

Metadata as a Catalyst: Experiments with Metadata and Search Engines in the Internet Journal, *First Monday*

ROBIN HENSHAW

Metadata Editor, *First Monday*

EDWARD J. VALAUSKAS

Chief Editor, *First Monday*

This study examines metadata as a means to enhance information retrieval in a suite of seven search engines, AltaVista, Excite, Google, HotBot, InfoSeek, Lycos, and Northern Light. Papers were selected from issues of the Internet-only journal *First Monday*, and examined by portions of author and title

in seven search engines, without metadata. Metatags were then added to these papers and the searches were repeated five months later. Metadata alone did not play a significant role in increasing the likelihood of a given paper being indexed or highly ranked by any specific search engine.

Introduction

Individuals have long been gathering collections of information; whether in the form of a library, a corporate filing system, a series of government documents, or simply organized stacks of papers and folders on a desk. Once a collection has been established, the question becomes one of access. How does a user get the most from this assortment of information (Lucas 2000)? Metadata is one solution to the problem of retrieving information about a stored collection of digital information (Turner and Brackbill 1998). Today, metadata most commonly refers to descriptive information about World Wide Web and other electronic resources (Dublin Core Metadata Initiative 1998). It provides a user with a means to discover that a resource exists, and provides details on how it might be obtained or accessed.

The hypertext mark-up language (HTML) provides the means to embed metadata in digital documents. HTML is simply a way of giving directions to Internet-based software on how to display a document on a computer, and how to

relate it to other documents and files with a series of hyperlinks (Musciano and Kennedy 1998). If we assume that HTML is primarily a tool to permit the rapid dissemination of information via the World Wide Web (Turner et al. 1996), then metadata is that part of the entire Internet publishing process that makes digital information highly accessible in a concise fashion.

Hence, HTML metatags are one method of organizing metadata that may aid in the resource discovery process. Metatags are non-displaying, or hidden, HTML tags. Meta information not defined by other HTML HEAD elements may be encoded and subsequently identified by the use of these tags. This practice is intended to offer some degree of control over how a Web page is indexed by search engines. Multiple problems exist with this procedure, most notably, the absence of a single agreed upon way of identifying items of electronic information on the Internet (Paskin 1997). In addition, not all search engines currently take advantage of metadata (Richmond 2001).

This paper examines the effectiveness of metatags applied to a set of articles in *First Monday*, a

peer-reviewed journal on the Internet (<http://firstmonday.org>). A metadata policy was first developed for the publication, using a basic framework for applying metatags in the header of a Web page's HTML (Henshaw 1999), and combined with elements from the Dublin Core Initiative (Dublin Core Metadata Initiative 1999). Metadata was then applied to select articles in *First Monday*, to test if metadata would raise the ranking of these articles in certain Internet search engines. Periodic searches for those specific tagged articles were performed in several popular search engines over a one-year period. The rankings were then evaluated to determine the effectiveness of applying standard and Dublin Core metadata in general, and over time (Milstead and Feldman 1999).

The problem

With over seven million unique Web pages (OCLC Office of Research 2000), there is a real problem for any given Web-based resource to find readers. This abundance of digital information leads to several significant troubles that are still largely unresolved. As noted by December (1994), the ease of creating Internet resources means that "information space" – that is, the finite area of servers, network connections, storage capacity, and software provided by a given Internet protocol and serviced by a broad and global audience – is both saturated and polluted with information. This saturation and pollution effectively hides significant resources from individuals and organizations, both in casual and highly-organized and focused searches.

Search engines were invented as a solution to this problem. Search engines are Internet-based databases that allow you to search for Internet-based resources (largely information found on the World Wide Web) based on entry of a few keywords, phrases, or natural language queries. The most significant and popular search engines include AltaVista (<http://www.altavista.com>), Google (<http://www.google.com>), HotBot (<http://www.hotbot.com>), Lycos (<http://www.lycos.com>), Northern Light (<http://www.northernlight.com>), and Yahoo (<http://www.yahoo.com>), in part because of longevity, scope, and usability (Glossbrenner and Glossbrenner 2001).

Most search engines use a variety of schemes to index the contents of the Internet on a routine

basis, with the use of software tools called spiders or crawlers or some combination of human-software collaboration. Resulting indices are based largely on unique terms and phrases, which in turn are accessible to users in well-designed interfaces allowing for unsophisticated as well as advanced searching, e.g. differing interpretations of Boolean searching. Search results are often displayed in terms of a "relevancy ranking", the output of algorithms analyzing the match of search terms to indexing in a given search engine master directory. This ranking can often lead to confusing results, because search engines and their algorithms inconsistently weigh different components of an Internet resource. Some search engines may value a word or phrase in a title more highly, while others may favor (or disfavor) words or phrases repeated frequently or occurring in close proximity (Bradley 2000). These differentials in search results and "ranking" consistently confuse both casual and expert search engine users.

Some have argued that search engines fail to index a large portion of the Internet and that they routinely incorporate biases into their indexing schemes. Some of this subjectivity manifests itself in geography (Internet resources in the United States favored over resources outside the United States); commercial content (in large part due to more than 80% of Web sites being commercial in nature); and, self-referencing (some search engines show a preference for digital resources that hyperlink or otherwise connect to them) (Lawrence and Giles 1999).

Metadata has been proposed as one solution to this problem. Metadata is essentially information about information, a means for content providers to make their resources more accessible via search engines and other content indexers and a vehicle for Internet users to find exactly what they need (Iannella and Waugh 1997).

The hypothetical solution

In this paper, we test the effectiveness of metadata by examining the retrieval of selected papers in the virtual pages of *First Monday*, an Internet-only, peer-reviewed journal, freely available at <http://firstmonday.org>. Papers were downloaded more than one million times from *First Monday's* server in the year 2000, accessed by about 340,000 unique hosts using the Internet from 160 different

Figure 1: Example of metatags at the article level, *First Monday*, February 2001 (Odlyzko, A. 2001)

```
<META NAME = "Description" CONTENT = "The primacy of Internet connectivity over content will likely mean that the dangers of balkanization are smaller than is often feared. The conclusion is that huge sums being invested by carriers in content are misdirected.">
<META NAME = "Keywords" CONTENT = "Internet content, Internet connectivity, digital convergence, content is king, value of communications, communication industry, content industry, connectivity, point-to-point communication, business model, article">
<META NAME = "DC.Title" CONTENT = "Content is not king">
<META NAME = "DC.Creator" CONTENT = "Odlyzko, Andrew">
<META NAME = "DC.Subject" CONTENT = "Internet content, Internet connectivity, digital convergence, content is king, value of communications, communication industry, content industry, connectivity, point-to-point communication, business model, article">
<META NAME = "DC.Description" CONTENT = "The primacy of Internet connectivity over content will likely mean that the dangers of balkanization are smaller than is often feared. The conclusion is that huge sums being invested by carriers in content are misdirected.">
<META NAME = "DC.Publisher" CONTENT = "Valauskas, Edward J.">
<META NAME = "DC.Publisher" CONTENT = "Dyson, Esther">
<META NAME = "DC.Publisher" CONTENT = "Ghosh, Rishab Aiyer">
<META NAME = "DC.Date" CONTENT = "2001-02-05">
<META NAME = "DC.Type" CONTENT = "text">
<META NAME = "DC.Format" CONTENT = "text/html">
<META NAME = "DC.Identifier" CONTENT =
"http://firstmonday.org/issues/issue6_2/odlyzko/index.html">
<META NAME = "DC.Language" CONTENT = "en">
<META NAME = "DC.Relation" CONTENT = "IsPartOf First Monday, vol. 6, no. 2">
```

countries. Since its debut in May 1996, *First Monday* has published 285 papers, written by 342 different authors, in its five-year history, as well as reviews and interviews. All 60 issues (as of June 2001) are available on the Web.

A scheme for adding metadata to *First Monday*, at the article level, was proposed in 1999 (Henshaw 1999). Since the acceptance of the proposal, metadata has been added to the contents of 30 issues (issues released in 1999, 2000, and 2001) out of all 60 issues (published since May 1996). The metadata program includes the use of Dublin Core metadata. Figure 1 provides an example of metadata from an article in a recent issue.

The use of metadata in *First Monday* logically led us to ask if metadata would lead more readers to the pages of the journal. *First Monday* is well linked on the Internet, with 4,142 hyperlinks to the original address <http://www.firstmonday.dk>, 4,482 links to its shortname <http://firstmonday.dk>, 1,847 links to its current name <http://firstmonday.org>, and 517 links to <http://www.firstmonday.org> (based on search results from AltaVista, 20 May 2001). Our logs indicate that *First Monday* is well read (see Figure 2), but is it because of content, hyperlinks, metadata, or some combination of these and other factors?

Figure 2: Use of *First Monday*, based on log analysis of <http://firstmonday.org> server.

Year	1999	2000
Successful requests	1,055,678	9,900,908
Distinct hosts served	157,669	338,039

We decided to experiment with some of the content of *First Monday* to see if metadata made a difference in search results over time (allowing for activity of spiders and other robots in indexing *First Monday's* content).

Methodology

Search engines use unique words, combinations of words, names, and phrases to index, identify and otherwise collate Internet-based information. We decided to search a series of words and phrases from a group of articles published over the course of *First Monday's* history. These terms included unique author names and phrases that are part of the titles of selected papers (see Figure 3a).

Seven search engines were used (see Figure 4). In advance of each search, "help" pages from each search engine's Web site were examined in order to understand the specific indexing practices and searching procedures (such as phrase parameters and sensitivity to capitalization).

During the course of searching, the results were saved as text files for later reference and analysis. There was no attempt to use special features in any search engine to somehow enhance the results during the course of the search. The total number of "hits" for each search was recorded. The first 100 Web sites (if 100 or more were found) were examined and saved as text files. The ranking of the first specific reference to the exact paper from *First Monday* was noted. Retrieved searches that point to <http://www.firstmonday.dk> were considered exactly equivalent to <http://firstmonday.org>. All search results were saved as text files, dated, and time-stamped.

Given the nature of searching in each of the engines, each phrase was searched with double quotes, so that the entire phrase was searched, rather than each word of the phrase. However,

Figure 3a: Phrases and names selected from selected issues of *First Monday*, published 1996–1999, for detection in AltaVista, Excite, HotBot, InfoSeek, Lycos, and Northern Light.

search term or phrase	source in <i>First Monday</i>
"law review in the age of cyberspace"	Hibbitts, 1996
"Bernard Hibbitts"	Hibbitts, 1996
"lex networkia"	Valauskas, 1996
"Edward Valauskas"	Valauskas, 1996
"spirituality and technology"	Bauwens, 1996
"Michel Bauwens"	Bauwens, 1996
"computers as tutors"	Bennett, 1996
"Frederick Bennett"	Bennett, 1996
"agency of the infozone"	McInnes, 1997
"Alice McInnes"	McInnes, 1997
"shroud of lecturing"	DeLong, 1997
"Stephen DeLong"	DeLong, 1997
"virtual harassment"	Ferganchick-Neufang, 1998
"Julia Ferganchick-Neufang"	Ferganchick-Neufang, 1998
"homesteading the noosphere"	Raymond, 1998b
"Eric S. Raymond"	Raymond, 1998b
"Internet cultural phenomenon"	Hakim et al., 1999
"Toufic Hakim"	Hakim et al., 1999
"between chaos and order"	Hamza and Alhalabi, 1999
"Bassem Alhalabi"	Hamza and Alhalabi, 1999

Figure 4: Internet search engines tested in this study.

Name	URL
AltaVista	http://www.altavista.com
Excite	http://www.excite.com
Google	http://www.google.com
HotBot	http://www.hotbot.com
InfoSeek	http://www.infoseek.go.com/
Lycos	http://www.lycos.com/
Northern Light	http://www.northernlight.com/

HotBot is an exception to this rule; rather than using double quotes, it uses an "exact phrase" option as part of its search interface. All searches were made within the same 24-hour period in all search engines on personal computers with similar speeds of Internet connectivity. After the searches were completed, the results were tabulated and summarized for later comparative use.

The articles selected for this study represent a sample of all articles published in *First Monday*. They were selected for the unique nature of a portion of their titles or the name of an author. An initial search was made of these phrases in the search engines in December 1999, before metatags were inserted into each of the articles searched. Metatags were then devised by *First Monday's* Metadata Editor (see examples in Figure 5) and added within 48 hours of the initial search.

Figure 5: Metadata at the article level, *First Monday*, September 1996 (Hibbitts, B. 1996)

```
<META NAME="Description" CONTENT="An analysis of the development of the law review genre leads to the conclusion that legal writers self-publish on the Internet.">
<META NAME="Keywords" CONTENT="law review, legal writers, student edited law journals, faculty edited law journals, law reviews online, legal online services, law review distribution, scholarly communication, electronic law journals, self-publishing, legal scholarship, article">
<META NAME="DC.Title" CONTENT="Last writes">
<META NAME="DC.Title" CONTENT="The law review in the age of cyberspace">
<META NAME="DC.Creator" CONTENT="Hibbitts, Bernard">
<META NAME="DC.Subject" CONTENT="law review, legal writers, student edited law journals, faculty edited law journals, law reviews online, legal online services, law review distribution, scholarly communication, electronic law journals, self-publishing, legal scholarship, article">
<META NAME="DC.Description" CONTENT="An analysis of the development of the law review genre leads to the conclusion that legal writers self-publish on the Internet.">
<META NAME="DC.Publisher" CONTENT="Munksgaard International Publishers Ltd., Copenhagen">
<META NAME="DC.Date" CONTENT="1996-09-02">
<META NAME="DC.Type" CONTENT="text">
<META NAME="DC.Format" CONTENT="text/html">
<META NAME="DC.Identifier" CONTENT="http://firstmonday.org/issues/issue3/hibbitts/index.html">
<META NAME="DC.Language" CONTENT="en">
<META NAME="DC.Relation" CONTENT="IsPartOf First Monday, vol. 1, no. 3">
```

Searches were then repeated with the same phrases, in the same fashion, with the same sorts of computers and Internet connections, five months later in April 2000. All searches again were completed within a 24-hour period, with information examined, retrieved, and saved for comparison to early results. During this five-month period between December, 1999 and April, 2000, metatags were systematically entered into a number of archival issues of *First Monday*, along with the routine placement of metatags into the contents of each new issue within two weeks of its appearance on the Internet. An additional search was completed in Google after the initial experiment to secure additional comparative data.

Experimental results

While some search engines demonstrated some sensitivity to metatags, notably AltaVista, Excite, and HotBot, there was no clear evidence in the experiments that metatags greatly enhanced the ranking of selected papers from *First Monday*. In each of these search engines, it was possible to note the use of *First Monday's* metatags, by examining the summary description in the search

Figure 6a: Change in ranking of 20 searches between December 1999 and April 2000, based on author and title excerpts from *First Monday*, 1996–1999.

Search engine	Loss in ranking	No change	Gain in ranking
AltaVista	3	17	5
Excite	13	5	2
HotBot	5	10	5
InfoSeek	1	10	9
Lycos	0	19	1
Northern Light	7	10	3

Figure 6b: Change in ranking of 12 searches between July 2000 and May 2001, based on a different suite of author and title excerpts from *First Monday*, 1997–1998, for the search engine Google.

Search engine	Loss in ranking	No change	Gain in ranking
Google	3	6	3

results. However, depending on the search engine’s own timetable for examining and indexing the content of *First Monday*, not all metatagged articles were captured, hence leading to some discrepancies in search results.

Examining the results of the experiments in Appendices 1–6 and the summary in Table 1, there is no universal evidence for the ranking of a given specific paper from *First Monday* improving over time, with the addition of metatags. If there is any trend, each search engine seems to gather more information into its own indexing scheme, which in turn reduces the likelihood of any single paper gaining notable ground in its ranking (see Figure 6a). In other words, more hits means more work in searching for a given specific paper, in spite of the metatags. Given that each of these search engines attempts to index a large part of the World Wide Web, it is not surprising that it becomes more difficult over time to find a given reference to a work, as more and more data enters the “files” of the search engine. Metadata alone does not seem to be the solution.

To test the validity of our assumption that more indexed information tends to defeat the ranking scheme, we tried a new search with different terms. We selected a different suite of papers from *First Monday* (see Figure 3b) and searched for these terms in the highly rated search engine Google. We searched for the terms

Figure 3b: Phrases and names selected from selected issues of *First Monday*, published 1997–1998, for detection in Google.

search term or phrase	source in <i>First Monday</i>
“attention economy and the net”	Goldhaber, 1997
“michael h. goldhaber”	Goldhaber, 1997
“filtering the internet in american libraries”	Bastian, 1997
“jeannette allis bastian”	Bastian, 1997
“filters and the public library”	Minow, 1997
“mary minow”	Minow, 1997
“digital diploma mills”	Noble, 1998
“david f. noble”	Noble, 1998
“cathedral and the bazaar”	Raymond, 1998a
“eric s. raymond”	Raymond, 1998a
“book jacket as access mechanism”	O’Connor and O’Connor, 1998
“brian c. o’connor”	O’Connor and O’Connor, 1998

Figure 7: Comparison of search results for hyperlinks to the *First Monday* Web site in AltaVista and Google with link: search option, May 2001.

	Number of hyperlinks
<i>AltaVista</i>	
www.firstmonday.dk	4,142
firstmonday.dk	4,482
www.firstmonday.org	517
firstmonday.org	1,847
<i>Google</i>	
www.firstmonday.dk	1,490
firstmonday.dk	1,490
www.firstmonday.org	1,490
firstmonday.org	1,490

in July 2000 and then in May 2001, with metatags added to the HTML of the specific papers in *First Monday* shortly after the first sequence of searches. The results are described in detail in Appendix 7 and Figure 6b.

In Google, the rating of most searches did not improve as Google added more sites to its overall database. Two papers, by Minow (1997) and Raymond (1998a), lost ground in their ranking; two papers gained, by Goldhaber (1997) and O’Connor and O’Connor (1998) gained some ground in their ranking in Google. Raymond’s work on Linux and open source development is quite popular, with thousands of references in sites, which may in part explain his loss in ranking, as more and more sites are indexed by Google. However, the same explanation does not exactly work for Minow’s (1997) paper, which saw only a modest increase in the number of sites from July 2000 to May 2001.

Table 1: Summary of the changes in ranking by the search engines listed of the selected articles between December 1999 and April 2000.

Searched phrase (citation)	AltaVista	Excite	HotBot	Infoseek	Lycos	Northern Light
“law review in the age of cyberspace” (Hibbitts 1996)	No change	6	No change.	Not ranked to Number 2.	No change	No change
“Bernard Hibbitts” (Hibbitts 1996)	No change	No change.	No change.	4	No change	-1
“lex networkia” (Valauskas, 1996)	No change	-14	-1	No change	No change	No change
“Edward Valauskas” (Valauskas, 1996)	No change	Number 37 to not ranked.	No change.	No change	No change	No change
“spirituality and technology” (Bauwens, 1996)	+1	No change.	Not ranked to Number 62.	1	Not ranked to Number 1.	No change
“Michel Bauwens” (Bauwens, 1996)	No change	No change.	Not ranked to Number 54.	-1	No change	4
“computers as tutors” (Bennett, 1996)	No change	Number 15 to not ranked.	Number 6 to not ranked.	1	No change	-2
“Frederick Bennett” (Bennett, 1996)	No change	Number 19 to not ranked.	Number 1 to not ranked.	No change	No change	-4
“agency of the infozone” (McInnes, 1997)	Not ranked to Number 1.	-1	No change.	No change	No change	No change
“Alice McInnes” (McInnes, 1997)	Not ranked to Number 3.	-5	1	No change	No change	No change
“shroud of lecturing” (DeLong, 1997)	No change	-21	Number 7 to not ranked.	No change	No change	1
“Stephen DeLong” (DeLong, 1997)	No change	Number 38 to not ranked.	No change.	No change	No change	No change
“virtual harassment” (Ferganchick-Neufang, 1998)	No change	-2	Number 8 to not ranked.	3	No change	No change
“Julia Ferganchick-Neufang” (Ferganchick-Neufang, 1998)	-2	Number 2 to not ranked.	Not ranked to Number 5.	No change	No change	1
“homesteading the noosphere” (Raymond, 1998b)	Not ranked to Number 16.	Number 1 to not ranked.	No change.	Not ranked to Number 12.	No change	-2
“Eric S. Raymond” (Raymond, 1998b)	No change	No change.	No change.	Not ranked to Number 10.	No change	-4
“Internet cultural phenomenon” (Hakim et al., 1999)	No change	Number 1 to not ranked.	Not ranked to Number 3.	Not ranked to Number 1.	No change	No change
“Toufic Hakim” (Hakim et al., 1999)	Not ranked to Number 3.	-15	No change.	No change	No change	-1
“between chaos and order” (Hamza and Alhalabi, 1999)	Number 14 to not ranked.	No change.	No change.	Not ranked to Number 3.	No change	-3
“Bassem Alhalabi” (Hamza and Alhalabi, 1999)	-2	Not ranked to Number 6.	No change.	No change	No change	No change

Google was successful in picking up the papers from *First Monday*, and in ranking them consistently in the first 20 search results. This in part may be due to Google’s idiosyncratic indexing strategy, in providing high ranking to those sites heavily linked by other Web sites (Google 2001). Hyperlinks to *First Monday* from other Web sites are well understood by Google (see Figure 7)

compared to a more traditional search engine like AltaVista. Since Google values these hyperlinks more than other search engines, it is able to better use these links to develop search results that more reasonably match user expectations, and reduce information noise.

Nevertheless, it would be unreasonable to assume that metatags in *First Monday* led to an im-

provement in the ranking in search results of papers by Goldhaber (1997) and O'Connor and O'Connor (1998). Other more complex factors come into play, some of which are hidden in the basic technologies of search engines and their methodologies for capturing information on the World Wide Web. There simply was no evidence that metatags made a difference in the search results for Google, a conclusion we also reached with the results for AltaVista, Excite, HotBot, InfoSeek, Lycos, and Northern Light.

Discussion

How do search engines work? Most search engines include three essential features: a) a means for scanning some portion of the Internet for content; b) a database for storing information about this scanned content; and, c) a vehicle for delivering results to users, providing access to the contents of the database. For this study, certainly the first aspect of search engine activity is crucial.

What exactly do search engines look for in a given Web site? Each search engine examines the contents of a site with a very specific purpose, looking for unique information to identify contents using a minimum of indexing overhead (that is, the number of keywords or phrases to exactly identify a given portion of a site) and time. Some search engines place an emphasis on words in the title of a site, such as AltaVista (Bradley 2000). Other search engines look for special or unusual words or phrases, or the number of times a given word or phrase is repeated in a given number of words on a site. Clearly, no search engine singles out metatags – and metadata – as being the sole means to isolate and identify an Internet resource.

How does this affect search results? As seen in our brief experiments, no single search engine consistently provides search results that are comparable to those from another search engine. Since each search engine uses special kinds of algorithms and other tools, search results can be widely different, depending on the true purpose of a given search engine (such as to draw traffic to a site to view banner ads, rather than yield meaningful results in the first screen). It is no wonder that Internet users are both baffled and troubled by the growth of digital information.

Should metadata matter? Metadata, and its application to the Web in the form of metatags in

HTML, should matter (Lynch 1997; Thomas and Griffin 1998), but obviously it has not had a great deal of impact – yet – on the search for just the right digital document. In part, without a consensus on a metadata standard for Web-based documents and files, there is little hope that search engines will place a greater emphasis on metatags, as part of their overall indexing strategy. Without a systematic routine for applying this hypothetical standard to the contents of the Web, there will be little incentive for search engines to take seriously those few (few in terms of the total number of files on the Internet) documents on the Web that contain content-rich metatags.

Conclusion

For search engines to succeed, metadata has to matter. But for metadata to make a difference, it has to be widely accepted, standardized, and applied in a routine fashion. As Thomas and Griffin (1998) asked, who really will create metadata for the future?

Libraries and librarians have an important role to play in the development of a fundamental metadata standard that will be accepted and used universally by all Internet publishers. Without a standard, and vehicles both automated and human to apply it to new documents and all legacy files, search engines will continue to be a frustrating experience for all parties, from casual Web users to advanced Boolean experts (as noted by Stefik 1999, Borgman 2000, and many others). Projects that apply that Dublin Core metadata to electronic journals, at the Manchester University Press (Apps and MacIntyre 2000), in mathematical journals (Perovic et al. 2000), and *First Monday* (Henshaw 1999), ultimately will prove the value of metadata as a catalyst for information retrieval. Unfortunately, these efforts in isolation will have little effect on the functionality of search engines without the development of effective and utilitarian metadata standards that are then widely used and accepted.

Acknowledgements

This paper benefited from numerous discussions with Karen Zuidema and Nancy R. John on the use of metatags and the significance of metadata.

References

- Apps, A. and R. MacIntyre. 2000. Dublin Core metadata for electronic journals. *Proceedings*, Fourth European Conference on Research and Advanced technology for Digital Libraries, 18–20 September 2000, Lisbon. Available at <http://www.bn.pt/org/agenda/ecdl2000/proceedings/19230093/19230093.pdf>
- Bauwens, M. 1996. Spirituality and Technology: Exploring the Relationship. *First Monday* 1. Available at <http://firstmonday.org/issues/issue5/bauwens/>
- Bennett, F. 1996. Computers as Tutors: Solving the Crisis in Education. *First Monday* 1. Available at <http://firstmonday.org/issues/issue6/intro/>
- Borgman, C. L. 2000. From Gutenberg to the Global Information Infrastructure: Access to information in the networked world. Cambridge, Mass.: MIT Press.
- Bradley, P. 2000. The Relevance of underpants in searching the Web. *Ariadne* 24. Available at <http://www.ariadne.ac.uk/issue24/search-engines/>
- December, J. 1994. Challenges for Web information providers. *Computer-Mediated Communication Magazine* 1: 8–14. Available at <http://www.december.com/cmc/mag/1994/oct/webip3.html>
- DeLong, S. E. 1997. The Shroud of Lecturing. *First Monday* 2. Available at http://firstmonday.org/issues/issue2_5/delong/
- Dublin Core Metadata Initiative. 1999. Dublin Core Metadata Element Set. Available at http://purl.oclc.org/dc/about/element_set.htm
- Dublin Core Metadata Initiative. 1998. A user guide for simple Dublin Core. Available at http://purl.org/DC/documents/working_drafts/wd-guide-current.htm
- Ferganchick-Neufang, J. K. 1998. Virtual Harassment: Women and Online Education. *First Monday* 3. Available at http://firstmonday.org/issues/issue3_2/fergan/
- Glossbrenner, A. and E. Glossbrenner. 2001. Search engines for the World Wide Web. 3rd ed. Berkeley, Calif.: Peachpit Press.
- Goldhaber, M. H. 1997. The Attention Economy and the Net. *First Monday* 2. Available at http://firstmonday.org/issues/issue2_4/goldhaber/
- Google 2001. Why use Google. Available at <http://www.google.com/technology/>
- Hakim, T., Pamela Prentice, Stephen Baker, and John Pauly. 1999. The Internet Cultural Phenomenon: A Bonafide Curricular Theme. *First Monday* 4. Available at http://firstmonday.org/issues/issue4_1/hakim/
- Hamza, M. K. and B. Alhalabi. 1999. Technology and Education: Between Chaos and Order. *First Monday* 4. Available at http://firstmonday.org/issues/issue4_3/hamza
- Henshaw, R. 1999. The First Monday Metadata Project. *Libri* 49: 125–131.
- Hibbitts, B. 1996. Last Writes? Re-Assessing the Law Review in the Age of Cyberspace. *First Monday* 1. Available at <http://firstmonday.org/issues/issue3/hibbitts/>
- Iannella, R. and A. Waugh. 1997. Metadata: Enabling the Internet. Available at <http://www.dstc.edu.au/RDU/reports/CAUSE97/>
- Lawrence, S. and L. Giles 1999. Accessibility and distribution of information on the Web. *Nature* 400: 107–109. Summary available at <http://www.wwwmetrics.com/>
- Lucas, M. 2000. Demystifying metadata. *Mappa.Mundi Magazine*, April. Available at <http://mappa.mundi.net/trip-m/metadata>
- Lynch, C. 1997. Searching the Internet. *Scientific American* 276: 50–56. Available at <http://www.sciam.com/0397issue/0397lynch.html>
- McInnes, A. 1997. The Agency of the InfoZone: Exploring the Effects of a Community Network. *First Monday* 2. Available at http://firstmonday.org/issues/issue2_2/mcinnes/
- Milstead, J. and S. Feldman. 1999. Metadata: Cataloging by any other name ... *Online* 23: 24–31. Available at <http://www.onlineinc.com/onlinemag/OL1999/milstead1.html>
- Minow, M. 1997. Filters and the Public Library: A Legal and Policy Analysis. *First Monday* 2. Available at http://firstmonday.org/issues/issue2_12/minow/
- Musciano, C. and B. Kennedy. 1998. HTML The Definitive guide. 3rd ed. Sebastopol, Calif.: O'Reilly & Associates.
- OCLC (Online Computer Library Center) Office of Research. 2000. The Web Characterization Project. Available at <http://wcp.oclc.org/>
- O'Connor, B. C. and M. K. O'Connor 1998. Book Jacket as Access Mechanism: An Attribute Rich Resource for Functional Access to Academic Books. *First Monday* 3. Available at http://firstmonday.org/issues/issue3_9/oconnor/
- Odlyzko, A. 2001. Content is not king. *First Monday* 6. Available at http://firstmonday.org/issues/issue6_2/odlyzko/
- Paskin, N. 1997. Information Identifiers. *Learned Publishing* 10: 135–156. Available at <http://www.elsevier.co.jp/inca/homepage/about/infoident/>
- Perovic, A., M. Jost, and B. Wegner. 2000. Metadata for electronic journals in the Electronic Library of Mathematics. EULER: European Libraries and Electronic Resources in Mathematical Sciences; Telematics for Libraries Project LB-5609. Available at <http://www.emis.de/projects/EULER/Reports/D24.html>

- Raymond, E. S. 1998a. The Cathedral and the Bazaar. *First Monday* 3. Available at http://firstmonday.org/issues/issue3_3/raymond/
- Raymond, E. S. 1998b. Homesteading the Noosphere. *First Monday* 3. Available at http://firstmonday.org/issues/issue3_10/raymond/
- Richmond, A. 2001. META Tagging for Search Engines. Available at <http://www.wdvl.com/Search/Meta/Tag.html>
- Stefik, M. 1999. The Internet edge: Social, technical, and legal challenges for a networked world. Cambridge, Mass.: MIT Press.
- Thomas, C. F. and L. S. Griffin 1998. Who Will Create The Metadata For The Internet? *First Monday* 3. Available at http://firstmonday.org/issues/issue3_12/thomas/
- Turner, R. C., T. A. Douglass, and A. J. Turner. 1996. README.1ST SGML for writers and editors. Upper Saddle River, N. J.: Prentice Hall PTR.
- Turner, T. P. and L. Brackbill. 1998. Rising to the top: Evaluating the use of the HTML META tag to improve retrieval of World Wide Web documents through Internet search engines. *Library Resources & Technical Services* 42: 258-271.
- Valauskas, E. J. 1996. Lex Networkia: Understanding the Internet Community. *First Monday* 1. Available at <http://firstmonday.org/issues/issue4/valauskas/>

Appendix 1: Experimental Results, AltaVista (<http://www.altavista.com>)

Searched phrase (citation)	Number of sites retrieved		Appearance of specific paper in first 100 listings? If Yes, highest ranking in the first 100 retrieved sites, ranking in []		Change in ranking (citation) between December 1999 and April 2000
	December 1999	April 2000	December 1999	April 2000	
“law review in the age of cyberspace” (Hibbitts 1996)	130	142	No	No	No change
“Bernard Hibbitts” (Hibbitts 1996)	173	505	No	No	No change
“lex networkia” (Valauskas 1996)	17	18	Yes [1]	Yes [1] ^a	No change
“Edward Valauskas” (Valauskas 1996)	67	80	No	No	No change
“spirituality and technology” (Bauwens 1996)	120	29	Yes [4]	Yes [3] ^b	+1
“Michel Bauwens” (Bauwens 1996)	146	166	No	No	No change
“computers as tutors” (Bennett 1996)	345	329	No	No	No change
“Frederick Bennett” (Bennett 1996)	137	153	No	No	No change
“agency of the infozone” (McInnes 1997)	5	8	No	Yes [1] ^c	Not ranked to Number 1.
“Alice McInnes” (McInnes 1997)	9	13	No	Yes [3] ^d	Not ranked to Number 3.
“shroud of lecturing” (DeLong 1997)	31	32	Yes [1]	Yes [1] ^e	No change
“Stephen DeLong” (DeLong 1997)	16	19	No	No	No change
“virtual harassment” (Ferganchick-Neufang 1998)	36	36	No	No	No change
“Julia Ferganchick-Neufang” (Ferganchick-Neufang 1998)	14	17	Yes [14]	Yes [16] ^f	-2
“homesteading the noosphere” (Raymond 1998b)	653	776	No	Yes [16] ^g	Not ranked to Number 16.
“Eric S. Raymond” (Raymond 1998b)	10,478	12,415	No	No	No change
“Internet cultural phenomenon” (Hakim et al. 1999)	11	12	Yes [1]	Yes [1] ^h	No change
“Toufic Hakim” (Hakim et al. 1999)	11	12	No	Yes [3] ⁱ	Not ranked to Number 3.
“between chaos and order” (Hamza and Alhalabi 1999)	32	34	Yes [4]	No	Number 14 to not ranked.
“Bassem Alhalabi” (Hamza and Alhalabi 1999)	15	17	Yes [1]	Yes [3] ^j	-2

a, e, f, j: no indication of metatags noted by AltaVista.

b, c, d, g, h, i: metatags noted by Altavista.

Appendix 2: Experimental Results, Excite (<http://www.excite.com>)

Searched phrase (citation)	Number of sites retrieved		Appearance of specific paper in first 100 listings? If Yes, highest ranking in the first 100 retrieved sites, ranking in []		Change in ranking (citation) between December 1999 and April 2000
	December 1999	April 2000	December 1999	April 2000	
"law review in the age of cyberspace" (Hibbitts 1996)	67	Not provided	Yes, [26]	Yes, [20] ^a	6
"Bernard Hibbitts" (Hibbitts 1996)	633	First 100 documents reviewed	No	No	No change.
"lex networkia" (Valauskas 1996)	17	Not provided	Yes, [1]	Yes, [15] ^b	-14
"Edward Valauskas" (Valauskas 1996)	260	First 100 documents reviewed	Yes, [37]	No	Number 37 to not ranked.
"spirituality and technology" (Bauwens 1996)	133	First 100 documents reviewed	No	No	No change.
"Michel Bauwens" (Bauwens 1996)	143	First 100 documents reviewed	No	No	No change.
"computers as tutors" (Bennett 1996)	187	First 100 documents reviewed	Yes, [15]	No	Number 15 to not ranked.
"Frederick Bennett" (Bennett 1996)	98	76	Yes, [19]	No	Number 19 to not ranked.
"agency of the infozone" (McInnes 1997)	9	Not provided	Yes, [1]	Yes, [2] ^c	-1
"Alice McInnes" (McInnes 1997)	11	Not provided	Yes, [2]	Yes, [7] ^d	-5
"shroud of lecturing" (DeLong 1997)	27	Not provided	Yes, [19]	Yes, [40] ^e	-21
"Stephen DeLong" (DeLong 1997)	54	49	Yes, [38]	No	Number 38 to not ranked.
"virtual harassment" (Ferganchick-Neufang 1998)	14	Not provided	Yes, [1]	Yes, [3] ^f	-2
"Julia Ferganchick-Neufang" (Ferganchick-Neufang 1998)	150	First 100 documents reviewed	Yes, [2]	No	Number 2 to not ranked.
"homesteading the noosphere" (Raymond 1998b)	281	First 100 documents reviewed	Yes, [1]	No	Number 1 to not ranked.
"Eric S. Raymond" (Raymond 1998b)	1,890	First 100 documents reviewed	No	No	No change.
"Internet cultural phenomenon" (Hakim et al. 1999)	13	7	Yes, [1]	No	Number 1 to not ranked.
"Toufic Hakim" (Hakim et al. 1999)	18	Not provided	Yes, [3]	Yes, [18] ^g	-15
"between chaos and order" (Hamza and Alhalabi 1999)	1,597	First 100 documents reviewed	No	No	No change.
"Bassem Alhalabi" (Hamza and Alhalabi 1999)	7	Not provided	No	Yes, [6] ^h	Not ranked to Number 6.

a, f, h: metatags noted by Excite.

b, c, d, g: no indication of metatags noted by Excite.

Appendix 3: Experimental Results, HotBot (<http://www.hotbot.com>)

Searched phrase (citation)	Number of sites retrieved		Appearance of specific paper in first 100 listings? If Yes, highest ranking in the first 100 retrieved sites, ranking in []		Change in ranking (citation) between December 1999 and April 2000
	December 1999	April 2000	December 1999	April 2000	
“law review in the age of cyberspace” (Hibbitts 1996)	74	87	No	No	No change.
“Bernard Hibbitts” (Hibbitts 1996)	100	117	No	No	No change.
“lex networkia” (Valauskas 1996)	7	7	Yes, [4]	Yes, [5] ^a	-1
“Edward Valauskas” (Valauskas 1996)	28	38	No	No	No change.
“spirituality and technology” (Bauwens 1996)	113	134	No	Yes, [62] ^b	Not ranked to Number 62.
“Michel Bauwens” (Bauwens 1996)	60	79	No	Yes, [54] ^c	Not ranked to Number 54.
“computers as tutors” (Bennett 1996)	142	174	Yes, [6]	No	Number 6 to not ranked.
“Frederick Bennett” (Bennett 1996)	58	67	Yes, [1]	No	Number 1 to not ranked.
“agency of the infozone” (McInnes 1997)	3	3	Yes, [1]	Yes, [1] ^d	No change.
“Alice McInnes” (McInnes 1997)	3	4	Yes, [2]	Yes, [1] ^e	1
“shroud of lecturing” (DeLong 1997)	12	18	Yes, [7]	No	Number 7 to not ranked.
“Stephen DeLong” (DeLong 1997)	1	7	No	No	No change.
“virtual harassment” (Ferganchick-Neufang 1998)	8	11	Yes, [8]	No	Number 8 to not ranked.
“Julia Ferganchick-Neufang” (Ferganchick-Neufang 1998)	5	5	No	Yes, [5] ^f	Not ranked to Number 5.
“homesteading the noosphere” (Raymond 1998b)	162	238 ¹	No	No	No change.
“Eric S. Raymond” (Raymond 1998b)	>1,000	>5,000 ¹	No	No	No change.
“Internet cultural phenomenon” (Hakim et al. 1999)	3	3	No	Yes, [3] ^g	Not ranked to Number 3.
“Toufic Hakim” (Hakim et al. 1999)	5	3	No	No	No change.
“between chaos and order” (Hamza and Alhalabi 1999)	77	301	No	No	No change.
“Bassem Alhalabi” (Hamza and Alhalabi 1999)	4	5	No	No	No change.

a, b, c, d, e, f, g: metatags noted by HotBot.

1: first 100 documents reviewed.

Appendix 4: Experimental Results, InfoSeek ([http:// www.infoseek.go.com](http://www.infoseek.go.com))

Searched phrase (citation)	Number of sites retrieved		Appearance of specific paper in first 100 listings? If Yes, highest ranking in the first 100 retrieved sites, ranking in []		Change in ranking (citation) between December 1999 and April 2000
	December 1999	April 2000	December 1999	April 2000	
"law review in the age of cyberspace" (Hibbitts 1996)	13	26	No	Yes, [2]	Not ranked to Number 2.
"Bernard Hibbitts" (Hibbitts 1996)	20 ^a	122	Yes, [11]	Yes, [7]	4
"lex networkia" (Valauskas 1996)	7	10	Yes, [1]	Yes, [1]	No change
"Edward Valauskas" (Valauskas 1996)	23 ^b	28	No	No	No change
"spirituality and technology" (Bauwens 1996)	10 ^c	62	Yes, [2]	Yes, [1]	1
"Michel Bauwens" (Bauwens 1996)	20 ^d	96	Yes, [11]	Yes, [12]	-1
"computers as tutors" (Bennett 1996)	10 ^e	72	Yes, [2]	Yes, [1]	1
"Frederick Bennett" (Bennett 1996)	10 ^f	51	Yes, [3]	Yes, [3]	No change
"agency of the infozone" (McInnes 1997)	3	8	Yes, [1]	Yes, [1]	No change
"Alice McInnes" (McInnes 1997)	7	12	Yes, [1]	Yes, [1]	No change
"shroud of lecturing" (DeLong 1997)	6	20	Yes, [2]	Yes, [2]	No change
"Stephen DeLong" (DeLong 1997)	7	9	No	No	No change
"virtual harassment" (Ferganchick-Neufang 1998)	7	15	Yes, [4]	Yes, [1]	3
"Julia Ferganchick-Neufang" (Ferganchick-Neufang 1998)	2	6	No	No	No change
"homesteading the noosphere" (Raymond 1998b)	57	106	No	Yes, [12]	Not ranked to Number 12.
"Eric S. Raymond" (Raymond 1998b)	10 ^g	1,977	No	Yes, [10]	Not ranked to Number 10.
"Internet cultural phenomenon" (Hakim et al. 1999)	1	6	No	Yes, [1]	Not ranked to Number 1.
"Toufic Hakim" (Hakim et al. 1999)	7	9	Yes, [1]	Yes, [1]	No change
"between chaos and order" (Hamza and Alhalabi 1999)	100 ^h	98	No	Yes, [3]	Not ranked to Number 3.
"Bassem Alhalabi" (Hamza and Alhalabi 1999)	7	11	Yes, [1]	Yes, [1]	No change

a: No total number of sites provided, first 20 sites examined.

b: 23 Web sites retrieved, first 10 examined.

c: No total number of sites provided, first 10 sites examined.

d: No total number of sites provided, first 20 sites examined.

e, f, g: No total number of sites provided, first 10 sites examined.

h: Unknown number of Web sites retrieved, no references to Hamza and Alhalabi (1999) in first 100 examined.

Appendix 5: Experimental Results, Lycos (<http://www.lycos.com>)

Searched phrase (citation)	Number of sites retrieved		Appearance of specific paper in first 100 listings? If Yes, highest ranking in the first 100 retrieved sites, ranking in []		Change in ranking (citation) between December 1999 and April 2000
	December 1999	April 2000	December 1999	April 2000	
“law review in the age of cyberspace” (Hibbitts 1996)	1,343	3,056	No	No	No change
“Bernard Hibbitts” (Hibbitts 1996)	219	484	No	No	No change
“lex networkia” (Valauskas 1996)	4	7	No	No	No change
“Edward Valauskas” (Valauskas 1996)	63	101	No	No	No change
“spirituality and technology” (Bauwens 1996)	5,896	15,085	No	Yes [1]	Not ranked to Number 1.
“Michel Bauwens” (Bauwens 1996)	72	104	No	No	No change
“computers as tutors” (Bennett 1996)	2,202	4,880	No	No	No change
“Frederick Bennett” (Bennett 1996)	3,387	5,710	No	No	No change
“agency of the infozone” (McInnes 1997)	34	52	No	No	No change
“Alice McInnes” (McInnes 1997)	48	48	No	No	No change
“shroud of lecturing” (DeLong 1997)	25	33	No	No	No change
“Stephen DeLong” (DeLong 1997)	458	769	No	No	No change
“virtual harassment” (Ferganchick-Neufang 1998)	1,588	2,485	No	No	No change
“Julia Ferganchick-Neufang” (Ferganchick-Neufang 1998)	24	44	No	No	No change
“homesteading the noosphere” (Raymond 1998b)	124	124	No	No	No change
“Eric S. Raymond” (Raymond 1998b)	5,646	9,590	No	No	No change
“Internet cultural phenomenon” (Hakim et al. 1999)	1,557	2,544	No	No	No change
“Toufic Hakim” (Hakim et al. 1999)	5	6	No	No	No change
“between chaos and order” (Hamza and Alhalabi 1999)	10,305	15,802	No	No	No change
“Bassem Alhalabi” (Hamza and Alhalabi 1999)	5	5	No	No	No change

Appendix 6: Experimental Results, Northern Light (<http://www.northernlight.com>)

Searched phrase (citation)	Number of sites retrieved		Appearance of specific paper in first 100 listings? If Yes, highest ranking in the first 100 retrieved sites, ranking in []		Change in ranking (citation) between December 1999 and April 2000
	December 1999	April 2000	December 1999	April 2000	
"law review in the age of cyberspace" (Hibbitts 1996)	58	249	Yes, [3]	Yes, [3]	No change
"Bernard Hibbitts" (Hibbitts 1996)	466	453	Yes, [6]	Yes, [7]	-1
"lex networkia" (Valauskas 1996)	24	26	Yes, [1]	Yes, [1]	No change
"Edward Valauskas" (Valauskas 1996)	102	118	No	No	No change
"spirituality and technology" (Bauwens 1996)	2,092	2,155	Yes, [2]	Yes, [2]	No change
"Michel Bauwens" (Bauwens 1996)	371	376	Yes, [12]	Yes, [8]	4
"computers as tutors" (Bennett 1996)	652	695	Yes, [3]	Yes, [5]	-2
"Frederick Bennett" (Bennett 1996)	216	267	Yes, [3]	Yes, [7]	-4
"agency of the infozone" (McInnes 1997)	12	13	Yes, [1]	Yes, [1]	No change
"Alice McInnes" (McInnes 1997)	12	17	Yes, [1]	Yes, [1]	No change
"shroud of lecturing" (DeLong 1997)	47	46	Yes, [2]	Yes, [1]	1
"Stephen DeLong" (DeLong 1997)	37	34	No	No	No change
"virtual harassment" (Ferganchick-Neufang 1998)	40	45	Yes, [1]	Yes, [1]	No change
"Julia Ferganchick-Neufang" (Ferganchick-Neufang 1998)	78	59	Yes, [2]	Yes, [2]	1
"homesteading the noosphere" (Raymond 1998b)	769	908	Yes, [3]	Yes, [5]	-2
"Eric S. Raymond" (Raymond 1998b)	7,000	8,113	Yes, [8]	Yes, [12]	-4
"Internet cultural phenomenon" (Hakim et al. 1999)	19	25	Yes, [1]	Yes, [1]	No change
"Toufic Hakim" (Hakim et al. 1999)	32	36	Yes, [1]	Yes, [2]	-1
"between chaos and order" (Hamza and Alhalabi 1999)	515	528	Yes, [22]	Yes, [25]	-3
"Bassem Alhalabi" (Hamza and Alhalabi 1999)	39	45	Yes, [1]	Yes, [1]	No change

Appendix 7: Experimental Results, Google (<http://www.google.com>)

Searched phrase (citation)	Number of sites retrieved		Appearance of specific paper in first 100 listings? If Yes, highest ranking in the first 100 retrieved sites, ranking in []		Change in ranking (citation) between July 2000 and May 2001
	July 2000	May 2001	July 2000	May 2001	
“attention economy and the net” (Goldhaber 1997)	79	177	No	Yes, [1]	Not ranked to Number 1.
michael h. goldhaber (Goldhaber 1997)	257	305	No	Yes, [4]	Not ranked to Number 4.
“filtering the internet in american public libraries” (Bastian 1997)	111	136	Yes, [1]	Yes, [1]	No change
jeannette allis bastian (Bastian 1997)	80	97	Yes, [1]	Yes, [1]	No change
“filters and the public library” (Minow 1997)	167	194	Yes, [1]	Yes, [1]	No change
mary minow (Minow 1997)	193	228	Yes, [5]	Yes, [15]	-10
digital diploma mills (Noble 1998)	1,480	2,110	Yes, [1]	Yes, [1]	No change
david f. noble (Noble 1998)	831	1,840	Yes, [1]	Yes, [1]	No change
cathedral and the bazaar (Raymond 1998a)	7,500	13,500	Yes, [5]	Yes, [8]	-3
eric s. raymond (Raymond 1998a)	17,200	35,000	Yes, [5]	Yes, [16]	-11
“book jacket as access mechanism” (O’Connor and O’Connor 1998)	13	13	Yes, [1]	Yes, [1]	No change
brian c. o’connor (O’Connor and O’Connor 1998)	82	66	Yes, [12]	Yes, [7]	5