Information Systems 35 (2010) 215–236

Contents lists available at ScienceDirect

Information Systems

journal homepage: www.elsevier.com/locate/infosys

On social Web sites

Won Kim [a,*], Ok-Ran Jeong [a,1], Sang-Won Lee [b]

[a] Kyungwon University, Department of Software Design & Management, IT College, San 65, Bokejeong-dong, Sujeong-gu, Seongnam, Gyeonggi-do 461-701, Republic of Korea
[b] Sungkyunkwan University, Suwon, Republic of Korea

ARTICLE INFO

Article history:
Received 8 July 2009
Received in revised form
28 August 2009
Accepted 5 September 2009
Recommended by: D. Shasha

Keywords:
Social networking web sites
Social media web sites

ABSTRACT

Today hundreds of millions of Internet users are using thousands of social Web sites to stay connected with their friends, discover new "friends," and to share user-created contents, such as photos, videos, social bookmarks, and blogs. There are so many social Web sites, and their features are evolving rapidly. There is controversy about the benefits of these sites, and there are social issues these sites have given rise to. There are lots of press articles, Wikipedia articles, and blogs—in varying degrees of authoritativeness, clarity and accuracy—about some of the social Web sites, uses of the sites, and some social problems, and business challenges faced by the sites. In this paper, we attempt to organize the status, uses, and issues of social Web sites into a comprehensive framework for discussing, understanding, using, building, and forecasting the future of social Web sites.

1. Introduction

During the past 10 years, hundreds of millions of Internet users all over the world have visited thousands of social networking sites and social media sites. They have taken advantage of the free services of such sites in order to stay connected online with their offline friends and new online acquaintances, or to share user-created contents, such as photos, videos, bookmarks, blogs, etc.

There are so many social networking sites and social media sites that there is even a search engine for them [30]. Further, there are Web sites, such as Ning and KickApps, which make it possible for users to create social networking sites, and host those Web sites. Such US-based sites as MySpace and Facebook each claim over 250 million registered users [83]. YouTube, the video-sharing site, is the third most visited site among all Web sites, right behind Yahoo and Google. Barack Obama's election campaign for Presidency of the United States extensively used the Internet and social networking sites to get his message out, drum up donations, organize campaign volunteers, and get people to vote [95,44]. Paul Potts and Susan Boyle, contestants on the "Britain's Got Talent" talent-search television program, became overnight international singing mega-stars when their auditions posted on YouTube drew over 10–20 million views in a matter of weeks. Even a course on "Facebook" is offered in Stanford University, in which students are to build Facebook applications and find ways to attract users [14]. Twitter, the social networking site that features 140-character messages, called microblogs, are so popular in the US that it has spawned such words as twitter, tweed, and twitterati [64]. Someone even created an interactive book, for digital reading devices, that includes text, online video, and Twitter update stream [84]. Kind-hearted people now use Facebook and Twitter to locate the owners of lost items they find, such as wallets, mobile phones, digital cameras, etc. [85].

Refs. [33,100] highlight some of the social networking sites popular outside the US. Orkut is wildly popular in Brazil; half the Internet users in Brazil are registered

* Corresponding author. Tel.: +82 31 750 5517; fax: +82 31 750 5662.
E-mail address: wonkimtx@gmail.com (W. Kim).
[1] Tel.: +82 31 750 5831; fax: +82 31 750 5662.

0306-4379/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.is.2009.08.003

members of Orkut. It even inspired a popular song "I'm going to delete you from my Orkut." The 87% of the Internet users in the Philippines are members of Friendster. Orkut and Friendster originated in the US, but have been overshadowed by other sites and instead enjoy popularity elsewhere. Skyrock Blog was the communication vehicle among the youth during the 2005 riots in France. A significant portion of the traffic for Facebook comes from Arab countries. In particular, Egypt has the highest number of users, after the US, Canada, and UK. Facebook has become an organizing tool of choice for people in Arab countries attempting to promote freedom of speech, human rights, and democracy.

Social networking sites and social media sites have had both positive and negative impact. Many users spend excessive amounts of time creating and/or viewing self-aggrandizing or totally trivial updates on tiny details of their and their so-called online friends' activities. Many of them lose sense of what is appropriate, and end up losing employment or college entrance opportunities because of their having posted inappropriate remarks or photos on social networking sites or social media sites. Many pedophiles join popular social networking sites to hunt for preys. Recently, MySpace removed, under pressure from the US government, 90,000 members who are known sex offenders [106]. Many users post copyrighted materials without authorization, or pornographic or other illegal contents, forcing site operators to hire staff to remove them or risk getting sued by copyright holders or even have their sites banned by governments.

We set out to create a knowledge-sharing Web site. As part of the work, we wanted to understand what social Web sites are, what their essential features are and will be, how they benefit people, what problems they have posed, and how social Web sites are likely to evolve in the foreseeable future. We realized quickly that there is a dearth of academic papers dealing with social networking sites and social media sites, but there are lots of press articles, Wikipedia articles, and blogs about them. However, we found that many of them focus only one a small number of popular sites (Facebook, MySpace, Twitter, YouTube, Flickr), and most articles also get rapidly outdated, as the sites evolve fast. Further, most articles do not provide adequate explanations about how things work, such as the Facebook Platform and the Google Open Social. Even worse, many articles report what are only plans and hypes released by the sites' marketing departments. We had to study many press articles, blogs and Wikipedia articles. We had to cross reference them to understand what they say and also ensure accuracy of the information. Further, we joined a few dozen sites and experimented with most of their many features.

In this paper, we organize and summarize the four aspects of social Web sites that we believe would serve as a comprehensive framework for understanding the status and future of social Web sites. We believe such a framework would be useful to people who use, study, or build such sites. The four aspects are taxonomy of social Web sites, taxonomy of their essential features, taxonomy of the uses and benefits of social Web sites, taxonomy of the issues and challenges facing the sites, and a prognosis of how the sites are likely to evolve in the near future. To the best of our knowledge, there are very few papers or articles that attempt to bring together all these aspects of social Web sites. Papers or articles that do attempt to cover all these aspects of social Web sites, however, run the risk of giving a rather skewed view of the social Web sites, and also of quickly becoming obsolete. The sheer number of social Web sites makes it a challenge to classify them. Further, most of the social Web sites are moving targets, in terms the number of users, user demographics, features provided, business strategies, etc. In order to overcome these challenges, we have attempted to cast the discussions of each of the four aspects of social Web sites in terms of taxonomy that is simultaneously general enough and specific enough so as not to be highly susceptible to changes to social Web sites in the near future, say, the next 4–5 years. In other words, we feel that the taxonomies we use to cover the four key aspects of social Web sites should serve as a framework for discussing, studying, using, and even building social Web sites for the foreseeable future.

The remainder of this paper is organized as follows. In Section 2, we provide our definition and taxonomy of social Web sites. In Section 3, we review and provide taxonomy of essential features found in many of the social Web sites. In Section 4, we discuss how some of the social Web sites make their data and some of the features available to outside programs. In Section 5, we provide technical insights into the implementation and prototypical architecture of the social Web sites. In Sections 6 and 7, we provide taxonomy of the benefits and uses of social Web sites, and taxonomy of the problems facing social Web sites, respectively. In Section 8, we provide a prognosis of social Web sites based on today's trends. We conclude the paper in Section 9.

2. Taxonomy of social web sites

We define social Web sites as those Web sites that make it possible for *people* to form online *communities*, and *share user-created contents* (*UCCs*). The *people* may be the users of the open Internet or may be restricted to those who belong to a particular organization (e.g., corporation, university, professional society, etc.). The *community* may be a network of offline friends (whose friendship is extended to online), online acquaintances, or one or more interest groups (based on school attended, hobby, interest, cause, profession, ethnicity, gender, age group, etc.). The *UCC* may be photos, videos, bookmarks of Web pages, user profiles, user's activity updates, text (blog, microblog, and comments), etc. The *sharing* of the UCC includes, at the minimum, the posting, viewing, and commenting of the UCC, and may also include voting on, saving, and re-transmitting of the UCC. Roughly, we regard social Web sites as a union of social networking sites and social media sites.

The terms "social networking sites" and "social media sites" have already been loosely and widely used in press articles, blogs, press releases from the sites, etc., and the features of such sites are rapidly evolving. As such, we do

not feel that efforts to define social Web sites (for that matter, social networking sites and social media sites as well) more precisely than above are warranted.

Roughly, social networking sites are Web sites that allow people to stay connected with other people in online communities (see [16,102,110] for attempts to define social networking sites rather more precisely). Some of the most widely used social networking sites in the world today include MySpace, Facebook, Windows Live Spaces, Habbo, etc. Social media sites are Web sites that allow people to share UCCs. Some of the most widely used social media sites include YouTube, Flickr, Digg, Metacafe, etc. Users of many of the popular social Web sites are dominated by teens and those in low twenties. The numbers of male and female users are roughly equal [63].

Historically, social networking sites started before social media sites. Classmates.com (1995) and SixDegrees.com (1997) were the first social networking sites. Friendster (2002), MySpace, Bebo, and Facebook (2004) were the next batch of social networking sites. Social media sites Flickr (2004) and Youtube (2005) followed. Refs. [16,102,98] provide a history of social networking sites, and Naone [59] gives a history of microblogging sites, social networking sites that use microblogs for keeping "friends" updated and connected.

Our definition of social Web sites, although fairly loose, does exclude certain types of Web sites and parts of Web sites that allow people to post UCCs and share them. For example, the groups in portal sites (such as Yahoo Groups, South Korea's Naver cafes), blogs, online news sites, and dating sites do not, at least today, meet the definition of social Web sites, since they do not allow the users to form communities. Such sites as Ning and KickApps are sites that allow the users to create social networking sites and host such social networking sites. We do not regard them as social Web sites.

In this paper, for two reasons, we have chosen to study social Web sites, that is, both social networking sites and social media sites together, rather than focusing on one or the other. One reason is that the two types of sites have a large and significant intersection that is worth studying, that is, both types of sites are online social gathering places for people. Another reason is that the distinction between the two types of sites is fast disappearing, as social networking sites are adding primary features of social media sites features, that is, the sharing of UCCs; and social media sites are adding primary features of social networking sites, that is, personal profiles and forming communities.

Table 1 shows our taxonomy of social Web sites. Social networking sites are divided into general-purpose sites and vertical sites. Both the general-purpose sites and vertical sites are divided into open sites and closed sites. Open sites are open to the general users of the Internet, while closed sites are for members of particular groups or enterprises [26]. A general-purpose site, such as Facebook or LinkedIn, may include closed sites that are used, for example, as corporate intranets. There are various ways to define vertical sites. Ref. [97] shows one way, as follows:

sites for connecting with people based on schools attended, profession, etc.
- *classmates,*
- *business contacts*

sites for connecting with people based on gender, age group, ethnicity, etc.
- *women,*
- *children/kids, teen,*
- *ethnic groups*

sites for connecting with people based on common hobbies & interests
- *arts, cars/auto, cooking/food, education/books, event planning,*
- *foreign languages, family, fashion/clothing, finance, games,*
- *health/medical, movies, music, pets, politics, pop culture,*
- *real estate, shopping, religion, social action, technology, travel,*
- *miscellaneous.*

Each subcategory above can be further divided. Each subcategory has many sites. For example, health/medical networking sites include the likes of PatientsLikeMe, SoberCircle, DailyStrength, iMedix, I'm Too Young for This, MDJunction, MedHelp, WEGOHealth, TauMed, Healia, ICYou, SugarStats, Trusera, TuDiabetes, etc. [29].

Media types that divide the social media sites include blogs, bookmarks, consumer reviews, music, news, photos, videos, etc.

To give some idea about the global popularity of social Web sites, in Table 2 we provide a partial list of social Web sites that are popular in different continents/regions/countries of the world. The list is adapted from social marketing goldmine: top 100 social networking sites [27], the list: the world's top social networking sites [33], Lewin [47], the top ten social networking sites [101] and list of social networking websites [103].

Table 3, taken from Ref. [103], lists social Web sites that today have more than 30 million registered users. We note that Ref. [103] lists only what it defines as social networking sites, and omits such popular social media sites as YouTube and Flickr. Further, we caution that these numbers to some extent are only rough, since they keep changing and also many user accounts are "dormant" accounts (not actively used after registration and some initial use). However, even if we are to take only perhaps one-fourth of the published numbers, they are substantial, and the relative orderings of the sites may tend to be unchanged.

Table 1
Categories of social Web sites.

Social networking sites	General sites	Open sites Closed sites
	Vertical sites	Open sites Closed sites
Social media sites	Media types	

Table 2
Popular social Web sites by continent/region/country.

Continent/region	Dominant social Web sites
Africa	Hi5 (Angola, Central Africa), Facebook (Egypt)
America (North)	MySpace, Facebook, YouTube, Flickr, Nexopia (Canada), Netlog (Canada)
America (Central and South)	Orkut (Brazil), Migente, Hi5, Sonico, Facebook (Panama)
Asia	Friendster (Southeast Asia), Orkut (India, Pakistan), Xianonei (China), Xing (China), Cyworld (S. Korea), hi5 (Thailand), YouTube (Japan), Mixi (Japan), Hi5 (Mongolia)
Europe	Badoo (UK, Europe), Bebo (UK, Ireland), Friends Reunited (UK), Facebook (UK), Hi5 (Portugal, Cyprus, Romania), Tagged, Xing, Skyrock (France, French speaking region), Studivz (Germany), Hyves (the Netherlands), iWiW (Hungary), Nasza-klasa.pl (Poland), IRC-Galleria (Finland), LunarStorm (Sweden), Netlog, Nettby (Norway), playahead (Sweden, Denmark, Norway), Odnoklassniki.ru (Russia, former Soviet republics), V Kontakte (Russia),
Middle East	Facebook (most Arab countries)
Pacific Islands	Bebo (including New Zealand)

Table 3
Social Web sites with over 30 million registered users.

Site name	Users (in millions)
Adult FriendFinder	31
Bebo	40
BlackPlanet	50
Facebook	309
Flixter	63
Friendster	90
Habbo	117
Hi5	80
LinkedIn	35
MySpace	253
Netlog	36
Odnoklassniki.ru	30
Orkut	67
Reunion.com	51
Tagged	70
Windows Live Spaces	120

Table 4
Top 10 trafficked social media sites.

Site name	Alexa ranking	Primary shared media
YouTube	3	Videos
Flickr	39	Images
Digg	127	Bookmarks
Metacafe	143	Videos
Stumbleupon	380	Cool content
Technorati	440	Blogging
Scribd	538	Articles
Blogcatalog	784	Blogs
Delicious	1096	Bookmarks
Propeller	1140	News

Table 4, taken from Prelovac [65], is a list of top social media sites, measured in terms of Alexa ranking (rather than the number of registered users). Alexa rank measures Web site's popularity based solely on traffic to that site.

Table 5
Top 10 trafficked social bookmarking sites.

Dig.com
Yahoo Buzz
Technorati
Delicious
Propeller
Stumbleupon
Reddit
Mix
Fark
MyBlogLog
Slashdot

For example, YouTube is the third most visited Web site at this time, after Yahoo and Google. Flickr is the 39th most visited Web site, etc.

Table 5, taken from Ref. [28], shows the top 10 bookmarking sites, where the ranking is based on a combination of inbound links, Google page rank, Alexa rank, and US traffic data from Compete and QuantCast.

3. Essential features of social web sites

There are thousands of social Web sites, and the sites differ in details and layouts of the features they provide. Further, they continue to add new features and make changes to existing features. It is not feasible, or even meaningful, to describe full details of all the features of all the sites. However, if we start from the primary objectives of social Web sites, we are able to extract essential features of the sites. The objectives of the sites are to enable the formation of online communities, interactions among members of such communities, and the sharing of UCCs. In this section, we will provide taxonomy of essential features that can be extracted from today's popular sites. To make the description of each feature concrete, we provide brief descriptions of how a few representative sites manifest the feature. Table 6 lists the essential features that we have extracted from today's most popular social Web sites. We note that although all social Web sites support all these features, they manifest these in different ways and in varying degrees of sophistication. Readers interested in learning about each feature in detail as supported in any particular site should join the site and use it. Further, we note that the list of features may grow somewhat as social Web sites evolve.

3.1. Personal profiles

Most social Web sites have members create and manage personal profiles, that is, homepages. However, they differ in the types of information included. Further, some sites have members specify privacy settings in order to control who (e.g., everyone, friends only) may access what types of information on their personal profiles. Personal profiles on social networking sites used to be more elaborate than those on social media sites. However, this distinction is becoming blurred, as social media sites are making the personal profile more elaborate.

Table 6
Essential features of social Web sites.

1. Personal profiles
2. Establishing online connections
3. Participating in online groups
4. Communicating with online connections
5. Sharing UCCs
6. Expressing opinions
7. Finding information
8. Holding the users

For example, the personal profile on Twitter, a social networking site, is currently very bare; it only includes the name and location of the member. The personal profile on YouTube, a social media site, includes basic information, such as name, photo, birthday, gender and e-mail address. In addition, it includes the member's interests, such as favorite movies & shows, favorite music, and favorite books. YouTube computes and displays "recommended videos" based on such information. The personal profile on Facebook is elaborate. It includes 4 types of information: basic information, personal information, contact information, and education & work information. The basic information includes the name, photo, age, birthday, relationship status, gender of interest, and type of relationship desired (friendship or dating). Personal information includes interest, favorite music & TV shows, movies, books, and quotations. Contact information includes mobile phone, landline phone, school mailbox, address, etc. Education and work information includes the names of schools attending/attended (high school, college), and current employer. Based on the information in the personal profile, Facebook computes and displays potential friends.

3.2. Establishing online connections

Many social Web sites provide facilities for a member to discover connection ("friend") candidates from existing members. The facilities include automatic discovery of existing members of a site from the email and messenger address books of a new member, browsing of all existing groups on the site, a friend-recommendation engine that suggests friends of friends, and a keyword-based search engine for looking up members' names.

Such sites as Flickr, Facebook, Twitter, LinkedIn, Yelp, etc., search the address books of popular email or messenger systems that new members may be using, such as Gmail, Yahoo mail, AOL mail, Windows Live Hotmail, Windows Live Messenger, to identify existing members of the sites (i.e., offline friends). Such sites as Facebook, LinkedIn, MySpace, etc., also assign new members to some of the existing groups, based on their schools attended, current employers, physical locations, etc. Further, most sites allow members (and often, non-members as well) to browse all the groups. While browsing the groups, the member may select certain existing members and try to connect with them.

Many social Web sites, such as Facebook, MySpace, LinkedIn, etc., send "friend request" notices to existing members whom the new member asks to connect with. Usually, only when the existing member confirms the friend request notice, the two members are established as online friends, and they can view each other's personal profiles. Some sites, such as Twitter, do not require confirmation from existing members to make them "friends" (or become a "follower", in Twitter's parlance). (There is an exception to this practice, but we will not elaborate on it.)

3.3. Participating in online groups

As remarked earlier, many social Web sites support a small number of default groups and assign new members to one or more of them. They also allow members to explicitly form new groups, and/or join them. An explicit group is assigned a manager. Members and non-members can both view all the UCCs in all the groups. However, only members may post UCCs.

For example, LinkedIn provides several categories of default groups, including alumni group, corporate group, conference group, networking group, non-profit group and professional group. It allows members to create new subgroups within these types of groups. By using the personal profiles, LinkedIn also facilitates the forming and joining of groups by linking current and past colleagues and classmates (invite contacts, import contacts). Sites such as Facebook, MySpace, Flickr, etc., host thousands of groups that members have created. Flickr, MySpace, Bebo, etc., allow three types of groups to be formed: public, public but by invitation only, and private. Any member may join a public group, while only those who receive an explicit invitation or consent from the group's manager or existing members may join a by-invitation public group. Only members who can present certain credentials, such as a user ID and password, may join a private group. A private group cannot be browsed by anyone who is not a member of the group.

YouTube members may create groups called "channels," and have other members join ("subscribe" to) them. Facebook members may also create groups called "pages," and have other members join as "fans." These are often used by celebrities and businesses to promote their brands.

3.4. Communicating with online connections

Social Web sites provide various facilities for members to use to communicate with their online connections, that is, friends and other members. These include email, instant messaging, text messaging, and public and private bulletin boards, and even Internet phone services. Such sites as MySpace and Facebook allow their members to use the messaging and phone call facilities of Internet phone services, such as Skype. Further, on behalf of the members, the sites send member updates and notices (e.g., friend request notice) using emails or text messages to members' friends. They also send updates and notices to members of groups (e.g., notices regarding a group's status and activities). The sites also display friends' updates to members' public and private boards. Some

sites try to block spam by requiring the member to verify a Completely Automated Public Turing Test to tell Computers and Humans Apart (CAPTCHA) code before sending an email or text message.

For example, Twitter allows members to send messages to friends on what the members are doing. A message, which is restricted to 140 characters, may be sent to friends' mobile devices and their Twitter accounts. A member may have one or more "followers," and many members may be "following" a member. Facebook provides a mini bulletin board (called a "wall") to post a member's message for all friends to see and respond to. LinkedIn provides an "answers" function to allow members to answer questions posted by other members, and to refer the questions to their online connections.

3.5. Sharing UCCs

Most social Web sites allow members to post various types of UCCs, such as blogs, microblogs, photos, images, music, video, bookmarks, and text. Friends and others may view or play these UCCs, send their links (URLs) to their online connections and even offline friends, and save them into private collections for future viewing and sharing with others. Social media sites provide richer facilities for sharing UCCs than social networking sites.

For example, Twitter primarily allows the sharing of short text messages. The only non-text UCCs it allows two photos of the member: one head and shoulder photo and one background photo. The member may save in his/her folder the links to selected ("favorite") text messages. MySpace allows the sharing of blogs, photos, videos, playlists of music contents, etc. YouTube supports the posting and viewing of videos, including TV program clips, music videos, user-created video blogs, and short original videos. Members may add titles and tags when posting videos. YouTube adds the titles and tags to its search engine to support keyword-based search of the videos.

3.6. Expressing opinions

Most social Web sites allow members to leave comments on the UCCs. Some sites allow members to vote on them, too. The voting may take the form of ranking (e.g., checking 3 stars out of 5 stars), or marking the UCC as a "favorite," or flagging it as spam or inappropriate. Some sites present all comments as a flat list of comments, while others organize comments as a 2-level hierarchy, that is, they allow comments on a comment. Some sites display the comments in posting timestamp order, and some in reverse timestamp order. Some sites give a few additional sorting options for displaying the comments, such as the number of comments, the number of emails that sent the UCC, the number of favorable votes received, etc.

For example, Digg provides two clickable buttons, "digg it" and "bury," to allow members to express strong liking and disliking, respectively, of the linked contents (primarily news articles), and even of the comments other members have posted. YouTube allows members to post video responses and text comments on videos. YouTube also allows members to flag videos and comments as spam. As we will discuss shortly below, Facebook provides open application programming interface to third-party developers. Facebook makes such third-party developed applications accessible to its members. Facebook members may rank the applications, and recommend them to other members.

3.7. Finding information

Both members and non-members have two types of facilities to find the information they need on social Web sites. The facilities are keyword-based search engines and browsing. The search engines can be used to look for the names of people, names of groups, and particular UCCs. Browsing can be done on selected groups and UCCs in a particular category. Most sites allow keyword-based search of text-based contents. Some search engines are fairly powerful, while others are crude. The users can also browse the results of keyword-based searches, and groups and UCCs that are suggested as related to the group and UCCs the users currently view. All social Web sites provide categories for the UCCs stored, so that users may browse UCCs in a specific category.

For example, Twitter supports search of only people's names. LinkedIn supports keyword-based search for each of the categories of people, jobs, companies, answers, inbox, and groups. LinkedIn also supports advanced search of the people, jobs, and answers categories. Advanced search of the people category admits the first name, last name, title, company, school and location. Search results may be ranked in terms of "relationship," "relationship and recommendation" (combined), and user-supplied keyword. YouTube supports keyword-based search for the three areas "all," "channels," and "playlists." Basic search is based on keywords included in the user-furnished titles, tags and comments. Facebook shows search results by people, Facebook-designated pages (e.g., iPhone page), groups, events and applications. "People" is the primary order for displaying search results, followed company, school or e-mail.

Many social Web sites organize their data in terms of people, groups, and UCCs, and allow users to browse each category of data. Some sites organize the UCCs in terms of a relatively small number of categories. For example, Digg has 8 categories: technology, world and business, science, gaming, lifestyle, entertainment, sports, and offbeat. Digg further organizes each category in terms of most recent, top in 24 h, 7, 30, and 365 days. The categories merely represent topics of greatest interest to the users (as determined by the site operators); they do not represent a rigorous ontology.

3.8. Holding the users

Many social Web sites provide various features that are designed to have the users spend a long time on the sites, and have them return frequently. Many sites display data related to the data the users specifically seek. Many sites

also apply special designations, such as "popular" and "recent," to data that may pique the interest of the users who may not necessarily be searching for such data. Further, many sites provide various special interest pages, such as the "Market Place" on Facebook.

For example, YouTube displays thumbnails of 30 related videos, related channels and related playlists, when a user watches a video. If a user clicks on a recommended video titled, say, "Connie Talbot: over the rainbow," a singing performance featured on a television program, Britain's Got Talent, thumbnails of other videos are shown, including other videos on Connie Talbot and popular videos of performances on the same show, such as "Paul Potts" and "Susan Boyle." YouTube displays on the main page the "Recommended for You" videos related to the videos that the user watched earlier. The related videos include popular and eye-catching spotlight videos, featured videos, and rising videos, enticing the user to keep watching video responses, related videos, promoted videos, etc. Further, YouTube maintains "my videos," "my favorites," "my playlists," "my channel" and "my subscriptions." Facebook and MySpace provide various types of updated information on members' friends and friends' friends. The information includes which groups the friends have just subscribed to, whom the friends have communicated with, whom the friends have made friends with, which friends updated their profiles, etc. Flickr shows a "folksonomy" of popular keywords on the main page, enticing users to click on some keywords, related tags and photos in sub-directories. Flickr also displays new members and popular members on the "people" column at the bottom of the main page, inducing users to click and spend time on some of the members.

4. Opening social web sites to external programs

Many application developers wanted to be able to develop applications that can make use of the enormous amounts of data stored on social Web sites, and make those applications available to the members of those sites. Many users of sites have found it bothersome to go through similar procedures and provide a similar set of data to establish online connections on a site with connections they have already established on other sites, because the sites do not allow the users to directly export their data to other sites [58]. Many users have also come to want to be able to keep track of the activities of their online connections on one site from another site. In response to these demands, Facebook published an open Application Programming Interface (API), and Google followed suit with its own API called OpenSocial [42]. The two subsequently offered open source software for linking a social Web site with other social Web sites. In this section, we discuss these two developments with social Web sites.

4.1. Open API

MySapce, Facebook, Flickr, Twitter are some of the social Web sites that have published an open API, opening up their software to third-party application developers. Google, in collaboration with MySpace and some other sites, also published an open API, known as Open Social. Below we highlight the Facebook API and the Open Social API.

Facebook published its API in May 2007 [36,17,18]. The Facebook API, or Platform, consists of a Facebook's variant of HTML, named Facebook Markup Language (FBML), and a Facebook variant of SQL, named FQL (Facebook Query Language). In other words, the Facebook API is not based on open standards. Such popular social Web sites as Bebo and Meebo support the Facebook Platform.

On one hand, the applications developed using the Facebook API can make it possible for the developers to reach a segment of the Facebook members. Facebook members may find the applications interesting or useful, since they can access the Facebook database of members, friends, groups and events. On the other hand, the applications expand the repertoire of services that Facebook makes available to its users. For example, iLike, a music site, by porting its software to the Facebook Platform, has made it possible for Facebook users to use the site's services while on Facebook. Users of iLike can automatically import their Facebook friends data.

Fig. 1 illustrates the interworking of Facebook and a third-party application (iLike in particular) developed in the Facebook API. Let us consider four cases of a user (John)'s membership in Facebook and iLike.

(case 1) John is a member of Facebook, but is not a member of iLike.

John can use iLike on Facebook via the "iLike sidebar." John can use all the functions provided by iLike. John's personal profile is stored on a Facebook server, and John's iLike activities data are stored on an iLike server. John's Facebook friends data is accessible only on Facebook, and John's iLike friends data is accessible only on iLike.

(case 2) John is not a member of Facebook, but is a member of iLike.

John's information (profile, friend, interest, etc.), is saved on an iLike server. iLike recommends John to link to Facebook. If John links to Facebook, John is registered as a Facebook member. Then case 2 becomes case 4 below.

(case 3) John is a member of both Facebook and iLike, but does not link the two accounts.

If John does not link the two accounts, John is regarded as two different users.

(case 4) John is a member of both Facebook and iLike, and links the two accounts.

John's data is shared by Facebook and iLike, and the data exists on both servers. Key shared data includes recently played songs, favorite music, saved points (iLike Challenge score), favorite artists, etc.

One year after Facebook launched its Platform, Facebook reported 24,000 applications, and 350,000 developers developing on the Platform [48]. The phenomenal acceleration in the growth of Facebook membership occurred shortly after the launching of the Facebook Platform. Kincaid [43] reports, however, that, although there have certainly been successful applications, most applications have turned out to be pretty useless.

About a half year after Facebook's launch of the Facebook Platform, Google launched Open Social. Unlike

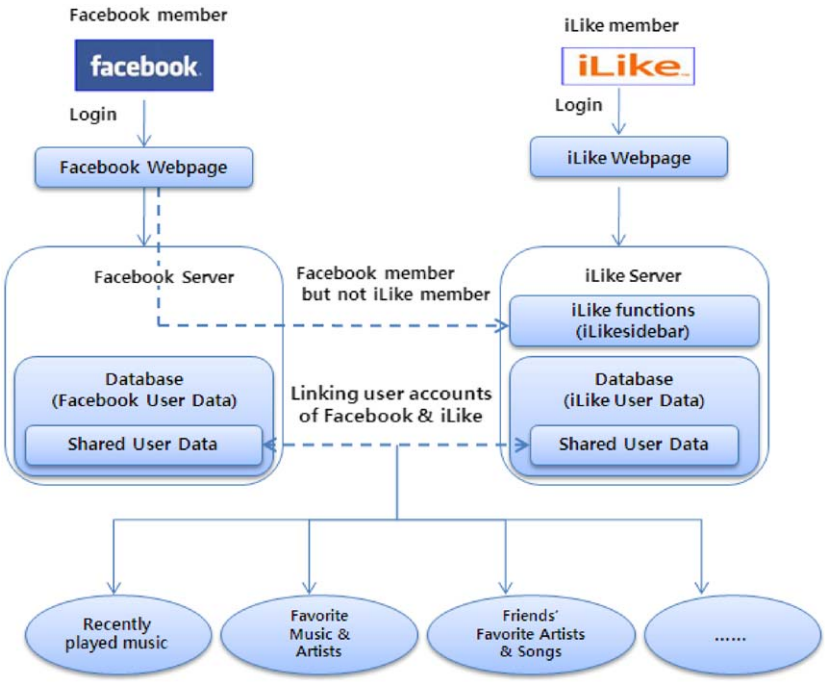

Fig. 1. Interworking of Facebook and a third-party application.

the Facebook Platform, Open Social provides a set of APIs that are based on open standards JavaScript, and HTML. The set of three APIs that comprise Open Social include functions for accessing member profiles, friends data, and activities data. In other words, Open Social, when embedded in a Web site, instantly makes the Web site a social Web site. The Web site is empowered with such basic social networking functions as member registration, member profile, review, comment, photo-sharing, etc., by accessing the corresponding data on the site. Good [36] characterizes Open Social as consisting of a "container" and "plug-in applications" embedded in the container. The container is the Web site that embeds Open Social. Popular social Web sites that have adopted Open Social, that is, have become containers for Open Social, include MySpace, Hi5, Plaxo, LinkedIn, Orkut, Friendster, Six Apart, etc. Plugged-in applications of Open Social include iLike, Slide, Flixter, Rock You, etc. [36].

Facebook, in a separate move, has also made software available for third-party developers to use to develop applications that can run inside Web browsers on a wide variety of mobile devices [86]. Further, the users of the applications do not need to log on to Facebook. Twitter has from the start allowed third-party developers develop applications that can access and post updates to Twitter. Using these software, some developers have developed applications that create special groups and see their updates, blend together Facebook and Twitter updates, etc.

4.2. Connecting with other sites

The open platforms that various social Web sites provide make it possible for third-party developers to develop applications on the platforms that make use of the sites' databases. However, the users of multiple social Web sites have the onerous chores of having to establish friends multiple times, and to log into different sites to connect up with their friends. Facebook and Google each extended the Facebook Platform and Open Social to address this problem. They are Facebook Connect [48] and Google Friend Connect [73,18].

Facebook Connect and Google Friend Connect make it possible for a member of one social Web site (say, Facebook) to link to another social Web site (say, Yelp), using a single sign-on credential, and share member data across the two sites. Using a secure authorization API, the member of one site (say, Facebook) can connect and interact with new friends on the other site (say, Yelp).

There are two ways in which a Web site may be connected to Facebook. One way is for users to join a Facebook-connected site (say, www.techcrunch.com) and explicitly connect with Facebook using a "Facebook Connect" button provided on the site. When connection with Facebook is made, the user is connected with Facebook members who are also members of the site (www.techcrunch.com) and get to view their Facebook "walls" and status updates. When the user leaves comments on the site, they are automatically published on the user's Facebook wall. Further, the Facebook-connected site creates a profile page for the user using the user's Facebook data (personal profile, friends, photos, events, groups, etc.). Another way requires users to log in only as Facebook users. For example, Course Hero (www.coursehero.com), a social e-learning site, does not provide its own accounts to users. A user can only log on to Course Hero using his Facebook account information. Then Course Hero creates a profile page for the user using

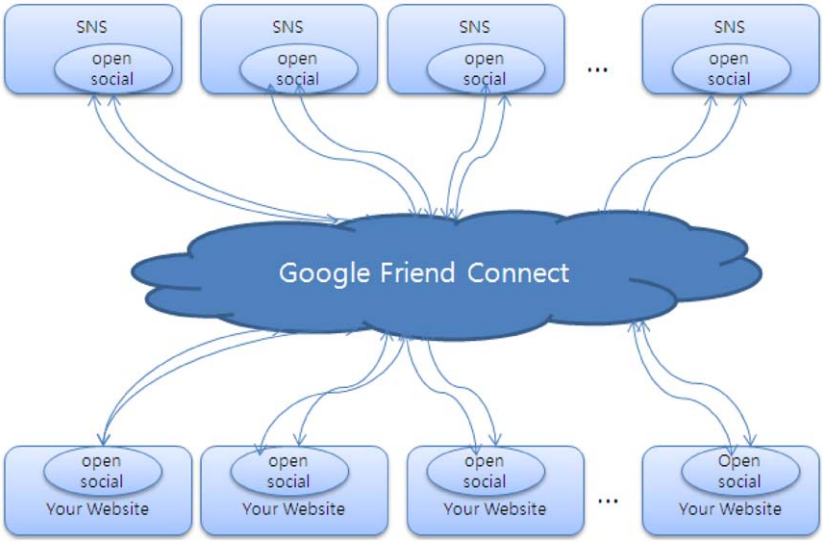

Fig. 2. Google friend connect.

the user's Facebook data (personal profile, friends, photos, events, groups, etc.).

Facebook Connect supports a trusted authentication mechanism, transfer of friends list, and transfer of privacy settings across sites. Google Friend Connect uses open standards OAuth for secure data transfer between sites, and OpenID for single sign-on in multiple sites. Fig. 2 illustrates the architecture of Google Friend Connect.

5. Implementation and architecture of social web sites

Most social Web sites, including such gigantic ones as Facebook, MySpace, YouTube and Flickr, have been developed, at least initially, by a surprisingly small group of engineers. Typically, the group included a few architecture designers, a few feature developers, a few network engineers, a few system administrators, and one database administrator. The engineers who created the sites had evidently no idea that the sites would grow, in terms of traffic, number of users, and amount of data posted and viewed, so explosively in a matter of a few years. It is incredible that a relatively small group of young engineers, totally caught off guard, has managed to address the tremendous stresses on performance, scalability, and availability of their site, and has kept their site running. In the absence of definitive and current descriptions of the architectures of the major social Web sites in the open literature, we have examined a number of PowerPoint presentations and blogs on the architectures of several well-known sites. The architecture descriptions of the sites are often from different perspectives and with different emphasis. There is no common format for even the site statistics (e.g., number of requests per second, number of accesses to the database, number and size of the UCCs stored, etc.). Further, the available descriptions of the architectures and site statistics on some sites are recent, while those on others are dated. As such, in this section, we are only able to provide general characteristics of the implementation and architectures of the major sites.

5.1. Implementation

Facebook, YouTube anf Flickr have been developed on commodity hardware using the LAMP open source software stack, namely, the Linux operating system, the Apache Web server, the MySQL relational database system, and the PHP Web programming language [3,112,9,4]. MySpace has been developed using Microsoft technologies, including the Windows operating system, Active Server Pages (ASP).NET 2.0, the Internet Information Services (IIS) Web server, and the SQL Server 2005 database system [7] Twitter has been developed using Linux, the Ruby programming language on the Rails application server, the Mongrel Web server, and MySQL. LinkedIn has been developed using Sun Microsystems' Solaris operating system, Java, the Tomcat and Jetty open source Java application servers, the Lucene open source text search software, and the Oracle and MySQL database systems [5].

One reason for the adoption of the open source software and commodity hardware is obviously cost consideration, since most sites have started as venture-financed startups and had to be careful with expenses. Another reason is the relative maturity and reliability of the software. Further, when the engineers waged mighty struggles to address performance and scalability bottlenecks, the availability of the source code made it possible for them to make necessary changes to the software [9].

Most major sites were initially tiny, running on a small number of servers. They typically had one master database and one backup database. The design of the early versions of these sites, and the sites that remain small today, centers around various types of data and functions (or methods in object-oriented parlance) that apply to each type of data. The types of data include

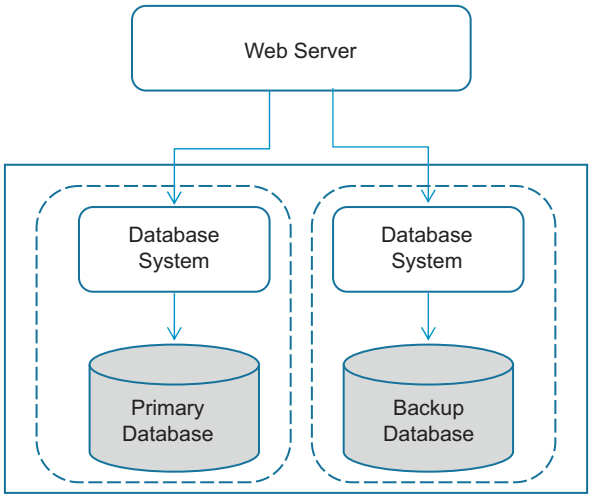

Fig. 3. Prototypical architecture of small social Web sites.

member registration data, videos posted by members, photos posted by members, member activities data, member's friends data, etc. The member registration function reads member ID (or email address), checks for duplication against the member registration data, reads password; saves the member ID and password; and creates an index entry for the member ID (for subsequent access to the member ID). The video posting function reads and stores the video being posted (against the member ID of the member who is posting the video), along with the title, description, and tags supplied by the member for the video.

The prototypical architecture of small social Web sites is shown in Fig. 3. It consists of a Web server and a database server. The Web server includes a script code and a script engine to run the script code, besides a file system for storing static contents, such as HTML documents and images. The script code generates dynamic HTML pages from the data stored in the database. For performance, scalability, and availability reasons (to be discussed further shortly), the Web server, the database server, and the database may be duplicated.

5.2. Architecture challenges

Although Twitter is not a gigantic site today, we show its site statistics below [8]:

- over 350,000 users
- 600 requests per second
- average 200–300 connections to the Web server per second (800 per second at peak)
- 2400 requests per second to MySQL
- Over 16 Gigabytes of memcache used (to be discussed later)
- 1 MySQL server on a single 8-core computer and 1 slave server
- 8 SUN × 4100 servers

The following are the site statistics on the gigantic MySpace [7].

- 300 million users
- processes 1.5 billion page views per day, and 2.3 million users per day
- pushes 100 Gigabits per second to the Internet
- over 4500 Web servers
- over 1200 cache servers (with 16 Gigabytes cached in RAM)
- over 500 database servers

As sites such as YouTube, MySpace, Facebook, Flickr, LinkedIn, Twitter, etc., explosively grew, the engineers had to solve three architecture challenges: performance, scalability, and availability. To understand and solve the bottlenecks on performance and scalability, the engineers have had to start using system and network monitoring tools, and Web traffic and log file analysis tools. The performance challenge is to provide real-time response to every user request. User requests include searching for a user or a UCC or a group using keywords, registering a new member, logging in an existing member, posting and viewing a video, posting and viewing a photo, discovering friends from an email or messenger address book, browsing the members of a group, browsing the search results, etc. As the traffic to a site increases, the number of members increases, and/or the amount of data stored in the site increases, performance of the site is bound to degrade. The scalability challenge is to maintain the performance at the same level even as the traffic, number of members, and the amount of data stored increase. The availability challenge is to have the site provide all services and access to all the data stored to all users at all times.

Fig. 4 shows the prototypical architecture of gigantic sites, such as YouTube, MySpace, Facebook, Flickr, etc. Below, we describe the key components of the architecture, and techniques used in meeting the three architecture challenges.

5.2.1. Speeding up the web servers

As any access to a Web site and resources linked to it (i.e., contents) must go through the Web server, the sites need to ensure that the Web server does not become overloaded. The sites manage incoming requests by using various means, such as firewalls and HTTP traffic manager to block or redirect unwanted or ill-formed traffic, and a traffic shaper to smooth out peaks in requests. Beyond this, such major sites as YouTube, Facebook, and MySpace have adopted the lightweight Web server lighttpd, instead of Apache, for serving videos. However, the engineers have determined that the key bottleneck for the Web servers is the time it takes waiting for responses to remote procedure calls (RPC).

For higher performance, gigantic sites use Web server farms, consisting of multiple, more powerful computers with more main memory and hard disk drive capacity.

Gigantic sites have also introduced a load balancer and/or reverse proxy servers in front of the Web server

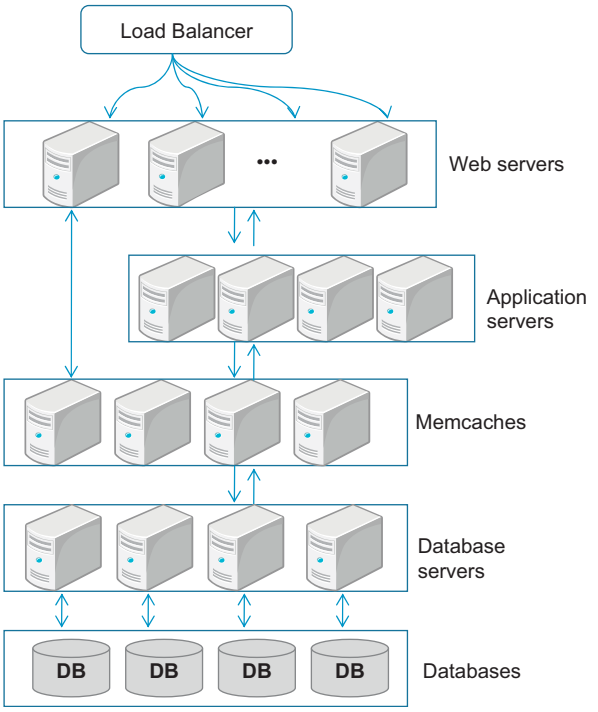

Fig. 4. Prototypical architecture of major social Web sites.

farm. A reverse proxy server is installed to load balance the Web server farm and to cache both static and dynamic content. It also provides an additional layer of security. Such major sites as Facebook, YouTube and Flickr have adopted the open source Squid reverse proxy server.

5.2.2. Use of main-memories as cache

Major sites, such as MySpace, Facebook, Flickr, YouTube, use memcaches extensively. Memcache is an in-memory hash table that is distributed across multiple servers. It has become an essential component of the architecture to meet the performance and scalability challenges. It listens to the TCP (transport control protocol) socket for requests. Hash tables are maintained in memcaches to rapidly respond to metadata keyword searches, such as the member ID. The size of memcaches is often over 16 Gigabytes.

Further, some major sites, such as YouTube and MySpace, store most frequently accessed contents and less frequently accessed contents on separate server farms, serving most frequently accessed contents out of main memories of one server farm. In particular, MySpace employs more than 1200 cache servers [9].

5.2.3. Database partitioning and replication

Major sites have to deal with huge amounts of data of various types, including metadata (member data, friends data), photo data, thumbnail data, video data, etc. The sites assign different server farms to manage different types of data.

Major sites, such as YouTube, MySpace, Facebook, Flickr, LinkedIn, etc., partition the database based on services (e.g., member profiles, emails, news, videos, etc.). They use a different application server farm for each service. This naturally distributes the load on the application servers and database servers. Further, the sites keep multiple copies (i.e., replicate) all the databases. This is to address the performance, scalability, and availability challenges.

Major sites extend database partitioning and replication to the level of relational tables. That is, they partition (i.e., decompose) a huge relational table (i.e., rows and columns) into multiple smaller tables comprising of some of the columns (i.e., "vertical decomposition") and/or rows (i.e., "horizontal decomposition") of the original table. They also maintain multiple copies of some of the table partitions. For example, such sites as YouTube, Flickr, LinkedIn, and MySpace decompose the members table into separate partitions, and replicate each partition [9,7]. MySpace keeps one million users in each partition; and needs 300 partitions to store all 300 million registered members. Partitioning and replication can each help meet the performance, scalability, and availability challenges. However, if done poorly, partitioning can result in needless and expensive joins; and replication creates the new problem of synchronizing all copies when one gets updated.

We note that most, if not all, major sites do not guarantee integrity and consistency of their databases. For example, some data posted by the users (e.g., comments on discussion boards, friends status, UCCs) get lost when hard disk drives fail, some updates to data (e.g., friends status, online connections) are not reflected when they logically should, etc. Today's relational database systems, such as Oracle, DB2, SQL Server, and MySQL, have been designed to guarantee database integrity and consistency in the presence of multiple concurrent transactions (some writing data to the database, and some reading from the database) and computer failures (i.e., servers and hard disk drives). This is because they have been designed to support "mission-critical" online transaction processing (OLTP) applications, such as bank deposit and debit, order entry, etc. However, to guarantee database integrity and consistency, database systems have to do many things, which hinder performance. Since keeping track of friends data and UCCs, etc. on social Web sites is not exactly "mission critical," some loss of data integrity and some data inconsistency, while annoying to many users, can be relatively more tolerable. As such, the engineers fighting the performance and scalability battles in the major sites have taken liberty with data integrity and consistency.

5.2.4. Use of server farms

Gigantic sites use server farms for each of the three key components of the site, rather than merely doubling it. For each server farm, there is a control server that manages distribution of work, load balancing, and dynamic reconfiguration of the server farm (i.e., adding a new server, taking out a server). The use of server farms helps meet the performance, scalability, and availability challenges.

5.2.5. Organizing hard disk drives

Some of the major sites have adopted disk arrays, and MySpace has adopted virtualized storage using storage

area network (SAN) to reduce disk I/O bottleneck and improve storage availability [7].

5.2.6. Redundant servers and database backup

As a major site consists of thousands or tens of thousands of commodity hardware, server failure occurs frequently. Major sites pair each server with a backup server. This helps to meet the availability challenges. When a server crashes, its backup takes over the processing load.

All sites back up the database on a regular basis in order to recover from crashes of hard disk drives and servers. They back up the database in its entirety or selected parts once a day, a week, every 2 weeks, or a month.

5.2.7. Moving to the 64-bit architecture

MySpace has moved to the 64-bit processor architecture, including 64-bit Windows 2003 and 64-bit SQL Server 2005, in order to improve performance [7].

6. Uses of social web sites

Social Web sites have managed to attract hundreds of millions of people and mountains of UCCs from all over the world. Now many users can derive benefits from the availability of the enormous number of people and the enormous amounts of UCCs on social Web sites.

In this section, we provide taxonomy of the uses of social Web sites. We group the users of social Web sites into three types: namely, individual users, businesses, and government. We examine the uses and benefits of social Web sites for each of these three types of users. Table 7 summarizes the uses of social Web sites. We note that each use requires some or all of the essential features discussed in Section 3. The list of features that enable the use is summarized in the "enabling features" column of Table 7.

6.1. Individual users

6.1.1. A new means of communication

Social Web sites have become a new means for people to communicate with others, beyond email, text messaging, and the mobile phone. These sites offer a new asynchronous (that is, not requiring real-time response) means of communication between a user and the user's online community. The user can broadcast a message or UCC to a large number of people, and view and respond at any time to messages and UCCs posted by other people. The online community for a user may consist of offline friends who are members of the same site, new online friends (actually, merely acquaintances [67]), and online groups the user may join. The user may post new UCCs and update the user's profile, so that the user's online network (friends, members and non-members) may view them at any time. Conversely, the user may view at any time the UCCs that online friends, groups and other users post; and receive updates to the status and activities of the user's friends and groups. Although many of the UCCs and updates are irrelevant or trivial to a user, some can be helpful and interesting. In any case, it appears that many people, especially the teenagers, have become addicted to this new mode of communication. Logging onto social Web sites, along with checking emails, has become a pre-breakfast "must activity" for many teenagers and families in the US [87].

Many celebrities, such as, US President Barack Obama, Oprah Winfrey (a US talk show hostess), Shaquille O'Neil (a US basketball player), Britney Spears (a US singer), Martha Stewart (a US television personality), Demi Moore (a US movie actress), etc. have joined social networking sites, such as Facebook and Twitter [52,107]. One of their motivations may be self-promotion, that is, to sustain their high visibility with the general public. Further, many of them may have other people ghost write for them [24]. In any case, these people are using social Web sites as a new means of communicating with their fan bases.

Table 7
Uses of social Web sites.

Users	Uses	Enabling features	Essential features of social Web sites
Individual users	New means of communication	1, 2, 3, 4, 5, 6	1. Personal profiles 2. Establishing online connections 3. Participating in online groups 4. Communicating with online connections 5. Sharing UCCs. 6. Expressing opinions 7. Finding information 8. holding the users
	New source of knowledge	2, 3, 4, 5, 6, 7, 8	
	Participating in online special interest groups	3, 4, 5, 6, 7, 8	
	New source of entertainment	1, 2, 3, 4, 5, 6, 7, 8	
	New venue for self expression	1, 2, 3, 4, 5, 6	
	Online directories of people	1, 7	
	Miscellaneous		
Businesses	Marketing and customer relationship management	1, 2, 3, 4, 5, 6, 7, 8	
	Corporate intranets	1, 2, 3, 4, 5, 6, 7	
	External peer networking	1, 2, 3, 4, 5, 6, 7	
	Miscellaneous		
Government	Governance	1, 2, 3, 4, 5, 6, 7	
	Law enforcement	1, 3, 4, 5, 6, 7	
	Election campaign	1, 2, 3, 4, 5, 6, 7	
	Legal proceedings	1, 3, 5, 6, 7	

6.1.2. A new source of knowledge

The UCCs make social Web sites a new source of collective knowledge, beyond the knowledge sources that the Internet search engines can dig into. The UCCs include photos, videos, music, blogs, comments, etc. Further, the user may engage in a one-on-one or one-to-many dialogs with his online network in order to seek or give answers to specific questions of importance to him (e.g., Why am I experiencing bad side effects of a certain drug after surgery? Can anyone recommend a good dentist in a particular area of a city?).

6.1.3. Participating in online communities

Social Web sites are in essence online communities, giving members opportunities to interact with a large number of people, keeping up with the world, receiving help, and occasionally helping others. Although such gigantic sites as Facebook, MySpace, YouTube, Flickr, etc., and other major sites receive most of the coverage in the media, there are thousands of small sites that serve the needs and interests of people from all walks of life. Clifford [21] even reports that many older people, who are dying of boredom, have found a "reason to keep going" from the social Web sites.

Almost all social Web sites make it possible for users to join existing groups and start new groups. The existing groups include those that the sites provide and those that members have started. The groups that the sites provide are based on a small number of natural attributes of the members, including schools attended, employers, cities of residence, etc. The groups that members start are based on any type of interest, hobby, and cause [75,111]. Popular general sites host thousands of interest groups. Members can interact with lots of people in the group who can help one another with information, advice, and personal networks.

A 16-year old started a Facebook group named "1,000,000 Strong for Stephen T. Colbert for President of the United States" as, we think, a jest. The group managed to gather 750,000 supporting members in less than a week. (In comparison, Barack Obama's election campaign took 8 months to gather 380,000 supporters [77].) This rather silly case nonetheless demonstrates the enormous potential of popular social Web sites as a vehicle for organizing a large interest group quickly.

Many people have used social Web sites to promote a large number of causes. As we observed earlier, Facebook has become an organizing tool of choice in Arab countries for causes such as human rights, freedom of speech, and democracy. The users have formed or joined numerous groups, such as "Free Egypt," and "No to Arresting Journalists" [33]. Young Egyptians vented on Facebook their anger over Israel's air strikes against Hamas positions in the Gaza Strip [72]. Besides political causes, there are groups for many other causes, such as charity, environment, support for political candidates, etc.

6.1.4. A new source of entertainment

Most people need to entertain themselves to enjoy life, to recharge themselves, and to pass idle time. People use social Web sites as a new source of entertainment. Many people have accounts on several social Web sites, and visit them rather diligently and regularly. They spend time creating and posting UCCs. They update their personal profiles, and write updates on details of their daily lives for their online networks to read and respond to. They view photos and videos on Flickr, YouTube, Hulu, etc. They listen to music on MySpace, YouTube, iLike, etc. They read news articles on Digg. They scan microblogs (called "tweeds") on Twitter. They read updates on Facebook, and watch photos posted by their online networks.

The enormous popularity of social Web sites as a source of entertainment is demonstrated by the speed with which some gifted but obscure people have been catapulted to worldwide fame after videos of their auditions were posted on YouTube. Paul Potts and Susan Boyle are just two such people. The YouTube videos of their performances on the "Britain's Got Talent" television program have amassed over 50 million views. An obscure aspiring singer in the US bolted to fame through her MySpace profile and music postings [74].

6.1.5. A new venue for self-expression

It appears that a surprisingly large number of people have had a strong pent-up desire for self-expression and desire for self-satisfaction that comes from helping others. There are three types of opportunities for self-expression online. First is for people to post UCCs and status updates. Second is to respond to UCCs and updates posted by others, and to participate in discussion boards. Third is to reach out and help others online.

Many people post personal photos (e.g., travels, parties, family gatherings) and detailed updates on their daily lives and thoughts (e.g., where and what they had for lunch, what parties they attended, what movies they watched, when they are going on vacation, etc.). They appear to derive a sense of self-assurance and belonging from such self-expression.

Many people appear to be ready to reach out to others online, believing that those unseen people whom they have never met are good people and potential friends. They take the time to answer questions posed by others online. This mindset may have been a major reason for the success of Wikipedia, the online encyclopedia. The 10 million articles (2.3 million in English) on Wikipedia [96] have been contributed by thousands of volunteers who wanted to help others, without any financial incentives.

6.1.6. Online directories of people

Just as people today turn to an Internet search engine for answers to many types of questions, people are now turning to social Web sites to find anybody who may reside anywhere in the world. This is a natural consequence of the fact that many sites have gathered tens or hundreds of millions of members. Indeed the numbers of members of the popular sites are larger than the populations of many of the countries. As such, the member databases of such sites have become "worldwide directories" of people. Employers search social Web sites for the names of job applicants in order to learn more about them; school administrators search these sites for

the names of admission applicants; and the police search the sites for the names of known sex offenders or criminal suspects.

Some people are now searching social Web sites in order to identify the owners of lost items they found. Stone [85] reports various cases of lost items, such as wallets, mobile phones, digital cameras, etc., being returned to their owners. People who found those items used social Web sites, such as Facebook, to find the owners or owners' friends. A person who found a lost digital camera posted on a social Web site some of the photographs found on the camera. Some of the people who looked at the photographs provided clues to the locations where the photographs were taken, and soon the owner was found.

6.1.7. Miscellaneous

The enormous amounts of social data stored on social Web sites, and the features, and their increasing sophistications, social Web sites provide, have some uses in themselves. The data can, if used properly, yield new knowledge. By analyzing the data, computer science researchers in the data mining field can improve ways to automatically classify UCCs and personal profiles. The results of such research can be used in many ways; for example, screening spam embedded in UCCs, helping online ad networks to more sharply target ads, etc. Social scientists can analyze social data (users, their friends, their groups, communications among them) to understand better, for example, how people's tastes, habits and values affect the formation of social relationships, and vice versa [70].

The features provided by social Web sites, and the sophistication of how they can be used, can help young students and non-technical users develop information technology skills, editing and customizing content, making use of various means of Internet-based communication, thinking about Web page layout, etc. [71]

6.2. Businesses

6.2.1. Marketing and customer relationship management

Social Web sites have become a powerful additional means that businesses and non-profits can use to market their products and services, and manage customer relationships [40,69,111], from the start, has had independent musicians post and market their music on the site. Many businesses now post product release announcements, photos, and videos to such sites as Facebook, MySpace, Twitter, YouTube, Flickr, LinkedIn, etc. They also monitor and respond to both positive and negative comments by members about their products and services. There are said to be 260,000 businesses in the US that use social Web sites to promote business [51].

Businesses can maintain profiles on most social Web sites free of charge. However, they need to make an appropriate level of investment in their social Web site initiatives. They need to have contents prepared professionally. They also need to assign qualified employees to manage their presence on social Web sites to deal with user comments and requests, gauge the tenor of reactions from the users, etc. Negative reactions on social Web sites can spread widely and quickly.

For example, Amazon.com reacted quickly to an uproar on social Web sites. Amazon.com re-classified books on gays and lesbians as "adult" books. This led to an uproar on Twitter. Amazon.com quickly issued an apology and undid the re-classification of those books [54].

Further, marketing efforts on social Web sites, if not done right, can sometimes backfire. One reported case was a user-created advertising video contest on YouTube that a business named Malibu sponsored to promote its liquor product named Banana Rum [90]. There was an apparent discrepancy in the contest rules published by Malibu and the public relations agency that managed the contest for Malibu. Some of the contestants felt that the contest did not follow the rules, and voiced their displeasure loudly on YouTube; one contestant even posted a 6-min conspiracy video on YouTube.

6.2.2. Corporate intranets

Since social Web sites provide a rich set of features for online networking and sharing UCCs, small businesses are now using social Web sites essentially as company intranets. In other words, many small businesses conduct everyday operations on social Web sites. The management posts announcements to employees; and encourages employees to share work-related documents, post PPT presentations, and exchange messages [99]).

To use social Web sites as corporate intranets, the businesses usually need "closed" social networking Web sites. There are two ways in which businesses may operate closed social Web sites. One is to create a closed enterprise social network on an open social network site. LinkedIn provides "company groups," and Facebook has "Facebook for Business" for this purpose. LinkedIn services for enterprises are not free. Another way is to create a closed enterprise social network by using enterprise social networking software, such as IBM Lotus Connections, Microsoft Sharepoint, NewsGator (Sharepoint add-ons), Jive Software, SelectMinds, Contact Software, Leverage Software, etc. [41,26].

6.2.3. Peer networking

There are two broad classes of peer networks. One is an internal network, that is, a network of peers within a business. Another is an external network, that is, a network of peers from different organizations (other businesses, universities, government, etc.).

Businesses can sometimes benefit from the external networks their employees maintain on open social Web sites. The employees can use the networks to recruit employees, and obtain advice and information. There is a downside to this, though. The employees may give away trade secrets to their external peers, provide unfavorable portrayals of their employers, and/or may be recruited by their peers, too [41].

6.2.4. Miscellaneous

A user's profile, network of friends, groups joined, UCCs posted, etc., can add to what the user shows on a resume

or may have shown on the job. Increasingly, employers and school administrators, etc., make use of such additional information in hiring, firing, and admission decisions. School administrators sometimes search prospective students on social Web sites. If they discover inappropriate postings by the prospective students, they often proceed to deny admission. Employers do similar things with job applicants and existing employees. They may decide not to hire the applicants, or take disciplinary measures or even fire the employees, if the employees, even in off hours, are found to be in violation of the employer's code of conduct and confidentiality rules. Over 20% of hiring managers chose not to hire job applicants based on the information found on social Web sites, such as drinking, using drugs, negative comments about a former employer, lies about qualification, etc. [91,22,56]. Hoover [41] cites a Forrester Research report that 14% of companies disciplined employees and 5% fired them for offenses involving social networking sites.

6.3. Government

6.3.1. Governance

Just as many businesses do, some government agencies, such as US Federal Emergency Management Administration (FEMA) and Environment Protection Agency (EPA), have presence on social Web sites to easily reach a large number of people. They can both disseminate information, and receive feedback on it. For example, the EPA has set up a group on Facebook, and allows people to make use of all social networking facilities provided.

6.3.2. Law enforcement

Just as employers and school administrators, law enforcement agencies have started using social Web sites in two different ways. One is simply to search the profiles, updates, and UCCs on social Web sites to gather incriminating evidence on suspects. One gang member in Los Angeles was arrested in part because of an update he posted on MySpace in which he boasted that he had shot someone.

Another way is to post surveillance photos and other information, and seek help from users of social Web sites. The police in Maine, US, posted on Facebook images from a surveillance video that showed three teenagers vandalizing a hotel spa, and were able to arrest the teenagers based on tips received from Facebook users [82].

6.3.3. Election campaign

The use of the Internet, including social Web sites, has become a key part of election campaign strategy in many countries. The presidential election in 2003 in South Korea was decided by Internet-savvy liberal activists who used Internet portal cafes (groups) to urge young voters in their twenties to go to the polls. As a result, a then-dark horse candidate was elected President. Howard Dean's and Barack Obama's election campaigns for US President in 2004 and 2008, respectively, made effective use of blogs and social Web sites to convey messages to the public, news media, volunteers, and donors; organize volunteers; manage news media; appeal to donors, etc. [[13,95,57]. Social Web sites are a two-edged sword with respect to election campaigns is a two-edged sword, in that supporters of a candidate can try to sway the positioning of the candidate [68].

6.3.4. Legal proceedings

The US Supreme Court has recently admitted a YouTube video link as a citation in a petition filed with the court [49]. In the future, UCCs discovered on social Web sites will be used in legal proceedings [46].

Earlier, in a different case, the Supreme Court viewed a video showing a high-speed police chase of a car that ended in the police car ramming the car and leaving the driver paralyzed. The federal appeals court ruled for the driver. However, the Supreme Court ruled for the police, rather than sending the case back to the appeals court, as they often do. As the Supreme Court justices see video evidence themselves, they may become less inclined to defer to the findings of jurors and conclusions of lower-court judges [82].

7. Issues and challenges

Social Web sites as we know them today have only a 5–6 year history. The growth of some of the sites, in terms of the number of users, the level of daily traffic, and the amount of UCCs stored, has been absolutely incredible. The general sense is that even the founders of those sites did not envision such growth. As social Web sites have evolved, various unforeseen issues and challenges have surfaced. We can group them into two types. As social Web sites have been new experiences to both the users and the sites, we may group the issues into two types: user misuse issues and site issues. User misuse issues are those that misuses of social Web sites have brought about to the users and the society, while site issues are those that the site operators have had to cope with as viable businesses in order to keep up with the growth of the sites. Table 8 summarizes both types of issues and challenges. In this section, we discuss each of the issues.

7.1. User misuse issues

7.1.1. Social issues

The misuse of social Web sites has brought about various forms of damage to people. In terms of whom the damage is inflicted, we can group the social issues into two types: damage to self, and damage to others. Below we discuss various manifestations of social issues. We expect that there are other manifestations, as people seem to be rather creative in finding ways to misuse social Web sites.

7.1.2. Damage to self

Many users do not seem to suspect that their behaviors on social Web sites may get them into trouble. They appear to believe that others who may read and view their postings, their interactions with online "friends," the groups they join, etc., will only take them benignly

Table 8
Issues and challenges of social Web sites.

Type	Issues
User misuse issues	Social issues Security issues Legal issues Privacy issues
Site issues	Profitability issues Online advertising issues Operation issues Legal issues

[37,62,67]. Because of indiscrete postings and membership in quirky groups, applicants to colleges have been denied admission, job seekers have been denied employment opportunities, criminals have been arrested, court cases have been lost, etc. [56,41,91,22]. The Ref. [22] provides some common-sense guidelines for job seekers on what to do and what not to do on social Web sites, such as "update the profile regularly," "not badmouth current or past employer," "remember others can see your friends."

Many users show signs of addiction to the Internet in general and social Web sites in particular. Many users spend what seems like inordinate amounts of time on social Web sites. Many users have joined multiple social Web sites and spend hours on them every day. There are such enormous amounts of contents on popular social Web sites, and enormous amounts of new contents are being added every day that just one major social Web site can keep a person fully occupied for the rest of the person's life. For example, more than six billion videos were viewed on YouTube in January 2009, and it is estimated that 15 h of new videos are posted to the site every minute [108]. Flickr has more than 3 billion photos [31].

Many users suffer from reduced productivity by spending time on social Web sites while at work. Many corporations in the US, including Citigroup, Goldman Sachs, JPMorgan, UBS, etc., restrict access to Facebook. The city of Toronto blocked their staff's access to social Web sites [41].

The ultimate damage to self that the users of social Web sites is the ending of life. There are many suicide Web sites in South Korea, and some members of such sites have committed suicide together during the past several years. One suicide site operator even sold cyanide to members. Now volunteer counselors are working to discourage suicide pacts.

7.1.3. Damage to others

Many users callously do irresponsible and harmful things to others, apparently feeling comfortable hiding behind fictional online identities. There are no means of validating the accuracy or veracity of personal profiles on most social Web sites. Most social Web sites only require email addresses for confirming the "real-ness" of members. South Korea now requires all Web sites with more than 100,000 average daily visitors to verify the true identity of members [35].

Many users spread false rumors. The police in South Korea reported 10,000 libels in 2007. In early 2008, there were prolonged, massive, and hysterical protests against the decision by the newly installed government to re-allow imports of beef from the US. The protests substantially gained in size and intensity by the false or exaggerated stories spread through Internet discussion boards and social Web sites. The stories claimed that beef from the US, and any byproducts (including cosmetics and feminine menstrual protections), guaranteed transmission of mad-cow disease. The government had to shake up the cabinet to calm the public anger. Now beef from the US is enjoying brisk sales in South Korea.

Many users instigate or participate in cyber bullying and cyber stalking [50,104]. These have led to occasional high-profile suicides, such as a MySpace teenager in the US [76], and a South Korean actress [19]. Many users also exhibit meanness and callousness towards those who may need help. Many users viewed and encouraged a 19-year-old student in Florida to swallow antidepressant pills to commit suicide, a man in Arizona overdosing on drugs while writing about his actions in a chat room, and a man in the UK hanging himself while chatting online and webcasting [78].

Many users spread false information. Apple Computers' stock fell 5% after a citizen-journalism site published a false report that Steve Jobs, the company's legendary CEO, had been rushed to the emergency room. The false report was amplified by its promotion on Digg, the news bookmark site. Other false news stories have been promoted on Digg [23]. To address this problem, online news media try to vet their news sources, and ratings systems may be created to evaluate accuracy and trustworthiness of news articles. Such efforts, however, cannot cope with the speed and breadth of the spread of false information on the Internet.

Many users post problematic materials, including pornographic materials or politically sensitive materials that are expressly prohibited by certain governments or materials that violate privacy or data protection laws of certain countries. United Arab Emirates has banned Orkut, Flickr, Hi5, MySpace and YouTube for containing sexually explicit materials [66]. Four Google executives are now on trial in Italy on criminal charges of defamation and privacy violation. Someone had posted on Google's Italian site a 3-min video showing four youths in Turin bullying a boy with Down syndrome. Google quickly removed the video from the site, and the four executives were not involved directly in handling the video. However, Italian prosecutors charged that the video should not have been published in the first place [38]. Moldova's government cut off Twitter service in its capital after more than 10,000 young Moldovans, many of whom communicated via Twitter, staged violent protests against the government [15].

7.1.4. Security issues

There are two types of security issues for social Web sites. One is the security of people. Another is the security of the computers people use and data they store in computer systems.

The presence of sex predators on social Web sites poses a real problem for the physical and mental security of teens. Under subpoena from the states of Connecticut and North Carolina in the US, MySpace recently removed 90,000 members who are convicted sex offenders using their real names [106]. Making use of services of a firm with identity and age verification technology, MySpace was able to sift these members. However, it is anybody's guess as to how many pedophiles and sex offenders who are not using real names, or who have not been convicted, are members of MySpace and other social Web sites popular among teens. The states of New Jersey, Nevada and Florida in the US have recently banned some sex offenders from using social Web sites [12]. Those barred are to submit to periodic examination of their computers, and install equipment on their computers so their use can be monitored. "Social Networking sites: safety tips for tweens and teens," FTC (US Federal Trade commission) Facts for consumers [34], Hargis [39], Social Networking Websites [60] and Social Networking [25] offer common-sense tips for parents to use to protect their kids on social Web sites.

Since social Web sites have enormous numbers of users and store enormous amounts of data, they are natural targets of spammers, and phishing and malware attacks. Unsuspecting users often download viruses; for example, the "You've got a new friend" notice carried malware. Users infect their computers at work with viruses they download from social Web sites. Users may compromise their employer's business secrets; for example, they may reveal passwords to phishing attacks. Recently, fake updates were made to 33 Twitter member accounts, including those of Barack Obama, Britney Spears, and a CNN reporter [105]. There is even a new type of spam on Twitter called "follow spam [1]." A user "follows" a huge number of people for the purpose of having them view the user's UCCs or click on URLs included in the UCCs. Many social Web sites use the Completely Automated Public Turing Test to tell Computers and Humans Apart (CAPTCHA) code to prevent spamming programs from sending mass emails or obtain accounts on the sites. However, some spammers have broken the CAPTCHA code and obtained bogus accounts. Because of such dangers, besides the loss of productivity mentioned earlier, many companies restrict access to social Web sites [62].

7.1.5. Legal issues

Misuse of social Web sites has led to some legal issues. In due course, these will be resolved by the courts, legislatures, sites themselves, employers, and the users. We review some of them below.

The issue of truthfulness of personal profiles has become a subject of debate. Many users provide incorrect data in their profiles, such as fake names, ages, schools attended, employment information, hobbies, interests, etc. This is partly motivated by privacy needs, and is even recommended by the US Federal Trade Commission as Internet safety tips for parents and kids [34]. However, recently a US federal court convicted a woman who created a phony profile on MySpace of misdemeanor charges of computer fraud [76]. The woman, mother of a teen age girl masqueraded as a teenage boy and caused her daughter's classmate to commit suicide, with a bullying message. MySpace's terms of service require users to submit "truthful and accurate" registration information, and the prosecutors argued that a phony profile amounts to "unauthorized access" to the site and violated the Computer Fraud and Abuse Act of 1986 [79]. Although the woman's behavior was reprehensible, criminalizing untrustworthy profiles is troublesome, especially in view of legitimate privacy needs and also the fact that most users do not even read or care about terms of use of social Web sites [79].

Of course, spammers, those who launch phishing and malware attacks, and those who post pornographic materials involving minors to social Web sites are all violating laws, and are subject to prosecution.

When police in Maine, US, posted on Facebook images from a surveillance video that showed three teenagers vandalizing a hotel spa, as we discussed earlier, they had to blur the faces of the teenagers. The reason was that the Maine state law prohibited identifying minors as crime suspects [82].

7.1.6. Privacy issues

Many social Web sites provide "privacy" settings, so that users may control who may view what parts of their profiles. However, a majority of the users have little anxiety about the "digital print" they leave on social Web sites [81,93]. On most social Web sites, a "friend" is simply any user who communicates a wish to be one and receives consent. Many users reveal all sorts of details about their activities, personal networks, and thoughts to these "friends" and, often, to everyone who may visit social Web sites.

7.2. Site issues

7.2.1. Profitability issues

During the past few years, many articles have been written questioning whether social Web sites can be viable businesses [61,94,98]. However, there are many profitable social Web sites, including LinkedIn, MySpace, Bebo, Cyworld (South Korea), QQ (China), Mixi (Japan), MobileGameTown (Japan), Xing (Europe), etc. [94]. LinkedIn charges fees for job postings and for hosting closed social networks for businesses, expert search services, and banner ads. Most social Web sites generate revenue by selling online ads, virtual gifts (e.g., virtual birthday cakes, popping corks of champagne, etc.); ringtones, artist merchandise, concert tickets (MySpace), etc. [94].

There have also been some mega business deals involving some of the most popular social Web sites. However, some of those deals have had disappointing results so far. In 2006, Google signed a 3-year $900 million agreement with MySpace to sell ads on MySpace. It was apparently not a good investment for Google [94]. In 2007, Microsoft bought 1.6% of the shares of Facebook for $240 million; however, in 2008, Facebook may have lost $150 million, with revenue of $50 million and expenses of $200

million [98]. In 2006 (Recently, a Russian investment firm, Digital Sky Technologies, invested $200 million to buy 1.96% of the shares of Facebook [55].), Google acquired YouTube for $1.6 billion, but YouTube's revenue has been small. One report estimates that in 2009 YouTube will incur operating expenses of $711 million, and bring in $240 million in online ad revenue [109], while another report expects the loss to be at $174 million [20].

Such social Web sites as Facebook, YouTube and Twitter are trying to rapidly increase traffic to their sites and/or number of members. One major difficulty this strategy poses is that the major increase has come from the parts of the world, such as some Asian countries, the Middle East, Latin America, Africa, and Eastern Europe, where Internet bandwidth is expensive, ad rates are very low, and people are even less inclined to click on ads than those in the US and Western Europe. To deliver the UCCs to people in those parts of the world, the sites have had to shoulder the huge costs of extra servers to overcome the limited and expensive Internet bandwidth. Some of the major social Web sites may end up shutting off services in certain non-profitable parts of the world, or use lower bandwidth to deliver the UCCs [88].

Further, many social Web sites need to modify their business strategy over time, as the demographics of the members evolve, and unforeseen situations present themselves. For example, Yelp, the customer review site, started out accepting reviews from customers of local businesses, and did not allow businesses to respond to negative reviews. Over time, Yelp has learned that some reviews are not fair, and that they needed online advertising revenue from local businesses. So, Yelp now allows businesses to respond to customer reviews [53]. Such sites as Friendster and Orkut started in the US, but have not succeeded in the US. They have become highly successful in South Asia, and Brazil, respectively, and changed business strategy to cater to the needs and preferences of the members in those regions.

7.2.2. Online advertising issues

Today, various popular social Web sites, including Facebook, YouTube, Twitter, have not been able to generate online advertising revenues commensurate with the enormous traffic and enormous numbers of members they have generated.

Urstadt [98] provides an excellent discussion of the inherent difficulties of targeted advertising on social Web sites. First, users of social Web sites are not looking for information to buy things; they are looking to connect and view UCCs posted by their online networks. As a result, while 2 percent of Google search engine users click on ads, only 0.04 percent of Facebook (and presumably other social Web sites) click on ads. Second, only 3% of Internet users are willing to have their friends become targets of online ads [92]. This is one of the reasons for the failure of Facebook's Beacon advertising program (to be discussed shortly). Third, many businesses do not want their ads placed on such quirky social groups as "I've had sex with someone on Facebook," "Nikki AKA Death Angel," etc. Fourth, as remarked earlier, many user profiles (some say as much as one-third) contain inaccurate data, and they in turn make targeted advertising inaccurate. As a result, social Web sites have great difficulty generating revenue from online ads. Facebook charges only 14 cents per thousand times an ad is served, while MySpace charges $2 for a banner ad per thousand times it is served.

When Facebook launched an advertising program called Beacon [98], users' privacy concerns were raised. Working with commercial Web sites like Blockbuster (video rental) and eBay (auction), Beacon tracked all Facebook users' purchases and displayed them to all of their Facebook "friends." After online petitions from the users and negative press, Facebook scaled back the program.

7.2.3. Operation issues

As popular social Web sites acquired enormous numbers of members and enormous amounts of UCCs in rather short periods of time, various business operation issues have challenged them, and led to negative press and even lawsuits.

One of the business operation issues is the IT infrastructure issue. As Facebook and Twitter have experienced the fastest growth, most of the press coverage of the IT infrastructure issues is, fairly or unfairly, about them. Their data centers sometimes were not able to keep up with the growth, and as a result experienced occasional failures and slow downs.

Facebook and MySpace have had difficulties obliterating departing members' data [10]. Although some people suspect that the sites deliberately did not purge all traces of departing members, we tend to think that their software and databases have become tangled up due to fast-paced changes and extensions, and made it difficult for their engineering staff to untangle cleanly.

Facebook also apparently tried to change its membership contract to claim perpetual ownership of all data posted by the members. After loud protests from angry users, Facebook claimed that it was a misunderstanding and restored an earlier version of the membership contract [89].

Facebook was sued by someone who received text messages that Facebook sends to members' friends. The person was not a friend to the Facebook member, but happened to be assigned a phone number from a mobile phone company after another customer's service ended [11]. The content of the text message was upsetting to the person, and, to add insult to injury, the person was charged 10 cents per text message, a share of which Facebook received. Facebook agreed to make it easier for recipients of text messages to block future messages from the site, and work with mobile phone companies to check the lists of recycled phone numbers.

After Facebook released its Platform, many developers responded, unfortunately, with many useless applications. Facebook users deleted such applications. The developers then developed additional useless applications to replace the ones the users had deleted. As the applications sent out "invite" notices to Facebook users, many of these third-party applications became spammers [43]. Facebook introduced limits on the number of invites an application may send out, and modified its metrics listing most

popular applications to measure actual use rather than just the install count.

7.2.4. Legal issues

People who do not abide by laws create legal problems not only for themselves but also the sites they use. Social Web sites have been used as platforms for fomenting or organizing antigovernment dissents in various countries, such as South Korea, France, Egypt, China, etc. The governments' responses have ranged from bringing civil and criminal charges to the leaders involved to blocking the sites.

As remarked earlier, many users post copyrighted materials without authorization, pornographic materials, materials that may violate privacy laws, etc. The sheer volume of the UCCs being posted every day to popular social Web sites makes it practically impossible for the site operators to remove such materials before they are viewed and spread on the Internet. This in turn often leads to legal problems for the sites and some of the users with copyright holders or governments.

8. Trends and prognosis

On the basis of the recent trends reported in press articles, we feel that we can fairly safely predict how social Web sites will evolve. Below we summarize our predictions.

8.1. Support for mobile devices and internet phones

Popular social Web sites, such as Twitter, Facebook Mobile, YouTube, and Orkut, have taken the lead in allowing the users to receive friends updates and UCCs on mobile phones, and post updates and UCCs from mobile phones. Other sites will follow suit. The ability to keep in contact with online connections is likely to be extended to other mobile devices that people may wear and carry.

Further, MySpace and Facebook include Skype the Internet phone service as one of the applications accessible to their users. These and other sites will include access to Internet phone services, such as Ribbit, Gizmo, Aim, Jahjah, Jaxter, etc., to expand the repertoire of communication facilities available to their members.

8.2. Location-based and data-mining-based services

Major sites are likely to augment their support for mobile devices with location-based services. Location-based services will include delivering information of interest to the users based on their present geographical locations. The information may include a list of friends in the vicinity, and notices to and from them. It may also include advertisements, coupons, directions, etc., for stores, restaurants, movies, concerts, etc., that may be of interest to the users.

Today, major sites already use data mining techniques to target advertisements to users with matching profiles (although, as remarked earlier, the users have shown remarkable disinterest in looking at ads on social Web sites). Major sites also have fairly simple friend-recommendation engine to present a list of members whom a member may be interested in including in his friends' list. Major sites are likely to make use of data mining techniques to enhance accuracy of such services, and to provide new services based on the results of data mining.

8.3. Support for a broadcast feed to multiple social web sites

As many users (and their friends) have joined multiple social Web sites, there is a need for a user to funnel all social Web site activities into a single broadcast for all the sites they belong to, and also selectively have their friends receive those updates. Such social broadcast sites are appearing, including FriendFeed, Iminta, Plaxo, Readr, Mugshot, etc. [80]. However, we think the social broadcast facility is highly likely to become a feature of Facebook Connect and Google Friend Connect.

8.4. Social features added to non-social web sites

The Google Open Social makes it almost trivial for any site to instantly become a simple social Web site. However, even without the Google Open Social, many sites are likely to add social features in order to grow their user bases. For example, there are a number of social e-learning sites that create online study groups [32]. These sites include Cramster, Course Hero, and Koofers.

8.4.1. Merging of social networking sites and social media sites

As remarked earlier, the distinction between social networking sites and social media sharing sites is fast becoming blurred. Social networking sites will increasingly add UCC-sharing features, and social media sites will increasingly add social networking features. The Google Open Social will make it easier for social media sites to add social networking features.

8.4.2. Emergence of enterprise social web sites

Businesses can take corporate knowledge management to the next level by leveraging social features found in social Web sites, such as personal profiles, communication among employees, updates, sharing of UCCs (in particular, blogs and wikis), to their corporate knowledge management. Such social features can be used to create a continuously updated centralized knowledgebase, which in turn can allow employees to search for an internal network of knowledgeable colleagues and form internal communities based on special subjects, etc. [26].

There are two ways businesses can leverage social features. One is to make use of the closed "company" or "business" groups on general-purpose open social Web sites, such as LinkedIn and Facebook. Another is the adoption of closed enterprise social Web sites provided by a number of emerging business social software vendors. The vendors include IBM (Lotus Connections), Microsoft (SharePoint), Contact Networks, Leverage Software, SelectMinds, Jive Software, etc. [41,26]. Businesses may use

one or both of these options. The use of closed company groups on major open social Web sites offers two advantages: low cost of ownership, and a very wide audience (assuming that the groups are only partially closed). Its disadvantages include relative lack of facilities for oversight and control (for security and governance) and analytics (usage monitoring and measurements). Further, it makes integration with other enterprise software problematic, such as customer-relationship management system, enterprise resource planning system, etc. [41]. It can probably serve the needs of small to medium businesses. However, large enterprises are likely to need to adopt closed enterprise social Web sites.

8.4.3. Implementation and architecture

As remarked earlier, most major sites have been developed by a surprisingly small group of engineers who were caught totally off guard by the explosive growth in the level of traffic, the amount of data to manage, and the number of users to serve. As the engineers struggled mightily to keep the sites running, they had to make changes to some of the open source software they had selected, and incorporated various clever software and systems tricks. In trying to keep up with the incredible performance and scalability demands stressing their sites, the engineers have uncovered some serious performance bottlenecks in existing open source software, such as the MySQL relational database system and Apache Web server; and fundamental problems with such database techniques as concurrency control and replication of data [112,3,4,9]. Fundamental solutions to such problems should be developed and incorporated into these open source software.

Further, some, if not all, sites may benefit from a thorough re-architecting and re-implementation starting with the understanding of the huge traffic, huge number of users, and huge amounts of data of various types. In particular, the tangled and difficult-to-manage code in some of the sites needs to be cleaned up, and opportunities for finer parallel processing of database access requests and other service requests should be examined, and solutions should be incorporated into the architecture of the sites.

All major sites depend heavily on relational database systems. All relational database systems store data on hard disk drives, and all major database management algorithms, such as query optimization and query processing, access methods (indexing, hashing, sequential search), update logging, locking, etc., have been designed in consideration of the performance characteristics of the hard disk drives (e.g., seek time, rotational latency, and data transfer time). During the past several years, flash memory has significantly gone up in capacity and gone down in price, and many people now expect that at some point in the near future, the flash memory can replace the hard disk drives in many situations. When that happens, all major algorithms used in relational database systems have to be significantly modified, since the performance characteristics of the flash memory are very different from those of the hard disk drives [45]. This will in turn force changes in key architecture decisions made in major social Web sites.

8.5. Social web sites as businesses

Although various social Web sites claim to be profitable, it is not clear how they do accounting and whether their claims are truthful. For most sites, online advertising has not become the primary source of revenue. Many, including such popular sites as Twitter, manage to operate almost exclusively on venture capital financing. Many sites have shut down and many more will cease operation.

Some popular social Web sites have been acquired by large Internet or media companies. These include MySpace (acquired by News Corp in 2005 for $650 million), Friends Reunited (acquired by UK ITV in 2005 for $208 million), Flickr (acquired by Yahoo in 2005 for an undisclosed but apparently a rather small amount), YouTube (acquired by Google in 2006 for $1.65 billion), Bebo (acquired by AOL in 2008 for $850 million), Plaxo (acquired by Comcast in 2008 for an undisclosed amount), etc. There have been rumors that Facebook, Twitter, and other popular social Web sites may be acquired by Internet or media companies whose business can benefit from a large number of Internet users.

When popular social Web sites do not discover a way to generate substantial revenue from online targeted advertising, they are likely to charge subscription fees to businesses for the use of closed company and business groups (following the lead by LinkedIn), and even to the general public. Charging subscriptions to the general public is likely to chase away a lot of the members; however, it can help solve some of the problems that social Web sites have today.

9. Concluding remarks

Social Web sites have rapidly become an integral part of daily lives of hundreds of millions of Internet users all over the world. This is nothing short of a major phenomenon. Since there are very few academic papers on social Web sites, and many of the press articles, Wikipedia articles, and blogs deal with limited subjects related to social Web sites, and many of the press articles and blogs are difficult to comprehend or outdated or inaccurate, we tried to provide a comprehensive paper that deals with the status, uses, issues and the future of social Web sites somewhat rigorously. Although social Web sites are evolving rapidly, we feel that the taxonomies we provide in the paper will remain valid for the near future. We believe that the taxonomies will serve as a basis for understanding, discussing, adopting, and developing social Web sites.

Acknowledgments

Several graduate students at Sungkyunkwan University did an initial feature analysis of various representative social Web sites. They are Seungwoo Choo, Hyeong-Joon

Kim, Gang-Nyeon Kim, Sung-Up Moon, Byung-Woo Nam, Eu-Teum Park, and Ki-Tae Yoon.

We thank the Editor-in-Chief for suggesting the inclusion of discussions of architectures of the social Web sites. His suggestions led to our adding Section 5, and expanding Section 8.

This research was supported by the Ministry of Knowledge Economy (MKE), Korea, under the Information Technology Research Center (ITRC) support program supervised by the National IT Industry Promotion Agency (NIPA) (NIPA-2008-(C1090-0801-0046)).

References

[1] The dangers of spam on social networking sites, 〈http://www/allspammedup.com/2008/12/the-dangers-of-spam-on-social-networking-sites/〉.
[3] 〈http://perspectives.mvdirona.com/default,date,2008-07-16.aspx〉.
[4] 〈http://www.scribd.com/doc/35222/Flickr-Architecture-Presentation〉.
[5] 〈http://hurvitz.org/blog/2008/06/linkedin-architecture〉.
[7] 〈http://perspectives.mvdirona.com/2008/12/27/MySpaceArchitectureAndNet.aspx〉.
[8] 〈http://poorbuthappy.com/ease/archives/2007/04/29/3616/the-top-10-presentation-on-scaling-websites-twitter-flickr-bloglines-vox-and-more〉.
[9] 〈http://video.google.com/videoplay?docid=-6304964351441328559〉.
[10] M. Aspan, After stumbling, Facebook finds a working eraser, The New York Times, February 18, 2008.
[11] The Associated Press, Facebook Settles Text-messaging Suit, The New York Times, December 18, 2007.
[12] The Associated Press, Sex Offenders Are Barred From Internet by New Jersey, The New York Times, December 28, 2007.
[13] M. Bai, The Web users' campaign, The New York Times, December 9, 2007.
[14] C. Baldwin, Forget the Classics, Try a Facebook College Course, InformationWeek, September 28, 2007.
[15] E. Barry, Protests in Moldova explode, with help of Twitter, The New York Times, April 8, 2009.
[16] D. Boyd, N. Ellison, Social network sites: definition, history, and scholarship, Journal of Computer-Mediated Communication 13 (1) (2007) article 11.
[17] D. Chartier, Bebo, Meebo team up with Facebook against Google, 〈http://arstechnica.com/software/news/2007/12/bebo-meebo-team-up-with-facebook-against-google.ars〉.
[18] D. Chartier, Google friend connect launches, eyes Facebook connect, 〈http://arstechnica.com/old/content/2008/12/google-friend-connect-launches-eyes-facebook-connect.ars〉.
[19] Sang-Hun Choe, Korean star's suicide reignites debate on Web regulation, The New York Times, October 13, 2008.
[20] T. Claburn, Is YouTube close to profitability? The New York Times, June 17, 2009.
[21] S. Clifford, Online, 'a reason to keep on going', The New York Times, June 2, 2009.
[22] 〈http://edition.cnn.com/2008/LIVING/wprklife/11/05/cb.social.networking〉.
[23] N. Cohen, Spinning a Web of lies at digital speed, The New York Times, October 13, 2008.
[24] N. Cohen, When stars Twitter, a ghost may be lurking, The New York Times, March 27, 2009.
[25] Social networking, 〈http://www.commonsense.com/internet-safety-guide/social-networking.php〉.
[26] A. Conry-Murray, Can enterprise social networking pay off? March 21, 2009, 〈http://internetevolution.com/dpcument.asp?doc_id=173854〉.
[27] Social marketing goldmine: top 100 social networking sites, 〈http://www.dulcenegosyante.com/social-marketing-goldmine-top-100-social-networking-sites〉.
[28] 30 Largest social bookmarking sites, 〈http://www.ebizmba.com/articles/social30〉.
[29] 〈http://experientia.com/blog/health-related-social-networking-sites/〉.
[30] Social networking sites search engine, 〈http://findasocialnetwork.com/search.php〉.
[31] 〈http://en.wikipedia.org/wiki/Flickr〉.
[32] L. Foderaro, Psst! Need the answers to no. 7? Click here, The New York Times, May 18, 2009.
[33] The list: the world's top social networking sites, 〈http://www.foreignpolicy.com/story/cms.php?story_id=4134〉.
[34] Social networking sites: safety tips for tweens and teens, FTC (US Federal Trade Commission) Facts for Consumers, 〈http://www.ftc.gov/bcp/edu/pubs/consumer/tech/tec14.shtm〉.
[35] Google scales back YouTube Korea, InformationWeek, April 13, 2009.
[36] R. Good, Open social: Google's new social networking platform—what is it and why IT matters, 〈http://www.masternewmedia.org/social_networking/social-networking-platforms/Open-Social-Google-social-networking-platform-what-is-it-20071102.htm〉.
[37] S. Hansell, One subpoena is all it takes to reveal your online life, The New York Times, July 7, 2007.
[38] S. Hansell, Google executives face jail time for Italian video, The New York Times, February 2, 2009.
[39] M. Hargis, Social networking sites dos and don'ts, 〈http://www.edition.cnn.com/2008/LIVING/worklife/11/05/cb.social.networking〉.
[40] M. Higgins, Twitter comes to the rescue, The New York Times, July 5, 2009.
[41] J.N. Hoover, Social networking: a time waster or the next big thing in collaboration? InformationWeek, September 22, 2007.
[42] O. Kharif, Social networking sites open up, Business Week, Technology, February 13, 2007.
[43] J. Kincaid, The Facebook platform, one year later, 〈http://www.techcrunch.com/2008/05/24/facebook-platform-one-year-later/〉.
[44] A. Lagourney. The '08 campaign: sea change for politics as we know it, The New York Times, November 4, 2008.
[45] Sang-Won Lee, Bongki Moon, Chanik Park, Jae-Myung Kim, Sang-Woo Kim, A case for flash memory SSD in enterprise database applications, in: Proceedings of the SIGMOD International Conference, 2008, pp. 1075–1086.
[46] R. Levine, S. Swatski-Lebson, Are social networking sites discoverable? Legal Technology, November 13, 2008. 〈http://www.law.com/jsp/legaltechnology/pubArticleLT.jsp?id=1202425974937〉.
[47] J. Lewin, The top 20 social networking sites, 〈http://www.podcastingnews.com/2009/01/01/the-top-20-social-networking-sites/〉.
[48] C. Li, Facebook connect—another step to open social networks, 〈http://blogs.forrester.com/groundswell/2008/05/facebook-connec.html〉.
[49] A. Liptak, Supreme Court enters the YouTube era, The New York Times, March 3, 2009.
[50] C. Maag, When the bullies turned faceless, The New York Times, December 16, 2007.
[51] M. Meece, Small businesses are taking tentative steps toward online networking, The New York Times, June 4, 2009.
[52] J. Metcalfe, The celebrity Twitter ecosystem, The New York Times, March 29, 2009.
[53] C. Miller, The review site Yelp draws some outcries of its own, The New York Times, March 3, 2009.
[54] C. Miller, Putting Twitter's world to use, The New York Times, April 14, 2009.
[55] C. Miller, Russian firm buys a stake in Facebook, The New York Times, May 27, 2009.
[56] One in 10 employers will use social networking sites to review job candidate information, 〈http://naceweb.org/press/display.asp?year=&prid=244〉.
[57] A. Nagourney, The '08 campaign: sea change for politics as we know it, The New York Times, November 4, 2008.
[58] E. Naone, Who owns your friends, Technology Review 111 (4) (2008) 44–48.
[59] E. Naone, A brief history of microblogging, Technology Review 111 (5) (2008) 26.
[60] Social networking websites, 〈http://www.ncpc.org/topics/by-audience/parents/social-networking-websites〉.
[61] J. Owyang, The many challenges of social networking sites, 〈http://www.web-strategist.com/blog/2008/02/11/the-many-challenges-of-social-networks/〉.
[62] Social networking sites concern cyber-security experts, 〈http://physorg.com/news148139711.html〉.
[63] 5 facts about social networking sites, 〈http://pipl.com/statistics.social-networks/5-facts/〉.
[64] D. Pogue, Twitter? It's what you make it, The New York Times, February 12, 2009.
[65] V. Prelovac, Top list of social media sites, 〈http://www.prelovac.com/vladimir/top-list-of-social-media-sites〉.
[66] S. Rahimi, D. Bowman, Social network site banned in UAE, 〈http://www.arabianbusiness.com/index.php?option=com_contents&view=article&id=495784:uae-bans-orkut-com&itemid=72〉.
[67] J. Randerson, Social networking sites don't deepen friendships, 〈http://www.guardian.co.uk/science/2007/sep/10/socialnetwork〉.

[68] J. Risen, Obama voters protest his switch on telecom immunity, The New York Times, July 2, 2008.
[69] J. Rosen, Be it Twittering or blogging, it's all about marketing, The New York Times, March 12, 2009.
[70] S. Rosenbloom, On Facebook, scholars link up with data, The New York Times, December 17, 2007.
[71] Educational benefits of social networking sites uncovered, 〈http://www.sciencedaily.com/release/2008/06/080620133907.htm〉.
[72] S. Shapiro, Revolution, Facebook-style, The New York Times, January 25, 2009.
[73] M. Shore, A friend connected web," 〈http://googleblog.blogspot.com/2008/05/friend-connected-web.html〉.
[74] D. Sipher, Singing her way from obscurity to fame on the Internet, The New York Times, December 18, 2007.
[75] R. Silverman, A new generation reinvents philanthropy, The Wall Street Journal, August 7, 2007.
[76] J. Steinhauer, Verdict in MySpace suicide case, The New York Times, November 27, 2008.
[77] B. Stelter, The Colbert nation quickly colonizes Facebook, The New York Times, October 29, 2007.
[78] B. Stelter, Web suicide viewed live and reaction spur a debate, The New York Times, November 25, 2008.
[79] B. Stelter, Guilt verdict in cyberbullying case provokes many questions over online identity, The New York Times, November 28, 2008.
[80] B. Stone, Friends may be the best guide through the noise, The New York Times, May 4, 2008.
[81] B. Stone, Our paradoxical attitudes toward privacy, The New York Times, July 2, 2008.
[82] B. Stone, Police take the fight onto the Web, The New York Times, March 8, 2009.
[83] B. Stone, Is Facebook growing up too fast, The New York Times, March 29, 2009.
[84] B. Stone, Is this the future of the digital book? The New York Times, April 5, 2009.
[85] B. Stone, Lost in the real world, found via cyberspace, The New York Times, April 8, 2009.
[86] B. Stone, Tinker away, Facebook says, The New York Times, May 3, 2009.
[87] B. Stone, Breakfast can wait. The day's first stop is online, The New York Times, August 10, 2009.
[88] B. Stone, M. Helft, In developing countries, Web grows without profit, The New York Times, April 27, 2009.
[89] B. Stone, B. Stelter, Facebook withdraws changes in data use, The New York Times, February 19, 2009.
[90] L. Story, Outcome of an ad contest starts an uproar on YouTube, The New York Times, June 27, 2007.
[91] R. Stross, How to lose your job on your own time, The New York Times, December 30, 2007.
[92] R. Stross, Advertisers face hurdles on social networking sites, The New York Times, December 14, 2008.
[93] R. Stross, When everyone's a friend, is anything private? The New York Times, March 8, 2009.
[94] J. Swartz, Social-networking sites work to turn users into profits, USA Today, May 11, 2008.
[95] D. Talbot, How Obama really did it, Technology Review 111 (5) (2008) 78–83.
[96] 〈http://techcrunchies.com/how-many-articles-exist-on-wikipedia/〉.
[97] List of social media and social networking sites, 〈http://traffikd.com/social-media-websites/〉.
[98] B. Urstadt, Social networking is not a business* * But it might be soon, Technology Review 111 (4) (2008) 36–43.
[99] M. Villano, The online divide between work and play, The New York Times, April 26, 2009.
[100] The future of social networking sites like Orkut, Ryze, Friendster, 〈http://www.webtalkguys.com/020704.shtml〉.
[101] The top ten social networking sites, 〈http://webtrends.about.com/od/socialnetworking/social_network.htm?p=1〉.
[102] Social network service, 〈http://en.wikipedia.org/wiki/Social_network_service〉.
[103] List of social networking websites, 〈http://en.wikipedia.org/wiki/List_of_social_networking_websites〉.
[104] Cyberstalking, 〈http://en.wikipedia.org/wiki/Cyberstalking〉.
[105] J. Wortham, Twitter-savvy hackers tweak the Twitterati, The New York Times, January 5, 2009.
[106] J. Wortham, MySpace turns over 90,000 names of registered sex offenders, The New York Times, February 4, 2009.
[107] J. Wortham, With Oprah onboard, Twitter grows, The New York Times, April 17, 2009.
[108] 〈http://en.wikipedia.org/wiki/YouTube〉.
[109] E. Zeman, YouTube will steal $471 million from Google in 2009, 〈http://www.informationweek.com/blog/main/archives/2009/04/youtube_will_st.html〉.
[110] Social network sites: my definition, 〈http://www.zephoria.org/thoughts/archives/2006/11/10/social_network_1.html〉.
[111] K. Zezima, Web space where religion and social networking meet, The New York Times, June 30, 2007.
[112] 〈http://www.infoq.com/presentations/Facebook-Software-Stack〉.