# The Future of Social Web Sites: Sharing Data and Trusted Applications with Semantics

SHEILA KINSELLA

*Digital Enterprise Research Institute, National University of Ireland, Galway, Ireland*

ALEXANDRE PASSANT

*Digital Enterprise Research Institute, National University of Ireland, Galway, Ireland*

JOHN G. BRESLIN

*Digital Enterprise Research Institute, National University of Ireland, Galway, Ireland*

*Department of Electronic Engineering, National University of Ireland, Galway, Ireland*

STEFAN DECKER

*Digital Enterprise Research Institute, National University of Ireland, Galway, Ireland*

AJIT JAOKAR

*University College London, United Kingdom*

**Abstract**

In recent years, there has been an explosion in the number of Social Web sites which allow the creation of knowledge through simplified user contributions via blogs, wikis, and the deployment of online social networks. As more Social Web

ADVANCES IN COMPUTERS, VOL. 76
ISSN: 0065-2458/DOI: 10.1016/S0065-2458(09)01004-3

sites 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 the objects in these sites: both within and across Social Web sites. Also, to better enable user access to multiple sites, interoperability among Social Web sites is required in terms of both the expressed data (content objects, person-to-person networks, etc.) and the social applications in use (e.g., widgets) on each site. This requires representation mechanisms for data and applications on the Social Web in an interoperable and extensible way. The Semantic Web provides such representation mechanisms: it can be used to link people and objects by representing the heterogeneous ties that bind us all to each other (either directly or indirectly). In this chapter, we will describe methods that build on agreed-upon Semantic Web formats to describe people, content objects, the connections that bind them together explicitly or implicitly, and embeddable application widgets on Social Web sites, thereby enabling these sites to interoperate by appealing to some common semantics. We will also focus on how a social aspect can be added to data such as software project and widgets descriptions, so that one can combine social networking, trust, and relationship aspects with those representation models. We will also look at how developers can use the Semantic Web to augment the ways in which they create, reuse, and link content on social networking sites and Social Web sites. In particular, we will see how both data and applications can be shared on the Web, thanks to these semantics.

## 1. Introduction

Since it was founded, the Web has been used to facilitate communication not only between computers but also between people. Usenet mailing lists and Web forums allowed people to connect with each other and enabled communities to form, often around topics of interest. The social networks formed via these technologies were not explicitly stated, but were implicitly defined by the interactions of the people involved. Later, technologies such as IRC (Internet Relay Chat), instant messaging and blogging continued the trend of using the Internet to build communities.

Social networking sites (SNSs) such as Friendster (an early SNS previously popular in the US, now widely used in Asia), Orkut (Google's SNS), LinkedIn (an SNS for professional relationships), and MySpace (a music- and youth-oriented service)—where explicitly stated networks of friendship form a core part of the Web site—have become part of the daily lives of millions of users, and 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. Boyd and Ellison [10] recently described the history of SNSs, and suggested that in the early days of SNSs, when only the SixDegrees service existed, there simply were not enough users: "While people were already flocking to the Internet, most did not have extended networks of friends who were online." A graph from Internet World Stats[1] shows the growth in the number of Internet users over time. Between 2000 (when SixDegrees shut down) and 2003 (when Friendster became the first successful SNS), the number of Internet users had doubled.

Content-sharing sites with social networking functionality such as YouTube (a video-sharing site), Flickr (for sharing images), and last.fm (a radio and music community site) have enjoyed similar popularity. The basic features of an SNS are profiles, friend's listings and commenting, often along with other features such as private messaging, discussion forums, blogging, and media uploading and sharing. Many content-sharing sites such as Flickr and YouTube also include some social networking functionality. In addition to SNSs, other forms of Social Web sites include wikis, forums, and blogs. Some of these publish content in structured formats enabling them to be aggregated together. The Social Web or "Web 2.0" has enabled community-based knowledge acquisition with efforts like the Wikipedia demonstrating the "wisdom of the crowds" in creating the world's largest online encyclopedia. Although it is difficult to define the exact boundaries of what structures or abstractions belong to the Social Web, a common property of such sites is that they facilitate collaboration and sharing between users with low technical barriers, although usually on single sites.

A limitation of current Social Web sites is that they are isolated from one another like islands in a sea. For example, different online discussions may contain complementary knowledge and topics, segmented parts of an answer that a person may be looking for, but people participating in one discussion do not have ready access to information about related discussions elsewhere. As more and more Social Web sites, communities, and services come online, the lack of interoperation among them becomes obvious: a set of single data silos or "stovepipes" has been created, that is, there are many sites, communities, and services that cannot interoperate with each other, where synergies are expensive to exploit, and where reuse and interlinking of data is difficult and cumbersome. The main reason for this lack of interoperation is that for the most part in the Social Web, there are still no common standards for knowledge and information exchange and interoperation available. RSS (Really Simple Syndication), a format for publishing recently updated Web content such as blog entries, could be a first solution for interoperability among Social Web sites, but it has various limitations that make it difficult to be used efficiently in such a context, as we will see later.

However, the Semantic Web effort aims to provide the tools that are necessary to define extensible and flexible standards for information exchange and interoperability. The Scientific American article from Berners-Lee et al. [4] defined the Semantic

[1] http://www.internetworldstats.com/emarketing.htm.

Web as "an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation." The last couple of years have seen large efforts going into the definition of the foundational standards supporting data interchange and interoperation, and currently a well-defined Semantic Web technology stack exists, enabling the creation of defining metadata and associated vocabularies. The Semantic Web effort is in an ideal position to make Social Web sites interoperable. The application of the Semantic Web to the Social Web can lead to a "Social Semantic Web" (Fig. 1), creating a network of interlinked and semantically rich knowledge. 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 modeled using machine-readable formats, so that it can be used for purposes that the current state of the Social Web cannot achieve without difficulty.

A semantic data "food chain" (see Fig. 2), that is, producers, collectors, and consumers of semantic data from social networks and Social Web sites, 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.

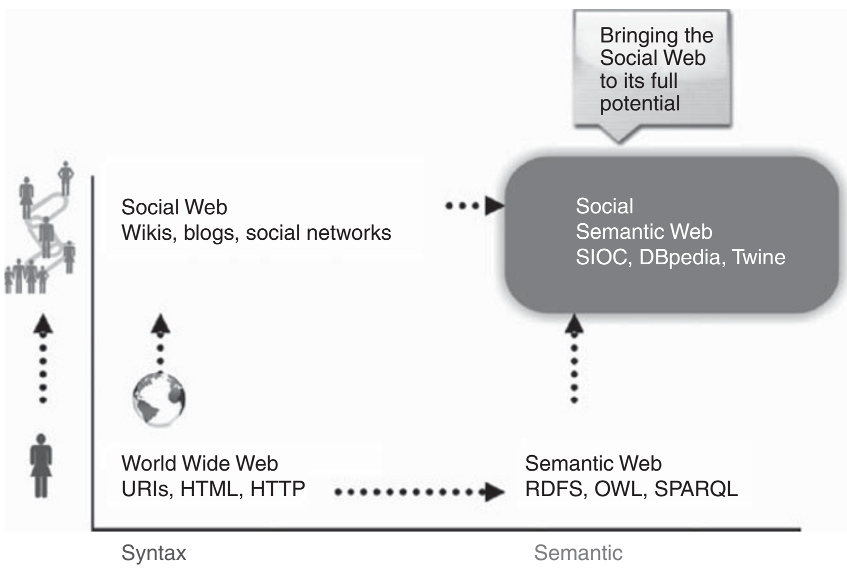


FIG. 1. The Social Semantic Web.

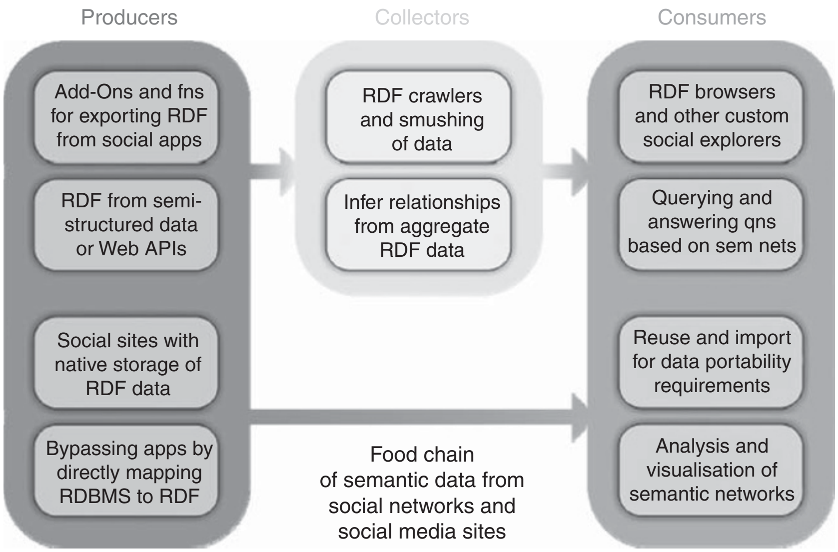


FIG. 2. A food chain for semantic data on the Social Web.

Applying semantic technologies to Social Web sites can greatly enhance the value and functionality of these sites. The information within these sites is forming vast and diverse networks which can benefit from Semantic Web technologies for representation and navigation. Additionally, to easily enable navigation and data portability across sites, mechanisms are required to represent data in an interoperable and extensible way. These are termed semantic data producers.

An intermediary step which may or may not be required is for the collection of semantic data. In very large sites, this may not be an issue as the information in the site may be sufficiently linked internally to warrant direct consumption after production, but in general, may users make small contributions across a range of services which can benefit from an aggregate view through some collection service. Collection services can include aggregation and consolidation systems, semantic search engines, or data lookup indexes.

The final step involves consumers of semantic data. Social networking technologies enable people to articulate their social network via friend connections. A social network can be viewed as a graph where the nodes represent individuals and the edges represent relations. Methods from graph theory can be used to study these networks, and we will describe how social network analysis (SNA) can consume semantic data from the food chain.

Also, representing social data in RDF (Resource Description Framework), a language for describing Web resources in a structured way, enables us to perform queries on a network to locate information relating to a person or people. Interlinking social data from multiple sources may give an enhanced view of information in distributed communities, and we will describe applications to consume this interlinked data.

In this chapter, we will begin by describing various social networking sites and Social Web sites, along with some of their limitations and initial approaches to leverage semantics in social networks, blogs, wikis, tagging, and software descriptions. We will discuss the representation methods that can be used by semantic producers to represent data (user profiles, feeds, content) and applications (widgets) for porting and sharing amongst users and sites. We will then describe the collection stage in a "semantic data food chain," giving examples of queries that can be used to consolidate aggregates of data from Social Web sites. We will also discuss how trust mechanisms in consuming applications can be leveraged via the distributed social graph, so that users can decide who to accept any new data or applications from. Finally, we will give our conclusions and ideas for future work.

# 2. Social Web Sites and Approaches to Add Semantics

## 2.1 Social Networks

The "friend-of-a-friend effect" often occurs when someone tells someone something and they then tell you—linked to the theory that anybody is connected to everybody else (on average) by no more than six degrees of separation. This number of six degrees came from a sociologist called Stanley Milgram who conducted an experiment in the late 1960s. Random people from Nebraska and Kansas were told to send a letter (via intermediaries) to a stock broker in Boston. However, they could only give the letter to someone that they knew on a first-name basis. Amongst the letters that found their target (around 20%), the average number of links was around 5.5 (rounded up to 6). While this experiment does not prove the theory of six degrees of separation, it does demonstrate that most individuals are not separated by many links. Some other related ideas include the Erdös number (the number of links required to connect scholars to mathematician Paul Erdös, a prolific writer who coauthored over 1500 papers with more than 500 authors), and the Kevin Bacon game (the goal is to connect any actor to Kevin Bacon, by linking actors who have acted in the same movie).

It is often found that even though one route is followed to get in contact with a particular person, after talking to them there is another obvious connection that was not previously known about. This is part of the small-world network theory [45], which says that most nodes in a network exhibiting small-world characteristics (such as a social network) can be reached from every other node by a small number of hops or steps.

There has been a proliferation of SNSs which Boyd and Ellison [10] define as a category of Web sites consisting of user profiles, which other users can comment on, and a traversable social network originating from publicly articulated lists of friends. The idea behind such services is to make people's real-world relationships explicitly defined online—whether they be close friends, business colleagues, or just people with common interests. Most SNSs allow one to surf from a list of friends to find friends-of-friends, or friends-of-friends-of-friends for various purposes. While the majority of these sites are for purely social reasons, others have additional purposes such as LinkedIn which is targeted toward professionals.

Before 2002, most people networked using online services such as OneList (a mailing list service), ICQ (an instant messaging program), or eVite (a site for sending invitations). The first big SNS in 2002 was Friendster; in 2003, LinkedIn and MySpace appeared; then in 2004, Orkut and Facebook (by a college student for college students) were founded; these were followed by Bebo (targeting both high school and college students) in 2005. Social networking services usually offer the same basic functionalities: network of friends listings (showing a person's "inner circle"), person surfing, private messaging, discussion forums or communities, events management, blogging, commenting (sometimes as endorsements on people's profiles), and media uploading. In general, these sites do not usually work together and therefore require you to re-enter your profile and redefine your connections when you register for each new site.

Some motivations for SNS usage include building friendships and relationships, arranging offline meetings, curiosity about others, arranging business opportunities, or job hunting. People may want to meet with local professionals, create a network for parents, network for social (dating) purposes, get in touch with a venture capitalist, or find out if they can link to any famous people via their friends.

In addition to relationship management, social networks are sometimes used for viral marketing [34], although recent results indicate that this might be less effective than often assumed. For example, Knorr-Cetina [31] reports that "the additional purchases that resulted from recommendations are just a drop in the bucket of sales" and that "marketers should take heed that even if viral marketing works initially, providing excessive incentives for customers to recommend products could backfire by weakening the credibility of the very same links they are trying to take advantage of."

A key feature of these sites is community-contributed content that may be tagged and can be commented on by others. That content can be virtually anything: blog entries, board posts, videos, audio, images, wiki pages, user profiles, bookmarks, events, etc. Already, sites are being proposed where live multiplayer video games will appear in browser-embedded windows just as YouTube does for videos, with running commentaries going on about the games in parallel. Tagging is common to many social networking Web sites—a tag is a keyword that acts like a subject or category for the associated content. Folksonomies (a portmanteau of the words "folks" and "taxonomies," meaning collaboratively generated, open-ended labeling systems) emerge from the use of tagging on a given platform and enable users of these sites to categorize content using the tags system, and to thereby visualize popular tag usages via "tag clouds" (visual depictions of the tags used on a particular Web site, similar to a weighted list in visual design, that provides an overview of the different categories and topics used within a community).

Even in a small-sized SNS, there can be a lot of links available for analysis, and these data are usually meaningless when viewed as a whole, so one usually needs to apply some SNA techniques.[2] Apart from comprehensive textbooks in this area [44], there are many academic tools for examining social networks and performing common SNA routines. For example, the tool Pajek[3][3] can be used to drill down into various social networks. A common method is to reduce the amount of relevant social network data by clustering. One can choose to cluster people by common friends, by shared interests, by geographic location, by tags, etc.

In SNA, people are modeled as nodes or "actors." Relationships (such as acquaintanceship, coauthorship, friendship, etc.) between actors are represented by lines or edges. This model allows analysis using existing tools from mathematical graph theory and mapping, with target domains such as movie actors, scientists and mathematicians (as already mentioned), sexual interaction, phone call patterns, or terrorist activity. There are some useful tools for visualizing these models, such as Vizster[4] by Heer and Boyd [26], based on the Prefuse[5] open-source toolkit.

## 2.2 Leveraging Semantics in "Object-Centered" Social Networks

Social networks exist all around us—at workplaces as well as within families and social groups. They are designed to help us work together over common activities or interests, but anecdotal evidence suggests that many SNSs lack such common

[2] http://lrs.ed.uiuc.edu/tse-portal/analysis/social-network-analysis/.
[3] http://vlado.fmf.uni-lj.si/pub/networks/pajek/.
[4] http://jheer.org/vizster/.
[5] http://prefuse.org/.

objectives [27]. Instead, users often connect to others for no other reason than to boost the number of friends they have in their profiles.[6] Many more browse other users' profiles simply for curiosity's sake. These explicitly established connections become increasingly meaningless because they are not backed up by common objects or activities.

The act of connecting sometimes becomes a site's primary (only) activity. In fact, some sites act simply as enhanced address books: although potentially useful for locating or contacting someone, they provide little attraction for repeat visits. This is a flaw with the current theory. As Jyri Engeström, cofounder of the Jaiku.com microblogging site (microblogging is a lightweight form of blogging that consists of short message updates), put it, "social network theory is good at representing links between people, but it doesn't explain what connects those particular people and not others." Indeed, many are finding that SNSs are becoming increasingly boring and meaningless.

Another problem is that the various SNSs do not usually work together. You thus have to re-enter your profile and redefine your connections from scratch when you register for each new site. Some of the most popular SNSs probably would not exist without this sort of "walled garden" approach, but some flexibility would be useful. Users often have many identities on different social networks. Reusable profiles would let them import existing identities and connections (from their own home page or another site they are registered on), thereby forming a single global identity with different views (using systems such as OpenID,[7] e.g., an open standard that enables users to log in to many Web sites using a single sign-on).

Engeström has theorized[8] that the longevity of Social Web sites is proportional to the "object-centered sociality" occurring in these networks, that is, the degree to which people are connecting via items of interest related to their jobs, workplaces, favorite hobbies, etc. Similarly, Jordan and colleagues [28] advocate *augmented social networks*, in which citizens form relationships and self-organize into communities around shared interests.

On the Web, social connections are formed through the actions of people—via the content they create together, comment on, link to, or for which they use similar annotations. Adding annotations to items in social networks (using topic tags, geographical pinpointing, etc.) is particularly useful for browsing and locating interesting items and people with similar interests. Content items such as blog entries, videos, and bookmarks serve as the lodestones for social networks, drawing

[6] http://www.russellbeattie.com/notebook/1008411.html.
[7] http://www.openid.net/.
[8] http://www.zengestrom.com/blog/2005/04/why_some_social.html.

people back to check for new items and for updates from others in their network. For many of the Social Web sites, success has come from enabling communities formed around common interests, where the users are active participants who as well as consuming information also provide content and metadata. In this way, it is probable that people's SNS methods will continue to move closer toward simulating their real-life social interaction, so that people will meet others via something they have in common, not by randomly approaching each other—eventually leading toward more realistic interaction methods with friends.

Virtual worlds such as Second Life have already begun to provide a user experience which is more faithful to reality. Users interact via avatars in a three-dimensional environment where they can move between different areas and socialize with other residents. An important aspect of Second Life is that the world is largely user-created. Residents can buy land, construct houses, and create objects. It is also possible to trade with other users, as well as buy or sell using the world's internal currency, the Linden Dollar. Second Life's world encourages residents to meet and stay in touch with other users with similar interests via themed areas and events—a prime example of object-centered sociality.

Figure 3 illustrates an object-centered social network for three people. Bob and Carol are connected through bookmarked Web sites that both have annotated, as well as through events they are both attending. Alice and Bob have matching tags on media items, and they subscribe to the same blogs.

Although object-centered social networks can fix one problem (that of sites becoming boring), the remaining challenge is how to achieve interoperability among SNSs and, ultimately, content-creation facilities on the Web. As more social networks form around connections between people and their objects of interest, and as these object-centered social networks grow bigger and more diverse, more intuitive methods are needed for representing and navigating the information in these networks—within and across SNSs. Also, to better enable navigation across sites, interoperability among SNSs is required in terms of both the content objects and the person-to-person networks expressed on each site. That requires representation mechanisms to interconnect people and objects on the Web in an interoperable, extensible way [11].

Semantic Web representation mechanisms are ideally suited to describing people and the objects that link them together in such object-centered networks, by recording and representing the heterogeneous ties that bind each to the other. By using agreed-upon Semantic Web formats to describe people, content objects, and the connections that bind them together, social networks can also interoperate by appealing to common semantics. Developers are already using Semantic Web technologies to augment the ways in which they create, reuse, and link content on social networking and Social Web sites [15]. These efforts include the Friend-of-a-Friend

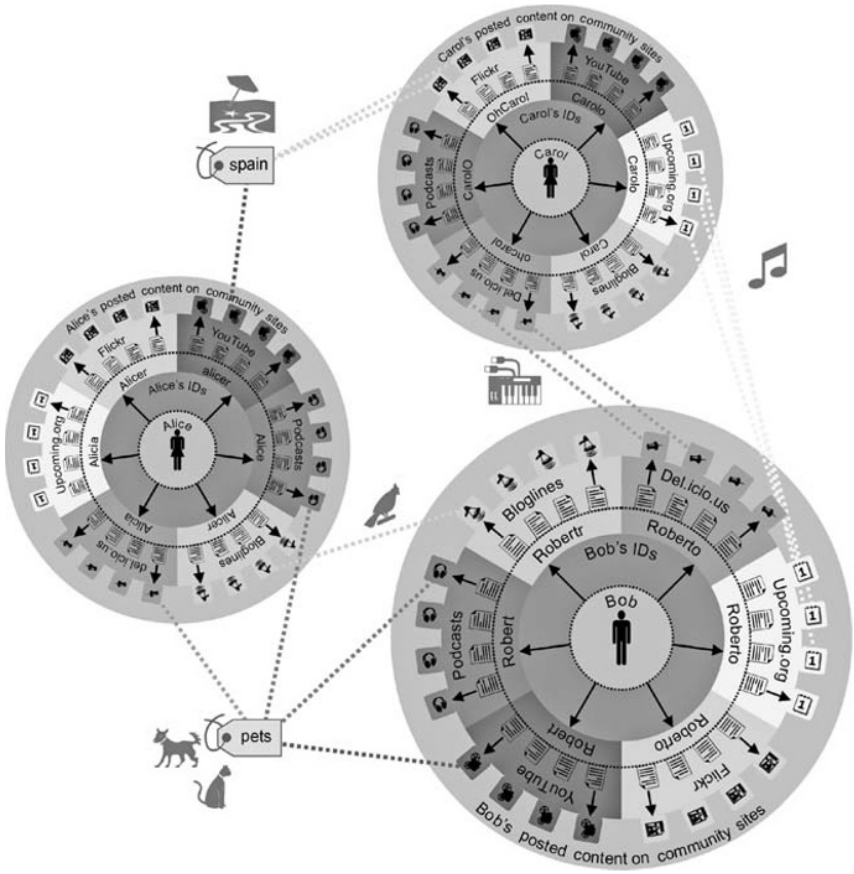


FIG. 3. Users form object-centered social networks (using their possibly multiple online accounts) around the content items they act on via social Web sites.

(FOAF) project[9] for describing people and relationships, the Nepomuk social semantic desktop[10] which is a framework for extending the desktop to a collaborative environment for information management and sharing, and the Semantically Interlinked Online Communities (SIOC) initiative[11] for representing online

[9] http://www.foaf-project.org/.
[10] http://nepomuk.semanticdesktop.org/.
[11] http://sioc-project.org/.

discussions. Some SNSs, such as Facebook, are also starting to provide query interfaces to their data, which others can reuse and link to via the Semantic Web.[12]

The Semantic Web is a useful platform for linking and for performing operations on diverse person—and object-related data gathered from heterogeneous SNSs. In the other direction, object-centered networks can serve as rich data sources for Semantic Web applications. This linked data can provide an enhanced view of individual or community activity in localized or distributed object-centered social networks. In fact, since all these data are semantically interlinked using well-given semantics (e.g., using the FOAF and SIOC ontologies), in theory it makes no difference whether the content is distributed or localized. All of these data can be considered as a unique interlinked machine-understandable graph layer (with nodes as users and related data and arcs as relationships) over the existing Web of documents and hyperlinks, that is, a Giant Global Graph as Tim Berners-Lee recently coined.[13] Moreover, such interlinked data allow advanced querying capabilities, for example, "show me all the content that Alice has acted on in the past three months."

As Tim Berners-Lee said in a 2005 podcast,[14] 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." For example, SNS users are already creating extensive vocabularies and annotations through folksonomies [38]. 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.

## 2.3 Blogs

A blog, or Weblog, is a user-created Web site consisting of journal style entries displayed in reverse chronological order. Entries may contain text, links to other Web sites, and images or other media. Often, there is a facility for readers to leave comments on individual entries, which make blogs an interactive medium. Blogs may be written by individuals, or by groups of contributors. A blog may function as a personal journal, or it may provide news or opinions on a particular subject.

The growth and take-up of blogs over the past 5 years has been impressive, with a doubling in the size of the "blogosphere" every 6 or so months (according to statistics from Technorati[15]). Over 100,000 blogs are created everyday, working out at about

[12] http://www.openlinksw.com/blog/~kidehen/?id=1237.
[13] http://dig.csail.mit.edu/breadcrumbs/node/215.
[14] http://esw.w3.org/topic/IswcPodcast.
[15] http://technorati.com/weblog/2007/04/328.html.

one a second. Nearly, 1.5 million blog posts are being made each day, with over half of bloggers still contributing to their sites 3 months after the blog's creation.

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 (more later) 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 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 Dublin 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 Uniform Resource Locators—URLs) mentioned in blog posts into your own applications.

## 2.4 Adding Semantics to Blogs

There have been some approaches to tackle the issue of adding more information to blog posts, so that queries can be made and the things that people talk about can be reused in other posts or applications (because not everyone is being served well by the lowest common denominator that we currently have in blogs). One approach is called "structured blogging,"[16] and the other is "semantic blogging": both approaches can also be combined together.

Structured blogging is an open-source community effort that has created tools to provide microcontent from popular blogging platforms such as WordPress and Movable Type. The term microcontent indicates a unit of data and associated metadata communicating one main idea and accessible at a URI. Sources of microcontent include microformats,[17] which enable semantic markup to be embedded directly within XHTML. Microformats therefore provide a simple method of

[16] http://structuredblogging.org/.
[17] http://microformats.org/.

expressing content in a machine-readable way, facilitating reuse and aggregation. An example of a microformat is hReview, which allows for the structured description of reviews within Web pages. Another approach to annotating XHTML documents is RDFa (Resource Description Framework in attributes)[18] which makes it possible to embed semantics in XHTML attributes in such a way that enables the data to be mapped to RDF.

Although the original effort has tapered off, structured blogging is continuing through services like LouderVoice,[19] a review site which integrates reviews written on blogs and other Web sites. 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.

Take the scenario where someone (or a group of people) is reviewing some soccer games that they watched. Their after-game soccer reports will typically include information on which teams played, where the game was held and when, who were the officials, what were the significant game events (who scored, when and how, or who received penalties and why, etc.)—it would be easier for these blog posters if they could use a tool that would understand this structure, presenting an editing form with the relevant fields, and automatically create both HTML and RSS with this structure embedded in it. Then, others reading these posts could choose to reuse this structure in their own posts, and their blog reading/writing application could make this structure available when the blogger is ready to write. As well as this, reader applications could begin to answer questions based on the form fields available—"show me all the matches from South Africa with more than two goals scored," etc.

At the moment, structured blogging tools provide a fixed set of forms that bloggers can fill in for things like reviews, events, audio, video, and people—but there is no reason that people could not create custom structures, and news aggregators or readers could autodiscover an unknown structure, notify a user that a new structure is available, and learn the structure for reuse in the user's future posts.

[18] http://www.w3.org/TR/xhtml-rdfa-primer/.
[19] http://www.loudervoice.com/.

Semantic Web technologies can also be used to enhance any available post structures in a machine-readable way for more linkage and reuse. Blog posts are usually only tagged on the blog itself by the post creator, using free-text keywords such as "scotland," "movies," etc. (or can be tagged by others using social bookmarking services like del.icio.us or personal aggregators like Gregarius). Technorati, the blog search engine, aims to use these keywords to build a "tagged Web." Both tags and hierarchical categorizations of blog posts can be further enriched using the SKOS (Simple Knowledge Organization Systems) framework for representing vocabularies. However, there is often much more to say about a blog post than simply what category it belongs in.

This is where semantic blogging comes in. 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. 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 (a format for syndicating Web content) 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.

In structured blogging, microcontent such as microformats or RDFa is positioned inline in the (X)HTML (and subsequent syndication feeds) and can be rendered via CSS. Structured blogging and semantic blogging do not compete, but rather offer metadata in slightly different ways (using microcontent and RDF, respectively). There are already mechanisms such as GRDDL (Gleaning Resource Descriptions from Dialects of Languages)[20] which can be used to move from one to the other and allows one to provide RDF data from embedded RDFa or microformats. Extracted RDF data can then be reused as would any native RDF data, and so it may be processed using common Semantic Web tools and services.

The question remains as to why one would choose to enhance their blogs and posts with semantics. Current blogging offers poor query possibilities (except for searching by keyword or seeing all posts labeled with a particular tag). There is little or no reuse of data offered (apart from copying URLs or text from posts). Some linking of posts is possible via direct HTML links or trackbacks, but again, 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?). Semantic blogging aims to tackle some of these issues, by facilitating better (i.e., more precise) querying when compared with

[20] http://www.w3.org/TR/grddl/.

keyword matching, by providing more reuse possibilities, and by creating "richer" links between blog posts.

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, audio metadata in MP3 files). Semantic blogging can be used to "lift" or release these data onto the Web, as in the semiBlog[21] application (now called Shift) which allows users to reuse metadata from Apple Mac desktops in blog posts. For example, Aidan can write a blog post which he annotates using metadata about events and people from his desktop calendaring and address book applications. He publishes this post onto the Web, and John, reading this post, can reuse the embedded metadata in his own desktop applications. As well as semiBlog, other semantic blogging systems have been developed by HP,[22] the National Institute of Informatics, Japan[23] and MIT.[24]

Also, conversations often span multiple blog sites in blog posts and their comments, and bloggers may respond to the entries of other users in their own blogs. The use of semantic technologies can also enable the tracking of these distributed conversations. Links between units of conversation could even be enhanced to include sentiment information, for example, who agrees or disagrees with the initial opinion.

## 2.5 Wikis

A wiki is a Web site which 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, and change existing ones, even those created by other members. One of the most well-known wikis is the Wikipedia free online encyclopedia. Wikis are also being used for free dictionaries, book repositories, event organization, and software

[21] http://semiblog.semanticweb.org/.
[22] http://www.hpl.hp.com/personal/Steve_Cayzer/semblog.htm.
[23] http://www.semblog.org/.
[24] http://theory.csail.mit.edu/~dquan/iswc2004-blog.ppt.

development. They have become increasingly used in enterprise environments for collaborative purposes: research projects, papers and proposals, coordinating meetings, etc. SocialText[25] 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. This is why some argue that Enterprise 2.0 [36], that is, the use of social software in or within companies, raises more philosophical issues than technical ones.

There are hundreds of wiki software systems now available, ranging from MediaWiki, the software used on the Wikimedia family of sites, and PurpleWiki, where fine-grained elements on a wiki page are referenced by purple numbers, to OddMuse, a single Perl script wiki install, and WikidPad, a desktop-based wiki for managing personal information. 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.

The Wikipedia project consists of over 250 different wikis, corresponding to a variety of languages. The English-language one is currently the biggest, with over 2 million pages, but there are wikis in languages ranging from Gaelic to Chinese. A typical wiki page will have two 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 so that the contents can be reverted or fixed by the community. Thus, while there is no predefined hierarchy in most wikis, content is autoregulated, thanks to an emergent consensus within the community, ideally in a democratic way (for instance, most wikis include discussions pages where people can discuss sensible topics). There is a certain amount of ego-related motivation in contributing to a wiki—people like to show that they know things, to fix mistakes, and fill in gaps in underdeveloped articles (stubs), and to have a permanent record of what they have contributed via their registered account. By providing a template structure to input facts about certain things (towns, people, etc.), wikis also facilitate this user drive to populate wikis with information.

[25] http://www.socialtext.com/.

## 2.6 Adding Semantics to Wikis

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. 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 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 and properties of domain ontologies (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. There are already about 20 semantic wikis in existence, and one of the largest ones is Semantic MediaWiki, based on the popular MediaWiki system. 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.

Let us take an example of providing structured access to information in wikis. 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 could 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 author of the page about Harry Potter and the Deathly Hallows" but rather that "the person JK Rowling is author of the novel Harry Potter and the Deathly Hallows."

Since Semantic MediaWiki is completely open in terms of the wiki syntax for annotating content, extracted data may be subject to heterogeneity problems. For instance, some users will use [[author of:xxx]] while others will prefer [[has written: xxx]], leading to problems when querying data. Other wikis such as OntoWiki, IkeWiki, or UfoWiki assist the user when modeling semantic annotations, to avoid those heterogeneity issues and provide data that are based on predefined ontologies.

Some semantic wikis also provide what is called inline querying. A question such as "?page dc:creator EyalOren" (or find me all pages where the creator is Eyal Oren) is processed as a query when the page is viewed and the results are shown in the wiki page itself. 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 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).

## 2.7 Tags, Tagging, and Folksonomies

Apart from providing a means to define and manage social networks, one of the most important features of Social Web sites is the ability to upload and share content with others, either with anyone subscribed to (or just browsing) the Web site or else within a restricted community. Various media files can be shared, such as pictures, videos, bookmarks, slides, etc. To make this content more easily discoverable, users can add free-text keywords, or tags, to any content that they upload. For example, this chapter could be tagged with "Semantic Web," "social networks," and "SIOC" on a scientific bibliography management system such as bibsonomy.org. While the same content can be tagged by various users on the same system, anyone can use their own tags. Yet, most services suggest existing tags for a given item when someone begins tagging it.

The main advantage of tagging for end users is that one does not have to learn a predefined vocabulary scheme (such as a hierarchy or taxonomy) and one can use the keywords that fit exactly with his or her needs. Web sites that support tagging benefit from the "wisdom of the crowds" effect. Tags evolve quickly according to the needs of the users, and these tags, combined with the tagging actions and the frequency with which they are used, lead to the emergence of a folksonomy, that is, a user-driven, open and evolving classification scheme. Moreover, tags can be used for various purposes, and Golder and Huberman [24] have identified seven different functions that tags can play for end users, from topic definition to opinion forming and even self-reference.

In spite of its advantages when annotating content, tagging leads to various issues in information retrieval. Since a single tag can refer to various concepts, it can lead to ambiguity. For instance, "paris" can refer to a city in France, a city in the USA or even a person. Moreover, various tags can be used to define the same idea, so that a user must run various queries to get the content related to a given concept. Such heterogeneity is mainly caused by the multilingual nature of tags (e.g., "Semantic Web" and "Web semantique") but also due to the fact people will use acronyms or shortened versions ("sw" and "semweb"), as well as linguistic and morphosyntactic variations (synonyms, plurals, case, etc.). Finally, since a folksonomy is essentially a flat organization of tags, the lack of relationships between tags makes it difficult to suggest related content, especially when there is a gap of expertise between people tagging content and the ones looking for it. Someone searching for the tag "Semantic Web" will not easily be able to find content tagged with "RDFa," even though there is a clear relationship between both.

## 2.8 Adding Semantics to Tags and Related Objects

Numerous works related to the links between tags, the tagging process, folksonomies, and the Semantic Web have been published during the last couple of years. We can divide these into two general approaches: the ones aiming to define, mine, or automatically link to ontologies from existing folksonomies, and works based on defining Semantic Web models for tags and related objects (e.g., tagging, tag clouds, etc.). Again, the border between both is not very precise and some approaches combine both.

The first set of approaches is based on the idea that emergent semantics naturally appears through the use of tags, relying on various methods to achieve this goal. For example, Specia and Motta [43] combine automatic tag filtering, clustering, and mapping with ontologies already available on the Web to extract ontologies from existing folksonomies in a completely automated approach. Another approach involving a social aspect is the one defined by Mika [38], which uses SNA to extract ontologies from the Flickr folksonomy, based on the way that the community shares and uses tags.

Regarding the second approach, various models have been proposed to define Semantic Web vocabularies for tagging. Representing tags using Semantic Web technologies offer various advantages: providing a uniform, machine-readable and extendable way to represent tags as well as other concepts such as tagging actions, tag clouds, the relationships between tags and the meanings that they carry. While tag-based search is the only way to retrieve tagged content at the moment (and leads to the aforementioned problems), these new models allow advanced querying capabilities such as “retrieve all the content tagged with something relevant to the Semantic Web field” or “give me all the tags used by Bob on Flickr and Alice on del.icio.us.” Moreover, having tags and tagged content published in RDF allows one to easily link to it from other Semantic Web data, and to reuse it across applications.

The Tag Ontology[26] provides an initial model to represent tags and tagging actions in RDF, based on the ideas of Gruber [25] and on a common mathematical model of tagging that defines it as a tripartite relationship involving a “Tag,” a “User,” and a tagged “Resource.” This ontology defines the Tag class by subclassing skos:Concept, which means that each tag has a given URI (Uniform Resource Identifier). This offers the ability to interlink tags together with semantic relationships, as this model permits. SCOT (Social Semantic Cloud of Tags) [29] aims to represent tag clouds, and so defines a model to represent the use and co-occurrence of tags on a given social platform, allowing one to move his or her tags from oneservice to another and to share tags with others. Finally, MOAT (Meaning of a Tag) [40] aims to represent the meaning of tags using URIs of existing domain ontology instances from existing public knowledge bases (such as Geonames, a geographical database, or DBpedia, a data set of structured information extracted from Wikipedia), thus creating a bridge between folksonomies and existing ontologies or knowledge bases. It also provides a framework using this model, the goal of which is to let people easily bridge the gap between simple free-text tagging and semantic indexing.

Some tools already used some of these models to provide advanced and more precise querying tag-based capabilities to their users, including gnizr (a tag-sharing application), SweetWiki (a wiki engine), int.ere.st (a tag-sharing service based on SCOT), and LODr[27] (a tag aggregation and interlinking application based on MOAT).

## 2.9 Software Project Descriptions

Software descriptions are also required for the embeddable applications or “widgets” that are now proliferating many of the big social networking Web sites. Third-party developers are now creating their own applications that can be added

[26] http://www.holygoat.co.uk/projects/tags/.
[27] http://lodr.info.

by users to their own social networking profiles. For example, a user may choose to add a widget to their profile showing a map of places they have visited in the world, or enabling some other functionality which may not be natively offered by the Social Web site. Soon after Facebook added a developer's interface to their site, 4000 third-party applications had been made available and 70,000 developers had signed up to the developer community. Facebook's active user count also jumped 70% in the 4 months after this contributable application layer was added. In parallel, Google has initiated the OpenSocial project,[28] which allows developers to create application widgets that can be deployed across a range of OpenSocial-enabled social networking sites. However, there is an important question in relation to these widgets: how does one trust the source of an application? For example, does a user have to browse the complete source code (as a developer would), or can they just rely on some social networking aspect, that is, trusting applications from people they know?

Before widgets, many applications were already produced on the Web, mainly from open-source developer communities. In these communities, the social aspects of software project hosting and directory services are present but may not be immediately obvious. Web sites like SourceForge,[29] Savannah,[30] or BerliOS Developer[31] offer tools for developers to manage their projects (source code repositories, versioning, FTP space, etc.); Freshmeat[32] or Ohloh[33] allow them to reference and give visibility to their projects; and Slashdot[34] provides the latest "hot" news from the developer community and information on some projects. Yet, as with many Social Web sites, one problem is that developers must subscribe to each hosting Web site independently, filling in their personal details on each one, and entering their project description again and again on each directory-like Web site.

Beyond project hosting, these Web sites generally offer various social interaction tools for project tracking (such as blogs, wikis, and mailing lists) which can provide a social aspect to a software project. Thus, while the software development itself does not necessary involve a social aspect (for instance, source code write access might be delegated to only a restricted of users), users can be part of the process, for instance by reporting bugs and participating on the mailing list, answering blog posts, or editing a project wiki page to suggest new functionalities. Software development can thus benefit from the participation of online communities in the

[28] http://opensocial.org/.
[29] http://sourceforge.net/.
[30] http://savannah.gnu.org/.
[31] http://developer.berlios.de/.
[32] http://freshmeat.net/.
[33] http://ohloh.net/.
[34] http://slashdot.net/.

development process, even if users are not directly "in touch" with the source code itself. Moreover, if those tools are not provided by the project hosting service itself, developers can easily set them up using freely available tools on the Web.

## 2.10 Adding Semantics to Software Project Descriptions

As for blogs and wikis, a project description that describes a software application usually depends on the Web site it has been created on. There is thus a need for a common metadata modeling scheme for describing applications, in order to provide a unified way to represent it wherever it comes from.

DOAP[35] (Description of a Project) is an RDF vocabulary that aims to achieve this goal. It defines a "Project" class with various properties, such as its maintainers, its license, subversion access, etc. Moreover, since it is RDF-based, DOAP can be reused with existing vocabularies. In particular, from a social networking point of view, DOAP can be linked to FOAF to specify the developers of a project (with their associated identifying URIs) rather than just having a plain-text name, which can often raise ambiguity or heterogeneity problems.

If a user decides to install a widget or application on their social networking service, they usually have to trust some third-party service that may provide them with a certificate which they can decide whether to trust or not. An alternate approach is to leverage the social graphs of publishers and consumers of application widgets. Let us suppose someone writes a Facebook or OpenSocial widget and they want to distribute it, using this new approach. A user may choose to trust applications written by people connected to them in their (distributed) social graph by no more than two degrees of separation.

It is possible to use semantics to represent the various parts required in this scenario: FOAF can be used to describe people and their (distributed) social graph; while DOAP can be used to describe software projects, with the widget or application as a component of this software projects. We then connect the application project and the person together using FOAF–DOAP relationships.

By using such representations, the social graph (that is used here to determine whether to install a widget or not) does not have to be locked into one site, but rather can be distributed across any site that can be part of the larger interconnected social graph. As long as a publisher is part of the FOAF network, they do not even have to be on the particular social networking service where you install the application. This

[35] http://doap-project.org/.

means that one can trust an OpenSocial widget on one social networking site if its author is someone he or she knows on another Social Web site, where both sites have representations on the Semantic Web.

## 3. Producers of Social Semantic Data

Applying Semantic Web technologies to online social spaces allows for the expression of different types of relationships between people, objects and concepts. By using common, machine-readable ways of expressing individuals, profiles, social connections, and content, they provide a way to interconnect people and objects on the Web in an interoperable, extensible way.

On the conventional Web, navigation of social data across sites 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 may have a different network of friends on each. The information existing in these spaces is generally disconnected, lacking in semantics, and centrally controlled by single organizations. Individuals generally lack control or ownership of their own data.

Social spaces on the Web are becoming bigger and more distributed. This presents new challenges for navigating such data. Machine-readable descriptions of people and objects, and the use of common identifiers, would allow for linking diverse information from heterogeneous SNSs. This would create a starting point for easy navigation across the information in these networks.

The use of common formats allows interoperability across sites, enabling users to reuse and link to content across different platforms. This also provides a basis for data portability, where users could have ownership and control over their own data and could move profile and content information between services as they wish. Recently, there has been a push within the Web community to make data portability a reality.

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, links, blogs posts, etc., social networks and semantics can complement each other. Already within online communities, common vocabularies or folksonomies for tagging are emerging through of a consensus of community members.

There are also a number of semantically enabled social applications appearing that have been enhanced with extra features due to the rich content being created in social software tools by users. The Twine application from Radar Networks is a recent example of a system that leverages both the explicit (tags and metadata) and implicit semantics (autotagging of text) associated with content items. Twine is 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 containers through a variety of methods. The underlying semantic data can be exposed as RDF by appending "?rdf" to any Twine URL. The DBpedia represents structured content from the collaboratively edited Wikipedia in semantic form, leveraging the semantics from many social content contributions by multiple users. DBpedia allows you to perform semantic queries on these data, and enables the linking of this socially created data to other data sets on the Web by exposing it via RDF. Revyu.com combines Web 2.0 interfaces and principles such as tagging with Semantic Web modeling principles to provide a reviews Web site that is integrated with linked data principles—a set of best practice guidelines for publishing and interlinking pieces of data on the Semantic Web. Anyone can review objects defined on other services (such as a movie from DBpedia), and the whole content of the Web site is available in RDF, therefore it is available for reuse by other applications.

## 3.1 FOAF

Semantic Web technologies allow for a more expressive description of a social network, enabling the use of heterogeneous nodes and link denoting different types of objects and different types of relationships. This enables us to express a model of an object-centered network where content and other items of interest can be described along with people.

The FOAF project was started 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. Anyone can create their own FOAF file describing themselves and their social network, and the information from multiple FOAF files can easily be combined to obtain a higher-level view of the network across various sources, as shown in Fig. 4. 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, SKOS, etc. Some prominent social networking services that expose data using FOAF include Hi5 (a social networking site), LiveJournal (a social networking and blogging community site), Vox (a social networking and blogging service), Pownce (a social networking and microblogging site), and MyBlogLog

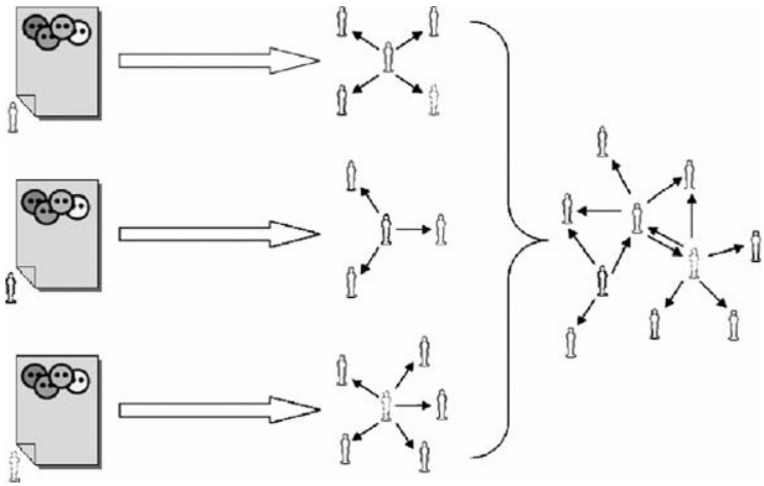

FIG. 4. Integrating social networks by using FOAF as a common representation format and having unique URIs for people.

(an application which adds community features to blogs). People can also create their own FOAF document and link to it from their homepage, and exporters are available for some major Social Web sites as Flickr, Twitter (a microblogging service), and Facebook. Such FOAF documents usually contain personal information, links to friends, and other related resources. The structure of the social network formed by relations expressed in FOAF documents on the Web has been studied in Ding et al. [16], particularly the small-world characteristics of the graph.

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

```
<foaf:Person rdf:about="#JB">
    <foaf:name>John Breslin</foaf:name>
    <foaf:mbox rdf:resource="mailto:john.breslin@deri.org"/>
    <foaf:homepage rdf:resource="http://www.johnbreslin.com/"/>
    <foaf:nick>Cloud</foaf:nick>
    <foaf:depiction rdf:resource="http://www.johnbreslin.com/
      images/foaf_photo.jpg"/>
    <foaf:interest>
         <rdf:Description rdf:about="http://dbpedia.org/
         resource/SIOC" rdfs:label="SIOC"/>
    </foaf:interest>
```

```
    <foaf:knows>
          <foaf:Person>
            <foaf:name>Sheila Kinsella</foaf:name>
            <foaf:mbox rdf:resource="mailto:sheila.kinsella
            @deri.org"/>
          </foaf:Person>
    </foaf:knows>
    <foaf:knows>
          <foaf:Person>
            <foaf:name>Stefan Decker</foaf:name>
            <foaf:mbox rdf:resource="mailto:stefan.decker
            @deri.org"/>
          </foaf:Person>
    </foaf:knows>
</foaf:Person>
```

There have been a lot of complaints in recent years about the walled gardens that are social network sites. Some of the most popular SNSs would not exist without the walled garden approach, but some flexibility would be useful. Users may have many identities on different social networks, where each identity was created from scratch. A reusable profile would allow a user to import their existing identity and connections (from their own homepage or from another site they are registered on), thereby forming a single global identity with different views.

"Social network portability" is a related term that has been used to describe the ability to reuse one's own profile across various social networking sites and applications. The founder of the LiveJournal blogging community, Brad Fitzpatrick, wrote an article[36] in August 2007 from a developer's point of view about forming a "decentralized social graph," which discusses some ideas for social network portability and aggregating one's friends across sites. "The Bill of Rights for Users of the Social Web" was authored in September 2007 for Social Web sites who wish to guarantee ownership and control over one's own personal information.[37] Dan Brickley, the cocreator of the FOAF vocabulary, wrote an article entitled "The World is Now Closed" which talked about how SNSs should not define one's relationships in absolute terms and that even an aggregate social graph cannot be so clearly defined.[38]

The evolving need for distributed social networks and reusable profiles has been highlighted by several recent notable efforts. DataPortability[39] is a group whose aim

[36] http://bradfitz.com/social-graph-problem/.
[37] http://opensocialweb.org/2007/09/05/bill-of-rights/.
[38] http://danbri.org/words/2007/09/13/194.
[39] http://www.dataportability.org/.

is to advance standards enabling data sharing between services. DiSo (Distributed Social Networking applications) is a project from Google which aims to implement distributed social networks. Google's Social Graph API indexes publicly articulated social connections and allows users to view their social network across multiple services. These initiatives make use of existing and open standards like FOAF, microformats, and OpenID.

## 3.2 SIOC

The SIOC initiative is aimed at interlinking related online community content from platforms such as blogs, message boards, and other Social Web sites. 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" or hierarchical categories). 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.

Since disconnected Social Web sites require ontologies for interoperation, and due to the fact that there is a lot of social data with inherent semantics contained in these sites, there is potential for high impact through the successful deployment of SIOC. Many online communities still use mailing lists and message boards as their main communication mechanisms, and the SIOC initiative has created a number of data producers for such systems to lift these communities to the Semantic Web. As well as having applications to Social Web sites, there is a parallel lack of integration between social software and other systems in enterprise intranets. So far, SIOC has been adopted in a framework of 50 applications or modules[40] deployed on over 400 sites.

A sample fragment of SIOC RDF is shown below, representing a blog post, its metadata and associated follow-up comments.

```
<sioc:Post rdf:about="http://johnbreslin.com/blog/2006/09/
  07/creating-connections-between-discussion-clouds-with-
  sioc/">
    <dc:title>Creating connections between discussion clouds
      with SIOC</dc:title>
```

[40] http://rdfs.org/sioc/applications.

```
    <dcterms:created>2006-09-07T09:33:30Z</dcterms:created>
    <sioc:has_container rdf:resource="http://johnbreslin.
      com/blog/index.php?sioc_type=site#weblog"/>
    <sioc:has_creator>
        <sioc:User rdf:about="http://johnbreslin.com/blog/
        author/cloud/" rdfs:label="Cloud">
            <rdfs:seeAlso rdf:resource="http://johnbreslin.
            com/blog/index.php?sioc_type=user&sioc_id=1"/>
        </sioc:User>
    </sioc:has_creator>
    <sioc:content>SIOC provides a unified vocabulary for
      content and interaction description: a semantic layer
      that can coexist with existing discussion platforms.
      </sioc:content>
    <sioc:topic rdfs:label="Semantic Web" rdf:resource="http://
      johnbreslin.com/blog/category/semantic-web/"/>
    <sioc:topic rdfs:label="Blogs" rdf:resource="http://
      johnbreslin.com/blog/category/blogs/"/>
    <sioc:has_reply>
        <sioc:Post rdf:about="http://johnbreslin.com/blog/
        2006/09/07/creating-connections-between-discussion-
        clouds-with-sioc/#comment-123928">
            <rdfs:seeAlso rdf:resource="http://johnbreslin.com/
            blog/index.php?sioc_type=comment&sioc_id=123928"/>
        </sioc:Post>
    </sioc:has_reply>
</sioc:Post>
```

So far, work on SIOC has focused on producing social semantic data, but the augmentation of these data with rules to aid with reasoning is the next step (e.g., as discussed in Aleman-Meza et al. [2] by members of the ExpertFinder initiative,[41] a project to improve publication of metadata on Web pages to help automated identification of experts on particular topics). By combining information from one's explicitly defined social network and from implicit connections that may be derived through common activities (e.g., commenting on each other's content, participating in the same community areas), the suggestion of experts can be enhanced. An interesting aspect of SIOC is that it goes beyond pure Web 2.0

[41] http://expertfinder.info/.

services and can be used in other use cases involving the need to model social interaction within communities, either in corporate environments,[42] or for argumentative discussions [33] and scientific discourse representation, as illustrated by recent efforts[43] to align SIOC and SWAN[44] (Semantic Web applications in neuromedicine).

## 3.3 DOAP

As introduced in the previous section, the DOAP project provides an RDFS (RDF Schema) vocabulary for defining metadata related to software projects. As with FOAF and SIOC, this is a lightweight vocabulary, and this makes it easy for software developers who want to provide open and common descriptions of their projects using Semantic Web technologies. For instance, the next snippet of code identifies metadata about the SIOC PHP API, defined as an instance of a doap: Project, and assigned a specific URI.

```
<doap:Project rdf:about="http://sw.deri.org/svn/sw/2005/08/
  sioc/phpapi/doap.rdf#siocexportapi">
    <doap:name>SIOC PHP Export API</doap:name>
    <doap:shortname>sioc-export-api</doap:shortname>
    <doap:shortdesc xml:lang="en">PHP API to create SIOC
      exporters</doap:shortdesc>
    <doap:description xml:lang="en">SIOC PHP Export API pro-
      vides an easy to write SIOC exporters for any PHP
      application.</doap:description>
    <doap:homepage rdf:resource="http://esw.w3.org/topic/SIOC/
      PHPExportAPI"/>
    <doap:download-page rdf:resource="http://esw.w3.org/topic/
      SIOC/PHPExportAPI"/>
    <doap:programming-language>PHP</doap:programming-
      language>
    <doap:license rdf:resource="http://usefulinc.com/doap/
      licenses/gpl"/>
    <doap:maintainer rdf:resource="http://apassant.net/alex"/>
```

[42] http://www.w3.org/2001/sw/sweo/public/UseCases/EDF/.
[43] http://esw.w3.org/topic/HCLSIG/SWANSIOC.
[44] http://swan.mindinformatics.org/.

```
    <doap:maintainer rdf:resource="http://captsolo.net/semweb/
      foaf-captsolo.rdf#Uldis_Bojars"/>
    <doap:developer rdf:resource="http://apassant.net/alex"/>
    <doap:developer rdf:resource="http://captsolo.net/semweb/
      foaf-captsolo.rdf#Uldis_Bojars"/>
    <doap:repository>
        <doap:SVNRepository>
          <doap:location  rdf:resource="http://sw.deri.org/
          svn/sw/2005/08/sioc/phpapi/"/>
        </doap:SVNRepository>
    </doap:repository>
</doap:Project>
```

While DOAP descriptions can be created by hand, various DOAP exporters for major free software development Web sites have been written by developers (see also the RDF exporter for Ohloh[45]). These exporters allow software metadata to be available on the Web, described in a uniform way using the DOAP vocabulary (rather than just being embedded in Web pages which makes it difficult for automatic reuse by software agents).

As one can see in the above example, there are various ties between FOAF and DOAP. Since any project can have various developers or maintainers, DOAP offers the ability to use not only a name to define an author, but their URI, that is, his or her identifier on the Semantic Web, generally associated with a FOAF profile. Thanks to URI identification, and in spite of the fact that these profiles are distributed on the network, the software graph (DOAP), the identity graph (FOAF), and even the content graph (SIOC) can be connected together, providing a complete overview of the online activity and identity of people working on a given project. For instance, Fig. 5 shows how different graphs, related mainly to FOAF, SIOC, and DOAP can interact together to provide a complete Semantic Web description of a network, a widget description and a related blog post by various people, in a distributed but interlinked way.

Moreover, projects can have various topics. Here, once again, instead of relying on text strings, people can use URIs and properties from Dublin Core, a vocabulary for information resource description, to define project topics in a machine-understandable way. A good practice would be to use URIs of topics as defined on DBpedia, or other data sets from the Linked Open Data movement to make open data sets available in RDF format. The link between the project and a topic URI can

[45] http://rdfohloh.wikier.org/.

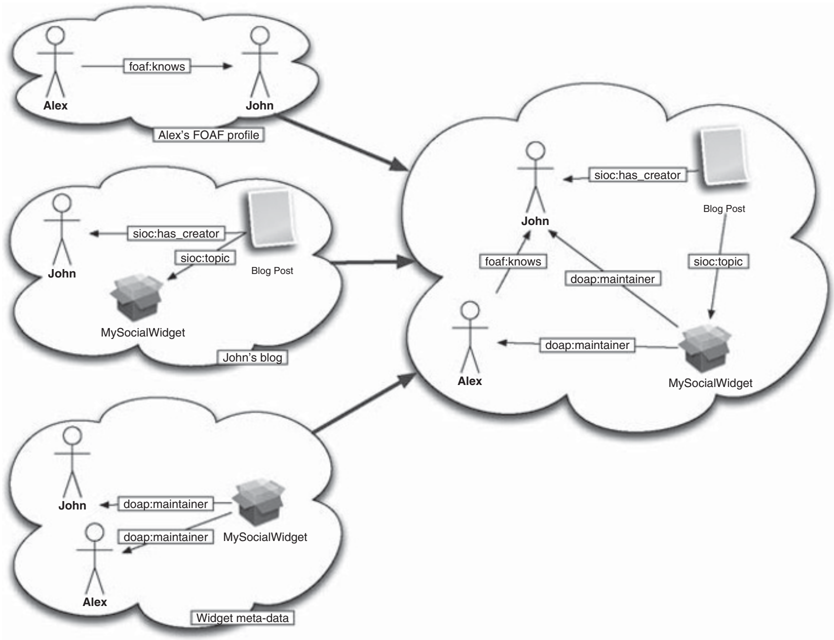


FIG. 5. Linking people, content, and social widgets across Web sites.

be defined directly by the project's author, may be extracted from the project's textual description using an NLP (natural language processing) algorithm, or can by added by the author via free-text keyword tagging using MOAT as explained earlier. For instance, since our example project is related to Semantic Web technologies, and particularly to the SIOC vocabulary, the following code mentions the links between the project and those topics, uniquely identified with their DBpedia URIs.

```
<doap:Project rdf:about="http://sw.deri.org/svn/sw/2005/08/
  sioc/phpapi/doap.rdf#sioc exportapi">
    <dc:subject rdf:resource="http://dbpedia.org/resource/
      Semantic_Web"/>
    <dc:subject rdf:resource="http://dbpedia.org/resource/
      SIOC"/>
</doap:Project>
```

Once again, and with reference to the earlier tagging section, expressing these URIs offers new capabilities regarding information exchange and modeling (we will also exemplify this later).

# 4. Collectors of Social Semantic Data

The semantic social data available on the Web are distributed across numerous sources and are stored in many different formats. In some cases, these data may be published in such a way that it can be consumed directly by applications, for example, in an RDF store with a SPARQL (Simple Protocol and RDF Query Language) endpoint. Alternatively, it may be necessary to first gather and process the data, for example, when it is stored in documents which need to be crawled and indexed. In the following, we describe issues with interpreting social data from mined the Web, inferring relations from semantic data, and technical aspects of collecting data.

## 4.1 The Web as a Source of Social Network Data

Common traditional methods of collecting social network information include administering questionnaires, conducting interviews or performing observational studies, and studying archival records. There are some fundamental differences between the networks acquirable by these methods and the networks retrievable from the Internet. Extracting data from the Web present a different set of challenges but also offer some advantages over traditional methods.

A major advantage of mining online social networks for analysis is the much lower cost of acquiring data due to the reduced time and effort involved. Also, the scale of the social information available online is unprecedented. In the past, acquisition of social network data of the order of millions of nodes would have been impossible; with the social data now freely available on the Internet, it is easy. In addition, networks collected from the Web are evidence-based and objective. Unlike interviews or questionnaires, results are not dependent on the accurate recall of the subjects, who may interpret questions differently, or may be unwilling to cooperate. Furthermore, while it is unlikely you will get a 100% participation rate in a survey, especially on a large network, if you have access to a full Web data set you can analyze a whole network. Finally, electronic data collection easily enables longitudinal studies, allowing the dynamics of networks to be investigated, as opposed to surveying, where repeated data collection would be time-consuming and maybe impossible if the subjects are unwilling or unable to repeat the survey.

However, the accuracy of social network data mined from the Internet can be highly questionable. People can easily misrepresent themselves or others. Depending on Internet usage habits, some people will have far more information available about them online than others. This means that the social networks extracted from the Web may not give a balanced representation of real-life social networks. There is also the question of how exactly to interpret information from the Internet, for example, the strength of the relationship implied. The people on an individual's contact list on an

SNS may encompass a spectrum from close friends to distant acquaintances or even strangers. Another problem is that there are likely to be errors in Web data, for example, resulting from typos, inconsistent spelling of names, and variations on names.

Semantic Web technologies can greatly assist the process of harvesting social networks. The use of common, structured formats means that social network data can easily be aggregated from multiple, heterogeneous sources. References to the same person or resource can be identified across multiple sources and consolidated. Much of the effort needed to construct a model of a social network is removed and the need for human effort is lessened. It is possible to do reasoning on the data and infer relations from certain properties. Additionally, it is possible to extract a network of typed nodes and links.

Harvesting and analyzing social data from the Web raises important ethical issues. It involves using data for purposes which were not intended by the users who uploaded for their use and that of their friends. Trust and provenance of information are important aspects that should be taken into consideration. At a technical level, the ability to confirm the origin of data is important, and at a more social level, a means to express trust in sources is also required [14, 23]. We believe that advanced policies are also needed to let users define who can access which part of their social data, and to which extent it can be reused.

## 4.2 Collecting and Aggregating Data

Data on the Semantic Web are published in different ways, so different methods may be required to collect them. Additional processing may also be required to merge data from multiple sources.

**Crawling.** Due to the linked nature of social networks, given URIs to seed members of the network, we can follow links from these nodes to their friends, and then their friends-of-friends and so on. This can be done by simply following rdf:seeAlso links. Additional knowledge about the structure of the data can be used to improve the task. For example, the SIOC Crawler [5] uses knowledge of the ontology's structure to incrementally retrieve new SIOC data in threads. For widgets and project descriptions, crawling is also important since there is a need to easily find a software project without having to manually browse the complete Semantic Web. We will later detail an architecture that can be used to achieve this goal. To ease the crawling of published data, site suppliers can provide a semantic sitemap[46] on their Web site, so that crawling agents know where to find related RDF data.

**Exporters.** For some platforms, exporters are available which generate a structured RDF representation of the data. These allow information in a relational database or

[46] http://sw.deri.org/2007/07/sitemapextension/.

other structured stores to be automatically transformed into RDF. Exporters make it easy for users to maintain semantic representations of their data. For example, there are SIOC exporters available for platforms including mailing lists [17], Web forums and blogs [12], and existing Web 2.0 services such as Flickr.

**Object consolidation.** An important task in extracting social data from the Web is merging identifiers of equivalent instances occurring across different sources. This involves identifying instances representing the same object, and unifying them into one entity. Object consolidation (or "smushing") can be performed for instances which share the same value for inverse functional properties or IFPs [37], for example, using foaf:mbox.[47] Another option is to provide explicit identification using instances of the OWL (Web Ontology Language) sameAs property between various resources that identify the same person or data, in spite of different URIs. This best practice allows one to unify all of their identities from various exporters (e.g., Flickr, Twitter, Facebook, etc.) and to then query their complete social network with a single entry point, as Fig. 6 shows. Finally, it can also be achieved

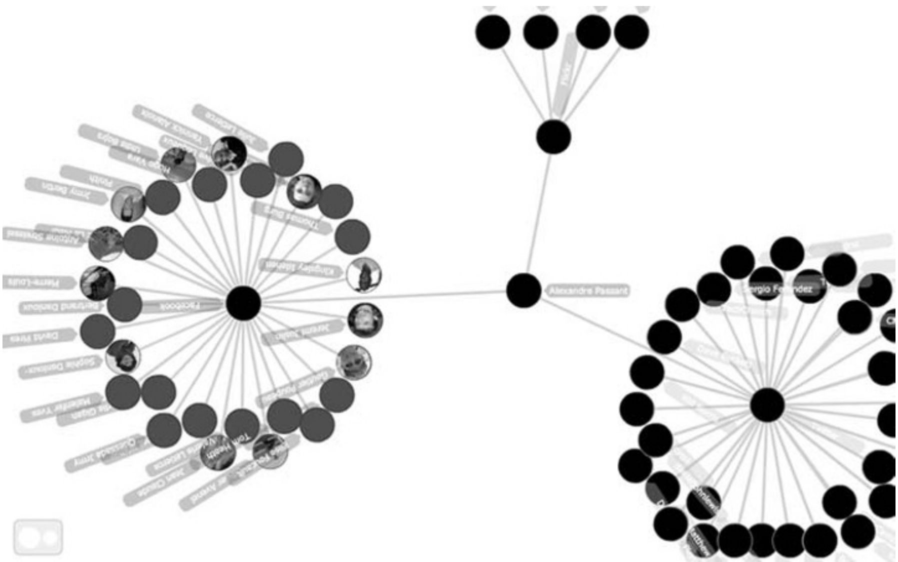

FIG. 6. Identity consolidation and social network browsing using data exported from various social Web sites.[48] (Partial view of the social network.)

[47] Defining a property as inverse functional (owl:InverseFunctionalProperty) implies that if two resources share the same value for that property, they are the same even if they have different URIs. FOAF defines various IFPs (foaf:mbox, foaf:opened).

[48] http://apassant.net/home/2008/01/foafgear.

by considering various alternative criteria and if a certain threshold is reached in similarity between two instances, they can be considered equal [1]. Yet, while one can define such rules within his or her own restricted social graph, it may lead to unexpected results on the complete Web (for instance, since different people will sometimes have the same name) and identity management on the Semantic Web is a vast research topic.

## 4.3 Crawling and Browsing Software Descriptions

As with FOAF profiles or any RDF data, DOAP files may be distributed over the network, which can make it difficult for end users or developers to discover them. To solve this problem, an architecture was proposed by Bojārs et al. [6] involving various components acting together (1) a Firefox plug-in, called Semantic Radar, whose goal is to discover RDF documents from HTML pages (either using auto-discovery links or thanks to embedded RDFa); (2) a ping service for Semantic Web documents, called PingTheSemanticWeb[49] (PTSW), which stores a fresh listing of RDF files it has received pings about; and (3) a collaborative and open directory of DOAP projects, called doap:store.[50] In fact, while all of these components were developed separately, they all act with each other to provide a complete Semantic Web food chain.

When people browse the Web using Semantic Radar, the plug-in sends a ping to PTSW each time an RDF file is found. PTSW then stores a link to this RDF file in its database, and provides a list of pinged documents to developers (which may then be organized by type). In this system, discovering documents and storing pings is not only dedicated to DOAP, but can be useful for people who are looking for FOAF or SIOC files. Finally, in this architecture, doap:store fetches the list of new DOAP files on a regular basis to provide a directory of DOAP projects that can then be queried and browsed. doap:store was one of the first tools to use this architecture, but anyone can benefit from it, by focusing on creating the application rather than finding and crawling the data. An interesting point in this workflow is the social process it involves. Since anyone can contribute just by browsing the Web, this means that any user can be a part of the Semantic Web document discovery process, weaving the "architecture of participation" principle from Web 2.0 into the Semantic Web.

[49] http://pingthesemanticweb.com.
[50] http://doapstore.org.

## 4.4 Inferring Relationships from Aggregated Data

The simplest way of extracting a social network from the Web is to look at explicitly stated connections. Social networking sites and other types of social software allow users to express lists of friends. Blogging platforms may allow users to add a blogroll which is a list of favorite blogs. Depending on the platform, these connections may indicate a directed or undirected link between users. For example, blogroll links are frequently unreciprocated, and are therefore directed, but many SNSs require both users to consent to the link, creating undirected ties. A sample query for extracting the social network formed by explicit foaf:knows relationships follows using the SPARQL query language.

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf:<http://xmlns.com/foaf/0.1/>

SELECT ?s ?o
WHERE {
    ?s rdf:type foaf:Person.
    ?o rdf:type 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. Links between people may be inferred due to links to some common objects, for example, appearing in the same pictures, tagging the same documents, and replying to each other's blog posts. These connections indicate relationships of varying strengths—for example, email communication may be interpreted as stronger evidence of a real tie than the case of one person replying to another's blog post. Co-occurrence of names in documents would be an even weaker sign of a relation. A sample query for extracting the implicit social network formed by replies to posts follows.

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX sioc:<http://rdfs.org/sioc/ns#>
PREFIX foaf:<http://xmlns.com/foaf/0.1/>

SELECT ?author1 ?author2
WHERE {
    ?post1 rdf:type sioc:Post.
    ?post1 foaf:maker ?author1.
    ?post1 sioc:has_reply ?post2.
    ?post2 rdf:type sioc:Post.
    ?post2 foaf:maker ?author2.
}
```

Instead of running queries to retrieve those implicit relationships, we can define rules to make them explicit and to state the acquaintance of users on a Weblog. For instance, we can consider that there is a formal agreement relationship between two users (modeled with an arg:agreedWith relationship) as soon as one replies to a post from the other one using "I agree" in his or her answer.[51] To model this rule, we rely here on the SPARQL CONSTRUCT pattern, which can be used to produce new statements from existing ones. Thus, we can apply the following query on our triple store, and then put the created RDF graph in the store itself, so that the relationship will become explicit. The produced statements may then be used to extract a more precise social network within a blogging community when querying data.

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX sioc:<http://rdfs.org/sioc/ns#>
PREFIX foaf:<http://xmlns.com/foaf/0.1/>

CONSTRUCT {
    ?author2 arg:agreedWith ?author1.
} WHERE {
    ?post1 rdf:type sioc:Post.
    ?post1 foaf:maker ?author1.
    ?post1 sioc:has_reply ?post2.
    ?post2 rdf:type sioc:Post.
    ?post2 foaf:maker ?author2.
    FILTER REGEX(?post2, "I agree", "i").
}
```

While the above examples result in simple networks of people and untyped ties, more complex social networks consisting of multiple node and link types can also be studied. These examples are only possible through linking people and content in and across sites. Traditional, nonsemantic queries like in SQL would be limited to one site and would require some kind of join on a user/content table. However, the use of shared semantically rich vocabularies makes it possible to perform operations like these on data originating from many different sources.

## 5. Consumers of Social Semantic Data

Once data have been collected and aggregated, or made directly accessible through a SPARQL endpoint, it can be studied or used in applications. As the information is in a structured format, it can easily be converted into the formats required by popular

[51] Ideally, more advanced pattern matching and NLP methods should be used to define agreement between two users on a Weblog.

social network analysis and visualization tools. RDF data can also be queried directly to return some set of items that fit certain criteria that a user is interested in. In the following, we describe these two ways of using semantic social data.

## 5.1 Social Network Analysis

SNA uses methods from graph theory to study networks of individuals and the relationships between them. The individuals are often referred to as nodes or actors, and they may represent people, groups, countries, organizations, or any other type of social unit. The relations between them can be called edges or ties, and can indicate any type of link, for example, acquaintance, friendship, coauthorship, and information exchange. Ties may be undirected, in which case the relationship is symmetric, or directed, in which case the relationship has a specific direction and may not be reciprocated.

The nodes in a social network can be seen as analogous to entities in an RDF graph, where a <subject, predicate, object> triple indicates a directed tie from the subject node to an object node, and the predicate indicates the type of the relationship. While SNA methods are generally applied to social networks, they can be used to analyze any kind of networked data.

We can apply mathematical measures from SNA to get interesting information about a social network. The more complex methods of network analysis cannot be performed directly on a graph in RDF format, but must be converted to a representation more suited to network analysis. An RDF graph can be loaded into a network analysis program such as Pajek or UCINET [9] which can perform various measures and visualizations. Alternatively, a library like JUNG [39], which provides analysis and visualization methods, can be used to develop custom analytic or visual tools.

**Locating important individuals.** Centrality measures can be used to locate key players in a network [44]. Degree centrality is based on the number of connections a person has. This measure locates individuals who are connected to a large number of others. In a directed graph, indegree is the number of incoming connections and outdegree is the number of outgoing connections. Closeness centrality is calculated based on the total shortest distance to all other nodes in the network. This measure can be an indicator of people who can most quickly communicate information to the whole network. Betweenness centrality is based on the number of shortest paths on which a node lies. A node which scores highly according to this metric may occupy a strategic position and function as a bridge between different parts of the network. Flink [37] applies these measures to a social network of Semantic Web researchers in order to investigate whether the network position of a scientist is related to their performance.

**Extracting communities.** We may be interested in finding subgraphs or small communities within a larger graph. This enables the restriction of network to a manageable size for performing further analysis. Algorithms exist for partitioning a network into different groups, for example, that of Girvan and Newman [21]. Alternatively, if there is a particular individual of interest we can extract their ego network, the area of the graph focused around them. For example, spreading activation algorithms can activate an input node or nodes, and propagate the activation from these to locate those individuals which are most strongly connected and therefore receive the most activation [30].

**Characterizing a social network.** There are some interesting whole network properties that can be investigated to gain an understanding of the overall structure of the network [44]. Centralization measures the degree to which the network has a leader. Cohesiveness measures the well-connectedness of the network. These measures can also be used to make comparisons between different networks.

**Visualizing a social network.** By creating a pictorial image of a social network, it may be possible to get an improved insight into the structure of the graph. A visual representation can help analysts to understand the network better themselves, and also aid in explaining features of the network to others [19]. Flink provides visualizations of the ego networks of individual researchers and allows users to browse members of the Semantic Web research community.

## 5.2 Querying an RDF Graph

By representing social data in RDF and putting it in a store with a SPARQL endpoint (i.e., an access point where remote SPARQL queries can be run via HTTP), we can perform queries to extract interesting information about users, communities, and content. In the following, we discuss some example scenarios and illustrate them with sample queries.

**Finding a person's ego network.** Identifying an ego-centric network centered around a focus person involves finding all people to whom they are connected to online. This means searching over all their accounts, and across all SNSs of which they are a member. Below is a simple example query over FOAF data to get all friends of persons with a particular email address sha1sum. We use the hash of an email address as an identifier (since the foaf:mbox_sha1sum is defined as an owl: InverseFunctionalProperty in FOAF), as the focus person is likely to have different URIs on different sites.

```
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
SELECT DISTINCT ?o
WHERE {
```

```
    ?s foaf:mbox_sha1sum "9a348bd34fe67b15f388c95c2cb9b4bfc
      9073797".
    ?s foaf:knows ?o.
}
```

**Finding a person's implicit social links.** While locating a person's explicitly stated connections goes some way to locating their social network, they may have more acquaintances with whom they are implicitly linked. It is possible to identify additional potential acquaintances of a person via objects to which they are both connected. The example below shows a query to find all people with the same workplace, school, or project as the focus person. We could also consider people who are coauthors of some documents, or who have replied to each other's SIOC-enabled posts.

```
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
SELECT DISTINCT ?s
WHERE {{
    <http://sw.deri.org/~sheila/foaf.rdf#me>foaf:workplace-
      Homepage ?o.
    ?s foaf:workplaceHomepage ?o.
} UNION {
    <http://sw.deri.org/~sheila/foaf.rdf#me>foaf:schoolHo-
      mepage ?o.
    ?s foaf:schoolHomepage ?o.
} UNION {
   <http://sw.deri.org/~sheila/foaf.rdf#me>foaf:project?o.
    ?s foaf:project ?o.
}}
```

We can carry out simple reasoning by expressing a set of rules to describe when such implicit links create a social connection between people and when they may not. For example, we may decide that two people are socially connected if one posts a comment on someone else's blog post; alternatively, we may conclude that a weak link exists if two people posted on the same lengthy discussion thread and that no social connection exists.

**Aggregating a person's Web contributions.** This means retrieving content that a person has contributed to various sources on the Web; for example, all blog posts and comments on other blogs, chat logs, mailing list, and forum posts. This is a difficult problem to perform with a normal search engine as people may share their name with other people, or may use different account names on different sites. A sample query over SIOC data is shown below, to get all posts created by a particular user.

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
PREFIX sioc:<http://rdfs.org/sioc/ns#>
SELECT DISTINCT ?post
WHERE {
    ?post rdf:type sioc:Post.
    ?post sioc:has_creator <http://www.mindswap.org/blog/
      author/hendler/#foaf>.
}
```

Yet, since this query is based on a precise URI, it will not retrieve content created by the same user while using another URI (for instance, http://example.org/hendler). One option to retrieve this content is to define owl:sameAs statements between this URIs and other URIs of the same user, such as:

```
<http://example.org/hendler>owl:sameAs<http://www.mind-
  swap.org/blog/author/hendler/#foaf>.
```

Then, by adding these statements in the triple store that holds the data, and assuming it supports reasoning based on owl:sameAs, the query will also retrieve posts that have http://example.org/hendler as a sioc:has_creator.

A second way to do retrieve the person's contributions is to run the query not based on the URI, but based on an IFP, such as the foaf:mbox or foaf:openid. Since OpenID aims to become a standard for authentication on the Web, this can be a useful way to retrieve all the contributions of a given user no matter which Social Web site it comes from—providing the person signs in using the same OpenID URL—and this method is shown in the following query:

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
PREFIX sioc:<http://rdfs.org/sioc/ns#>
SELECT DISTINCT ?post
WHERE {
    ?post rdf:type sioc:Post.
    ?post sioc:has_creator ?user.
    ?user foaf:openid<http://example.org/hendleropenid>.
}
```

**Locating a community around a topic.** We may be interested in extracting a community centered around a certain topic, using tags, keywords, and other metadata to find people who are talking about a certain thing. The query below locates posts with the topic "Semantic Web" and returns the URIs of the authors of these posts.

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX sioc:<http://rdfs.org/sioc/ns#>
PREFIX foaf:<http://xmlns.com/foaf/0.1/>

SELECT DISTINCT ?author
WHERE {
    ?post rdf:type sioc:Post.
    ?post foaf:maker ?author.
    ?post sioc:topic ?post_topic.
    ?post_topic rdfs:label "semantic web".
}
```

Yet, this query will not retrieve posts written in French, for example, using a "Web semantique" string instead of the "Semantic Web" phrase. However, if people were encouraged to use a precise URI instead of the simple tag, such as http://dbpedia.org/resource/Category:Semantic_Web, we would then be able to retrieve all related posts. Moreover, using those URIs, we can run even more advanced queries, as in the example of retrieving all posts related to the Semantic Web, we could also show those for which the topic is directly related to this URI (e.g., RDFa, SKOS, etc.), as the following query does, emphasizing the benefits of combining data from various data sets, interlinked together in the whole Semantic Web graph.

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX sioc:<http://rdfs.org/sioc/ns#>
PREFIX foaf:<http://xmlns.com/foaf/0.1/>

SELECT DISTINCT ?author
WHERE {
    ?post rdf:type sioc:Post.
    ?post foaf:maker ?author.
    ?post sioc:topic ?topic.
    ?topic ?rel <http://dbpedia.org/resource/Category:
      Semantic_Web>.
}
```

As with the example queries in Section 4, the queries above can be performed on data originating from various diverse sources.

**Locating software projects from people you trust.** If we consider that a user will only trust software applications written by people that they have added as personal connections (represented on the Semantic Web using FOAF), the following query will retrieve projects in which one of the maintainers of a project is in their network, where the original user is identified with $uri:

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX sioc:<http://rdfs.org/sioc/ns#>
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
PREFIX doap:<http://usefulinc.com/doap/ns/doap#>

SELECT DISTINCT ?project ?friend
WHERE {
    ?project rdf:type doap:Project.
    ?project doap:maintainer ?friend.
    <$uri>foaf:knows ?friend.
}
```

Moreover, as explained earlier, instead of giving a URI, one can use an IFP to identify themselves, such as an email address or an OpenID URL.

A similar query can be used if one decides to trust not only their direct friends, but also their friends-of-friends as shown below, retrieving the project, its maintainer, and the person that acted as an intermediary connection:

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX sioc:<http://rdfs.org/sioc/ns#>
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
PREFIX doap:<http://usefulinc.com/doap/ns/doap#>

SELECT DISTINCT ?project ?friend ?friendofafriend
WHERE {
    <$uri>foaf:knows ?friend.
    ?friend foaf:knows ?friendofafriend.
    ?project rdf:type doap:Project.
    ?project doap:maintainer ?friendofafriend.
}
```

Moreover, the query could be extended to express various degrees of connectivity. The current SPARQL specification only allows node–arc–node queries, which means that for each desired path length, the query must be adapted. However, a SPARQL "path" extension like SPARQLer [32] can be used with appropriate SPARQL engines, allowing us to write queries like "find all projects from people I'm connected to via a path of between 1 and 3 (inclusive) foaf:knows relationships."

**Locating a software project related to a particular topic.** Similar to the earlier example of blog posts and associated topics, where projects are related to some topics using URIs rather than keywords, projects around a particular topic can easily be found. Once again, we show how various data sets interlinked with URIs in this "Giant Global Graph" can enable us to perform advanced queries. Moreover, this can be combined with a social networking aspect. The following query will retrieve all projects with a topic related to the Semantic Web created by people known to a user with the identifier $uri:

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX sioc:<http://rdfs.org/sioc/ns#>
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
PREFIX doap:<http://usefulinc.com/doap/ns/doap#>

SELECT DISTINCT ?project ?friend
WHERE {
    ?project rdf:type doap:Project.
    ?project doap:maintainer ?friend.
    ?project foaf:topic ?topic.
    ?topic ?rel <http://dbpedia.org/resource/Category:
      Semantic_Web>.
    <$uri>foaf:knows ?friend.
}
```

# 6. Future Work

## 6.1 Leveraging Semantics in Multimedia-Enabled Social Web Sites

A key feature of the new Social Web is the change in the role of user from just a consumer of content, to an active participant in the creation of content. For example, Wikipedia articles are written and edited by volunteers; Amazon.com uses information about what users view and purchase to recommend products to other users; Slashdot moderation is performed by the readers. One area of future work in relation to social networks on the Semantic Web is the application of semantic techniques to take even more advantage of community input to provide useful functionality. As an example, we will look at the area of multimedia management.

There is an ever increasing amount of multimedia of various formats becoming available on the Internet. Current techniques to retrieve, integrate, and present these media to users are deficient and would benefit from improvement. Semantic technologies make it possible to give rich descriptions to media, facilitating the process of locating and combining diverse media from various sources. Making use of online communities can give additional benefits. Two main areas in which social networks and semantic technologies can assist multimedia management are annotation and recommendation. Some efforts such as DBTune[52] already provide musical content exported to the Semantic Web, and recent work has been done in order to use that interlinked musical content for music-based recommendations [41].

Social bookmarking systems like del.icio.us allow users to assign shared free-form tags to resources, thus generating annotations for objects with a minimum amount of effort. The informal nature of tagging means that semantic information cannot be directly inferred from an annotation, as any user can tag any resource with whatever strings they wish. However, studying the collective tagging behavior of a large number of users allows emergent semantics to be derived [46]. Through a combination of such mass collaborative "structural" semantics (via tags, geotemporal information, ratings, etc.) and extracted multimedia "content" semantics (which can be used for clustering purposes, e.g., image similarities or musical patterns), relevant annotations can be suggested to users when they contribute multimedia content to a community site by comparing new items with related semantic items in one's implicit/explicit network.

Another way in which the wisdom of crowds can be harnessed in semantic multimedia management is in providing personalized social network-based recommender systems. Liu et al. [35] present an approach for semantic mining of personal tastes and a model for taste-based recommendation. Ghita et al. [20] explore how a group of people with similar interests can share documents/metadata and can provide each other with semantically rich recommendations. The same principles can be applied to multimedia recommendation, and these recommendations can be augmented with the semantics derived from the multimedia content itself (e.g., the information on those people depicted or carrying out actions in multimedia objects[53]).

## 6.2 Privacy and Deliberate Fragmentation

Some challenges must also be overcome regarding the online identity aspect and authentication/privacy for users of Social Web sites. An interesting aspect of social networking and media-sharing Web sites is that most people use various Web sites

52 http://dbtune.org.
53 http://acronym.deri.org/.

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 personas, 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 IFPs) allow us to merge these data and provide vocabularies, methods, and tools for data portability among Social Web sites [7, 8], this identity fragmentation must be taken into account. It implies a need for new ways to authenticate queries or carry out inferencing, by delivering data in different manners depending on, for instance, which social subgraph the person requesting the data belongs to (family, coworker, etc.). Here, Web 2.0 efforts like oAuth[54] are of interest. oAuth is an open protocol which enables users to allow applications access their protected data stored in accounts they hold with other services. Also relevant is the recent proposal for RDFAuth.[55] 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 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.

## 6.3 Using Wikipedia as a Reputation System with Embedded Semantics

As a global, independent and neutral framework to which we can all contribute content, Wikipedia could serve as the basis for a de-facto global and open reputation system. At the moment, Wikipedia does not provide much information on people's reputations, that is, those who make changes to articles are not very visible on

[54] http://oauth.org/.
[55] http://blogs.sun.com/bblfish/entry/rdfauth_sketch_of_a_buzzword.

Wikipedia and are not treated as experts as such. On the Wikipedia Web site, it is often the case that the contributor who may know the most about an article is not clearly identified in the Wikipedia article as being the foremost expert.

There have been various attempts to establish reputation sites on the Web, for example, Naymz, which may help a person to improve their visibility in search engines. However, there is a problem with these sites in that a person's reputation can only be truly reflected online if they regularly contribute to the site and maintain an up-to-date version of their profile with all of their achievements. Another issue is that people who already have a good reputation will most probably not join these sites, perhaps due to time constraints, or if reputation is related to the number of connections or endorsements one has (which may be by invitation).

Wikipedia can be improved by the addition of a global reputation system with embedded semantics. This could be achieved by placing larger emphasis on the discussion pages in the Wikipedia, and by introducing threaded structures in these pages from which expertise would emerge. For example, experts could emerge from their actions in discussion pages when their suggested changes have been accepted, highlighting those who made the best changes on the article page itself.

If we include microcontent such as microformats or RDFa in these pages, we solve two problems at one stroke (1) Wikipedia benefits from a richer reputation framework where people can be motivated to add contextual semantic information to make their content better searchable (directly benefiting their own reputations) and (2) this can also move forward the Semantic Web, by solving the issue of who will be motivated to add the semantics to the Semantic Web and why. This information can also be used to power services like QDOS that aim to measure people's digital status online.

## 6.4 A Common Social Networking Stack

So far, SNSs use explicit representations of social networks primarily for visualization and browsing purposes. Yet, some research prototypes show that social networks are actually useful for more than just ego surfing to discover unexpected links in networks of friends. For example, some efforts are under way to examine email filtering and ranking based on social networks [18, 22] Explicitly represented social networking information can also provide a means for assessing a piece of information's importance and relevance for many other kinds of information filtering (e.g., in semantic attention management [42]) and routing, in general.

Rather than building a separate social networking layer into tools (with all the created maintenance problems), information space and application architects need to fold it into the technology stacks (see Fig. 7). Nepomuk does this for the desktop, but given the evolution toward ubiquitous computing and the so-called "Internet of

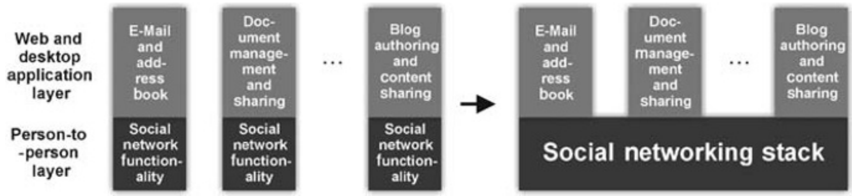


FIG. 7. Making social networking a shared component across various desktop and Web applications.

things," which will deliver much more information, the Internet infrastructure itself might need to be augmented to include social networking infrastructure to keep users from drowning in an ocean of unconnected and meaningless information. Just as the social semantic desktop Nepomuk[56] provides an operating system layer for representing and exchanging information on the desktop, information creation on the Web and the Internet should take existing connections between content objects and people into account to provide meaning for this information. For example, SNSs might include mechanisms to automate the creation of connections among information items or to route information based on existing relationships between people and content items.

A social networking stack needs to take into account a person's relevant objects of interest and provide some limited data portability (at the very least, for their most highly used or rated items). Through this, the actions and interactions of a person with other users and objects (exhibiting relevant properties) in existing systems can be used to create new user or group connections when a person registers for a new social networking site or application. Also, instead of having a fragmented view of one's network in each application, the social networking stack would let a user use all of their person-to-person connections in any application. To enable the sharing of existing contacts and to aid with the creation of new ones, the cross-application social networking stack will require a number of layers:

1. **Personal authentication and authorization layer.** This layer would use OpenID, Sxip,[57] or some other single sign-on mechanism to authenticate that an individual is who they claim to be, and would in turn ensure that they are authorized to make use of their social network connections (layer 2) and/or leverage previously created content items (layer 3).

[56] http://nepomuk.semanticdesktop.org/.
[57] http://sxip.com/.

2. **Social network access layer.** This layer would utilize the social networking contacts created by an individual across various platforms, for example, by collecting FOAF "knows" relationships from multiple sites. However, access control is required as social connections may not always be bidirectional: that is, there has to be some consent from both sides for certain transactions. For example, Alice may create a connection to Bob in order to view Bob's public content, but Bob may have to approve the connection in the reverse direction if Alice ever wants to send him a direct message. This layer would not only ensure that the required directional links exist for various interactions, but would also verify that the source of this social network information is valid.
3. **Content object access layer.** This layer would collect a person's relevant content objects, and verify that they are allowed to reuse data/metadata from these objects in the current application. This could be achieved using SIOC as a representation format, aggregating a person's created items (through their user accounts) from various site containers. For reputation purposes, this layer would also verify that these items were in fact created by the authenticated individual on whatever sites they reference. This may require provenance of information as well as signing of RDF graphs [13] and possibly advanced policies for dealing with identity theft.

For the implementation of a social networking stack, various architectural alternatives exist: the existing Domain Name System (DNS) system is an example of a possible architecture, but creates a central point of control. A peer-to-peer approach is another possibility which would be worthwhile to explore, especially since it preserves the distributed aspect.

The availability of a social networking stack would also have an effect on existing networking layers: social routing algorithms are able to deliver information directly to people for whom the information is relevant—email filtering and routing with social networks being just a simple example.

## 7. Conclusions

In this chapter, we have described the significance of community-oriented and content-sharing sites on the Web, the shortcomings of many of these sites as they are now, and the benefits that semantic technologies can bring to social networks and Social Web sites. Online social spaces encouraging content creation and sharing have resulted in the formation of massive and intricate networks of people and associated content. However, the lack of integration between sites means that these networks are disjoint and users are unable to reuse data across sites. As well as content, many third parties are producing application widgets that can be added by users to their Social Web site profiles, but mechanisms for trusting the source of these widgets can be improved or augmented with information derived from social network connections. There is a need for Semantic Web technologies that can solve some of these issues and improve the value and functionality of online social spaces. The process of creating and using semantic data in the Social Web can be viewed as a sort of food chain of producers, collectors, and consumers. Semantic data producers publish information in structured, common formats, such that it can easily be integrated with data from other diverse sources. Collectors, if necessary, aggregate and consolidate heterogeneous data from other diverse sources. Consumers may use these data for analysis or in end-user applications.

In this way, it becomes possible to integrate diverse information from heterogeneous sites, enabling improved navigation and the ability to query over data. There are also advantages for those interested in studying social networks, as the Semantic Web makes freely available large-scale, multirelational data sets for analysis. In this chapter, we described some methods by which consolidated facts and content can be extracted from people and content networks aggregated from multiple social networks and Social Web sites. We also presented some of our ideas for future work, including the need for more semantics as the focus of Social Web sites moves toward the provision of multimedia content; requirements for privacy and occasional fragmentation of a user's aggregated semantic content; and how a reputation system with embedded semantics could be deployed in a large-scale community site. Finally, based on observations that form and deployment are evolving toward object-centered networks and driven by the need to exploit information assessment methods, we described the direct integration of a social networking layer into the technology stack of clients (the desktop) and the Internet itself.

## Acknowledgment

This work was supported by Science Foundation Ireland under Grant No. SFI/02/CE1/I131.

## References

[1] Aleman-Meza B., Nagarajan M., Ramakrishnan C., Ding L., Kolari P., Sheth A. P., Arpinar I. B., Joshi A., and Finin T., 2006. Semantic analytics on social networks: Experiences in addressing the problem of conflict of interest detection. In *Proceedings of the 15th International Conference on the World Wide Web*, Edinburgh, Scotland.

[2] Aleman-Meza B., Bojars U., Boley H., Breslin J. G., Mochol M., Nixon L. J. B., Polleres A., and Zhdanova A. V., 2007. Combining RDF vocabularies for expert finding. In *Proceedings of the 4th European Semantic Web Conference (ESWC'07)*, June 2007, Innsbruck, Austria.

[3] Batagelj V., and Mrvar A., 1998. Pajek—Program for large network analysis. *Connections*, **21**(2): 47–57.
[4] Berners-Lee T., Hendler J. A., and Lassila O., May 2001. The semantic Web. *Scientific American*, **284**(5): 34–43.
[5] Bojārs U., Heitmann B., and Oren E., 2007. A prototype to explore content and context on social community sites. *The SABRE Conference on Social Semantic Web (CSSW 2007)*, September 2007, Leipzig, Germany.
[6] Bojārs U., Passant A., Giasson F., and Breslin J. G., 2007. An architecture to discover and query decentralised RDF data. In *The 3rd Workshop on Scripting for the Semantic Web at the 4th European semantic Web Conference (ESWC'07)*, June 2007, Innsbruck, Austria.
[7] Bojārs U., Breslin J. G., Finn A., and Decker S., 2008. Using the semantic Web for linking and reusing data across Web 2.0 communities. In *Special Issue on the Semantic Web and Web 2.0. The Journal of Web Semantics*, **6,** 21–28.
[8] Bojārs U., Passant A., Breslin J. G., and Decker S., 2008. Social network and data portability using semantic Web technologies. In *Proceedings of the BIS 2008 Workshop on Social Aspects of the Web*, May 2008, Innsbruck, Austria.
[9] Borgatti S. P., Everett M. G., and Freeman L. C., 2002. UCINET for Windows: Software for Social Network Analysis. Analytic Technologies, Harvard, MA.
[10] Boyd D. M., and Ellison N. B., 2007. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, **13**(1).
[11] Breslin J. G., and Decker S., November/December 2007. The future of social networks on the Internet: The need for semantics. *IEEE Internet Computing*, **11,** 86–90.
[12] Breslin J. G., Harth A., Bojārs U., and Decker S., 2005. Towards Semantically-Interlinked Online Communities. In *Proceedings of the 2nd European Semantic Web Conference (ESWC'05)*, May 2005, Heraklion, Greece, LNCS, vol. 3532, pp. 500–514.
[13] Carroll J. J., Bizer C., Hayes P., and Stickler P., 2005. Named graphs, provenance and trust. In *Proceedings of the 14th International Conference on the World Wide Web (WWW2005)*, pp. 613–622. ACM Press, Chiba, Japan.
[14] Choi H. C., Kruk S. R., Grzonkowski S., Stankiewicz K., Davis B., and Breslin J. G., 2006. Trust models for community aware identity management. In *Proceedings of the Identity, Reference and Web Workshop at the 15th International World Wide Web Conference (IRW2006, WWW2006)*, May 2006, Edinburgh, Scotland.
[15] Decker S., and Frank M., 2004. The social semantic desktop. Technical Report 2004-05-02, Digital Enterprise Research Institute, National University of Ireland, Galway, May 2004. (http://www.deri.ie/fileadmin/documents/DERI-TR-2004-05-02.pdf ).
[16] Ding L., Zhou L., Finin T., and Joshi A., 2005. How the semantic Web is being used: An analysis of FOAF documents. In *Proceedings of the 38th Hawaii International Conference on System Sciences (HICSS'05)*.
[17] Fernandez S., Berrueta D., and Labra J. E., 2007. Mailing lists meet the semantic Web. In *Proceedings of the BIS 2007 Workshop on Social Aspects of the Web*, April 2007, Poznan, Poland.
[18] Fisher D., Hogan B., Brush A. J., Smith M. A., and Jacobs A., 2006. Using social sorting to enhance email management. In *Proceedings of the Human–Computer Interaction Consortium (HCIC'06)*, ACM Press, New York, NY. (http://research.microsoft.com/research/pubs/view.aspx?type=Publication&id=1600).
[19] Freeman L. C., 2000. Visualizing social networks. *Journal of Social Structure*, **1**(1).
[20] Ghita S., Nejdl W., and Paiu W. R., 2005. Semantically rich recommendations in social networks for sharing, exchanging and ranking semantic context. In *Proceedings of the 4th International Semantic Web Conference*, November 2005, Galway, Ireland.

[21] Girvan M., and Newman M. E. J., 2002. Community structure in social and biological networks. In *Proceedings of the National Academy of Sciences of the United States of America*, **99**(12): 7821–7826.

[22] Golbeck J., and Hendler J., 2004. Reputation network analysis for email filtering. In *Proceedings of the 1st Conference on Email and Anti-Spam, Microsoft Research*. (http://www.ceas.cc/papers-2004/177.pdf ).

[23] Golbeck J., Parsia B., and Hendler J., 2003. Trust networks on the semantic Web. In *Proceedings of Cooperative Intelligent Agents*, August 2003, Helsinki, Finland.

[24] Golder S., and Huberman B. A., April 2006. The structure of collaborative tagging systems. *Journal of Information Sciences*, **32**(2): 198–208.

[25] Gruber T., 2007. Ontology of folksonomy: A mash-up of apples and oranges. *International Journal on Semantic Web and Information Systems*, **3**(2).

[26] Heer J., and Boyd D., 2005. Vizster: Visualizing online social networks. In *IEEE Symposium on Information Visualization (InfoVis 2005)*, October 2005, Minneapolis, MN.

[27] Irvine M., 9 October 2006. Breaking up with social-networking sites. *The Seattle Times*. (http://seattletimes.nwsource.com/html/living/2003292646_onlinebacklash09.html).

[28] Jordan K., Hauser J., and Foster S., 2003. The augmented social network: Building identity and trust into the next-generation Internet. *First Monday*, **8**(8). (http://www.firstmonday.org/issues/issue8_8/jordan/index).

[29] Kim H. L., Yang S. K., Breslin J. G., and Kim H. G., 2007. Simple algorithms for representing tag frequencies in the SCOT exporter. In *The IEEE/WIC/ACM International Conference on Intelligent Agent Technology*, pp. 536–539. IEEE Computer Society, Washington, DC.

[30] Kinsella S., Harth A., Troussov A., Sogrin M., Judge J., Hayes C., and Breslin J. G., 2007. Navigating and annotating semantically-enabled networks of people and associated objects. In *Accepted for the 4th Conference on Applications of Social Network Analysis (ASNA 2007)*, September 2007, University of Zurich, Switzerland.

[31] Knorr-Cetina K., 1997. Sociality with objects: Social relations in postsocial knowledge societies. *Theory, Culture & Society*, **14**(4): 1–30.

[32] Kochut K., and Janik M., 2007. SPARQLeR: Extended SPARQL for semantic association discovery. In *Proceedings of the 4th European Semantic Web Conference (ESWC'07)*, Lecture Notes in Computer Science, vol. 4519, pp. 145–159. Springer-Verlag, Berlin.

[33] Lange C., Bojārs U., Groza T., Breslin J. G., and Handschuh S., 2008. Expressing Argumentative Discussions in Social Media Sites. In *Proceedings of the 1st International Workshop on Social Data on the Web (SDOW 2008) at the 7th International Semantic Web Conference (ISWC 2008)*, October 2008, Karlsruhe, Germany (ISSN 1613-0073).

[34] Leskovec J., Adamic L. A., and Huberman B. A., 2006. The dynamics of viral marketing. In *Proceedings of the ACM Conference on Electronic Commerce*, pp. 228–237. ACM Press, New York, NY.

[35] Liu H., Maes P., and Davenport G., 2006. Unraveling the taste fabric of social networks. *International Journal on Semantic Web and Information Systems*, **2,** 42–71.

[36] McAfee A., 2006. Enterprise 2.0: The dawn of emergent collaboration. *MIT Sloan Management Review*, **47**(3): 21–28.

[37] Mika P., 2005. Flink: Semantic Web technology for the extraction and analysis of social networks. *Web Semantics: Science, Services and Agents on the World Wide Web*, **3**(2–3): 211–223.

[38] Mika P., 2005. Ontologies are us: A unified model of social networks and semantics. In *International Semantic Web Conference*, Lecture Notes in Computer Science, pp. 522–536. Springer-Verlag, Berlin.

[39] O'Madadhain J., Fisher D., White S., and Boey Y., 2003. The JUNG (Java Universal Network/ Graph) Framework. University of California, Irvine, CA.

[40] Passant A., and Laublet P., 2008. Meaning of a tag: A collaborative approach to bridge the gap between tagging and linked data. In *Proceedings of the WWW2008 Linked Data on the Web Workshop (LDOW2008)*, April 2008, Beijing, China.

[41] Passant A., and Raimond Y., 2008. Combining social music and semantic Web for music-related recommender systems. In *Proceedings of the 1st International Workshop on Social Data on the Web (SDOW 2008) at the 7th International Semantic Web Conference (ISWC 2008)*, October 2008, Karlsruhe, Germany (ISSN 1613-0073).

[42] Petrie C., 2006. Semantic attention management. *IEEE Internet Computing*, **10**(5): 93–96.

[43] Specia L., and Motta E., 2007. Integrating folksonomies with the semantic Web. In *Proceedings of the 4th European Semantic Web Conference (ESWC'07), semantic Web: Research and Applications*, May 2007, Innsbruck, Austria, LNCS, vol. 4519, pp. 624–639.

[44] Wasserman S., and Faust K., 1994. Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge.

[45] Watts D. J., and Strogatz S. H., 1998. Collective dynamics of 'small-world' networks. *Nature*, **393**(6684): 409–410.

[46] Wu X., Zhang L., and Yu Y., 2006. Exploring social annotations for the semantic Web. In *Proceedings of the 15th International Conference on World Wide Web*, May 2006, Edinburgh, Scotland.