

Saila Ovaska and Juha Leino

A Survey on Web 2.0



DEPARTMENT OF COMPUTER SCIENCES
UNIVERSITY OF TAMPERE

D-2008-5

TAMPERE 2008

UNIVERSITY OF TAMPERE
DEPARTMENT OF COMPUTER SCIENCES
SERIES OF PUBLICATIONS D – NET PUBLICATIONS
D-2008-5, JUNE 2008

Saila Ovaska and Juha Leino

A Survey on Web 2.0

DEPARTMENT OF COMPUTER SCIENCES
FIN-33014 UNIVERSITY OF TAMPERE

ISBN 978-951-44-7389-0
ISSN 1795-4274

Abstract

Today's Internet is a far cry from the network of academic sharing as which it began. From the ruins of the dot-com bubble has risen a brave new Internet that O'Reilly has named Web 2.0 while others prefer such names as social net. We were interested in what characterizes today's Internet services and set out to study eleven Web 2.0 sites that encapsulated the new breed of Internet services.

We found that O'Reilly's definition of Web 2.0 describes well what is happening on the Internet today. Today's Internet is indeed about harnessing collective intelligence and about user-contributed content. Huge numbers of items require us to use social navigation with its recommender systems to find items of interest and users have advanced from being simple consumers of content to being a major source of the Web 2.0 content as well. Users contribute content directly by uploading text (in blogs, forums, and reviews), photos, and video clips, and in addition to such intentionally contributed content, the systems generate content by tracking user activities.

Moreover, today's Internet services are characterized by sociability. While some services merely provide means for communal discourse, many others, such as MySpace, LinkedIn, and Facebook, are based on building and maintaining social networks. Regrettably, the social aspects and user-contributed content of the services have also lead to multi-faceted privacy concerns and even such criminal activities as identity theft and child molestation. Furthermore, copyright violations have become an everyday phenomenon.

This survey offers examples of modern, state-of-the-art interface features in today's net and descriptions of the services from the user's viewpoint. The main goal of the presentation is to outline the current state of Internet services together with recent research findings about them. However, we have not shied away from using many blog posts and other writings on the Internet as source material because it is on the Internet where the web of the future is currently being woven.

Keywords: Web 2.0, social net, social networking sites, recommendations, blogging, tagging, privacy, trust, identity, social cues, social presence, user-generated content, citizen journalism

Contents

Preface: Seminar on Web 2.0.....	1
1 Introduction	2
2 Sites studied for this paper	6
3 Social networking and online communities.....	9
3.1 User profiles and online identity	10
3.2 Awareness and social presence	13
3.3 Motivation for user participation	16
3.4 Community policies, oversight, and coordination of action	17
3.5 Networking and making friends.....	18
3.6 Privacy concerns.....	23
4 Tools for harnessing collective intelligence	29
4.1 Social navigation.....	30
4.2 Personalization	32
4.3 Recommendation systems.....	33
4.3.1 Algorithm-based recommenders: Collaborative filtering.....	34
4.3.2 Human-to-human recommendations	36
4.3.3 Popularity-based recommendations, rankings, and ratings	36
4.4 Tagging	39
4.4.1 Tags and vocabulary	39
4.4.2 Geotagging.....	41
4.4.3 Tag lists and tag clouds in the interface	42
4.5 History information	44
4.6 Push and pull technologies: feeds, instant messaging, shoutboxes, and chats	46
4.7 Widgets, gadgets, mash-ups, and open APIs.....	49
5 Collective intelligence as content.....	51
5.1 System-generated content and user-generated content.....	52
5.2 Business value and user generated content.....	52
5.3 User-generated content concerns.....	54
5.4 Approaches to textual user-generated content	59
5.4.1 Discussion forums	59
5.4.2 Blogs	60
5.4.3 Wikis.....	61
5.4.4 Commenting	62
5.4.5 Reviewing.....	63
6 Conclusion	64
References	66

Preface: Seminar on Web 2.0

Web 2.0 is a vague concept, considered by many merely a buzzword, but relevant to all working in the fields of interactive technology and user interfaces. To understand better what today's web is and what kind of research challenges it presents, we arranged a seminar to study the phenomenon in the field and to survey research papers already written about its various facets. The seminar, led by Saila Ovaska and Juha Leino, was an advanced studies course in Interactive Technology at the University of Tampere. The other seminar participants were Vesa Huotari, Jarno Ojala, Hannamari Saarenpää, Jussi Saine, and Markus Tammi.

In the seminar, we studied and compared the Web 2.0 services that appear often in news headlines, such as Amazon, LinkedIn, MySpace, YouTube and Wikipedia. Mikael Johnson from Helsinki University of Technology gave us a speech about Habbo Hotel and his studies of its user community. While Habbo Hotel is by far not as popular as the other sites we studied, it is the only one with Finnish roots and developers. Furthermore, we chose to add MovieLens to our site collection as well. Although its visitor count does not reach millions, it is a site where many important recommender system studies have been conducted.

We started the seminar off with discussions about the phenomenon, by using the services ourselves, and by reading research papers about various facets of the phenomenon. Then we focused on eleven popular Web 2.0 services that we felt encapsulated the different aspects of Web 2.0 and started to look at them in depth to find common denominators and to see what each service's area of specialty is. However, we did not restrict our discussion only to these eleven sites but also discussed some other popular services when relevant. Meetings and discussions continued as we started to write about the features, such as user profiles, tagging, blogging, and collaborative filtering, in a shared Google docs workspace.

The seminar ended on June 18, 2007, after which Saila Ovaska and Juha Leino compiled the material generated during and after the seminar into this report.

Contact information:

Saila Ovaska (Saila.Ovaska@cs.uta.fi) and Juha Leino (Juha.Leino@uta.fi)
Tampere Unit for Computer-Human Interaction (TAUCHI)
Department of Computer Sciences
University of Tampere

1 Introduction

Today's World Wide Web is different from the web that started to gain popularity in the mid-1990s. Corporations have moved in, and the network for academic sharing and free movement of ideas has become a billion-dollar business. Moreover, censorship has entered the picture. For instance, Wikipedia and Flickr are blocked by "China's Great Firewall" (Reuters, 2007). YouTube was shortly blocked by Turkey in March 2007 for having clips that insulted modern Turkey's founder, and it continues to be blocked in Thailand for having clips critical of the country's monarch (Fuller, 2007). Interestingly, censorship seems to have increased coincidentally with the advent of user-generated content.

In 2003, Dale Dougherty, working for O'Reilly, coined the term "Web 2.0" to describe the post dot-com-bubble Internet that had again grown to be a thriving center of business and was on brink of a new era (Musser, O'Reilly, & O'Reilly Radar Team, 2006). However, others have objected to the use of the term. For instance, Slashdot founder Rob Malda says that "what people are calling Web 2.0 is just the realization of what the Internet was always meant to do" (Noyes, 2007). In the same way, Tim Berners-Lee and others have questioned the meaningfulness of the term as much of the technology that Web 2.0 uses has existed since the early days of the Internet (Wikipedia, 2007p). In fact, much of what today seems like a leap forward has been envisioned decades earlier by such men as Vannevar Bush and J.C.R. Licklider (Weiss, 2005).

What are the defining characteristics of the "new" Internet? Tim O'Reilly himself also underlines that it is not about technology: "Anybody who thinks that this is about AJAX is completely missing the boat" (Tweney, 2007). Technology, such as AJAX (Asynchronous JavaScript and XML) or Ruby on Rails, is just plumbing, "and most people don't care about plumbing" (Bricklin, 2000). Tim O'Reilly claims that Web 2.0 is about data (Tweney, 2007) and attitude (O'Reilly, 2005). Web 2.0 is a new approach that underlines the participation of the users. Users have become contributors and the services are harnessing their collective intelligence (O'Reilly, 2005). One central idea of the Web 2.0 services is that the more they are used, the better they get (Musser et al., 2006).

In this paper, we use the term Web 2.0 while noting that its exact meaning is unclear and that it has not reached absolute acceptance in the community. We adopt the term to describe today's popular Internet services, and because much of our paper in fact discusses what defines today's Internet services, we feel confident that the disagreements about the exact meaning of the term are not relevant here.

While O'Reilly suggests that 2001, the year the dot-com bubble burst, was the year when Web 1.0 came to an end and Web 2.0 was born (O'Reilly, 2005), that year more probably marks the change in the business paradigm of the Internet. New vigor emerged (Weiss, 2005) from the smoking ruins of the dot-com dreams as the developers were freed from the manacles of the old paradigm. Web 2.0 is about seeing it all with new eyes (O'Reilly, 2005).

Be that as it may, the term has caught on even if nobody can agree on what the term exactly means. As of June 22, 2007, a search in Google gave 208 000 000 hits for a search for "web 2.0" (with quotation marks). Whether the term is just a business buzz phrase for selling old stuff in a new package or not, Web 2.0 has come to denote such modern and new Internet

services as YouTube, Flickr, Wikipedia, MySpace and so on. These new services are extremely popular with millions of unique visitors each month (Table 1), and the number of unique visitors is still growing at amazing speed. The data in Table 1 is based on the monthly statistics collected by Compete.com, an US online traffic analysis company.

May 2007	MySpace.com	YouTube.com	Digg.com	Facebook.com
Unique visitors	67 654 880	43 798 702	22 637 952	20 284 357
Pageviews per visit	66	15	6	43
Visits per unique visitor	17	4	2	13
Growth (May 2007 vs. May 2006)	+29%	+215%	+1400%	+88%

Table 1. Fast growth of some Web 2.0 services (Meattle, 2007).

People have embraced the new methods of contributing. Blogging may not differ philosophically from an often-updated home page but the simple tools for having a blog without any need to know even elementary HTML has brought the means to contribute to practically everybody who has Internet access—and there were 1 133 408 294 of us by June 10, 2007 (internetworldstats.com, 2007). Moreover, broadband coverage is inching towards 50%, and 50% of the US adults have contributed content online (Musser et al., 2006). The success of Flickr, Del.icio.us, and Wikipedia all point out to the fact that there is a social order for this type of means of contribution. Perhaps the miracle is not that Web 2.0 services are growing so fast but the fact that it took us so long to create the tools to harness all this energy since the technology has been there from the start.

In the early days of web, if somebody made a new homepage, it was news and the few users around actually went to see the page. Nowadays, nobody knows how many web sites there are and nobody would try to visit them all. We need search engines to find the sites relevant to us. A similar situation has developed in most Web 2.0 services. The numbers of items in them are such that we need means to find the ones that are of interest to us. Collective intelligence is one way to do that. Not only can we see what is hot and popular but we can also be recommended items that are likely to be of interest just to us based on our behavior and the behavior of others in the service. Social navigation and personalization have become means to deliver us, the users, what we are interested in rather than leaving us to figure it out with millions of items to choose from.

Furthermore, Web 2.0 is about sharing and users networking with other users. Dedicated social networking sites and other sites providing tools and means for networking are growing fast by any standard (Table 1). In addition, awareness has become one of the central themes in today's web and in software applications used by more than one person. Especially in the Web 2.0 sites, we need various means for social awareness to be able to take part in social networking and to benefit from it.

The concept “social network” can have several connotations and meanings, depending on the context: social network as opposed to technical network underlines the fact that the network consists of human beings and their relationships. Social network as opposed to, say, professional network, emphasizes the nature of the relationships between participants. Within this paper, we use the term “social” as a neutral way to refer to ties between human beings that do not necessarily have to involve affection or friendship. Moreover, here social network

is understood to consist of human beings and social networking technology that enables forming and maintaining ties between them.

One aspect that is common to all the social networking sites is the users' willingness to produce content and share it with others. Content here can be as simple as building a network identity, a profile that enables the user to join a community or a group within a community. This is often called social networking: forming networks of people by linking to their profiles or to content they have made available to others. Online users can make more than just their profiles public. The content created and shared by users can be bookmarks, pictures, media files, music, video, or own writings—anything that they consider interesting to other people.

Table 2 summarizes some of the collective intelligence tools and approaches to user-generated content in the eleven sites that we studied. While it is not even meant to be exhaustive—creating such a table would probably be impossible in any case—it does afford a glimpse at what is going on today in the Web 2.0 services from the feature-content viewpoint. Features typical to Web 2.0 are discussed in detail in Chapters 4 and 5.

Web 2.0 typical features	Amazon	Del.icio.us	Flickr	Habbo	Last.fm	LinkedIn	MovieLens	MySpace	Technorati	Wikipedia	YouTube
Recommendation systems	X	X	X	X	X	X	X		X		X
Algorithmic matching	X	X			X	X	X				X
Ratings	X	X	X	X	X	X	X	X	X		X
Rankings	X	X	X	X	X		X	X	X		X
Wiki	X				X			X		X	
Instant Messaging / Chat				X	X			X		X	
Commenting	X		X					X	X		X
Reviewing	X		X								
Tagging	X	X	X		X		X		X		X
Discussion forum/board	X	X	X		X		X	X	X	X	X
Blog	X	X	X		X			X	X		X
Web Feeds	X	X	X		X	X	X		X	X	X
Newsletters and subscribed emails	X	X	X	X				X	X	X	X
Open API	X	X	X		X			X	X		X
Marking items as favorites			X		X			X			X

Table 2. Features in the eleven services studied for this paper in Spring 2007.

Much of today's feature development is based on both allowing and utilizing user-based actions and contributions. How can the users of a service contribute and how can these contributions be used to generate value? What user actions should be recorded and how to generate value out of them to the community? How to encourage user contributions? Tags, for instance, are one such approach. Users can add words to describe an item, be it a link,

photo, or book, and while the users manage their own links with their tags (part of the motivation), the site uses the tags for collaborative filtering and social navigation.

Much of this paper deals with user-based content, that is, content that is either explicitly provided by the users or figured out implicitly by the system based on what the users do within the site. However, both implicit and explicit collecting of information and the constant profiling of users have also introduced number of privacy issues in addition to the concerns of content quality and ownership that also need to be addressed in this paper.

In one sense, one could claim that the whole Web 2.0 is about supporting awareness. Awareness is a broad concept that is defined in biological psychology as “a human’s or an animal’s perception and cognitive reaction to a condition or event” that “does not necessarily imply understanding, just an ability to be conscious of, feel or perceive” it (Wikipedia, 2007b). Thus, awareness can be conscious, partially conscious, or sub-conscious. Most if not all widgets and features in the user interfaces of the modern sites support awareness one way or another. For instance, tag clouds, what’s hot, and new community member lists, all show the users where the action is and what is happening in the community.

In this paper, we limited our scope to the aspects of the Internet services that are used with a web browser, thus leaving the two billion mobile devices (Musser et al., 2006) in the world outside of our discussions. While social media is not limited to the Internet browsers by any means and most services are creating content for different platforms—for instance, Flickr has an interface for mobile phones as well—we simply had to draw the line somewhere.

Furthermore, while Web 2.0 certainly creates new business opportunities (Hintikka, 2007; Kangas, Toivonen, & Bäck, 2007; O’Reilly, 2005), we will not view the phenomenon from the business viewpoint as much as from the service and user viewpoint. Different viewpoints have much in common, however, and thus some business aspects are also touched upon when it comes to the huge numbers of users and user-generated content.

We start this paper off with brief descriptions of the eleven sites that we studied in-depth in Chapter 2. In Chapter 3, we discuss social networking and privacy issues before moving on to collective intelligence and content-related issues.

In Chapter 4 we look at the use of collective intelligence in terms of social navigation and personalization before discussing recommender systems and other features that characterize the sites that we studied. From collective intelligence, we move on to discussing content sources in today’s popular sites in Chapter 5. We look at different sources of content and then at approaches to allowing the contribution of user-generated content.

2 Sites studied for this paper

We chose eleven sites for closer scrutiny for this paper. Our main criteria were that the site represented some aspect of Web 2.0 and was popular. We tried to avoid picking competing sites that would at least to some extent repeat each other's approach and features. Thus, we picked MySpace but did not include Facebook.com because MySpace is still today more popular than Facebook, even though Facebook's popularity is growing at a very fast rate (Table 1). Our collection cannot possibly represent all the important aspect of the literally hundreds of popular services. However, with limited time and resources we had to start somewhere. Here are in alphabetical order the eleven sites that we selected.

Amazon.com (<http://www.amazon.com/>): Amazon is the biggest online retailer in the world, and it has constantly been an early adopter and developer of various approaches to recommender systems and user-generated content. Of all the sites we studied, Amazon offered the widest collection of Web 2.0 features.

Del.icio.us (<http://del.icio.us/>): Del.icio.us is a social bookmarking site that allows tagging and sharing of bookmarks. It allows its users to profit from collective intelligence and access their bookmarks from any computer with Internet connection.

Flickr (<http://www.flickr.com/>): Flickr is photo-sharing service in which tagging forms the backbone of the navigation. Flickr has also large number of user groups that have their own discussion forums.

Habbo (<http://www.habbo.fi/>): Habbo is a teen community implemented as a graphical chat. The community members, represented by their avatars ("Habbos"), meet people, play games, and create their own online rooms in the virtual hotel where the action takes place. While joining, building one's avatar, and chatting with others is free, other activities, such as buying furniture ("furni") or a pet, cost real money that is represented by Habbo coins. Habbo Hotel also has external fan sites (Johnson & Toiskallio, 2005) that host user groups. These fan sites are not affiliated with Habbo Hotel but are born out of fan activity.

Last.fm (<http://www.last.fm/>): Last.fm is a social web radio that uses "scrobbling" to collect information about what its users listen to so that it is able to give its users the kind of music they want to hear and allow them to discover new artists with collaborative filtering. Last.fm has a large user community with Friends, Neighbors, and Groups.

LinkedIn (<http://www.linkedin.com/>): LinkedIn is a service for connecting with people. While some connections are based on knowing the other person through a shared background, such as high school or job, the site supports connecting with new people as well. The social network in LinkedIn helps to keep aware of the changes that take place in the community, e.g. when a former colleague finds a new job.

MovieLens (<http://movielens.umn.edu/>): MovieLens is a movie recommendation site run by GroupLens Research at the University of Minnesota. It requires explicit ratings from its users

and uses collaborative filtering for recommending new movies. It also has some social networking features.

MySpace (<http://www.myspace.com/>): MySpace provides a full range of features supporting communities. It enables not only building one's profile but also contacting others, seeing videos and photos they have taken, reading their journal entries within the site, and so on.

Technorati (<http://technorati.com/>): Technorati is a search engine and monitor for blogs. It assigns them authority (based on the number of blogs linking to the blog in the last six months). Technorati uses tags extensively. Technorati also has listings of music and videos.

Wikipedia (<http://en.wikipedia.org/>): Wikipedia is a multilingual, collaboratively written online encyclopedia. It is based on wiki, a web application designed to allow multiple authors to add, remove, and edit content.

YouTube (<http://www.youtube.com/>): YouTube is "the world's most popular online video-sharing website where users can upload, view and share video clips" (theage.com.au, 2007). YouTube also has a large user community with Groups and Friends.

It is a difficult task to compare the popularity of the sites. Although several net traffic analysis companies collect clickstream statistics, the actual numbers can vary a lot depending on how the data is collected. Table 3 shows some statistics of the sites based on the data collected by Alexa.com (the first column) and Compete.com (the other columns). The number of unique visitors is indicative of the number of people interested in a site, but it can be influenced other factors as well, such as excessive advertising that generate one-time visits from people who stay only shortly and never return. The number of visits by these unique visitors shows how many times each user has returned to the site within the month. Another measure of engagement with a site is how many minutes a visit lasts. Table 3 has no data on Habbo and MovieLens since they are not present in the samples of clickstream data the companies collect.

Launched		Alexa traffic rank (June 2007)	Compete traffic rank (July 2007)	Unique visitors (July 2007)	Growth (July 06-July 07)	Visits (July 2007)	Average stay (July 2007)
Amazon	1994	31	10	47.1m	+12%	132.8m	7 min
Del.icio.us	2003	219	-	1.7m	+707%	4.5m	3 min
Flickr	2003	45	27	23.6m	+161%	54.8m	7 min
Habbo	2001	13491	-	N/A	N/A	N/A	N/A
Last.fm	2003	320	1450	1.2m	+28%	2.2m	6 min
LinkedIn	2002	164	611	2.4m	+769%	7.5m	7 min
MovieLens	1987	3512	-	N/A	N/A	N/A	N/A
MySpace	1995-6	6	6	68.3m	+24%	1.3b	27 min
Technorati	2002	197	417	3.2m	+74%	7.3m	2 min
Wikipedia	2001	9	12	41.4m	+53%	124.6m	9 min
YouTube	2005	4	9	50.2m	+168%	241.4m	16 min

Table 3. The sites compared (m: millions of visitors and visits; b: billions of visits).

Alexa (<http://www.alexa.com/>) determines the site traffic ranks based on clicks collected using a special toolbar that the users have to install in their browser. The smaller the number (rank), the more popular the site is. The traffic rank is based on the aggregation of clicks in the last three months, and it takes into account how many users within the Alexa toolbar user group are using the site and how many pages they view on the site. For instance, according to Alexa, in June 2007, YouTube was the fourth most often-visited site, right after Yahoo, MSN, and Google. However, the actual number of people who have installed the Alexa toolbar and their nationality are not revealed.

In comparison, Compete (<http://www.compete.com/>) bases its statistics on clickstream data generated by more than two million U.S. Internet users. Compete also requires a toolbar to be installed into the web browser. The toolbar gives the visited sites a trust score as well as describes the site profile and lights up if there are special sales promotions on the site.

As brought up already in Table 1 (page 3), the growth rates of some of the Web 2.0 sites are phenomenal (Meattle, 2007). While the true site popularity is hard to define when visitors may change their pseudonyms often and may not return to the site after their first exploration, the monthly number of unique visitors in MySpace has reached 67 million, and it is still growing fast at the annual growth rate of 24%. Nevertheless, some other sites are growing even faster. For instance, LinkedIn grew by 769% in one year (July 2006 – July 2007).

Overall, however, the traffic statistics and the ranks based on them should be approached with caution. The user panels of Compete and Alexa are opt-in panels that run on volunteers. The statistics are based on the clickstream of those who voluntarily install a toolbar to share their clickstreams with such companies as Alexa and Compete. The panels may be biased and not represent the whole Internet population, especially users and sites in countries outside Northern America. In fact, Compete's statistics are based entirely on the U.S. users. Not all users are willing to share such data with a company due to privacy concerns and fear of being logged. Furthermore, the toolbars are not even available for all the browsers. Nevertheless, we believe that these statistics do provide a general idea of the amazing popularity of the sites studied for this paper.

3 Social networking and online communities

Social networking has existed in the computer networks since the early days of the first online communities, such as WELL, a text-only conferencing system that came online in 1985 (Rheingold, 1998). However, tools for social interaction pre-date such communities. Among the first awareness support systems was finger (used in the UNIX systems since mid-1970s). Finger allowed the users of the system to see who else was online and contact them with talk, another UNIX software application. Since the advent of Internet Relay Chat (IRC), non-local group chats have become possible. Chats required new functionality to help users stay aware of their contacts and their online presence and availability. These early social networking tools and studies of them have had a pronounced impact on shaping the current services and the modern Web 2.0 functionality.

Awareness was first studied in the field of computer science in the context of Computer Supported Co-operative Work (CSCW) tools in the late 1980s and early 1990s. Various shared workspaces, both synchronous and asynchronous, were built to support joint work. At the same time, media space studies emphasized the social needs of the workers. These areas of research converge in the modern Internet where platforms and workspaces for communication, collaboration, and coordination between remote collaborators are today's reality.

In social networking sites, the whole site and its users can be seen and described as a large community (for instance, YouTube community, MySpace community). Such communities may not really have traditional communal attributes, such as a shared purpose (Preece, 2000), but in today's parlance, the word is commonly used to describe the whole site and all the people involved. Still, they are communities in the sense that the rules of the site apply to all members. By registering to the site, the users agree to the policies of the community even if they do not necessarily actually read the rules. However, some researchers claim that the concept of "virtual community" should be reserved to such communities where a sense of virtual community and behavior supporting the community has been observed among the participating people (Blanchard & Markus, 2004).

Some user-related information is collected automatically by the service by tracking the users' actions. This way, the service might be able to recommend contacts with other users that share similar taste for music or some other content in which other like-minded people have been interested. Algorithmic approaches to recommender systems are covered in Chapter 4. While the user can benefit from the recommendations given by the site, the automatic data collection for generating the recommendations does not involve explicit user input and own contributions. Thus, it is not clear if there actually is a feeling of community present.

The reasons for belonging to a community and taking part in its activities vary. It is not clear if the reasons have actually changed much since the early days of the WELL or other text-only discussion forums (Preece, 2000; Wellman & Gulia, 1999). The need for sociability is

innate to the human beings, and the Internet with all of its social networking sites appears to appeal strongly to our social instincts. Motivation for participation is a theme in Section 3.3.

Social web sites naturally benefit from the human tendency for social interaction. However, they also face the challenge of finding the right ways of presenting the awareness cues that enhance the feeling of being with others and that help the users with their tasks in the community context. The sites discussed in this paper have approached this challenge in various ways, some of which are described in this chapter.

To be successful, a social networking site needs to pay attention to basic social interface components that can foster social interactions. Study by Girgensohn and Lee (Girgensohn & Lee, 2002) found several such components. They discovered that the site must provide a common ground, that is, a shared understanding among collaborators, for instance by letting them introduce themselves with their own words and pictures in their user profiles. User profiles are discussed in detail in Section 3.1. Likewise, Girgensohn and Lee emphasized that it is important to support the community members' on-going awareness of each other by adding activity indicators and traces of activity, and to provide them with cues of other users' availability for a variety of interaction possibilities, together with the actual mechanisms of interaction in the user interface. Awareness and presence indicators are discussed in Section 3.2., and we return to the user interface mechanisms for making contact in Section 3.5.

Eventually, any community forms naturally a cultural and social understanding of the norms and practices that are appropriate in the particular network "place" (Girgensohn & Lee, 2002). These are discussed in Section 3.4.

3.1 User profiles and online identity

To join a social networking site and benefit from it, one typically needs to register and build a user profile. Although in many cases it is possible to view the content other users have made publicly available without registering, without a profile, it is impossible to start building a community identity that allows social networking.

The user profile is a collection of information items that a user chooses (or is required) to reveal about himself or herself to other people in the community and possibly more widely on the Internet. If made public, the profile information is shared with friends and strangers alike, and since the number of the user counts in many social networking services reaches millions, nobody can actually know how many people have an access to one's profile information even if it is not open to everybody on the Internet. Without such personal representations as profiles, however, many tasks of social networking are not possible, but revealing such information has raised concerns about privacy violations (for example, Gross, Acquisti, & Heinz, 2005; Privacy International, 2007).

The profile information is one type of user-contributed content. In addition to personal descriptions and facts, the information in many social networking sites also includes explicit social information, such as articulated "friend" relationships (boyd & Heer, 2006). Thus, in addition to the relatively static personal data, the profile also includes detailed information about the person's social contacts as long as those are formed and maintained within the system.

The profile data does not necessarily reveal any such information that could identify the person behind the profile. Still, some sites need to connect a particular user to a real person. For instance, in Amazon the user at some point needs to enter credit card information to be able to buy something, and in LinkedIn, job hunters cannot be contacted if they do not reveal their true identity. The sites encourage revealing identifiable information through technical

specifications, registration requirements, or social norms (Gross et al., 2005). Though the users seldom fill in all fields of a user profile, those fields with content entered by the user work to encourage connections and articulated relationships between users (Lampe, Ellison, & Steinfield, 2007).

The amount and level of identifiable information required in the profiles varies from site to site. It is a common practice of the service providers to verify an account by sending electronic mail to the email address entered as part of the profile. Of course, this does not necessarily connect the profile to any particular real world person since the email service providers do not generally have means to verify their users' identities, either.

Usually most profile items are voluntary to fill in. Sometimes the maintainers of the site have included filling in the profile fields into the registration process where the user acquires a pseudonym (user ID; login name, user name, or display name) to use the service. In many services, it is not always explicitly indicated whether each field in the registration form is required or not for getting the user account. To compare the 11 studied sites, we collected information about their registration processes and fields in their user profiles (Table 4). All the sites require the user to create a persistent user ID that can function as a pseudonym.

Profile details	Amazon	Del.icio.us	Flickr	Habbo	Last.fm	LinkedIn	MovieLens	MySpace	Technorati	Wikipedia	YouTube
Pseudo username	X	X	X	X	X	X	X	X	X	X	X
Name	X	X	X			X	X	X			
Email	X	X	X	X	X	X	X	X	X	X	X
Location (IP, ZIP, timezone) information	X		X		X	X		X	X	X	X
Gender			X		X			X			X
Marital status			X					X			
Photo	X				X			X	X		X
Other contact information (IM)			X								
Personal bio, description	X		X	X	X	X		X	X		X
Personal interests and favorites	X		X	X		X		X	X		
Birthday or year	X			X	X			X			X

Table 4. User profile items in the sites.

As pointed out by Ma and Agarwal (2007), identity has several facets, and a person's online identity can differ significantly from his or her offline (true) identity. Identity is much more than just the name of a person. It covers personal traits and motives, physical and cognitive abilities, and social roles one may have as student, worker, family and community member, among others.

Relating an online identity to an offline identity is many times impossible—even for researchers of network communities, which is why the research questions relating to online identity forming tend to be studied online without having a connection to the offline identities

and behavior of the people involved. Ma and Agarwal (2007) studied online identity relating to the forms of social interaction within the online communities. They found that the technology used affects how the participant can maintain an online identity and how the identity is verified within the interactions with the other members of the community. The technology artifacts that support building an online identity include such items as a persistent user ID, user profile pages, member directories, reputation or rankings with which the users can rate each other based on some criteria, such as trustworthiness, and tools supporting virtual co-presence, such as knowing who is online at the moment (Ma & Agarwal, 2007).

Numerous surveys have repeatedly indicated that the Internet users are very worried about privacy matters (Kobsa, 2007). However, practice demonstrates consistently that people are quite willing to publish personal information on the Internet (Kobsa, 2007). For instance, an early analysis on Facebook (<http://www.facebook.com>) showed that close to 90% of the profiles contained image, birthday, and high school information. Hometown, address, relationship status, and interests were entered in between 50% and 70% of the profiles, and 40% contained a telephone number. (Gross et al., 2005)

People's willingness to provide information depends on several factors, one being the type of information requested. Kobsa (2007) found that people were by far the most protective of their personal contact information and financial information. Furthermore, the less desirable a trait is in the context of a group, the less willing people are to reveal it. Demographic information and information on one's preferences, on the other hand, are given the easiest. (Kobsa, 2007)


According to Kobsa, Internet users fall into clear categories in relation to revealing information about themselves when it comes to stated attitudes. Interestingly, observations of behavior do not support the existence of these categories. People categorized as "privacy fundamentalists" according to their stated attitudes are not much more likely to withhold information than other categories. (Kobsa, 2007)

In Kobsa's study, the other factors that affect people's willingness to provide personal information were the perceived value of personalization gained by providing information, knowledge of and control over who are the users of personal information, trust in the collecting website and the reputation of the website operator. Young people tend to value personalization somewhat more than older users. Furthermore, the willingness of providing information depends on past positive experiences, design and usability of the website, and the presence of a privacy statement (although they are rarely read). (Kobsa, 2007)

While some sites make the profile data public to all by default, in some others it is by default not shown to other users but collected for the site records. Sometimes, some parts of the profile information are used as basis for automatic recommendations within the system. For instance, LinkedIn tracks other members who have studied in the same school with the user and provides the user with means to contact them. Likewise, the birthday date that a user enters in MySpace appears automatically as a reminder (if the date is close enough) for his or her designated friends when they log into the system.

Many sites ask the users to add a photo to the profile. MySpace community policy prohibits photos that contain nudity or are otherwise offending, or if the user does not have the copyright to the photo. However, it is ok to add photos from which it is easy to recognize the person. Of course, the images and accompanying information vary greatly in detail and style. For instance, "Tom" in Figure 1 has not given his last name or detailed contact information but is easily recognizable in his MySpace profile photo.

Tom



":-)"

Male
31 years old
Santa Monica,
CALIFORNIA
United States

Last Login:
6/24/2007

View My: [Pics](#) | [Videos](#)

Contacting Tom			
	Send Message		Forward to Friend
	Add to Friends		Add to Favorites
	Instant Message		Block User
	Add to Group		Rank User

Figure 1. Tom's user profile (<http://www.myspace.com/tom>) from MySpace (June 25, 2007).

In a study comparing Facebook and Friendster user profiles, the percentage of images obviously unrelated to a person ("joke image") was much lower in Facebook than in Friendster profiles: 12% vs. 23% (Gross et al., 2005). Gross et al. note that Friendster is more clearly oriented towards social contacts online. Having an account in Facebook is clearly a norm to college students in the USA. In such a service, it is natural to expect to be recognized in the real world (campus area) as well as in the service.

3.2 Awareness and social presence

In 1992, Dourish and Bellotti gave awareness an oft quoted (for example, Andersen, Jørgensen, Kold, & Skov, 2006; Liechti, 2000; Raento, 2007) definition: "awareness is an understanding of the activities of others, which provides a context for your own activity" (Dourish & Bellotti, 1992). In fact, most papers give the quote as the definition of social awareness, although Dourish and Bellotti were defining awareness. Consequently, nowadays this is seen as a de facto, broad definition of social awareness. However, a myriad of different concepts complement it to focus on specific aspects of awareness. Such concepts include, among others, social presence (Andersen et al., 2006; Preece, 2000), contextual awareness (Liechti, 2000), and situational awareness (Espinosa et al., 2000).

For our purposes here, we are comfortable using Dourish and Bellotti's broad definition, as it covers also awareness of the actions of others in the shared space, such as a web site, that are not necessarily directly related to our task at hand or the artifact we might be currently manipulating (Raento, 2007). Furthermore, it includes the history of the actions that have taken place in the site. Those actions have formed and shaped the information environment (Liechti, 2000) where we work, thus encompassing such concepts as social navigation, that is, navigational aids based on the actions of people in the information environment. Social awareness is here understood to include the context of the activities and people's presence in the information environment as well.

Prinz (1999) contrasts social awareness with task-oriented awareness in CSCW. He argues that social awareness "includes information about the presence and activities of people in a shared environment," and contrasts it with task-oriented awareness, that is, "the awareness that is focused (sic) on activities performed to achieve a specific shared task." He further points out that task-oriented awareness "can be promoted by change notifications or

information about the state of a certain document or a shared workspace,” and that it “allows users to coordinate their activities on the shared object.”

“The difference between task-oriented and social awareness is primarily determined by the shared context. For task-oriented awareness the shared context is established by an object that is part of a cooperative process, for social awareness it is the environment that is inhabited by the users.” (Prinz, 1999)

Thus, we have two wide approaches to seeing, supporting and studying awareness in the social web environments, one, social awareness, focusing on humans (their actions, presence, context etc.) and one, task-related awareness, focusing on artifacts (different types of changes in and to them, including creation and deletion, who has made the changes, consequences of the changes etc.) As Prinz (1999) states, we have to consider both and, in many if not all cases, design ways to support both.

While the division into task-oriented awareness and social awareness works well for analyzing purposes, the two types of awareness information are often combined in today’s interfaces. For instance, Figure 2 gives an example of member promotion in Technorati: the public user profile of “usabWS”. The username and date of joining Technorati are always public information. Favorites list the blogs that the user has marked as favorites, and the user is told if they have new content for the user. “Authority” indicates how many blogs have been linked to this blog within 180 days. The authority information is generated automatically. Thus, social awareness of the user is augmented with up-to-date information about the artifacts she is interested in. In a sense, the artifact information becomes social information about the user when it is combined with the user information in this manner.

Everything in the known universe about usabWS

usabWS (Saila O)
Joined on May 01, 2007

Favorites

Being Being: A Directory of Wonderful Things
by **ottomatik** · 5 hours ago
Authority: 23,272
(No tags)
<http://www.beingbeing.net>

Pinseri
by **samikki** · 8 days ago
Authority: 61
(No tags)
<http://www.pinseri.com>

[View all »](#)

Figure 2. A part of a Technorati user profile (username usabWS).

In our work with the popular social web sites, we also found many features that are designed to increase what we call trend awareness. Trend awareness features tell us, for instance, where the action is in tag clouds, what is popular or what is gaining or losing popularity (all kinds of “Top ten most popular” lists, or even which camera models have been used to take pictures and how this has changed over time as in Flickr’s Camera Finder. The trends can be of social or task-oriented in nature but they typically tell us what is happening in the community. Flickr’s camera finder, for instance, tells us which cameras are popular and how their popularity has shifted over time.

While popularity tells us about a trend in the community, it has more to do with an artifact (camera) used by the community than with the humans and their activities in the community. On the other hand, such features as tag clouds that tell us “where’s the party at”, as in BBC England’s message board (Murison, 2005), come clearly under social awareness. Thus, the division into task-oriented and social awareness can be somewhat labored at times when applied to the Web 2.0 interfaces and their features. However, its usefulness as a tool for analysis remains in spite of this.

Any site with larger number of users faces the situation where it cannot present all the awareness information to all the users. Instead of providing the users with the benefits of collective intelligence and social navigation, we would end up drowning them in mainly useless information as far as their task, interests, and context are concerned. Consequently, when we design awareness supporting systems, we need to understand the actions of others in relation to the user’s current task and context in order to be able to support the user with the relevant sub-set of awareness information (Liechti, 2000). Liechti calls this kind of awareness of the user “contextual awareness” and argues that we need to both determine “i) what information users should be made aware of, and ii) how they should be made aware of it.” In other words, we need to design proper awareness cues that provide the useful information with minimal or, at most, appropriate disturbance (Liechti, 2000).

The extreme case of awareness information being provided with minimal disturbance, sometimes called peripheral awareness, is where the user is provided the information without requiring them to focus their attention on the information.

One sub-category of social awareness that interests us here is social presence. Social presence or co-presence, as it is sometimes called, includes the “sense of being with others” (Wikipedia, 2007k; Preece, 2000). Awareness system studies have repeatedly found the users experiencing a feeling of not being alone or being physically close to the other users (Raento, 2007; Wikipedia, 2007k). Furthermore, Raento (2007) has found that the mere knowledge that somebody else is also using the system, even if not directly engaging us, produces such feelings. It appears that a sense of space emerges from socially aware systems, and that leads to the sensation of co-habiting that space (Raento, 2007).

Social presence information can exist in real time. “Sense of virtual co-presence” (Ma & Agarwal, 2007) is related to the awareness of other users in the community. Virtual co-presence is affected by all parts of the user interface that induce a subjective feeling of being together with others in a virtual environment—parts showing who is online at the moment, indicating who is allowing instant messages and chat, and showing updates of postings in real time. Ma and Agarwal have shown that the feeling of virtual co-presence increases motivation to participate. (Ma & Agarwal, 2007)

The sites we studied show social presence information in different ways. Flickr enhances social presence by showing up-to-the-minute information of the uploaded photos on its front page (Figure 3). Habbo lists how many users are currently logged in while MySpace shows in the user profile if the person is currently online and thus available for instant messaging. MySpace even implements mood indicators whereby the users can select a smiley face to represent their current mood.

2,195 photos uploaded in [the last minute](#) · 6,292 photos tagged with [fresco](#) · 2.0 million photos [geotagged](#) this month ·

Figure 3. Part of Flickr home page with social awareness cues.

In Habbo Hotel, a Habbo is informed if its group members are in the Habbo Hotel, and it can teleport to the location of the group member. Social presence of the avatars in one's close vicinity enables talking with them, but people further away in the room are not able to "hear", that is, see the contents of the speech bubbles, unless it is being "shouted" (see Section 4.7).

3.3 Motivation for user participation

Although united under the heading of Web 2.0, the eleven sites studied in this paper differ quite a lot from each other in their main goals and the functionality offered to the users. Furthermore, social networking services come in more flavors than these 11 web sites can illustrate. For instance, blogging as an individual journal writing activity is quite common in Web 2.0. As a starting point for interaction with other users, however, it differs greatly from belonging to a community, say, LinkedIn, and building up a LinkedIn profile and connections network.

Furthermore, people blog for many reasons. Nardi, Schiano, and Gumbrecht (2004) found five main reasons in their interviews. Some study participants used the blog as a means to keep someone updated of personal and other activities while some others were hoping to have an influence on something with the opinions they express. Some others wrote blog entries to clarify their thinking and to seek for feedback for their ideas. Finally, for some the blog was a place for releasing tensions and blowing off steam, thus serving the same function as a personal diary. (Nardi et al., 2004)

With all these different reasons for keeping up a blog, it seems quite natural that the motivations for registering in a community and contributing to it vary even more when we consider all types of social networking services in existence today.

While several studies approach the motivational aspects, their viewpoint is often restricted to only one type of social networking software. Moore and Serva (2007) propose a unifying framework for future studies of motivational aspects. One of the motivating factors is reputation. By making contributions that are respected within the community, a user may gain a social standing within the community (Moore & Serva, 2007). Sometimes the community members are especially promoted either by the administrators of the site or by other community members. In Amazon, for instance, a person who has written numerous, well-received reviews can be assigned a badge, such as Top 1000 Reviewer. Also, many sites regularly pick some members to be introduced to others. For instance, in Habbo Hotel's home page winners of a competition are introduced to all with links to their public homepages. Some of these competitions are based on votes by the community members. Similarly, Flickr recommends interesting photos and the people who took them.

The motivation for user participation in a community is tied to awareness of the community members and their interests. Lee (2006) studied how Del.icio.us gives its users means of forming social networks of people interested in the same topics, for instance, CSCW. When the users become more aware of each other's presence, they reveal more information in their user profiles about themselves, giving out not only username but also email and home page address, and participate in the bookmark sharing networks provided by del.icio.us. Though not in real time, this perceived social presence has an effect on the actions that the users make in del.icio.us. The results showed that if the users had strong perception of social presence, they showed more consideration to other users by including annotations that might help them. Furthermore, it is motivating to see the bookmark you have recommended appear in the bookmark lists of the contacts you have in your network page. By adding the bookmarks to their bookmark collections, others show that they value your contribution, which again increases the likelihood of making new contributions. (Lee, 2006)

The contributions a member makes within a community increase his or her satisfaction with the community. When a user is happy with the feedback the community gives for his or her contributions, the user perceives that his or her identity within that community is verified and this motivates further contributions (Ma & Agarwal, 2007).

Typically, many members “lurk” without making any explicit contributions to the community (Nonnecke & Preece, 2000). However, lurkers can still perceive a sense of virtual community. While they do not build up an identity online themselves by actively contributing to the community discussions, they can still recognize other members’ names and observe relationships between them (Blanchard & Markus, 2004).

Both information exchange and emotional support, even emotional attachment to the community, have been observed in studies that have focused on discussion forums (Blanchard & Markus, 2004; Moore & Serva, 2007; Rheingold, 1998). Especially the most active community members feel an obligation to respond to questions and express their support (Blanchard & Markus, 2004).

In many sites, the main contribution that a user can make is recommendations. For instance, in Amazon user-generated content exists in the form of reviews and ratings, and in MovieLens, the ratings are the core of the service that help both the users themselves and others find movies that match their taste. User-generated contributions are discussed in depth in Chapter 4.

In their analysis of the goals and tasks of the users of recommender systems, Herlocker et al. (2004) indicate that some users are motivated to use a recommender system to improve the ratings they get themselves. Others, however, are not looking forward to better matching predictions but simply aim at self-expression of their opinions, even influencing others in the community. Finally, some users wish to help other users by giving ratings. The study points out the variety of user goals and tasks that need to be supported by a recommender system. (Herlocker, Konstan, Terveen, & Riedl, 2004)

3.4 Community policies, oversight, and coordination of action

Communities need to give support to their members, especially new members or “newbies” (Preece, 2000). Such support takes many forms. In Habbo Hotel, there are Habbo X (eXpert) users who are there to help new Habbos. Likewise, in Wikipedia some users have signed up as voluntary mentors to Adopt-a-user program. In both cases, these users are members of the community themselves and they need to have some qualifications or fulfill certain requirements before becoming advisors. Some support can be automated. For instance, Flickr has a so-called shadow application that finds “Loneliest Users”, users who were not inviting friends to the service, and adds Flickr as a contact to those users to teach them how to make better use of the service (Musser et al., 2006).

The need for moderation has long been recognized in mailing lists and newsgroups (Preece, 2000). In Habbo, a filter changes improper language into nonsense words automatically in the chat messages and there are moderators to observe online if any facts that reveal a user’s real identity appear in the chat messages. Practically all social networking sites emphasize that no improper content should appear in the profiles or other content. While such statements exist in the community policy pages, the sites still find it necessary to resort to manual moderation. The sites commonly offer a link to reporting any inappropriate content to the administrators, thus relying partly on the users’ oversight in the moderation task.

Cosley et al. (Cosley, Frankowski, Kiesler, Terveen, & Riedl, 2005) conducted an experiment on the impact of oversight on the quality of the member contributions in member-maintained sites. They found oversight to be “an important social mechanism employed by successful member-maintained communities including Slashdot, Amazon, and Wikipedia.” According to the findings, the presence of oversight encourages and motivates people to make high-quality contributions to the community, increases the number of contributions, and helps reduce vandalism. Peer-oversight and expert-oversight both had the same degree of positive impact, and so the use of peer-oversight is encouraged as resource-effective means. (Cosley et al., 2005)

Wikipedia, an online encyclopedia, is an example of successful collaboration effort on the net. Wikipedia articles are written by tens of thousands of active users working on shared artifacts, that is, the articles. Currently, Wikipedia has more than 5 million wiki pages in several languages. (Wikipedia, 2007m)

Since all revisions of the Wikipedia pages are stored, the evolution of the material and its revision history can be viewed and analyzed. Viegas et al. (2007) have collected datasets from the Wikipedia article database in 2003 and 2005, witnessing a huge growth in size of the encyclopedia. While in 2003 there were roughly 170 thousand English-language articles, in 2005 their number had exceeded 1 500 000. The fast-repair mechanisms noted in 2003 were still in operation in 2005, especially for pages confronting malicious edits and vandalism. They were reverted to the earlier versions in a matter of minutes. (Viegas, Wattenberg, Kriss, & Ham, 2007)

To prevent the so-called Edit wars between competing editors with differing viewpoints, the Wikipedia community has voluntarily accepted a “three revert rule”. No more than three reverts are allowed to a given page in a 24 hour period. Furthermore, the Talk pages associated with each article page have been used extensively as places for planning and discussing article content. They function as places for coordination of action, and the messages often contain links to the community guidelines, writing policies, and even to polls about the wording of the text. (Viegas et al., 2007)

Wikipedia is an example of how a community over years has adopted policies to guide authors, for instance a guiding policy of writing style called NPOV (Neutral Point of View, <http://en.wikipedia.org/wiki/NPOV>). As Wikipedia grows in size, such policies have become even more important to resolve conflicts and maintain the quality of the articles.

However, the policies do not ensure that they are followed by the members. In addition to the actual community policies, the sites need to develop mechanisms through which these policies can be taught to the contributors and enforced in the contributions. The role of some of the contributors has changed increasingly into administrative moderators. (Viegas et al., 2007)

3.5 Networking and making friends

In social networking communities, the term “group” usually refers to a subset of people from the whole community. Groups, sometimes called “neighbor users,” are created based on interests in hobbies, bands, movies, politics, or anything else that combines two or more users.

Groups are a popular means of social networking, as the number of groups in various interest areas created within MySpace show (Figure 4). Typically, the group members are listed on the group profile page, and they contribute to the asynchronous discussion forums within the group.



Figure 4. MySpace groups by category.

Typically, the creation of a group requires activity—sometimes even money, as in Habbo Hotel—from the person who defines the group profile. Others can then choose to join the group although sometimes participation is by invitation only. This way the person who creates the group also has options to define how the group operates and who can take part.

Though the groups are easily created, their functioning is then based on how active their members are. Although the number of groups in any social networking site is huge, empirical studies reveal that very few of them are successful at retaining their members and motivating them to participate (Ma & Agarwal, 2007)

Sometimes a group exists first outside the social networking site. For instance, an international group with members from several countries can invite its members to join LinkedIn via an ordinary email message on the group's mailing list. The message contains a link to the group's page in LinkedIn. After joining, the new members can set if they allow the other group members to contact them and if the group membership information is visible in the profile to outsiders who are not members of the group.

However, groups can also be defined in other way, for instance by automatic tracking of the users' actions in the site. In these implicit or passively formed groups, the system generates groups based on users' shared interest or actions on the site. Because these groups do not require creation or subscription by the users, they are "passively formed". For example, in Last.fm, the system creates groups of listeners who listen to certain artists. Grouping is based on information that is gathered from users' listening data. Last.fm shows neighbors that listen to similar music to you but it is up to the individuals then to make contact with each other.

Obviously, one motivation for taking part in social networking is to find new contacts and make new friends. The sites often promote seemingly closer person-to-person relationships between the users. Social network sites are constructed in a way that both allows and requires people to indicate relationships with other members. These relationships can take many forms, such as groups, friends, or fans (Table 5). The name chosen in the interface for the relationship type does not necessarily reveal much about the true nature of the connection. For instance, they can be "contacts" (LinkedIn), "buddies" (MovieLens), "network" (del.icio.us), or "friends" (most of the other studied sites).

Relationship details	Friends	Reciprocal contacts	Groups	Public contacts	Fans or explicit recommendations	Neighbors
Amazon	X	X				
Del.icio.us	X			restricted	X	
Flickr			X			
Habbo			X	public		
Last.fm	X		X			X
LinkedIn		X	X	restricted	X	
MovieLens	X					X
MySpace	X		X	public	X	
Technorati					X	
Wikipedia			X			
YouTube			X			

Table 5. Friends and groups in the sites studied.

The ways of forming and maintaining relationships within the systems vary. The relationship does not require reciprocity in some of the services. It is also up to the site to protect privacy of the relationships. When the relationship is public (c.f. Table 5), the connection is shown to all. However, in some services it is up to the user to decide if the connections are public or not. Sometimes the visibility of the connections can be restricted (c.f. Table 5) to only the closest personal connections, but more commonly the setting is bimodal: either full visibility or no visibility at all. It depends on the site which setting is the default, full visibility or privacy.

In many sites that allow the users to articulate their social networks, the friendship links are reciprocal. This is accomplished through “friend” requests wherein one user asks another to accept the invitation. If the relationship is acknowledged, the users show up on each other’s friends list. For instance, in Facebook friendships are reciprocal (Lampe et al., 2007). Likewise, in Friendster one is asked to add another person as a friend. The answer is either yes or no. In practice, the everyday meaning of “friendship” is stretched in these systems. You do not need to know the person more than vaguely to accept the invitation to become a “friend” (boyd & Heer, 2006), if even that.

Friendship links are one way by which the users traverse through the network, using the links to travel from one profile to another (boyd, 2006). In some sites, the visibility of the details of friendship links can be restricted and contacts made only via the service. Of course, some people add links of their external home pages into their public profiles, making it much easier to get into contact.

Sometimes the relationships start offline and are then articulated, that is, defined as a connection between the persons and continued online. For instance, respondents to early surveys of Facebook members have indicated that they list mainly offline friends as friends in the service, and only rarely do they list people that they have met only online as friends (Lampe, Ellison, & Steinfield, 2006). However, even in Facebook this is changing. A friendship does not need to be tied to any offline social networks or individuals encountered

offline any longer. According to Lampe et al. (2007), now even fake profiles, such as the school mascots, are increasingly linked to as “friends”.

Such tight connections to the offline community shown in the early Facebook years are not that common any longer in social networking sites. However, sometimes the sites show their existence also in the physical world. For instance, Friendster created a buzz in the streets of San Francisco when it, as the first social networking service was launched to the public:

“Walking around San Francisco in the summer of 2003, it was impossible to ignore Friendster; the topic dominated bar and cafe culture and WiFi users would make a display out of surfing the site.” (boyd & Heer, 2006)

Since the profiles in Friendster were available to only those who had a Friendster identity, the information in the profiles was somewhat private anyway, despite the real world connections. The friend requests span strangers and long-forgotten acquaintances as well as people known from different social settings. Sometimes these friend requests pose social dilemmas: “Yet, how does one say no to a Friend request from one’s boss?” (boyd & Heer, 2006)

Boyd and Heer (2006) also report on the lack of social cues within the list of friends. Since all friends are equal, there is only one kind of friendship status, the list does not reveal any real world relationships between them. If the professor of a student is listed as “a friend”, for others viewing the student’s profile the professor is “just a friend”. MySpace allows listing “Top friends” separately from the rest of the “buddies”, but even this does not reveal the nature of the friendship.

LinkedIn has an in-built limitation that the users can only view profiles that are three degrees away, meaning that a user can maximally explore the profiles of their friends’ friends’ friends. Similar design was also used in Friendster. The limitation was designed to improve the level of trust within the system (boyd & Heer, 2006).

Your Network of Trusted Professionals

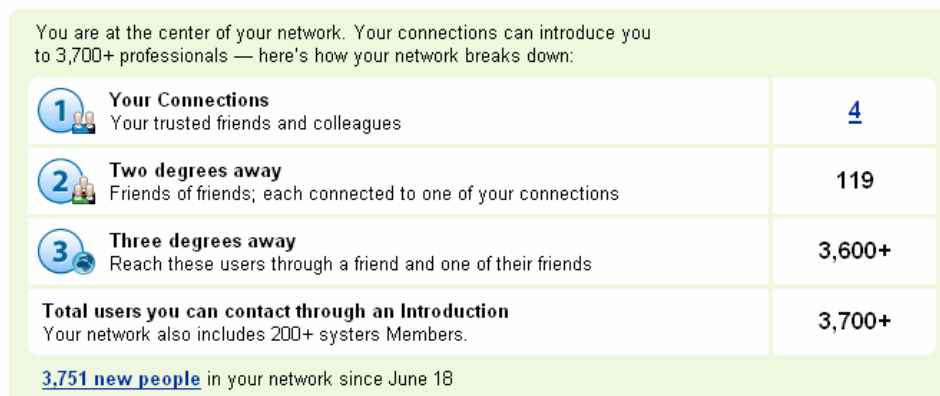


Figure 5. LinkedIn Network with 4 connections and network two or three degrees away.

In LinkedIn, the email addresses are shown only to the people directly connected to the user. The connections are reciprocal and by invitation only. The invitations will be sent by email if the receiving member has consented to that.

The network structure shown in Figure 5 is important when one wants to run a search within LinkedIn. All members in the LinkedIn community who allow public searches can be searched by their name, title, location and other information they have entered in their profiles. Consequently, it is possible to find potential new contacts. LinkedIn shows the full

names of people within the network in the search results. Getting into touch with the found users happens by asking somebody to introduce you (Figure 6) or by inviting the other users into your network with a message written and sent within the LinkedIn system. However, LinkedIn lists only profile data and no contact information for those who are not in the viewer's network, and the free basic user account owners cannot contact them.

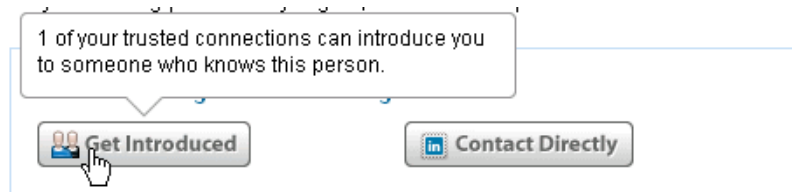


Figure 6. Asking to be introduced in LinkedIn.

Originally, MySpace also only showed the contacts of one's extended network and not all user profiles but this was soon changed since it slowed down the site remarkably to process each relationship every time a profile was viewed (Tom, 2007). In MySpace, forming a friendship relation with someone does not demand any acceptance of friend status. When a new user registers on MySpace, there is already one "friend" waiting: Tom (whose profile is displayed in Figure 1) is one of the support staff of the site. One does not need to do anything to acknowledge being friends with Tom. The number of members in MySpace is growing fast, as Tom's friend count shows (Figure 7).

Tom's Friend Space
Tom has 186053171 friends.

Figure 7. Part of Tom's MySpace profile showing his friend count (June 25, 2007).

It is not unusual to fake one's profile information (Kobsa, 2007). For instance, boyd and Heer (2006) reported that some of the profiles in Friendster were clearly faked ("Fakesters"). Surprisingly, other users showed appreciation for the creative talents and amusing ideas of the Fakester profiles. Fakesters gained large friend networks and increasing visibility, eventually becoming "social network hubs" with tens of thousands of friends. Later on, they were thrown out from Friendster, which caused many of them to move to MySpace where fake profiles, commercial profiles for example for products, and acting as Collector of friends are allowed (boyd, 2006).

In addition to the "friends" linkages, Friendster also includes testimonials, small messages sent to friends. If the friend accepts the testimonial, it is added to his or her user profile, thus supporting messaging between members. As a social norm, both the friend requests and testimonials are based on reciprocity, and so the other person is supposed to respond. (boyd & Heer, 2006)

According to boyd and Heer (2006), most social network sites allow Friends to leave messages on each other's profile, and these messages are visible to anyone who has access to the profile. In LinkedIn, the members can recommend others with small messages that are attached to the user's public profile. In MySpace, similar short messages, conversational in tone, are called Comments. When they are added to a user profile, those listed as "friends" can take part in the conversation (boyd, 2006).

Table 6 lists several incentives for being friends with both real world friends and previously unknown contacts in Friendster and MySpace (boyd, 2006). While blogs are persistent

writings attached to a profile, bulletins are small, non-persistent entries made by the member. Since they are shown only to friends, they are one of the reasons to list up as friends.

Getting into contact with people you know	<ol style="list-style-type: none"> 1. Actual friends 2. Acquaintances, family members, colleagues 3. It would be socially inappropriate to say no because you know them
Contacting others: both enabled by technical affordances and reflecting social factors within the community	<ol style="list-style-type: none"> 4. Having lots of Friends makes you look popular 5. It's a way of indicating that you are a fan (of that person, band, product, etc.) 6. Your list of Friends reveals who you are 7. Their Profile is cool so being Friends makes you look cool 8. Collecting Friends lets you see more people (Friendster) 9. It's the only way to see a private Profile (MySpace) 10. Being Friends lets you see someone's bulletins and their Friends-only blog posts (MySpace) 11. You want them to see your bulletins, private Profile, private blog (MySpace) 12. You can use your Friends list to find someone later
Being contacted	<ol style="list-style-type: none"> 13. It's easier to say yes than no

Table 6. Incentives mentioned for being friends, adapted from (boyd, 2006).

Lampe et al. (2007) studied Facebook profile entries when the service was still restricted to the students and faculty of colleges and universities. The number of items in the user profile correlated with the number of contacts users had in the social network. Those who had filled in all the profile fields had more contacts listed than those who had left the fields empty. The profile included, among others, fields for gender, status (undergraduate or graduate student, faculty), current and past schools attended, interests, and contact information on campus. While some fields supported maintaining the pre-existing social networks (for instance, high school attended), many fields helped in forming new relationships based on similar taste for music, literature or hobbies. Of all user types, undergraduate students had the most friends in their profiles. In general, the older the user account was, the more contacts were listed. (Lampe et al., 2007)

3.6 Privacy concerns

As seen, the fundamental functions in social networking sites are geared towards creating an online identity, managing contacts, sharing content, and staying aware of things happening in the community. The downside of all this, however, is that many things previously known only to your closest colleagues or friends are now openly available to others as well. Different social networking sites take the privacy concerns, such as the need for restricting the availability of content, identity, and other networking information, into consideration at varying degrees.

For instance, in Flickr the users have the option to control the visibility of all photos they have uploaded in one default setting or by changing the setting for each photo. A public photo is available to any visitor of the Flickr to search and view. This availability can be restricted in two ways, either by making the photo only available to the photo owner or by extending the availability to the designated "friends" or "family". Thus, Flickr has five privacy levels: private, family-only, friends-only, friends-and-family, and public. All the recent photos that

those listed as the user's contacts have uploaded and given the user permissions to see are shown in the user's Contacts page (Ahern et al., 2007).

The actual photo can be augmented with location information as well as with the names of the persons involved. For instance, in Facebook the user can upload whole photo collections for others to see, and mark the people photographed with links to their Facebook profiles.

Flickr and many other photo-sharing services support showing location data automatically if it is available in the image metadata. GPS-aware mobile phones with camera may add location data directly to the image metadata and some digital cameras also can also be set to add GSP information to it. This information is then made visible in the photo sharing services. As pointed out by Ahern et al. (2007), "some locations are more private than others". The identity of the person who is going to see the information affects the decision whether to disclose the location information. In some cases, the privacy settings need to be changed for each photo separately.

Perhaps because privacy has many meanings and various interpretations (Lederer, Hong, Dey, & Landay, 2004), the user concerns regarding online privacy in the social networking sites cover several distinct themes. Ahern et al. (2007) present a taxonomy of some of these privacy concerns (Table 7). The taxonomy is based on the viewpoints expressed by their study participants in interviews focusing on their photo-related privacy decisions. Each study participant brought up several of the concerns listed in Table 7. The first dimension of the taxonomy is the object of consideration, the photographer or others either appearing in the picture or otherwise related to the photo in some way, and the second dimension is the themes of concern that emerged in the interviews.

The first theme is online security, especially brought up by parents regarding their own or other children in the photos—you never know who is out there viewing the photo, and that is considered a good reason to make it private when uploading it to Flickr. The second theme is related to identity. Perhaps the photo is damaging to the online identity that the photographer or the persons in the picture want to maintain, and one does not always know if the people in the photo really want it to appear on the net or not. The third theme, social disclosure of the activity and whereabouts of the people in the photo to people they know was an immediate concern of many participants and a reason for making a photo private or restricting the access to it to a certain group of friends or family. The fourth theme is the convenience or ease of use by the photographer and the other people interested in the photo. For instance, making the photo non-public means that to be able to view it, one would need to be registered as a user and part of the group of friends for whom the photo is open in addition to logging in to view it. If, on the other hand, the photo is made public, no such extra operations are necessary to view the picture. Thus, for the study participants, competing considerations sometimes generated conflicts that needed to be resolved case by case. (Ahern et al., 2007)

		Theme			
		Security	Identity	Social Disclosure	Convenience
Object	Self	Exposing self to security hazards	Managing own online identity	Exposing socially sensitive information to contacts	Difficulty of sharing and viewing
	Other	Exposing other to security hazards	Influencing other's online identity	Exposing other's socially sensitive information	Other's ability to share and view

Table 7. Privacy considerations condensed into a taxonomy (Ahern et al., 2007).

In the following, several concerns introduced in the taxonomy will be discussed with examples from other fields than photo sharing.

Often the TCP/IP net address one uses to connect to the Internet is not considered personal information (Privacy International, 2007). However, it is information traceable to the user, as explained in the privacy policy of Wikipedia (Wikimedia Foundation, 2006)

“If you have not logged in, you will be identified by your network IP address. This is a series of four numbers which identifies the Internet address from which you are contacting the wiki. Depending on your connection, this number may be traceable only to a large Internet service provider, or specifically to your school, place of business, or home. It may be possible that the origin of this IP address could be used in conjunction with any interests you express implicitly or explicitly by editing articles to identify you even by private individuals.

It may be either difficult or easy for a motivated individual to connect your network IP address with your real-life identity.”

When the user has registered in Wikipedia and acquired a user pseudonym, the TCP/IP net address is not revealed any longer to others except the administrators (Wikimedia Foundation, 2006). However, when the users have acquired user IDs, the login data is often stored in cookies so that they do not need to authenticate themselves each time to gain access to the site from the same computer. In some cases, the site does not even work if cookies are disabled (Privacy International, 2007).

Another concern with login addresses is that many sites collect clickstream data based on the TCP/IP net address also from the visitors who do not log into the site. The sites do not always explain clearly what is done with the clickstream data they collect. Sometimes it is even shared with third party companies. (Privacy International, 2007)

User profile data is often public to any user viewing the community pages, sometimes even without him or her having to log into the site. The profile data may reveal potentially unsafe information of one’s identity, behavior and characteristics, friends and family, location, and hobbies.

The study by Ahern et al. (2007) on Flickr (c.f. Table 7) indicated that people are concerned about privacy issues when they are interviewed about them but that they do not always pay attention to them when using the web. For example, while entering their ZIP code level location information raised concerns in the study participants when interviewed, in practice none of them had configured their location settings to conceal the location information (Ahern et al., 2007). This is consistent with the results of numerous other studies that Kobsa (2007) summed up, indicating effectively that there is a significant say-do issue between people’s views on privacy and their actual actions.

Other research studies confirm these findings. In practice, users seem not overly concerned about their privacy but offer information generously and seldom change their default privacy settings (Gross et al., 2005). When the default settings do not match the current requirement for privacy, it goes easily unnoticed and can result in privacy-related information leaking to others. One possible solution might be to show the user a preview of the current settings in effect, and let the user decide if those settings are satisfactory.

However, dealing with the privacy settings in real life is not easy. For instance, when a user uploads photos to Flickr, their privacy settings should be decided immediately based on limited knowledge of the people potentially wishing to see the photo and uncertainty about the preferences of the people shown in the photo (Ahern et al., 2007). Also, our experience of

using the services shows that the place of the privacy settings in the interface is not always readily available nor are their interfaces self-explanatory. Sometimes, as in Amazon, the privacy settings are distributed in the interface, requiring the user to set privacy separately for various items. We found Flickr's approach of being able to set defaults and then changing them item-by-item base rather good. However, as all content shown openly increases the site's attraction, the sites have an interest in keeping things public. Flickr without public pictures would not be what it is today, as a rough example.

Moreover, with so many Web 2.0 services around requiring registration, aggregation of information across various services has potential privacy implications. It is possible to collect information about a user in one service and link that with the information provided in another service, over time in one service or across services. Such aggregation might enable building an increasingly complete profile or even reveal the user's real-life identity. While the information bits generated in a short period of time and within one web site might be harmless, the information bits collected in over time and encompassing several source sites might reveal too much to those interested. Today there are already such services as Wink (<http://wink.com/>) that enable searching for people simultaneously from multiple Web 2.0 services based on the public profiles built by the users in these services.

As the data on the Internet has become increasingly machine-readable for web crawlers and other automatic tools, it has become possible to build rather complete user profiles with the information available on the Internet. The more information the user reveals, the more complete the profile. The profile information can be misused to send tailored phishing messages or perform other type of security attacks.

Recently, researchers at Indiana University (Indiana University, 2007) conducted an experiment on social phishing on the campus. One group of students received e-mail messages from senders they thought to be friends from a social networking site they used while the other group received an e-mail from a stranger. The e-mail message asked the students to visit an external Web site and enter their university ID and password to log in. Sixteen percent of those approached by strangers visited the site and entered their ID and password information. A much larger percentage of participants (72%) were willing to visit the site and enter their confidential login information when approached by "friends" on a social networking site.

While some social networking sites encourage revealing personal data, in some other services, such as Habbo hotel, the personal Home page of a Habbo can contain no information on who the Habbo is in real life. Such information is banned not only in the Home page but also in the chats that take place in the site to protect the identity of the users. As Habbo caters mainly to young teenagers, it is essential that the potentially harmful contacts a Habbo gets in the virtual hotel will not approach or harass the child in real life.

In most other cases, however, the social networking sites are open to all, including even those with criminal records. Several cases have already been reported where the criminals have made contact with their to-be victims through MySpace (Jones, 2007) or FaceBook (Wikipedia, 2007n). In May 2007, MySpace was approached by eight attorney generals with a letter expressing their worries about criminals hiding behind pseudonyms in MySpace and requested that sex-offender information be crosschecked by the site owners (Jones, 2007). In response to a subpoena that it said it needed, MySpace removed the user profiles of 29 000 sexual offenders in July, 2007 in the U.S.A. (Richards, 2007).

However, Richards (2007) points out that the biggest danger to young people using the Internet lie in the information they reveal there about themselves. It is advised, for instance,

to only post pictures they would be happy their parents to see and to avoid giving facts in chat rooms and instant messaging conversations to people they do not know. (Richards, 2007)

Since most of the sites in our survey are run by US-based companies, they do not follow the same privacy policies as European companies. The European Commission's Directive on Data Protection went into effect in October, 1998. In order to bridge these different privacy approaches and provide a streamlined means for US organizations to comply with the Directive, the U.S. Department of Commerce developed the "Safe Harbor" framework. The organizations need to comply with the seven requirements of Safe Harbor—and publicly declare that they do so—before they can join the Safe Harbor. Joining is, however, voluntary. (International Trade Administration, 2007)

Table 8 introduces some privacy features of the sites we studied. Some of the sites are part of Yahoo or Google that have joined Safe Harbor. All the sites studied have a Privacy Policy. On the other hand, only some of them have a separate page for Safety tips. "Profile preview" in Table 8 means that the user profile can be viewed by its owner as it would be seen by other users. Profile preview helps users to become aware of the privacy implications and understand how the interface allows certain information fields to be hidden. While systems should not require excessive configuration to create and maintain privacy, the settings should not be buried deep into the interface, thus making them hard to use (Lederer et al., 2004).

Privacy features	Privacy policy	Safety tips	Profile preview	Signed to Safe Harbor
Amazon	X			X
Del.icio.us	X			(Through Yahoo)
Flickr	X			(Through Yahoo)
Habbo	X			(Not applicable)
Last.fm	X			
LinkedIn	X		X	X
MovieLens	X			
MySpace	X	X		
Technorati	X			
Wikipedia	X			
YouTube	X	X		(Through Google)

Table 8. Some privacy features related to the sites studied (Spring 2007).

There is a great need for studying privacy in the social networking sites. According to a recent study of privacy in twenty Web 2.0 sites conducted by UK-based Privacy International, Amazon, Friendster, LinkedIn, and MySpace "were generally privacy aware but demonstrated some notable lapses". Both Last.fm and Wikipedia were rated as "generally privacy aware," but YouTube was deemed to have "serious lapses in privacy protection," mainly because of not considering the video content personal information and giving out vague information in its policy statement about sharing personal information to affiliated

companies. None of the sites studied received the highest rating in the Privacy International's interim report of June 9, 2007. The interim report is to be replaced by an updated version in September 2007 after the sites have been contacted and given a chance to react to the problem areas. (Privacy International, 2007)

Overall, the advent of social networking sites together with user contributed content clearly call for better understanding of personal privacy since the sites have privacy-affecting content that can even lead to identity theft and other criminal activities in the wrong hands. We get back to some of these issues in Chapter 4.

4 Tools for harnessing collective intelligence

As discussed, one of the defining characteristics of Web 2.0 is that it harnesses the collective intelligence of the users in various ways. The applications become more useful and smarter as the number of people using them increases (Musser et al., 2006). Collective intelligence has effectively become an integral part of the services without which they could not function. For instance, how to create a folksonomy (Section 4.4) from tags if nobody tags? With massive numbers of people tagging photos, however, Flickr's tagging-based approach to categorizing photos and generating navigation has become a success story.

One important part of being able to use collective intelligence is to make it easy for people to "do their share." Like with user-generated content, it needs to be easy for people to contribute. This has to do with both motivation and ability. Not surprisingly, it is far easier to motivate people to do easy things than difficult things. In addition, people should not need any new skills to contribute. Otherwise, getting the contributions will be challenging, to say the least.

Collecting user preferences that form the basis of collective intelligence information can take place in two ways: explicitly and implicitly (Svensson, Höök, & Cöster, 2005). Explicit collecting means that the users have to provide the information actively. For instance, in MovieLens we first have to rate movies before the system can give us recommendations. Our ratings, however, simultaneously represent the collective intelligence of the system as they are also used to make recommendations for others. The motivation to contribute here is to improve the recommendations we receive. In the same way, Amazon asks us to rate books when we first sign up for an account to be able to give us personalized recommendations. Both likes and dislikes can be collected this way. For instance, in Last.fm we can mark songs with "Don't ever play me this track again" or with "Express your love for this track."

Implicit collecting means that we do not have to do anything other than use the service to provide the information needed for generating collective intelligence. The system tracks our actions to detect our likes. For instance, Amazon tracks which books we buy and recommends books based on this to us and others in such features as "Customers who bought this item also bought." In the same way, what we listen in Last.fm affects what we and others who in some way share our taste are recommended. Implicit information collection is easier for likes than dislikes. For instance, in the recipe collection of Kalas, printing or saving a recipe was interpreted as an implicit vote for liking that recipe (Svensson et al., 2005), though the actual food tasting is to come much later. Dislike information can be deduced, say, if we start to listen to a song and after a few seconds click the "Next" button. Implicit feedback, however, is not as easy to interpret reliably as an explicit expression of dislike.

However, just as people are wary of sharing personal data, the implicit collecting of information also raises privacy issues. Kobsa (2007) reports that vast majority of the Internet users feel uncomfortable being tracked across different sites and clearly over half of the users are concerned about being tracked at all. The majority of the users are also worried that the

sites might use information about them for undisclosed purposes, such as sharing it with third parties. However, the benefits of personalization and other benefits that the users get appear to figure significantly in the cost-benefit analysis, resulting in people trading this kind of data for the benefits even if uncomfortable with the privacy implication. Still, the lack of sufficient information about data collecting and handling results in uneducated privacy decisions. (Kobsa, 2007)

Most socially oriented web services today use both explicit and implicit approaches to profile the users for personalizing the service for individual users. Many recommender systems are based on collaborative filtering or social filtering (Heylighen, 2001), as it is sometimes called.

Before we discuss the typical Web 2.0 features that allow us to use collective intelligence, we discuss briefly social navigation that is both the result and the goal of many features that we discuss in this paper.

4.1 Social navigation

As the number of items, be it books in Amazon, photos in Flickr, songs in Last.fm, blogs in Technorati, or the 11 million registered users in LinkedIn (LinkedIn.com, 2007), is such that nobody can even glance at all of them, we need ways to navigate to the items of interest. Social navigation means using what other users have done or are doing in the service, that is, in the community, for navigating or assisting in the navigation in the service (Dieberger, Dourish, Höök, Resnick, & Wexelblat, 2000).

Heylighen (2001) compares the process of collaborative filtering to that of word-of-mouth. When we need to choose between options of which we have no prior experience, we seek advice from others who have the experience. In the era of the Internet, we often have millions of options. For instance, a book search in Amazon without a keyword returns more than 9.5 million books. Even by dividing the mass into sub-groups with keyword searches and such, the possibility of a user going through all the possibilities one by one becomes impossible for all practical purposes (Heylighen, 2001). Furthermore, how could the shop recommend books for the user? Showing, for instance, the best selling 100 books would lead inevitably to showing a large number of books that do not interest that particular individual at all.

While Dieberger et al. (Dieberger et al., 2000) discuss mainly using the whole community's information for individual users, today collaborative filtering increasingly allows us to match individual users to such sub-groups in the community that share more in common with the individual users than the community-at-large does, thus allowing the system to make better predictions for the individual user.

In today's social navigation, we show the user where the community-at-large, matching sub-group of the community, or a certain individual (matching by criteria or marked as friend etc.) is or has been, is currently interested in, or has been interested in. For instance, Amazon's "What Other Customers Are Looking At Right Now" and its bestselling lists, Technorati's "Where's the Fire? What's Hot, and Why", and YouTube's "Videos being watched right now" show the user what the community is interested in right now. On the other hand, Amazon's "Customers who viewed this item also viewed" and Last.fm's selections for the songs that are played in "<artist's name>'s Similar Artists" are based on the tastes of a sub-group that matches the user in certain ways. These types of recommendations are indirect, that is, they offer no direct contact between users, and are mostly based on anonymity of the users whose usage data has been aggregated (Dieberger et al., 2000).

Typically, the features based on one user are not anonymous. For instance, Last.fm's "Weekly Top Listeners for this artist" with lists of what they have been listening and Amazon's "Listmania!" lists help other users find items based on individual users. In addition, some services offer ways to find out about the people on whom the aggregated recommendation is based. For instance, Amazon reviews are linked to the profiles and other reviews made by the reviewer.

Sometimes social navigation can be based on direct contact between the users. Messaging, chatting, and other such features offer direct, sometimes even real-time contact between users. Real-time chats are possible only when the system indicates real-time presence and users are logged in at the same time (Svensson et al., 2005). Though messaging, chatting, and other such features offer direct contact between users and can be used for finding items, they primarily have more of a social rather than navigational function.

Direct contact includes also asking experts or other users. While many sites host forums that are dedicated for discussion only, even discussion forums attached directly to a product can be used for enquiries somewhat related to the product in question, as in Figure 8. People are creative and social by nature, and different features are often used socially independent of their original purpose.

The screenshot shows a forum thread on Amazon. At the top, it says "Customer Discussions > A Thousand Splendid Suns forum" and "please help i need recommendations". A navigation bar indicates "Showing 1-9 of 9 posts in this discussion" and "Sort: Oldest first | Newest first".

The first post is by H. Khan, dated Jun 1, 2007 4:10 PM PDT. The text of the post is: "hey i read kite runner and splendid suns both and absolutely LOVED them the problem is everytime i finish his books. i look for otehr books to read but whenever i find one it just doesnt compare? does anyone have any sggestions as to what i can read next?". Below the post are links for "Reply to this post", "Permalink", and "Report abuse", along with a poll asking "Do you think this post adds to the discussion?" with "Yes" and "No" buttons.

The second post is by Alyson Delvecchio, dated Jun 4, 2007 4:50 AM PDT. The text is: "Last year one of my absolute favorite books was Kite Runner. My other hands-down favorite was Shadow of the Wind. Have you".

On the right side, there is a summary box for "This discussion" for "A Thousand Splendid Suns forum". It lists "Participants: 9", "Total posts: 9", "Initial post: 12 days ago", and "Latest post: 1 day ago". Below this is a "Track this discussion" button and an "RSS Feed" link. At the bottom of the summary box, it asks "Is this discussion helpful?" with "Yes" and "No" buttons and a "(Report abuse)" link.

Figure 8. Discussion forum in Amazon used for social navigation.

Many of the features typical to Web 2.0 services are, in fact, one way or another related to social navigation. Using collective intelligence in information environments with vast collections of items, especially when discussing implicitly collected collective intelligence information, typically suits to solving navigation and item-finding challenges. Tagging and the resulting folksonomies (Section 4.4), recommendations, and interest-matching of various kinds all create links for us to follow based on a community or its subset. Some are matched specifically to us through collaborative filtering (Section 4.3) while some others are simply based on popularity and community-level ratings and rankings (Section 4.3.3) without being personalized for us, but both help us navigate to items of interest.

The big design question is how to make useful social navigation tools that allow themselves to be molded into true needs of the community. This is analogous to how to design pathways in a park so that people do not end up walking over the grass, developing new paths that serve their true needs. We need to design tools that allow people to develop the paths they need.

4.2 Personalization

Personalization or customization of services is also another typical feature of Web 2.0. When we come to Amazon.com, we are recommended books based on what we have bought and viewed previously—unless we have explicitly signed out after the preceding session (which, in case of Amazon, is made difficult on purpose by not providing a clear sign out button). Likewise, Last.fm gives you recommendations based on what you have listened to if you log in, and if you listen to one artist, you are given a list of other, similar artists to check out.

Personalization is based on both explicitly given and implicitly collected information about the user. The raw data is often analyzed through collaborative filtering to bring in the collective intelligence. In Amazon, you can evaluate books explicitly to improve the recommendations you get. However, at the same time Amazon collects implicitly information about what you view and buy. It is the same with Last.fm. If you download Last.fm's software, it creates a record of what you listen with your music players (Winamp, Windows Media Player etc.) in a process called "scrobbling" and thus profiles your music taste. However, you can also "Express your love" for a track you are listening to or block it from ever being played again. This way, Last.fm gets explicit information about your likes and dislikes. The result is better book and music recommendations that match your taste. Other Web 2.0 services have similar features to tailor the service to you.

An example of personalized content not put through collaborative filtering is personal history information. For instance, Amazon's "Your Recently Viewed Items" shows your personal viewing history that is not shown the same way to anybody else.

Another way many services, such as Last.fm, YouTube, and Technorati, allow you to explicitly construct your personalized experience is by allowing you to mark items (songs, videos and blogs respectively) as favorites, thus creating a shortcut list to them. Figure 2 on page 14 shows an example of favorites in Technorati.

In addition, many services offer different types of reminder systems and calendars to the users. For instance, Amazon has a feature that allows you to enter what you have given to whom as a present in addition to providing you with a calendar that will remind you of important occasions of which you wish to be notified, and Last.fm offers personalized concert recommendations at your location. Behind these features we can also see the earning mechanisms raising their heads. Naturally, at least the services wish to see it as a win-win situation where the users get value as well.

The beauty of this apparent win-win situation is also marred by privacy concerns (Kobsa, 2007). Generating personalized content and experience entails collecting vast amounts of privacy-sensitive data. As discussed earlier, while large numbers of users are concerned with the privacy questions inherent to this constant profiling, in practice users do give privacy-sensitive data rather carelessly (Kobsa, 2007; Ahern et al., 2007). While this is good news to the services as they need the data, it also means that the privacy issues inherent to Web 2.0 social services need to be studied further. The repercussions of today's actions will only be heard tomorrow, and so the impact of today's laxness about privacy-sensitive data might come back to haunt us later on.

Much of personalization involves personal history information, both explicitly given and implicitly collected, often put through collaborative filtering (Section 4.3) to generate recommendations and navigation for the user in question. Often, recommendations and navigation go hand in hand. For instance, if we are offered links to bands the system determines that we are likely to like, it is both a recommendation and means for social navigation. Personalization also involves providing tools for explicitly marking items as

favorites or for making reminders for ourselves, thus allowing us explicitly shape our experience and the interface options given to us. Variations are endless and in most cases, different sources of information are combined to generate the end result that we see. However, true to the Web 2.0 concept, all of our interactions with the service are typically used to shape our and other users' use experience in it.

4.3 Recommendation systems

Various recommendation systems represent a vast collection of different approaches for making collective intelligence useful. The amount of information and items in many services is staggering, and we need ways to find the instances that interest us (McNee et al., 2002). For instance, if we search for books in Amazon.com without a keyword, we get 9 595 516 hits (June 13, 2007). In Flickr, there are "millions of users, and hundreds of millions of photos" (Flickr.com, 2007b). Collective intelligence can be harnessed to provide recommendations based on the explicitly given and implicitly collected information about the individual user and the whole user community by matching users based on their likes and dislikes.

Recommendation systems can narrowly be understood simply to mean the algorithms that match individuals with individuals or individuals with sub-groups of the community and produce recommendations based on that. However, recommendations can be made in various other ways as well. They can be direct recommendations from one user to another user(s), such as Last.fm's "Recommend this track to your friends". They can be thumbs-up-thumbs-down kind of comments, such as Amazon's "Was this review helpful to you? Yes/No." The thumbs-up-thumbs-down ratings are counted as votes to organize book reviews in the "Most helpful customer reviews" section. Furthermore, recommendations can be made based on the popularity of the item, such as Amazon's "Bestsellers" or Last.fm's "Weekly Charts." Consequently, there are innumerable approaches to recommendation systems that all ultimately work to help us locate items of interest based on our and other users' likes and dislikes.

Trust is an interesting question in relation to the recommendations. Recommendation systems are open to financial, political and other kinds of influences (O'Donovan & Smyth, 2006). Herlocker et al. (2004) report that, unsurprisingly, recommenders are not trusted implicitly. Some people play around with and check the recommender systems to see if they produce good hits or not. On the other hand, there are also people who do not always stop to consider how and why the recommendations are brought to them (Svensson et al., 2005).

While recommendations are typically presented in the interface, they can be delivered to the user in other ways as well. Amazon emails recommendations to the registered users and many sites offer various types of web feeds, such as RSS and Atom, to the users. This way the recommendations reach the users even if they do not visit the site itself for a while. This type of recommendations also works as reminders to visit the site.

The recommendations can be given implicitly to the users as part of the interface without the users explicitly asking for them. The recommendation can appear in the ordering of the items based on their popularity. It is not always even recognized as a recommendation. The other extreme is that the users explicitly ask for recommendations from experts or other users.

In this paper, we divide recommendation systems into three broad categories: algorithm-based recommenders, human-to-human recommendations, and recommendations based on popularity and ranking. These categories do overlap as practically all recommendations are based on user actions and practically all are implemented with some kind of algorithm. When considering the algorithm-based recommenders, the most successful ones are based on

collaborative filtering (Herlocker et al., 2004). Thus, we first discuss collaborative filtering and then look at the other two approaches to making recommendations.

4.3.1 Algorithm-based recommenders: Collaborative filtering

Collaborative filtering is an umbrella term for various algorithms that generate automatic predictions or recommendations for an individual. For instance, MovieLens can recommend movies (Figure 9) as a result of explicit user ratings of a subset of the movies and collaborative filtering. The predictions become better the more the system knows about the user's opinions on the movies seen.

The screenshot shows a page titled "Page 1 of 597 | Go to page: 1...119...238...357...476...595...last" with a "page 2 >" link. The page is divided into four columns: "Predictions for you", "Your Ratings", "Movie Information", and "Wish List".

Predictions for you	Your Ratings	Movie Information	Wish List
★★★★★	Not seen	Killer (a.k.a. Bulletproof Heart) (1994) VHS info imdb add tag Drama, Thriller	<input type="checkbox"/>
★★★★★	Not seen	Young and the Passionate, The (I Vitelloni) (1953) DVD info imdb Drama - Italian [add tag] Popular tags: Federico Fellini Italian	<input type="checkbox"/>
★★★★★	Not seen	Red Circle, The (Le Cercle Rouge) (1970) DVD info imdb Crime, Thriller - French [add tag] Popular tags: atmospheric tense visceral	<input type="checkbox"/>
★★★★★	Not seen	Stray Dog (Nora inu) (1949) DVD VHS info imdb Drama, Film-Noir, Thriller - Japanese [add tag] Popular tags: black market baseball my dvds	<input type="checkbox"/>

Figure 9. Examples of movie recommendations in MovieLens.

Traditionally, a collaborative filtering recommender algorithm finds a group of individuals based on their similarities of preferences, ratings, or purchases. It then combines them into a sub-group of neighboring users, and uses the sub-group's or nearest neighbors' common preferences for generating predictions for an individual (Herlocker et al., 2004; Heylighen, 2001; Linden, Smith, & York, 2003; McNee et al., 2002; Sarwar, Karypis, Konstan, & Riedl, 2001; Wikipedia, 2007d). The preferences can be distilled from both dislikes and likes.

However, using collaborative filtering algorithms this way is computationally expensive, and the algorithm gets even slower when there are millions of users and millions of items to compare (Linden et al., 2003). Consequently, Amazon uses item-to-item collaborative filtering for generating such recommendations as "Customers who bought this item also bought". This type of item-centric collaborative filtering is based on building an item-to-item matrix of items that customers bought together. It finds items similar to each of the user's purchases and ratings, aggregates those items, and then combines the most popular ones into a recommendation list for the customer. This approach is computationally cheaper than comparing customer profiles since it can be run off-line on the product catalog items. Its time requirement depends only on the number of items the user has purchased or rated. (Linden et al., 2003)

Thus, collaborative filtering can be human-centric or item-centric or combinations of both. In addition to the computational requirements, the major challenge for the algorithms is naturally to produce good enough predictions (Herlocker et al., 2004; Linden et al., 2003; Sarwar et al., 2001).

In Wikipedia (Wikipedia, 2007d) collaborative filtering is divided into three categories, active, passive, and item-based filtering. Active filtering is based on a peer-to-peer approach where the peers with similar interests explicitly share information with each other in the

system. The system requires active participation from the users. On the other hand, item-based filtering is a method of consumer research: participants rate the products with a given rating scale, and the ratings are used to group the items for consumers to compare them (Wikipedia, 2007d). For instance, a shop might show SLR-cameras from different manufacturers in the price category of \$600-\$1000 with a cell of more than 6 mega pixels together with their ratings.

It is passive filtering, however, that is seen as the great promise for future (Wikipedia, 2007d). In passive filtering, the user preferences are recorded implicitly during the normal use (Heylighen, 2001; Wikipedia, 2007d). The user preferences arise from such actions as buying, printing, or saving an item, referring or linking to a site, or formulating a query content. Thus, while active filtering requires explicit actions from the user, such as indicating a value on a scale, implicit gathering of data only requires that the users use the service, thus reducing the burden on the user (Heylighen, 2001; Sarwar et al., 2001; Wikipedia, 2007d). Implicit collection of preferences eases the critical mass requirement for the system to be able to function, as it is not motivating for the early users to enter explicitly their ratings when the system produces poor recommendations (Heylighen, 2001). Furthermore, people do not have to do anything up front to see the system functioning. Different approaches are being employed, however. While for instance Amazon uses a lot of implicit preference gathering, MovieLens requires the users to explicitly rate movies. One method does not fit all.

The reason for various collaborative filtering approaches to exist is that no one approach can handle all the situations. The user goal, data set properties, and number of users in relation to the number of items, among several other factors, determine which approach works the best (Herlocker et al., 2004).

Perhaps we have a user, who wants to find a video to rent. In this case, the recommender works fine as long as it returns movies that the user likes. There is no need to list all the good items. This situation contrasts with the needs of a lawyer, who needs to find all the precedents to prepare his or her case reliably. The goal determines the desired characteristics of the collaborative filtering approach, as well as the important quality attributes against which the recommendations should be evaluated. (Herlocker et al., 2004)

In many ways, we agree with the argument that user satisfaction should be the measure of the recommendation system's goodness. While accuracy and relative error-freeness are certainly primary concerns in evaluating a collaborative filtering algorithm, we have to go beyond them to increase user satisfaction. For instance, it is useless to recommend a user a book that he or she has purchased from our store earlier on.

Herlocker et al. (2004) discuss two important non-accuracy metrics for recommendations generated by collaborative filtering, novelty and serendipity. Novelty and serendipity both describe the non-obviousness of the recommendation. If, for instance, we are searching for a video to rent, and the system only recommends movies directed by directors that we have indicated that we like, the recommendation is likely to be accurate. If we were not aware of a certain movie beforehand, the recommendation would have been novel, but even in that case, we would likely have found that movie anyway based on its director. However, if the system, based on our liking of a director, finds us movies by other directors that we like but were not aware of, the recommendation is serendipitous. Thus, a serendipitous recommendation is by definition novel, but a novel recommendation might or might not be serendipitous. If the recommendations are not at least novel, the system is not useful for the users although it might be very accurate.

The value of the recommendations depends on the user's familiarity with the area of application as well as on his or her experience level with the system. For a novice user, it

might be important to suggest things that the user already knows and would select based on earlier experience even without the recommendation. Although not giving novel recommendations, it builds trust in the system's abilities (Herlocker et al., 2004). In contrast, a user who has experience with the system prefers to get novel and serendipitous items and gets irritated if he or she has to go through a long list of non-novel items to get to the items that interest him or her.

4.3.2 Human-to-human recommendations

Many Web 2.0 services offer different ways for humans to make recommendations to other humans. Perhaps the most classical is the "Email this article to a friend" recommendation. Such features as "Recommend this track to your friends" in Last.fm, possibility of sharing bookmarks in Del.icio.us, and "Tell a friend" in Amazon are all examples of this. At this level, the user typically knows the person(s) to whom the recommendation is made. Using these features might require you first to create a social network for instance by naming friends, family or joining groups in the service. These types of features are close to what marketers call viral marketing.

A step further is the recommendation systems where recommendations are made to the community-at-large. Here the goal is to help everyone in the community. Amazon's "Listmania!" allows users to make lists of books for a certain topic. "So You'd Like To..." lists work the same way. In them, individual users list books for doing something, such as taking outdoor photographs.

In Last.fm, all registered users automatically have their own radio stations based on the music they listen to. Others can also listen to the station. In Last.fm's case, the station is automatically generated based on the user actions, that is, what the user listens to. The user is present as the username whose favorite music is played on that radio station. Thus, the feature represents implicitly collected data but it is still a human-to-human recommendation.

Recommendations can also come in written format, such as reviews and comments. We return to these in Section 5.4.

4.3.3 Popularity-based recommendations, rankings, and ratings

Popularity-based recommendations are in fact community-at-large recommendations to individual users but given without collaborative filtering. They are not especially made for any particular user but all get the same recommendation. They are in one way or another based on what is popular among the users of the service. The following are examples of popularity-based recommendations in the eleven services studied.

- Amazon "Bestsellers"
- Last.fm "Weekly Charts" and "Visitors recommendations"
- Technorati "Top favorite blogs" (blogs that the most people have marked as favorite) and "Top searches"
- Flickr "Interesting photos from the last 7 days" (interestingness is a concept that is algorithmically calculated) and "All time most popular tags" which is entirely based on how often the tag is used by the users.
- "Today's popular items" in Del.icio.us.

As is evident from these examples, popularity is based on some kind of ranking. It can be an explicit ranking or rating action, such as the users marking the item as a favorite or voting for

an item, or collected by tracking user actions, such as most bought books or most listened to songs. Finally, it can also be calculated from a mixture of user actions. Flickr's interestingness, for instance, is based on such factors as number of viewings, comments, tags, number of mentionings in the discussion groups and so on (Flickr.com, 2007a).

Ranking refers to “the process of positioning items such as individuals, groups or businesses on an ordinal scale in relation to others” (Wikipedia, 2007i). The items in a collection are evaluated based on some principle so that any two items can be compared to see which ones should be in the higher position (Wikipedia, 2007i).

The possible ranking principles are endless. We can rank items by sales (for instance, “Bestselling” lists in Amazon), by views (as “Most viewed” in YouTube), by number of discussions related to the item (as “Most discussed” in YouTube), by favorite markings (as “Top favorited blogs” in Technorati), and by the number of people who have added the link (as in Del.icio.us). Habbo lists the most popular rooms based on the number of visitors. The variations are endless but the central principle is to count something and see which item has the most, which the second most, etc., and show the resulting ordered list.

As with many other features, ranking information can be collected explicitly or implicitly. View-information, for instance, is collected implicitly, while marking an item as favorite requires explicit action from the users. However, many explicit actions carry benefits for the user. Marking a blog in Technorati or a video clip in YouTube as favorite allows you to have it on your list of favorites and thus access it easily. This way, it is easy to motivate the users to take action as both they and the whole community profit from it.

In many ways, different rankings work to show what is going on in the community and what is popular. Del.icio.us's “hotlist – what's hot right now on Del.icio.us” tells us about what the community at large is interested in. It is the same with Habbo showing the most popular rooms as that is literally where the action is. “Most Popular Furni” is telling what is popular with the users based on their explicit investment actions.

Rating refers to an “evaluation or assessment of something” in terms of quality, quantity, or some combination of them (Wikipedia, 2007i). Again, there are endless variations on the theme. In Amazon's product reviews, the reviewers rate the products with 1–5 star scale and then the readers of the reviews rate the reviews as useful or not. In Last.fm, the listeners rate a song with “Express your love for this track” and “Don't ever play this track again” buttons. In Digg.com, you either “digg” a link or “bury” it (thumb down). In YouTube (Figure 10), a video is rated with 1–5 stars, and the system shows the number of raters next to the current rating.

Ratings are typically done by individual users although the aggregation shown in the interface is naturally processed information. Many rating systems require the user to first sign in before they can rate an item. This not only to get the user registered—although that certainly does play a role in the equation—but also to stop people from voting several times for their favorite or even own item, be it web site, book, or photo. Competition is hard, and unethical means are by no means unheard of in the race for visitors.

"I Got a Crush...On Obama" By Obama Girl

Try out the **NEW (beta)** version of this page!

Added: **June 14, 2007**
 From: [heartbreakkid9](#)
 Provided By: [heartbreakkid9](#)
 Category: [Comedy](#)
 Tags: [crush](#) [on](#) [barack](#) [obama](#) [girl](#) [campaign](#) [song](#) [hillary](#) [2008](#) [election](#) [amy](#) [hot](#) [\(less\)](#)
 URL: <http://www.youtube.com/watch?v=Ydkk4d1x8>
 Embed: `<object width="425" height="350"><param nam`

Rate this video:
 ★★★★★ 38 ratings
 Save to Favorites | Share Video | Flag as Inappropriate
 Add to Groups | Post Video | Inappropriate

Views: **11,442** | Comments: **59** | Favorited: **32** times

Showing 1–20 of 24 [See All Videos](#)

"I Got a Crush...On Obama"
 By Obama Girl
 03:18
 From: [barelypolitical](#)
 Views: 1322112

Crush on Obama
 03:18
 From: [OTBMedia](#)
 Views: 30409

Figure 10. An example of a YouTube page with ratings and tags.

As is evident from the examples, ratings are often used for ranking items. For instance, in Amazon the number of “useful” votes in relation to “not useful” votes for a book review determines how high that review is displayed in “Most Helpful Customer Reviews”. The good side of using explicit ratings by the users is that it avoids the risk of wrong interpretation. For instance, we do not know if a person who viewed a video in YouTube liked it or not. With ratings, we know if—and sometimes how much—the user liked the item. However, ratings require explicit action by the user and the benefits are not always obvious for the user. Consequently, many sites advertise that by rating items the user gets recommendations that are more accurate as compensation. This is in keeping with Kobse’s (2007) recommendation that users need to be made aware of the benefits of providing information to encourage them provide data.

Naturally, the users use ratings also for selecting items for closer look, buying, listening, etc. The five-star scale, familiar from hotels, gives us a clear impression of quality or lack thereof. Once again, they are part of the user-generated information that guides our actions in the services.

With ratings, we again confront the question of trust. How many people are behind the rating? Who are they? Many services allow the users to find out the number and community identity—although often not the true identity—of the users who have rated an item. For instance, Amazon shows the number of reviewers on whose reviews the star rating shown on the item list page is based if you are signed in and the community identities of the reviewers on the item page. From the item page, you can go to reviewers profile or see other reviews the reviewer has written to get a clearer image of the reviewer. In Del.icio.us, you see the number of all the users who have bookmarked the link and you can get a list of them and their tags. From the list, you can move to all the tags by the user. In Technorati, you can likewise follow the Authority trail to individual bloggers.

On the other hand, Amazon does not allow you to see who the users who voted a review as useful or not useful are. We only get their number. Likewise, in Technorati we see the number of people linking to, say, a video but we have no way to find out who these people are.

Finding out the community identity of a user, such as user name, however, does not necessarily give us much information. Consequently, different sites are using various ways to improve the feeling of reliability of the raters and reviewers. Amazon has Real Name™ badge for showing that the reviewer goes by his or her real name. The identity is guaranteed by the name having been taken from the user's credit card. Real Name™ is only one of the badges that Amazon uses. The others include such badges as "THE" (given to celebrities such as, surprisingly, Amazon's founder Jeff Bezo), and "Top 10 Reviewer" and "Top 50 Reviewer" that denote a ranking of the reviewers. While these are certainly to encourage submissions, they also make the reviews and ratings more credible for other users.

4.4 Tagging

Tagging, that is, associating user selected keywords or phrases to describe items, be they books, photos, or blogs, is a new and fashionable way to collect and use collective intelligence. While the idea of using keywords, in this context called tags, to organize electronic material is not new (Furnas et al., 2006; Golder & Huberman, 2006), on the Internet it has only been used since 2004 (Hammond, Hannay, Lund, & Scott, 2005). Tagging on the Internet usually means that a user gives different meanings to certain items, such as photos in Flickr, blogs in Technorati, or artists in Last.fm, by linking one or more words to them.

Tags help people to organize information and share and discover new information that someone else has contributed to the system (Sen et al., 2006). Moreover, if the same resource is tagged by many users, it can be used as a sign that the resource is more likely to be valuable (Furnas et al., 2006).

Tags offer new ways to organize material, provide first impressions, create means for social navigation, and find items in communities where the number of items is high. Over half of the sites studied for this paper offer tagging features (c.f. Table 2 on page 4). In effect, tags represent one of the major ways of organizing content, navigating, and searching in Flickr and Del.icio.us. For others, such as Amazon and Last.fm, tagging represents additional way for navigating and searching for items of interest.

Tagging communities, such as Flickr and Del.icio.us, can be viewed as self-organizing communities where the system does not impose structure on the users (Weiss, 2005). However, tagging does not only bring benefits to the community. Individual users can also use their own tags to organize their own material, thus providing more motivation for using tags.

4.4.1 Tags and vocabulary

Members of a community that shares a common vocabulary get the most advantage from each other's tags (Sen et al., 2006). Common vocabulary makes tagging a powerful tool for social navigation, and searching and finding items of interest. In MovieLens, you can rate the tags given by others with thumbs up or down to indicate agreement with the word chosen.

Users take an active role in tagging. In many systems, users can freely choose whatever words they want to use for tagging an item. The users are known to have personal tendencies in their tagging (Sen et al., 2006). Figure 11 shows tags that users have added without

constraints as to the tags to a product, in this case a book called *A thousand splendid suns*. The fact that some users have tagged the book with its own name shows that not everybody understands tagging. Some guidance on how to use tags might prove useful in bringing out the full potential of tagging.

Tags customers associate with this product

First tag: [wish list book](#) ([Jason Macbeth](#) on Nov 29, 2006)
Last tag: [tragedy of women in afghanistan under se...](#)

Sort by:

Popularity

- Popularity
- Alphabetical: A to Z
- Alphabetical: Z to A
- Tag Date

fiction (6)	afghani women (1)
afganistan (5)	author to check out (1)
south africa (4)	bendib (1)
books to consider (3)	but no kite runner (1)
literature (3)	ccaine (1)
mightlike (3)	christian science monitor (1)
war (2)	classics (1)
women (2)	fiction (1)
1000siddiq (1)	for moms birthday (1)
a thousand splendid suns (1)	friendship (1)
a thousand splendid suns khaed hosseini (1)	housseini (1)
a thousand splendid suns--splendid (1)	islamic women (1)
afghaistan (1)	joe guse (1)
afghani fiction (1)	kabul (1)
	khaled (1)
	masterclass theater (1)
	memoir (1)

Figure 11. Tag list in Amazon with a sorting tool for organizing them.

It is hard for people to agree even on simple descriptive words for an item. Tags can give facts about the item, describe one's opinions about it, or be somehow meaningful only in personal use (Sen et al., 2006).

However, in some systems users can only select tags from a given set of words. This is already rather close to an ontology that is then applied to items by users. In some other systems, such as Flickr, the user can choose the words freely, but is also given suggestions, such as tags other users have used together with the tag that the user is entering (Weiss, 2005).

Personal tendencies of selecting tags change over time as personal experience about tagging systems increases. However, Sen et al. (2006) have shown that the tagging community has its effects on the tagging behavior as well. People observe how others act in the community and what tags others use and then start to copy that behavior because they believe it is the correct way to act. This is another example of self-regulating behavior observed in online communities. Furthermore, the actual tagging interface plays some role in deciding if novel ones are invented or if the tags used are borrowed from tags other people have already used. If the tags used by others are shown, they are more likely to be used. (Sen et al., 2006)

Tagging also brings in some linguistic problems, such as polysemy, synonymy, and basic level variation, into play (Golder & Huberman, 2006). Polysemy is a problem caused by the fact that some words have more than one meaning. This means that when searching with some special word, we also might get irrelevant results. Still, this is not that serious a problem because the relevant results are also found. In contrast, synonymy is a much more serious problem. It occurs when there are many words that denote the same thing. This kind of

inconsistency in words means that some relevant results are not found. This is the kind of a problem that cannot be prevented in tagging systems where users can freely choose which tags to use. A bit similar problem is variations in words, including such linguistic challenges, such as plurals and abbreviations. They can also cause some relevant information not to be found.

Consequently, it is hard to see a complete communal understanding to emerge from tagging as it is today. However, sub-communities can develop their own tagging vocabularies and profit from them. Collaborative filtering can help recognize people who tag in a similar way and thus allow them to use each other's tags more effectively. This is one of the strengths of the so called folksonomies, sort of "folk taxonomies" (Golder & Huberman, 2006) as they allow sub-communities to develop and have their own understanding of the material. Groups in Flickr, for instance, appear to be developing their own terms.

4.4.2 Geotagging

Geotagging was introduced by Flickr, and it means that users can assign a geographic location (latitude and longitude) for the photos they upload to Flickr. Typically, the geotag points to the place where the photo was taken. The photos can then be browsed by using a map (Figure 12). For example, if the photographer adds three tags to a photo:

```
geotagged
geo:lat=51.494434242907
geo:lon=-0.17341714682856
```

...the latitude and longitude in the geotag associate the photo with a location in Chelsea, London, in Great Britain. (Silver Smith, 2006)

The geotagging feature has been available in Flickr since August 28, 2006. Only 24 hours later, the developers posted a blog entry that more than 1 million photos had been geotagged in Flickr within the first day (Butterfield, 2006).

Automatically tracing the location (with a GPS-enabled camera or mobile phone with GPS and camera) is one option for getting the geotags attached to the photo. The active discussion groups on Geotagging in the Flickr community have also introduced several software packages for the job (Geotagging Flickr, 2007). Drag-and-drop of photos has also been implemented on Yahoo maps by Flickr, but this direct manipulation approach does not give the photos exact coordinates, which minimizes their relative value at least in the eyes of the most committed geotaggers (Geotagging Flickr, 2007).

Some other tag-like information can be extracted from the photos without the users actually entering it. Flickr extracts EXIF (Exchangeable Image File Format) information, such as camera model, f-stop, and focal length used, from the photos and shows it with the picture. Furthermore, the EXIF information is used to make statistics and trends of camera usage in the Flickr community in such features as "Most Popular Cameras in the Flickr Community." While showing the EXIF data automatically next to photos is a potential privacy violation, the option can be checked and unchecked in the user profile settings.

In Flickr, the privacy level for location can be set independent of other photo privacy settings. For instance, you can allow only family and friends to see the location information even if your photos are otherwise viewable for all.

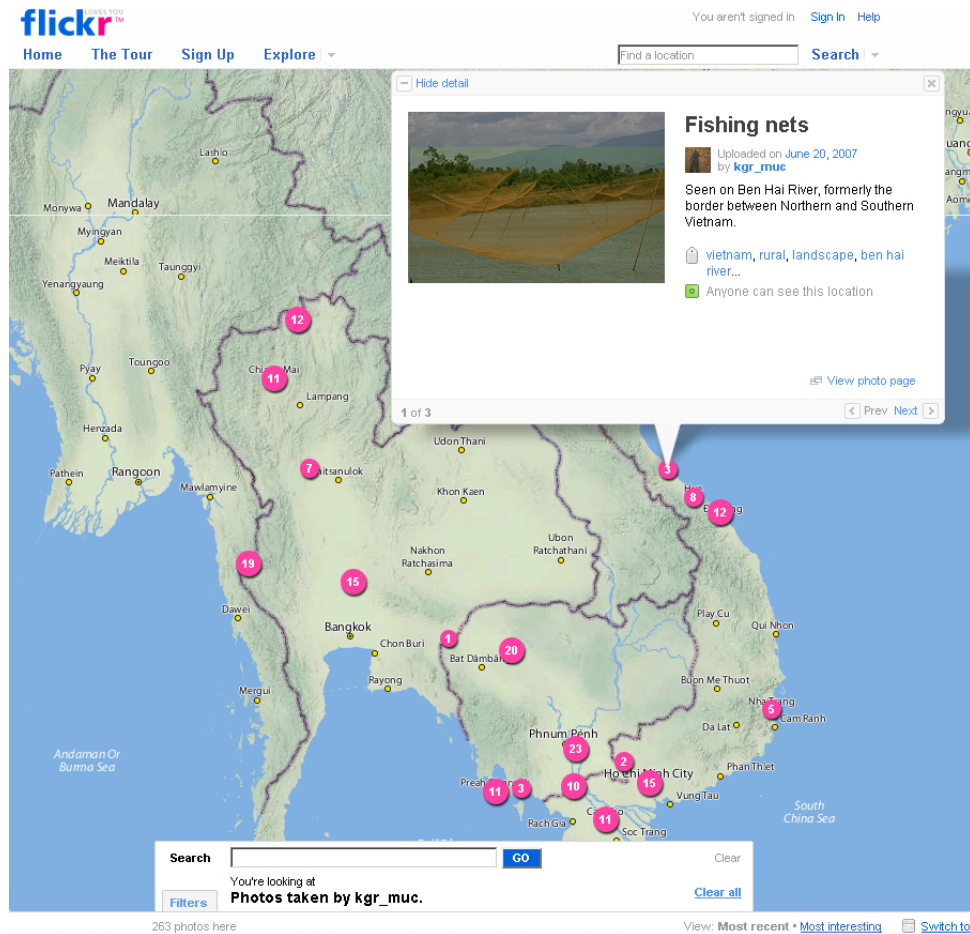


Figure 12. Geotagging interface in Flickr for viewing one user's photos.

The potential for geotagging is huge, especially in mapping services like Google maps. The users can add their own tags and even commercial advertisements to locations (Silver Smith, 2006). We are likely to see more services employing geotagging or similar features.

4.4.3 Tag lists and tag clouds in the interface

There are numerous interface design approaches for presenting tags. While tag clouds are becoming very common, tags are often also presented as lists. Sometimes, some visual formatting is applied to the lists so that more important tags stand out. Figure 13 is an example of two different layouts of the same set of tags a person has associated to items in dogear.com. While the top layout resembles a cloud with a fisheye view, the bottom presentation looks more like an ordinary list.

Impression forming or “gisting” means that, based on something, such as tags, we form “a general impression of the underlying data set or entity associated with it. This impression should include awareness of the most prevalent topics, but also knowledge of those that appear less frequently” (Rivadeneira, Gruen, Muller, & Millen, 2007). While the alphabetically ordered list allows for sequential scanning and searching, it is likely that the tag clouds make it easier to form impressions (Rivadeneira et al., 2007).



Figure 13. Two visually distinct versions of a tag cloud (Rivadeneira et al., 2007).

The tag lists are usually presented in some ranking order, typically with the most commonly used tag on the top and the rest in descending order. In Amazon, the default order is the popularity of the tag (Figure 11, p. 40) shown without applying any visual emphasis on the more popular items but the number of the people who used the tag on the item is shown in brackets after the tag. Sometimes some organizing or filtering tools are provided for working with the tags, as in Amazon where the tag lists can be sorted with such tools.

Of the sites studied for this paper, Del.icio.us, Flickr, Last.fm, Amazon (Figure 16), and Technorati have tag clouds. In their tag clouds, the tags are in alphabetical order and the font size denotes the popularity of the tag. Figure 14 is an example of Flickr's tag cloud of the most popular tags.

All time most popular tags

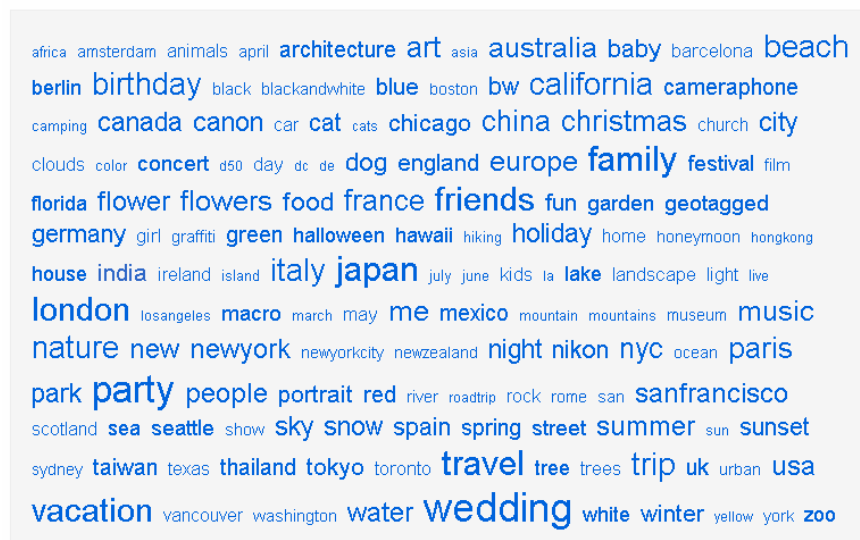


Figure 14. Flickr's tag cloud of most popular tags.

Interestingly, although tag clouds have very quickly become very popular and common, there are few studies about their effectiveness in the tasks they are meant to support (Rivadeneira et al., 2007).

Of the sites studied for this paper that use tag clouds, all but Last.fm use tag lists in addition to tag clouds. The lists are typically located in the item pages separately for each item. Last.fm uses tag clouds also on item pages. While Flickr uses list for “Hot tags” (“In the last 24 hours” and “Over the last week”), it presents them together with the corresponding tag cloud. Flickr uses boldness and order to denote popularity of a tag in the list.

In some services, tags constitute only one of several features and they are not as prominent in the interface as they are in Flickr and Technorati. YouTube, for instance, offers a tag list next to the video window but not in a very prominent position on the page (Figure 10). Similarly in MovieLens, the tag list is positioned below the title of the movie (Figure 9).

Numerous visual design options related to text features and word placement are available when constructing tag clouds. Text features include font weight, font size and font color while word placement includes sorting (alphabetically, by frequency, or by a predetermined algorithm), clustering (semantically or by user preferences), and spatial layout (lists or clouds). (Rivadeneira et al., 2007).

Both tag clouds and lists can be used for navigation in all the studied sites that had a tagging feature. Clicking on a tag performs a search and lists the content where the tag is used. Combining tags with search terms is not provided in most services. One of the studied sites, Last.fm, offers the tag search functionality (Figure 15). For instance, a user can search for “relaxing modern” from artists, album, track, tags, or labels. The search results for tags returned 1 253 tags matching either relaxing or modern, among which there are also such variations of “relaxing” as “so relaxing”, “very relaxing”, and “relaxing on the back porch” as separate tags. While the tag “relaxing” has been used by 4 381 people more than 20 000 times altogether, only two persons share the tag “relaxing on the back porch” and they have added it to four songs. Clearly, the search functionality increases its importance when the number of tags and tagged items grows.

Find Music

Artist Album Track **Tag** Label

relaxing modern Search

1 2 3 4 5 6 7 8 9 10 11 Next >

Tags

1,253 tags matching "relaxing modern"

Tag	Popularity	Tag	Popularity
relaxing	👍👍👍👍👍	modern music	👍👍👍
modern	👍👍👍	relaxing trance	👍👍👍
modern rock	👍👍👍	relaxing classical	👍👍👍
modern metal	👍👍👍	so relaxing	👍👍👍

Figure 15. Part of the search results for music tagged relaxing or modern in Last.fm.

4.5 History information

History information is typically collected implicitly as a service is being used. It is primary material for social navigation, as it tells where people typically go under certain circumstances, or is used to recommend items or paths (Dieberger et al., 2000). We have already discussed how implicitly collected history information can be used for recommendations. However, history information can be used also for other purposes than for collaborative filtering and popularity-based recommendations that typically aggregate it.

History information can be seen to be related to an artifact or a user. When Technorati front page tells us that there are 113 new links to Amy Winehouse's album or that 149 Del.icio.us users have bookmarked a link "Ian's Shoelace Site - Shoe Lacing Methods", the information is connected to the artifact. On the other hand, if we click to see what other links besides shoe lacing methods Del.icio.us user "mguth" has added to his bookmarks or what another listener of the song we are listening to has listened in Last.fm, we are looking at history information related to a user. Of course, the user and the artifact are here just two sides of the same coin.

In some services, we can navigate based on an individual user's history by selecting items from their history lists. Both the earlier examples of seeing what other links a Del.icio.us user has bookmarked or what other songs a Last.fm user has listened to are examples of this. Thus, social navigation and recommendations can be based on individual user's history as well as an aggregation of several users' data. However, when history tract leads us to individual users, we tend to face various privacy challenges (Dieberger et al., 2000). Not all of us are happy to let the world know what books we have bought, for instance.

Furthermore, we can use our own personal history to navigate. Amazon gives us a list of products we have viewed in "Your Browsing History" at the bottom of every product page. This way we can easily return to the items that we have viewed without having to remember any exact information about the product.

Another question is how long the history information is relevant (Dieberger et al., 2000). Should it fade in importance after a while, and at what rate? The answer is likely to be related to the user's purpose and application environment. If we want to see what music is popular right now, we do not want too much of listening history in the aggregation, but if we want to see long-term trend, the situation changes.

Consequently, Flickr, for instance, gives different history-divided views to its collection of photos, such as "Last 7 Days Interesting" and "A Year Ago Today", and "Most Recent Photos." Similarly, Last.fm allows us to see each user's "Top Artists this Week" and "Top Artists Overall." Amazon tries to convey in its tag cloud the recentness information in a subtle way by darkening the tags that have been used most recently in its "Most Popular Tags" (Figure 16). Resorting to visual variations is not a watertight option, though, since the small differences might go unnoticed by the users.

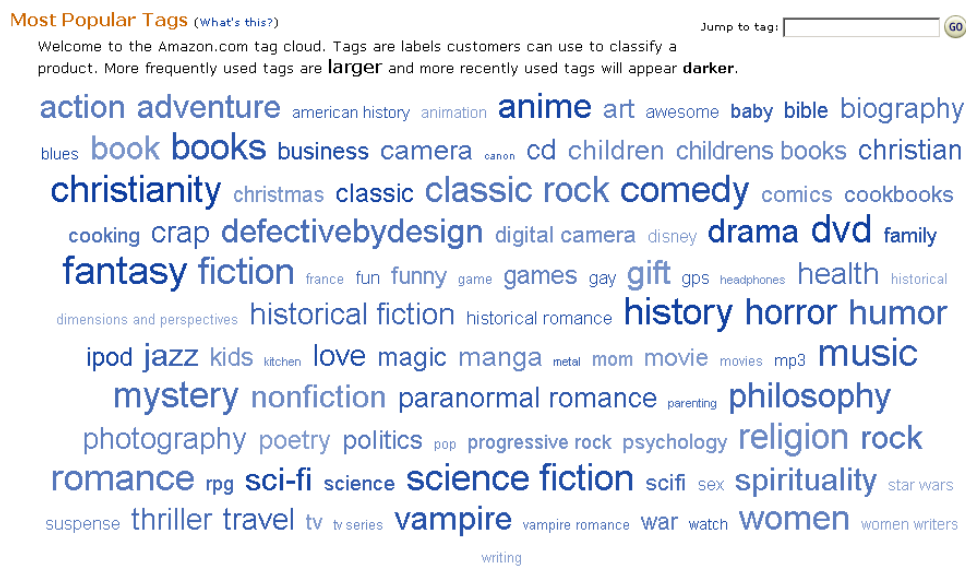
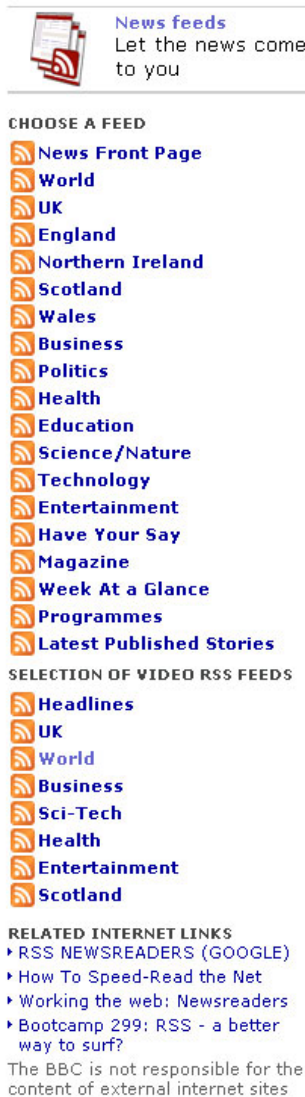


Figure 16. Amazon's "Most popular Tags" tag cloud.

4.6 Push and pull technologies: feeds, instant messaging, shoutboxes, and chats




















Push technologies have been part of the web since the days of the first mailing lists, as reported in the story of the first even if unintentional spam e-mail message in 1978 (Wikipedia, 2007l). Early subscribed push transactions consisted of newsletters and digests, among others, that arrived in the e-mail boxes of the subscribers. Mailing list concept is, in fact, very simple: once a user has signed up on the list to receive messages, the messages posted on the list are sent automatically either at once or first as a digest of messages.

Today we also have syndication and web feeds, such as RSS and Atom, in which users subscribe to the channels of interest. For example, the BBC web site publishes content in several feeds (Figure 17). To get the frequently updated content, users can subscribe to a feed with an aggregator program, also called a news reader or feed reader, which is running on their own machine. Subscribing takes place for instance by dragging the link from the browser window to the aggregator. The aggregator checks for new content periodically and provides the user with an updated view of the site content, typically as links or other HTML content (Wikipedia, 2007h).











News feeds
Let the news come to you

CHOOSE A FEED

-  [News Front Page](#)
-  [World](#)
-  [UK](#)
-  [England](#)
-  [Northern Ireland](#)
-  [Scotland](#)
-  [Wales](#)
-  [Business](#)
-  [Politics](#)
-  [Health](#)
-  [Education](#)
-  [Science/Nature](#)
-  [Technology](#)
-  [Entertainment](#)
-  [Have Your Say](#)
-  [Magazine](#)
-  [Week At a Glance](#)
-  [Programmes](#)
-  [Latest Published Stories](#)

SELECTION OF VIDEO RSS FEEDS

-  [Headlines](#)
-  [UK](#)
-  [World](#)
-  [Business](#)
-  [Sci-Tech](#)
-  [Health](#)
-  [Entertainment](#)
-  [Scotland](#)

RELATED INTERNET LINKS

- ▶ [RSS NEWSREADERS \(GOOGLE\)](#)
- ▶ [How To Speed-Read the Net](#)
- ▶ [Working the web: Newsreaders](#)
- ▶ [Bootcamp 299: RSS - a better way to surf?](#)

The BBC is not responsible for the content of external internet sites

Figure 17. List of news feeds for subscribing in the BBC web site.

Newsletters have been with us since the first mailing lists implemented in the Internet, but web feeds and syndication are part of today's Web 2.0. Seven of the eleven services we studied offered web feeds. If anything, the number is surprisingly low since, as discussed, these types of services keep the users up-to-date with the service even if they do not visit the site. Thus, they motivate the users to return to the site.

Web feeds can be and often are used in mash-ups, web applications combining content from many sources (Wikipedia, 2007g). Mash-ups are discussed in Section 4.7.

While the feeds appear to be push technology, that is, the transmission originates with the publisher, they are in fact better described as pull technology since the aggregator checks the server for content and the content is not sent unrequested (Wikipedia, 2007h). In this sense, we still do not have true push technologies besides spam and instant messaging. Considering such attempts as “the derided Pointcast service of the mid-1990's” and its justified demise after endless stream of headlines and advertisement (Quain, 2004), we probably should be grateful for that. On the other hand, instant messaging, discussed below, is one type of push technology that has spread in the web.

Instant messaging is “a form of real-time communication between two or more people based on typed text” that is conveyed over a networks (Wikipedia, 2007e). Instant messaging is one form of push technology that has become prevalent in the web. Again, the idea itself is old. Some Web 2.0 services, such as Gmail, are now integrating instant messaging into the site interface instead of the users needing to have separate software for it.

In Last.fm, visitors can send short messages to another user with a shoutbox (Figure 18) where the message waits for the user if he or she is not logged in currently. In effect, the Shoutbox, also known as saybox, tagboard, or chatterbox, is a chat-like feature for leaving short messages and they usually do not require registration (Wikipedia, 2007j). Thus, the shoutbox does not represent instant messaging as far as real-time aspect is concerned, but otherwise it looks and feels like instant messaging. In any case, Shoutbox certainly represents push technology, originated by other users of the site.



Figure 18. An example of Shoutbox in Last.fm.

Habbo hotel's Habbo talking is also a form of instant messaging without history (Figure 19). It is implemented in such a way that to be able to read the text appearing in a speech bubble one needs to be close to the speaking Habbo in the physical room layout at the time of the utterance—the bubble moves up and evaporates. Although shouting to the whole room is possible, it is considered rude behavior under most circumstances. While usually the movements of the avatars are not blocked or restricted in any way, a room can have special areas for “inside” group members that requires a password for entrance and outsiders cannot see what is said in that area. In addition to speaking aloud in speech bubbles, Habbos can also send console messages to named members. These messages are private. (Johnson, 2007)



Figure 19. A room in Habbo Hotel with speech bubbles (Johnson, 2007).

The major problem with instant messaging in the interface is the real-time aspect that requires people to be simultaneously online. Consequently, many services have opted for different types of messaging approaches where the simultaneity is not necessary and one user leaves a message for another to respond when he or she gets online. For instance, building connections with other users, such as asking them to be your friends, is often done this way. The messages are sent as email notifications to the person contacted by the service providers.

While Wikipedia supports IRC chat for the developers, the software is not integrated in the Wikipedia interface. Although each version of a Wiki page has a talk link to the author who has saved it, the “talk” is saved as a discussion forum entry that can be viewed later. Viegas et al. (2007) emphasize the importance of persistent talk pages as a tool for coordination in Wikipedia. Chat messages as in Habbo are not saved in history, but even Habbo enthusiasts have community sites that are external to the Habbo Hotel (Johnson & Toiskallio, 2005).

4.7 Widgets, gadgets, mash-ups, and open APIs

On December 30, 2006, Newsweek's Technology and science section suggested that year 2007 would be the year of widgets (Braiker, 2006). Widgets or gadgets are short pieces of code, kind of mini-applications, that can be dragged onto the user's desktop or copy-pasted to a web page to include content from another service, creating in effect a mash-up. A mash-up is "a website or application that combines content from more than one source into an integrated experience" (Wikipedia, 2007g). Web feeds can also be seen as content for mash-ups (Wikipedia, 2007g).

Mash-ups have become possible—and very popular—with many Web 2.0 services partially opening up their APIs (Application Programming Interface) so that others can access their content under certain terms. While many offer ready-made widgets with easily adjustable parameters, open APIs also make it possible to program custom-made widgets, c.f. Figure 20 for Flickr. Last.fm has a 3-step wizard for creating, embedding and sharing widgets. In its case, the widget is of course a radio station for listening. Of the sites that we studied, also Amazon, YouTube, Technorati, MySpace, and Del.icio.us have opened up their APIs for making mash-ups. Thus, altogether seven of the eleven sites studied have opened up their APIs.

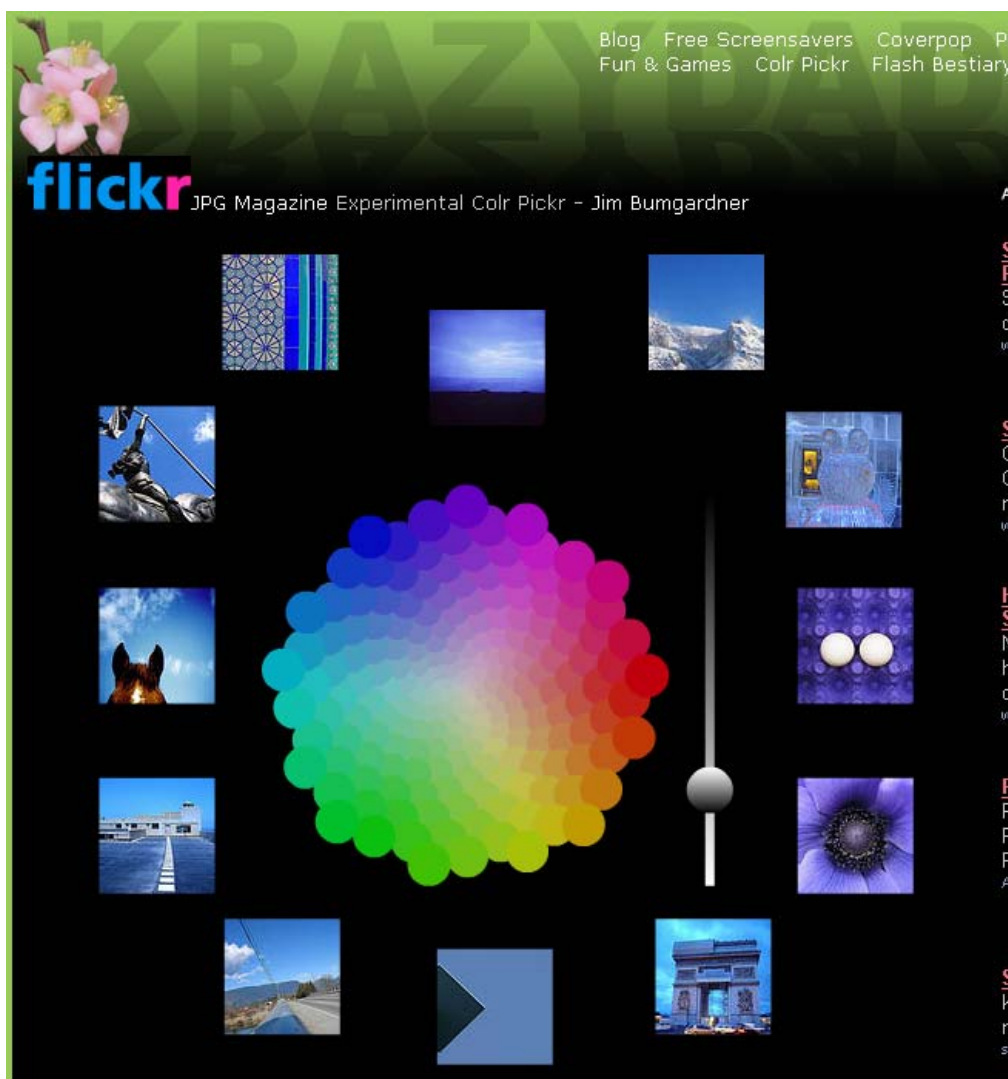


Figure 20. An example of a service that is built on top of Flickr's open API.
(<http://www.krazydad.com/colrpickr/index.php?group=jpgmag>)

In the example in Figure 20, Jim Bumgardner's Flash application uses Flickr's open API to find photos tagged with a certain color. The color is selected by clicking a color in the color picker. However, the application does not analyze the actual colors in the photos but merely the tags that the users have associated with the photos.

If we look at the Internet more widely, also Google, Yahoo, eBay, and Digg.com, among others, have opened up their APIs at least to some degree. It is easy to see how Google Maps, for instance, can be integrated into many services to show location, as in Housingmaps.com's real estate listings (Weiss, 2005), and in www.chicagocrime.org where Google Maps are combined with local crime statistics.

The opening of APIs has resulted in intense growth of development activity and collaboration on the Internet (Weiss, 2005). Not everybody is happy, though, as mash-ups, like any new approach, challenge the traditional ways. Mash-ups do not allow us to use page views for measuring site popularity as the content might be seen in another site and no hits are recorded in the original content-providing site (Braiker, 2006). However, mash-ups appear to be here to stay, and new ways are likely to develop around them as the approach matures.

5 Collective intelligence as content

As the preceding discussion on features shows, collective intelligence has become an important source of content. It is not only that recommendations and different lists can be generated from explicit and implicit user actions. Users are also important providers of what we read, watch, and listen to today on the Internet.

In fact, much of what we see in the interfaces in Web 2.0 services is based on user activity and contributions. Be it

- Amazon's star-ratings for products or bestselling lists
- Last.fm's "People online right now" listings of user profiles or the lists of "Up-and-coming stations"
- Del.icio.us's "Hotlist" links and the number of their respective votes,
- Flickr's "<number> uploaded in the last minute" and tag clouds or
- Technorati's "Popular videos" and the number of their respective links,

all are based on user actions and contributions. In short, without the users there would be little content in a typical Web 2.0 site.

In some services, practically all content is based on users or at least the site could not exist without the user activity. Wikipedia is an obvious case to the point but also Flickr would be nothing without the photos, tags, and comments added by the users. Even its ways of navigating are based on user-contributed material that has been used in the features that the site offers. Del.icio.us is in a very similar position. Users provide the links and tags without which the service would have no content and no navigation. Technorati is also based on user-contributed content as is YouTube by and large as well.

Some services, while based on both explicit and implicit contributions by the users, have some content as a starting point. Last.fm, while allowing musicians to add their music to the service, has a large collection of music for the listeners to enjoy. However, beyond this, much of the content is user-based. Last.fm's band descriptions are user-contributions to the site's wiki, its radio stations are based on its listeners, and the music played to you is based on collaborative filtering, tags and so on. Without user activity, Last.fm would hardly be able to play "Tina Turner's Similar Artists" or offer a station called "Play Listeners of Tina Turner."

Similarly, while Amazon has large collections of products for sale, many features depend on users and their contributions and activities for their existence. The users contribute the star ratings, Customer reviews, Amazon.com community (product wikis), and Customer Discussions, in addition to tags, "Help others find this item", "Customers who bought this item also bought", and Listmania! Lists, among others.

Overall, all the Web 2.0 services we looked at for this paper depend on user-activity and contributions for survival. No new videos from the users to YouTube and it would die. No

new photos and tags to Flickr and it would soon be unvisited. No links to Del.icio.us and it would be useless to its users. If there is a be-all-end-all for Web 2.0, it is the user-based and user-contributed content.

5.1 System-generated content and user-generated content

We can broadly divide user-based content into two categories: system-generated content and user-generated content. System-generated content refers to any content that is generated based on and derived from the user actions, be they implicit or explicit. However, the key is that no matter how the user information is collected and what it is, it is not shown as it is but the system generates and often aggregates the content for its interface presentation. On the other hand, user-generated content is something that users enter into the system explicitly and that is shown basically as it was entered.

For instance, the star ratings on the product list pages in Amazon are system-generated content as they are averages of the stars individual users gave the products as part of their reviews. The system counts the average. In contrast, the star ratings and the reviews in the product page are user-generated content as the stars in one review are what one user gave and the review is something one user wrote. The system only presents the user-generated content without using any kind of collaborative filtering or calculations on it.

Likewise, tag clouds in various services are system-generated content. They are aggregations of what various users have entered in the system and how the users have acted in the system. On the other hand, if we look at what tags a certain user has given to a product, we are viewing user-generated content.

User-generated content can further be divided into two categories: collaboratively generated content and individually generated content. Collaboratively generated content refers to any user-generated content that is generated collaboratively by several users, such as a Wikipedia article, while individually generated content refers to content generated by one user, such as a book review in Amazon.

User-generated content can also be examined from the point of view of who can edit or delete it. Collaboratively generated content can be edited or removed by a number of user-editors while individually generated content can only be modified or removed by the user who generated the content in the first place. However, this way of examining content leaves the role of a moderator outside of the consideration.

Of course, few categorizations are perfect without any gray area. Photos in Flickr are user-generated content as users enter them into the system. However, other users can define areas in the photos and add comments to them. Is the end result of a photo with comments user-generated content or system-generated content? Probably we would argue that it is user-generated as the photo and comments are user-generated content. In the same vein, Wikipedia articles are user-generated although they have multiple authors because the system is only an enabler in the equation. However, these examples show that our categorization is meant for discussion and does not represent absolute division that is clear-cut in every situation.

5.2 Business value and user generated content

Many of today's successful Web 2.0 services are, in fact, based on an idea of allowing some kinds of user contributions and a model how to generate value from them both for the users so that they will use the site and for the service provider. YouTube is about allowing users put their videos online while Flickr is about photos online. The crucial question for businesses in

Web 2.0 is how much value the site generates for the users who contribute the content by explicitly giving it or by habiting the site and thus providing an opportunity to collect implicit information. Flickr is not the only photo-site in the web and neither is YouTube the only place where you can put your videos for the world to see. The make-or-break question is the value that can be generated for the users.

While advertisement and big money behind the services matters, it is not always the whole story or even the decisive factor for success. For instance, YouTube was founded by three ex-employers of PayPal with angel-funding, private investor funding for companies that are too small to attract venture capital (Wikipedia, 2007a), in 2005. Today the service is owned by Google Inc. which acquired it for \$1.65 billion US in Google stock at the end of 2006 (Wikipedia, 2007q). This example shows the importance of the initial concept. The key is to reach the critical mass of users (Weiss, 2005). In the Internet-dominated world, the word does get around even without conventional marketing efforts.

User-generated content generates value in many ways. Most Web 2.0 services are business ventures and often the business is based on user-contributed content wholly or partially. Of the sites we studied, only Wikipedia is a non-profit foundation. Amazon and Sulake (Habbo Hotel's owner) generate their income from the sales of virtual or real commodities, LinkedIn by charging money for some user account levels. For the others, advertising revenue often plays a substantial role in their business models for generating income.

In 2007, according to Accenture's survey of 110 senior executives in the media and entertainment industry, "[m]edia and entertainment executives see the growing ability and eagerness of individuals to create their own content as one of the biggest threats to their business" (Accenture, 2007).

"This is just the beginning for a rapidly changing landscape where the media content environment grows more fractious and the user gains more control and power," said Gavin Mann, digital media lead for Accenture's Media & Entertainment practice. "Traditional, established content providers will have to adapt and develop new business and monetization models in order to keep revenue streams flowing. The key to success will be identifying new forms of content that can complement their traditional strengths." (Accenture, 2007)

At the same time that they saw user-generated content as a threat, the executives who took part in the Accenture study also believed that within three years their businesses would be making money out of social media (Accenture, 2007). Thus, Web 2.0 and especially user-generated content are seen as both threat and opportunity in today's mainstream media. Today, nobody can be quite sure what the equilibrium will be a few years down the line.

In any case, user-based content is challenging the old models of entertainment and information businesses. Wikipedia is taking on such established giants as Encyclopedia Britannica while YouTube is challenging television broadcasters. Flickr allows its users to sell their photos, thus challenging the photo agents. While the business models are going through changes, also our social world is changing as network-building is moving to web and people get to know each other through social services in addition to traditional ways.

In addition to user-based content, sites also have content specifically created for the site, which is of course the traditional model of web sites. Content can also be a commodity to be bought and sold or even given.

The key difference between bought and user-based content is that user-based content, generally speaking, does not cost money. This makes it very attractive for the services. It

should also—ideally—protect them from all kinds of licensing problems. For example, Pandora.com, a collaborative filtering based radio station similar to Last.fm, had to stop broadcasting to international users due to licensing constraints (Westergren, 2007) when a “contentious decision that would elevate royalty fees Webcasters must pay to record labels” was upheld in April, 2007 (Broache, 2007). It now only serves users from US and Canada although it originally served all users independent of location. That spells difficulties for other net radios as well:

“‘As a former touring musician myself, I’m no stranger to the challenges facing working musicians,’ Tim Westergren, founder of the Internet radio service Pandora, wrote in a Monday e-mail. ‘The issue we have with the recent ruling is that it puts the cost of streaming far out of the range of ANY Webcaster’s business potential.’” (Broache, 2007)

Some content is produced for Web 2.0 services for promotional purposes. For instance, some companies have produced video clips for YouTube and similar services. Having often sexy or witty promotional clips in the service with people emailing the links to each other means free advertising with a community of users taking care of promoting it virally. That spells good business. In the web, the same restrictions as in TV and radio advertising do not apply. However, as the Pandora example shows, the rules for Internet are just being made today.

Not all the Web 2.0 services simply take the user-generated content for free. For instance, Revver.com has a revenue-sharing ideology where the makers of the video get 50% of the advertisement revenue (Revver.com, 2007). Brightcove.com (2007) has a similar model. It “provides content producers with a number of ways to generate revenue, including placing adverts inside videos, and through charging for downloads, which can either be made available for rental or for purchasing outright. Brightcove will keep 50% from ad-revenue, and 30% from downloads” (O’Hear, 2006). Again, we see the new business model trying to replace the old model, in this case the model applied by YouTube.

Although YouTube has a huge advantage over its competitors due to the huge exposure factor it enjoys—for instance, the site had 46.4 million visitors in April, 2007 (Barlas, 2007)—it has apparently felt the pinch of the new models. “Speaking at the World Economic Forum, YouTube CEO Chad Hurley has revealed that the company plans to financially compensate users who produce and upload their content” (O’Hear, 2007). The details of the model, however, are not clear yet.

5.3 User-generated content concerns

User-generated content seems easy and effortless to collect and monetize, but that is not the whole truth. While those depicted in the material might feel their privacy violated, the ownership and other fair use questions also cause disputes.

It is easy to publish and distribute information on the net about oneself and other people. Audio-visual material entered in YouTube or written into personal blog entries, regardless of its truthfulness, might gain massive popularity and reach thousands of people in a matter of minutes. That the networked identity and exact copies of one’s words will persist is inherent to online contributions (boyd & Heer, 2006). Exact copies of one’s words once written somewhere can be quoted at will elsewhere in a manner that people seldom come to think about.

Some people may thoughtlessly or even intentionally upload content that contains private or confidential information to a public Internet service. After that, it is difficult to control the

information distribution or take back the content. Web pages that have been removed from the original server may still be visible via proxy servers or services like Google cache. Furthermore, online content is easily duplicable and the duplicates cannot be removed as easily as the original one. The freedom of expression is taken for granted, but there is no knowing where and by whom the words, photos, or videos will be shown or linked to in the future. There have already been examples of privacy violations and defamation trials where the defendant has been sentenced to jail and ordered to pay damages to the plaintiff. A recent example is that of a person posting sexual video material of a former girlfriend into YouTube (STT-IS, 2007).

As YouTube and others have found out, user-generated content occasionally tends to translate into copyrighted content put into the system as user-generated content (Andrews, 2007; Indiantelelevision.com Team, 2007; Martinson, 2006). One recent example surfaced in YouTube where the latest Michael Moore documentary “Sicko” was posted in 14 clips a week before its official release at least by two users (AP, 2007). The documentary was removed in 2–3 days because of copyright complaints by the distributors of the film. Apparently about 500 visitors had seen the film before it was removed.

The problem is of such magnitude that YouTube, while still struggling to consolidate a stream of revenue, has set aside a reserve of \$200 million US to cover the site’s lawsuit losses (Yang, 2007). Contrast that to the fact that Viacom is seeking \$1 billion US from YouTube in a lawsuit—and that is only one of the lawsuits in which YouTube is involved—and the true magnitude of the problem becomes clear.

The legalities and legal responsibilities are uncertain (Masnick, 2006), and while many services agree to remove copyright-violating content on notice, copyright holders have complained of tardiness and non-pro-active approach taken by the services (Indiantelelevision.com Team, 2007). The decisions to be made in courts around the world are likely to clarify the situation.

In any case, the content once removed tends to pop right back into the services. The threat of more legal action has prompted the services to start designing technologies to keep the once-removed copyrighted content out of the system (McCarthy, 2007).

MySpace started to implement a technology titled “Take Down Stay Down” in May, 2007. The technology allows copyright holders to take content down and uses “digital fingerprint” to make sure that the same content is not re-posted (Kaplan, 2007; McCarthy, 2007). YouTube that faces several legal challenges is also moving towards similar tools. YouTube’s filtering system is called “Claim your content” (Ali, 2007; McCarthy, 2007).

How well the technologies will work remains to be seen. Corynne McSherry, Electronic Frontier Foundation attorney, commented the MySpace initiative skeptically:

“It certainly is true that with every form of digital rights management that we’ve ever seen, it always gets hacked eventually, so I think it’s likely that eventually this too will be hacked. It’s just a matter of time.” (McCarthy, 2007)

Digg.com, ranked as “one of the top 100 Web sites” by Alexa.com (c.f. Chapter 2), faced another challenging aspect of user-generated content woes recently when a user posted “32-digit code that lets people crack HD-DVD copyright protection” in the site in May, 2007 (CNN, 2007). Initially, the company started to delete messages containing the code but when facing an angry community of “diggers” concerned about censorship, the company opted to let the code stay. Finally, the Digg founder Kevin Rose posted a blog entry with the code as the subject (Malone, 2007):

“Today was an insane day. And as the founder of Digg, I just wanted to post my thoughts...

In building and shaping the site I’ve always tried to stay as hands on as possible. We’ve always given site moderation (digging/burying) power to the community. Occasionally we step in to remove stories that violate our terms of use (eg. [*sic*] linking to pornography, illegal downloads, racial hate sites, etc.). So today was a difficult day for us. We had to decide whether to remove stories containing a single code based on a cease and desist declaration. We had to make a call, and in our desire to avoid a scenario where Digg would be interrupted or shut down, we decided to comply and remove the stories with the code.

But now, after seeing hundreds of stories and reading thousands of comments, you’ve made it clear. You’d rather see Digg go down fighting than bow down to a bigger company. We hear you, and effective immediately we won’t delete stories or comments containing the code and will deal with whatever the consequences might be.

If we lose, then what the hell, at least we died trying.” (Rose, 2007)

Of course, by then the code had already spread everywhere anyway (Moren, 2007). However, now we are waiting to see the possible legal ramifications to Digg. Incidentally, the code also popped up in Wikipedia from where it was promptly removed (Moren, 2007).

The Digg.com case underlines the fact that in Web 2.0, the users have more power than ever before in the short history of the Internet. Sites like Digg.com depend “on its users to act as contributors and editors” (CNN, 2007). Consequently, Macworld’s Editor’s Notes article on the Digg.com case was titled “Digg users show who’s the boss in Web 2.0 world” (Moren, 2007). Dependency has deepened to a new level, and we are likely to see further ramifications as a result.

One interesting question is whose job it is to moderate the web (Malone, 2007). Is the site responsible for the content that the users add to it or are individual users responsible for the content they add? The on-going and future court cases will set precedents that will perhaps clarify the distribution of responsibilities for the content the users put to the sites (Indiantelelevision.com Team, 2007).

The moderation issues do not end with copyright-related matters, however. We also face questions like how ethical it was to show the hanging of Saddam Hussein whose execution was uploaded to YouTube, Google video and Revver.com only hours after the actual execution. Some clips were unauthorized, grainy mobile phone footage while others were from Iraqi and Arabic television channels (Red Herring Staff, 2006). The cell phone clip appears in various versions in YouTube. One of them, named Saddam, had been viewed 705 555 times by June 24, 2007, and had 1600 ratings, averaging 4 out of 5 stars. It had been marked as favorite 2254 times.

YouTube does offer tools for flagging a video inappropriate. “Graphic or gratuitous **violence** [bolding in the original, not added] is not allowed. If your video shows someone getting hurt, attacked, or humiliated, don’t post it” (YouTube.com, 2007b). The clips flagged inappropriate are reviewed by YouTube administrators within 24–48 hours and removed “immediately” if the content violates the Terms of Use (YouTube.com, 2007a). Apparently, the Saddam Hussein clip does not violate the terms, or nobody has flagged the video as inappropriate since it has been there since December 30, 2006. One might be excused for thinking that somebody getting killed by hanging would come under the category of showing “somebody getting hurt,” but as the users are not offended by the video, it stays. The users are the boss.

In a similar vein, when a Virginia Tech graduate student filmed a 41-second video of the Virginia Tech massacre in April 2007 with his cell phone camera and sent it to CNN, he was then invited to narrate the recording on CNN (Ponn, 2007). The video and narration, taken from CNN broadcast, is (as expected) available in YouTube. While no graphic violence is shown on the clip, one might wonder how the families of the victims feel about it.

Thus, the public executions of yesterday have become both global and virtual in the world of user-generated content (Brea, 2007). Many major news outlets refused to show the Saddam Hussein execution, but citizen journalists felt no such compulsion. We might set age limits on movies but all this is freely available in the net as any underage person in practice can access any YouTube content.

Only the communities can police themselves in these types of issues as no human administrator or even an army of human administrators can moderate the flood of material rolling into the services. In any case, removing the videos after they have been made public means that they will pop right up somewhere else as they already have spread around the hard disks of the world.

On the other hand, citizen journalism (Outing, 2005) has a more positive side to it as well. For instance, Abu Ghraib prisoner abuses and other similar cases have been brought to the consciousness of the world through cell phone pictures. Recently, a cell phone film of campus police repeatedly tasing a Middle Eastern student made it to the YouTube and is now being used as key evidence in the lawsuit (Ponn, 2007). Nobody can any longer control the media although some countries, such as China and Saudi Arabia, continue their attempts to do so. Perhaps Web 2.0 will make the 1960s slogan “people have the power” finally come true.

The services that are facing new challenges because of the user-generated content are certainly not alone. For instance, controlling political image has gotten a lot harder as the presidential candidates for 2008 US presidential election are finding out. An unauthorized video clip “Crush on Obama” had more than 55 000 views the first day it was posted to YouTube (Figure 10 shows one version of the clip). In the video, a half-naked young lady lip-synchs Obama’s praises. While Obama camp is uncertain of what to think about it all—the clip after all does not in any way insult the candidate, in fact, she praises him as “the best candidate”, but there is the little matter of her being half-naked—Ms. Clinton was probably not amused with the 1984 treatment she was given in March, 2007. While the campaigners are happy with the low-cost possibilities of the services like YouTube, they are finding out that it is difficult to control the message in them. (Parsons & McCormick, 2007)

Moreover, the US army has wakened to the threats posed by user-generated content. US Defense Department blocked the access on its computers and networks of the soldiers in Iraq to 13 social networking web sites in May, 2007. Included were MySpace and YouTube. Besides bandwidth concerns, the military security issues were quoted as a reason for the ban. (Aun, 2007; Warner, 2007)

However, many speculate that the Abu Ghraib debacle is behind the blocking as military is trying to control its image. The policy change also means that “troops must now have approval from their commanding officer before posting to blogs or sending e-mails” (Warner, 2007). For the grunts, however, this means that their “porn tube”, BitTorrent, is no longer available, either (Warner, 2007).

Boyd (2007a) discusses another, partially parallel aspect of the military blocking the 13 social networking sites in her blog entry. She has found that Facebook users are typically better educated while MySpace users are lower socio-economic classes (see also Boyd, 2007b). Typically, officers with their college background use Facebook while less educated grunts use

MySpace. The military only blocked MySpace but not Facebook. Furthermore, she notes that before the ban took place, the previously “extremely pro-war, pro-guns, anti-Arab, anti-Muslim, pro-killing, and xenophobic” soldiers were starting to question the war in Iraq and wondering what exactly they were doing there in their MySpace profiles. Potential recruits see these statements, and this might be an important reason for the ban of MySpace. (boyd, 2007a)

Reliability of user-generated content is yet another concern expressed in many different contexts. It has been claimed that 90% of the videos of Saddam Hussein’s execution appearing in Google Video, YouTube and Revver were, in fact, fakes (Cashmore, 2006).

Similar concerns of trustworthiness span all content types, not only videos. Academic world is trying to make up its mind if it is proper to quote Wikipedia in academic work, in spite of its free editing policies and openness to all contributions. The openness and community editorial control in fact maintain quality, and when conflicts arise, the community members can solve the disputes within the talk pages (Viegas et al., 2007).

In contrast, when a blog entry raises discussion, it does not necessarily take place in the blog comment thread only but also expands to other blogs, discussion forums, and even personal emails (boyd, 2007b). Thus, it is much harder for a reader to learn about the controversy, let alone form a big picture of the dispute.

Casual users only occasionally pause to question the trustworthiness of information for example in discussion forums or blogs, or the authority behind the words (Ovadia, 2007). However, the media itself plays a role in determining what kind of content there will be even though the readers might expect otherwise. Although blog entries are informal in style and not as tightly composed as research articles, some readers take them as final articles in addition to taking them out of the blogging context (boyd, 2007b). Blog entries should not be taken as the final word since the main objects for blogging are to clarify one’s thoughts and express opinions possibly not shared by others (Nardi et al., 2004).

While conventional newspapers have strict editorial policies to confirm the reliability of their news sources, such policies do not exist on the Internet. To be able to judge the veracity of an author, reputation systems have been proposed (Resnick, Kuwabara, Zeckhauser, & Friedman, 2000). A reputation mechanism can be either user-driven, that is, based on user-to-user comments or ratings, or content-driven, that is, based on the quality of the contribution.

Researchers at the University of California in Santa Cruz have shown that it is possible to create a content-based reputation system for Wikipedia where the pervasiveness of one’s edits increases and their removal lowers one’s reputation (Adler & Alfaro, 2007). The longer your words stay in place on the Wikipedia page, the higher reputation you have in the community. The reputation information can be shown as a shadowing color on top of what you write, and the shade can change based on timely changes in the reputation. However, such a mechanism is impossible in the scale of the whole Internet. Furthermore, writing on some contentious questions might lead to one’s reputation getting lower in spite of high-quality contributions in such a system. With CIA and Vatican, among others, changing Wikipedia articles to represent their views might mean that a reliable source would in fact appear unreliable because some party has loaded interest in the topic of the article. In any case, the reliability weakness probably means that citizen journalism will continue to co-exist with mainstream media also in the future.

5.4 Approaches to textual user-generated content

The key to getting users submit content is, of course, making tools available for them and making those tools easy enough to use so that anybody who wants to contribute can do so. We do not focus on such contributions as photos, videos, or links here but concentrate instead on textual contributions. Tags are discussed elsewhere (Section 4.4) because of their current popularity, role in navigation, and the fact that they are very short (typically one or a couple of words) in comparison to the textual contributions discussed here.

We are not interested in the technical aspects, such as HTML forms, with which the textual contributions are technically made, but in the ways the contribution is presented and the context inherent to the approach. Different approaches to collecting and presenting user-generated content include discussion forums, blogs, wikis, reviews, and comments, and each of them creates special context for the contributions. We discuss their defining characteristics and show examples of their use in Web 2.0 services in this section. Table 2 on page 4 shows an overview of the functionality included in the eleven example services.

Different approaches offer different possibilities for self-expression. Users rarely receive any direct benefit from contributing text to a service—beyond having a large audience. The popularity of contributing content, however, shows that there is a social order for such means of self-expression. It seems that Shakespeare is right¹: all the Internet is a stage today, and all the men and women seem to have a word to say. A popular book gets hundreds of reviews in Amazon.com. Blogs, too, offer unheard-of possibilities for self-expression, and millions of bloggers are regularly making new entries in their journals.

Any social web service relies on and develops through user-contributed content. The contribution methods provided shape and define the way the community works and what it offers to the users. A Web 2.0 site is a combination of its content—content here includes all the information available to the site, thus including both explicitly and implicitly collected data—and tools for utilizing that content, such as recommendations, reviews, and rankings.

5.4.1 Discussion forums

Discussion forum, also know as Internet forum, web forum, message board, discussion board, discussion group, discussion forum, bulletin board, or simply forum (Wikipedia, 2007f), is the traditional and still popular way to organize discourse on the Internet. A forum typically has set categories that the average users cannot modify. The users post a topic to the certain category. After a topic is posted, other users can comment on it, and eventually the thread of comments generates a hierarchical tree structure. In some forums, the postings are moderated. While usually operating with pseudonyms, posting a topic often requires registering to the service. However, there are several different approaches: some forums require registration even for reading the postings while others allow posting without registration.

Typically, a larger percent of the users never post a topic (Nonnecke & Preece, 2000). Majority of the users in any given social web sites are probably non-contributing anyway if we consider only explicit contributions, such as postings.

Discussion forums have taught the Internet users the essential skills for using web applications because most of the postings take place through HTML-forms. Moreover, users have gained understanding for general netiquette by using forums.

¹ “All the world's a stage, And all the men and women merely players.” –From *As We Like It*.
<http://www.enotes.com/ayli/q-and-a/what-does-quote-all-worlds-stage-all-men-women-2475>

Although discussion forums are today facing serious competition from blogs, wikis, and commenting, they still have a strong position in Web 2.0 services. In addition, they are used in numerous and popular discussion-based services, such as Experts-Exchange and Slashdot. In an interesting example of peaceful co-existence, Wikipedia has adopted several discussion forums (for instance, Village pump, Help desk, User talk, and Wikipedia guidelines) to be used in general discussion about the site while each article has its own discussion in the talk page of the article. Last.fm has full-blown discussion forums for music and bands. Both Flickr and YouTube have group discussions as a core method of discourse between the users. Even Amazon allows its users to discuss its products in a discussion forum.

5.4.2 Blogs

Web logs or blogs are like public web diaries where the author uses blogging software to post dated topics that include text, images and links to another resources (Li, Xu, & Zhang, 2007). The readers of the blog can then comment these topics. Discussion develops in two ways. First, the reader can comment a certain post using commenting functionality of the blog. Second, the person who reads the blog writes an entry to his or her own blog and then links to it in the original blog. When the comments are posted to the original blog, the topic and related comments generate a hierarchical conversation similar to a discussion forum. The second way generates so-called blogosphere where each blog represents a node of the blogosphere and the links to the other blogs represent connections between the nodes, thus creating an interconnected network of blogs. In effect, the blogosphere encompasses all blogs as a community or social network (Wikipedia, 2007c).

A clear difference between discussion forums and blogs is that with blogs, only the authors can post new blog entries while with discussion forums, all members can typically post new topics. In a blog, there is a central person or entity, such as Last.fm personnel, who decides the topics and who is not limited by any pre-set categories as with forums.

Blogs started to become common in 2000 with the advent of such simple-to-use software as Blogger, LiveJournal and EditThisPage. In and onto themselves, blogs do not offer much new in comparison to a frequently updated homepage, but the new approach with simple updating—no HTML required—has brought net-authorship within everybody's reach. Virtually anybody who knows how to browse the Internet can now create a blog and thus a web community. (Weiss, 2005)

While there are sites dedicated to hosting blogs, also seven of the eleven sites studied for this paper had blogs. While blogs are a small part of the whole in some services, for such services as MySpace and Friendster, they constitute a significant part of the total service. For the likes of Last.fm and Del.icio.us, on the other hand, their blogs are used for telling the community what is going on in the service and the bloggers are the employees of the services. However, Last.fm users have “journals” where others can comment the user's dated entries—a blog by any other name.²

Technorati as a service is a search engine for searching blogs and the tags associated with the blog entries. It rates blogs using “authority” which is counted as the number of blogs linking to a blog in the last six months (Technorati.com, 2007). In April 2007, Technorati indexed over 75 million blogs.

² “What's in a name? That which we call a rose, By any other name would smell as sweet.”
From *Romeo and Juliet* (II, ii, 1-2)
<http://www.enotes.com/shakespeare-quotes/what-s-name-that-which-we-call-rose>

Blogs have also caused social uproar. In China, a sexually explicit blog caused a few raised eyebrows in 2003 while, naturally enough, the blog entries were published as a book in France. Today, the female blogger in question has moved to podcasting with sounds of her sexual encounters (The Singapore Internet Research Center, 2005) and material on her can also be found in YouTube. “Despite the Chinese government’s aggressive attempts at controlling publicly available information on her, her name is still one of the most popular searches on Chinese search engines” (The Singapore Internet Research Center, 2005).

Such things are not only the right of the less-liberal societies as China, however, as the case of the Washingtonienne showed. A congressional staff assistant blogged in detail about her lively sex life in 2004 (Wikipedia, 2007o) and caused wide interest with articles about her appearing even in the Washington Post. She also went on to write a book that, naturally enough in the world of Web 2.0, is available in Amazon.com.

Thus, while blogs can be used for informing the community of developments, as with the blogs of Last.fm and Del.icio.us, they seem to answer very well to the social order for a platform for self-expression. Technically easy, they allow anybody to communicate their thoughts to others. As pointed out by boyd (2007b), though, blogs should only be considered work-in-progress and not be taken out of blogging context. That, however, is not how the real world necessarily operates.

5.4.3 Wikis

Wiki is software for collective document writing that was first developed by Ward Cunningham in 1995. Wikipedia’s method to arrange interaction differs from the call-and-response interaction used in discussion forums and blogs (Weiss, 2005). In Wikipedia, everyone can add a new topic or article as it is called in Wikipedia and other users can edit all the topics. The other users do not respond to the topic by generating a separate comment. Instead, they change the original topic by adding their piece of information or by modifying or deleting existing information. In some cases, however, the access to the topics is restricted to special moderator users.

Wikis are, in effect, a clear example of collaborative content creation and collective intelligence. Wikipedia, for instance, consists of fragments of information that “are ultimately weaved into a whole” (Weiss, 2005). Wiki provides the tools for contribution and the rest is up to the community. The idea is that truth emerges from a consensus in the community (Weiss, 2005). However, their openness leaves them open also to vandalism and even governmental manipulation. This has led to discussion whether some of Wikipedia’s “mature” articles should be “frozen” to protect them (Weiss, 2005).

Wikipedia, the most famous example of a Wiki, was launched in 2001 as an open encyclopedia. English-language Wikipedia has currently 1 842 067 articles (June 19, 2007) while Encyclopaedia Britannica only had 65 000 articles in 2005 (Weiss, 2005).

Participation in Wikipedia may sometimes reach surprising forms. Although Wikipedia is not a news forum, it has contributors with timely information. For instance, a broadcast on FOXNews.com points out that an anonymous contributor posted information on a killing even before the police knew about it:

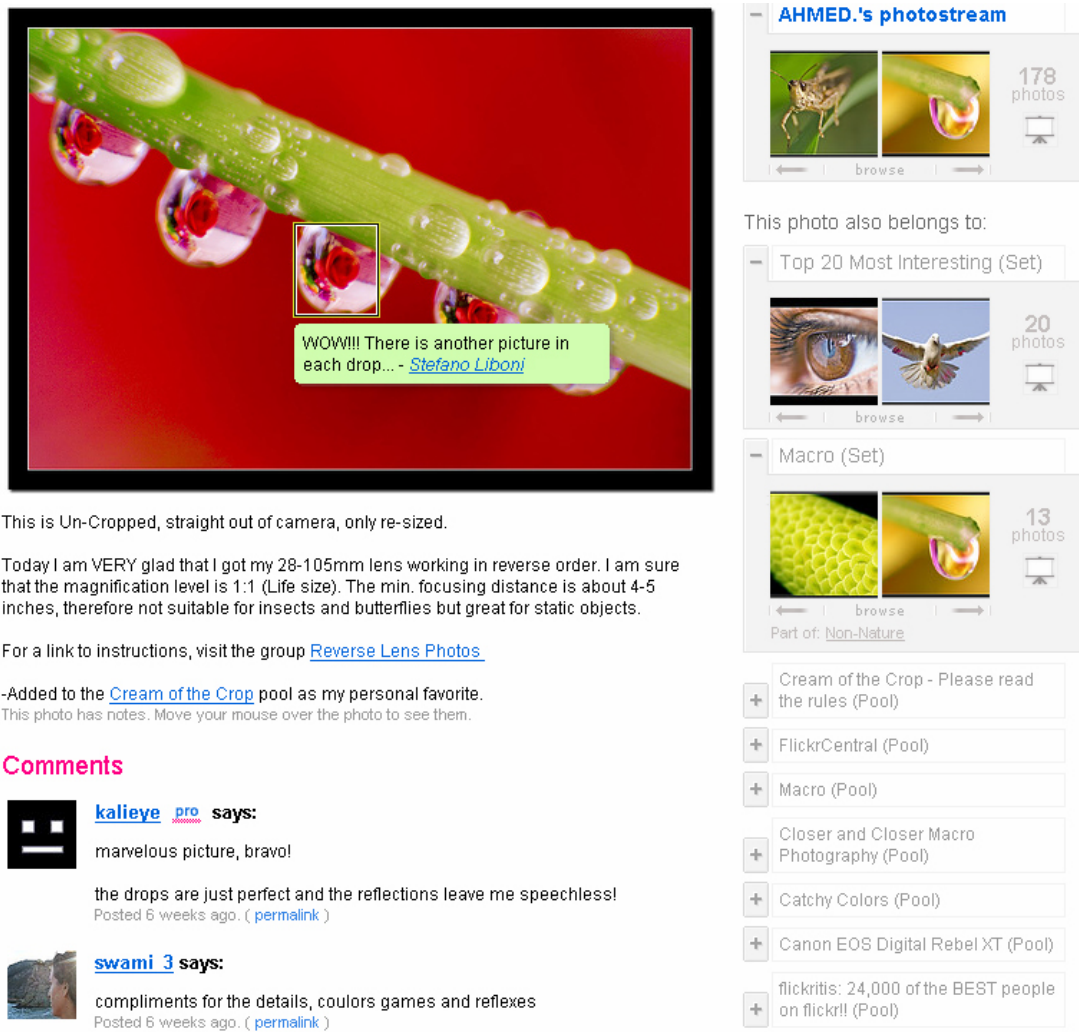
“An anonymous user operating a computer traced to Stamford, Conn.—home to World Wrestling Entertainment—posted an entry to pro wrestler Chris Benoit’s biography on Wikipedia.org announcing the death of his wife Nancy at least 13 hours before police in suburban Atlanta said they found her body along with her husband’s and that of their 7-year-old son.” (Bachelor, 2007)

While Wikis are often stand-alone services, such as Wikipedia, Wikis are also starting to pop up as features in the Web 2.0 sites. Amazon has recently added Amapedia (beta) to the tools that the users have available to comment on the site's products. Amapedia is a Wiki with a tagging functionality added in. In addition, Last.fm uses a Wiki for artist descriptions in the site.

5.4.4 Commenting

Commenting refers to making short remarks or annotations in relation to something, such as a blog posting. Comments are not one-word tags but they are not really full-blown reviews or articles, either. They are typically used as responses to something else. Without the context, they would mostly be unintelligible.

Users can embed comments to freely definable areas in the photos in Flickr (Figure 21) that can then be read by moving the mouse cursor over the defined area. Also, when posting a photo, the user can add longer comments to the picture. Other users can also comment the photos with the comments shown below the poster's text.



AHMED's photostream

178 photos

browse

This photo also belongs to:

Top 20 Most Interesting (Set)

20 photos

browse

Macro (Set)

13 photos

Part of: [Non-Nature](#)

+ Cream of the Crop - Please read the rules (Pool)

+ FlickrCentral (Pool)

+ Macro (Pool)

+ Closer and Closer Macro Photography (Pool)

+ Catchy Colors (Pool)

+ Canon EOS Digital Rebel XT (Pool)

+ flickritis: 24,000 of the BEST people on flickr!! (Pool)

This is Un-Cropped, straight out of camera, only re-sized.

Today I am VERY glad that I got my 28-105mm lens working in reverse order. I am sure that the magnification level is 1:1 (Life size). The min. focusing distance is about 4-5 inches, therefore not suitable for insects and butterflies but great for static objects.

For a link to instructions, visit the group [Reverse Lens Photos](#).

-Added to the [Cream of the Crop](#) pool as my personal favorite.
This photo has notes. Move your mouse over the photo to see them.

Comments

kalieye pro says:
marvelous picture, bravo!
the drops are just perfect and the reflections leave me speechless!
Posted 6 weeks ago. ([permalink](#))

swami 3 says:
compliments for the details, colors games and reflexes
Posted 6 weeks ago. ([permalink](#))

Figure 21. Flickr offers many ways to comment the photos.

In the same vein, Technorati allows users who post a link to WTFs (Where's the Fire) to write short comments about the link. Last.fm, on the other hand, allows other users to comment each user's journal entries. The Last.fm journals are anyway much like blogs where

commenting is normal. YouTube allows users to comment the video clips in “Comments & Responses” but as the name implies, the other users can respond, that is, comment, each other’s comments, so this section is already close to a discussion forum. MySpace allows commenting in various places, from user profile pages to movies and so on.

5.4.5 Reviewing

Reviewing refers to a critique, evaluation, or report about something, such as song or book. Whole sites, such as ePinions.com, have grown around customer reviews. Again, we are talking about collective intelligence as the reviews offer us the collective experiences of other users to guide our decisions about the items reviewed.

Amazon is big on “Customer Reviews” and in many cases they take the most space in the product page. It is not rare to have books with hundreds of reviews in the service, and consequently Amazon has recently added filtering and searching tools for finding the reviews relevant to the user. From the users’ point of view, the customer reviews are seen as information about the contents of the book, and especially the longer reviews are valued. In a field study of Amazon shopping, though, the participants also compared the reviewer’s needs and expertise against their own situation before selecting what to buy (Leino & Rähä, 2007).

Although reviewing is not strongly present in our eleven Web 2.0 sites, it is eminently part of the collective intelligence and user-generated content wave on which Web 2.0 rides. As with ePinions.com, reviewing is commonly present in the sites dedicated to reviewing and in numerous online stores, such as Amazon’s competitor in the book market, Barnes and Noble’s online store (<http://www.barnesandnoble.com/>).

6 Conclusion

While no clear-cut definition of Web 2.0 exists, most commentators seem to agree that it is not about any new technology but rather a new paradigm, a new understanding and approach to what has in fact been long available. Today's web services emphasize collective intelligence for content generations, as in Wikipedia and in Amazon's product reviews, and for generating recommendations, social navigation, and personalization of services. Today's web is about user-generated content and social networking. Neither is new but now the services are providing the tools necessary for generating content without any technical understanding and ways to develop networks and stay in touch with friends and relatives in addition to making new ones. With such services as LinkedIn, even career opportunities are moving online.

Web 2.0 is also about staying aware and finding items of interest from millions of possibilities. Many sites incorporate tools for showing where the action is and what others are interested in with rating and ranking services thrown in. Collaborative filtering allows the services to use user profiling to find sub-groups from communities that correspond to our own preferences to make various types of recommendations for us.

On the down side, constant profiling and information collecting are eroding our means of protecting our privacy. With so many different services and different approaches to privacy, our means of understanding the future privacy implications of our today's actions are limited. Regrettably, protecting privacy does not always make business sense to Web 2.0 services. "Buyers beware" is becoming "users beware." Further studies of privacy are necessary to empower designers to develop privacy-sensitive systems and for users to protect themselves.

User-generated content has also brought its own problems of ownership and copyright. While presidential candidates seem powerless to control their image building, photos and video clips of normal people doing things they would rather forget about also pop up in Web 2.0 services. It is probably no co-incidence that various countries are increasingly blocking such sites as Wikipedia and YouTube as user-generated content proliferates and challenges the powers that be³.

We are still in the middle of the Internet paradigm shift and it is not clear how the world will be after a new equilibrium is reached. For instance, user-generated content and traditional news agencies are learning to cope with each other as the users are learning to make up their minds about what information to consume in today's media world (Noyes, 2007). Every day there is news about different developments in the world of Web 2.0 as cited in the numerous examples in this report. The borders of what is legal and acceptable are being defined.

The fast-growing Web 2.0 services are part of today's business world. When YouTube broke the limit of 20 million visitors, Google acquired it four days later in for \$1.65 billion dollar

³ http://en.wikipedia.org/wiki/The_powers_that_be

deal. Now Facebook and Digg have also passed that magical marker and Facebook considers itself to be worth 8 billion US dollars (Utter, 2007). With huge amounts of money changing hands, how much profit can Web 2.0 services generate in the end remains an open question. Different ways to monetize the services in addition to advertising are being developed and tested.

Some of the features, such as user profiles, blogs, wikis, and mash-ups, of Web 2.0 are already entering the business world under the name of Enterprise 2.0 (Weisman, 2007). At the recent Enterprise 2.0 conference at Boston, such companies as IBM Corp., Microsoft Corp., Cisco Systems Inc., and SAP AG were demonstrating new collaboration products that are familiar to us from Web 2.0. While some dismiss the move as “software looking for a business application” and warn that ROI (return on investment) is missing from the sales pitches (Weisman, 2007), the trend does underline the power and popularity of social web.

Tomorrow’s world—both online and offline—is partially being formulated in the world of Web 2.0 services. Perhaps our old ways of finding music to listen to, choosing books to read and selecting movies to watch are being irrevocably changed. Our ways of finding information about world affairs are being changed. There are so many channels that we need collective intelligence to select whom to listen to. Perhaps even the way we make contacts with other humans is changing for good. Will we even date anybody without making a search on them nowadays? Are there videos about them in YouTube? How is their Facebook profile? What are they bookmarking? What are they writing about in their blog and how much authority their blog has?

For the researchers of human computer interaction and interactive technologies there is a plethora of things to discover. While the research on recommender systems and especially their technological advances flourishes, much less is being done to study the means of interaction and linkages between different media, for instance, mobile phones, in Web 2.0. The services demand new insight into motivational factors, trust building, and member oversight as well as privacy and security.

We need to study and understand the full implications of what is happening today to build a tomorrow that serves our need and interests and still protects our privacy. If we as the HCI research community are not at the helm of this development, then who is, and whose interests are being served?

References

- Accenture. (2007). *User-generated content is top threat to media and entertainment industry, Accenture survey finds*. Retrieved May, 23, 2007, from http://accenture.tekgroup.com/article_display.cfm?article_id=4534
- Adler, B. T., & Alfaro, L. d. (2007). A content-driven reputation system for the wikipedia. *WWW '07: Proceedings of the 16th international conference on World Wide Web*, Banff, Alberta, Canada. 261-270. from <http://doi.acm.org/10.1145/1242572.1242608>; <http://www.soe.ucsc.edu/~luca/papers/07/wikiwww2007.pdf>
- Ahern, S., Eckles, D., Good, N. S., King, S., Naaman, M., & Nair, R. (2007). Over-exposed?: Privacy patterns and considerations in online and mobile photo sharing. *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, San Jose, California, USA. 357-366. from <http://doi.acm.org/10.1145/1240624.1240683>
- Ali, R. (2007). *NAB: GoogleTube Close To Its "Claim Your Content" Filtering System*. Retrieved June 18, 2007, from <http://www.paidcontent.org/entry/419-nab-googletube-close-to-its-claim-your-content-filtering-system>
- Andersen, B. L., Jørgensen, M. L., Kold, U., & Skov, M. B. (2006). iSocialize: Investigating awareness cues for a mobile social awareness application. *OZCHI '06: Proceedings of the 20th conference of the computer-human interaction special interest group (CHISIG) of Australia on Computer-human interaction: design: activities, artefacts and environments*, Sydney, Australia. 7-14. from <http://doi.acm.org/10.1145/1228175.1228181>
- Andrews, R. (2007). *French Sport Bodies, Cherry Lane Join YouTube Copyright Suit*. Retrieved June 18, 2007, from <http://www.paidcontent.org/entry/419-french-sport-bodies-elvis-rightsholder-join-youtube-copyright-suit/>
- AP. (2007). *'Sicko' Makes Unauthorized Debut on YouTube*. Retrieved June 25, 2007, from <http://www.technewsworld.com/story/q8vSbDH09mkLud/Sicko-Makes-Unauthorized-Debut-on-YouTube.xhtml>
- Aun, F. J. (2007). *Soldiers Lose MySpace, YouTube Privileges*. Retrieved May 27, 2007, from <http://www.macnewsworld.com/story/social-networking/57368.html>
- Bachelor, B. (2007). *Web time stamps indicate Nancy Benoit's death reported on web at least 13 hours before police found bodies in her home*. Retrieved July 28, 2007, from <http://www.foxnews.com/story/0,2933,287194,00.html>
- Barlas, P. (2007). *Online Video Site Break.com May Be First To Charge Fee*. Retrieved June 18, 2007, from <http://www.investors.com/editorial/IBDArticles.asp?artsec=17&artnum=1&issue=20070525>

- Blanchard, A. L., & Markus, M. L. (2004). The experienced sense of a virtual community: characteristics and processes. *SIGMIS Database*, 35(1), 64-79. from <http://doi.acm.org/10.1145/968464.968470>
- boyd, d. (2006). Friends, Friendsters, and Top 8: Writing community into being on social network sites. *First Monday*, 11(12), December 2006. Retrieved June 18, 2007, from http://www.firstmonday.org/issues/issue11_12/boyd/
- boyd, d., & Heer, J. (2006). Profiles as Conversation: Networked Identity Performance on Friendster. *Proceedings of the Hawai'i International Conference on System Sciences (HICSS-39), Digital Media: Content and Communication Track. Persistent Conversation Minitrack*. Kauai, HI., 39. From <http://www.danah.org/papers/HICSS2006.pdf>; <http://csdl2.computer.org/comp/proceedings/hicss/2006/2507/03/250730059c.pdf>
- boyd, d. (2007a, June 24). Viewing American class divisions through Facebook and MySpace. Message posted to <http://www.danah.org/papers/essays/ClassDivisions.html>; http://www.zephorias.org/thoughts/archives/2007/06/24/viewing_america.html
- boyd, d. (2007b, July 25). Responding to responses to: "Viewing American class divisions through FaceBook and MySpace". Message posted to <http://www.danah.org/papers/essays/ResponseToClassDivisions.html>
- Braiker, B. (2006). *The Year of the Widget?* Retrieved June 25, 2007, from <http://www.msnbc.msn.com/id/16329739/site/newsweek/page/0/>
- Brea, J. (2007). *Execution, Voyeurism in the Age of Youtube*. Retrieved June 25, 2007, from <http://worldnews.about.com/b/a/257509.htm>
- Bricklin, D. (2000). *The Cornucopia of the Commons: How to get volunteer labor*. Retrieved June 25, 2007, from <http://www.bricklin.com/cornucopia.htm>
- Brightcove.com. (2007). *Liberate Your Video-Build Your Business*. Retrieved June 18, 2007, from http://corp.brightcove.com/content_owners/network/index.cfm
- Broache, A. (2007, April 16). Net radio operators lose appeal over fees. Message posted to http://news.com.com/8301-10784_3-6176631-7.html
- Butterfield, S. (2006, August 29). Geotagging - one day later. Message posted to <http://blog.flickr.com/en/2006/08/29/geotagging-one-day-later/>
- Cashmore, P. (2006). *Saddam's Execution Video Makes it to Google Video, YouTube, Revver*. Retrieved June 25, 2007, from <http://mashable.com/2006/12/30/saddams-execution-video-makes-it-to-google-video-youtube>
- CNN. (2007). *Digg.com in 'hacking' copyright row*. Retrieved June 18, 2007, from <http://edition.cnn.com/2007/TECH/internet/05/03/digg.copyright/>
- Cosley, D., Frankowski, D., Kiesler, S., Terveen, L., & Riedl, J. (2005). How oversight improves member-maintained communities. *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*, Portland, Oregon, USA. 11-20. from <http://doi.acm.org/10.1145/1054972.1054975>
- Dieberger, A., Dourish, P., Höök, K., Resnick, P., & Wexelblat, A. (2000). Social navigation: techniques for building more usable systems. *interactions*, 7(6), 36-45. from <http://doi.acm.org/10.1145/352580.352587>
- Dourish, P., & Bellotti, V. (1992). Awareness and coordination in shared workspaces. *CSCW '92: Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, Toronto, Ontario, Canada. 107-114. from <http://doi.acm.org/10.1145/143457.143468>

- Espinosa, A., Cadiz, J., Rico-Gutierrez, L., Kraut, R., Scherlis, W., & Lautenbacher, G. (2000). Coming to the wrong decision quickly: why awareness tools must be matched with appropriate tasks. *CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems*, The Hague, The Netherlands. 392-399. from <http://doi.acm.org/10.1145/332040.332463>
- Flickr.com. (2007a). *Explore / About Interestingness*. Retrieved June 18, 2007, from <http://www.flickr.com/explore/interesting/>
- Flickr.com. (2007b). *Share*. Retrieved June 18, 2007, from <http://www.flickr.com/tour/share/>
- Fuller, T. (2007). *Thailand Bans YouTube*. Retrieved June 25, 2007, from <http://www.nytimes.com/2007/04/05/business/worldbusiness/05tube.html?ex=1333425600&en=f8cc38541477b31e&ei=5124&partner=permalink&exprod=permalink>
- Furnas, G. W., Fake, C., Ahn, L. v., Schachter, J., Golder, S., Fox, K., et al. (2006). Why do tagging systems work? *CHI '06: CHI '06 extended abstracts on Human factors in computing systems*, Montréal, Québec, Canada. 36-39. from <http://doi.acm.org/10.1145/1125451.1125462>
- Geotagging Flickr. (2007, June 13). Current state of geotagging? Message thread posted to <http://www.flickr.com/groups/geotagging/discuss/72157600347253293/>
- Girgensohn, A., & Lee, A. (2002). Making web sites be places for social interaction. *CSCW '02: Proceedings of the 2002 ACM conference on Computer supported cooperative work*, New Orleans, Louisiana, USA. 136-145. from <http://doi.acm.org/10.1145/587078.587098>
- Golder, S. A., & Huberman, B. A. (2006). Usage Patterns of Collaborative Tagging Systems. *Journal of Information Science*, 32(2), 198-208.
- Gross, R., Acquisti, A., & Heinz, H. J. III. (2005). Information revelation and privacy in online social networks. *WPES '05: Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, Alexandria, VA, USA. 71-80. from <http://doi.acm.org/10.1145/1102199.1102214>
- Hammond, T., Hannay, T., Lund, B., & Scott, J. (2005). Social Bookmarking Tools (I): A General Review. *D-Lib Magazine*, 11(4), June 4, 2007 from <http://www.dlib.org/dlib/april05/hammond/04hammond.html>
- Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems* 22(1), 5-53. from <http://doi.acm.org/10.1145/963770.963772>
- Heylighen, F. (2001). *Collaborative Filtering*. Retrieved May 20, 2007, from <http://pespmc1.vub.ac.be/COLLFILT.html>
- Hintikka, K. A. (2007). *Web 2.0 - johdatus internetin uusiin liiketoimintamahdollisuuksiin..* Helsinki: TIEKE Tietoyhteiskunnan kehittämiskeskus ry. Julkaisusarjan raportti 28. April 25, 2007, from http://www.tieke.fi/mp/db/file_library/x/IMG/20815/file/julkaisu_28.pdf
- Indiana University. (2007). *Phishers can use social Web sites as bait to net victims: Informatics study*. Retrieved June, 25, 2007, from <http://newsinfo.iu.edu/news/page/normal/5726.html>
- Indiantelevision.com Team. (2007). *Premier League files lawsuit against Youtube*. Retrieved June 18, 2007, from <http://www.indiantelevision.com/headlines/y2k7/june/june133.php>
- International Trade Administration. (2007). *Safe Harbor*. Retrieved June 26, 2007, from <http://www.export.gov/safeharbor/>

- internetworldstats.com. (2007). *Internet Usage Statistics - The Big Picture*. Retrieved June 25, 2007, from <http://www.internetworldstats.com/stats.htm>
- Johnson, M. (2007). *Habbo Hotel ja Privacy. Esitelmä Tampereen yliopistossa 9.5.*
- Johnson, M., & Toiskallio, K. (2005). Fansites as sources of user research: Case Habbo Hotel. *Proceedings of the 28th Conference on Information Systems Research in Scandinavia (IRIS28)*. Retrieved June 26, 2007, from http://www.soberit.hut.fi/~johnson/Johnson_IRIS_2005.pdf
- Jones, K. C. (2007). *States demand answers about sex offenders on MySpace*. Retrieved June 25, 2007, from <http://www.informationweek.com/news/showArticle.jhtml?articleID=199600284>
- Kangas, P., Toivonen, S., & Bäck, A. (toim.). (2007). *Googlen mainokset ja muita sosiaalisen median liiketoimintamalleja* (VTT tiedotteita - Research notes No. 2369). Espoo: VTT Technical Research Centre of Finland. May 15, 2007, from <http://www.vtt.fi/inf/pdf/tiedotteet/2007/T2369.pdf>
- Kaplan, D. (2007). *MySpace Enhances Copyright Protection Tool; Prevents Repeated Uploading*. Retrieved June 18, 2007, from <http://www.paidcontent.org/entry/419-new-myspace-feature-offers-to-take-down-copyright-violators/>
- Kobsa, A. (2007). *Privacy-Enhanced Web Personalization*. In: Brusilovsky, P., Kobsa, A., and Nejdl, W. (Eds.): *The Adaptive Web*, LNCS 4321, 628-670. Springer-Verlag Berlin Heidelberg, 2007. Retrieved September 12, 2007, from http://dx.doi.org/10.1007/978-3-540-72079-9_21
- Lampe, C., Ellison, N., & Steinfield, C. (2006). A face(book) in the crowd: Social searching vs. social browsing. *CSCW '06: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, Banff, Alberta, Canada. 167-170. from <http://doi.acm.org/10.1145/1180875.1180901>
- Lampe, C., Ellison, N., & Steinfield, C. (2007). A familiar Face(book): Profile elements as signals in an online social network. *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, San Jose, California, USA. 435-444. from <http://doi.acm.org/10.1145/1240624.1240695>
- Lederer, S., Hong, J. I., Dey, A. K., & Landay, J. A. (2004). Personal privacy through understanding and action: five pitfalls for designers. *Personal and Ubiquitous Computing*, 8, 440-454.
- Lee, K. J. (2006). What goes around comes around: an analysis of del.icio.us as social space. *CSCW'06: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, Banff, Alberta, Canada. 191-194. from <http://doi.acm.org/10.1145/1180875.1180905>
- Leino, J., & Rähkä, K. (2007). Case Amazon: Ratings and Reviews as Part of Recommendations. In *Proceedings of the 2007 ACM Conf. on Recommender Systems*, Minneapolis, Minnesota, 137-140. from <http://doi.acm.org/10.1145/1297231.1297255>
- Li, B., Xu, S., & Zhang, J. (2007). Enhancing clustering blog documents by utilizing author/reader comments. *ACM-SE 45: Proceedings of the 45th annual southeast regional conference*, Winston-Salem, North Carolina. 94-99. from <http://doi.acm.org/10.1145/1233341.1233359>
- Liechti, O. (2000). Awareness and the WWW: an overview. *SIGGROUP Bull.*, 21(3), 3-12.
- Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, (January-February), 76-80.

- LinkedIn.com. (2007). *LinkedIn.com Front page*. Retrieved June 18, 2007, from <http://www.linkedin.com/>
- Ma, M., & Agarwal, R. (2007). Through a Glass Darkly: Information Technology Design, Identity Verification, and Knowledge Contribution in Online Communities. *Information Systems Research*, 18(1), 42-67.
- Malone, M. S. (2007). *The First Amendment vs. Patents in Web 2.0*. Retrieved June 18, 2007, from http://abclocal.go.com/wpvi/story?section=sci_tech&id=5269200
- Martinson, J. (2006). *Google faces copyright fight over YouTube*. October 13, 2006. Retrieved June 18, 2007, from <http://business.guardian.co.uk/story/0,,1921154,00.html>
- Masnick, M. (2006). *And Here Come The YouTube Copyright Lawsuits*. Retrieved June 18, 2007, from <http://www.techdirt.com/articles/20060717/1852217.shtml>
- McCarthy, C. (2007). *New MySpace copyright tech turns heads, raises brows*. Retrieved June 18, 2007, from http://news.com.com/New+MySpace+copyright+tech+turns+heads,+raises+brows/2100-1030_3-6183162.html
- McNee, S. M., Albert, I., Cosley, D., Gopalkrishnan, P., Lam, S. K., Rashid, A. M., et al. (2002). On the recommending of citations for research papers. *CSCW '02: Proceedings of the 2002 ACM conference on Computer supported cooperative work*, New Orleans, Louisiana, USA. 116-125. from <http://doi.acm.org/10.1145/587078.587096>
- Meattle, J. (2007, June 26). Digg overtakes FaceBook; Both cross 20 million unique U.S. visitors. Message posted to <http://blog.compete.com/2007/06/20/digg-overtakes-facebook-cross-20-million-visitors/>
- Moore, T. D., & Serva, M. A. (2007). Understanding member motivation for contributing to different types of virtual communities: a proposed framework. *SIGMIS-CPR '07: Proceedings of the 2007 ACM SIGMIS CPR conference on 2007 computer personnel doctoral consortium and research conference*, St. Louis, Missouri, USA. 153-158. from <http://doi.acm.org/10.1145/1235000.1235035>
- Moren, D. (2007, May 02). Digg users show who's the boss in Web 2.0 world. Message posted to <http://www.macworld.com/weblogs/editors/2007/05/digg/index.php>
- Murison, J. (2005). Messageboard topic tagging: user tagging of collectively owned community content. *DUX '05: Proceedings of the 2005 conference on Designing for User eXperience*, San Francisco, California. 5.
- Musser, J., O'Reilly, T., & O'Reilly Radar Team. (2006). *Web 2.0 Principles and Best Practices*. O'Reilly Radar Report, O'Reilly Media, Inc. from http://www.oreilly.com/catalog/web2report/chapter/web20_report_excerpt.pdf
- Nardi, B. A., Schiano, D. J., & Gumbrecht, M. (2004). Blogging as social activity, or, would you let 900 million people read your diary? *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work*, Chicago, Illinois, USA. 222-231. from <http://doi.acm.org/10.1145/1031607.1031643>
- Nonnecke, B., & Preece, J. (2000). Lurker demographics: counting the silent. *CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems*, The Hague, The Netherlands. 73-80. from <http://doi.acm.org/10.1145/332040.332409>
- Noyes, K. (2007). *Slashdot Founder Rob Malda: Looking Beyond Web 2.0*. Retrieved June 25, 2007, from <http://www.technewsworld.com/story/social-networking/57862.html>
- O'Donovan, J., & Smyth, B. (2006). Is trust robust?: an analysis of trust-based recommendation. *IUI '06: Proceedings of the 11th international conference on*

- Intelligent user interfaces*, Sydney, Australia. 101-108. from <http://doi.acm.org/10.1145/1111449.1111476>
- O'Hear, S. (2006, October 31). Brightcove launches a revenue-sharing rival to YouTube. Message posted to <http://blogs.zdnet.com/social/?p=6>
- O'Hear, S. (2007, January 27). YouTube to pay for User Generated Content. Message posted to <http://blogs.zdnet.com/social/?p=75>
- O'Reilly, T. (2005). *What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software*. Retrieved April 10, 2007, from <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html?page=1>
- Outing, S. (2005). *The 11 layers of Citizen Journalism*. Retrieved August 14, 2007, from http://www.poynter.org/content/content_view.asp?id=83126
- Ovadia, S. (2007) Digg.com and Socially-Driven Authority. *Library Philosophy and Practice*, LPP Special Issue on Libraries and Google, June 2007. <http://www.webpages.uidaho.edu/~mbolin/ovadia.pdf>
- Parsons, C., & McCormick, J. (2007). *Obama Groupies, Image Control and the Veracity of Viral Video*. Retrieved June 25, 2007, from <http://www.technewsworld.com/story/social-networking/57878.html>
- Ponn, J. (2007). *Cameras on phones rival digital cousins*. Retrieved June 25, 2007, from <http://www.gainesville.com/apps/pbcs.dll/article?AID=/20070529/LOCAL/705290317/-1/news>
- Preece, J. (2000). *Online Communities: Designing Usability, Supporting Sociability*. West Sussex, England: John Wiley & Sons.
- Prinz, W. (1999). NESSIE: an awareness environment for cooperative settings. *ECSCW'99: Proceedings of the sixth conference on European Conference on Computer Supported Cooperative Work*, Copenhagen, Denmark. 391-410.
- Privacy International. (2007). *A race to the bottom: Privacy ranking of Internet service companies*. Consultation Report. London: Privacy International. June 13, 2007, from <http://www.privacyinternational.org/issues/internet/interimrankings.pdf>; [http://www.privacyinternational.org/article.shtml?cmd\[347\]=x-347-553961](http://www.privacyinternational.org/article.shtml?cmd[347]=x-347-553961)
- Quain, J. R. (2004). *Fine-Tuning Your Filter for Online Information*. Retrieved June 18, 2007, from <http://www.nytimes.com/ref/technology/circuits/03basi.html>
- Raento, M. (2007). Exploring privacy for ubiquitous computing: Tools, methods and experiments. Ph.D. thesis, University of Helsinki, Department of Computer Science. <http://urn.fi/URN:ISBN:978-952-10-3986-7>
- Red Herring Staff. (2006). *Saddam's YouTube Execution*. Retrieved June 25, 2007, from <http://www.redherring.com/Article.aspx?a=20519>
- Resnick, P., Kuwabara, K., Zeckhauser, R., & Friedman, E. (2000). Reputation systems. *Communications of the ACM*, 43(12), 45-48. from <http://doi.acm.org/10.1145/355112.355122>
- Reuters. (2007). *Blocked China Web users rage against Great Firewall*. Retrieved June 25, 2007, from <http://www.reuters.com/article/inDepthNews/idUSPEK21813920070620>
- Revver.com. (2007). *What is Revver?* Retrieved June 18, 2007, from <http://one.revver.com/go/about>

- Rheingold, H. (1998). *The Virtual Community*. Retrieved June 26, 2007, from <http://www.rheingold.com/vc/book/>
- Richards, J. (2007). *Sex offenders can use social sites, say police*. Retrieved July 31, 2007, from http://technology.timesonline.co.uk/tol/news/tech_and_web/the_web/article2137973.ece
- Rivadeneira, A. W., Gruen, D. M., Muller, M. J., & Millen, D. R. (2007). Getting our head in the clouds: toward evaluation studies of tagclouds. *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, San Jose, California, USA. 995-998. from <http://doi.acm.org/10.1145/1240624.1240775>
- Rose, K. (2007, May 01). Digg This: 09-f9-11-02-9d-74-e3-5b-d8-41-56-c5-63-56-88-c0. Message posted to <http://blog.digg.com/?p=74>
- Sarwar, B. M., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based Collaborative Filtering Recommendation Algorithms. *Proceedings of WWW'10: Tenth International World Web Conference*, Hong Kong, China. Retrieved May 20, 2007, from <http://www.www10.org/cdrom/papers/pdf/p519.pdf>
- Sen, S., Lam, S. K., Rashid, A. M., Cosley, D., Frankowski, D., Osterhouse, J., et al. (2006). tagging, communities, vocabulary, evolution. *CSCW '06: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, Banff, Alberta, Canada. 181-190. from <http://doi.acm.org/10.1145/1180875.1180904>
- Silver Smith, C. (2006, April 26). Google Maps for Europe & the rise of GeoTagging. Message posted to <http://www.naturalsearchblog.com/archives/2006/04/26/google-maps-for-europe-the-rise-of-geotagging/>
- STT-IS. (2007). *Kotipornoa levittänyt joutuu maksamaan jättikorvaukset*. February 23, 2007. Retrieved June 25, 2007, from <http://www.iltasanomat.fi/uutiset/kotimaa/uutinen.asp?id=1324732>
- Svensson, M., Höök, K., & Cöster, R. (2005). Designing and evaluating Kalas: A social navigation system for food recipes. *ACM Transactions on Computer-Human Interaction*, 12(3), 374-400. from <http://doi.acm.org/10.1145/1096737.1096739>
- Technorati.com. (2007). *What is Authority?* Retrieved June 25, 2007, from <http://support.technorati.com/faq/topic/71>
- The Singapore Internet Research Center. (2005). *Muzi Mei posts controversial podcast*. Retrieved June 25, 2007, from http://internetinasia.typepad.com/blog/2005/12/podcast_of_sexu.html
- theage.com.au. (2007). *Google considers launching YouTube in SKorea*. Retrieved June 25, 2007, from <http://www.theage.com.au/news/Technology/Google-considers-launching-YouTube-in-SKorea/2007/06/22/1182019342640.html>
- Tom. (2007, July 13). Extended network. Message posted to <http://blog.myspace.com/index.cfm?fuseaction=blog.view&friendID=6221&blogID=287450663>
- Tweney, D. (2007). *Tim O'Reilly: Web 2.0 Is About Controlling Data*. Retrieved June 25, 2007, from http://www.wired.com/techbiz/people/news/2007/04/timoreilly_0413?currentPage=1
- Utter, D. (2007). *Digg, Facebook hit magical mark*. Retrieved June 25, 2007, from <http://www.webpronews.com/insiderreports/2007/06/21/digg-facebook-hit-magical-mark>

- Viegas, F. B., Wattenberg, M., Kriss, J., & van Ham, F. (2007). Talk Before You Type: Coordination in Wikipedia. *Proceedings of the 40th Annual Hawaii International Conference on System Sciences HICSS'40*, p. 78. From <http://doi.ieeecomputersociety.org/10.1109/HICSS.2007.511>
- Warner, B. (2007). *Why didn't the US army ban bloggers before?* Retrieved June 25, 2007, from http://technology.timesonline.co.uk/tol/news/tech_and_web/article1803202.ece
- Weisman, R. (2007). *Social-network sites give businesses ideas for new collaboration*. Retrieved June 25, 2007, from http://www.boston.com/business/personaltech/articles/2007/06/25/social_network_sites_give_businesses_ideas_for_new_collaboration/
- Weiss, A. (2005). The power of collective intelligence. *netWorker*, 9(3), 16-23 from <http://doi.acm.org/10.1145/1086762.1086763>
- Wellman, B., & Gulia, M. (1999). Virtual Communities as Communities: Net Surfers Don't Ride Alone. In P. Kollock, & M. Smith (Eds.), *Communities in Cyberspace* (pp. 167-194). New York: Routledge. Retrieved April 14, 2007, from <http://www.chass.utoronto.ca/~wellman/publications/netsurfers/netsurfers.pdf>
- Westergren, T. (2007). *Dear Pandora Visitor (Open letter)*. Retrieved June 12, 2007, from <http://www.pandora.com/restricted>
- Wikimedia Foundation. (2006). *Privacy policy*. Retrieved June 27, 2007, from http://wikimediafoundation.org/wiki/Privacy_policy
- Wikipedia. (2007a). *Angel investor*. Retrieved July 31, 2007, from http://en.wikipedia.org/wiki/Angel_investor
- Wikipedia. (2007b). *Awareness*. Retrieved May 14, 2007, from <http://en.wikipedia.org/wiki/Awareness>
- Wikipedia. (2007c). *Blogosphere*. Retrieved June 25, 2007, from <http://en.wikipedia.org/wiki/Blogosphere>
- Wikipedia. (2007d). *Collaborative filtering*. Retrieved April 16, 2007, from http://en.wikipedia.org/wiki/Collaborative_Filtering
- Wikipedia. (2007e). *Instant messaging*. Retrieved June 18, 2007, from http://en.wikipedia.org/wiki/Instant_messaging
- Wikipedia. (2007f). *Internet forum*. Retrieved June 25, 2007, from http://en.wikipedia.org/wiki/Internet_forum
- Wikipedia. (2007g). *Mashup (web application hybrid)*. Retrieved June 25, 2007, from http://en.wikipedia.org/wiki/Mashup_%28web_application_hybrid%29
- Wikipedia. (2007h). *Push technology*. Retrieved June 18, 2007, from http://en.wikipedia.org/wiki/Push_technology
- Wikipedia. (2007i). *Rating*. Retrieved June 18, 2007, from <http://en.wikipedia.org/wiki/Rating>
- Wikipedia. (2007j). *Shoutbox*. Retrieved July 27, 2007, from <http://en.wikipedia.org/wiki/Shoutbox>
- Wikipedia. (2007k). *Social presence*. Retrieved May 14, 2007, from http://en.wikipedia.org/wiki/Social_presence
- Wikipedia. (2007l). *Spam*. Retrieved June 18, 2007, from <http://en.wikipedia.org/wiki/Spamming>

- Wikipedia. (2007m). *Statistics*. Retrieved June 25, 2007, from <http://en.wikipedia.org/wiki/Special:Statistics>
- Wikipedia. (2007n). *Use of social network websites in investigations*. Retrieved July 31, 2007, from http://en.wikipedia.org/wiki/Use_of_social_network_websites_in_investigations
- Wikipedia. (2007o). *Washingtonienne*. Retrieved June 25, 2007, from <http://en.wikipedia.org/wiki/Washingtonienne>
- Wikipedia. (2007p). *Web 2.0*. Retrieved June 25, 2007, from http://en.wikipedia.org/wiki/Web_2.0
- Wikipedia. (2007q). *YouTube*. Retrieved June 18, 2007, from <http://en.wikipedia.org/wiki/Youtube>
- Yang, X. (2007, June 19). YouTube Goes International. Message posted to http://blog.washingtonpost.com/posttech/2007/06/youtube_goes_international.html?nav=rss_blog
- YouTube.com. (2007a). *How does YouTube handle inappropriate content?* Retrieved June 25, 2007, from <http://www.google.com/support/youtube/bin/answer.py?answer=55777&ctx=sibling>
- YouTube.com. (2007b). *YouTube Community Guidelines*. Retrieved June 25, 2007, from http://www.youtube.com/t/community_guidelines