

# Internet search engines' response to metadata Dublin Core implementation

Jin Zhang and Alexandra Dimitroff

*School of Information Studies, University of Wisconsin-Milwaukee, USA*

Received 10 March 2004  
Revised 14th April 2004

## Abstract.

The research described here examined performance of major search engines with regard to two groups of web pages: those with metadata and those without metadata. Introduction of metadata Dublin Core, especially the Subject element, should improve retrievability of a web page and, subsequently, its visibility. The visibility of a web page in a search engine results list is defined and used in this study to measure retrievability performance of major search engines. Results show there are significant differences among these two groups in most search engines. The reasons for the differences are analyzed.

**Keywords:** information retrieval; subject searching; world wide web; web sites; metadata; Dublin Core; comparative studies; search engines; recall; visibility

## 1. Introduction

The richness and dynamics of Internet information resources have resulted in a revolution in how

---

*Correspondence to:* School of Information Studies, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53211, USA. E-mail: jzhang@uwm.edu, or dimitrof@uwm.edu

individuals search for and retrieve information. The Internet has arguably become a primary means for people to get information. It is reported that more than 500 billion web pages are available on the Internet and it is not surprising that the number of search engine searches has reached an unprecedented 550 million searches per day [9]. It is apparent that Internet search engines are predominant among information search tools. About 80% of Web users surveyed claim to be using search engines or some kind of search tools to find specific information of interest [14].

There are numerous search engines available on the Internet and their features and performance vary. The evaluation of search engines has attracted the attention of many researchers. Because of the unique characteristics of the Internet, such as its dynamics and diversity, evaluation of Internet search engines is quite different from evaluation of traditional information retrieval systems in many respects. Internet users tend to focus on pages that appear at the top of results lists [13]. Interface quality and user satisfaction with the interface have also been considerations of search engine evaluations [10]. Other related studies include signal detection analysis in examining the quality of search engine results [17], search engine result inconsistency [12], user-oriented evaluation studies on search engines [19, 6], and applying traditional information retrieval system evaluation precision and recall methods to Web search engines [4, 16].

Metadata is widely used in information organization. Metadata describes an information resource or guides access to an information resource. Basically, metadata provides an effective mechanism for describing and locating data that is relevant to a particular interest [1]. Organization of metadata is designed in a flexible and structural electronic form; it allows for multi-faceted

logical organization of information; and it facilitates Internet information identification and extraction [18]. Metadata enhances cross-disciplinary discovery, offers a conceptual framework that supports the modular development of auxiliary metadata components, is used as a global standard for resource discovery, and defines a common ground of principles, procedures, core semantics, and a framework to support interoperable metadata [21]. Metadata should make Internet resources more organized, informative, searchable, and accessible [2]. Metadata techniques have been applied to digital libraries [8, 9, 11], digital museums [22], data/Web mining [3, 15], and information retrieval [5].

It has been asserted that if the general Internet searching public adopts and increases use of a simple metadata standard such as the Dublin Core [7], the precision of information retrieved by search engines should improve substantially [14]. Basically, introduction of Dublin Core is an effort to improve web page visibility in a search engine. However, few studies have been done to test this assertion. While it is clear that there have been many studies done to evaluate search engines, studies of the response of search engines to implementation of metadata Dublin Core have not been done.

The primary objective of this research was to investigate whether and to what degree search performance of major search engines is improved after metadata is implemented.

The proposed general hypothesis was:

H: There is no significant difference for the search engines with respect to search engine visibility performance of web pages before and after the metadata Dublin Core elements are implemented in these web pages.

Search engine visibility of a web page refers to its position in the returned results list of a search engine. Its valid value should be a positive integer.

## 2. Research approach

In order to examine the performance of a search engine, the following steps were planned:

- (a) Identify the web page factors which have played a role in search visibility in a search engine.
- (b) Analyze the characteristics of these factors.
- (c) Select an appropriate test web page.
- (d) Design a group of web pages from the test web page based on these characteristics, making them searchable and available on the Internet. Additional test web pages were derived from the

selected web page. Web pages with different keyword frequencies assigned to Subject fields were derived. The content of the original web page was slightly revised when some derived web pages were generated.

- (e) Choose a group of representative test search engines.
- (f) Submit the derived web pages to these search engines.
- (g) Formulate a special query to search the selected search engines.
- (h) Record responses of these search engines to the designed web pages in terms of their search visibility. The data were collected on a weekly basis.
- (i) Finally, compare and analyze their performance.

In this study, search result visibility of a web page with a certain characteristic was used to measure and judge performance of a search engine with respect to that characteristic.

Toward this aim, two groups of web pages representing two distinctive characteristics (one without metadata and the other with metadata subject implementation) were produced. These pages were derived from an original test web page. After it was slightly revised and modified, the derived web pages were generated for the experimental study. These changes included discarding or adding some keywords to title, full text, or metadata subject field.

For the first group, keywords in both title and full text of a web page were considered. In addition, their frequencies in the title and full text were also considered because a related study showed that keyword occurrence in a title and full text make a remarkable contribution to search visibility in a search engine [23]. The maximum keyword frequencies for title and full text were set to 4 and 5, respectively. That is, 4 web pages for keywords with occurrences 1, 2, 3, and 4 in title, and 5 web pages for keywords with occurrences 1, 2, 3, 4, and 5 in full-text were created for the study, respectively.

For the second group, metadata elements were implemented within the derived web pages. In the Dublin Core [7], Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, and Rights are defined as basic metadata elements for describing web-based information. Among these Dublin Core elements, Title, Subject, and Descriptor play a role in search engine retrieval improvement [24, 25, and 26]. However, the impact of each element on search engine result visibility varies. Impact of some elements is stronger while impact of the others is weaker. Only the

Subject field was identified for this experiment because the element is designed to hold subject and/or keywords which reflect the topic(s) of a web page. Keywords and/or subjects in a Subject field are usually directly crawled by search engines and they are further indexed as indexing terms in their databases. According to previous findings [24], among these Dublin Core elements, Subject field plays a major and crucial role in determining its search visibility in a search engine compared to other elements like the Title and Descriptor fields. The maximum frequency for a keyword in a Subject field was set to 4. In other words, web pages with keyword frequencies 1, 2, 3, and 4 in the Subject field were derived, respectively.

After all test pages were derived, these derived web pages were posted in an Internet public domain so that search engines could access and crawl them. Then, all addresses of the derived web pages were submitted to the selected search engines through each individual web page's http submission feature.

Seven search engines – AllTheWeb ([www.alltheweb.com/](http://www.alltheweb.com/)), EntireWeb ([www.EntireWeb.com](http://www.EntireWeb.com)), Google ([www.google.com](http://www.google.com)), Lycos ([www.lycos.com](http://www.lycos.com)), AltaVista ([www.altavista.com](http://www.altavista.com)), Yahoo ([www.yahoo.com](http://www.yahoo.com)), and Infospace/Fast ([www.infospace.com](http://www.infospace.com)) – were identified and selected. Five of them come from a highly recommended search engine list [20]. These search engines offer a free http address submission mechanism, through which publishers can directly submit their web page addresses. This increases the chances that the posted web pages would be indexed by the selected search engines.

Based on the content of the original web page, a search strategy (query) for the study was formulated. The search query was used to search all of the selected search engines on a weekly basis after the web pages were submitted to the search engines. Web page file name, position in a results list (visibility), search engine name, and other related information were recorded if a web page appeared in a search engine results list. The experiment ended after 21 weeks of data collection.

### 3. Findings and analysis

In this study, we assumed that the involved dependent variable (web page search visibility) in a search engine is normally distributed, the population variances of the dependent variable are the same for all cells, and the case represents random samples and the values of the dependent variable are independent of each other.

The position of a web page in a search engine results list was used to measure performance of the test web pages. Retrieved web pages located at the top of a results list (e.g., in the first few screens) have better visibility than those situated at the bottom of the list. In other words, the smaller the position number of a returned web page in a search results list, the better performance it achieves.

The significance level ( $p$ ) for all tests was 0.05. Regardless of what specific statistic was used, if  $p$  was smaller than 0.05, the finding was statistically significant and the null hypothesis was rejected.

#### 3.1. Examination of hypothesis

The hypothesis states that there is no difference for a search engine with respect to search engine visibility performance between two groups: web pages where the metadata elements are not implemented and web pages where the metadata elements are implemented. To test the proposed hypothesis, each individual search engine was examined. A t-test was used because only two variables were involved. A t-test assesses whether the means of two groups are statistically different from each other. The standard deviation is a statistic that tells people how tightly all the various examples are clustered around the mean in a group of data. The estimate of the standard deviation of sample means is called the standard error of means.

The independent variables for this study were the identified search engines and the dependent variable was web page search visibility in a search engine. The variable  $S\_Engine$  in the following result tables stands for search engine. Its valid values are AllTheWeb (1 and 11 for without metadata and with metadata, respectively), EntireWeb (2 and 12 for without metadata and with metadata respectively), Google (3 and 13 for without metadata and with metadata respectively), Lycos (4 and 14 for without metadata and with metadata respectively), AltaVista (5 and 15 for without metadata and with metadata respectively), Yahoo (6 and 16 for without metadata and with metadata respectively), and Infospace/Fast (7 and 17 for without metadata and with metadata respectively).

Tables 1 and 2 show the statistical results for the t-test. The variable POSITION is defined as returned web page position in a search engine.

From Table 2, Levene's  $F$  is statistically significant ( $F = 46.436$ ,  $p = 0.000 < 0.05$ ). Variances are significantly different so the assumption of equal variances is violated. The null hypothesis was rejected. The means for web pages with metadata elements and those

without metadata elements are 9.3617 and 9.2182 respectively (see Table 1). Notice that even though the hypothesis was rejected, the visibility performance mean of the web pages with metadata elements is larger than the mean of those without metadata elements (negative mean difference  $-0.1435$  in Table 2). This indicates that poor performance of web pages with metadata elements leads to rejection of the hypothesis. In addition, the standard deviations for web pages with metadata elements and those without metadata elements are 7.14577 and 4.67358. This suggests that performance of web pages with metadata elements is not as stable as performance of web pages without metadata elements according to their standard deviations. Figure 1 displays the test results visually. In Figure 1, a boxplot summarizes one or more numeric variables. Each box shows the median, quartiles, and extreme values for one of the summary variables.

Tables 3 and 4 show the statistical results for the EntireWeb t-test. The variable POSITION is defined the same as before. S\_Engines 2 and 12 refer to web pages without metadata elements and those with metadata elements, respectively.

From Table 4, Levene's F is statistically significant ( $F = 19.020, p = 0.000 < 0.05$ ). Variances are significantly different so the assumption of equal variances is violated. The null hypothesis was rejected. The means for the web pages with metadata elements and those without metadata elements are 4.1633 and 8.3141 respectively (see Table 3). This shows that the web

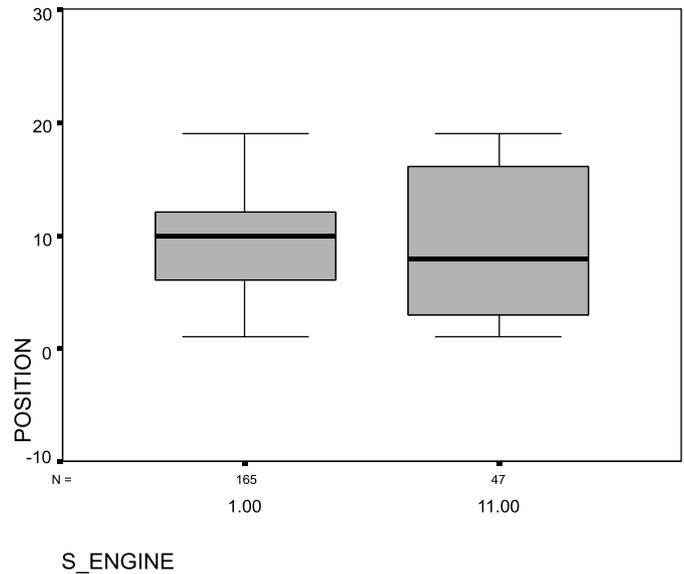


Fig. 1. Boxplots of group distribution changes for AllTheWeb.

pages with metadata elements achieved better visibility performance than those without metadata elements. Performance of web pages with metadata elements (Std. Deviation = 2.58479) is much more stable than performance of web pages without metadata elements (Std. Deviation = 4.56325). The findings confirm the claim that implementation of metadata in a web page

Table 1  
Group statistics for AllTheWeb

	Group statistics				
	S_ENGINE	N	Mean	Std. deviation	Std. error mean
POSITION	1	165	9.2182	4.67358	.36384
	11	47	9.3617	7.14577	1.04232

Table 2  
Independent samples test for AllTheWeb

		Independent samples test				
		Levene's test for equality of variances		t-test for equality of means		
		F	Sig.	t	df	Mean difference
POSITION	Equal variances assumed	46.436	.000	-.163	210	-.1435
	Equal variances not assumed			-.130	57.653	-.1435

Table 3  
Group statistics for EntireWeb

	Group statistics				
	S_ENGINE	N	Mean	Std. deviation	Std. error mean
POSITION	2	156	8.3141	4.56325	.36535
	12	49	4.1633	2.58479	.36926

Table 4  
Independent samples test for EntireWeb

		Independent samples test				
		Levene's test for equality of variances		t-test for equality of means		
		F	Sig.	t	df	Mean difference
POSITION	Equal variances assumed	19.020	.000	6.063	203	4.1508
	Equal variances not assumed			7.991	144.961	4.1508

would improve its visibility in the search engine EntireWeb. Figure 2 displays the results visually.

An independent-samples t test was calculated comparing the mean score of web pages with metadata to the mean score of web pages without metadata in the search engine Google results lists. The statistical data show a difference between the means of the two groups

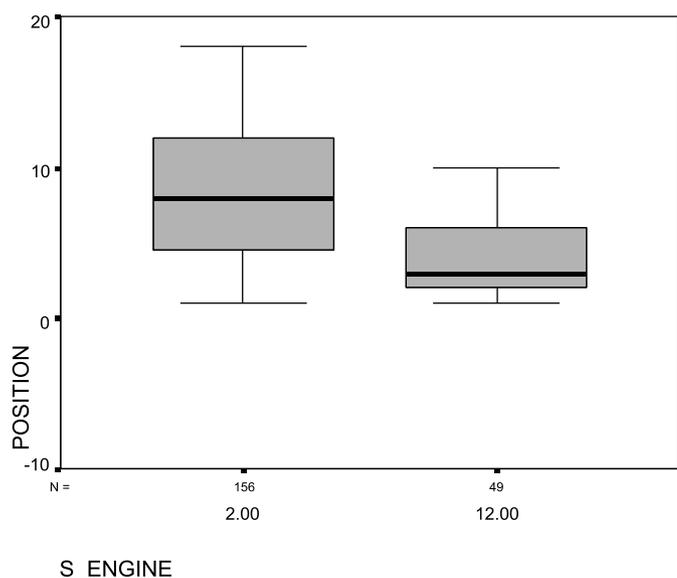


Fig. 2. Boxplots of group distribution changes for EntireWeb.

( $F = 26.522, p = 0.000$  [ $< 0.05$ ]) – see Table 6. The web pages with metadata (Mean = 7.3913, Std. Deviation = 7.03071) achieve better performance than those without metadata (Mean = 7.4588, Std. Deviation = 4.58324) in the search engine Google search results lists (see Table 5). However, the higher standard deviation value of web pages with metadata elements show performance was relatively unstable compared to those without metadata elements. Figure 3 displays the results of this group visually.

The test results for the search engine Lycos are presented in Tables 7 and 8, and Figure 4. It is evident that the hypothesis for the search engine Lycos was rejected because of  $F = 29.143$  and  $p = .000$  ( $< 0.05$ ). This suggests that the variances for the two groups were substantially different. The mean values for web pages without metadata and web pages with metadata are 7.7034 and 3.0357 respectively. Therefore the latter achieves better performance than the former in terms of the search visibility. The column Mean Difference in Table 8 shows to what degree the two groups performed differently in the test. In this case, the mean difference is 4.6677, a relatively large figure.

Tables 9 and 10 demonstrate the statistical results for the search engine AltaVista. The variable POSITION is defined the same as before. S\_Engines 5 and 15 refer to the two groups: web pages without metadata elements and those with metadata elements, respectively.

In Table 10, Levene's F is statistically significant

Table 5  
Group statistics for Google

	Group statistics				
	S_ENGINE	N	Mean	Std. deviation	Std. error mean
POSITION	3	194	7.4588	4.58324	.32906
	13	23	7.3913	7.03071	1.46600

Table 6  
Independent samples test for Google

		Independent samples test				
		Levene's test for equality of variances		t-test for equality of means		
		F	Sig.	t	df	Mean difference
POSITION	Equal variances assumed	26.522	.000	.063	215	.0675
	Equal variances not assumed			.045	24.266	.0675

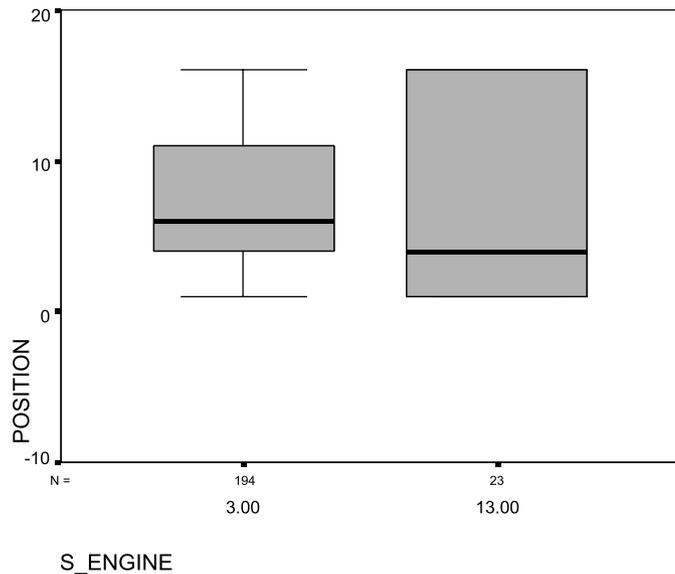


Fig. 3. Boxplots of group distribution changes for Google.

( $F = 1.228$ ,  $p = .271$  [ $< 0.05$ ]). Variances are not significantly different so the assumption of equal variances was accepted. In other words, the proposed null hypothesis was accepted. The means for the web pages with metadata elements and those without metadata elements are 3.9091 and 4.3810 respectively (see Table 9). The standard deviations for web pages with metadata

elements and web pages without metadata elements are 1.70027 and 1.94628, respectively. Their mean difference is 0.4719. It suggests that even though the mean of web pages with metadata elements is smaller than that of web pages without metadata elements, the difference is not small enough to make the difference significant.

Figure 5 displays the AltaVista results visually.

Tables 11 and 12 show the statistical results for Yahoo. The variable POSITION is defined the same as before. S\_Engines 6 and 16 refer to web pages without metadata elements and those with metadata elements, respectively.

From Table 12, Levene's  $F$  is statistically significant ( $F = 7.480$ ,  $p = 0.008$  [ $< 0.05$ ]). Variances are significantly different so the assumption of equal variances is disproved. The means for web pages with metadata elements and those without metadata elements are 1.4000 and 2.5217 respectively (see Table 11). The mean difference in this case is 1.1217 (see Table 12). This shows that web pages with metadata elements achieved better visibility performance than those without metadata elements. Performance of web pages with metadata elements (Std. Deviation = 0.51640) seems more stable than performance of web pages without metadata elements (Std. Deviation = 1.20626). The findings confirm the claim that implementation of metadata in a web page would improve its visibility in Yahoo.

Figure 6 displays the results visually.

Table 7  
Group statistics for Lycos

	Group statistics				
	S_ENGINE	N	Mean	Std. deviation	Std. error mean
POSITION	4	118	7.7034	4.02622	.37064
	14	28	3.0357	1.66627	.31490

Table 8  
Independent samples test for Lycos

		Independent samples test				
		Levene's test for equality of variances		t-test for equality of means		
		F	Sig.	t	df	Mean difference
POSITION	Equal variances assumed	29.143	.000	6.001	144	4.6677
	Equal variances not assumed			9.597	106.474	4.6677

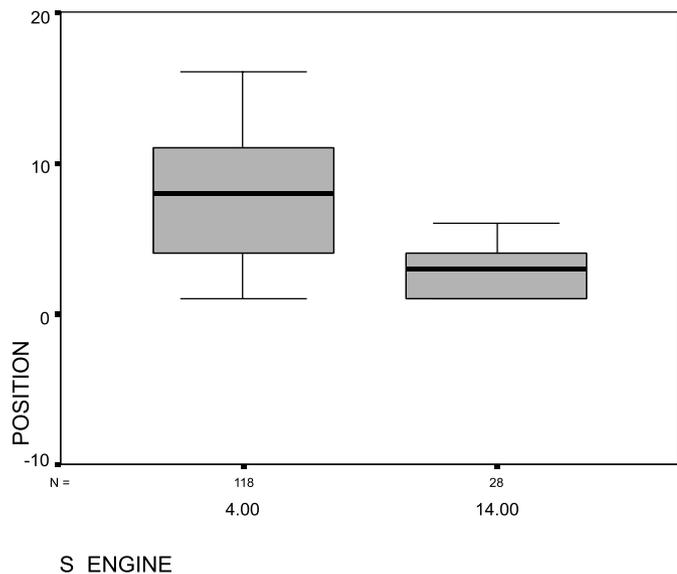


Fig. 4. Boxplots of group distribution changes for Lycos.

Finally, the test results for the search engine Infospace/Fast are presented in Tables 13 and 14, and Figure 7. It is clear that the proposed hypothesis for the search engine Infospace/Fast was rejected because of  $F = 9.225$  and  $p = 0.003 (< 0.05)$  – see Table 14. This implies that the variances for the two compared groups are substantially different. The mean values for web pages

without metadata and web pages with metadata are 5.6034 and 2.6667 respectively (see Table 13). The corresponding mean difference is 2.9368 (see Table 14). As a result, the web page group with metadata elements achieves better performance than the web page group without metadata elements in terms of search results list visibility.

The proposed hypothesis was examined or tested in seven different search engines (AllTheWeb, EntireWeb, Google, Lycos, AltaVista, Yahoo, and Infospace/Fast). The hypothesis was accepted only in one search engine (AltaVista) and rejected in the other six search engines (AllTheWeb, EntireWeb, Google, Lycos, Yahoo, and Infospace/Fast). However, in the search engine AllTheWeb, web pages without metadata elements achieved a better performance than those with metadata elements. In the other five search engines, web pages with metadata elements achieved better performance than those without metadata elements. This implies that metadata implementation in web pages can improve their visibility in the search results lists of most search engines but not all of them.

Basically a mean difference is an indicator that demonstrates the degree to which the web page group with metadata elements achieves better (or worse) performance than the web page group without metadata elements. The larger a mean difference value, the better the group's performance. A negative mean difference indicates worse performance and a positive

Table 9  
Group statistics for AltaVista

	Group statistics				
	S_ENGINE	N	Mean	Std. deviation	Std. error mean
POSITION	5	63	4.3810	1.94628	.24521
	15	11	3.9091	1.70027	.51265

Table 10  
Independent samples test for AltaVista

		Independent samples test				
		Levene's test for equality of variances		t-test for equality of means		
		F	Sig.	t	df	Mean difference
POSITION	Equal variances assumed	1.228	.271	.754	72	.4719
	Equal variances not assumed			.830	14.973	.4719

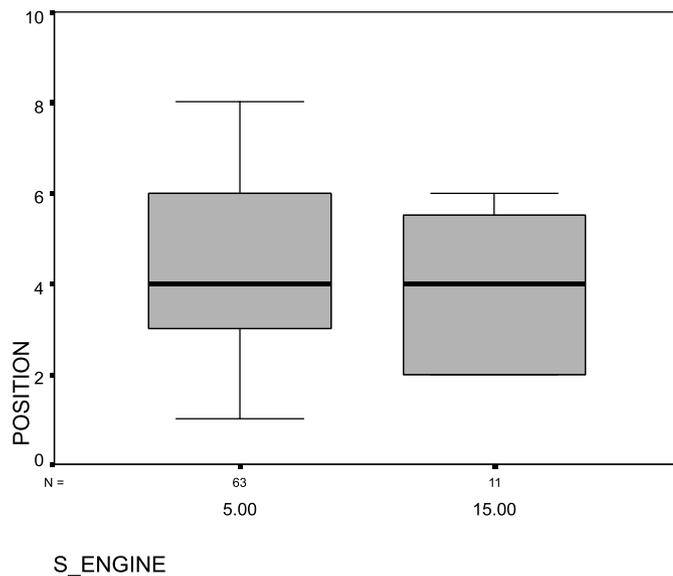


Fig. 5. Boxplots of group distribution changes for AltaVista.

value means better. All mean differences of the examined search engines are listed in Table 15. Lycos (4.6677) and EntireWeb (4.1508) achieve the largest mean differences. In other words, implementation of metadata in web pages has a stronger positive impact on their visibility in search results lists in search engines Lycos and EntireWeb. It has less positive

impact on search engines AltaVista and Google. Notice that implementation of metadata in web pages has a negative impact on their visibility in search results lists in the search engine AllTheWeb due to the negative mean difference value (−0.1435). On the other hand, in spite of the positive mean difference value for AltaVista (0.4719) the difference is not statistically significant. Therefore, the proposed hypothesis for AltaVista was accepted.

Notice that although the mean difference in AltaVista (0.4719) is larger than that in Google (0.0675), the proposed hypothesis for AltaVista was accepted and the proposed hypothesis for Google was rejected. This is possible because the value of a t-test is not only determined by the mean difference, but the variability of groups which is related to the data distribution also makes a contribution to the t-test result. The quite different standard deviations for AltaVista (1.94628 and 1.70027) and for Google (4.58324 and 7.03071) confirm this.

#### 4. Conclusion

The use of Internet search engines is one of the most important means available for users to search for information and a variety of search engines are available on the Internet. Evaluation of search engines has become a challenging research topic. Evaluation of search

Table 11  
Group statistics for Yahoo

	Group statistics				
	S_ENGINE	N	Mean	Std. deviation	Std. error mean
POSITION	6	46	2.5217	1.20626	.17785
	16	10	1.4000	.51640	.16330

Table 12  
Independent samples test for Yahoo

		Independent samples test				
		Levene's test for equality of variances		t-test for equality of means		
		F	Sig.	t	df	Mean difference
POSITION	Equal variances assumed	7.480	.008	2.868	54	1.1217
	Equal variances not assumed			4.646	33.569	1.1217

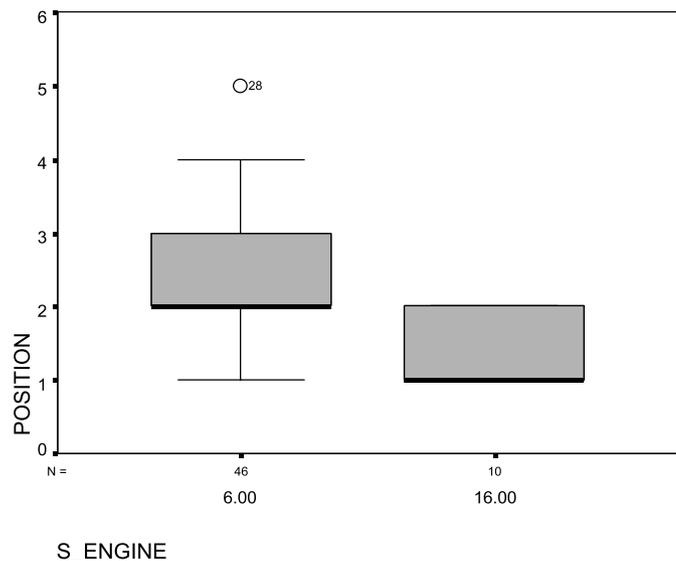


Fig. 6. Boxplots of group distribution changes for Yahoo.

engines can be approached from many perspectives, ranging from search engine interface, to functionality and features, to search visibility, and so on. Metadata was developed as a means to improve access to web page content characteristics, which would result in improvement of a web page's visibility in a search engine's results list (given a reasonable search query).

This study focused on how a search engine responds to implementation of metadata elements (Dublin Core) in web pages. This is a crucial factor by which to measure a search engine, a factor not previously examined.

In this study, a general hypothesis was proposed and it was examined using seven different search engines (AllTheWeb, EntireWeb, Google, Lycos, AltaVista, Yahoo, and Infospace/Fast). Two distinct web page test groups (with metadata and without metadata) were generated for examination. Visibility performance of search engines was examined. Final statistical results show that the hypothesis was accepted in only one search engine (AltaVista) and rejected in the other six search engines (AllTheWeb, EntireWeb, Google, Lycos, Yahoo, and Infospace/Fast). It is worth pointing out that although the proposed hypothesis was accepted in AltaVista (it suggests that there is no significant difference between the web group with metadata and the web group without metadata), the web group with metadata still outperformed the web group without metadata but not significantly. In the group of search engines where the proposed hypothesis was rejected, notice that AllTheWeb rejected the hypothesis in a totally different way from the other search engines. Contrary to our expectation, the web group without metadata outperformed the web group with metadata. It is not surprising that each search engine responds to metadata differently because each search engine has its own algorithm to handle metadata. Generally speaking,

Table 13  
Group statistics for Infospace/Fast

	Group statistics				
	S_ENGINE	N	Mean	Std. deviation	Std. error mean
POSITION	7	58	5.6034	2.88942	.37940
	17	6	2.6667	.81650	.33333

Table 14  
Independent samples test for Infospace/Fast

		Independent samples test				
		Levene's test for equality of variances		t-test for equality of means		
		F	Sig.	t	df	Mean difference
POSITION	Equal variances assumed	9.225	.003	2.463	62	2.9368
	Equal variances not assumed			5.815	22.966	2.9368

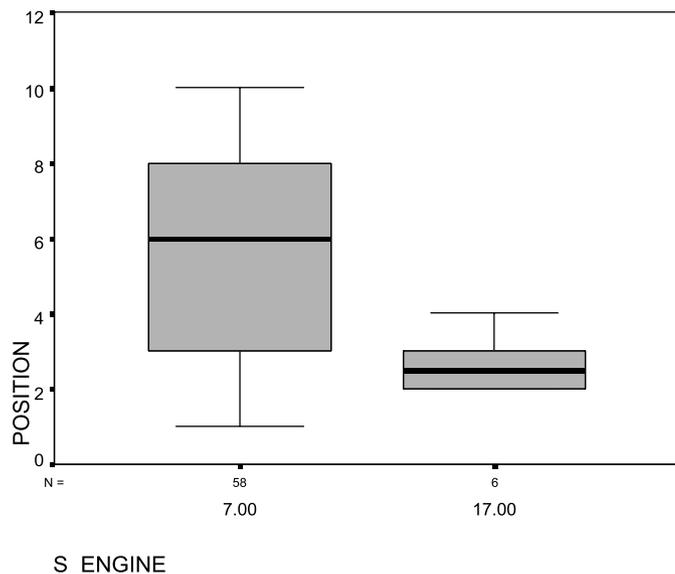


Fig. 7. Boxplots of group distribution changes for Infospace/Fast.

the introduction of metadata Subject content to web pages will improve visibility performance in most search engines but not all.

In a prior study, findings [24] demonstrate that implementation of metadata can effectively improve

visibility of search results lists in search engines. The difference between this study and the prior study is that in the prior study all involved search engines were treated as a whole while in this study each search engine was treated separately.

As we mention in Section 2, Subject was the only element from the Dublin Core set used in this study. Other Dublin Core elements such as Title and Descriptor may also play a role in terms of visibility of search engine results lists. If these elements had factored in the study, the study results may have been different.

Future research on this topic includes, but is not limited to, testing other metadata elements such as Title and Descriptor and increasing the number of test search engines.

### Acknowledgement

This research was sponsored by an IMLS National Leadership Grant (#NR-10012-01).

### References

- [1] K. Burnett, K.B. Ng and S. Park, A comparison of the two traditions of metadata development, *Journal of the American Society for Information Science* 50(13) (1999) 1209–1217.

Table 15  
Mean differences for the search engines

	Lycos	EntireWeb	Infospace/Fast	Yahoo	AltaVista	Google	AllTheWeb
Mean difference	4.6677	4.1508	2.9368	1.1217	0.4719	0.0675	-0.1435

- [2] K.S. Candan, H. Liu, and R. Suvarna, Resource description framework: metadata and its applications, *ACM SIGKDD Explorations Newsletter* 3(1) (2001) 1–19.
- [3] S. Chakrabarti, Data mining for hypertext: a tutorial survey, *ACM Newsletter of SIGKDD* 1(2) (2000) 1–11.
- [4] S.J. Clarke and P. Willett, Estimating the recall performance of Web search engines, *ASLIB Proceedings* 49 (1997) 184–189.
- [5] E.M. Cortez, Use of metadata vocabularies in data retrieval: from USDA's Research, Education, Economic Information System, *Journal of the American Society for Information Science* 50(13) (1999) 1218–1223.
- [6] W. Ding and G. Marchionini, A comparative study of web search service performance. In: S. Hardin (Ed.), *The American Society for Information Science Meeting 1996: Proceedings of 59th Annual Meeting of the American Society for Information Science* (Softbound, Baltimore, 1996).
- [7] *Dublin Core Initiative and Metadata* (2004). Available at: <http://dublincore.org/> (accessed 22 March 2004).
- [8] N. Dushay, Models and tools for generating digital libraries: localizing experience of digital content via structural metadata. In: W. Hersh (ed.), *Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries* (ACM, Portland, 2002).
- [9] L. Grossman, Search and Destroy, *Time* 162(25) (2003) 46–50.
- [10] M. Hearst, Interfaces for searching the Web, *Scientific American* (1997) 68–72.
- [11] L.L. Hill, G. Janeé, R. Dolin, J. Frew and M. Larsgaard, Collection metadata solutions for digital library applications: Alexandria Digital Library, *Journal of the American Society for Information Science* 50(13) (1999) 1169–1181.
- [12] *Inconsistencies Reports* (2004). Available at: [www.searchengineshowdown.com/inconsistent.shtml](http://www.searchengineshowdown.com/inconsistent.shtml) (accessed 11 January 2004).
- [13] J.M. Kleinberg, Authoritative sources in a hyperlinked environment, *Journal of the ACM* 46(5) (1999) 604–632.
- [14] M. Kobayashi and K. Takeda, Information retrieval on the Web, *ACM Computing Surveys* 32(2) (2001) 144–173.
- [15] R. Kosala and H. Blockeel, Web mining research: a survey, *ACM SIGKDD Explorations Newsletter* 2(1) (2000) 1–15.
- [16] H.V. Leighton and J. Srivastava, First 20 precision among World Wide Web search services (search engines), *Journal of the American Society for Information Science* 50(10) (1999) 870–881.
- [17] C. Schlichting and E. Nilsen, *Signal Detection Analysis of WWW Search Engines* (2004). Available at: [www.microsoft.com/usability/webconf/schlichting/schlichting.htm](http://www.microsoft.com/usability/webconf/schlichting/schlichting.htm) (accessed 22 March 2004).
- [18] T.R. Smith, The Meta-Information Environment of digital Libraries, *D-Lib Magazine* July/August (1996). Available at: [www.dlib.org/dlib/july96/new/07smith.html](http://www.dlib.org/dlib/july96/new/07smith.html) (accessed 22 March 2004).
- [19] L.T. Su, A comprehensive and systematic model of user evaluation of Web search engines: I. theory and background, *Journal of the American Society for Information Science and Technology* 54(13) (2003) 1175–92.
- [20] *The Major Search Engines* (2004). Available at: [www.searchenginewatch.com/links/major.html#others](http://www.searchenginewatch.com/links/major.html#others) (accessed 11 January 2004).
- [21] S. Weibel, The Dublin Core metadata initiative: mission, current activities, and future directions, *D-Lib Magazine* 6(12) (2000). [www.dlib.org/dlib/december00/weibel/12weibel.html](http://www.dlib.org/dlib/december00/weibel/12weibel.html) (accessed 22 March 2004).
- [22] M.L. Zeng, Metadata elements for object description and representation: a case report from a digitized historical fashion collection project at Kent State University Museum, *Journal of the American Society for Information Science* 50(13) (1999) 1193–1208.
- [23] J. Zhang and A. Dimitroff, The impact of web page content characteristics on the web page visibility in search engine results (Part I), *Information Processing and Management* (2004). Accepted.
- [24] J. Zhang and A. Dimitroff, The impact of metadata implementation on the web page visibility in search engine results (Part II), *Information Processing and Management* (2004). Accepted.
- [25] R. Henshaw and E.J. Valauskas, Metadata as a catalyst: experiments with metadata and search engines in the Internet journal, *First Monday, Libri* 51(2) (2001) 86–101.
- [26] T.P. Turner and L. Brackbill, 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 and Technical Services* 42(4) (1998) 258–271.