

## The Use of Search Engines for Locating Information on Cultural Heritage of Croatia

Danijel Radošević

University of Zagreb, Faculty of Organization and Informatics  
Pavlinska 2, 42000 Varaždin, Croatia  
goran.bubas@foi.hr

Goran Bubaš

University of Zagreb, Faculty of Organization and Informatics  
Pavlinska 2, 42000 Varaždin, Croatia  
igor.balaban@foi.hr

Igor Balaban

University of Zagreb, Faculty of Organization and Informatics  
Pavlinska 2, 42000 Varaždin, Croatia  
miran.zlatovic@foi.hr

### Summary

*In this paper the authors present an overview of how search engines work and the results of the use of different search engines for locating information on the web related to specific artifacts of the cultural heritage of Croatia. It was found that the Google search engine outperformed the competition in locating web sites with information on specific cultural heritage artifacts. In another analysis, selected web pages with general information on Croatian cultural heritage were examined regarding several elements of search engine optimization. Finally, a group of students of information systems were asked what would be their typical activity when they needed information related to cultural heritage and most of them indicated that they would first search the Internet/web. Recommendation is provided regarding the means for increasing the "visibility" of information related to cultural heritage on the Internet/web and the likelihood of this information being found by the users of this medium.*

**Key words:** search engine optimization, cultural heritage

### Introduction

There is considerable interest for developing standards and systems for distributed search and retrieval of cultural heritage information (see, for instance: Moen, 1998; Kando and Adachi, 2004). The worldwide population of Internet

users reached 1.17 billion in June 2007 with Internet penetration of 40% and 69% in the regions of Europe and North America respectively (for the latest statistics see: InternetWorldStats, 2007). The Internet enables cultural heritage information to be shared among a wide range of audiences and this makes immediate access to information, channel capacity, media type, display mode, and adaptability to the type of device of the users some of the important issues that have to be considered in online presentation of cultural heritage (see: Smeulders, 2002). Recently, use of visual navigation with web maps and personalization to user interest (Mac Aoidh *et al.*, 2007), use of the semantic web techniques for searching and annotating cultural heritage collections (Ossenbruggen, 2007) and development of web-based recommender systems (Rutledge *et al.*, 2007) are being developed to facilitate search for cultural heritage information.

The way Internet users locate cultural heritage information on the web is related to their search strategies. Some of the most evident aspects of Internet users web search behaviour can be described as follows (see: White and Iivonen, 1999):

- Users tend to focus on the web sites that are known to them and subsequently proceed in their search by tracing links from those web sites.
- Users browse the web by quickly selecting and following links, perform quick searches and browse through the results when they do not immediately locate what they are searching for.
- Web search often starts with the use of a search engine.
- The simple search terms/statements are preferred in the use of search engines and the modification of those terms/statements is a common reaction to dissatisfying search results.
- The success rate of web searches variable and depends on the search questions and sometimes the users believe that they have located the right information even when that is not true.

More than 60% of adult Internet users in the U.S.A. use a search engine on an average day (Rainie and Shermak, 2005), they trust their search engines (Fallows, 2005) and most of them click on a link within the first page of results, but only 12% of the users continue their search beyond the third page of results (iProspect, 2006). Today, there are a number of web search engines based mostly on keywords that enable full text search. Some of these are world popular, like *Google* and *Yahoo!* etc., but some are local like our Croatian search engines *www.hr* and *Pogodak!*.

This paper will focus on the way search engines collect documents on the web and process them for later retrieval, as well as on web page optimization by use of metadata and other techniques. Also, a brief analysis is performed of what would be typical behavior of an information science student in search of information about popular cultural heritage artifacts of Croatia. Finally, the results of

comparative analysis of several search engine performance in locating selected cultural heritage artifacts is presented.

### **How search engines work?**

Web search engines collect html and other web documents under its domain using so called web crawlers. Web crawlers read web documents, put them into temporary storages (caches) and process them in many ways e.g. by extracting keywords and other features for later retrieval, automatic classification of documents, forming document indexes etc. Therefore search engines try to find needed documents among documents collected by crawlers.

### **Use of search engines**

To facilitate the location of specific documents on the web with a search engine the documents have to be well described. To locate a specific document users try to describe them in a meaningful way and they mostly use queries or topic catalogues for this purpose. Queries consist of keywords – words that describe documents for searching. It is not always easy to find appropriate keywords that describe documents in the best way to perform a web search, but that is a skill which could be improved by practice. On the other hand, topic catalogues are rarely used because the user seldom knows the full topic hierarchy for a specific document, and also there are many ways in which some documents can be classified (e.g. a historical document can be classified using both a geographic criterion and a chronological criterion).

### **Web crawling**

Web crawlers (sometimes called *web spiders*) collect web documents for later searching. Crawling the web is a continuous process because new documents appear on the web daily. Except for that, the documents that have been updated/changed since they were last accessed by the web crawler should also be updated in search engine database.

Crawlers typically use sets of Uniform Resource Locators (URLs; Thompson, 2002.) which represent addresses of web sites and read their contents. Web sites usually have entry documents, like `index.html` or `default.htm`, so the crawling of a particular site could start from that document. Except for collecting document texts, it's important to collect all links to other documents. These links will lead the crawling engine to other documents to be crawled. But, the number of links can grow rapidly (e.g. 10 links on the first level could lead to 10 links on each document on the second level, etc.) so the number of links at lower levels could be too big for effective processing. Therefore crawling should be bounded by some parameters like number of levels or by using single domain name (e.g. crawling only links under the same domain as entry URL). Also, a crawler engine should be capable of collecting documents (and information from these documents) in different file formats like html, doc, pdf, ppt etc. Another request

to crawlers is their ability to support different communication protocols like HTTP and FTP.

### **Search engine indexing**

Crawled documents should be indexed for later searching. Of course, the size of indexes has to be significantly smaller than the size of crawled documents, because of the efficiency of searching. There are many different techniques used for indexing, but some of them are most common:

- Ignoring stop words and other extremely common words (Thompson, 2002.). Stop words and extremely common words are useless for searching documents (e.g. "the", "and", "it" etc.). Other extremely common words include different formatting words which depend on document type (e.g. html documents use html tags like <html>, <body>, <title>, etc.).
- Using document metatags. Html documents can contain some additional data except for their information content, like keywords, author's name, time when document was last updated, software used in document production etc.

Html documents often contain keywords which are invisible when viewing through web browser, but visible for web crawlers and for indexing purposes, e.g.

```
<meta name="keywords" content="computing, programming, C++, C#, Pascal, Java">
```

means that this document should be indexed using keywords computing, programming, C++, Pascal, C#, and Java. But, despite metatags make indexing easier, there are some possibilities of their misuse, for instance by using too many keywords, very long keywords etc.

### **Improving search engine ranking**

Users often create web pages using tools which are focused mostly on the appearance of a particular web page. Creating and publishing a web page is often not enough for it to be found by other people. Special techniques are used for web pages to be found by search engines crawlers and favorably ranked in the results of search engine use.

There are several established techniques that can be used to increase web page ranking in the results of a search engine, which makes the web contents of a specific web page more approachable for potential viewers interested in its content (Coopee, 2000):

- submitting a web site to the search engine simply by filling out an online form (even though crawlers collect majority of all web contents, it's also the way to put information about our site to web search engines and catalogues, especially those which have no crawlers and support manual fill only;

- using metatags (there are more than 50 types of metatags available for use within HTML pages, but most commonly used by search engines for indexing and ranking purposes are the description and keywords tags.

The *description metatag* includes a short description of a site that a search engine would display in a list of search results:

```
<meta name="description" content="Programming and programming languages">
```

The above mentioned tag means that the text “*Programming and programming languages*” will appear next to the URL in search engine results.

When using *keywords tag*, a combination of unique and common keywords should be used (Coopee, 2000). The idea is to anticipate which keywords the potential visitors would use to find the site.

Calculated *linking to the web page* from other web sites (Coopee, 2000) is another technique for influencing page rank. Some search engines (Google) use web page popularity in ranking, which is determined by how often a particular web page is linked to by other pages in the index.

Optimizing by *doorway pages* is also popular by web site promoters. Doorway pages are pages optimized for one search engine and 1-3 keywords (Shapiro and Lehoczky, 2007). Other names for doorway pages are “gateway”, “bridge”, “entry”, “jump” or “supplemental” pages. These pages are separated from the rest of the site (stand on their own), containing only the link to other contents. This method was popular in the past, and often misused for spamming so it should be used carefully to avoid search engines *antispam* protection.

Some frequent mistakes that search engines don’t like are *frames*, *dynamic content* and *Flash intros* which make web page indexing difficult for the search engine, while *spam* may cause penalization of the web page by the search engine in relation to rank in search results (Shapiro, 2007). *Spamindexing* is a term that denotes the methods used to increase the web page rank by manipulating the web page ranking algorithms of search engines. The following techniques are usually considered as spam:

- meta refresh tags,
- invisible text and overuse of tiny text,
- irrelevant keywords,
- excessive repetition of keywords,
- overuse of mirror sites,
- submitting too many pages in one day,
- identical or nearly identical pages,
- submitting to an inappropriate category (for directories),
- link farms (a group of web sites that hyperlink to all group members).

An overview of various factors which influence web page relevancy and ranking on different search engines is presented in Table 1.

Table 1. Factors that influence ranking of web pages on Google, MSN and Yahoo! (Shapiro, 2007)

SEARCH ENGINE AND WHAT IT LIKES	Elements that block indexing	POTENTIAL POSITIVE INFLUENCE ON RANKING				
		Content and location of keywords	HTML title	Meta tags	Keyword frequency	Link popularity
<b>Google</b> Link popularity, keywords near each other, keywords in URLs and link text, themes	Not mentioned, spammers	Keywords in text or links; keywords should be close to each other	Not mentioned, but seems to be a factor	No	Not mentioned	Very important, especially from relevant pages
<b>MSN</b> Theme present throughout the site, site popularity	Spammers, frames (<noframes> tag needed)	Not mentioned	Important, should contain keywords	Important, both description and keywords	Important, 4-12 times	Important
<b>Yahoo</b> Concise/accurate descriptions and keywords, appropriate categorization	Spammers	Worthy of indexing as determined by editors; in appropriate category	No, but the title filled in plays a role, it should be concise	No, but the description and keywords play a role	No	Very important

From the data presented in Table 1 it can be concluded that, after the content and location of keywords, the HTML title of a web page and link popularity are the second most important elements in relation to relevancy ranking by listed search engines. HTML titles can be easily defined by the author of the web page. On the other hand, link popularity is harder to achieve because it means that a specific web site is linked from other web sites.

Link popularity can also be used as a method for manipulating with ranking results. This type of activity is often called a *link bomb* or *Google bomb* when the attempt is to bias the results of the Google search engine by linking to a target web page from words that are not in consistence with the content of this web page (thus a wrong impression of the target can be formed). Simplified, a Google bomb is created if a large number of sites contain an anchor text that leads to a specific target site. Some specific requirements for Google search engine are presented in Table 2.

Table 2. Specific requirements for Google search engine (Shapiro, 2007)

Favored document length	Keywords	Location of keywords	HTML title	Meta tags	What's spam?	Other info
Wide range, from 50-600 words	Weight and proximity matter most	<h> tags, bold text	Keywords here, up to 90 characters	No	Use of link farms, cloaking, excessive repetition	Link popularity is the most important factor

To improve the ranking of a particular web site, the following should be considered (Shapiro, 2007; Heng, 2006, 2007):

- *HTML title* should contain properly placed keywords;
- meta tags should be defined, especially description and keywords tags, despite the fact that some search engines, like Google do not consider them relevant;
- *keyword density* is important because if a particular keyword has a higher density on the web page, then the likelihood of that page obtaining a better search engine ranking increases;
- *keywords in the URL or file name* of the web page;
- *Alt tags* help search engines to recognize what's in the image displayed within the web page;
- *link popularity* is increased if our pages are linked to from other sites; for example, Google uses the *Page Ranking Algorithm* which ranks pages according to the number and quality of links leading to that page;
- *themes* of web pages which are consistent;
- *design* of a web page which makes it easy to navigate;
- no spam or frames since search engines penalize such content;
- dynamically generated pages could be indexed by some search engines, like Google, but it is still recommended to put some links to them from other pages inside the web site.

It must be noted that an experimental analysis proved that using metadata elements increases rank order of web pages (this depends on the quality of metadata description) and that the use of metadata is comparable to using the Dublin core which is much less popular for the authors of web pages (see: Khaled, 2006).

### **Locating Croatian cultural heritage information with search engines**

Different search engines use various mechanisms for web crawlers and different algorithms and principles for indexing and ranking of links to located documents on the web. Therefore, in August 2007 a test was performed of search engines Google, MSN, Yahoo! and Pogodak! to determine which search engine reports more locations on the web in relation to selected names of Croatian cultural heritage artifacts. Also, the first 100 web sites that were linked on the list of search results were examined regarding the usefulness and quality of information on the specific item of Croatian cultural heritage and the more useful links were counted. The results of this test are presented in Table 3.

As can be observed from the data presented in Table 3, the largest number of web sites in relation to most of the items of Croatian cultural heritage that were searched for was reported in the results of Google search engine use. MSN and Yahoo! performed rather poorly in comparison to Google. Finally, on the average, Pogodak! performed slightly better than MSN and Yahoo! search engines.

In addition, for every search term in Table 3 an inspection was performed of the web sites on the list of first 100 links in the search results of all four previously mentioned search engines and, in most cases, a larger number of *quality sites* (with sufficient textual information on a specific cultural heritage artifact) was found in the results of Google search. It can be concluded that Google search engine is probably the best choice for locating information about Croatian cultural heritage artifacts on the web.

Specialized institutional web sites and portals in Croatia provide general information on cultural heritage and links to related web sites. The users of the Internet who are interested in Croatian cultural heritage could locate those portals with a search engine and use of appropriate key words (e.g. "cultural heritage Croatia").

In the previous analysis (see Table 3) it appeared that Google was the best search engine for locating information on most of the specific cultural heritage artifacts of Croatia. It must be emphasized that Google is the search engine with most search queries (for instance, in August 2007 in the U.S.A. it had about 54% of share of searches, while Yahoo! had 20% and MSN 13% of share of searches; see Nielsen/NetRatings, 2007). Therefore, a Google search was performed for web sites with general information related to Croatian cultural heritage and the results are displayed in Table 4. Also, the web page source was analyzed for the selected web pages and the use of meta tags was assessed together with the count of key words in textual information on the web page.

The data in Table 4 lead to the conclusion that the three web pages with the highest rank had key words "cultural heritage Croatia" either in the title of the web page (marked with "<title>" in page source) or in the meta tags and document description. However, of those only the *Culturenet Croatia* website, the virtual portal to Croatian culture project funded by the Ministry of Culture of the Republic of Croatia (for more information about this project see: Uzelac, 2005), could be regarded as a web site of a formal cultural or heritage institution, while the other three web sites ranked among the first two pages of Google search results were of the nonprofit organization *Croatian Camping Union (Kamping udruženje Hrvatske)*, tourism agency *adriatica.net*, tourism portal *Croatica.net*, and *Croatian National Tourist Board*. Within the first 100 results there were also the web page "Croatia UNESCO Heritage" of the *Find Croatia* tourism portal and several pages with partly related content of the *Ministry of Foreign Affairs and European Integration of Croatia* (not shown in Table 4). The homepage of the more important website of the *Museum Documentation Center* with links to many web sites of Croatian museums was not in the first 100 search results.



Table 3. The number of links found in the results of search engines Google, MSN, Yahoo! and Pogodak! with the use of search terms related to Croatian cultural heritage (the results with and without the use of diacritical symbols in Croatian language are displayed; performed in September 2007)

Searched term*	Google	MSN	Yahoo!	Pogodak!
Bašćanska ploča	27,000	634	626	4,230
Bascanska ploca	31,000	430	334	3,070
Plominski natpis	347	25	157	84
Višeslavova krstionica	336	28	209	150
Viseslavova krstionica	231	2	14	150
Crkva sv. Križa Nin	13,000	160	231	504
Crkva sv. Kriza Nin	13,000	54	73	505
Vinodolski zakonik	46,000	576	411	650
Povelja kneza Trpimira	376	15	64	65
Vučedolska golubica	798	290	708	386
Vucedolska golubica	797	14	178	45
Knežev dvor	66,600	3,431	3,180	5,898
Knezev dvor	45,000	580	468	5,898
Katedrala Sv. Jakova	118,000	1,046	2,610	1,217
Eufrazijeva bazilika	29,800	1,514	18,000	1,916

\* Bašćanska ploča (*The Basca Tablet*), Plominski natpis (*The Plomin Tablet*), Višeslavova krstionica (*The Baptistry of Višeslav*), Crkva sv. Križa (*The Church of the Holy Cross*), Vinodolski zakonik (*The Law Code of Vinodol*), Povelja kneza Trpimira (*Trpimir's Charter*), Vučedolska golubica (*Vučedol dove*), Knežev dvor (*Duke's Palace*), Katedrala Sv. Jakova (*St Jacobs Cathedral*), Eufrazijeva bazilika (*Euphrasian Basilica*).

In search for web sites with general information on the cultural heritage of Croatia it was found that the Ministry of Culture of the Republic of Croatia does not have an English version of its web site (however, it does fund the *Culturenet Croatia* website). Interestingly, the web page "Croatia UNESCO Heritage" of the *Find Croatia* tourism portal had a rather low rank (59) despite the use of keywords in the title of web page and in meta tags. This could have been caused by too many (21) keywords in the meta tag of that web page that may have been perceived as *spamindexing* and penalized by the Google search engine. It must also be mentioned that the occurrence of keywords in the text of the analyzed web pages did not appear to have much impact on page ranking. Finally, one of the most interesting results of the analysis presented in Table 4 was that not more than 0-1 links were found pointing to the listed web pages on Croatian cultural heritage from outside of their web domain. This means that most of the

web sites in Table 4 can be located only by Internet users who already know of them, or by those who perform a search on one of the search engines.

Table 4. Analysis of the web page source regarding use of key words "cultural heritage Croatia" in meta tags and in the text of the web page

Document title for web page description and web address	Rank after Google search <sup>1</sup>	All keywords in title of web page ("<title>")	Document description in meta tags <sup>2</sup>	Number of links to web address <sup>3</sup>	Occurrence of text <sup>4</sup> "cultur"/ "heritage"/ "Croatia"
"culturenet.hr – Panorama – Art – Croatian cultural heritage – monuments and sites" <a href="http://www.culturenet.hr/v1/english/panorama.asp?id=67">http://www.culturenet.hr/v1/english/panorama.asp?id=67</a>	1	Yes	None	0	5/2/3
"Kamping udruženje Hrvatske – Cultural heritage of Croatia" <a href="http://www.camping.hr/druga.aspx?stranica=914&amp;pid=69">http://www.camping.hr/druga.aspx?stranica=914&amp;pid=69</a>	3	Yes	None	0	4/4/5
"adriatica.net &gt; Tourist guide – Cultural heritage" <a href="http://www.adriatica.net/common/destinations/features-culture_en.htm">http://www.adriatica.net/common/destinations/features-culture_en.htm</a>	4	Partly	Very good	0	6/9/12
"Croatia Information – Croatica.net Traveller" <a href="http://www.croatia.net/en/portal/hrvatska/">http://www.croatia.net/en/portal/hrvatska/</a>	9	No	No	1	9/3/22
"Croatia – Croatian National Tourist Board" <a href="http://www.croatia.hr/English/Home/Naslovna.aspx">http://www.croatia.hr/English/Home/Naslovna.aspx</a>	17	No	No	0	3/3/4
"Croatia UNESCO Heritage" <a href="http://www.find-croatia.com/unesco-heritage.html">http://www.find-croatia.com/unesco-heritage.html</a>	59	Partly	Good	1	8/5/26
"MDC   Home page" <a href="http://www.mdc.hr/index_en.aspx">http://www.mdc.hr/index_en.aspx</a>	>100 8*, 55*	No	None	3	3/3/23
"Ministarstvo kulture Republike Hrvatske – KULTURNA BAŠTINA" <a href="http://www.min-kulture.hr/default.aspx?id=6">http://www.min-kulture.hr/default.aspx?id=6</a>	?	No web page in English	No web page in English	1	0/0/0

<sup>1</sup> Keywords used: cultural heritage Croatia; finding of another web page under the same web domain is marked with an asterisk (\*)

<sup>2</sup> *None* – no keywords and no description tags are used within page meta tags.

*Poor* – inadequate number and poor selection of used keywords within page meta tags; no description tags.

*Good* – adequate number and proper selection of used keywords, but no description tags; or, adequate number but poor selection of used keywords, with description tags.

*Very good* – proper usage of meta tags within page; adequate number of properly selected keywords and description tags.

<sup>3</sup> A Google search was performed of web address enclosed in quotation marks.

<sup>4</sup> The occurrences were not counted if found in HTML, meta tags, web page title and URLs.

### **Search for information on cultural heritage by informatics students**

It can be assumed that most internet users who needed to find information on cultural heritage would first consult the Internet and then perhaps go to a library or a museum. To verify this assumption a group of 48 students information systems at the Faculty of organization and informatics, University of Zagreb, Croatia, were asked to write in their own words what would be the first, second and third type of activity that they would perform if they needed to find additional information on a Croatian cultural heritage artifact of their choice (they had to select one item from a list of artifacts similar to the one presented in Table 3). The response of 92% of the students was that they would first look on the Internet and 8% responded that they would first go to the library. The list of their second and third choices was rather long and included going to the library, use of Internet/Wikipedia, consulting literature/encyclopedia, asking friends, asking a professor, going to a museum, asking one's parents, going to a tourist agency, etc.

There was much diversity in what students in our brief survey would do as their second and third activity, but the most common sequence of activities reported by 71% of students was to look on the Internet first, and then go to the library. This could reflect the typical behavior on an average Internet user to whom the search of the Internet is an activity that can be performed with much less effort than a visit to the library.

### **Conclusion**

This paper examined various means for influencing the web page rank in search engine results. Since web search often starts with the use of a search engine (White and Iivonen, 1999) and since most Internet users do not proceed in their search for information past the third page of search results (iProspect, 2006) it is important to ensure the highest possible web page rank on the list of results of search engine use for web pages and web sites of culturally valuable information. As can be concluded from the data presented in Table 4, most of the analyzed web pages with content related to general information on cultural heritage of Croatia were not optimized for search engines. Among those sites that were analyzed, the *Culturenet* web site provided a significant amount of information on cultural heritage, but a closer inspection of this web site revealed that it needed improvement regarding usability and visual design to facilitate access to information. The *Museum Documentation Center* web site with its numerous online museums (and texts in Croatian and English language) is the premier resource on the web regarding Croatian cultural heritage and one of the most successful regional projects in its domain. However, this web site was not optimized for search engines at the time of our analysis and difficult to locate with the use of Google search engine.

Since most Internet users would first use this medium to find information about cultural heritage, and because their likely activity would be to use a search en-

gine, it is important not only to present cultural heritage information of the web, but also to optimize the cultural heritage web sites for search engines.

## References

- Coopee, Todd. *How to climb the search engine rankings*. // InfoWorld Test Center. InfoWorld Media Group, San Francisco, California, U.S.A. 2000. <http://www.infoworld.com/articles/mt/xml/00/06/12/000612mtsearch.html> (22 August 2007)
- Fallows, Deborah. *Search Engine Users*. Washington, DC, USA: Pew Internet and American Life Project, 2005. [http://www.pewinternet.org/pdfs/PIP\\_Searchengine\\_users.pdf](http://www.pewinternet.org/pdfs/PIP_Searchengine_users.pdf) (22 August 2007)
- Heng, Christopher. *How to Improve Your Search Engine Ranking on Google*. thesitewizard.com. 2007. <http://www.thesitewizard.com/archive/google.shtml> (22 August 2007)
- Heng, Christopher. *Keyword Density in Search Engine Ranking*. thesitewizard.com. 2006. <http://www.thesitewizard.com/archive/keyworddensity.shtml>, thesitewizard.com. 2007. (22 August 2007)
- InternetWorldStats. *Internet usage statistics – The Big Picture: World Internet Users and Population Stats* [Online]. // Internet World Stats, Miniwatts Marketing Group. 2007. <http://www.internetworldstats.com/stats.htm> (22 August 2007)
- iProspect. *Search Engine User Behavior Study*. Watertown, MA, USA: iProspect.com, inc. 2006. [http://www.iprospect.com/premiumPDFs/WhitePaper\\_2006\\_SearchEngineUserBehavior.pdf](http://www.iprospect.com/premiumPDFs/WhitePaper_2006_SearchEngineUserBehavior.pdf) (22 August 2007)
- Kando, Noriko; Adachi, Jun. *Cultural heritage online: Information access across heterogeneous cultural heritage in Japan*. // *International Symposium on Digital Libraries and Knowledge Communities in Networked Information Society, DLKC'04*. Tsukuba, Ibaraki, Japan. 2004. <http://www.kc.tsukuba.ac.jp/dlkc/e-proceedings/papers/dlkc04pp136.pdf> (22 August 2007)
- Khaled, A.F. Mohamed. *The impact of metadata in web resources discovering*. // *Online Information Review*, 30 (2006), 2, 155-167
- Mac Aoidh, Eoin; Bertolotto, Michela; Wilson, C. David. *Capturing spatial interactions to personalise cultural heritage access*. // *International Workshop on Personalization Enhanced Access to Cultural Heritage*. Corfu, Greece. 2007. <http://www.cs.vu.nl/~laroyo/PATCH/Proceedings/7-Aoidh-final.pdf> (22 August 2007)
- Moen, E. William. *Accessing distributed cultural heritage*. // *Communications of the ACM*, vol. 41, no. 4, 1998, 45-48
- Nielsen/NetRatings. *Nielsen/NetRatings Announces August U.S. Search Share Rankings* [Online]. Nielsen/NetRatings, The Nielsen Company, Haarlem, The Netherlands. 2007. [http://www.netratings.com/pr/pr\\_070919.pdf](http://www.netratings.com/pr/pr_070919.pdf) (19 September 2007)
- Ossenbruggen, Jacco; Amin, Alia; Hardman, Lynda; Hildebrand, Michiel; Assem, Mark; Ome-layenko, Borys; Schreiber, Guus; Tordai, Anna; de Boer, Victor; Wielinga, Bob; Wielemaker, Jan; de Niet, Marco; Taekema, Jos.; Orsouw, Marie-France; Teesing, Annemiek. *Searching and Annotating Virtual Heritage Collections with Semantic-Web Techniques*. // *Proceedings of the Museums and the Web 2007: The International Conference for Culture and Heritage* [Online]. San Francisco, California, U.S.A. 2007. <http://www.archimuse.com/mw2007/papers/ossenbruggen/ossenbruggen.html> (22 August 2007)
- Rainie, Lee; Shermak, Jeremy. *Search engine use shoots up in the past year and edges towards email as the primary internet application*. [Data Memo]. Washington, DC, USA: Pew Internet and American Life Project, 2005. [http://www.pewinternet.org/pdfs/PIP\\_SearchData\\_1105.pdf](http://www.pewinternet.org/pdfs/PIP_SearchData_1105.pdf) (22 August 2007)
- Rutledge, Lloyd; Aroyo, Lora; Brussee, Rogier; Stash, Natalia; Wang, Yiwen; Cramer, Henriette; Evers, Vanessa; Ramlal, Satyan. *Recommendation with semantics for cultural heritage*. // *International Workshop on Personalization Enhanced Access to Cultural Heritage*, Corfu, Greece. 2007. <http://www.cs.vu.nl/~laroyo/PATCH/Proceedings/PATCH.pdf> (22 August 2007)
- Shapiro, Yelena. *How do I Improve Search Engine Rankings?* SearchEngines.com. 2007. <http://www.searchengines.com/searchEnginesRankings.html> (22 August 2007)

- Shapiro, Yelena; Lehoczky, Etelka. Factors that Influence Search Engine Rankings. SearchEngines.com. 2007. [http://www.searchengines.com/ranking\\_factors.html](http://www.searchengines.com/ranking_factors.html) (22 August 2007)
- Smeulders, W.M. Arnold; Hardman, Lynda; Schreiber, Guus; Geusebroek, Jan-Mark. An integrated multimedia approach to cultural heritage e-documents. // *4th Intl. Workshop on Multimedia Information Retrieval, in conjunction with ACM Multimedia 2002*. Juan-les-Pins, France. 2002. <http://homepages.cwi.nl/~media/publications/Smeulders02a.pdf> (22 August 2007)
- Thompson, Dave. *The influence of metatags on web-based search retrieval, ranking and relevancy*. Dalhousie University, Halifax, Nova Scotia, Canada. 2002. <http://www.ecomm.dal.ca/emec/research/emec-final-april-22.pdf> (22 August 2007)
- Uzelac, Aleksandra. Cultural Networks and Cultural Portals – is there a difference? // *Colloquium “Interactive Culture: Culture and Online Information”*. Nantes, France. 2005. <http://www.culturenet.hr/v1/english/about.asp> (22 August 2007)
- White, Marilyn Domas; Iivonen, Mirja. Factors Influencing Web Search Strategies. // *ASIS Annual Conference: Proceedings of the 62nd Annual Meeting*. Washington, DC, U.S.A. 1999. [http://www.wam.umd.edu/~whitemd/ASIS\\_pap.pdf](http://www.wam.umd.edu/~whitemd/ASIS_pap.pdf) (22 August 2007)