The Social Semantic Web

John G. Breslin · Alexandre Passant · Stefan Decker

The Social Semantic Web

Springer

John G. Breslin
Electrical and Electronic Engineering
School of Engineering and Informatics
National University of Ireland, Galway
Nuns Island
Galway
Ireland
john.breslin@nuigalway.ie

Alexandre Passant
Digital Enterprise Research
Institute (DERI)
National University of Ireland, Galway
IDA Business Park
Lower Dangan
Galway
Ireland
alexandre.passant@deri.org

Stefan Decker
Digital Enterprise Research
Institute (DERI)
National University of Ireland, Galway
IDA Business Park
Lower Dangan
Galway
Ireland
stefan.decker@deri.org

ISBN 978-3-642-01171-9 e-ISBN 978-3-642-01172-6
DOI 10.1007/978-3-642-01172-6
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2009936149

ACM Computing Classification (1998): H.3.5, H.4.3, I.2, K.4

Cover design: KuenkelLopka GmbH

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

1 Introduction to the book

1.1 Overview

The Social Web - encompassing social networking services such as MySpace, Facebook and orkut, as well as content-sharing sites (that also offer social networking functionality) like Flickr, Last.fm and del.icio.us - has captured the attention of millions of users as well as billions of dollars in investment and acquisition. As more social websites form around the connections between people and their objects of interest (to avoid these sites becoming boring), and as these 'object-centred networks' (where people connect via these objects of interest) grow bigger and more diverse, more intuitive methods are needed for representing and navigating the content items in these sites: both within and across social websites. Also, to better enable user access to multiple sites and ultimately to content-creation facilities on the Web, interoperability among social websites is required in terms of both the content objects and the person-to-person networks expressed on each site. This requires representation mechanisms to interconnect people and objects on the Web in an interoperable and extensible way (Breslin and Decker 2007).

Semantic Web representation mechanisms are ideally suited to describing people and the objects that link them together in such object-centred 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 websites. These efforts include the Friend-of-a-Friend (FOAF) project[1] for describing people and relationships, the Nepomuk social semantic desktop[2] 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[3] for representing online discussions (Breslin et al. 2005). Some social networking services (SNSs), such as FriendFeed, are also starting to provide query interfaces to their data, which others can reuse and link to via the Semantic Web.

The Semantic Web is a useful platform for linking and for performing operations on diverse person- and object-related data (as shown in Figure 1.1) gathered from heterogeneous social websites (in what is termed 'Web 2.0').

[1] http://www.foaf-project.org/ (URL last accessed 2009-06-09)

[2] http://nepomuk.semanticdesktop.org/ (URL last accessed 2009-06-09)

[3] http://sioc-project.org/ (URL last accessed 2009-06-09)

J.G. Breslin et al., *The Social Semantic Web*, DOI 10.1007/978-3-642-01172-6_1,

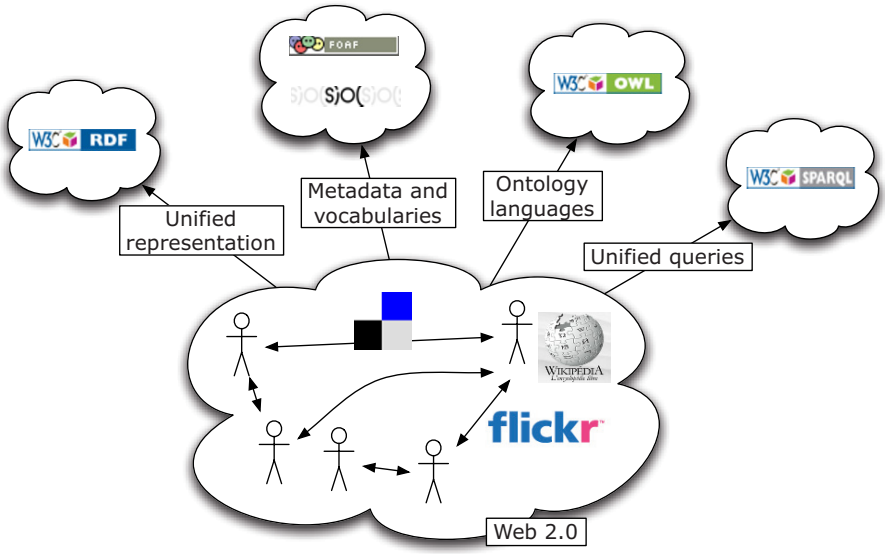

Fig. 1.1. Interconnecting and reusing distributed Web 2.0 data with semantic technologies

In the other direction, object-centred networks and user-centric services for generating collaborative content can serve as rich data sources for Semantic Web applications (Figure 1.2).

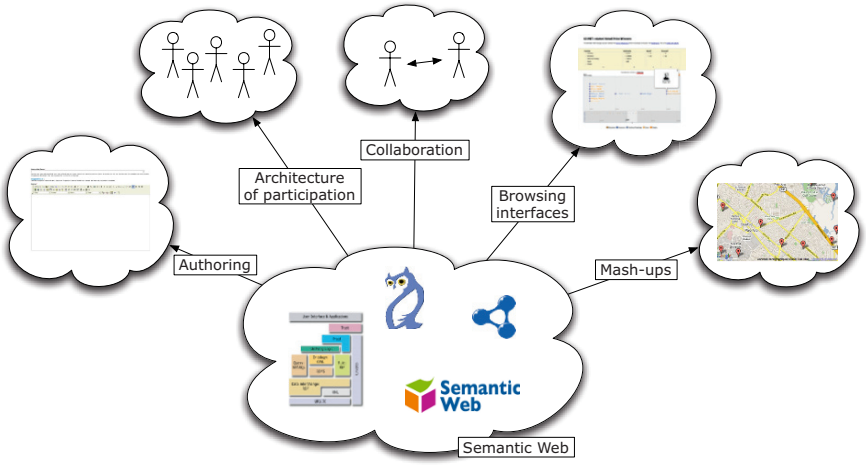

Fig. 1.2. Powering semantic applications with rich community-created content and Web 2.0 paradigms

This linked data can provide an enhanced view of individual or community activity in localised or distributed object-centred social networks. In fact, since all this data can be 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 localised. All of this data can be considered as a unique interlinked machine-understandable graph layer (with nodes as users or related data and arcs as relationships) over the existing Web of documents and hyperlinks, i.e. a Giant Global Graph as Tim Berners-Lee recently coined[4]. Moreover, such interlinked data allows advanced querying capabilities, for example, 'show me all the content that Alice has acted on in the past three months in any SNS'.

In this book, we will begin with our motivations followed by overviews of both the Social Web and the Semantic Web. Then we will describe some popular social media and social networking applications, list some of their strengths and limitations, and describe some applications of Semantic Web technologies to address current issues with social websites by enhancing them with semantics.

Across these heterogeneous social websites, we will demonstrate a twofold approach towards integrating the Social Web and the Semantic Web: in particular, (1) by demonstrating how the Semantic Web can serve as a useful platform for linking and for performing operations on diverse person- and object-related data gathered from these websites, and (2) by showing that in the other direction, social websites can themselves serve as rich data sources for Semantic Web applications.

We shall conclude with some observations on how the application of Semantic Web technologies to the Social Web is leading towards the 'Social Semantic Web', forming a network of interlinked and semantically-rich content and knowledge.

1.2 Aims of the book, and who will benefit from it?

Initially, we aim to educate readers on evolving areas from the world of collaboration and communication systems, social software and the Social Web. We shall also show connections with parallel developments in the Semantic Web effort. Then, we will illustrate how social software applications can be enhanced and interconnected with semantic technologies, including semantic and structured blogging, interconnecting community sites, semantic wikis, and distributed social networks. The goal of this book is that readers will be able to apply Semantic Web technologies to these and to other application areas in what is termed the Social Semantic Web.

This book is intended for computer science professionals, researchers, academics and graduate students interested in understanding the technologies and research issues involved in applying Semantic Web technologies to social software. Applications such as blogs, social networks and wikis require more automated ways for information distribution. Practitioners and developers interested in such applica-

[4] http://dig.csail.mit.edu/breadcrumbs/node/215 (URL last accessed 2009-06-09)

tion areas will also learn about methods for increasing the levels of automation in these forms of web communication.

For those who have background knowledge in the area of the Semantic Web, we envisage that this book will help you to develop application knowledge in relation to social software and other widely-used related Social Web technologies. For those who already have application knowledge in web engineering or in the development of systems such as wikis, social networks and blogs, we hope this book will inspire you to develop and create ideas on how to increase the usability of social software and other web systems using Semantic Web technologies.

1.3 Structure of the book

We shall now give an introduction to the chapters in this book and explain the logical chapter layout and flow (Figure 1.3).

Following an overview of the motivation for combining the Social Web and the Semantic Web, we will proceed with an introduction to various technologies and trends in both the Social Web and the Semantic Web domains.

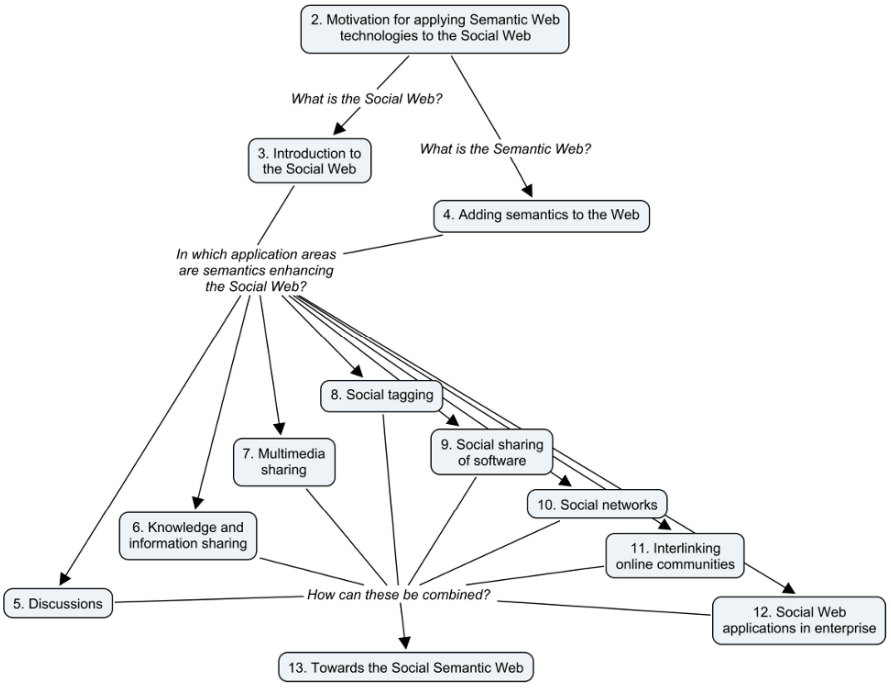

Fig. 1.3. Chapter flow for the book

This will be followed by a series of chapters whereby various Social Web application areas will be introduced, and semantic enhancements to these areas will be described. The areas we focus on are: online discussion systems such as forums, blogs and mailing lists; knowledge sharing services such as wikis and other sites for (mainly textual) information storage and recovery; multimedia services for sharing images, audio and video files; bookmarking sites and similar services organised around tagging functionality; sites for publishing and sharing community software projects; online social networking services; interlinked online communities; and enterprise applications. These chapters will have varying ratios of semantic implementations to non-semantic ones where state-of-the-art semantic techniques may have achieved more traction in some application areas.

Finally, in the last chapter we will describe approaches to integrate these social semantic applications in what we have termed 'Social Semantic Information Spaces'.

1.3.1 Motivation for applying Semantic Web technologies to the Social Web

This part will focus on the motivation for applying Semantic Web technologies to the Social Web, as summarised in the introductory description just given.

1.3.2 Introduction to the Social Web (Web 2.0, social media, social software)

We shall begin with an overview of social websites, looking at common Social Web technologies and methods for collaboration, content sharing, data exchange and representation (enhancing interaction and exchange with AJAX and mashups, how content is being categorised via tagging and folksonomies, etc.). We shall also discuss existing structured content that is available from social websites, mainly via content syndication whereby people can keep up to date with published material using RSS, Atom and other subscription methods. Then we will introduce the notion of object-centred sociality (referencing the observations of Jyri Engeström and Karen Knorr-Cetina), where social websites are organised around the objects of interest that connect people together.

1.3.3 Adding semantics to the Web

In this chapter, we will examine state of the art in the Semantic Web such as metadata and ontology standards and mashups, as well as some efforts aimed at providing semantic search and leveraging linked data. We shall talk about why object-centred sociality provides a meaning for representing Social Web content using semantics. The chapter will focus not only on the 'uppercase' Semantic Web (where formal specifications such as OWL and RDF are used to represent ontologies and associated metadata), but will also look at the 'lowercase' semantic web (where developer-led efforts in the microformats community are creating simple semantic structures for use by 'people first, machines second').

1.3.4 Discussions

We shall describe the area of blogging, one of the most popular Social Web activities. Blogs are online journals or sets of chronological news entries that are maintained by individuals, communities or commercial entities, and can be used to publish personal opinions, diary-like articles or news stories relating to a particular interest or product. We shall begin by describing current approaches to blogging, and detail how semantic technologies improve both the processes of creating and editing blog posts, and of browsing and querying the data created by blogs (via structured blogging and semantic blogging). We shall also discuss forums, mailing lists, and other web-based discussion systems such as microblogging, a recent trend regarding lightweight and agile communication on the Web.

1.3.5 Knowledge and information sharing

Wikis are collaboratively-edited websites that can be updated or added to by anyone with an interest in the topic covered by the wiki site, and have been used to create online encyclopaedias, photo galleries and literature collections. We shall describe the Social Web application area of wikis, and describe how adding semantics to wikis can offer distinct benefits: augmenting the language text in wiki articles with structured data and typed links enables advanced querying and browsing. We shall examine popular semantic wikis in usage today (e.g. Semantic MediaWiki), and we will look at semantic services that leverage structured information from wikis (such as the DBpedia). We shall briefly detail how a reputation system with embedded semantics could be deployed in a large-scale community site like the Wikipedia. We shall also look at the latest wave of knowledge net-

working and information sharing services (including Twine, Freebase, and OpenLink Data Spaces).

1.3.6 Multimedia sharing

We shall begin by looking at Social Web applications for storing and sharing photographs and other images (Flickr, Zooomr, etc.), and describe an application called FlickRDF that exports semantic data from the Flickr service. We shall then describe both audio and video podcasting, and give some ideas for the application of semantics to this area (e.g. through metadata descriptions and applications like ZemPod). We shall finish the chapter with a description of how semantic technologies can be applied to social music services and websites like Last.fm, through projects such as DBTune and the Music Ontology.

1.3.7 Social tagging

This chapter will discuss social tagging and bookmarking services on the Web. We shall look at tagging and how semantics can assist the tagging process as well as enhancing related aspects such as tag clouds. We shall look at annotated social bookmarks, where sites like del.icio.us are allowing people to publicly publish textual descriptions of their favourite links along with associated annotations of use to others, and we will describe different issues related to tagging behaviours. We shall describe how semantics can be added to tagging systems, both by defining models to represent tagging activities or particular behaviours and by extracting a hierarchy of concepts or vocabularies from tags. Semantic social bookmarking and tagging applications (e.g. int.ere.st, Revyu, LODr) will also be described to emphasise how different aspects of tagging applications can be augmented thanks to Semantic Web technologies.

1.3.8 Social sharing of software

The Social Web allows us to not only share data or multimedia content, but also applications, especially free-software applications or lightweight add-ons to web pages such as widgets. We shall look at how interoperability among social websites is possible not just in terms of the expressed content but also in terms of the social applications in use (e.g. widgets) on each site. We shall give an overview of existing ways to share software on the Web, focussing on how a social aspect can be added to data such as software projects or widget descriptions. We shall follow

this with a description of methods for describing software projects using semantics, and we will see how applications can be identified and discovered on the Web thanks to these semantics. We shall also discuss how trust mechanisms for consuming applications can be leveraged via the distributed social graph so that users can decide who to accept any new data or applications from.

1.3.9 Social networks

We shall begin with an overview of social networks, and look at current developments regarding the 'social graph'. We further describe the idea of object-centred sociality as introduced in Chapter 3. We shall then discuss initiatives from major Web companies to provide interoperability between social networking applications such as Facebook Connect and Google's OpenSocial and Social Graph APIs. We shall finish the chapter with a description of how open and distributed semantic social networks can be created through definitions such as Friend-of-a-Friend (FOAF) or XHTML Friends Network (XFN), enabling interoperability between different SNSs.

1.3.10 Interlinking online communities

We shall describe the usage of Semantic Web technologies for enhancing community portals and for connecting heterogeneous social websites - SIOC is currently being used for information structuring as well as for export and information dissemination. We shall describe current standardisation activities as well as research prototype applications and commercial implementations. We shall also show how SIOC can be combined with other ontologies (including FOAF, SKOS, and Dublin Core) in architectures for community site interoperability. We will look at current projects that enable one to query for topics or to browse distributed discussion content across various types of social websites (e.g. the SIOC Explorer, Sindice SIOC Widget).

1.3.11 Social Web applications in enterprise

We shall begin with an overview of Enterprise 2.0, looking at how Social Web applications are being used internally and externally by companies. We shall then examine the application of Semantic Web technologies to Enterprise 2.0 ecosystems. In particular, we will look at the usage of semantics in integrated enterprise social software suites as well as how the Semantic Web can help us to integrate

the various components that are being used in Enterprise 2.0 ecosystems. For example, we will show how collaborative work environments can be enhanced through the application of semantics (e.g. SIOC4CWE).

1.3.12 Towards the Social Semantic Web

Finally, we will discuss and present current approaches to realize the ideas of Vannevar Bush (Bush 1945) and Doug Engelbart (Engelbart 1962) on distributed collaboration infrastructures, towards both the Social Semantic Web and the Social Semantic Desktop (together, we term these as Social Semantic Information Spaces). We can combine the semantically-enhanced social software applications described in previous chapters into a Social Semantic Information Space. In the spirit of seminal visions such as Bush's Memex and Engelbart's open hyperdocument system (OHS), this chapter will detail how previous perspectives on group forming, network modelling and algorithms, and innovative IT-based interaction with feedback are driving new initiatives for creating semantic connections within and between people's information spaces.

2 Motivation for applying Semantic Web technologies to the Social Web

Many will have become familiar with popular Social Web applications such as blogging, social networks and wikis, and will be aware that we are heading towards an interconnected information space (through the blogosphere, inter-wiki links, mashups, etc.). At the same time, these applications are experiencing boundaries in terms of information integration, dissemination, reuse, portability, searchability, automation and more demanding tasks like querying. The Semantic Web is increasingly aiming at these applications areas - quite a number of Semantic Web approaches have appeared in recent years to overcome the boundaries in these application areas, e.g. semantic wikis (Semantic MediaWiki), knowledge networking (Twine), embedded microcontent detection and reuse (Operator, Headup, Semantic Radar), social graph and data portability APIs (from Google and Facebook), etc. In an effort to consolidate and combine knowledge about existing efforts, we aim to educate readers about Social Web application areas and new avenues open to commercial exploitation in the Semantic Web. We shall give an overview of how the Social Web and Semantic Web can be meshed together.

2.1 Web 2.0 and the Social Web

One of the most visible trends on the Web is the emergence of the Web 2.0 technology platform. The term Web 2.0 refers to a perceived second-generation of Web-based communities and hosted services. Although the term suggests a new version of the Web, it does not refer to an update of the World Wide Web technical specifications, but rather to new structures and abstractions that have emerged on top of the ordinary Web. While it is difficult to define the exact boundaries of what structures or abstractions belong to Web 2.0, there seems to be an agreement that services and technologies like blogs, wikis, folksonomies, podcasts, RSS feeds (and other forms of many-to-many publishing), social software and social networking sites, web APIs, web standards[1] and online web services are part of Web 2.0. Web 2.0 has not only been a technological but also a business trend: according to Tim O'Reilly[2]: 'Web 2.0 is the business revolution in the computer in-

[1] http://www.webstandards.org/ (URL last accessed 2009-06-09)

[2] http://radar.oreilly.com/archives/2006/12/web-20-compact.html (accessed 2009-06-09)

J.G. Breslin et al., *The Social Semantic Web*, DOI 10.1007/978-3-642-01172-6_2,

dustry caused by the move to the Internet as platform, and an attempt to understand the rules for success on that new platform'.

Social networking sites such as Facebook (one of the world's most popular SNSs), 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 website - have become part of the daily lives of millions of users, and have generated huge amounts of investment since they began to appear around 2002. Since then, the popularity of these sites has grown hugely and continues to do so. (Boyd and Ellison 2007) have described the history of social networking sites, 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[3] 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.

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

A common property of Web 2.0 technologies is that they facilitate collaboration and sharing between users with low technical barriers – although usually on single sites (e.g. Technorati) or with a limited range of information (e.g. RSS, which we will describe later). In this book we will refer to this collaborative and sharing aspect as the 'Social Web', a term that can be used to describe a subset of Web interactions that are highly social, conversational and participatory. The Social Web may also be used instead of Web 2.0 as it is clearer what feature of the Web is being referred to[4].

The Social Web has applications on intranets as well as on the Internet. On the Internet, the Social Web enables participation through the simplification of user contributions via blogs and tagging, and has unleashed the power of community-based knowledge acquisition with efforts like Wikipedia demonstrating the collective 'wisdom of the crowds' in creating the largest encyclopaedia. One outcome of such websites, especially wikis, is that they can produce more valuable knowledge collectively rather than that created by separated individuals. In this sense, the Social Web can be seen as a way to create collective intelligence at a Web-scale

[3] http://www.internetworldstats.com/emarketing.htm (URL last accessed 2009-06-09)

[4] http://en.wikipedia.org/wiki/Social_web (URL last accessed 2009-06-09)

level, following the 'we are smarter than me' principles[5] (Libert and Spector 2008).

Similar technologies are also being used in company intranets as effective knowledge management, collaboration and communication tools between employees. Companies are also aiming to make social website users part of their IT 'team', e.g. by allowing users to have access to some of their data and by bringing the results into their business processes (Tapscott and Williams 2007).

2.2 Addressing limitations in the Social Web with semantics

A limitation of current social websites is that they are isolated from one another like islands in the sea (Figure 2.1). 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 websites, communities and services come online, the lack of interoperability between them becomes obvious. The Social Web creates a set of single data silos or 'stovepipes', i.e. 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 most Social Web applications, communities, and domains, there are still no common standards for knowledge and information exchange or interoperation available. RSS (Really Simple Syndication), a format for publishing recently-updated Web content such as blog entries, was the first step towards interoperability among social websites, but it has various limitations that make it difficult to be used efficiently in such an interoperability context, as we will see later.

Another extension of the Web aims to provide the tools that are necessary to define extensible and flexible standards for information exchange and interoperability. The Scientific American article (Berners-Lee et al. 2001) from Berners-Lee, Hendler and Lassila defined the Semantic 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 quite well-defined Semantic Web technology stack exists, enabling the creation of defining metadata and associated vocabularies.

[5] http://www.wearesmarter.org/ (URL last accessed 2009-06-09)

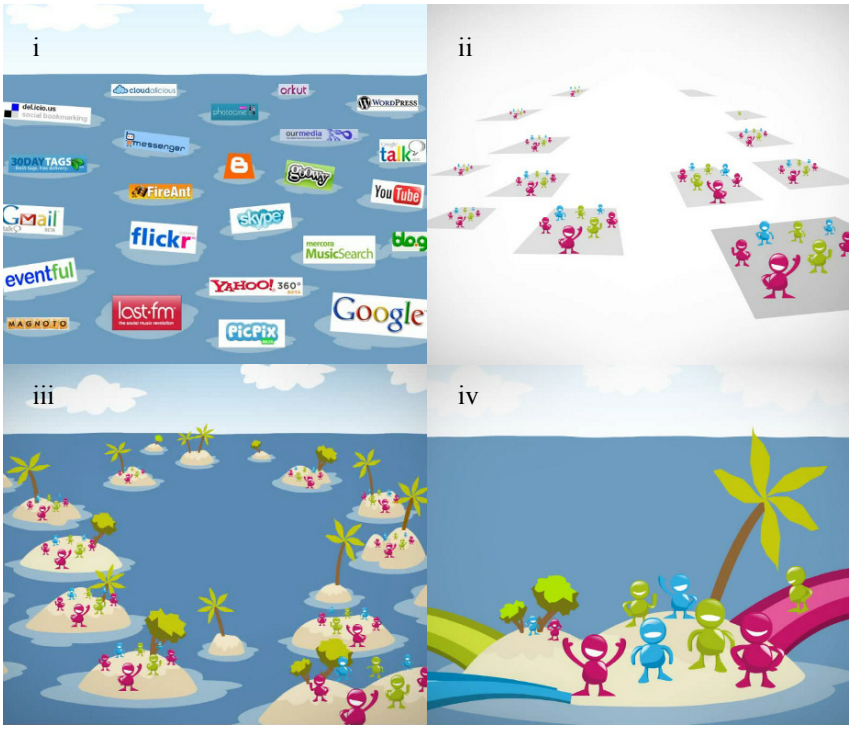

Fig. 2.1. Creating bridges between isolated communities of users and their data[6]

A number of Semantic Web vocabularies have achieved wide deployment – successful examples include RSS 1.0 for the syndication of information, FOAF, for expressing personal profile and social networking information, and SIOC, for interlinking communities and distributed conversations. These vocabularies share a joint property: they are small, but at the same time vertical – i.e. they are a part of many different domains. Each horizontal domain (e.g. e-health) would typically reuse a number of these vertical vocabularies, and when deployed the vocabularies would be able to interact with each other.

The Semantic Web effort is in an ideal position to make social websites interoperable by providing standards to support data interchange and interoperation between applications, enabling individuals and communities to participate in the creation of distributed interoperable information. The application of the Semantic Web to the Social Web is leading to the 'Social Semantic Web' (Figure 2.2), 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 modelled using machine-readable formats so that it can be used for purposes that the

[6] Images courtesy of Pidgin Technologies at http://www.pidgintech.com/

current state of the Social Web cannot achieve without difficulty. As Tim Berners-Lee said in a 2005 podcast[7], 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, social website users are already creating extensive vocabularies and semantically-rich annotations through folksonomies (Mika 2005a).

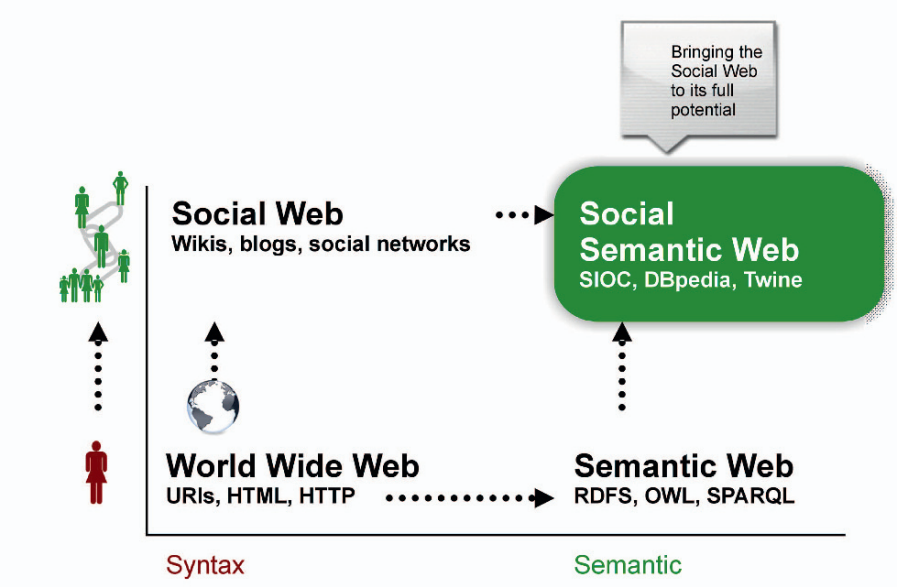

Fig. 2.2. The Social Semantic Web

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

2.3 The Social Semantic Web: more than the sum of its parts

The combination of the Social Web and Semantic Web can lead to something greater than the sum of its parts: a Social Semantic Web (Auer et al. 2007, Blu-

[7] http://esw.w3.org/topic/IswcPodcast (URL last accessed 2009-06-09)

mauer and Pellegrini 2008) 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.

In this book, we will describe various solutions that aim to make social websites interoperable, and which will take them beyond their current limitations to enable what we have termed Social Semantic Information Spaces[8]. Social Semantic Information Spaces are a platform for both personal and professional collaborative exchange with reusable community contributions. Through the use of Semantic Web data, searchable and interpretable content is added to existing Web-based collaborative infrastructures and social spaces, and intelligent use of this content can be made within these spaces - bringing the vision of semantics on the Web to its most usable and exploitable level.

Some typical application areas for social spaces are wikis, blogs and social networks, but they can include any spaces where content is being created, annotated and shared amongst a community of users. Each of these can be enhanced with machine-readable data to not only provide more functionality internally, but also to create an overall interconnected set of Social Semantic Information Spaces. These spaces offer a number of possibilities in terms of increased automation and information dissemination that are not easily realisable with current social software applications:

- By providing better interconnection of data, relevant information can be obtained from related social spaces (e.g. through social connections, inferred links, and other references).
- Social Semantic Information Spaces allow you to gather all your contributions and profiles across various sites ('subscribe to my brain'), or to gather content from your friend / colleague connections.
- These spaces allow the use of the Web as a clipboard to allow exchange between various collaborative applications (for example, by allowing readers to drag structured information from wiki pages into other applications, geographic data about locations on a wiki page could be used to annotate information on an event or a travel review in a blog post one is writing).
- Such spaces can help users to avoid having to repeatedly express several times over the same information if they belong to different social spaces.
- Due to the high semantic information available about users, their interests and relationships to other entities, personalisation of content and interface input mechanisms can be performed, and innovative ways for presenting related information can be created.
- These semantic spaces will also allow the creation of social semantic mashups, combining information from distributed data sources together that can also be enhanced with semantic information, for example, to provide the geolocations of friends in your social network who share similar interests with you.

[8] http://www2006.org/tutorials/#T13 (URL last accessed 2009-06-09)

- Fine-grained questions can be answered through such semantic social spaces, such as 'show me all content by people both geographically and socially near to me on the topic of movies'.
- Social Semantic Information Spaces can make use of emergent semantics to extract more information from both the content and any other embedded metadata.

There have been initial approaches in collaborative application areas to incorporate semantics in these applications with the aim of adding more functionality and enhancing data exchange - semantic wikis, semantic blogs and semantic social networks. These approaches require closer linkages and cross-application demonstrators to create further semantic integration both between and across application areas (e.g. not just blog-to-blog connections, but also blog-to-wiki exchanges). A combination of such semantic functionality with existing grassroots efforts such as OpenID[9] (a single sign-on mechanism) or OAuth[10] (an authentication scheme) can bring the Social Web to another level. Not only will this lead to an increased number of enhanced applications, but an overall interconnected set of Social Semantic Information Spaces can be created.

2.4 A food chain of applications for the Social Semantic Web

A semantic data 'food chain', as shown in Figure 2.3, consists of various producers, collectors and consumers of semantic data from social networks and social websites. Applying semantic technologies to social websites 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, in order to easily enable navigation and data exchange across sites, mechanisms are required to represent the 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, many 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.

[9] http://openid.net/ (URL last accessed 2009-06-09)

[10] http://oauth.net/ (URL last accessed 2009-06-09)

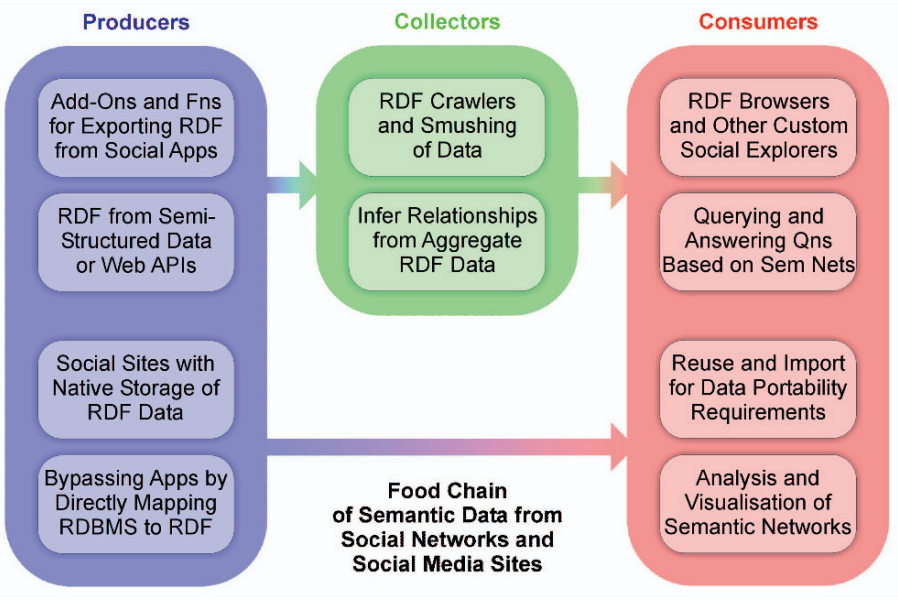

Fig. 2.3. A food chain for semantic data on the Social Web

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 use to study these networks, and we refer to initial work by (Ereteo et al. 2008) on how social network analysis 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 people and to the content that they create. RDF can be used to structure and expose information from the Social Web allowing the simple generation of semantic mashups for both proprietary and public information. HTML content can also be made compatible with RDF through RDFa (RDF annotations embedded in XHTML attributes), thereby enabling effective semantic search without requiring one to crawl a new set of pages (e.g. the Common Tag[11] effort allows metadata and URIs for tags to be exposed using RDFa and shared with other applications). Interlinking social data from multiple sources may give an enhanced view of information in distributed communities, and we will describe applications to consume and exchange this interlinked data in future chapters.

[11] http://commontag.org/ (URL last accessed 2009-07-07)

2.5 A practical Social Semantic Web

Applying Semantic Web technologies to social websites allows us to express different types of relationships between people, objects and concepts. By using common, machine-readable ways for expressing data about individuals, profiles, social connections and content, these technologies provide a way to interconnect people and objects on a Social Semantic Web in an interoperable, extensible way.

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

Social websites are becoming more prevalent and content is more distributed. This presents new challenges for navigating such data. Machine-readable descriptions of people and objects, and the use of common identifiers, can allow for linking diverse information from heterogeneous social networking sites. This creates 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 can have ownership and control over their own data and can move profile and content information between services as they wish. Recently there has been a push within the web community to make data portability (i.e. the ability for users to port their own data wherever they wish) a reality[12].

Additionally, the Social Web and social networking sites can contribute to the Semantic Web effort. Users of these sites often provide metadata in the form of annotations and tags on photos, ratings, blogroll links, etc. In this way, social networks and semantics can complement each other. Already within online communities, common vocabularies or folksonomies for tagging are emerging through a consensus of community members.

In this book we will describe a variety of practical Social Semantic Web applications that have been enhanced with extra features due to the rich content being created in social software tools by users, including the following:

- The Twine application from Radar Networks is an example of a system that leverages both the explicit (tags and metadata) and implicit semantics (automatic tagging of text) associated with content items. The underlying semantic data can also be exposed as RDF by appending '?rdf' to any Twine URL.

[12] http://www.dataportability.org/ (URL last accessed 2009-07-21)

- The SIOC vocabulary is powering an ecosystem of Social Semantic Web applications producing and consuming community data, ranging from individual blog exporters to interoperability mechanisms for collaborative work environments.
- The DBpedia represents structured content from the collaboratively-edited Wikipedia in semantic form, leveraging the semantics from many social media contributions by multiple users. DBpedia allows you to perform semantic queries on this 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-type interfaces and principles such as tagging with Semantic Web modelling principles to provide a reviews website that follows the principles of the Linking Open Data initiative (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 website is available in RDF, therefore it is available for reuse by other Social Semantic Web applications.

As Metcalfe's law defines, the value of a network is proportional to the number of nodes in the network. Metcalfe's law is strongly related to the network effect of the Web itself: by providing various links between people, social websites can benefit from that network effect, while at the same time the Semantic Web also provides links between various objects on the Web thereby obeying this law (Hendler and Golbeck 2008).

Therefore, by combining Web 2.0 and Semantic Web technologies, we can envisage better interaction between people and communities, as the global number of users will grow, and hence the value of the network. This will be achieved by (1) taking into account social interactions in the production of Semantic Web data, and (2) using Semantic Web technologies to interlink people and communities.

3 Introduction to the Social Web (Web 2.0, social media, social software)

Web 2.0 is a widely-used and wide-ranging term (in terms of interpretations), made popular by Tim O'Reilly who wrote an article on the seven features or principles of Web 2.0. To many people, Web 2.0 can mean many different things. Most agree that it can be thought of as the second phase of architecture and application development for the Web, and that the related term 'Social Web' describes a Web where users can meet, collaborate, and share content on social spaces via tagged items, activity streams, social networking functionality, etc. There are many popular examples that work along this collaboration and sharing meme: MySpace, del.icio.us, Digg, Flickr, Upcoming.org, Technorati, orkut, 43 Things, and the Wikipedia.

3.1 From the Web to a Social Web

Since it was founded, the Internet has been used to facilitate communication not only between computers but also between people. Usenet mailing lists and bulletin boards 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), web forums, instant messaging, blogging, social networking services, and even MMOGs or MMORPGs (massively multiplayer online [role playing] games) have continued the trend of using the Internet (and the Web) to build communities.

The structural and syntactic web put in place in the early 90s is still much the same as what we use today: resources (web pages, files, etc.) connected by untyped hyperlinks. By untyped, we mean that there is no easy way for a computer to figure out what a link between two pages means. Beyond links, the nature of the objects described in those pages (e.g. people, places, etc.) cannot be understood by software agents. In fact, the Web was envisaged to be much more (Figure 3.1). In Tim Berners-Lee's original outline for the Web in 1989, entitled 'Information Management: A Proposal'[1], resources are connected by links describing the type of relationships between them, e.g. 'wrote', 'describes', 'refers to', etc. This is a precursor to the Semantic Web which we will come back to in the next chapter.

[1] http://www.w3.org/History/1989/proposal.html (URL last accessed 2009-06-09)

J.G. Breslin et al., *The Social Semantic Web*, DOI 10.1007/978-3-642-01172-6_3,

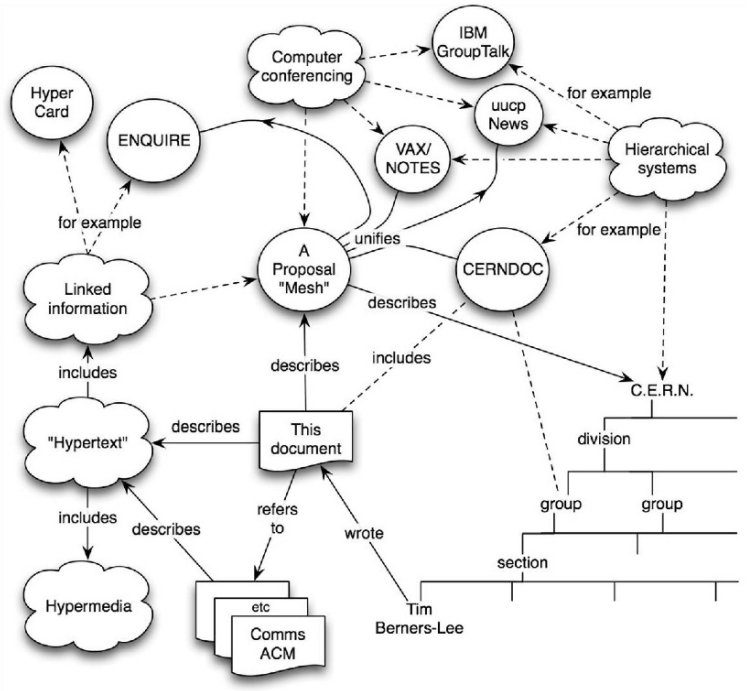

Fig. 3.1. Adapted from 'Information Management: A Proposal' by Tim Berners-Lee

Over the last decade and a half, there has been a shift from just 'existing' or publishing on the Web to participating in a 'read-write' Web. There has been a change in the role of a web 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, and Slashdot moderation is performed by the readers.

Web 2.0[2] is a widely-used and wide-ranging term (certainly in terms of interpretations) made popular by Tim O'Reilly. O'Reilly defined Web 2.0 as 'a set of principles and practices that ties together a veritable solar system of sites that demonstrate some or all of those principles, at a varying distance from that core'. While this definition is quite vague, he defined seven features or principles of Web 2.0, to which some have added an eighth: the long tail phenomenon (i.e. many small contributors and sites outweighing the main players). Among these features, two points seems particularly important: 'the Web as a platform' and 'an architecture of participation'. Actually, in spite of the 2.0 numbering, this vision is close to the original idea of Berners-Lee for the Web, i.e. that it should be a par-

[2] http://tinyurl.com/7tcjz (URL last accessed 2009-06-09)

ticipative medium. For example, the first Web browser called WorldWideWeb[3] was already a read-write browser, while current ones are generally read-only.

The first idea from O'Reilly of 'the Web as a platform' considers the Web and its principles as a way to provide services and value-added applications in addition to generally static contents. In some cases, the Web can even be seen as a transit layer for information to the desktop or mobile devices, for example, using RSS. We can also consider that 'the Web as a platform' refers to the migration of traditional desktop services such as e-mail and word processing to web-based applications, for example, as provided by Google with Gmail and Google Docs. In that context, the vision of 'an architecture of participation' emphasises how applications can help to produce value-added content and synergies by simply using them, thanks to the way they were designed. As people begin to use Web 2.0 applications for their own needs (uploading pictures, writing blog posts, tagging content), they enhance the global activity of the system and this can be a benefit for everyone. O'Reilly hence makes a comparison with open-source development principles and peer-to-peer architectures in relation to how they are providing the same kind of architecture of participation.

The evolution of the Web is - in our opinion - mostly a sociological and economic one, as referred to in the book 'Wikinomics' (Tapscott and Williams 2007). However, thanks to the strong interactions between services and users, it has led to interesting practices in terms of software development. O'Reilly in particular incites application developers to go further than they would in traditional development processes and to constantly deliver new features, leading to 'the perpetual beta', considering that 'users must be treated as co-developers'. Agile developments methods are therefore becoming popular on the Web, as well as languages that adhere to such software development principles (e.g. Ruby on Rails).

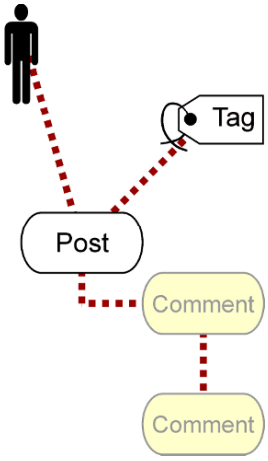

Fig. 3.2. The Social Web in simple terms: users, content, tags and comments

[3] http://www.w3.org/People/Berners-Lee/WorldWideWeb.html (URL last accessed 2009-06-09)