

User Generated Meta Data Aggregation in the Context of Web Search and Discovery

Christoph Burgdorfer

A dissertation submitted in partial fulfillment
of the requirements of Sussex University for the
degree of MA Networked Media Environments

September 2007

Part I

Abstract

In the early days of the World Wide Web, search engines used to read so called "Meta Tags" within the website source code in order to determine the content of a page. Nowadays' search engines do not consider those tags as very relevant anymore. Instead they use other, more sophisticated algorithms to determine not only the content but also the context around an Internet site. This dissertation and research paper argues we have to go a step further: The content of a website should not be classified by machines but by the people who use it. Other than machines, humans can determine the context and nature of a website intuitively. With the dawn of the participative Web the time is right to introduce user generated classification of websites. This dissertation discusses the development and rationale behind the proposed way of search and discover of online content.

Part II

Keywords and Definitions

The following terms are discussed within this dissertation:

Firefox Extension - A Firefox extension relates to the web browser part of the Mozilla Foundation's¹ software suite: *Mozilla Firefox*².

The number of extensions has increased significantly over the past years. Firefox extensions are generally used to extend the *Mozilla Firefox* browser's standard functionalities with features such as for example meteorological data within the browser bar, blockage of advertising on websites or utilities and tools to develop websites. The programming language used in Firefox Extensions is typically JavaScript³ and XUL⁴, a XML⁵ descendant to describing user interfaces.

Tagging - Tagging describes the act of adding keywords (tags) to items and can be considered as a form of categorization. Tagged content can be found and filtered more easily. Tagging has become popular through community sites such as Flickr⁶ and del.icio.us⁷ where the users can tag photos or website URLs. A set of tags would be referred to as a "tag cloud" suggesting that the tags describe the object in a blurry way. Tagging can also be referred to as "folksonomy". (Please refer to "Aggregating Meta-Information via User Generated Tags" in section 3.2 for more details.)

Mashup - A mashup is a website that facilitates new content through seamless (re-) combination of existing content i.e. contents of the Web such as text, images or videos are remixed in a new way. Mashups can be seen as a significant example of the participative Web 2.0. They make use of open Application Programming Interfaces (APIs) which are provided for third party web applications generally for free. Typical examples of Mashups are embedded *Google Maps*⁸ or *YouTube*⁹

¹<http://www.mozilla.org>

²<http://www.mozilla.org/firefox/>

³<http://developer.mozilla.org/en/docs/JavaScript>

⁴<http://www.mozilla.org/projects/xul/> (XML User Interface Language)

⁵Extended Markup Language

⁶<http://www.flickr.com>

⁷<http://del.icio.us>

⁸<http://maps.google.com>

⁹<http://www.youtube.com>

videos.

Metadata - Metadata or Meta-Information (Meta: Greek for "after", "beyond", "with") describes in general data which contains information about other data. The described data often includes larger collections of data such as documents, databases or files. Attributes and properties of an object like keywords can be referred to as metadata. In the context of the World Wide Web, metadata describes for example the context, author, date or even the geographical location, the so called geo-location or GeoURL¹⁰ related to a particular website.

User Generated Content (UGC) - User Generated Content refers to various kinds of media content that are created by end-users. This content can include text as well as multimedia such as audio, video and images. Typically user generated content will be provided to a community for free.

The term has become main stream in 2005 and is often used in the discourse of the development of the Internet and its future development (Web 2.0, see below). It can be assumed that there is a connection between the increase of Internet bandwidth, the decrease of data storage costs, the development of sophisticated computer equipment and the user's skills that lead to the popularity of User Generated Content. Amongst the most popular examples of User Generated Content sites are *YouTube*, *Wikipedia*¹¹ or *Flickr* but also social networking sites as *Facebook*¹² or *MySpace*.¹³

Semantic Web - The concept of the Semantic Web bases on a suggestion of Sir Tim Berners-Lee, founder of the World Wide Web (WWW). The Semantic Web is considered to be an extension of the WWW to define the semantics of the content formally. Information in the Web should not only be human readable but also machine readable and processable in order to enable programs to operate on the data. This allows machines to process information on the base of the semantic content as opposed to processing on the the notation. Such "data about data" is often being referred to as *Metadata*.

Web 2.0 - Web 2.0 is a term for a number of interactive and collaborative phenomena in the Internet, particularly the WWW. The term has become popular in 2003 through Tim O'Reilly, the owner of the publishing company *O'Reilly Media*. 2.0 refers to

¹⁰<http://geourl.org/add.html>

¹¹<http://www.wikipedia.org>

¹²<http://www.facebook.com>

¹³<http://www.myspace.com>

the second generation of web-based communities and hosted services such as social networking sites or wikis which facilitate collaboration and sharing between users. *"Web 2.0 is the business revolution in the computer industry caused by the move to the Internet as platform, and an attempt to understand the rules for success on that new platform."*[25]

The term "Web 2.0" also incorporates aspects of the Semantic Web: social networks such as FOAF¹⁴ and XFN¹⁵, the development of folksonomy, weblogs, social bookmarkings and Wikis. The approach however is different: The folksonomy or tagging of Web 2.0 applications is typically developed by the users as opposed to the authors in the Semantic Web.

Critics of the term "Web 2.0" such as Sir Tim Berners-Lee claim that the *"Web 2.0 is of course a piece of jargon, nobody even knows what it means"* and that *"Web 1.0 was all about connecting people"*. [8] He claims that the idea of the Web has always been accessing, creating and publishing content. The developers of the popular Web browsers however did not adapt this vision which is the reason why the Web stayed a "passive medium" for a long time. Only the introduction of collaborative techniques such as Wikis facilitated Sir Tim Berners-Lee's idea.

¹⁴<http://www.foaf-project.org> (Friend of a Friend)

¹⁵<http://gmpg.org/xfn/> (XHTML Friends Network)

Part III

Acknowledgements

I would like send my thanks to everybody who helped me during this dissertation including my friends Abraham Alaka, Alex Nako, Andreas Hunziker, Andrea Cruz, Georg Fasching, Marco Klingmann, Mario Aeby, Stephen Breakey.

The Ravensbourne College of Design and Communication staff including Armin Medosch, Fiona Heavey, James Smith, Jamie King, Jim Wood, Karel Dudesek, Lisa Haskel, Martin Schmitz, Neal White and my fellow students.

Everybody who beta-tested my application.

The MINICK Crew in Zurich, Switzerland and London, United Kingdom.

My girlfriend Katia, my mother and family.

Contents

I	Abstract	2
II	Keywords and Definitions	3
III	Acknowledgements	6
IV	Main Part	11
1	Introduction	11
1.1	Context and Scope	11
1.2	Problem Definition and Objectives	12
1.3	Methodology	12
1.4	Hypothesis	14
2	Collecting Data	14
3	Course of Investigation	17
3.1	The Internet and Democracy	17
3.2	Aggregating Meta-Information via User Generated Tags	17
3.3	The Badge	20
3.4	The Firefox Extension	21
4	Developing the Concept	22
4.1	Considerations	23
4.1.1	User Database and Logins	23
4.1.2	Critical Mass	24
4.1.3	Distribution	24
4.1.4	Quality Assurance and Abuse Prevention	26
4.2	Related Existing Projects	27
4.2.1	del.icio.us	27
4.2.2	Stumbleupon.com	29
4.2.3	Technorati.com	30
4.3	Challenges	31

4.3.1	Anonymity	31
4.3.2	Quality of Tags	32
4.4	The Firefox Extension	33
4.4.1	Interface Design Considerations	33
4.5	The Search Application	36
4.6	The Back End	36
4.7	Out of Scope Functionalities	37
4.8	Branding	38
4.9	Target Audience and Technology Adoption	39
4.10	Monetization	40
5	Illumination	41
5.1	Refinement	41
5.2	Verification	42
5.3	Critical Context	43
5.4	Websearch in the Future	43
6	Conclusion and Recommendations	44
V	Appendices	46
A	Source Code	46
A.1	Firefox Extension: content/firefoxOverlay.xul, v.0.65	46
A.2	Firefox Extension: content/options.xul, v.0.65	47
A.3	Firefox Extension: content/overlay.js, v.0.65	47
A.4	Firefox Extension: install.rdf, v.0.65	49
B	Statistical Data	51
B.1	Mozilla's Firefox Global Usage	51
C	Communication	53
C.1	Focus Group Mail	53
D	Screenshots	54
E	Links	56

List of Figures

1	Serendipity method chosen in this dissertation	13
2	BBC News Website and attached Tag Cloud.	18
3	Rating or voting is biased positive or negative.	19
4	With tagging, the bias is extracted from the amount of tags. The meaning of the tags contain additional meta-information about the object.	19
5	The benefit of tagging is it can be reversed. Websites can be found by browsing tags.	19
6	Prototype of an early tagging interface	21
7	Second Prototype of an early tagging interface	21
8	Blog with implemented badge	22
9	Del.icio.us Screenshot	27
10	Del.icio.us tagging interface	28
11	StumbleUpon home page	29
12	StumbleUpon profile page	30
13	Most Popular Technorati tags overview	31
14	Screenshot of Firefox browser showing the "Tag!" caption at the bottom right.	35
15	Screenshot of Firefox browser showing the tagging interface at the bot- tom right.	35
16	Synopse logo	38
17	Technology Adaption Lifecycle[3]	39
18	Crossing the Chasm[1]	40
19	Synopse Home Screenshot	54
20	Synopse Results Screenshots (beta)	55
21	Synopse Nothing Found Screenshot (beta)	55

Part IV

Main Part

1 Introduction

1.1 Context and Scope

The context of this dissertation revolves around the Internet, particularly the World Wide Web and its nature. The focus hereby lies in the fact that the Internet is a many-to-many communication medium where individuals can exchange information easily and nearly independently from space and time.

Furthermore, the Internet has been described as the "Global Brain"[9], a storage space where information gets filed, remixed, put into context and interconnected. There are unlimited ways how this information can be processed. Key to this dissertation is the paradigm shift whereby the Internet has moved from a medium where people passively retrieve information, to a medium where anybody can contribute, participate and collaborate. This phenomenon is also referred to as the "Web 2.0"[26] in the mainstream media. The notation 2.0 is used in software development as a versioning number in order to indicate a later, more mature release.

Internet and Web pioneers spearheaded by Sir Tim Berners-Lee, the inventor of the World Wide Web, would claim that the Internet has always been about collaboration[8] but the reason why it didn't take off earlier was down to the tools such as web browsers and technical limitations. Only phenomena such as popular Wikipedia shifted people's mind[31] and made more and more Internet users aware of the fact that the Internet is about mass-participation.

The focus of the research evolved around creating an application that would blend into the current flow of development within the Web- and Internet space. This application should be beneficial to the Internet users and contribute to known problems such as "Search and Find" or quality assurance of Information.

It is not necessarily the idea to create an application that is unique in its approach but certainly novel in the way users interact with the Web, retrieve data and add structure to the seemingly chaotic nature of the Web.

1.2 Problem Definition and Objectives

The Web and Internet being a medium where literally everybody can contribute has its good aspects. However it also has its negative sides. For example, since everybody can "dump" content onto the Web we more and more face "trash", content of no or little relevance, untrue or misleading nature, or simply content and information that propagates hostile intentions such as instructions on how to create weapons. After all, whether the networked society is beneficial or destructive for society yet has to be found out. For sure it has a significant impact.

One of the most critical problems of the Internet is finding qualitatively good and relevant information. Many companies have tried to solve this problem before and the one that seems to do particularly well is California based and Stanford graduates founded search company *Google Inc.*[7]¹⁶ Google's approach differentiated from the other search engines insofar as co-founder Larry Page implemented an algorithm called PageRank[20] which ranked the websites according to how often they have been referenced as opposed to just the textual contents and simple Metadata of the website which other search engines used to assess the quality of the content. This "outside-the-box" box thinking made Google immune - at least for a while - against spamming. Search engine spamming was a major problem in the late 1990ies for all other search engines.[7]

It is the objective of this dissertation project to propose and implement an **alternative solution for search and discovery** on the Web. The method is novel in the way that the links between websites are made not by machines but **by the users** who use the Web. In that sense, the goal is to create an application which is in line with the Web 2.0 trend and that proposes a solution for the search and discovery as well as the quality assurance process on the Web.

1.3 Methodology

Due to the novelty of the matter, it is yet challenging to find solid and established research papers and publications in the area. Therefore a different, more reflective approach has to be taken in order to elaborate sound results on which the project can progress. Even though the main direction and the goal of the research stays the same, each step towards this goal has to be reflected carefully and necessary refinements of

¹⁶<http://www.google.com>

the concepts have to be tried in a *trial and error* way. New media is fast paced and agile. A project can turn from a private homepage to a multi billion dollar company within 18 months which is what happened to YouTube.com¹⁷. Reasons why projects turn into a success can be amongst others be the fact that they "get it" right from the outset, i.e. have the perfect setup at the perfect time with the perfect user base and the perfect marketing channels. Or they can be flexible enough to be changed "on the fly" so the demand can be satisfied by fitting into the users' need deftly. It is not entirely absurd to make connections to Charles Darwin's biological evolution theory of the "Survival of the Fittest"[11] in this context.

The method to be chosen in this field of research and development is the method of Serendipity. Serendipity is described as an aptitude for making desirable discoveries by accident or even by looking for something else.[2]

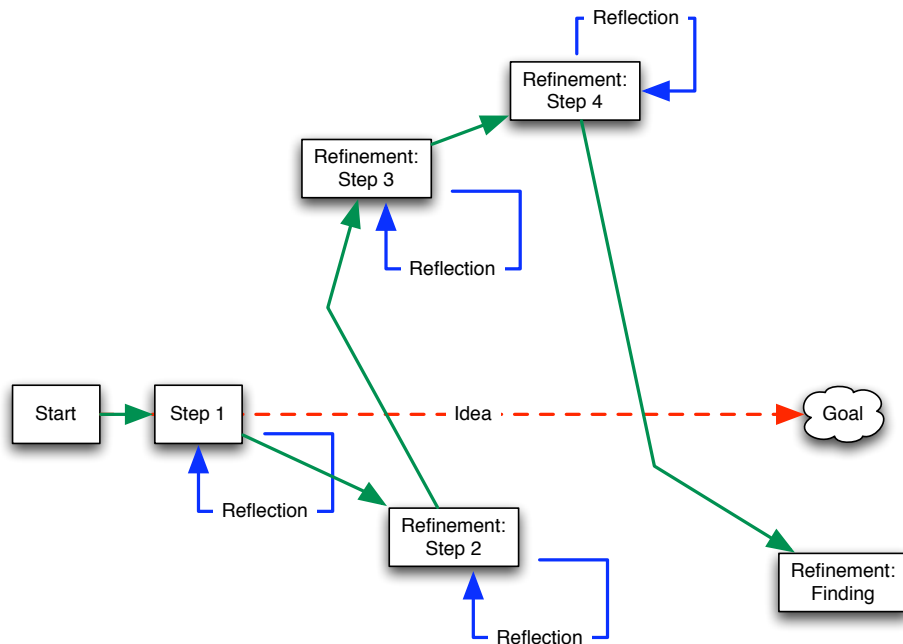


Figure 1: Serendipity method chosen in this dissertation

Throughout this document a lot of changes in the course of investigation, reflections and serendipitous approaches will be revealed.

¹⁷<http://www.youtube.com>

1.4 Hypothesis

The hypothesis this dissertation is claiming reads in brief: *"There is a need for an efficient search and discovery method on the Internet"* and *"the aggregation needed for this method has to be human driven as opposed to machine driven."* The claim is furthermore that the current climate of broad engagement with networked media supports the user generated content approach.

2 Collecting Data

The collection of data in the context of this dissertation predominantly consists of finding case studies of disruptive projects basing on networked media related concepts and ideas. The most popular case studies shall be illustrated here:

eBay - Founded in 1995 eBay¹⁸ can nowadays be seen as one of the pioneers in decentralized web based "user generated" shopping. While in the days before on-line auctions customers had to go to a shop or a market to purchase goods, these days with a couple of mouse clicks every customer can become a merchant itself. This decentralized way of doing business stimulates a whole range of new mechanisms that impact the traditional idea of doing business.[23] What mainstream media refers to as "The Long Tail"[5] has become all-day reality in our everyday lives more than ever before.

Napster - When Napster¹⁹ was sued by the management of Heavy Metal band Metallica in 2000²⁰, it has already been operational for almost a year. Napster paved the way for more peer-to-peer file-sharing applications such as KaZaa²¹, Limewire²² or iMesh²³. By trying to circumvent legal issues with hosting copyrighted content Napster developed a decentralized hosting network that would later revolutionize an entire industry.[4]

Flickr - Flickr²⁴ is one of the most popular photo sharing communities in the Internet. It

¹⁸<http://www.ebay.com>

¹⁹<http://www.napster.com>

²⁰http://www.theregister.co.uk/2000/04/14/metallica_sues_napster/

²¹<http://www.kazaa.com>

²²<http://www.limewire.com>

²³<http://www.imesh.com>

²⁴<http://www.flickr.com>

gained its strength and popularity by propagating open Application Programming Interfaces (APIs)²⁵ which allowed third parties to integrate very easily. Flickr's rise paralleled the ripening of the digital camera industry. People could share their digital pictures with friends as well as strangers very easily who then would comment, bookmark or share on.

Podcasting - Radio as well as classical terrestrial TV is a broadcasting medium which makes their audience listen to or view the same content at the same time. This leads to a number of consequences such as Prime Time, the most expensive time to switch advertising because of the high number of consumers. The invention of Podcasting, which in terms of lexicology is a artificial word made up of the word *Pod* from iPod²⁶ and *Casting* from broadcasting, however origins from the idea that content could be downloaded to the output device such as an Apple iPod at one time and actually consumed at another. This idea disrupted not only the concept of Prime Time but also solved the problem of limited channel space. Content could now be downloaded on demand, suited to the personal taste and interest and consumed at any suitable time. The download of the actual content happens in the background in idle time.

Wikipedia - Wikipedia indubitably was one of the most amazing and also surprising phenomena on the Internet.[31] Most experts would have thought that the life span of a user generated encyclopedia would last just as long until the first vandal destroys the content and therefore the project's credibility and integrity. To everyone's surprise, the contrary happened. In general, Wikipedia is considered to be a reliable source of information and vandalism usually gets corrected quickly. The successful concept bases on the idea that any change can be reverted easily without great effort. Furthermore everybody who is a consumer of Wikipedia sites is also a potential contributor and editor. This seems to successfully prevent users from vandalizing. The constant review and monitoring process also seems to maintain the quality of the content: The Nature Magazine²⁷ has compared the quality of Wikipedia versus Encyclopedia Britannica and found *"Only eight serious errors, such as misinterpretations of important concepts, were detected in the pairs of articles reviewed, four from each encyclopaedia. But reviewers also found many factual errors, omissions or misleading statements: 162 and 123 in Wikipedia and Britannica, respectively."*[18]

²⁵<http://www.flickr.com/services/api/>

²⁶<http://www.apple.com/ipod/>

²⁷<http://www.nature.com>

RSS Syndication - Wherever regular publication of content occurs, RSS is not far off.

This applies for example for blogs, podcasts, news, job or video sites. RSS stands for "Really Simple Syndication" and bases on the XML standard. It has become a very popular way on the Internet to transfer or aggregate mainly text and image material. RSS is typically delivered in the form of a feed and aggregated in a RSS feed reader such as Sage²⁸, NetNewsWire²⁹ or even any modern Web browser. Those readers constantly crawl the Internet and indicate to the user where new content can be found. It is considered to be an easy way of keeping track of regularly changing content. RSS is disruptive to the search and discovery behavior. Instead of checking every website that interests you manually every day, the feed reader can cope with hundreds of sources and check them for new content on hourly basis. The fresh content will then be presented to the user within the application without him or her go out to the Web.

The above described projects resulted in a shift of paradigms in the context of how Media, the Internet or the Web would be used. They all make use of the interconnected society which is able to act, re-act, push and pull data in a decentralized way. Decentralization, instantness, peer-to-peer and independence from space and time are the key elements that drive the proposed solution.

²⁸<http://sage.mozdev.org>

²⁹<http://www.newsgator.com/NetNewsWire.aspx>

3 Course of Investigation

This section of the document will describe how the development of this project's concept evolved. The origins of the idea lie in the fact that the Internet has to be seen as a truly interactive medium. Even though the popularity of interaction has only emerged with the newer technologies, it has always been the idea that the Internet is a User Generated Content container. It is now the duty of the creators to add meaning to the content and feed the content back to the users in a sensible way.[16] Other than in classic media such as broadcasting or print, the Internet users can be engaged into a loop of back-and-forth, many-to-many communication.

3.1 The Internet and Democracy

The fact that the Internet allows its peers to communicate in many layers and directions instantly and transparently leads to the idea that this phenomenon could be the ultimate solution to solve one of the oldest problems of humankind: organizing society via truly democratic and transparent processes[29].

After closer investigation of projects that attempted to address exactly this issue such as ThinkTools³⁰ which went bankrupt after the Dotcom Bubble and NetzNetz in Vienna³¹ who repeatedly seem to fail setting up an effective online voting mechanism, the course of research quickly abandoned the idea doing the same. Investigating why online voting tools seem to cause more disruption than support effective decision making[24] in groups would cover a dissertation on its own and would most certainly exceed the scope of this work.

3.2 Aggregating Meta-Information via User Generated Tags

Internet sites such as Flickr³² and Del.icio.us³³ have popularized the phenomenon of *Folksonomy*. The term Folksonomy is a made-up word generally accredited to information architect Thomas Vander Wal and a combination of the the words "Folk" and

³⁰<http://www.thinktools.com>

³¹<http://netznetz.net>

³²<http://www.flickr.com>

³³<http://del.icio.us>

"Taxonomy". *"Folksonomy is the result of personal free tagging of information and objects (anything with a URL) for one's own retrieval. The tagging is done in a social environment (usually shared and open to others). Folksonomy is created from the act of tagging by the person consuming the information.*

The value in this external tagging is derived from people using their own vocabulary and adding explicit meaning, which may come from inferred understanding of the information/object. People are not so much categorizing, as providing a means to connect items (placing hooks) to provide their meaning in their own understanding.

In a few conversations around folksonomy and tagging in 2004 I stated, 'folksonomy is tagging that works'. This is still a strong belief in the three tenets of a folksonomy: 1.) tag; 2.) object being tagged; and 3.) identity, are core to disambiguation of tag terms and provide for a rich understanding of the object being tagged."[34]



Figure 2: BBC News Website and attached Tag Cloud.

Tagging and rating are related to each other. Whilst rating is a statement about a certain quality aspect of an object, tagging additionally carries information about which particular aspect the rating relates to. In addition to that, the tagging mechanism can be reversed: Objects cannot only be added metadata by tagging, they can actually be found via tags.

In their paper "Harvesting Social Knowledge from Folksonomies"[36], Harris Wu, Mohammad Zubair and Kurt Maly draw the attention to a number of challenges related to

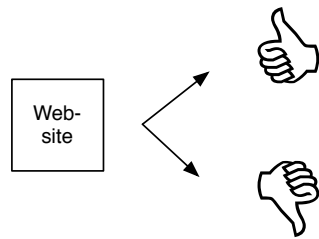


Figure 3: Rating or voting is biased positive or negative.



Figure 4: With tagging, the bias is extracted from the amount of tags. The meaning of the tags contain additional meta-information about the object.

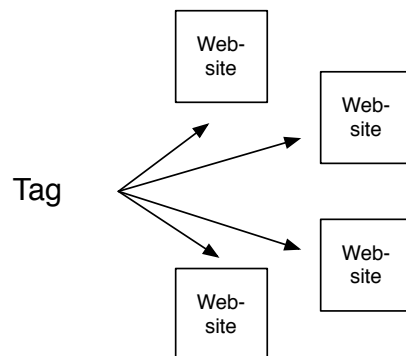


Figure 5: The benefit of tagging is it can be reversed. Websites can be found by browsing tags.

tagging and folksonomies.

"Folksonomies are criticized to have flaws that formal classification systems are designed to eliminate, including polysemy, words having multiple related meanings, and synonymy, multiple words having the same or similar meanings [...]. In addition, folksonomies invite deliberately idiosyncratic tagging, also called meta noise, which burdens users and decreases the system's utility."[36]

However they conclude that

"As tags are intermediate objects associating users and documents, polysemy and synonymy change the associative routes but not the end results of spectral analyses. It is also worth noting that [...] idiosyncratic tagging, or meta noise, tends to have a negligible impact as the number of users get large."[36]

3.3 The Badge

During the research phase, the focus shifted from creating a voting and rating application which could help groups of people to make decisions based on feedback, rating and bias to developing a collaborative feedback and rating engine that allowed anyone to add Metadata not only to persons but to all existing websites. The feedback submitted in the form of tags were meant to have a bias, either positive or negative, in order to stimulate a discussion. Owners of websites and blogs should be able to implement a "badge", a tiny window frame with interactive functionalities, to their sites in order to give their visitors an opportunity to leave feedback.

After a review of this method, it turned out that this particular method is not working. Closer investigation particularly by questioning potential users³⁴ have identified the problems as:

Lack of Simplicity - Implementing the badge to the user's site would require minimal HTML editing skills. The code of the site would have to be extended by two lines of HTML code which could be achieved via simple copy-paste. Even though the look-and-feel of the badge could be edited and adapted to the owner's site design, this would require more effort and time to be spent.

³⁴Students of Ravensbourne College of Design and Communications, Tutors, Teachers, Members of various Internet related groups such as Geekdinner, MINIBar or MobileMonday London as well as colleagues and friends

Lack of Trust - People who understand the mechanics of website would realize that inserting code could potentially harm or manipulate the content of the hosting site. Since there is no relationship of trust to the application website owners would be hesitant to implement the code.

Lack of Penetration - The penetration of the application was limited to the audience of the site where the badge was implemented. There room for scale was very limited.

Lack of Incentive - The incentive for adding the badge was missing. Owners of websites would not take the effort adding the badge to bother their visitors. Also because the application itself doesn't have a wide user base, they would not extend their own reach by adding it to the site.

account			
subject			
tag 0		upvalue ↻	downvalue ↻
tag 1		upvalue ↻	downvalue ↻
tag 2		upvalue ↻	downvalue ↻
checksum			
cmd	add_rating		
<input type="button" value="submit"/>			

Figure 6: Prototype of an early tagging interface

	<input type="text"/>
	<input type="text"/>
<input type="button" value="Rate!"/>	

Figure 7: Second Prototype of an early tagging interface

3.4 The Firefox Extension

The low uptake of the badge-application therefore forced the course of investigation and research to take a different, more scalable approach which led to the development of

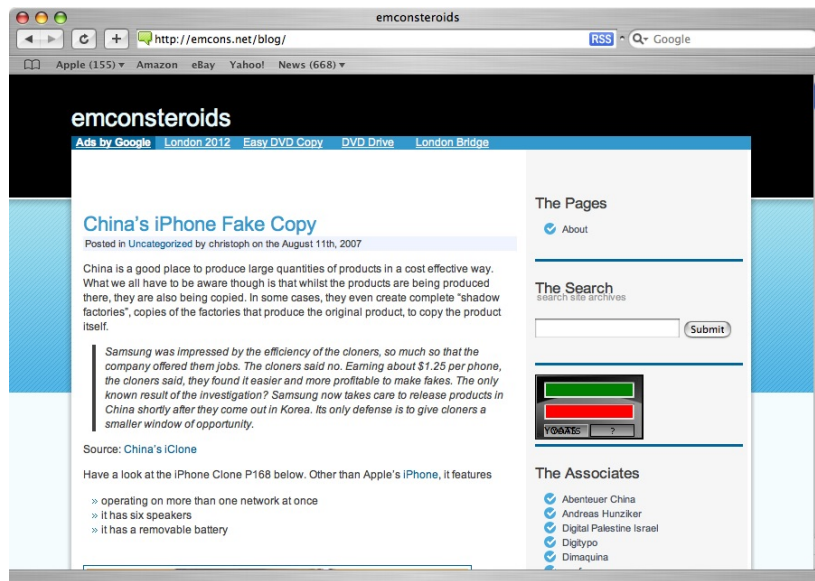


Figure 8: Blog with implemented badge

a Firefox extension. This decision allowed a more flexible scaling of the application in two dimensions: 1.) Across all Firefox users and 2.) every website becomes instantly applicable.

4 Developing the Concept

The decision to develop a Firefox extension was based on a number of circumstances:

Firefox Penetration - *"It seems that some users in the USA, Canada, Germany, Italy and Belgium are switching from Internet Explorer to Mozilla Firefox"*[10] said Niels Brinkman, co-founder of OneStat.com³⁵. Microsoft Internet Explorer is losing market share mostly to Mozilla Firefox across all major territories. (cp. Appendix B.1) It is safe to assume that in the future Firefox will grow even more in market share, particularly if the number of Internet Explorer Security issues does not decrease and users lose trust and confidence into Microsoft's product.

Availability of Source Code - The Firefox Extension programming code is, like in HTML and JavaScript, easily visible, modifiable, re-mixable and extensible. Transparency of source code is important to the context of openness around the project.

³⁵<http://www.onestat.com>

Simplicity of Technology - With Firefox Extensions programming the developers have instant access to a number of very powerful tools and a solid base. Furthermore, the sophisticated version management and auto-update functionality facilitates a comfortable way of updating the clients remotely without great effort for developers as well as users.

4.1 Considerations

4.1.1 User Database and Logins

Nowaday's Web is full of personalized sites that require usernames, passwords, logins, e-mail addresses or any other form of authentication. The concept of the application deliberately excludes any sort of identification. There are a number of reasons why this decision had been taken:

Inhibitions - The registration process is considered to be the biggest hurdle for people interacting with the service. Users do not want to share information and go through the hassle of filling out web forms.

Simplicity and Convenience - Not having to register but still being able to use an application is very rare. It is the "out of the box" feeling that lets users interact with the application straight away which makes it interesting.[21]

Anonymity - Even though there is no direct connection between a real and a "virtual" person online, most online applications allow some form of analysis of their users in the back end. Instead of asking people to register (with unverified information) why not letting them using it completely anonymously in the first place? The downside is they cannot backtrack their tags easily unless they use a "static tag. (For details please refer to section 4.3.1, Anonymity)

Spam - Typical registrations also incorporate sharing the e-mail address. This opens a door for potential spammers. Users generally prefer to keep their e-mail addresses as private as possible.

Experiment - This approach of getting users to utilize the application interactively without registration is novel. The outcome of this experiment is not foreseeable which can lead to interesting new conclusions.

The fact that users are not identified clearly opens doors to abuse the system. It is difficult to tell in advance whether users will misapply the application or whether they understand that abuse is pointless because of the self moderating mechanism in place that allows other users to correct potential misuse, spam and profanity. The mechanism is obviously inspired by Wikipedia which seems to to be working to everyone's surprise.[18]

4.1.2 Critical Mass

In order to get a community kick-started, it needs to have "critical mass" to hit the "tipping point".[17] There are a number of ways to reach the tipping point. The most popular is to pre-populate the application with data pre-launch. This gives possible users the impression that the application is already being used by other people which makes it look less "idle" or "superfluous". Activity attracts more activity. In order to get that ball rolling it is important to simulate or mock activity. Whereas in social networking sites this tends to be more difficult as their members may behave very randomly, in the case of the proposed application this is easy. There will be a crawler that crawls the Internet for tags, for example Technorati Tags³⁶ and adds them to the database as "dummy tags". Furthermore there are plans to develop a small automated application that will migrate tags for important or popular URLs from del.icio.us to the proposed application in order to help getting the needed content to increase initial attraction. In addition to that, there was a small group of around 50 beta testers who used and tested the application. While testing, they also populated the database with sites relevant to them. The findings of the beta tests can be found in section 5.2, Verification.

4.1.3 Distribution

Scale Free Networks[6] such as the Web offer numerous ways of distribution. Albert-Laszlo Barabasi describes a "Scale Free" network as

*"governed by two laws: growth and preferential attachment."*³⁷

The consequence is that networks grow around a smaller core networks of highly connected hubs. In the web, an example for highly connected hubs could be Wikipedia,

³⁶<http://www.Technorati.com>

³⁷Barabasi: 86

search or web indexes. Also certain particularly popular blogs can be considered as highly connected or frequently cited nodes.

In terms of efficiency, it makes sense to piggyback the distribution of the application on those highly connected hubs in order to get the application distributed. A number of strategies have been outlined:

Facebook Plugin - Facebook³⁸ offers third parties the possibility to develop what they call "Facebook Applications". Facebook applications hook into the Facebook framework and are allowed to have access to certain user data. The advantage is that an application can spread quickly piggybacking on the respective Facebook user's social networks.

"Applications are tools that allow you interact with your friends and networks. Examples of applications that Facebook has built are Photos, Notes, and Groups. Now, Facebook allows third parties to develop applications that abide by your privacy settings and live within the site. Adding applications allows you to expand the usefulness of the site."[14]

Wordpress Plugin - Wordpress³⁹ is a weblog provider which allows extending the blog with a number of plugins⁴⁰. A plugin allows the user to enhance the functionality of the blog e.g. with special commenting, spam filtering, automatic translation, mapping integration or special syndication. Third parties can provide plugins for Wordpress for free and those plugins will be listed in the Wordpress plugins directory. This is a good opportunity to distribute the badge and therefore allow users of blogs to tag the sites without having the Firefox extension installed.

Public Firefox Extension - Just like Wordpress has a database for plugins, Firefox or Mozilla have a database for extensions⁴¹. Here too, any third party can submit an extension which will then be reviewed by the Mozilla team before released to the public in the directory. Once it's uploaded and appears in the categories however, there is a good chance that Firefox users will either stumble upon the extensions by accident or explicitly search for an extension that allows tagging of websites. In order to submit a Firefox extension, the developer has to be registered at the

³⁸<http://www.facebook.com>

³⁹<http://www.wordpress.com>

⁴⁰http://codex.wordpress.org/Plugins#Plugins_by_category

⁴¹<https://addons.mozilla.org/en-US/firefox/browse/type:1>

Mozilla Addons site ⁴² before the application can be submitted⁴³.

Digg.com - With over 1m new users and 4000 stories submitted per day⁴⁴, Digg.com⁴⁵ is presently one of the most exciting digital hubs and link collections. It is certainly difficult to get a story on the front page as it's the users who will rate and vote for stories. However if a story gets on the front page, immediate traffic and attention is guaranteed.⁴⁶

4.1.4 Quality Assurance and Abuse Prevention

With the users being able to contribute anonymously, there is a high risk of abuse and loss of quality. Inspired by Wikipedia, a three self regulating mechanisms should prevent users from abusing the system:

Feeds - Every site or domain can be monitored via a standard RSS feed. This feed contains all new tags not only for a specific URL but for an entire domain. This RSS monitoring allows abuse to be caught very quickly and therefore corrected instantly via the web interface which allows anybody to delete malicious tags.

Self Moderation - The standard web interface allows any visitor to remove tags from a site. This may prevent spammers and abusers from wasting their time trying to vandalize the site. Every change will be recorded so in case important tags get removed unintentionally they are still retrievable.

History - The concept envisages a recorded history for all changes of the site.

These concepts are heavily inspired by the Wikipedia concept. However other than in Wikipedia, there is no registration needed for this application. This makes the finding of particular users or repeating abusers more difficult. It is subject to the concept that at a later stage, additional authentication mechanisms can be implemented if needs be.

⁴²<http://addons.mozilla.org>

⁴³<https://addons.mozilla.org/en-US/firefox/developers>

⁴⁴http://www.businessweek.com/magazine/content/06_33/b3997001.htm

⁴⁵<http://www.digg.com>

⁴⁶On http://www.intensivstation.ch/stats/usage_200607.html is an example from a site hit by the "Digg-Effect" on 15th of July 2006.

4.2 Related Existing Projects

4.2.1 del.icio.us

Del.icio.us⁴⁷ (pronounced "Delicious") is a website created by 33 year old Joshua Schachter in late 2003. The site became the prime example of what today is known as a "social bookmarking" site. While normal bookmarks are typically kept within the local browser, the idea of social bookmarking is to share the bookmarks online with friends. This does not only have the advantage that users can see what their friends are bookmarking, but it also helps measuring the popularity of certain sites and last but not least, the bookmarks become more inter-portable between computers. Being able to see what other people are bookmarking can be particularly useful if people's profiles are being compared. Similar interests can then be clustered and RSS feeds of friend's bookmarks can be subscribed to.

Del.icio.us was sold to Yahoo! Inc.⁴⁸ in December 2005 for an undisclosed sum which is generally estimated around 15-30 million USD.[35]

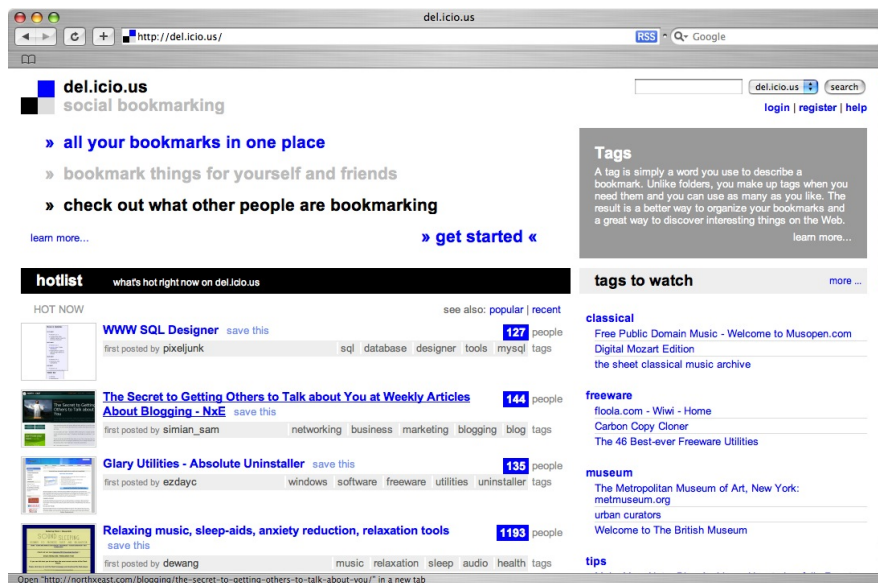


Figure 9: Del.icio.us Screenshot

Del.icio.us URLs can be added to the user account in various different ways:

⁴⁷<http://del.icio.us>

⁴⁸<http://www.yahoo.com>

Web Interface - The URL's can be posted to a web form⁴⁹ on del.icio.us. Fields such as the URL, the Title, a brief description and the tags have to be filled out. The new bookmark is instantly visible to the whole community after submission.

Bookmark Link - Users can set a bookmark in the browser's bookmark bar and hit this button in order to save the current site to the user's del.icio.us account. The interface will also suggest tags based on an unknown algorithm and show popular tags used by other users for the current URL.

Bookmark button/icon - Third party website providers can integrate a button or a small icon for visitors to add the website to their del.icio.us accounts via a single mouse click. This method has amongst others also been adapted by many mainstream media companies such as the British Broadcasting Company⁵⁰.

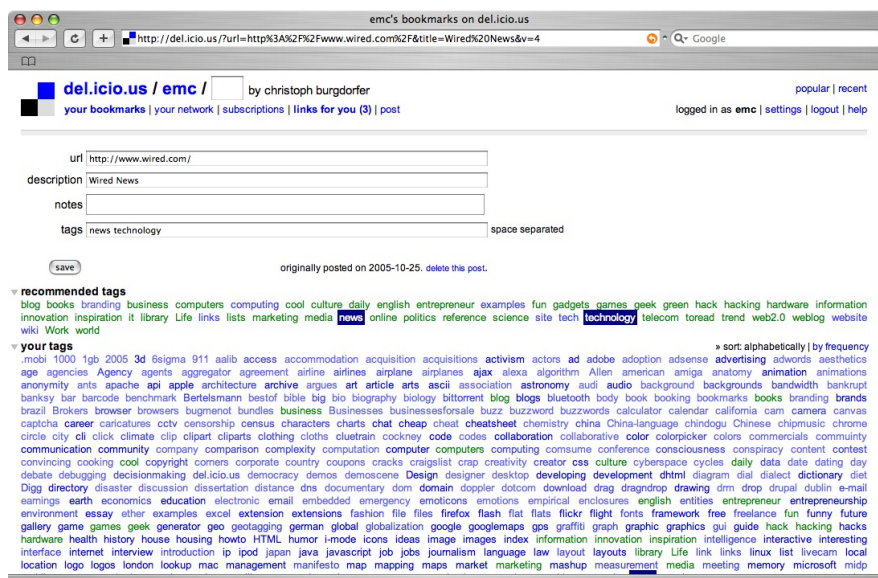


Figure 10: Del.icio.us tagging interface

Del.icio.us can also be used to search or group websites according to tags. However this does not seem to be the main functionality of the site as the performance when searching for tags is exceptionally poor. There is also hardly any logic behind displaying the search results.

It yet remains unclear how del.icio.us funds its operational costs however it is likely that due to the joint venture with Yahoo! there are mutual agreements in place.

⁴⁹<http://del.icio.us/post/> (You have to be logged in in order to be able to post.)

⁵⁰<http://news.bbc.co.uk/1/hi/help/6915817.stm>

4.2.2 Stumbleupon.com

Stumbleupon.com⁵¹ is a 3.3 million strong community (August 2007) and consists of a browser plugin/extension in the form of a bar that lets users adding tags and ratings to websites. As the name suggests, the main part of the bar is the "Stumble!" button upon which a user will be displayed a seemingly random website which is in fact chosen by the StumbleUpon application according to a number of factors such as personal preferences (what the user has tagged before), personal settings and the ratings/tags of the user's friend. StumbleUpon is usually mentioned in the context of online social networking and online automated recommendation system sites. Stumbleupon was acquired by Yahoo! Inc. in May 2007 for a sum of 75 million USD.[15] StumbleUpon is mainly used for the web-equivalent of "TV channel surfing"; discovering random but possibly relevant content by "zapping" through the web. Secondly, StumbleUpon can be used to remember previously visited websites and share them with online social contacts.

StumbleUpon's revenue comes from highly targeted advertising. In less than 2% "stumbles" are actually links to sponsored sites related to the user's topics of interest.



Figure 11: StumbleUpon home page

⁵¹<http://www.stumbleupon.com>

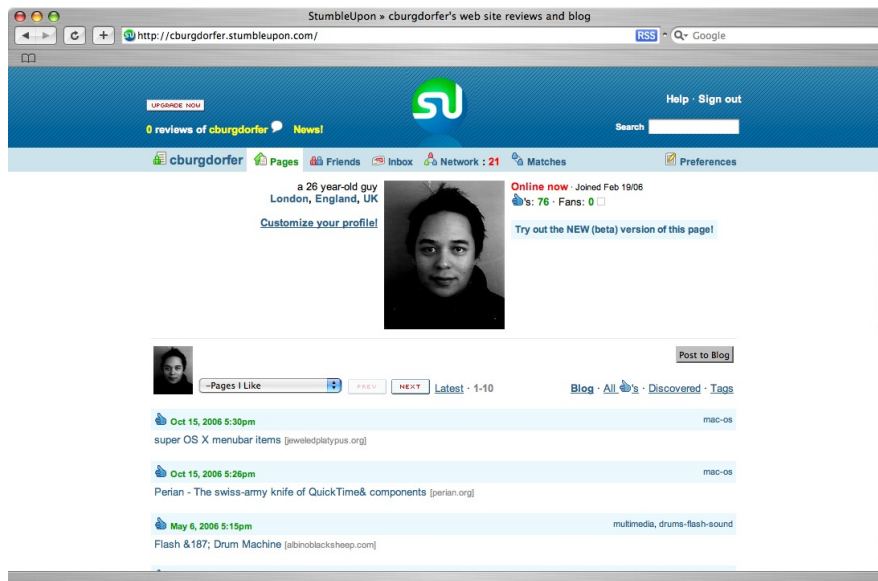


Figure 12: StumbleUpon profile page

4.2.3 Technorati.com

Technorati was founded by David L. Sifry in 2002 and is headquartered in San Francisco, California, USA. Its original mission was to become the search engine for blogs, whereby Technorati aggregates live RSS feeds and incorporates them into the search. Technorati is *"currently tracking 100.1 million blogs and over 250 million pieces of tagged social media."*[32] Besides the fact that Technorati regularly checks blogs, it can so be "pinged" meaning notified, when a change in a blog occurs or when it gets updated. Technorati's slogan is *"Who's saying what. Right now."*[32] In early 2005 Technorati unveiled its "Technorati Tag" service.[12] Tags are aggregated from del.icio.us, flickr or specially formatted XML sections within the RSS feed. Most blogging software nowadays support a categorizing functionality for blog posts. Those categories are then taken on as tags in Technorati. There are also a number of plugins and extensions for already existing blogging software that allow the bloggers to tag and categorize their posts with the specific Technorati formatted tags.

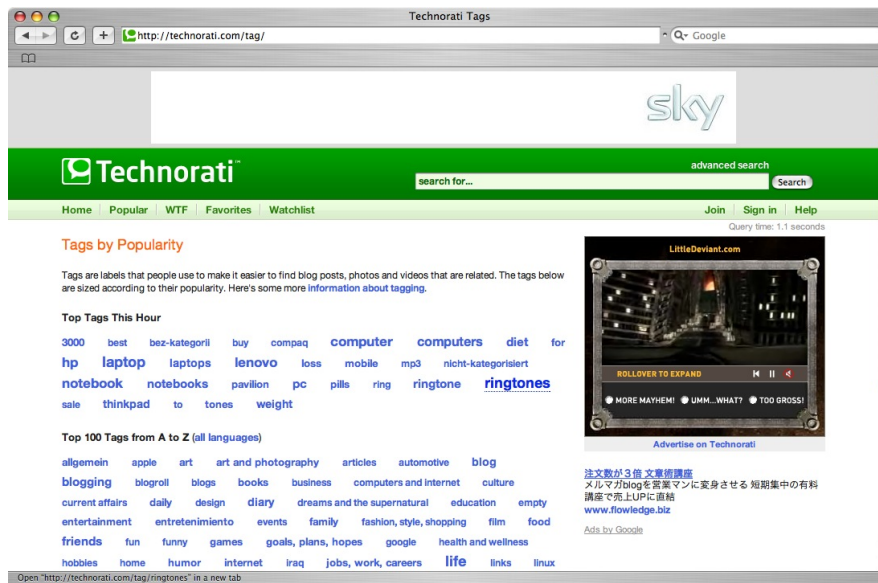


Figure 13: Most Popular Technorati tags overview

4.3 Challenges

4.3.1 Anonymity

The fact that users will be able to interact with the application bears a big risk of abuse. Technically it will not be possible to track back actions to specific users and individuals cannot be banned nor punished for their behaviors. This could lead into complete chaos with untrustworthy tags, spam and abuse. The risk of abuse also increases proportionally to the amount of active users. The wider the reach is, the more interesting it is for spammers and vandals to use the system inadequately.

If it turns out that the system would not work, logging and personalization features can easily be added at a later stage. There will also be profanity filters automatically refusing any sorts of offensive or destructive tags. The quality and ability to recognize faulty content through profanity filter increases as the usage of the applications grow. This spam filter will work according to the Bayesian filtering algorithms[13] described by Thomas Bayes. Bayesian email filters take advantage of Bayes' theorem. Bayes' theorem, in the context of spam, says that the probability that an email is spam, given that it has certain words in it, is equal to the probability of finding those certain words in spam email, times the probability that any email is spam, divided by the probability of

finding those words in any email. The filter doesn't know these probabilities in advance, and must first be trained so it can build them up. To train the filter, the user must manually indicate whether a new email is spam or not.

The absence of user profiles also makes it difficult to assign actions to persons in order to find the actions at a later stage. The way this limitation is supposed to be worked around will be what can be referred to as an optional "static tag". A static tag, e.g. a username or a randomly picked word, is a tag that always gets added to the tags left behind in a site. This allows the user to find the tags again at a later stage by searching for this unique identifiable keyword in the search engine. This system cannot be considered as 100% fool proof nor save and secure but the significant benefit in terms of simplicity is worth taking the risk.

4.3.2 Quality of Tags

As described in their 2006 paper "Harvesting Social Knowledge from Folksonomies" [36], Harris Wu et al. describe a number of problems that evolve around folksonomies or tagging systems:

Polysemy - The word "polysemy" derives from Greek "poly" ("many" or "multiple") and "sema" (sign) and describes the ambiguity of words. Polyseme words can have similar meanings which in many cases bases on the fact that they have similar etymological roots. They can however also have completely different meanings. An example for a polyseme word in English would be the word "mole". It can refer to an animal, a chemical unit, a spy, a skin marking or even a name. Folksonomies rely on tags entered by the users and ambiguous tags can falsify the result of searching users. E.g. if a user searches for the tag "mole" in the search engine, the system will not be able to determine which of the above options the user means. Only when adding another tag, e.g. "animal" the computer software is able to narrow down the results to what the user is looking for.

Synonymy - Unlike polysemy where one word can have multiple meanings there is synonymy where one meaning can have multiple words. This causes problems with folksonomies insofar as the users will not necessary use the same synonyms to describe the same meanings. Which word people will be using to tag depends on the level of sophistication of the vocabulary, creativity and randomness. This problem can hardly be addressed but one can assume that with the increase of

the usage and as the aggregation of tags increases, the core relevant tags will establish while variety and synonymity will level out.[30]

Meta Noise - Wu et. al. write *"Collaborative tagging suffers from idiosyncratic tagging, also called meta noise, which burdens users and decreases the system's information retrieval utility."*[36]

Other than that, there are also language overlapping issues that have to be considered. For example the term "bad" does not mean the same in English than it does in German (Bath). Generally, English has to be suggested as the standard language of tagging mainly because of it's simplicity and ubiquity.

Wu et. al. also note that *"Folksonomies claim to have many advantages over controlled vocabularies or formal taxonomies. Tagging has dramatically lower costs because there is no complicated, hierarchically organized nomenclature to learn. Users simply create and apply tags on the fly. Folksonomies are inherently open-ended and therefore respond quickly to changes and innovations in the way users categorize content. Collaborative tagging is regarded as democratic metadata generation where metadata is generated by both the creators and consumers of the content."*[36]

4.4 The Firefox Extension

4.4.1 Interface Design Considerations

Throughout the project, there is a heavy focus on simplicity and ease of use for both, the Firefox extension as well as the web search interface. In nowadays' complex and media flooded world where 75% of the web surfers would not return to a site if it takes longer than 4 seconds to load[27], simplicity is key. John Battelle, author of "The Search", would even argue that simplicity of Google's search interface contributed significantly to it's success.[7] The interface design and interactivity patterns follow the basic rules of Bruce Tognazzini's "First Principles of Interaction Design"[33]:

Anticipation - The user must be able to anticipate automatically how the application works. Long and complicated explanations must not be needed.

Autonomy - Make the user the "owner" of the application and give him the feeling that he is in control while still giving enough visual and interactive guidance.

Colors - The use of colors are thought through and are a compromise of drawing attention and not getting on the users' nerves. Yet the color scheme has to leave a visual impact.

Consistency - In order to make the application as easy as possible, the consistency is given particular attention. One single-line input fields and no unnecessary buttons and interactivity possibilities.

Defaults - A number of settings within the application can be changed. E.g. static tags or the behavior of the tagging bar. The user will be able to change these settings however it is assumed that more than two third will not do so.

User Efficiency - In order to keep the database server power strong enough there may be hardware investments. It is important not to underestimate the importance of a fast infrastructure and not to waste the user's time.

Tagging should be made as easy as possible for the users, meaning ideally the number of interactions via mouse have to be kept to a minimum. Mouse interactions take a lot of time and effort from the user as he has to move his hand away from the keyboard to the external device and back. This is a complex action and requires coordination between hand and eyes. Therefore, the interface design deliberately excludes buttons. The tags can be submitted by hitting the "Enter" button. Furthermore the tagging bar can be triggered by a shortcut and the focus is automatically set to the input field.

Another important point is the visibility of the tagging bar: In terms of attention span, it is not desirable to interrupt the users in the middle of an action, e.g. when he/she is reading a long article online. Instead, try to catch the attention in the milliseconds between the actions, e.g. when the user has finished reading the article, scrolled to the bottom and is about to scroll back up. This is exactly the time-frame the applications aims at getting. As soon as the user has scrolled to the bottom of the page, the application assumes that 1.) he/she has read or at least seen the content of the page and 2.) has a spare second to be interrupted and asked for some feedback. The tagging bar automatically appears whenever the user scrolls to the bottom of a page.

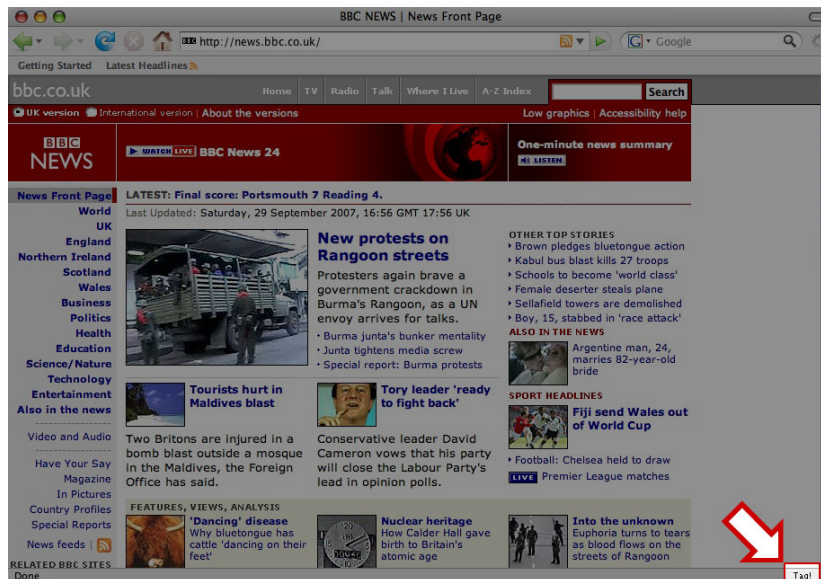


Figure 14: Screenshot of Firefox browser showing the "Tag!" caption at the bottom right.

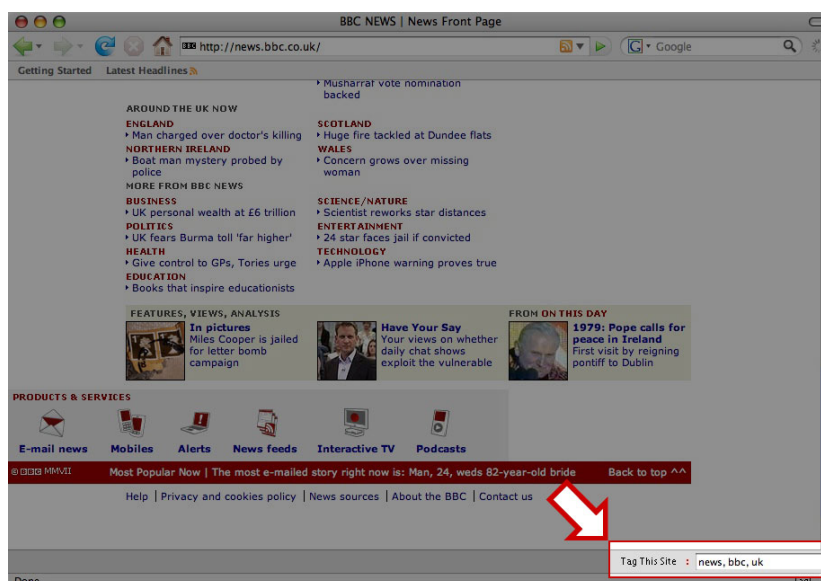


Figure 15: Screenshot of Firefox browser showing the tagging interface at the bottom right.

4.5 The Search Application

The web search application which allows users to find and explore pages according to tags also follows the rules of simplicity. It adapts the basic concept of a search engine interface for the benefits of user friendliness and intuitive interaction patterns. It bases on a single lined input field and a search button. The results are displayed in a list form underneath the search interface and the search can be re-iterated easily and without effort. The technology of the interface bases on the AJAX (Asynchronous Javascript and XML) principle which works with asynchronously exchanged data without the user's notification. This data allows to change the contents of the website without re-loading or even re-rendering the website. The downside of this technology is that unlike in a synchronous request where all the data is echoed according to the HTTP GET or HTTP POST⁵² request, the current URL in the browser's location bar would not change and the visible page therefore cannot be shared with other people. There is a work-around in place for this which is being used in this interface: The URL in the location bar can be changed without the page to reload under one single circumstance which is if the change is AFTER a hash (#) symbol. Hashes usually indicate anchors within the page. Jumping to an anchor does not require the page to reload. This circumstance can be exploited to indicate the setting, i.e. the current search term within the location bar allowing the URL to be shared amongst other individuals by simply forwarding the information in the browser.

4.6 The Back End

The application back end bases on an Apache 2.0⁵³ Server run on a GNU/Linux Debian⁵⁴ infrastructure. The database server is run by MySQL⁵⁵ and the application logic is written in PHP⁵⁶. This setup has a proven track record of being able to cope with these kinds of projects. It is generally referred to as the LAMP setup (Linux, Apache, MySQL and PHP) and all components are available as open source software. The

⁵²HTTP GET and HTTP POST requests are two ways of requesting information from a web server. The difference between GET and POST is that GET requests are submitted via URL and parameters within the urls whereas POST requests are submitted via URL and a message body. GETs are usually used for idempotent actions like retrieving data whereas POSTS are used to actually alter information.

⁵³<http://www.apache.org/>

⁵⁴<http://www.debian.org>

⁵⁵<http://www.mysql.com>

⁵⁶<http://www.php.net>

reasons why this technology is the technology of choice are:

Costs - LAMP systems are very cost effective in comparison with its competitors from corporations such as Microsoft or Sun.

Documentation - The documentation for PHP and MySQL is very widely available on the internet and easy to find. Because of the large user base of this technologies, it is also easy to find information in forums and chat rooms.

Scalability - LAMP systems are known for scaling nicely. Most hosting centers support LAMP systems and the infrastructure can easily be ramped up. Also because of its popularity it is easier to find project participants in the future should this be necessary.

The reason why the Debian distribution of GNU/Linux has been chosen bases on the fact that Debian is known for running stable and having an effective package management system to update new software.[19]

The software is hosted on Christoph Burgdorfer's private server due to specific requirements such as Apache 2.0⁵⁷ with the PHP5⁵⁸ module enabled and MySQL Server 5.0. In order to have root access to the server to include PEAR Modules⁵⁹ to enhance the efficiency of the development.

4.7 Out of Scope Functionalities

The application leaves a lot of options to be extended. The following functionalities can be considered as good supplements or extensions to the existing setup:

Open API - An Open API gives third parties the opportunity to develop their own application on top of the existing infrastructure. The benefit of third parties developing applications for APIs is that the ranges of ideas can be scaled better. Other people use their own creativity to extend the functionality of the core infrastructure. The downside of an open API is that it needs to be secured properly. A concept of access restrictions and permissions has to be in place and the documentation needs to meet a certain quality level. This is not within the scope of this dissertation but will be added at a later stage.

⁵⁷<http://httpd.apache.org/docs/2.0/>

⁵⁸<http://devzone.zend.com/tag/PHP5>

⁵⁹<http://pear.php.net>

Sophisticated Evaluation - Connecting links with tags and tags with links opens many new opportunities and possibilities to examine the relations between URLs and their context. An example evaluation could be a hierarchical tree of related sites via tags or a "thread" of sites that are connected via common tags. Also linguistic analysis of tags, context and frequency patterns could bring interesting information up. It is not within the scope of this dissertation to elaborate sophisticated statistical information about the usage of the application however there are plans to implement that at a later stage.

Graphical Visualization - Even though originally planned, the idea of displaying the relative relations between tags and links in a visual (e.g. a tree diagram) way, this functionality is not implemented in the application. Even though it would be "nice-to-have", there is not enough benefit in it. Third parties will be able to implement these kind features via the open API in the future.

4.8 Branding

The name of the project is "Synopsis", an artificial word made up of "synopsis" (general overview or summary of an article or text) and "synapse" (the connectors of brain cells). These two words combined reflect the functionality of the application. Firstly, the tags should give an overview or even a summary of the content of a specific site. Secondly by grouping websites with tags, the tagging mechanism should act as a connector between two sites. A common tag generally refers to the fact that there must be some sort of contextual link between two sites. The logo is kept simple and therefore easily recognizable. The colors have to be bold and associate the right connections.



Figure 16: Synopsis logo

The caption is held in simple grey shades fading out towards the right climaxing in the heavy red comma. The comma refers to the commas used to separate tags. For the sake of consistency, all punctuation such as colons, semicolons, commas, periods,

exclamation marks as well as question tags are held in bold red. Any other caption is held in various shades of grey.

The size of the interface is notably bigger than the size of similar interfaces. This emphasizes the simplicity aspect and makes it easier for the user to navigate. Furthermore, the application looks "bolder" and emanates more self-esteem.

Slick icons indicate functionalities such as tags, RSS feeds and links.

4.9 Target Audience and Technology Adoption

Due to the fact that the plugin relies on the Firefox browser, the target audience is already limited to the Firefox user base. Firefox users however fall into the same category this application is targeted at: mainly male customers, technophile, early adapters of Web 2.0 technologies, 18-24 years old and University Students.[28] The reason why this is the primary target group bases on the fact that this application heavily relying on what Geoffrey A. Moore describes as the "Early Adopters". People, who help "Crossing the Chasm"[22] in the Technology Adoption Lifecycle curve.

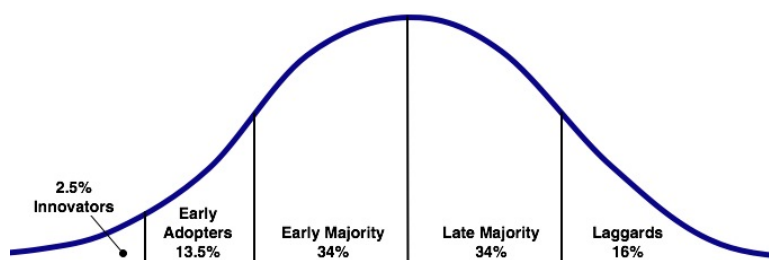


Figure 17: Technology Adaption Lifecycle[3]

Moore points out that there is an invisible "chasm" within the early adopters section of the technology lifecycle which has to be crossed and the best method of crossing this chasm is by keeping it small. This can be achieved by good usability and simplicity of the application.

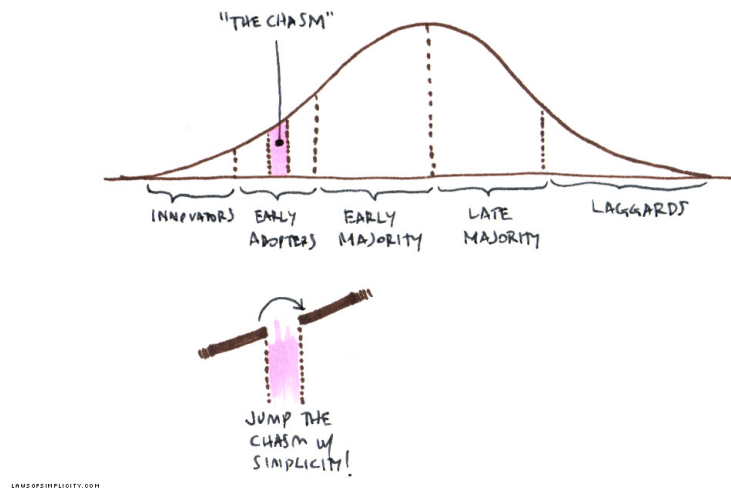


Figure 18: Crossing the Chasm[1]

4.10 Monetization

Because of the lack of easy and straight forward payment mechanisms on the Web monetization can generally be considered as a challenge. The most common ones are credit card billing and Paypal⁶⁰. The advantage of Paypal is that no delicate information such as credit card information has to be shared with or exposed to the third party. The payment transaction is between the user and Paypal only and bidirectional. This makes Paypal particularly interesting for micro-payments or casual transactions such as donations.

Donations will be a way of financing the project. Furthermore there may be a Google AdSense⁶¹ window. Google AdSense will detect the user, his or her interests and even the context and search intentions. Google will then target the advertising accordingly and in real-time so there is a bigger likelihood that the user will click on an advertising. For each click, money will be transferred to account associated with the project.

The third revenue stream will be via sponsoring. The contact link at the bottom of the search page allows third parties to get in contact in order to discuss sponsoring.

⁶⁰<http://www.paypal.com>

⁶¹<http://www.google.com/adsense/>

Supporter of the idea can contribute either monetary or via donations of goods. In return they can place a logo on the search site accompanied by the text "supported by" and a link.

The fourth and last step of monetizing the project is by exiting to an investor in return of stakes in the company. This is a particularly interesting option once the project takes off and the user base is flourishing. The chance however that this will happen is very slim, nonetheless it should not be ruled out from the outset.

5 Illumination

5.1 Refinement

The progress of this project was based on constant feedback. During the development of the concept, it was important to keep releasing applications and ask for immediate feedback. The more feedback available, the easier it is to steer the application's progress and next release. Having a variety of people with different backgrounds giving feedback is particularly effective. Even more if they are part of the target audience Synopsis is geared at.

Being able to react to feedback quickly however also requires the technical setup being flexible and scalable enough to adapt or extend the functionalities. This can be achieved by splitting the backend and frontend into various modules that can be put together that can either be altered individually or stucked together in different ways like Lego bricks. Constant refinement was an essential part of the whole development process and included e.g.

Feedback on the badge as the medium - *"Why should I add a badge to my blog if the user base is small?"* lead to the *Firefox Extension* approach.

Feedback on incentive concerns - *"Why should I use this if I can't find my own tags anymore?"* lead to the *Static Tag* feature.

Feedback on abuse and vandalism concerns - *"How can you make sure this doesn't get abused and vandalised?"* lead to the *RSS* and *self moderation* feature.

5.2 Verification

In the beginning of September 2007 a focus group test had been launched. An E-Mail (Appendix C.1) had been sent out to around 100 random people including IT specialists, techies, friends, acquaintances in various age groups and working in a wide range of industries. The feedback was manifold and included questions as well as answers:

Why is it different from del.icio.us? - The main difference is: del.icio.us is bookmarking and archiving, synopsis is tagging and searching. del.icio.us suggests tags and therefore biases where synopsis is more random. del.icio.us needs registration and personalization whereas synopsis is completely anonymous.

How can I tag? - Many beta testers did not notice the "Tag!" button at the bottom of the browser. The idea of the tiny "Tag!" button was to keep the application as unobtrusive as possible. The disadvantage is, people would not find it easily. The FAQ and instructions have been adapted accordingly.

How do you deal with spam? - A very popular question which is answered in this dissertation document. Whether or not this method will work in the future is still subject to more extensive tests.

Sometimes the application returns error messages - This was down to special characters in URLs such as &, ? and #. This has been fixed in the next release.

Why is the search engine not implemented into the browser extension? - This possibility certainly has to be incorporated into the future development road map.

The Popup saying "Thanks for your contribution" is annoying - this functionality will be removed in future releases.

What search algorithm are you using, how do you order results? - The search algorithm is very trivial: The concept suggests that the tags are ordered alphabetically and the websites are ordered by the time they have been tagged with the respective tag with the newest at the top. The tag cloud of an URL will grow over time, meaning that the description of a URL will increase in accuracy. Therefore, if users want to find URL's with more accuracy, they have to search for two or more tags at the same time. This filters out unwanted other results. In other words: The older an URL is (and hence the further down the list) the more tags it will have. The more tags it has, the more likely is it to be found by more accurate searches.

All the feedback was quickly adopted and will be incorporated in future releases.

5.3 Critical Context

Throughout the whole dissertation project, a critical view on the context of the work was important. Ultimately, it is what drives the project forward and allows or even enforces changes for the better. During the course of investigation and implementation, it turned out that critical feedback was the most valuable feedback as it had the most significant impact on the development. Having chosen "serendipity" as the methodology of choice it made any critical feedback even more valuable as it allows re-thinking the status quo from different angles. After a number of iterations, the final product emerged which seemed to have a good uptake in the the final beta test runs.

5.4 Websearch in the Future

With Google's success in "organizing the world's information"⁶² a lot of researchers, companies and "Gyro Gearlooses" have tried to get their head around of how search in the future might work. There are many concepts of what and how search will change particularly with the help and ubiquity of mobile devices. Google is expanding their offer by adding location based search⁶³, distributed indexing⁶⁴ and currently there are even rumors about a possible "Google Phone" with GPS circulating.⁶⁵ Notable in the context of future search is also a concept developed by Ex-Microsoft technology evangelist, blogger and writer Robert Scoble⁶⁶. He explains in three Kyte videos⁶⁷ the concept of "Social Graph Based Search". In his opinion, social networks and a tight knit web of trust can help solving the problem caused by the Search Engine Optimization (SEO) industry who mainly makes its money illegitimately by tricking search engines. Scoble claims that a social network and a web of trust does would not allow SEOs to enter the domain and therefore spoil search engine results with fraudulent trickery. The base of such a search network could for example be Facebook. Even though the concept of this search mechanism is antipode to the approach proposed in this dissertation, namely a completely anonymous tagging but self regulatory approach, it will be interesting to

⁶²<http://www.google.com/corporate/>

⁶³<http://maps.google.com/>

⁶⁴<http://www.google.com/services/local-business-referrals/>

⁶⁵<http://gizmodo.com/gadgets/rumor/google-phone-confirmed-by-htc-insider-294277.php>,
<http://www.last100.com/2007/08/29/the-gphone-is-coming-how-google-could-rewrite-the-rules/>

⁶⁶<http://scobleizer.com/>

⁶⁷<http://www.kyte.tv/channels/view.html?uri=channels/6118/47141>, <http://www.kyte.tv/channels/view.html?uri=channels/6118/47146> and <http://www.kyte.tv/channels/view.html?uri=channels/6118/47151>

see how the search industry develops in the future. Both approaches have their legitimization and it cannot even be ruled out that there will be both systems in parallel or a symbiosis of both systems in the future.

6 Conclusion and Recommendations

The research in the context of this dissertation led to the following conclusions:

Always ask questions - The digital media space is changing so rapidly, it is nearly impossible to predict the future. Working on a concept requires constant self-criticism and review of the goals. What may look like a perfect solution on one day might be a completely wrong idea on the next. It is important not to get stuck in a fixed idea and look to the right and left what is going on and consider this development within the own context.

Rapid development is essential - In order to test new concepts and ideas, it is crucial to get them out there, even if they are not 100% finished. The adoption can therefore be assessed quickly and necessary measures can be taken in order to react to the user's responses. Changes have to be implemented quickly before users decide to abandon the service. It is hard to regain users once they are lost.

Test, test, test - Quality assurance is important. Testing also includes taking extensive feedback from the users and incorporate this in future releases. It is only during excessive testing where usability and/or conceptual flaws can be identified and corrected. The extensive testing phase of Web 2.0 applications is also often referred to as the "perpetual beta"⁶⁸.

Challenges are not technology but the users - Therefore simplicity is key for adoption. Technologically, the possibilities are not limited. In software development, everything is just a matter of time, money and effort. However in many cases, the interface between the machine and the human, the user interface, gets neglected. Going the extra mile is worth the effort because it will lower the barrier making the users engage with the software.

Crowdsourcing and Folksonomies are effective for contextual browsing - Even though there are critics about the efficiency of random tagging and folksonomies, I have found tagging a very effective concept to "browse" or "surf" through tag clouds. It

⁶⁸http://en.wikipedia.org/wiki/Perpetual_beta

is much easier to jump from context to context by linking tags between tag clouds than linking subjects in a hierarchical tree-like category based structure. The fact that there are more overlaps between the tag clouds automatically makes it easier to connect seemingly random contexts.

The dynamics of networked media is so unpredictable that is impossible to tell in advance, whether new concepts and ideas will pick up or extinct. A big part of success is certainly down to luck, however after having tried a number of different approaches I come to the conclusion that simplicity is absolute key for success. Users need to understand within seconds what the project is all about, otherwise they will not bother taking the time and have a closer look. This is particularly important if there is no imminent benefit to the user using the service.

After all, only time can tell whether *synopse* is a sprout that will carry flowers or not - and you can contribute to contribute to its fate: visit <http://emcons.net/synopse/>, download the Firefox extension and start tagging straight away.

Thank you.

Part V

Appendices

A Source Code

The whole project encompasses over 10'000 lines of code of which the most interesting in terms of functionality, namely the ones of the Firefox extension, are displayed here below. The PHP and SQL code cannot be released openly at current stage due to security restrictions and risks.

A.1 Firefox Extension: content/firefoxOverlay.xul, v.0.65

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet href="chrome://synopse/skin/overlay.css" type="text/css"?>
<!DOCTYPE overlay SYSTEM "chrome://synopse/locale/synopse.dtd">
<overlay id="synopse-overlay"
    xmlns="http://www.mozilla.org/keymaster/gatekeeper/there.is.only.xul">
    <script src="overlay.js"/>
    <stringbundle id="stringbundle">
        <stringbundle id="synopse-strings" src="chrome://synopse/locale/synopse.properties"/>
    </stringbundle>

    <menupopup id="menu_ToolsPopup">
        <menuitem id="synopse-hello" label="&synopse.label;"
            oncommand="synopse.onMenuItemCommand(event);"/>
    </menupopup>

    <statusbar id="status-bar">
    <statusbarpanel id="synopse-button" label="Tag!" onclick="synopse.control_synopse_bar()" />
    </statusbar>

    <popupset>
    <tooltip id="synopse-howtotag">
        <description value="Use comma to separate tags. E.g. news, bbc, britain"/>
    </tooltip>
    </popupset>

    <window id="main-window">
    <hbox id="synopse-bar" hidden="true" align="right" insertbefore="status-bar" tooltip="synopse-howtotag">
    <label id="synopse-bar-label">Tag This Site</label><label id="synopse-bar-label-punctuation">:</label>
    <textbox id="synopse-tags" onkeypress="synopse.check_enter_key(event.keyCode)"/>
    <!--<button id="synopse-button" oncommand="synopse.submit_tags()" label="Tag!" /> -->
```

```

</hbox>
</window>
</overlay>

```

A.2 Firefox Extension: content/options.xul, v.0.65

```

<?xml version="1.0"?>
<?xml-stylesheet href="chrome://global/skin/" type="text/css"?>

<prefwindow xmlns="http://www.mozilla.org/keymaster/gatekeeper/there.is.only.xul" \
id="synopse-prefs" title="Options">

<script src="overlay.js"/>

<prefpane id="synopse-prefs-pane" label="Preferences">
<preferences>
  <preference id="static_tags" name="static_tags" type="string"/>
</preferences>

  <hbox align="center">
    <label control="symbol" value="Enter your static tag(s) separated by commas: "/>
    <textbox preference="static_tags" id="static_tags"/>
  </hbox>

<button label="Save" oncommand="synopse.save_prefs(); return false;"/>
</prefpane>
</prefwindow>

```

A.3 Firefox Extension: content/overlay.js, v.0.65

```

var synopse = {

  prefs: null,
  static_tags: "",

  onLoad: function() {
    // initialization code
    this.initialized = true;
    this.manually_enabled = false;
    this.strings = document.getElementById("synopse-strings");

    this.prefs = Components.classes["@mozilla.org/preferences-service;1"]
      .getService(Components.interfaces.nsIPrefService)
      .getBranch("extensions.synopse@christoph.burgdorfer.");
    this.prefs.QueryInterface(Components.interfaces.nsIPrefBranch2);

    this.prefs.addObserver("", this, false);
  }
};

```

```

// document.getElementById("debug").value = this.prefs.getCharPref("static_tags").toLowerCase();

},

onMenuItemCommand: function(e) {
    var promptService = Components.classes["@mozilla.org/embedcomp/prompt-service;1"]
        .getService(Components.interfaces.nsIPromptService);
    promptService.alert(window, this.strings.getString("helloMessageTitle"),
        this.strings.getString("helloMessage"));
},

control_synopse_bar: function () {
    if(document.getElementById("synopse-bar").hidden == false) {
        document.getElementById("synopse-bar").hidden = true;
        this.manually_enabled = false;
    } else {
        document.getElementById("synopse-bar").hidden = false;
        this.manually_enabled = true;
    }
},

check_enter_key: function(keycode) {
    if(keycode == 13) { this.submit_tags(); }
},

submit_tags: function() {
    var req = new XMLHttpRequest();

    req.open('POST', 'http://emcons.net/synopse/submit.php?', true);
    req.setRequestHeader("Content-Type", "application/x-www-form-urlencoded");

    req.onreadystatechange = function (aEvt) {
        if (req.readyState == 4) {
            if(req.status == 200)
                alert(req.responseText);
            else
                alert("Error loading page\n");
        }
    };

    // alert("static tags: "+this.prefs.getCharPref("static_tags").toLowerCase());
    var static_tags = "";
    if(this.prefs.getCharPref("static_tags")) {
        static_tags = this.prefs.getCharPref("static_tags").toLowerCase()+" ";
    }
    req.send('s='+escape(content.location)+'&u='+escape(static_tags+document. \
        getElementById("synopse-tags").value)+'&cmd=add_rating');
    document.getElementById("synopse-tags").value = "";
},

```



```

save_prefs: function() {
this.prefs.setCharPref("static_tags", document.getElementById("static_tags").value);
alert("saved: "+document.getElementById("static_tags").value);
}
};

window.addEventListener("load", function(e) { synopsis.onLoad(); }, false);

window.addEventListener("scroll", function(e) {

// Yoffset -> how much has been scrolled down?
// innerHeight -> the height of the window
// scrollHeight -> height of the whole page

// when scrolling hits the bottom.
if(((content.window.pageYOffset+content.window.innerHeight)>=content.document.body.scrollHeight) &&
(content.document.body.scrollHeight != content.window.innerHeight)) {
// show the bar
document.getElementById("synopse-bar").hidden = false;
}

// if scrolling up.
else if(((content.window.pageYOffset+content.window.innerHeight)<content.document.body.scrollHeight-20)
&& (synopse.manually_enabled == false)) {
// hide the bar
document.getElementById("synopse-bar").hidden = true;
}

}, false);

```

A.4 Firefox Extension: install.rdf, v.0.65

```

<?xml version="1.0" encoding="UTF-8"?>
<RDF xmlns="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:em="http://www.mozilla.org/2004/em-rdf#">
  <Description about="urn:mozilla:install-manifest">
    <em:id>synopse@christoph.burgdorfer</em:id>
    <em:name>synopse</em:name>
    <em:version>0.65</em:version>
    <em:creator>Christoph Burgdorfer</em:creator>
    <em:description>Tag the Web!</em:description>
    <em:homepageURL>http://emcons.net</em:homepageURL>
    <em:updateURL>http://emcons.net/synopse/update.rdf</em:updateURL>
    <em:targetApplication>
      <Description>
        <em:id>{ec8030f7-c20a-464f-9b0e-13a3a9e97384}</em:id> <!-- firefox -->
        <em:minVersion>1.5</em:minVersion>
        <em:maxVersion>2.0.0.*</em:maxVersion>
      </Description>
    </em:targetApplication>
  </Description>
</RDF>

```

```
</em:targetApplication>  
<em:optionsURL>chrome://synopse/content/options.xul</em:optionsURL>  
</Description>  
</RDF>
```

B Statistical Data

B.1 Mozilla's Firefox Global Usage

Worldwide	June 2007	January 2007	Difference
Explorer	84.66%	85.81%	-1.15%
Firefox	12.72%	11.69%	1.03%
Safari	1.79%	1.64%	0.15%
Opera	0.61%	0.58%	0.03%
Netscape	0.11%	0.13%	-0.02%

USA	June 2007	January 2007	Difference
Explorer	75.69%	78.13%	-2.44%
Firefox	19.65%	16.11%	3.54%
Safari	3.77%	3.68%	0.09%
Opera	0.61%	0.73%	-0.12%
Netscape	0.17%	0.18%	-0.01%

Canada	June 2007	January 2007	Difference
Explorer	75.76%	79.00%	-3.24%
Firefox	16.47%	14.13%	2.34%
Safari	5.72%	4.70%	1.02%
Opera	0.69%	0.71%	-0.02%
Netscape	0.13%	0.14%	-0.01%

UK	June 2007	January 2007	Difference
Explorer	86.00%	86.72%	-0.72%
Firefox	11.22%	10.86%	0.36%
Safari	1.61%	1.78%	-0.17%
Opera	0.53%	0.49%	0.04%
Netscape	0.10%	0.10%	0.00%

Australia	June 2007	January 2007	Difference
Explorer	66.42%	65.71%	0.71%
Firefox	26.32%	26.68%	-0.36%
Safari	1.86%	1.77%	0.09%
Opera	4.05%	4.28%	-0.23%
Netscape	0.24%	0.24%	0.00%

Germany	June 2007	January 2007	Difference
Explorer	68.84%	69.47%	-0.63%
Firefox	26.69%	25.66%	1.03%
Safari	1.25%	1.22%	0.03%
Opera	1.88%	2.10%	-0.22%
Netscape	0.19%	0.22%	-0.03%

France	June 2007	January 2007	Difference
Explorer	83.32%	83.29%	0.03%
Firefox	14.05%	13.53%	0.52%
Safari	1.68%	1.41%	0.27%
Opera	0.78%	0.86%	-0.08%

Italy	June 2007	January 2007	Difference
Explorer	78.11%	79.48%	-1.37%
Firefox	18.38%	15.89%	2.49%
Safari	2.08%	2.10%	-0.02%
Opera	1.24%	1.41%	-0.17%
Netscape	0.14%	0.33%	-0.19%

Belgium	June 2007	January 2007	Difference
Explorer	76.07%	80.92%	-4.85%
Firefox	19.91%	15.38%	4.53%
Safari	2.90%	2.31%	0.59%
Opera	0.95%	0.67%	0.28%
Netscape	0.11%	0.11%	0.00%

the Netherlands	June 2007	January 2007	Difference
Explorer	88.49%	88.66%	-0.17%
Firefox	9.75%	9.41%	0.34%
Safari	1.25%	1.19%	0.06%
Opera	0.39%	0.36%	0.03%
Netscape	0.08%	0.08%	0.00%

Source: OneStat.com[10]

C Communication

C.1 Focus Group Mail

Hello dear friend,

I hope you are doing well! We may not have spoken for a while. Anyways, you may or may not know that I am currently working hard on a dissertation here in the UK. I am heading towards my hand-in date and the time has come to perform some beta-testing of my recent work. I would like to invite you to take part in the beta testing of my application. The idea is simple:

If you are using Firefox as your browser, you can install the extension which I have developed. Just click here⁶⁹. This extension lets you tag websites. No need to register, no need to login, no password to remember, no frills. Cut the crap. Out of the box. Straight to the point.

Now you may think, what happens to the tags? Of course, there is a search engine (which I am still working on at the moment) that allows you to browse tags and group sites. You find it under <http://emcons.net/synopse/>

Feel free to play around with it! But please remember: The quality is only as good as the tags you leave.

But how do I find my own tagged sites? Here's the trick: In Firefox under "Tools" - "Add-Ons" - "Synopsis" - "Preferences", you can setup your personal "Static Tag". This tag gets added to all the sites you're tagging. So if you look for your static tag in the search engine, you will find your sites.

There's much more you can do with this application. Use your imagination. More features will come in the near future such as RSS feeds for tags and domains as well as the ability to report faulty tags or even delete them straight away. You'll be thrilled ... and Google scared! :)

Any beta testing and critical feedback is warmly welcome!

Thanks for your help testing my project and I am looking forward to hear from you soon - in any case - whether you tested or not :)

Best wishes, Christoph

PS: Feel free to share with your friends, forward, digg it, reddit it, del.icio.us it, create a facebook fan group, write a book or blog about it. PPS: Don't even think of hacking it - it's beta! ;) PPPS: If you happen to be one of my friends-who-work-at-google: shhhh! ;)

⁶⁹<http://emcons.net/downloads/latest.php>

D Screenshots

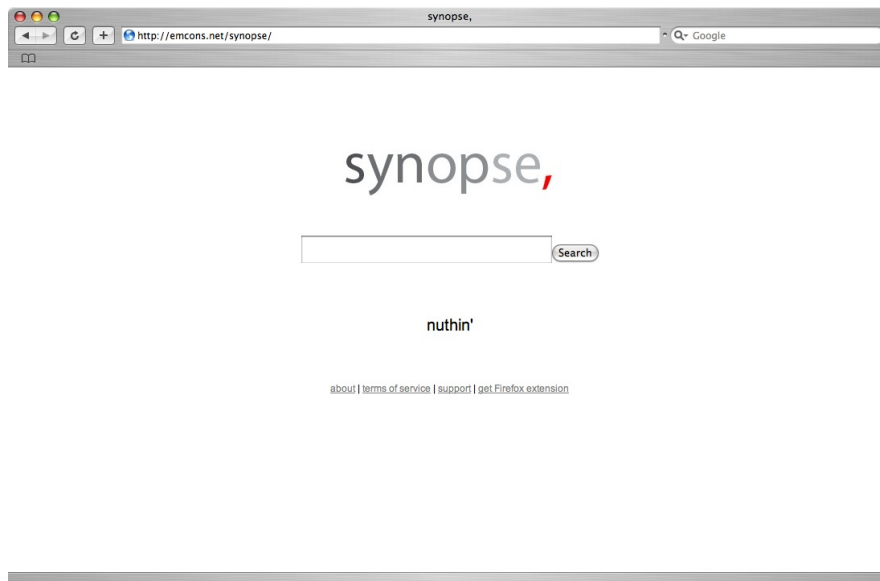


Figure 19: Synopse Home Screenshot



Figure 20: Synopsis Results Screenshots (beta)

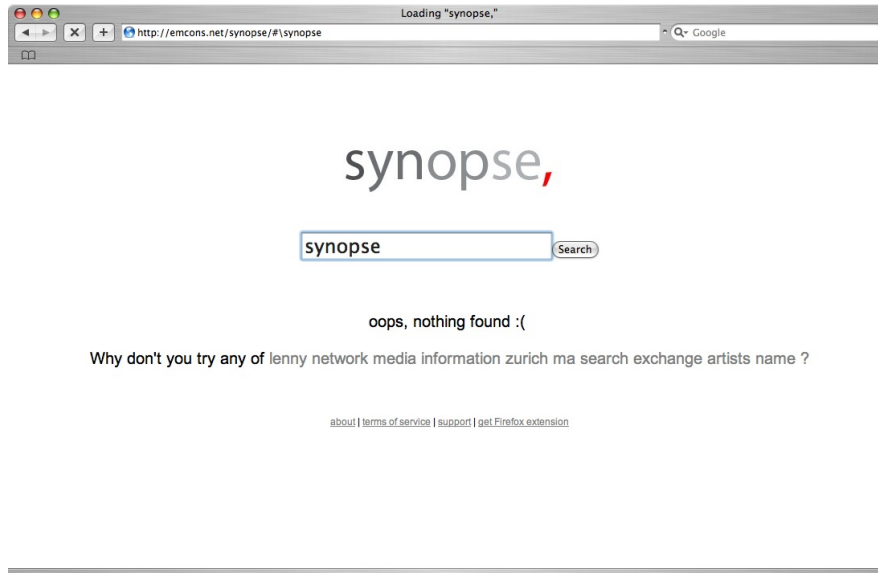


Figure 21: Synopsis Nothing Found Screenshot (beta)

E Links

The following links are relevant to the project:

Homepage - <http://www.emcons.net/synopse/>

About the Project - <http://emcons.net/synopse/about.php#about>

Terms of Service - <http://emcons.net/synopse/about.php#terms>

Support and FAQ - <http://emcons.net/synopse/about.php#support>

Goodies - <http://emcons.net/synopse/goodies.php>

Download Firefox Extension - <http://emcons.net/synopse/download/latest.php>

Part VI

References and Lists

References

- [1] Crossing the chasm. Available from World Wide Web: <http://lawsofsimplicity.com/wp-content/uploads/2006/09/chasm.jpg> [cited 1st September 2007].
- [2] Dictionary: Serendipity. Available from World Wide Web: <http://dictionary.reference.com/search?r=2q=Serendipity> [cited 1st September 2007].
- [3] Technology adaption curve. Available from World Wide Web: <http://upload.wikimedia.org/wikipedia/en/4/45/DiffusionOfInnovation.png> [cited 1st September 2007].
- [4] John Alderman. *Sonic Boom: Napster, P2P and the Battle for the Future of Music*. Fourth Estate, new ed edition, May 2002.
- [5] Chris Anderson. *Long Tail, The: Why the Future of Business Is Selling Less of More*. Hyperion, July 2006.
- [6] Albert-Laszlo Barabasi. *Linked: How Everything Is Connected to Everything Else and What It Means*. Plume, April 2003.
- [7] John Battelle. *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*. Portfolio Hardcover, September 2005.
- [8] Tim Berners-Lee. Developerworks interviews. Available from World Wide Web: <http://www-128.ibm.com/developerworks/podcast/dwi/cm-int082206.txt> [cited 7th August 2007].
- [9] Howard Bloom. *Global Brain: The Evolution of Mass Mind from the Big Bang to the 21st Century*. Wiley, new ed edition, September 2001.
- [10] Niels Brinkman. Mozilla's firefox global usage share is still growing according to onestat.com. Available from World Wide Web: http://www.onestat.com/html/aboutus_pressbox53-firefox-mozilla-browser-market-share.html [cited 11th August 2007].
- [11] Charles Darwin. *The Origin of Species*. Gramercy Books, new ed edition, June 1995.
- [12] Cory Doctorow. Boingboing - technorati tags: three great services on one page. Available from World Wide Web: http://www.boingboing.net/2005/01/13/technorati_tags.thre.html [cited 25th August 2007].
- [13] S. T. Dumais, J. Platt, D. Heckerman, and M. Sahami. A bayesian approach to filtering junk e-mail. In *Learning for Text Categorization: Papers from the 1998 Workshop, AAAI Technical Report*, 1998.
- [14] Facebook. Applications help. Available from World Wide Web: <http://www.facebook.com/help.php?page=25> [cited 18th August 2007].
- [15] Yahoo! Finance. ebay acquires stumbleupon. Available from World Wide Web: <http://biz.yahoo.com/bw/070530/20070530006201.html> [cited 22nd August 2007].
- [16] Dan Gillmor. *We the Media: Grassroots Journalism By the People, For the People*. O'Reilly Media, Inc., new ed edition, 2006.
- [17] Malcolm Gladwell. *The Tipping Point: How Little Things Can Make a Big Difference*. Back Bay Books, 2002.

- [18] Nature Magazine Jim Giles. Internet encyclopaedias go head to head. Available from World Wide Web: <http://www.nature.com/news/2005/051212/full/438900a.html> [cited 9th of August 2007].
- [19] Martin Krafft. *The Debian System: Concepts and Techniques*. No Starch Press, 1 edition, September 2005.
- [20] Rajeev Motwani Lawrence Page, Sergey Brin and Terry Winograd (1999). The pagerank citation ranking: Bringing order to the web. Available from World Wide Web: <http://dbpubs.stanford.edu:8090/pub/showDoc.Fulltext?lang=en&doc=1999-66&format=pdf&compression=> [cited 8th August 2007].
- [21] John Maeda. *The Laws of Simplicity*. The MIT Press, August 2006.
- [22] Geoffrey A. Moore. *Crossing the Chasm*. Collins, revised edition, August 2002.
- [23] Daniel Nissano. *Futureshop: How the New Auction Culture Will Revolutionize the Way We Buy, Sell, and Get the Things We Really Want*. Penguin Press, 2006.
- [24] Donald A. Norman and Tamara Dunaeff. *Things That Make Us Smart: Defending Human Attributes in the Age of the Machine*. Perseus Books, U.S., new ed edition, March 1994.
- [25] Tim O'Reilly. Web 2.0 compact definition: Trying again. Available from World Wide Web: http://radar.oreilly.com/archives/2006/12/web_20_compact.html [cited 7th of August 2007].
- [26] Tim O'Reilly. What is web 2.0 - design patterns and business models for the next generation of software. Available from World Wide Web: <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html> [cited 8th of August 2007].
- [27] Akamai Press Release. Akamai and jupiterresearch identify '4 seconds' as the new threshold of acceptability for retail web page response times. Available from World Wide Web: http://www.akamai.com/html/about/press/releases/2006/press_10606.html [cited 26th August 2007].
- [28] ComScore Press Release. Firefox vs. internet explorer. Available from World Wide Web: http://www.comscore.com/blog/2007/04/firefox_vs_internet_explorer.html [cited 1st September 2007].
- [29] Douglas Rushkoff. *Open Source Democracy: How Online Communication is Changing Offline Politics*. Demos, October 2003.
- [30] James Surowiecki. *The Wisdom of Crowds: Why the Many Are Smarter Than the Few*. Abacus, new ed edition, March 2005.
- [31] Don Tapscott and Anthony D. Williams. *Wikinomics: How Mass Collaboration Changes Everything*. Portfolio Hardcover, December 2006.
- [32] Technorati. Technorati: About us. Available from World Wide Web: <http://www.technorati.com/about/> [cited 25th August 2007].
- [33] Bruce Tognazzini. First principles of interaction design. Available from World Wide Web: <http://www.asktog.com/basics/firstPrinciples.html> [cited 26th August 2007].
- [34] Thomas Vander Wal. Folksonomy coinage and definition. Available from World Wide Web: <http://www.vanderwal.net/folksonomy.html> [cited 7th August 2007].
- [35] Wikipedia. Wikipedia: del.icio.us. Available from World Wide Web: <http://en.wikipedia.org/wiki/Del.icio.us> [cited 22nd August 2007].
- [36] Harris Wu, Mohammad Zubair, and Kurt Maly. Harvesting social knowledge from folksonomies. In *Hypertext*, pages 111–114, 2006.