that people create). Some of the most important properties are `foaf:knows` (used to create an acquaintance link), `foaf:mbox_sha1sum` (a hash over the eMail-address, often used as an identifier for a person and defined as an `owl:InverseFunctionalProperty` to allow smushing between instances), and `foaf:topic_interest` (used to point to resources representing an interest that a person may have).

foaf:knows is one of the most used FOAF properties: it acts as a simple way to create social networks through the addition of knows relationships for each individual that a person knows. For example, Bob may specify knows relationships for Alice and Caroline, and Damien may specify a knows relationship for Caroline and Eric; therefore Damien and Bob are connected indirectly via Caroline.

Anyone can create their own FOAF file describing themselves and their social network, using tools such as FOAF-a-matic [40] or FOAF Builder [41] from QDOS. In addition, the

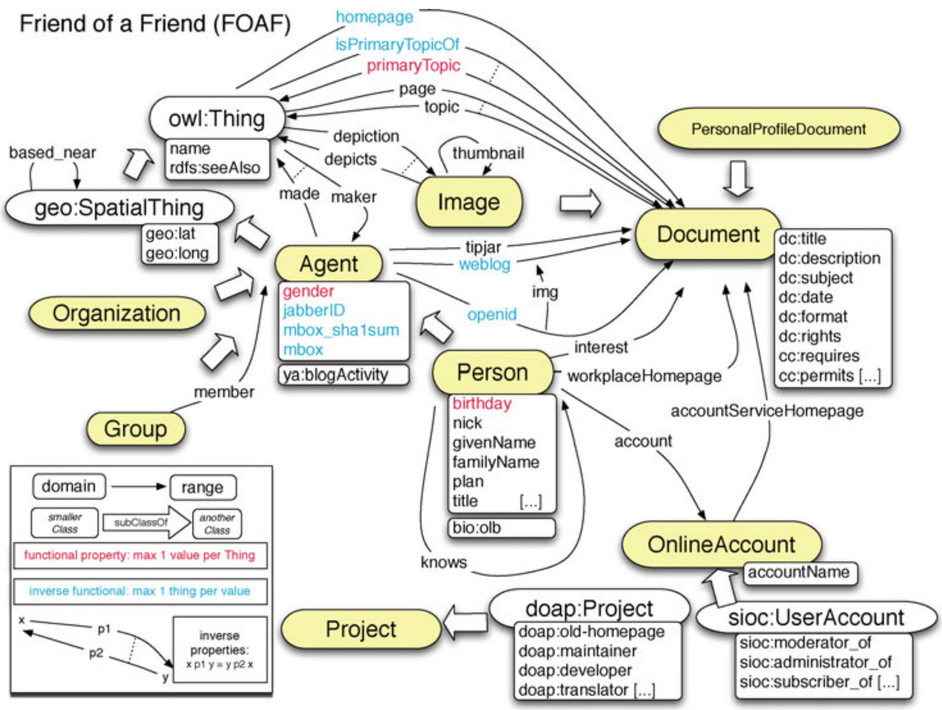


Fig. 12.2

Friend-of-a-friend terms: Updated from Dan Brickley's CC picture (http://www/flickr.com/photos/danbri/1855393361/)

information from multiple FOAF files can easily be combined to obtain a higher-level view of the network across various sources. This means that a group of people can articulate their social network without the need for a single centralized database, following the distributed principles used in the architecture of the Web.

FOAF can be integrated with any other Semantic Web vocabularies, such as SIOC (described below) and SKOS – Simple Knowledge Organization System [123]. Some prominent social networking services that expose data using FOAF include *Hi5* (a social networking site [58]), *LiveJournal* (a social networking and blogging community site [71]), *Identi.ca* (a microblogging site [61]), and *MyBlogLog* (an application that adds community features to blogs [85]). People can also create their own FOAF document and link to it from their homepage. Aggregations of FOAF data from many individual homepages are creating distributed social networks; this can in turn be connected to FOAF data from larger online social networking sites. Third-party exporters are also available for major social websites including *Flickr* [38, 103], *Twitter* [146], *MySpace*, and *Facebook* [34, 111].

The knowledge representation of a person and their friends would be achieved through a FOAF fragment similar to that below.

```
@prefix foaf: <http://xmlns.com/foaf/0.1.>.
<http://www.johnbreslin.com/foaf/foaf.rdf#me> a foaf:Person;
   foaf:name "John Breslin";
   foaf:mbox <mailto:john.breslin@deri.org>;
   foaf:homepage <http://www.johnbreslin.com/>;
   foaf:nick "Cloud";
   foaf:depiction <http://www.johnbreslin.com/images/foaf_photo.jpg>;
   foaf:topic_interest <http://dbpedia.org/resource/SIOC>;
   foaf:knows [
     a foaf:Person;
     foaf:name "Sheila Kinsella";
     foaf:mbox <mailto:sheila.kinsella@deri.org>
   ];
   foaf:knows [
     a foaf:Person;
     foaf:name "Smitashree Choudhury";
     foaf:mbox <mailto:smitashree.choudhury@deri.org> ].
```

hCard and XFN. hCard [55] is a microformat used to describe people, organizations, and contact details for both. It was devised by Tantek Çelik, and Brian.

Suda based on the vCard IETF format [148] for describing electronic business cards. Like FOAF, hCard can be used to define various properties relating to people, including "bday" (a person's birth date), "email," "nickname," and "photo," where these properties are embedded within XHTML attributes. The specification for hCard also incorporates the Geo microformat, which is used to identify the coordinates for a location or "adr" (address) described within an hCard. For example, the hCard for John Breslin is:

```
<div class="vcard">
<div class="fn">John Breslin</div>
<div class="nickname">Cloud</div>
<div class="org">National University of Ireland, Galway</div>
<div class="tel"> +35391492622</div>
<a class="url" href="http://johnbreslin.org/">
      http://johnbreslin.org/</a>
</div>
```

XFN (XHTML Friends Network) [154] is another social network–oriented microformat, developed by Tantek Çelik, Eric Meyer, and Matthew Mullenweg in 2003 just before the creation of the microformats community. XFN allows one to define relationships and relationship types between people, for example, "friend," "neighbor," "parent," "met," etc. XFN is also supported through the WordPress blogging platform: when adding a new blogroll link, one can use a form with checkboxes to specify additional metadata regarding the relationship between the blog owner and the person who is being linked to (which is then exposed as metadata embedded in the blog's resulting XHTML). For example, an XFN "colleague"-type link to Uldis Bojārs would be written as:

```
<a href="http://captsolo.net" rel="colleague">Uldis Bojārs</a>
```

When combined with XFN, hCard provides similar functionality to FOAF in terms of describing people and their social networks. The different types of person-to-person relationships available in XFN allow richer descriptions of social networks to be created as FOAF voluntary only has a "knows" relationship. However, FOAF can also be extended with richer relationship types via the XFN in RDF vocabulary [155] developed in 2008 by Richard Cyganiak or the RELATIONSHIP vocabulary [108], which includes a variety of terms including `siblingOf`, `wouldLikeToKnow`, and `employerOf`.

SIOC – Semantically Interlinked Online Communities: SIOC aims to interlink related online community content from platforms such as blogs, message boards, and other social websites, by providing a lightweight ontology to describe the structure of and activities in online communities, as well as providing a complete food chain for such data. In combination with the FOAF vocabulary for describing people and their friends, and the SKOS model for organizing knowledge, SIOC lets developers link discussion posts and content items to other related discussions and items, people (via their associated user accounts), and topics (using specific "tags," hierarchical categories, or concepts represented with URIs).

As discussions begin to move beyond simple text-based conversations to include audio and video content, SIOC is evolving to describe not only conventional discussion platforms but also new Web-based communication and content-sharing mechanisms. At the moment, a lot of the content being created on social websites (events, bookmarks, videos, etc.) is being commented on and annotated by others. If you consider such content items to be the starting point for a discussion about the content (similar to a text-based thread starter in a forum or blog), and if the content item being created is done so in a container linked to a user or topic, then SIOC is quite suitable for describing metadata about these content items as well.

Since disconnected social websites require ontologies for interoperation, and due to the fact that there are a lot of social data with inherent semantics contained in these sites, there is potential for high impact through the successful deployment of a SIOC ontology. The development of SIOC was also an interesting process to explore how an ontology can be developed for and bootstrapped on the Semantic Web. Feedback from the research and development community to the ontology development process was increased through the development of a W3C Member Submission for SIOC [122].

Many online communities still use mailing lists and message boards as their main communication mechanisms, and the SIOC initiative has also created a number of data producers for such systems in order to lift these communities to the Semantic Web. So far, SIOC has been adopted in a framework of about 60 applications or modules ranging from exporters for major Social Web platforms to applications in neuromedicine research, and has been deployed on hundreds of sites. Since the W3C Member Submission, SIOC has gained even more success and attention from interested parties. For example, SIOC was recently chosen by Yahoo! SearchMonkey [116] as a suitable reference vocabulary to describe the activities of online communities, along with FOAF to describe the social networking stack.

An interesting aspect of SIOC is that it goes beyond pure Social Web systems and can be used in other use cases involving the need to model social interaction within communities, either in corporate environments (where there is a parallel lack of integration between social software and other systems in enterprise intranets), or for argumentative discussions and scientific discourse representation (e.g., via the SWAN/SIOC [135] initiative).

The ontology consists of the SIOC Core ontology, an RDF-based schema consisting of 11 classes and 53 properties, and 5 ontology modules: SIOC Access, SIOC Argumentation, SIOC Services, SIOC Types, and SWAN/SIOC. The SIOC Core ontology defines the main concepts and properties required to describe information from online communities on the Semantic Web. The main terms in the SIOC Core ontology are shown in *Fig. 12.3*. The basic concepts in SIOC have been chosen to be as generic as possible, thereby allowing to describe many different kinds of user-generated content.

The SIOC Core ontology was originally created with the terms used to describe Web-based discussion areas such as blogs and message boards: namely Site, Forum, and Post. UserAccounts create Posts organized in Forums that are hosted on Sites. Posts have reply Posts, and Forums can be parents of other Forums. In parallel with the evolution of new types of social websites, these concepts became subclasses of higher-level concepts – data spaces (`sioc:Space`), containers (`sioc:Container`), and content items (`sioc:Item`) – which were added to SIOC as it evolved. These classes allow one to structure the information in online community sites and distinguish between different kinds of objects. Properties defined in SIOC allow one to describe relations between objects and the attributes of these objects. For example:

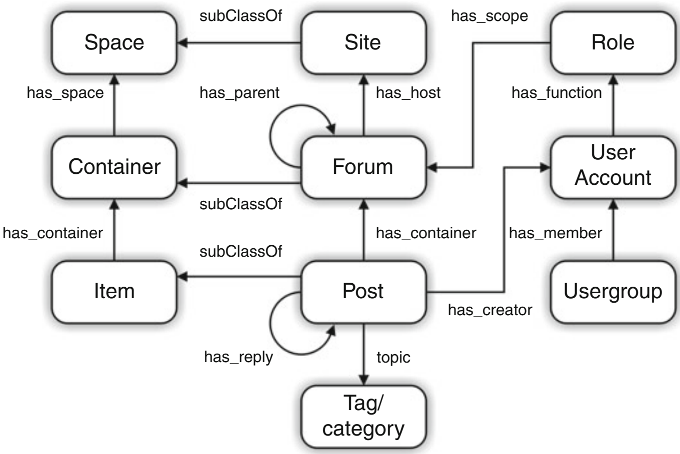


Fig. 12.3
The SIOC ontology

- The `sioc:has_reply` property links reply posts to content that they are replying to.
- `sioc:has_creator` and foaf:maker link user-generated content to additional information about its authors.
- The `sioc:topic` property points to a resource describing the topic of content items, for example, their categories and tags.

The high-level concepts `sioc:Space`, `sioc:Container`, and `sioc:Item` are at the top of the SIOC class hierarchy, and most of the other SIOC classes are subclasses of these. A data space (`sioc:Space`) is a place where data reside, such as a website, personal desktop, shared file space, etc. It can be the location for a set of Container(s) of content Item(s). Subclasses of Container can be used to further specify typed groupings of Item(s) in online communities. The class `sioc:Item` is a high-level concept for content items and is used for describing user-created content. Usually these high-level concepts are used as abstract classes, from which other SIOC classes can be derived. They are needed to ensure that SIOC can evolve and be applied to specific domain areas where definitions of the original SIOC classes such as `sioc:Post` or `sioc:Forum` can be too narrow. For example, an address book, which describes a collection of social and professional contacts, is a type of `sioc:Container` but it is not the same as a discussion forum.

A sample instance of SIOC metadata from a forum (message board) in Drupal is shown below in Turtle. This forum has a title, a taxonomy topic in Drupal, a description, and is the container for one or more posts. More information on the posts can be obtained from the referenced URI (e.g., if it has replies, related posts, who wrote it, etc.).

```
<http://sioc-project.org/forum/13> a sioc:Forum;
   dc:title “Developers Forum”;
   dc:description “Developers Forum at sioc-project.org”;
   sioc:link <http://sioc-project.org/forum/13>;
   sioc:topic <http://sioc-project.org/taxonomy/term/13>;
   sioc:container_of <http://sioc-project.org/node/185>.
<http://sioc-project.org/node/185> a sioc:Post;
   rdfs:label “Microformats and SIOC”;
   rdfs:seeAlso <http://sioc-project.org/sioc/node/185>.
```

A separate SIOC Types module defines more specific subclasses of the SIOC Core concepts, which can be used to describe the structure and various types of content of social websites. This module defines subtypes of SIOC objects needed for more precise representation of various elements of online community sites (e.g., `sioc_t:MessageBoard` is a subclass of sioc:Forum), and introduces new subclasses for describing different kinds of Social Web objects in SIOC. The module also points to existing ontologies suitable for describing full details of these objects (e.g., a `sioc_t:ReviewArea` may contain Review(s), described in detail using the Review Vocabulary). Examples of SIOC Core ontology classes and the corresponding SIOC Types module subclasses include: sioc:Container (AddressBook, AnnotationSet, AudioChannel, BookmarkFolder, Briefcase, EventCalendar, etc.); sioc:Forum: (ChatChannel, MailingList, MessageBoard, Weblog); and sioc:Post (BlogPost, BoardPost, Comment, InstantMessage, MailMessage, WikiArticle). Some additional terms (Answer, BestAnswer, Question) were also added for question-and-answer-type sites like Yahoo! Answers [156], whereby content from such sites can also be lifted onto the Semantic Web.

Community sites typically publish Web service interfaces for programmatic search and content management services (typically SOAP and/or REST). These services may be generic in nature (with standardized signatures covering input and output message formats) or service specific (where service signatures are unique to specific functions performed, as can be seen in current Web 2.0 API usage patterns). The SIOC Services ontology module allows one to indicate that a Web service is associated with (located on) a sioc:Site or a part of it. This module provides a simple way to tell others about a Web service, and should not be confused with Web service definitions that define the details of a Web service. A `sioc_s:service_definition` property is used to relate a `sioc_s:Service` to its full Web service definition (e.g., in WSDL, the Web Services Description Language).

A SIOC Access module was created to define basic information about users' permissions, access rights, and the status of content items in online communities. User access rights are modeled using Roles assigned to a user and Permissions on content items associated with these Roles. This module includes terms like `sioc_a:Status` that can be assigned to content items to indicate their publication status (e.g., public, draft, etc.), and `sioc_a:Permission`, which describes a type of action that can be performed on an object (e.g., a `sioc:Forum`, `sioc:Site`) that is within the scope of a `sioc:Role`.

An argumentation module extension to SIOC has been provided to allow one to formulate agreement and disagreement between SIOC content items. The properties and classes defined in this SIOC Argument module are related to other argumentation models such as SALT [134] and IBIS [60]. Some related work has also been performed by aligning SIOC with the SWAN ontology for scientific discourse in neuromedicine in the SWAN/SIOC joint initiative [135], providing a common framework to model online conversations in these communities, from the item level to the conversational layer.

Ontologies for Semantic Tagging: While they have been wrongly opposed in some discussions, ontologies and free-tagging practices can be efficiently combined together. One approach to bridge folksonomies and the Semantic Web is to use RDF(S)/OWL modeling principles to represent tags, tagging actions, and other related objects such as tag clouds. While tag-based search is the only way to retrieve tagged content at the moment (leading to problems if people use different tags for the same meaning, or if they use the same tag for different meanings), these new models allow advanced querying capabilities such as "which items are tagged 'semanticweb' on any platform," "what are the latest ten tags used by Denny on del.icio.us," "list all the tags commonly used by Alex on SlideShare and by John on Flickr," or "retrieve any content tagged with something relevant to the Semantic Web field." Having tags and tagged content published in RDF also allows one to easily link this to or from other Semantic Web data, and to reuse it across applications in order to achieve the goal of a global graph of knowledge.

While it has not been implemented in RDF, Gruber [53] defined one of the first approaches to model folksonomies and tagging actions using a dedicated ontology. This work considers the tripartite model of tagging and extends it with (1) a space attribute, aimed at modeling the website in which the tagging action occurred and (2) a polarity value in order to deal with spam issues. His proposal provides a complete model to represent tagging actions, but also considers the idea of a tag identity, such that various tags can refer to the same concept while being written differently, introducing the need to identify some common semantics in the tags themselves.

Mika discusses the emergent semantics of folksonomies and how they can be materialized through formal models, that is, ontologies [79]. In particular, he defines how ontologies can be mined from existing social bookmarking services by combining social network analysis with a tripartite model for representing ontologies, and therefore introduces a social component into ontology mining. The Tagora project [137] has also leveraged masses of data from collaborative tagging systems to examine the behavior of human agents on the Web and to help develop an understanding of the dynamics of information in online communities.

The Tag Ontology [138] was the first RDF-based model for representing tags and tagging actions, based on the initial ideas of Gruber and on the common theoretical model of tagging that was mentioned earlier. This ontology defines the "Tag" and "Tagging" classes with related properties to create the tripartite relationship of tagging. In order to represent the user involved in a tagging action, this ontology relies on the FOAF vocabulary. An important feature of this model is that it defines a Tag class, hence implying that each tag will have a proper URI so that tags can be used both as the subject and object of RDF triples. Since Tag is defined as a subclass of skos:Concept from the Simple Knowledge Organization System (SKOS), tags can be linked together, for example, to model that the "rdfa" tag is more specific than "semanticweb."

The Social Semantic Cloud of Tags (SCOT) ontology [65, 115] is focused on representing tag clouds and defines ways to describe the use and co-occurrence of tags on a given social platform, allowing one to move his or her tags from one service to another and to share tag clouds with others. SCOT envisions data portability not for the content itself but for the tagging actions and the tags of a particular user. SCOT reuses the Tag Ontology as well as SIOC to model tags, tagging actions, and tag clouds. An important aspect of the SCOT model is that it considers the space where the tagging action happened (i.e., the social platform, e.g., Flickr or del.icio.us), as suggested by Gruber's initial proposal. SCOT also provides various properties to define spelling variants between tags, using a main spellingVariant property and various subproperties such as acronym, plural, etc.

Finally, Meaning of a Tag (MOAT) [81] aims to represent the meaning of tags using URIs of existing domain ontology instances or resources from existing public knowledge bases [104], such as those from the Linking Open Data project (a set of best practice guidelines for publishing and interlinking pieces of data on the Semantic Web, see ❯ Semantic Annotation and Retrieval: Web of Data). To solve issues with free-form tagging regarding information retrieval, MOAT allows us to model facts such as: In this blog post, Bob uses the tag "apple" and by that he means the computer brand identified by `dbpedia:Apple_Inc.`, while the "apple" tag on that other picture means the fruit identified by `dbpedia:Apple`. MOAT is more than a single model, as it also provides a framework [82] to let people easily bridge the gap between simple free-form tagging and semantic indexing, helping users to annotate their content with URIs of Semantic Web resources from the tags that they have already used for annotated content. MOAT can also be automated as applied by [1] in the GroupMe! system.

More recently, the Common Tag initiative [19] (involving AdaptiveBlue, Faviki, Freebase, Yahoo!, Zemanta, Zigtag and DERI, NUI Galway) developed a lightweight vocabulary with a similar goal of linking tags to well-defined concepts (represented with their URIs) in order to make tagging more efficient and interconnected. In particular, it focuses on a simple approach allowing site owners to publish RDFa tag annotations, as well as providing a complete ecosystem of producers and consumers of Common Tag data that can help end users to deploy applications based on this format.

Other Ontologies: Various other models for modeling Social Web content have also been deployed. For example, in the wiki domain, WIF (Wiki Interchange Format) and WAF (Wiki Archive Format) have been developed [149] as common models to exchange and archive data between different wikis, as well as the WikiOnt vocabulary [54], with a more complete list of wiki-based models available [100].

Other application-specific models include SAM [42] and NABU [101] for instant messaging, as well as mle [107] and SWAML [36] for mailing list representation using Semantic Web technologies. These last two applications rely on SIOC, and the SWAML ontology has been completely integrated with SIOC.

In 2008, the Online Presence Ontology (OPO) [96] project was initiated for modeling online presence information. Whereas FOAF is mainly focused on static user profiles and SIOC has been somewhat oriented toward threaded discussions, OPO can be used to model dynamic aspects of a user's presence in the online world (e.g., custom messages, IM statuses, etc.). By expressing data using OPO, online presence data can be exchanged between services (chat platforms, social networks, and microblogging services). The ontology can also be used for exchanging IM statuses between IM platforms that use different status scales, since it enables very precise descriptions of IM statuses. The maintainers of OPO and SIOC are also working together to define alignments such that semantic descriptions of online presence and community-created content can be effectively leveraged on the Social Semantic Web.

APML [5], or Attention Profiling Markup Language, is an XML-based format that allows people to share their own personal "attention profile," similar to how OPML (Outline Processor Markup Language) allows the exchange of reading lists between sites and news readers. APML compresses all forms of attention-related data into a portable file format to provide a complete description of a person's rated interests (and dislikes). Efforts are also underway to link APML into the Semantic Web by creating an APML-RDF schema.

12.2 Example Applications

12.2.1 Semantic Blogging

A blog, or weblog, is a user-created website consisting of journal-style entries displayed in reverse-chronological order. Entries may contain text, links to other websites, and images or other media. Often there is a facility for readers to leave comments on individual entries, which makes blogs a very interactive medium. The growth and take-up of blogs over the past 5 years has been dramatic, with a doubling in the size of the blogosphere every 6 months or so (according to statistics from Technorati [141]). Over 120,000 blogs are created every day, with nearly a million blog posts being made each day. Technorati counted 133 million blogs in 2008 [140]. Bloggers are often at the forefront of information, where traditional media cannot act as fast as the online "wisdom of crowds."

Similar to accidentally wandering onto message boards and Web-enabled mailing lists, when searching for something on the Web, one often comes across a relevant entry on someone's blog. RSS feeds are also a useful way of accessing information from your favorite blogs, but they are usually limited to the last 15 or 20 entries, and do not provide much information on exactly who wrote or commented on a particular post, or what the post is talking about. Some approaches like SIOC (outlined earlier) aim to enhance the semantic metadata provided about blogs, forums, and posts, but there is also a need for more information about what exactly a person is writing about. Blog entries often refer to resources on the Web and these resources will usually have a context in which they are being used, and in terms of which they could be described. For example, a post which critiques a particular resource could incorporate a rating, or a post announcing an event could include start and end times. When searching for particular information in or across blogs, it is often not that easy to get it because of "splogs" (spam blogs) and also because of the fact that the virtue of blogs so far has been their simplicity – apart from the subject field, everything and anything is stored in one big text field for content. Keyword searches may give some relevant results, but useful questions such as "find me all the Chinese restaurants that bloggers reviewed in Washington DC with a rating of at least 5 out of 10" cannot be posed, and you cannot easily drag-and-drop events or people or anything (apart from URLs) mentioned in blog posts into your own applications.

Blog posts are sometimes categorized (e.g., "Japan," "Movies") by the post creator using predefined categories or tags, such that those on similar topics can be grouped together using free-form tags/keywords or hierarchical tree categories. Posts can also be tagged by others using social bookmarking services like Delicious [28] or personal aggregators like Gregarius [52]. Other services like Technorati can then use these tags or keywords as category names for linking together blog posts, photos, links, etc. in order to build what they call a "tagged Web." Using Semantic Web technology, both tags and hierarchical categorizations of blog posts can be further enriched and exposed in RDF via the SKOS framework.

Blogging at present offers poor query possibilities (except for searching by keyword or seeing all posts labeled with a particular tag). Some linking of posts is possible via direct HTML links or trackbacks, but nothing can be said about the nature of those links (are you agreeing with someone, linking to an interesting post, or are you quoting someone whose blog post is directly in contradiction with your own opinions?). There have been some approaches to tackle the issue of adding more information to blog posts, such that: (1) advanced (more precise) queries can be made regarding the posts' content; (2) the things that people talk about can be reused in other posts or applications (addresses, events, etc.); and (3) "richer" links can be created between blog posts (going beyond the previously described techniques involving categories or tags). One approach is called structured blogging [133] (mainly using microformats to annotate blog content), and the other is semantic blogging (using RDF to represent both blog structures and blog content): both approaches can also be combined together.

Structured blogging is an open-source community effort that has created tools to provide microcontent (e.g., microformats or RDFa, see Semantic Annotation and Retrieval: Web of Hypertext – RDFa and Microformats) from popular blogging platforms such as WordPress and Movable Type. In structured blogging, microcontent is positioned inline in the (X)HTML (and subsequent syndication feeds) and can be rendered via CSS. Although the original effort has tapered off, structured blogging is continuing through services like LouderVoice [73], a review site that integrates reviews written on blogs and other websites. In structured blogging, packages of structured data are becoming post components. Sometimes (not all of the time) a person will have a need for more structure in their posts – if they know a subject deeply, or if their observations or analyses recur in a similar manner throughout their blog – then they may best be served by filling in a form (which has its own metadata and model) during the post creation process. For example, someone may be writing a review of a film they went to see, or reporting on a sports game they attended, or creating a guide to tourist attractions they saw on their travels. Not only do people get to express themselves more clearly, but blogs can start to interoperate with enterprise applications through the microcontent that is being created in the background.

Semantic Blogging Applications: Semantic Web technologies can also be used to enhance any available post structures in a machine-readable way for more linkage and reuse, through various approaches in what is termed semantic blogging. Steve Cayzer [17] envisioned an initial idea for semantic blogging with two main aspects that could improve blogging platforms: a richer structure both for blog post metadata and their topics – using shared ontologies – and richer queries in terms of subscription, discovery, and navigation. He later defined a Snippet Manager service implementing some of these features. [64] gave some other ideas about "what would it mean to blog on the Semantic Web." They argued that such tools should be able to produce structured and machine-understandable content in an autonomous way, without any additional input from the users. They also provided a first prototype based on the Haystack platform [106] that showed new ways to navigate between content thanks to these techniques. Traditional blogging is aimed at what can be called the "eyeball Web" – that is, text, images, or video content that is targeted mainly at people [83]. Semantic blogging aims to enrich traditional blogging with metadata about the structure (what relates to what and how) and the content (what is this post about – a person, event, book, etc.). Already RSS and Atom are used to describe blog entries in a machine-readable way and enable them to be aggregated together. However, by augmenting these data with additional structural and content-related metadata, new ways of querying and navigating blog data become possible.

It is not simply a matter of adding semantics for the sake of creating extra metadata, but rather a case of being able to reuse what data a person already has in their desktop or Web space and making the resulting metadata available to others. People are already (sometimes unknowingly) collecting and creating large amounts of structured data on their computers, but these data are often tied into specific applications and locked within a user's desktop (e.g., contacts in a person's address book, events in a calendaring application, author and title information in documents, and audio metadata in MP3 files). Semantic blogging can be used to "lift" or release these data onto the Web (e.g., using applications like Shift [120]). For example, looking at the picture in *Fig. 12.4* [84], Andreas writes a blog post, which he annotates using content from his desktop address book application. He publishes this post on the Web, and someone else reading this post can reuse the embedded metadata in his or her own desktop applications (i.e., using the Web as a clipboard).

SparqlPress [127] is another prototype that leverages Semantic Web technologies in blogs. It is not a separate blogging system but rather an open-source plug-in for the popular WordPress platform [153], and it aims to produce, integrate, and reuse RDF data for an enhanced user experience. SparqlPress mainly relies on the FOAF, SIOC, and SKOS Semantic Web vocabularies. One interesting feature that SparqlPress provides is its linking of a FOAF social networking profile and an OpenID identity. This can be used to display extra information about the user on the blog, and it can also be useful for the blocking of blog comment spam. Finally, Zemanta [159] provides client-side and server-side tools that enrich the content being created by bloggers or publishers, allowing them to automatically add hyperlinks, choose appropriate tags (e.g., using the Common Tag framework), and insert images based on an analysis of the content being posted.

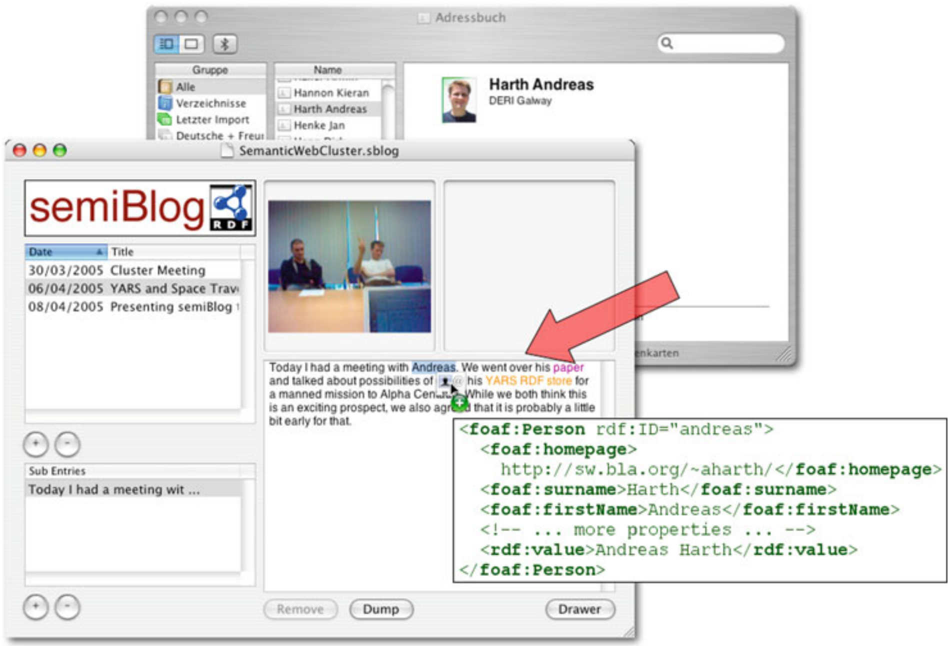


Fig. 12.4

Annotating a blog entry with an address book entry [83]

12.2.2 Semantic Microblogging

Microblogging is a recent social phenomenon on the Social Web, with similar usage motivations (i.e., personal expression and social connection) to other applications like blogging. It can be seen as a hybrid of blogging, instant messaging, and status notifications, allowing people to publish short messages (usually fewer than 140 characters) on the Web about what they are currently doing. These short messages, or microblog posts, are often called "tweets" (due to the most popular microblogging platform, Twitter [145]) and have a focus on real-time information. As a simple and agile form of communication in a fluid network of subscriptions, it offers new possibilities regarding lightweight information updates and exchange. Twitter is now one of the largest microblogging services, and the value of microblogging is demonstrated by its popularity and that of other services such as FriendFeed [45], StatusNet [129], etc. Some microblogging services also have SMS integration, allowing one to send updates and receive microblog posts from friends via a mobile phone.

Individuals can publish their brief text updates using various communication channels such as text messages from mobile phones, instant messaging, e-mail, and the Web. The simplicity of publishing such short updates in various situations or locations and the creation of a more flexible social network based on subscriptions and response posts makes microblogging an interesting communication method. Similar to how blogging led to "grassroots journalism," microblogging has led to grassroots reporters, especially Twitter, as updates can be posted in many ways and from different devices (e.g., via text message from mobile phones). Hence, it was one of the first media to report the May 2008 Sichuan earthquake in China [9] and the November 2008 terrorist attacks in Mumbai [142].

Microblogging is quite useful for getting a snapshot of what is going on in and for interacting with your community or communities of interest. Similar to using a blog aggregator and scanning the titles and summaries of many blogs at once, thereby getting a feel for what is going on at a particular point in time, microblogging allows one to view status updates from many people in a compact (screen) space. If you are subscribed to a few hundred people it can be somewhat difficult to see all that is relevant since even the most interesting microbloggers will not be talking about stuff that is interesting to you all the time. However, Twitter clients like TweetDeck [143] do allow various searches to be set up in separate columns, such that updates relevant to a certain keyword or combination of keywords (e.g., "galway OR ireland," "Semantic Web") can be monitored quite easily, irrespective of whom one is following.

This communication method is also promising for corporate environments in facilitating informal communication, learning, and knowledge exchange (e.g., Yammer [157] is an enterprise microblogging platform). Its so far untapped potential can be compared to that of company-internal wikis some years ago. Microblogging can be characterized by rapid (almost real-time) knowledge exchange and fast propagation of new information. For a company, this can mean real-time questions-and-answers and improved informal learning and communication, as well as status notifications, for example, about upcoming meetings and deliveries. However, the potential for microblogging in corporate environments still has to be demonstrated with real use cases (e.g., IBM has recently deployed an internal beta microblogging service called Blue Twit). It is expected that a trend of corporate microblogging will emerge in the next years similar to what happened with blogging, wikis, and other Enterprise 2.0 services (see eScience).

SMOB – Semantic MicrOBlogging: On the technology blog TechCrunch, Michael Arrington wrote a post [7] about the need for a "decentralised Twitter" via open alternative microblogging platforms, which was picked up by technologists Dave Winer, Marc Canter, and Chris Saad amongst others. The SMOB or semantic microblogging platform [125] developed in DERI, NUI Galway is an example of how Semantic Web technologies can provide an open platform for decentralized and distributed publishing of microblog content, mainly using the FOAF and SIOC vocabularies.

One aim of SMOB is to demonstrate how such technologies can provide users with a way to control, share, and remix their own data as they want to, not being solely dependent on the facilities provided by a third-party service, since SMOB-published data always belongs to the user who created it. As soon as someone writes some microblog content using a SMOB client, the content is spread through various microblogging servers or aggregators (including SMOB, Twitter, and Laconica), but remains available locally to the user who created it. Therefore, if one aggregator closes for some reason, the user can still use their local data somewhere else as it is primarily hosted by him or her and then aggregated by the third-party service.

In order to represent microblogging data, SMOB uses FOAF and SIOC to model microbloggers, their properties, account and service information, and the microblog updates that users create. A multitude of publishing services can ping one or a set of aggregating servers as selected by each user, and it is important to note that users retain control of their own data through self-hosting. The aggregate view of microblogs uses ARC [6] for storage and querying, and MIT's Exhibit faceted browser [80] for the user interface as shown in *Fig. 12.5*. It therefore offers a user-friendly interface for displaying complex RDF data aggregated from distributed sources. Furthermore, microblog posts can also embed semantic tags, for example, geographical tags, which can leverage the GeoNames database [47] to power new visualizations such as the map view in *Fig. 12.6*. At the moment, the complete dataset of updates is publicly available and can be browsed using any RDF browser (Tabulator, Disco, etc.).

Other Initiatives: Other microblogging platforms leveraging semantics include smesher [124] and StatusNet [129] (formerly Laconica). smesher is a semantic microblogging client with local storage, that integrates with Twitter and Identi.ca (another popular microblogging website). It features structure identification and a dashboard for custom filters, and has a SPARQL API for querying (see Querying the Semantic Web: SPARQL). StatusNet, an open-source platform that powers Identi.ca, also publishes both FOAF (describing people) and SIOC data (as SIOC-augmented RSS feeds for users and groups). StatusNet also allows users to create friend connections across installations.

Related to semantic microblogging are various approaches for embedding semantics in microblog posts, including microturtle [78], the "Star Priority Notation" [128], and microsyntax [77].

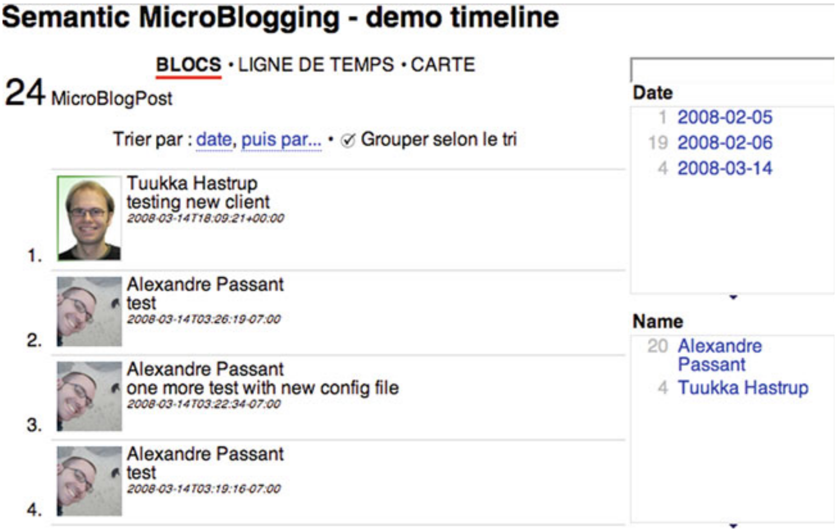


Fig. 12.5

Latest SMOB updates rendered in Exhibit

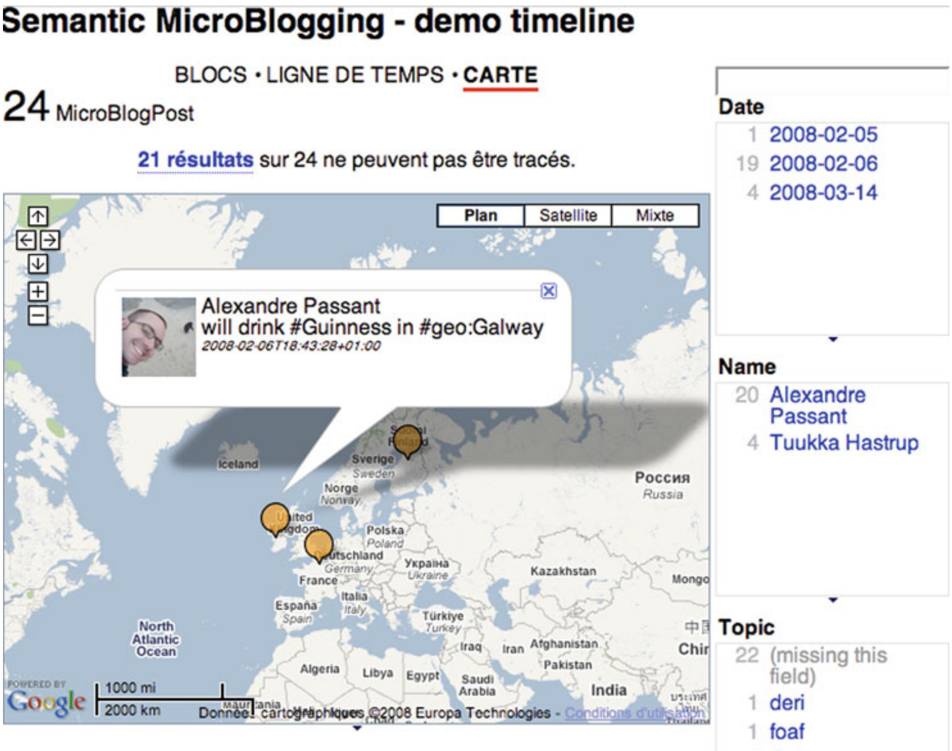


Fig. 12.6

Map view of latest updates with Exhibit

12.2.3 Semantic Wikis

Many people are familiar with Wikipedia [151], but less know exactly what a wiki is. A wiki is a website that allows users to edit content through the same interface they use to browse it, usually a Web browser, while some desktop-based wikis also exist. This facilitates collaborative authoring in a community, especially since editing a wiki does not require advanced technical skills. A wiki consists of a set of Web pages which can be connected together by links. Users can create new pages (e.g., if one for a certain topic does not exist), and they can also change (or sometimes delete) existing ones, even those created by other members. The WikiWikiWeb was the first wiki, established by Ward Cunningham in March 1995, and the name is based on the Hawaiian term wiki, meaning "quick," "fast," or "to hasten." Wikis often act as informational resources, like a reference manual, encyclopedia, or handbook. They amass to a group of Web pages where users can add content and others can edit the content, relying on cooperation, checks, and balances of its members, and a belief in the sharing of ideas. This creates a community effort in resource and information management, disseminating the "voice" amongst many instead of concentrating it upon few people. Therefore, contrary to how blogs reflect the opinions of a predefined set of writers (or a single author), wikis use an open approach whereby anyone can contribute to the value of the community.

Wikis are also being used for free dictionaries, book repositories, event organization, writing research papers, project proposals, and even software development or documentation. In this way, the openness of wiki-based writing can be seen as a natural follow-up to the openness of source-code modification. Wikis have become increasingly used in enterprise environments for collaborative purposes: research projects, papers and proposals, coordinating meetings, etc. SocialText [126] produced the first commercial open-source wiki solution, and many companies now use wikis as one of their main intranet collaboration tools. However, wikis may break some existing hierarchical barriers in organizations (due to a lack of workflow mechanisms, open editing by anyone with access, etc.), which means that new approaches toward information sharing must be taken into account when implementing wiki solutions.

There are hundreds of wiki software systems now available, ranging from MediaWiki [76], the software used on the Wikimedia family of sites, and Eu-gene Eric Kim's PurpleWiki [105], where fine-grained elements on a wiki page are referenced by purple numbers (a concept from Doug Engelbart), to Alex Schröder's OddMuse [92], a single Perl script wiki install, and WikidPad [150], a single-user desktop-based wiki for notes and personal information management. Many are open source, free, and will often run on multiple operating systems. The differences between wikis are usually quite small but can include the development language used (Java, PHP, Python, Perl, Ruby, etc.), the database required (MySQL, flat files, etc.), whether attachment file uploading is allowed or not, spam prevention mechanisms, page access controls, RSS feeds, etc.

A typical wiki page has two specific buttons of interest: "Edit" and "History." Normally, anyone can edit an existing wiki article, and if the article does not exist on a particular topic, anyone can create it. If someone messes up an article (either deliberately or erroneously), there is a revision history – as in most wiki engines – so that the contents can be reverted or fixed by the community. Thus, while there is no predefined hierarchy in most wikis, content is auto-regulated thanks to an emergent consensus within the community, usually achieved in a delicate mix of democracy and meritocracy (e.g., most wikis include discussions pages where people can discuss sensible topics).

Going further than what was discussed previously, in semantic blogging it is not just blog posts that are being enhanced by structured metadata and semantics – this is happening in many other Social Web application areas. Wikis such as the Wikipedia have contained structured metadata in the form of templates for some time now, and at least 20 semantic wikis [118] have also appeared to address a growing need for more structure in wikis. In his presentation on "The Relationship Between Web 2.0 and the Semantic Web" [51], Mark Greaves said that semantic wikis are a promising answer to various issues associated with semantic authoring, by reducing the investment of time required for training on an annotation tool and by providing incentives required to providing semantic markup (attribution, visibility, and reuse by others).

Typical wikis usually enable the description of resources in natural language. By additionally allowing the expression of knowledge in a structured way, wikis can provide advantages in querying, managing, and reusing information. Wikis such as the Wikipedia have contained structured metadata in the form of templates for some time now (to provide a consistent look to the content placed within article texts), but there is still a growing need for more structure in wikis (e.g., the Wikipedia page about Ross Mayfield links to about 25 pages, but it is not possible to answer a simple question such as "find me all the organizations that Ross has worked with or for"). Templates can also be used to provide a structure for entering data, so that it is easy to extract metadata about the topic of an article (e.g., from a template field called "population" in an article about London). Semantic wikis bring this to the next level by allowing users to create semantic annotations anywhere within a wiki article's text for the purposes of structured access and finer-grained searches, inline querying, and external information reuse. Generally, those annotations are designed to create instances of domain ontologies and their related properties (either explicit ontologies or ontologies that will emerge from the usage of the wiki itself), whereas other wikis use semantic annotations to provide advanced metadata regarding wiki pages. Obviously, both layers of annotation can be combined to provide advanced representation capabilities. By allowing people to add such extra metadata, the system can then show related pages (either through common relationships or properties, or by embedding search queries in pages). These enhancements are powered by the metadata that the people enter (aided by semantic wiki engines).

A semantic wiki should have an underlying model of the knowledge described in its pages, allowing one to capture or identify further information about the pages (metadata) and their relations. The knowledge model should be available in a formal language as RDFS or OWL, so that machines can (at least partially) process and reason on it. For example, a semantic wiki would be able to capture that an "apple" is a "fruit" (through an inheritance relationship) and present you with further fruits when you look at the apple article. Articles will have a combination of semantic data about the page itself (the structure) and the object it is talking about (the content). Some semantic wikis also provide what is called inline querying. For example, in SemperWiki [119], questions such as "?page dc:creator EyalOren" (find me all pages where the creator is Eyal Oren) or "?s dc: subject todo" (show all me all my to do items) are processed as a query when the page is viewed and the results are shown in the wiki page itself [98]. Also, when defining some relationships and attributes for a particular article (e.g., "foaf:gender male"), other articles with matching properties can be displayed along with the article.

Moreover, some wikis such as IkeWiki [114] feature reasoning capabilities, for example, retrieving all instances of foaf:Person when querying for a list of all foaf:Agent(s) since the first class subsumes the second one in the FOAF ontology. Finally, just as in the semantic blogging scenario, wikis can enable the Web to be used as a clipboard, by allowing readers to drag structured information from wiki pages into other applications (e.g., geographic data about locations on a wiki page could be used to annotate information on an event or a person in your calendar application or address book software, respectively).

Semantic MediaWiki: The most widely used semantic wiki is Semantic MediaWiki [68], an extension to the popular MediaWiki system as used on the Wikipedia. Semantic MediaWiki allows for the expression of semantic data describing the connection from one page to another, and attributes or data relating to a particular page. Semantic MediaWiki is completely open in terms of the terms used for annotating content, such that the underlying data model, that is, the different ontologies used to model the instances, evolve according to user behavior.

Let us take an example of providing structured access to information via a semantic wiki. There is a Wikipedia page about JK Rowling that has a link to "Harry Potter and the Deathly Hallows" (and to other books that she has written), to Edinburgh because she lives there, and to Scholastic Press, her publisher. In a traditional wiki, you cannot perform fine-grained searches on the Wikipedia data set such as "show me all the books written by JK Rowling," or "show me all authors that live in the UK," or "what authors are signed to Scholastic," because the type of links (i.e., the relationship type) between wiki pages are not defined. In Semantic MediaWiki, you can do this by linking with [[author of::Harry Potter and the Deathly Hallows]] rather than just the name of the novel. There may also be some attribute such as [[birthdate::1965-07-31]], which is defined in the JK Rowling article.

Such attributes can be used for answering questions like "show me authors over the age of 40" or for sorting articles, since this wiki syntax is translated into RDF annotations when saving the wiki page. Moreover, page categories are used to model the related class for the created instance. Indeed, in this tool, as in most semantic wikis that aim to model ontology instances, not only do the annotations make the link types between pages explicit, but they also make explicit the relationships between the concepts referred to in these wiki pages, thus bridging the gap from documents plus hyperlinks to concepts plus relationships. For instance, in the previous example, the annotation will not model that "the page about JK Rowling is the author of the page about Harry Potter and the Deathly Hallows" but rather that "the person JK Rowling is the author of the novel Harry Potter and the Deathly Hallows, and the pages describe these entities."

The data within the wiki are exported using RDF and other export formats, so that the knowledge within the wiki can be reused in external applications. An example of such a reuse is the Beers of the World website [10], providing a sophisticated beer selector based on Exhibit [80] and drawing the data from an external Semantic MediaWiki installation.

Semantic MediaWiki is supported by a vibrant developer community, creating further extensions on top of the core system, and used in several hundred sites. The extensions add the ability for form-based editing, graphical querying, keyword-based structural queries, integration with mapping and geocoding services, exhibit visualization, and many others. The system has been localized to more than 40 languages.

OntoWiki: OntoWiki [94] is a semantic wiki developed by the AKSW research group at the University of Leipzig that also acts as an agile ontology editor and distributed knowledge engineering application. Unlike other semantic wikis, OntoWiki relies more on form-based mechanisms for the input of structured data rather than using syntax-based or markup-based inputs. One of the advantages of such an approach is that complicated syntaxes for representing structured knowledge can be hidden from wiki users and, therefore, syntax errors can be avoided. OntoWiki visually presents a knowledge base as an information map, with different views on available instance data. It aims to enable intuitive authoring of semantic content, and also features an inline editing mode for editing RDF content, similar to WYSIWIG for text documents. As with most wikis, it fosters social collaboration aspects by keeping track of changes and allowing users to discuss any part of the knowledge base, but OntoWiki also enables users to rate and measure the popularity of content, thereby honoring the activities of users. OntoWiki enhances the browsing and retrieval experience by offering semantically enhanced search mechanisms. Such techniques can decrease the entrance barrier for domain experts and project members to collaborate using semantic technologies. OntoWiki is open source and is based on PHP and MySQL.

SweetWiki: SweetWiki (Semantic Web Enabled Technology Wiki) [136] is a semantic wiki prototype featuring augmented-tagging features for end users. In contrast to the other wikis mentioned above, it is not designed for creating and maintaining ontology instances, but rather uses Semantic Web technologies to augment the user experience and navigation between pages. One relevant feature of SweetWiki regarding the work described in this chapter is the ability to organize tags as a hierarchy of concepts. This hierarchy is then modeled in RDFS so that it can be reused in other applications, while the wiki model itself is defined using a particular OWL ontology. Most importantly, this hierarchy of tags is not a personal one but is built and shared amongst all the users of the wiki. In this way, SweetWiki provides a social and collaborative approach to maintaining hierarchies of concepts that can be seen as lightweight ontologies. Moreover, users can define two tags as synonyms in order to solve heterogeneity issues. From a tagging point of view, tags can be not only assigned to Web pages but also to pictures and embedded videos, and these are then used to retrieve or browse content, while similar and related tags are used to augment the navigation process by suggesting related pages. Finally, SweetWiki models all of its data using RDFa. Hence, an application that wants to reuse it is only required to extract and parse an XHTML page, since all the required RDF annotations are embedded in it and can be extracted using GRDDL.

DBpedia: While not a semantic wiki per se, DBpedia benefits from wiki principles to build a large machine-readable knowledge base of structured and interlinked data [8]. It provides an RDF export of the Wikipedia and can be seen as one of the core components of the Linking Open Data [12] project (see Semantic Annotation and Retrieval: Web of Data). Notably, various datasets in the Linking Open Data cloud (see *Fig. 12.7* [72]) link to DBpedia, since it is considered as being a central point [13] in efforts toward linking structured data on the Semantic Web.

DBpedia is created by exporting the "infoboxes" (i.e., metadata entered on Wikipedia articles using predefined template structures) from various language versions of Wikipedia and linking them together. By weaving Wikipedia articles and related objects into the Semantic Web, DBpedia defines URIs for many concepts so that people can use them in their semantic annotations. These various URIs can be used for instance in semantic tagging and semantic social bookmarking applications, as proposed in the MOAT framework (see Sect. 12.1.4) or in Faviki [35] (see Sect. 12.2.4). Another use case, especially related to social networking, is the reuse of these URIs to indicate interests of people, in combination with the foaf:topic interest property from the FOAF vocabulary. For example, to indicate that one is interested in Semantic Web technologies and the Social Web, he or she can use the following snippet of code. In this way, such information can be retrieved using SPARQL (see Querying the Semantic Web: SPARQL), providing better ways to find people who are sharing a common interest in an online community.

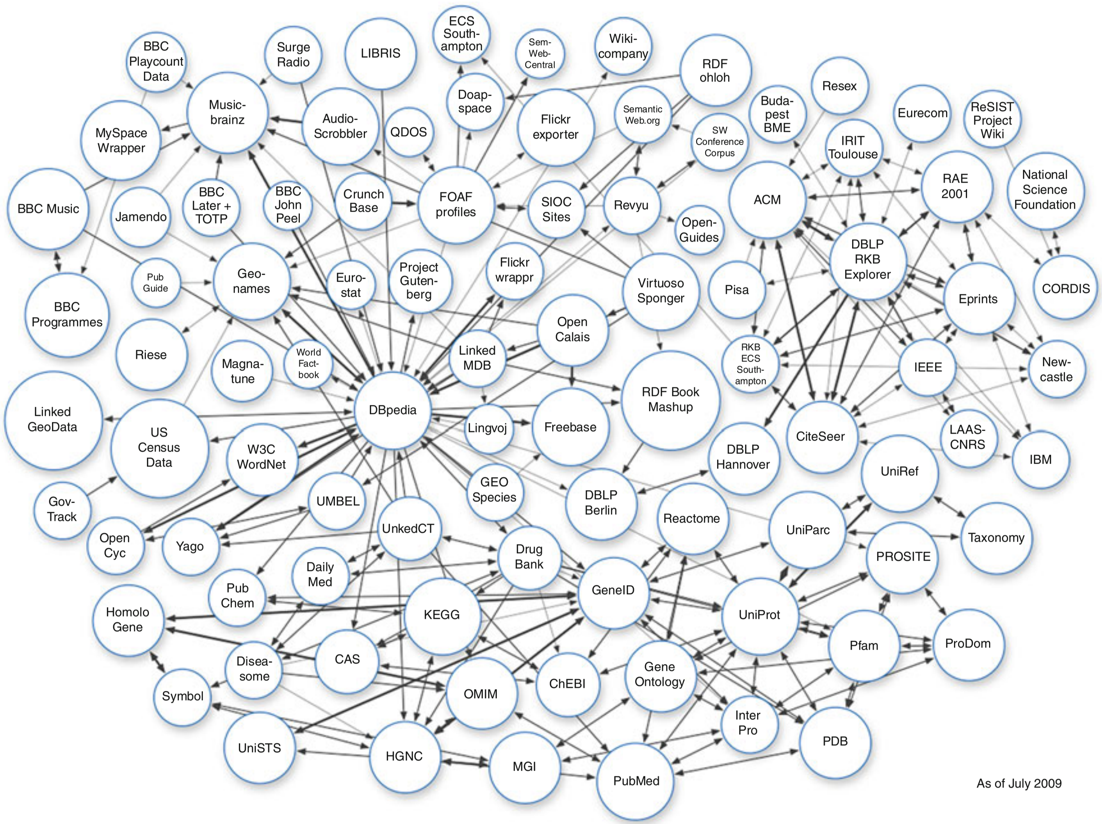
```
@prefix foaf: <http://xmlns.com/foaf/0.1/>.
@prefix dbpedia: <http://dbpedia.org/resource/>.
@prefix : <http://example.org/>.

:me foaf:topic_interest dbpedia:Semantic_Web, dbpedia:Social_web.
```

In addition, the DBpedia dataset is freely available for download and it also provides a public SPARQL endpoint so that anyone can interact with it for advanced querying capabilities [26]. Other ways to navigate DBpedia include a faceted browser [23], where people can restrict content by type (using predefined facets as well as an auto-completion system), for example, "Scientist," and then refine their queries based on various criteria, for example, birthdate [24].

An interesting application related to DBpedia is DBpedia Mobile [27], a "location-centric DBpedia client application for mobile devices" that consists of a map interface, the Marbles Linked Data Browser [74], and a GPS-enabled launcher application. The application displays nearby DBpedia resources (from a set of 300,000) based on a users' geolocation as recorded through his or her mobile device. Efforts are also ongoing toward allowing DBpedia to feed new content back into the Wikipedia [67] (e.g., by suggesting new values for infoboxes, or by contributing back new maps created via DBpedia Mobile).

12.2.4 Semantic Social Bookmarking

Twine: Radar Networks is one of a number of startup companies that is applying Semantic Web technologies to social software applications. Radar's flagship product



■ **Fig. 12.7**

The Linking Open Data Cloud. CC by Richard Cyganiak and Anja Jentzsch – http://lod-cloud.net

is called Twine [144], and the company is led by CEO Nova Spivack. The Twine service allows people to share what they know and can be thought of as a knowledge networking application that allows users to share, organize, and find information with people they trust. People create and join "twines" (community containers) around certain topics of interest, and items (documents, bookmarks, media files, etc., that can be commented on) are posted to these twines through a variety of methods. Twine has a number of novel and useful functions that elevate it beyond the social bookmarking sites to which it has been compared, including an extensive choice of twineable item types, twined item customization ("add detail" allows user-chosen metadata fields to be attached to an item) and the "e-mail to a twine" feature (enabling twines to be populated through messages sent to a custom e-mail address).

The focus of Twine is these interests. Where Facebook [33] is often used for managing one's social relationships and LinkedIn [70] is used for connections that are related to one's career, Twine can be used for organizing one's interests. With Twine, one can share knowledge, track interests with feeds, carry out information management in groups or communities, build or participate in communities around one's interests, and collaborate with others. Twine allows people to find things that might be of interest to them based on what they are doing.

Twine performs natural language processing on text, mainly providing automatic tagging with semantic capabilities. It has an underlying ontology with a million instances of thousands of concepts to generate these tags (at present, Twine is exposing just some of these). Radar Networks is also working on statistical analysis and machine-learning approaches for clustering related content to show people, items and interests that are related to each other (e.g., to give information to users such as "here is a selection of things that are all about movies you like").

Everything in Twine is generated from an ontology. Even the site itself – user interface elements, sidebars, navigation bar, buttons, etc. – come from an application-definition ontology. Similarly, the Twine data is modeled on a custom ontology. However, Twine is not just limited to these internal ontologies, and Radar Networks is beginning the process of bringing in other external ontologies and using them within Twine. They offer a lightweight ontology editing tool to allow people to make their own ontologies (e.g., to express domain-specific content) resulting in the Twine community having a more extensible infrastructure.

Twine search also has semantic capabilities. For example, bookmarks can be filtered by the companies they are related to, or people can be filtered by the places they are from. Underneath Twine, a lot of research work on scaling has been carried out, but it is not trying to index the entire Web. However, Twine does pull in related objects (e.g., from links in an e-mail), thereby capturing information around the information that you bring in and that you think is important. In terms of data interoperability, semantic data can be obtained from Twine in RDF for reuse elsewhere (by appending "?rdf" to the end of any Twine URL). Having already hardcoded some interoperability with services like Amazon [3] and provided import functionality from Delicious [28], Radar Networks is also looking at potential adaptors to other services including Digg [29], desktop bookmark files, Outlook contacts, Lotus Notes, Exchange, and Freebase [43].

Twine is aiming at mainstream users, so the interface has to be simple so that someone who knows nothing about structured data or automatic tagging should be able to figure out in a few minutes or even seconds how to use it. Individuals are Twine's first target market, allowing them to author and develop rich semantic content. For example, this could be a professional who has a need for a particular interest in some technical subject that is outside the scope of what they are doing at the moment. However, such a service becomes more valuable when users are connected to other people, if they join groups, thereby giving a richer network effect. The main value proposition for these users is that they can keep track of things they like, people they know, and capturing knowledge that they think is important.

When groups start using Twine, collective intelligence begins to take place (by leveraging other people who are researching material, finding items, testing, commenting, etc.). It is a type of communal knowledge base similar to other services like Wikipedia or Freebase. However, unlike many public communal sites, in Twine more than half of the data and activities are private (60%). Therefore privacy and permission control is very important, and it is deeply integrated into the Twine data structures. Since Twine left beta in 2008, public twines have become visible to search engines and SEO has been applied to increase the visibility of this content.

With such a service, there is a requirement for duplication detection. Most people submit similar bookmarks and it is reasonably straightforward to identify these, for example, when the same item enters the system through different paths and has different URLs. However, some advanced techniques are required when the content is similar but comes from different locations on the Web.

Referring to the theory of object-centered sociality [66] and others (i.e., people are networked through shared objects of interest), there is great potential in the community aspects of twines. These twines can act as "social objects" that will draw people back to the service in a far stronger manner than other social bookmarking sites currently do (in part, this is due to there being a more identifiable home for these objects and also due to the improved commenting facilities that Twine provides).

Faviki: Faviki [35] is a social bookmarking service that uses a controlled vocabulary for its tags, namely, the resources defined in Wikipedia. Hence, it provides features such as multilingual tagging (with various tags being automatically linked to the same concept), a related tags suggestion service (based on the relationships between these concepts), and it can also display tag descriptions. Faviki relies on DBpedia [25], Zemanta [159], and Google Language APIs [48] to provide its service. In addition, Faviki exposes its data in RDFa using the Common Tag format.

12.2.5 Review Websites

Revyu.com: Revyu.com [56, 110] is an online service dedicated to creating reviews for various items ranging from conference papers to pubs or restaurants (*Fig. 12.8*). It reuses some well-known principles and features of Social Web applications such as tags,

Fig. 12.8
The Revyu service

tag clouds, and stars ratings, and it provides a JavaScript bookmarklet to ease the publication of new reviews for end users when browsing the Web. Most importantly, Revyu.com is completely RDF based.

Each review is modeled using the RDF Review vocabulary [109] (compatible with the hReview microformat), and tags as well as tagging actions are represented using the Tag Ontology.

As Revyu.com also provides a SPARQL endpoint for querying its data, it also allows one to reuse tagged data from the website in any other application, as well as enabling mash-ups with existing content. Two important features of Revyu.com regarding the use of Semantic Web technologies are:

- *Integration and Interlinking with Other Data Sets*: Thanks to different heuristics, Revyu.com integrates identity links (using owl:sameAs properties) to resources already defined on the Semantic Web, especially resources being described in datasets from the Linking Open Data cloud (see Semantic Annotation and Retrieval: Web of Data). For example, most reviews regarding research papers are linked to the paper definition from the Semantic Web Dogfood project [30], while reviews about movies can be automatically linked to their DBpedia URI. Thus, it provides global interlinking of Semantic Web resources rather than defining new URIs for existing concepts.
- *The ability to consume FOAF-based user profiles*: While many Social Web applications require the user to fill in their personal details when subscribing, with those details having already been filled in on other platforms, Revyu.com allows one to simply give

his or her FOAF URI so that the information contained therein is automatically reused. Consuming FOAF profiles in Web-based applications provides a first step toward solving data portability issues between applications on the Social Web.

Therefore, Revyu.com combines Web 2.0-type interfaces and principles such as tagging with Semantic Web modeling principles to provide a reviews website that follows the principles of the Linking Open Data initiative. Anyone can review objects defined on other services (such as a movie from DBpedia), and the whole content of the website is available in RDF, therefore it is available for reuse by other Social Semantic Web applications.

12.2.6 Social Semantic Web Applications for Sharing Scientific Research

Science Collaboration Framework: The Science Collaboration Framework (SCF) is a reusable, open-source platform for structured online collaboration in biomedical research that leverages existing biomedical ontologies and RDF resources on the Semantic Web. The SCF GPL software [113] consists of the Drupal core content management system and customized modules. SCF supports structured Social Web–type community discourse amongst biomedical research scientists that is centered on a variety of interlinked heterogeneous data resources available to them (both formal and informal content, including scientific articles, news items, interviews, and various other perspectives).

The first instance of the SCF framework is being used to create an open-access online community for stem cell research called StemBook [130]. StemBook was developed based on requirements from the Harvard Stem Cell Institute. A second community is being planned for PD Online, a Web community for Parkinson's disease researchers sponsored by the Michael J. Fox Foundation (MJFF). The developers of SCF have cited significant overlaps between PD Online requirements and existing features built for StemBook, suggesting that the framework will achieve feature convergence through successive community implementations.

The architecture [22] makes it possible to define common schemas in OWL for a set of Web communities and to enable interoperability across biological resources, SWAN research statements or other objects of interest defined in the shared schemas. It is planned to make these graphs available via RDFa embedded within the HTML, and this work is being carried out in parallel with efforts to integrate RDFa into Drupal core [20]. myExperiment [87] is a collaborative environment where scientists can publish their experimental results and the workflows they used to produce these results. The myExperiment team reuses existing vocabularies for publishing and sharing experimental data by scientists. In David Newman's myExperiment ontology [86], concepts from Dublin Core, FOAF, and SIOC are reused since they are closely related to the two main functions of myExperiment: a social networking framework for researchers, and a metadata registry for experiments.

12.3 Future Issues

12.3.1 Searching the Social Semantic Web

As has been described already, RDF can be used to structure and expose information from the Social Web allowing the simple generation of semantic mash-ups and integrated views for both proprietary and public information. HTML content can also be made compatible with RDF through RDFa (RDF annotations embedded in XHTML attributes), thereby enabling effective semantic search without requiring one to crawl a new set of pages (e.g., the Common Tag effort allows metadata and URIs for tags to be exposed using RDFa and shared with other applications).

Search engine companies have recently started to publish recommended ontologies so that site owners and publishers can use these to annotate their content with RDFa or microformats. The result is that more relevant content will be displayed in search results, and the results themselves become more visually appealing (by using metadata to show people or organizations on a map, to display stars for review ratings, etc.), thereby encouraging click throughs.

Yahoo! SearchMonkey [116] has published a list of recommended vocabularies (including FOAF, GoodRelations, hReview, SIOC, vCalendar, and vCard) that publishers can use to create structured data and thereby drive more traffic to their sites. Google's "Rich Snippets" initiative (introduced in May 2009 [49]) has a similar aim, albeit using Google's own RDF vocabularies rather than popular existing ones like FOAF. Rich Snippets also promotes the use of the hReview, hProduct, and hCard microformats for annotating reviews, products, and people or organizations, respectively.

By providing RDFa-enabled HTML templates for popular social software applications with metadata relevant for search results (leveraging experience gained from creating RDF/XML exporters for WordPress, vBulletin, etc.), a very important step toward the formation of the Social Semantic Web can be taken. The RDFa in Drupal initiative is one of the first efforts to do this.

The SPARQL query language can be used for searching not just for keywords but for relationships between people and objects in aggregated Social Semantic Web data. Using RDF and SPARQL, it becomes possible to integrate diverse information from heterogeneous social websites, enabling improved navigation and the ability to query over these data. There are also advantages for those interested in studying social networks or looking for less obvious connections between people, as the Semantic Web makes large-scale, multi-relational datasets freely available for analysis. While the majority of SPARQL interfaces are designed for use within application architectures, more specialized user-oriented interfaces to custom SPARQL queries could provide network visualizations based on implicit and explicit relationships. There may also be new business models based on SPARQL queries across these data aggregates, for example, to provide topic-centric advertising on one site based on the related linked objects (and associated topics) from other sites.

A sample query for extracting the social network formed by explicit foaf:knows relationships follows using the SPARQL query language:

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT DISTINCT ?s ?o
WHERE {
  ?s a foaf:Person.
  ?o a foaf:Person.
  ?s foaf:knows ?o.
}
```

In addition to explicitly stated person-to-person links, there are many implicit social connections present on the Web. A sample query for extracting the implicit social network formed by replies to posts follows (using the has reply property from SIOC).

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX sioc: <http://rdfs.org/sioc/ns#>
SELECT ?author1 ?author2
WHERE {
  ?post1 a sioc:Post;
    foaf:maker ?author1;
    sioc:has_reply ?post2.
?post2 a sioc:Post;
    foaf:maker ?author2.
}
```

12.3.2 Trust and Privacy on the Social Semantic Web

Some challenges must also be overcome regarding the online identity aspect of the Social Web, as well as authentication and privacy for users of social websites. An interesting aspect of social networking and social websites is that most people use various websites because they want to fragment their online identity: uploading pictures of friends on MySpace, forming business contacts on LinkedIn, etc. Under each persona, a user may reveal completely different facets of their personality. People may wish to share many of their identities with certain contacts, but retain more privacy when dealing with others. For example, many people are careful to keep their personal life distinct from their professional life. However, just as people may wish to keep separate identities for some purposes, it can also be beneficial to be able to connect these persons as and when desired. Members of online communities often expend a lot of effort into forming relationships and building their reputation. Since reputation determines how much trust other people will place in an individual, it can be of very real value and therefore the ability to maintain a reputation across different identities could be very beneficial.

While the Semantic Web and, in particular, reasoning principles (such as leveraging Inverse Functional Properties (IFPs)) allow one to merge these data and provide vocabularies, methods, and tools for data portability among social websites, this intentional identity fragmentation must be taken into account on the Social Semantic Web. It implies a need for new ways to authenticate queries or carry out inferencing, by delivering data in different forms depending on, for instance, which social subgraph the person requesting the data belongs to (family, coworker, etc.). Here, Web 2.0 efforts like OAuth [91] are of interest. OAuth is an open protocol that enables users to grant applications access to their protected data stored in accounts they hold with other services. Also relevant is the recent proposal for FOAF+SSL [131, 132]. Moreover, advanced social aspects of contextualizing information delivery may be added later. The nature of each relationship (e.g., work, family, romantic, friendship) could be taken into account, as well as the current status, location, or even the mood of a user. In some cases, external influences such as the political climate in a country may be considered in determining what kind of information to share about an individual. Additionally, as relationships evolve over time, the processing of requests could be updated accordingly.

Besides the issue of security, the issue of privacy also arises due to the possibility to more easily combine existing data sources. In the USA, donors for political campaigns, be it for a candidate or a proposition, are listed publicly. This improves transparency and aims at countering hidden influences in policy-making and the democratic process. In November 2008 in California, voters passed proposition 8, which overturned the right of same-sex couples to marry. The campaigns concerning the propositions raised a record of over 80 million US Dollars – and as always, the list of donors was released to the public. The list contained names and addresses of the donors. The addresses were geo-coded and a mash-up with Google Maps was created to provide a map of all donors for proposition 8 (see screenshot in *Fig. 12.9* [32]). This is just one example how data can be merged. In the future, we expect the Semantic Web technology to simplify the creation of such mash-ups even further, so that even nontechnical users will be able to create such views on data on-the-fly. The impact of these technologies on privacy is yet only marginally understood, and protocols like P3P [102] (the Platform for Privacy Preferences) as well as specifications like CC REL [18] (for including licenses in content) are important here.

12.3.3 Integrating with the Social Semantic Desktop

The vision of the Social Semantic Web has emerged in parallel with another prominent research area, that of the Social Semantic Desktop [117], where users and their peers can share and interlink multimedia content, calendars, e-mails, or documents. There is a convergence occurring between these two areas, for example, in the areas of information models (for personal or shared content), collaboration and knowledge exchange, domain-specific structures, expert finding, argumentative discussions, and semantic authoring and publishing. The combination of these two areas is leading to what has been termed Social Semantic Information Spaces [62].

In Sect. 12.2.1 an example of an overlap between the Social Semantic Web and the Social Semantic Desktop was given, where a blogging application allowed one to leverage

■ **Fig. 12.9**
A screenshot of EightMaps, a mash-up of Proposition 8 donors and Google Maps [32]

information from the desktop (e.g., in an address book or event calendar) and lift the associated metadata onto the Web for reuse by others. Similar integrations can occur with desktop-based microblogging solutions or mobile phone photo applications. One particular project of interest and referenced earlier is the Nepomuk social semantic desktop [90], a European Union-funded project that focused on how semantic technologies can help people to find and add structure to information on their personal computers, and to share that information with other users. Nepomuk allows users to give meaning to documents, contact details, pictures, videos, and other data files stored on a user's computer, regardless of the file format, associated application, or language used, making it easier and quicker to find information and to identify connections between different items. When information is added to the desktop, Nepomuk asks users to annotate the information so that it can be correctly situated, and it also crawls the user's computer to search for existing information thereby establishing connections between different content items. Nepomuk is available in KDE4, a popular desktop environment for Linux.

12.4 Cross-References

- Knowledge Management in Large Organizations
- Ontologies and the Semantic Web

- Querying the Semantic Web: SPARQL
- Semantic Annotation and Retrieval: RDF
- Semantic Annotation and Retrieval: Web of Data

Acknowledgments

The work presented in this chapter was supported by the Líon-2 project supported by Science Foundation Ireland under Grant No. SFI/08/CE/I1380.

Supported by the EU IST project ACTIVE, http://www.active-project.eu.

References

1. Abel, F.: The benefit of additional semantics in folksonomy systems. In: Proceeding of the Second Ph.D. Workshop on Information and Knowledge Management (PIKM 2008), Napa Valley, pp. 49–56. ACM Press, New York (2008)
2. http://www.google.com/adsense
3. http://www.amazon.com
4. Ankolekar, A., Vrandečić, D., Krötzsch, M., Tran, D.T.: The two cultures: mashing up web 2.0 and the semantic web. In: Proceedings of the 16th International Conference on the World Wide Web (WWW 2007), Banff. ACM Press, New York (2007)
5. http://apml.areyoupayingattention.com
6. http://arc.semsol.org
7. http://www.techcrunch.com/2008/05/05/twitter-can-be-liberated-heres-how
8. Auer, S., Bizer, C., Lehmann, J., Kobilarov, G., Cyganiak, R., Ives, Z.: Dbpedia: a nucleus for a web of open data. In: Proceedings of the Sixth International Semantic Web Conference and Second Asian Semantic Web Conference (ISWC/ASWC 2007), Busan. Lecture Notes in Computer Science, vol. 4825, pp. 722–735. Springer, Berlin
9. http://www.bbc.co.uk/blogs/technology/2008/05/twitter_and_the_china_earthqua.html
10. http://beer.geekworks.de
11. http://www.bibsonomy.org
12. Bizer, C., Heath, T., Berners-Lee, T.: Linked data – the story so far. Int. J. Semant. Web Inf. Syst. **5**, 1–22 (2009)
13. Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., Hellmann, S.: DBpedia – a crystallization point for the web of data. J. Web Semant. **7**, 154–165 (2009)
14. Breslin, J.G., Harth, A., Bojārs, U., Decker, S.: Towards semantically-interlinked online communities. In: Proceedings of the Second European Semantic Web Conference (ESWC 2005), Heraklion. Lecture Notes in Computer Science, vol. 3532, pp. 500–514. Springer, Heidelberg (2005)
15. http://danbri.org/words/2007/11/04/223
16. Brickley, D., Miller, L.: FOAF vocabulary specification. Namespace document, FOAF project. http://xmlns.com/foaf/0.1/ (2004)
17. Cayzer, S.: Semantic blogging and decentralized knowledge management. Commun. ACM **47**(12), 47–52 (2004)
18. http://wiki.creativecommons.org/CC_REL
19. http://www.commontag.org
20. Corlosquet, S., Delbru, R., Clark, T., Polleres, A., Decker, S.: Produce and consume linked data with drupal! In: Proceedings of the Eighth International Semantic Web Conference (ISWC 2009), Chantilly. Lecture Notes in Computer Science, vol. 5823, pp. 763–778. Springer, Berlin (2009)
21. http://www.craigslist.org
22. Das, S., Greena, T., Weitzman, L., Lewis-Bowen, A., Clark, T.: Linked data in a scientific collaboration framework. In: Proceedings of the 17th International Conference on the World Wide Web (WWW 2008), Beijing (2008)
23. http://dbpedia.neofonie.de/browse/
24. http://dbpedia.neofonie.de/browse/rdf-type:Scientist/
25. http://dbpedia.org
26. http://dbpedia.org/sparql/
27. http://wiki.dbpedia.org/DBpediaMobile
28. http://www.delicious.com
29. http://digg.com

30. http://data.semanticweb.org
31. Dolog, P., Krötzsch, M., Schaffert, S., Vrandečić, D.: Social web and knowledge management. In: King, I., Baeza-Yates, R.A. (eds.) Weaving Services and People on the World Wide Web, pp. 217–227. Springer, Berlin (2008)
32. http://www.eightmaps.com
33. http://www.facebook.com
34. http://www.dcs.shef.ac.uk/˜mrowe/foafgenerator.html
35. http://www.faviki.com
36. Fernández, S., Berrueta, D., Labra, J.E.: Mailing lists meet the semantic web. In: Proceedings of the BIS 2007 Workshop on Social Aspects of the Web (SAW 2007), Poznan. CEUR, vol. 245. CEUR-WS.org (2007)
37. http://www.flickr.com
38. http://apassant.net/blog/2007/12/18/rdf-export-flickr-profiles-foaf-and-sioc/
39. http://www.foaf-project.org
40. http://www.ldodds.com/foaf/foaf-a-matic
41. http://foafbuilder.qdos.com
42. Franz, T., Staab, S.: SAM: semantics aware instant messaging for the networked semantic desktop. In: Proceedings of the First Workshop on the Semantic Desktop (SemDesk 2005), Fourth International Semantic Web Conference (ISWC 2005), Galway. CEUR, vol. 175. CEUR-WS.org (2005)
43. http://www.freebase.com
44. http://www.freedb.org
45. http://friendfeed.com
46. http://www.friendster.com
47. http://www.genoames.org
48. http://code.google.com/apis/ajaxlanguage/
49. http://www.google.com/support/webmasters/bin/answer.py?answer=99170
50. http://maps.google.com
51. http://rease.semanticweb.org/ubp/PUSH/search@srchDetailsLR?lrID=lr-lear-diederich- 118647867 1106
52. http://gregarius.net
53. Gruber, T.: Ontology of folksonomy: a mash-up of apples and oranges. Int. J. Semant. Web Inf. Syst. **3**(2), 1–11 (2007)
54. Harth, A., Gassert, H., O'Murchu, I., Breslin, J.G., Decker, S.: WikiOnt: an ontology for describing and exchanging wikipedia articles. In: Proceedings of Wikimania 2005 – The First International Wikimedia Conference, Frankfurt (2005)
55. http://microformats.org/wiki/hcard
56. Heath, T., Motta, E.: Revyu.com: a reviewing and rating site for the web of data. In: Proceedings of the Sixth International Semantic Web Conference and Second Asian Semantic Web Conference (ISWC/ASWC 2007), Busan. Lecture Notes in Computer Science, vol. 4825, pp. 895–902. Springer, Heidelberg (2007)
57. Hendler, J.A., Golbeck, J.: Metcalfe's law, web 2.0, and the semantic web. J. Web Semant. **6**(1), 14–20 (2008)
58. http://hi5.com
59. http://www.housingmaps.com
60. http://hyperdata.org/xmlns/ibis/
61. http://identi.ca
62. http://www2006.org/tutorials/#T13
63. http://esw.w3.org/topic/IswcPodcast
64. Karger, D.R., Quan, D.: What would it mean to blog on the semantic web? In: The Semantic Web: Third International Semantic Web Conference (ISWC 2004), Hiroshima. Lecture Notes in Computer Science, vol. 3298, pp. 214–228. Springer, Heidelberg (2004)
65. Kim, H.L., Yang, S.-K., Breslin, J.G., Kim, H.-G.: Simple algorithms for representing tag frequencies in the SCOT exporter. In: Proceedings of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology, Silicon Valley, pp. 536–539. IEEE Computer Society, Washington, DC (2007)
66. Knorr-Cetina, K.: Sociality with objects: social relations in postsocial knowledge societies. Theory Cult. Soc. **14**(4), 1–30 (1997)
67. http://videolectures.net/www09_kobilarov_dbpldh/
68. Krötzsch, M., Vrandecic, D., Völkel, M.: Semantic MediaWiki. In: Proceedings of the Fifth International Semantic Web Conference (ISWC 2006), Athens, GA. Lecture Notes in Computer Science, vol. 4273, pp. 935–942. Springer, Heidelberg (2006)
69. http://last.fm
70. http://www.linkedin.com
71. http://www.livejournal.com
72. http://richard.cyganiak.de/2007/10/lod/
73. http://www.loudervoice.com
74. http://marbles.sourceforge.net
75. Mcafee, A.P.: Enterprise 2.0: the dawn of emergent collaboration. MIT Sloan Manag. Rev. **47**(3), 21–28 (2006)
76. http://www.mediawiki.org
77. http://www.microsyntax.org

78. http://buzzword.org.uk/2009/microturtle/
79. Mika, P.: Ontologies are us: a unified model of social networks and semantics. In: Proceedings of the Fourth International Semantic Web Conference (ISWC 2005), Sardinia. Lecture Notes in Computer Science, vol. 3729, pp. 522–536. Springer, Heidelberg (2005)
80. http://www.simile-widgets.org/exhibit/
81. http://www.moat-project.org
82. http://www.moat-project.org/architecture
83. Möller, K., Bojars, U., Breslin, J.G.: Using semantics to enhance the blogging experience. In: Proceedings of the Third European Semantic Web Conference (ESWC 2006), Budva. Lecture Notes in Computer Science, vol. 4011, pp. 679–696. Springer, Heidelberg (2006)
84. Möller, K., Decker, S.: Harvesting desktop data for semantic blogging. In: Decker, S., Park, J., Quan, D., Sauermann, L. (eds.) Proceedings of the Semantic Desktop (SD 2005), Workshop at the ISWC 2005, Galway (2005)
85. http://www.mybloglog.com
86. http://users.ecs.soton.ac.uk/drn05r/release/stable/myexp.owl
87. http://www.myexperiment.org
88. http://www.myspace.com
89. http://network.nature.com
90. http://nepomuk.semanticdesktop.org
91. http://www.oauth.net
92. http://www.oddmuse.org
93. http://www.opengraphprotocol.org/
94. http://www.ontowiki.net
95. http://www.openid.net
96. http://www.milanstankovic.org/opo/
97. O'Reilly, T.: O'Reilly network: what is web 2.0: design patterns and business models for the next generation of software. http://www.oreillynet.com/lpt/a/6228 (2005)
98. Oren, E., Völkel, M., Breslin, J.G., Decker, S.: Semantic wikis for personal knowledge management. In: Proceedings of the 17th International Conference on Database and Expert Systems Applications (DEXA 2006), Krakow. Lecture Notes in Computer Science, vol. 4080, pp. 509–518. Springer, Heidelberg (2006)
99. http://www.orkut.com
100. Orlandi, F., Passant, A.: Enabling cross-wikis integration by extending the SIOC ontology. In: Proceedings of the Fourth Workshop on Semantic Wikis (SemWiki 2009), Heraklion (2009)
101. Osterfeld, F., Kiesel, M., Schwarz, S.: Nabu – a semantic archive for XMPP instant messaging. In: Proceedings of the First Workshop on the Semantic Desktop (SemDesk 2005), Fourth International Semantic Web Conference (ISWC 2005), Galway. CEUR, vol. 175. CEUR-WS.org (2005)
102. http://www.w3.org/P3P/
103. Passant, A.: me OWL:sameAs flickr:33669349@N00. In: Proceedings of the WWW 2008 Workshop Linked Data on the Web (LDOW 2008), Beijing. CEUR, vol. 369. CEUR-WS.org (2008)
104. Passant, A., Laublet, P., Breslin, J.G., Decker, S.: A URI is worth a thousand tags: from tagging to linked data with MOAT. Int. J. Semant. Web Inf. Syst. **5**, 71–94 (2009)
105. http://purplewiki.blueoxen.net
106. Quan, D., Huynh, D., Karger, D.R.: Haystack: a platform for authoring end user semantic web applications. In: Proceedings of the Second International Semantic Web Conference (ISWC 2003), Sanibel Island. Lecture Notes in Computer Science, vol. 2870, pp. 738–753. Springer, Heidelberg (2003)
107. Rehatschek, H., Hausenblas, M.: Enhancing the exploration of mailing list archives through making semantics explicit. In: Proceedings of the Semantic Web Challenge 2007, Collocated with the Sixth International Semantic Web Conference (ISWC 2007), Seoul (2007)
108. http://vocab.org/relationship/
109. http://www.purl.org/stuff/rev#
110. http://revyu.com
111. Rowe, M., Ciravegna, F.: Getting to me – Exporting semantic social network from facebook. In: Proceedings of the First Workshop on Social Data on the Web (SDoW 2008), Karlsruhe. CEUR, vol. 405. CEUR-WS.org (2008)
112. http://web.resource.org/rss/1.0/
113. http://sciencecollaboration.org
114. Schaffert, S.: IkeWiki: a semantic wiki for collaborative knowledge management. In: Proceedings of the First International Workshop on Semantic Technologies in Collaborative Applications (STICA 2006), Manchester (2006)
115. http://www.scot-project.org
116. http://developer.yahoo.com/searchmonkey/
117. http://www.semanticdesktop.org
118. http://semanticweb.org/index.php/Semantic_Wiki_State_Of_The_Art
119. http://www.eyaloren.org/semperwiki.html

120. http://kantenwerk.org/shift
121. http://www.sioc-project.org
122. http://www.w3.org/Submission/2007/02/
123. http://www.w3.org/2004/02/skos/
124. http://www.smesher.org
125. http://smob.sioc-project.org
126. http://www.socialtext.com
127. http://bzr.mfd-consult.dk/sparqlpress
128. http://civilities.net/Star_Priority_Notation
129. http://status.net
130. http://www.stembook.org
131. http://blogs.sun.com/bblfish/entry/rdfauth_sketch_of_a_buzzword
132. Story, H., Harbulot, B., Jacobi, I., Jones, M.: FOAF+TLS: restful authentication for the social web. In: Proceedings of the First Workshop on Trust and Privacy on the Social and Semantic Web (SPOT 2009), Heraklion. CEUR Workshop Proceedings, vol. 447. CEUR-WS.org (2009)
133. http://www.structuredblogging.org
134. http://salt.semanticauthoring.org
135. http://www.w3.org/TR/hcls-swansioc/
136. http://sweetwiki.inria.fr/ontology
137. http://www.tagora-project.eu/
138. http://www.holygoat.co.uk/projects/tags/
139. Tapscott, D., Williams, A.D.: Wikinomics: How Mass Collaboration Changes Everything. Pearson Education, New York (2007)
140. http://technorati.com/blogging/state-of-the-blogosphere/
141. http://technorati.com/weblog/2007/04/328.html
142. http://www.telegraph.co.uk/news/worldnews/asia/india/3530640/Mumbai-attacks-Twitter-and-Flickr-used-to-break-news-Bombay-India.html
143. http://www.tweetdeck.com
144. http://www.twine.com
145. http://twitter.com
146. http://semantictweet.com/
147. http://upcoming.org
148. http://www.ietf.org/rfc/rfc2426.txt
149. Völkel, M., Oren, E.: Towards a Wiki Interchange Format (WIF) – opening semantic wiki content and metadata. In: Proceedings of the First Workshop on Semantic Wikis – From Wiki to Semantics (SemWiki 2006), Budva. CEUR, vol. 206. CEUR-WS.org (2006)
150. http://wikidpad.sourceforge.net
151. http://en.wikipedia.org
152. http://en.wikipedia.org/wiki/Metcalfe's_law
153. http://wordpress.org
154. http://www.gmpg.org/xfn
155. http://vocab.sindice.com/xfn.html
156. http://answers.yahoo.com
157. http://www.yammer.com
158. http://www.youtube.com
159. http://www.zemanta.com