JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY

JASIST VOLUME 59 NUMBER 11 2008

CONTENTS

Papers published online in Wiley InterScience, 8 May 2008 through 9 July 2008

Coden AISJB6 ISSN 1532-2882

WILEY-BLACKWELL

Journal of the American Society for Information Science and Technology

JASIST

The JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY, (Print ISSN: 1532-2882; Online: 1532–2890), is published monthly, except twice in January and February, on behalf of the American Society for Information Science and Technology by Wiley Subscription Services, Inc., a Wiley Company, 111 River St., Hoboken, NJ 07030-5774. The positions expressed by the contributors do not necessarily reflect the position or the official policy of the Society.

Subscription Price (Volume 59, 2008): Print only: $1,999.00 in US, $2,139.00 in Canada and Mexico, $2,258.00 outside North America. Electronic only (includes online ASIST Digital Library): $2,197.00 worldwide. A combination price of $2,354.00 in US, $2,494.00 in Canada and Mexico, $2,613.00 outside North America, includes the subscription in both electronic (with online ASIST Digital Library) and print formats. All subscriptions outside US will be sent by air. Payment must be made in US dollars drawn on a US bank. Claims for undelivered copies will be accepted only after the following issue has been received. Please enclose a copy of the mailing label. Missing copies will be supplied when losses have been sustained in transit and where reserve stock permits. Please allow four weeks for processing a change of address. Subscription inquiries should be addressed to John Wiley & Sons, Inc., Attn: Journals Admin. Dept. UK, 111 River Street, Hoboken, NJ 07030-5774; Tel: (201) 748-6645, E-mail: subinfo@wiley.com.

Postmaster: Send all address changes to JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY, Journal Customer Services, John Wiley & Sons, Inc., 350 Main St., Malden, MA 02148-5020.

Advertising Sales: Inquiries concerning advertising should be addressed to: Advertising Sales Manager, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030-5774; Tel: (201) 748-8832. Advertising Sales, European Contact: Jackie Sibley, John Wiley & Sons, Ltd., The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, England; Tel: 44 1243 770 351, Fax: 44 1243 770 432, E-mail: adsales@wiley.co.uk.

Reprints: Reprint sales and inquiries should be directed to the Customer Service Dept., Gale Krouser, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030-5774; Tel: (201) 748-8789, Fax: (201) 748-6326, E-mail: gkrouser@wiley.com.

Other Correspondence: Address all other correspondence to: Publisher, JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY, Wiley-Blackwell, 350 Main Street, Malden, MA 02148-5020.

The contents of this journal are indexed in the following: *Chemical Abstracts, CompuMath, Computer Literature Index, Current Contents/Engineering Technology and Applied Sciences, Current Contents/Social and Behavioral Sciences, Current Index to Journals in Education (CIJE), Engineering Index, Information Science Abstracts, INIST for the PASCAL database, ISI Compumath, Library and Information Science Abstracts (LISA), Library Literature, Research Alert (ISI),* and *Social Sciences Citation Index.*

Editorial Production, Wiley Periodicals, Inc.: Catherine Dunkerley

This paper meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper).

Journal of the American Society for Information Science and Technology

JASIST

VOLUME 59 NUMBER 11 SEPTEMBER 2008

CONTENTS

(*Continued on next page*)

Volume 59, Number 11 was mailed the week of August 18, 2008

In This Issue

In this issue, Thelwall compares the results of 1,587 single word searches from Google, Yahoo! and Live Search, evaluating the results in terms of the quantitative data likely to be used in Webometric investigations. The results indicate that the three search engines are reasonably consistent in their hit count estimates and the number of different top-level domains in the URLs . However, they have significant inconsistencies for the number of different URLs, sites and domains returned. Moreover, these inconsistencies sometimes manifest themselves in quite irregular ways, and there are inconsistencies in Yahoo! and Live Search that make their hit count estimates seem less reliable than those of Google. Based on the findings, the author makes recommendations as to which search engine(s) would be most appropriate for various Webometric investigations.

Meho and Rogers examine the differences between Scopus and Web of Science (WoS) in the citation counting, citation ranking, and h-index of 22 human-computer interaction (HCI) researchers from EQUATOR (a large British interdiscipinary Research Collaboration project). The results indicate that Scopus provides significantly more coverage of HCI literature than WoS, primarily due to coverage of relevant ACM and IEEE peer-reviewed conference proceedings. No significant differences exist between the two databases in terms of journal citations. Although broader coverage of the literature does not significantly alter the relative citation ranking of individual researchers, Scopus helps distinguish between the researchers in a more nuanced manner than WoS in both citation counting and h-index. Scopus also generates significantly different maps of citation networks of individual scholars. The authors also present a comparison of h-index scores based on Google Scholar with those based on a union of results from Scopus and WoS. The authors conclude that Scopus can be used as a sole data source for citation-based research and evaluation in HCI, especially if citations in conference proceedings are sought, and that researchers should calculate h scores manually, rather than relying on system calculations.

Espadas, Calero and Piattini present a method that Web administrators can use to quantitatively evaluate the visibility of a Web site. The method consists of four activities, each involving several tasks. Most of the tasks are accompanied by a set of defined measures that can help to determine if, and where, the Web site is failing from the positioning point of view. The efficacy of the method is demonstrated through the presentation of two case studies.

 DOI: 10.1002/asi.20933

Buente and Robbin describe a research study that attempts to answer the questions: who is embedding the Internet into their everyday lives, and what are the activities they pursue to facilitate everyday life? The study investigates trends in Internet information activities between 2000 and 2004 using repeated cross-sectional data from the Pew Internet and American Life surveys. Findings demonstrate the differential returns for Internet use, particularly in key demographic categories. The study also contributes to emerging research on the digital divide, namely emphasis on the study of use rather than access to technology. The authors conclude by discussing ways in which the identification of trends in key Internet use dimensions enables policymakers to target populations who underutilize the potential of networked technologies.

Vishwanath and Chen used the Galileo system of multidimensional scaling to systematically map the extensional associations with nine personal communication technologies across three cultures – the United States, Germany and Singapore. Samples of undergraduate communication students were polled using a Web survey. Data were collected between March and May 2006, resulting in 448 responses from the United States, 347 responses from Germany, and 73 responses from Singapore. The research focused on seven personal technologies: cell phone, personal email, blog URL, IM screen name, home page URL, home phone number, and office or school phone number. The findings indicate that the technologies closest to self were similar across the three cultures, suggesting a universality of associations with certain technologies. In contrast, the technologies farther from the self were significantly different across cultures. Moreover, the magnitude of associations with each technology differed based on the extensional association or distance from the self. Also, the antecedents to these associations differed significantly across cultures, suggesting a stronger influence of cultural norms on personal technology choice.

Lisée, Lariviére and Archambault examine the scientific impact and aging of conference proceedings compared to those of scientific literature in general. The relative importance of proceedings is diminishing over time and currently represents only 1.7% of references made in the natural sciences and engineering, and 2.5% in the social sciences and humanities. Although the scientific impact of proceedings is losing ground to other types of scientific literature in nearly all fields, it has grown from 8% of the references in engineering papers in the early 1980s to its current 10%. Proceedings play a particularly important role in computer sciences, where they account for close to 20% of the references. The authors also show that, not unexpectedly, proceedings age faster than cited scientific literature in general. The authors

conclude that proceedings have a relatively limited scientific impact, represent on average only about 2% of total citations, and become obsolete faster than scientific literature in general.

Shin compares national information infrastructure (NII) development in the United States and Korea, focusing on the role of government. Data were gathered from government officials, industry representatives and academic researchers through interviews and surveys. The author concludes that important similarities and differences can be demonstrated through comparisons of sociotechnical dimensions: government function, histories, visions, policy design, implementation plans, and realities and prospects.

Kayed et alia present a new approach to the building of ontological relationships in Web documents. It has been shown that extracting concepts is easier than building relationships among them. For a defined set of concepts, many existing algorithms produce all possible relationships for that set. This makes the process of refining the relationships almost impossible. The authors present a new algorithm which enables a domain-knowledge expert to refine the relationships linking a set of concepts. The proposed algorithm groups the concepts together according to the number of relationships with other concepts and provides formalization. The use of this new algorithm reduced the number of relationships among 273 concepts (which were obtained using text-mining techniques in the domain of e-commerce laws) from approximately 32,000 relationships to 29 relationships with two or three levels.

Leydesdorff and Schank examine the dynamic analysis of structural change in the organization of the sciences using an integration of multivariate and time-series analysis. Recent developments in multidimensional scaling may be used to distinguish the stress originating in each time-slice from the stress originating from the sequencing of time-slices, and thus locally optimize the trade-offs between these two sources of variance in the animation. Visualization programs such as Pajek and Visone show, not only the position of the nodes, but their relational attributes such as betweenness centrality. Betweenness centrality in a vector space can be considered as an indicator of interdisciplinarity. Using this indicator, the authors animated and assessed the dynamics of the citation-impact environments of the journals *Cognitive Science, Social Networks,* and *Nanotechnology* in terms of the interdisciplinarity among the disciplines.

Huang and Chang review the literature on the nature of research output in the social sciences and humanities (SS&H) in terms of characteristics of the research publications. Based on this review, the authors conclude that SS&H research output has characteristics distinct from the natural sciences (NS), and thus output-based evaluations for SS&H research need different methodologies than those for NS disciplines. An ideal evaluation of SS&H research should address issues including the diversity of publication types, the emphasis on book publishing, an emphasis on referencing books rather than journal articles, the local and regional concerns in SS&H research, and the tendency to cite older literature.

Wu, Chuang and Chen investigate the relations between users' motivation for using (i.e., trying out and continuing to use) a search engine and the engine's functional features. Based on Herzberg's two-factor theory, the features can be characterized as hygiene factors and motivation factors. Hygiene factors support the query process and provide a basic task context for information seeking that allows users to access relevant information. Motivation factors help users navigate (i.e., browse) and comprehend the retrieved information, related to the task-context aspect of information seeking. Given the consistent findings that hygiene factors induce work motivation for a shorter period of time, it was hypothesized that hygiene factors are more effective in attracting users, while motivation factors are more effective in retaining users. A survey was posted to the web in October 2005, resulting in758 usable responses. The authors conclude that the survey results provide substantial support for the hypothesis.

Bornmann and Daniel examine whether the peer review system fulfills its declared objective of selecting the "best" scientific work. The authors investigated the journal peer review process at *Angewandte Chemie: International Edition* (AC-IE), one of the prime chemistry journals worldwide. A citation analysis was conducted for Communications that were accepted by the journal ($n = 878$) and Communications that were rejected by the journal and published elsewhere ($n = 959$). The results of negative binomial-regression models show that, holding all other variables constant, acceptance by AC-IE increases the expected number of citations by up to 50%. A comparison of average citation counts (with 95% confidence intervals) of Communications rejected by AC-IE but published elsewhere indicates that mean citation counts below baseline values were significantly less frequent for accepted Communications than rejected Communications. The authors conclude that peer review at AC-IE does select the "best" scientific work with the highest impact for chemical research.

Rousseau and Ye, in a brief communication, propose a time-dependent h-type indicator. This indicator depends on the size of the h-core, the number of citations received, and recent change in the value of the h-index. As such, it tries to combine, in a dynamic way, older information about the source (e.g., a scientist or research institute that is evaluated) with recent information.

Zitt and Small, in a brief communication, introduce a new approach to the field normalization of the classical journal impact factor. This approach, called the *audience factor*, takes into consideration the citing propensity of journals for a given cited journal (specifically, the mean number of references of each citing journal) and fractionally weights the citations from those citing journals. Hence, the audience factor is a variant of a fractional citation-counting scheme, but computed on the citing journal rather than the citing article or disciplinary level. In contrast to other cited-side normalization strategies, this approach is focused on the behavior of the citing entities. A comparison with standard journal impact factors from Thomson Reuters shows a more diverse

DOI: 10.1002/asi

representation of fields within various quintiles of impact and significant movement in rankings for a number of individual journals, but nevertheless a high overall correlation with standard impact factors.

Cronin and Meho, in a brief communication, use the Author Affiliation Index (AAI) to determine whether faculty at the top ten North American library and information science (LIS) programs have a disproportionate presence in the premier journals of the field. The AAI of a journal (or set of journals) is defined as the percentage of authors publishing in that journal (or set) who are affiliated with a predetermined group of top-rated universities, or university departments. The results indicate that LIS may be both too small and too interdisciplinary for the AAI to provide reliable results.

Carol L. Barry
School of Library & Information Science
Louisiana State University
267 Coates Hall
Baton Rouge, LA 70803
E-mail: lsbary@lsu.edu

DOI: 10.1002/asi

Web Site Visibility Evaluation

Javier Espadas
Information Systems Department, Fund. Col. Thyssen-Bornemisza, Madrid, Spain.
E-mail: jespadas@museothyssen.org

Coral Calero and Mario Piattini
Alarcos Research Group, Information Technologies and Information Systems Department, University of Castilla-La Mancha, Paseo de la Universidad, 4, 13071 Ciudad Real, Spain. E-mail: {coral.calero, mario.piattini}@uclm.es

In recent years, the Internet has experienced a boom as an information source. The use of search engines is the most common way of finding this information. This means that less visible contents (for search engines) are increasingly difficult or even almost impossible to find. Thus, Web users are forced to accept alternative services or contents only because they are visible and offered to users by search engines. If a company's Web site is not visible, that company is losing clients. Therefore, it is fundamental to assure that one's Web site will be indexed and, consequently, visible to as many Web users as possible. To quantitatively evaluate the visibility of a Web site, this article introduces a method that Web administrators may use. The method consists of four activities and several tasks. Most of the tasks are accompanied by a set of defined measures that can help the Web administrator determine where the Web design is failing (from the positioning point of view). Some tools that can be used for the determination of the measure values also are referenced in the description of the method. The method is furthermore accompanied by examples to help in understanding how to apply it.

Introduction

In the last few years, the Internet has evolved from being a tool used by professionals with a technological profile to being widely used by all types of people. It is used not only from the point of view of information queries but also from that of information publication, thus becoming one of the main available sources of both direct and indirect information.

About 20% of the world's population uses the Web, and a large majority uses Web search engines to find information. Web search engines currently index tens of billions of Web pages and serve hundreds of millions of Web searches per day (Henzinger, 2007). This has made it possible to take the services and contents of companies or institutions to any corner of the world, achieving a huge potential volume of customers/users. In fact, business models have evolved from being product oriented to being customer oriented. These models also incorporate different phases, from customer attraction to customer loyalty (Cutler & Sterne, 2000).

Therefore, our objective is to ensure that a site's services and/or contents are visible to as many potential customers as possible so that customer royalty is achieved. Taking into account the fact that only 5% of a Web site's visitors actually consume/utilize the service or content offered (Nielsen, 2004), increasing the number of visits is the only way in which to increase the number of times that a company's services are consumed. Therefore, the number of visitors is a key factor in the success of a business or institution on the Web.

Our Web site must compete against thousands of other Web sites with the same or similar content, information, and even design and quality. If we wish to "win the battle," then we need to have an added value: We need to be found first. As noted by Wu and Davison (2006), the "search engine is the door to the Web today." (p. 1)

In this article, we present a method through which one can analyze the visibility of a Web site and detect any weak points that can be improved. The method defines a plan of action for the improvement of Web visibility and, therefore, the assurance of greater visitor numbers. It is accompanied by a set of measures which have been defined, from our own experience, to evaluate the results of the tasks. They were driven by the real needs of the company of one of the article's authors (J.E.) and are therefore based on actual requirements.

The method has been successfully applied to several Web sites, of which we highlight the Thyssen Museum, the workplace of the first author (http://www.museothyssen.org). The method has been implemented in a public tool called *Know*

Received April 12, 2007; revised February 8, 2008; accepted February 9, 2008

 DOI: 10.1002/asi.20865

Your Visibility (KYV; available at http://kyv.webportalquality.com), which calculates the visibility of a Web site and offers a report with corrective actions, if necessary. We also have applied the corrective actions given by KYV to the visibility improvement of the tool itself.

The method for improving Web site visibility is discussed in the next few sections. Then the KYV tool is presented, followed by the application of the method to two different Web sites. The final section includes our conclusions and notes on future work.

Visibility Analysis Method

The visibility analysis method attempts to synthesize visibility quality by using diverse metrics which focus on three aspects: (a) visitors that "visit" a site, (b) search engines as tools which allow Internet users to look for information, and (c) other sites that make recommendations as links. Once we have the information related to these aspects, most of the necessary information for improving the visibility of our Web site will be recovered.

We selected these three factors because, at present, the Web search engines use the analysis of the hyperlink structure of the Web as a basis for obtaining the search results. This new approach (as opposed to the previous one which was based solely upon text-only ranking algorithms) has improved the quality of the search results. The hyperlink structure analysis uses factors such us the frequency of the query terms on the Web page, geographic origin of the queries, and so on (Henzinger, 2007).

Therefore, the three aspects arise from a four-phase method (see Figure 1): Web site visibility, competitor identification, competitor visibility, and the definition of a plan of action through which to improve the visibility of the Web.

As can be seen in Figure 1, the first three phases have certain defined activities. For each activity, some tasks have been outlined. To evaluate each of the tasks, some measures have been drawn up. Note that some of the measures must be obtained by using internal information from the Web site by using the log files, the HTML code, and so on; others can be calculated by using public information.

The former internal measures can be applied only by the Web owner (or developer) who has access to the complete information of the Web site. The latter have been used as indicators of the visibility of a Web site and to create rankings. In fact, these measures are those which are used by the KYV tool. Although it may seem that the general measures are not sufficient to create a ranking, they actually are because they try to calculate how connected a site is, and hence its visibility.

For most of the measures used by our method, we give indications as to how and via which (generally available) tools they can be calculated. In this way, all the phases can be applied, and the method can thus be used. All the measures have been validated by using and refining them through their application in the improvement of different Web sites, particularly after some years of experience at the Thyssen Museum's Web site (museothyssen.org). We have used these measures to attempt to cover all the important factors related to the visibility of a Web site.

Another aspect taken into account when defining and mainly when applying the measures was the type of source used to obtain the necessary data with which to evaluate the visibility of a Web site. Two different types of sources can be used: our own engine or the information offered by search engines such as Google, Yahoo, MSN, and so on.

Multiple Internet analyses have used their own search engine. This kind of study needs a great technological effort and, in general, can be carried out only at research centers.

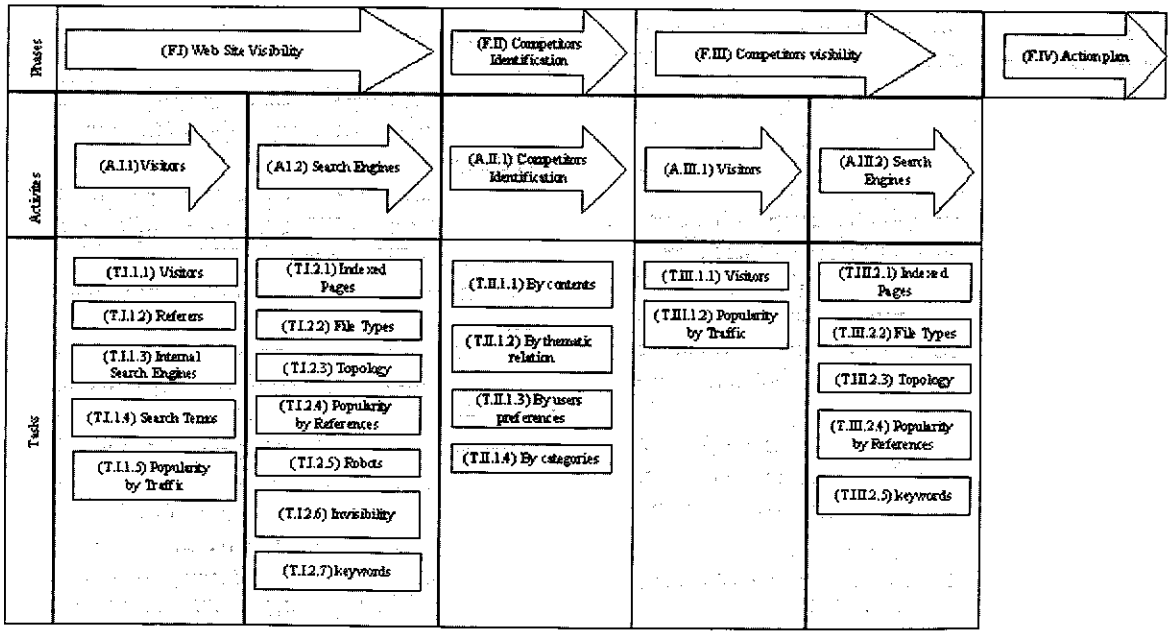

FIG. 1. Method for evaluating the visibility of a Web site.

DOI: 10.1002/asi

These results are usually more general, but they also need more technological effort, more investment, and so on. On the other hand, we can find several proposals and studies (Aguillo, Prieto, Ortega, & Granadino, 2005; Baeza-Yates & Poblete, 2003; Smith, 1999) based on the information offered by search engines.

As the intention of our proposal is to be as general as possible and to be able to be used by anyone, we decided to use the second approach: to use the information given by search engines. The advantage of our method is the necessity of a smaller technological effort and less investment in equipment and computer development. Our method uses cybermetric indicators (i.e., the defined measures) (Almind & Ingwersen, 1997; Dhyani, Ng, & Bhowmick, 2002; EICSTES, 2007; WISER, 2007), which allow the synthesis of a clear image of a Web site's visibility.

In summary, we can characterize our method as one which:

- is centered on the elaboration of quantitative cybermetric indices, which are obtained automatically and are driven by real needs and are based on actual requirements;
- allows the comparison of the visibility of one Web site with other Web sites; and
- is easy to use and whose application does not necessitate the development of a complex project.

Each of the phases of the method will be explained in depth in the following sections.

Phase 1: Analysis of the Web Site Positioning

The purpose of Phase 1 is the individual analysis of the Web site's position, without taking its competitors into account. This analysis is made up of two activities: analysis of visitors and analysis of search engines.

Activity 1.1: Analysis of Visitors

With regard to the visitors of a Web site, it is necessary to quantify not only the diffusion in a given period but also its evolution and tendency. This quantification will indicate, among other things, how many people visit a Web site and its most attractive contents.

The performance of an analysis which offers this information is very simple. There are several tools on the market—many of them free—which offer good analyses. Some of these tools read the Web log files and generate a report in a format such as Word, Excel, and so on, with data in tabular and graphic format that can be used to analyze the results.

This activity consists of five tasks: visitors, referrers, internal search engines, search terms, and popularity by traffic. Next, we will show the objectives of and the measures defined for each task.

Task 1.1.1: Visitors. Regarding visitors, it is important to know the number of visitors and the duration of visits as well as their geographical origin. Although we are conscious that the data concerning geographical origin are not always reliable (e.g., visits from Latin American countries can be considered as visits from the United States), we also think that it is necessary to take possible errors into account.

TABLE 1. Measures for Task 1.1.1.

Name	Description
No. of visits	No. of user sessions with a 30-s timeout
Visit duration	Average visit duration
% Visits by country	Distribution of visits by countries

The measures defined for this task can be found in Table 1. The software LogAnalyzer from WebTrends can be used to calculate the duration and the number of visits.

It is possible to use a specially created application to obtain the distribution of visits by countries. For the IP address georeferentation, we have used the Geolp Country Database (http://www.maxmind.com/app/country) of the MaxMind enterprise.

We do not believe that it is necessary to take into consideration other data such as pages, most accessed directory, most downloaded files, and so on. This kind of information is more related to the site quality in terms of content usability or quality and thus is not considered when evaluating visibility quality.

Task 1.1.2: Referrers. Regarding where and how the users access the Web site, it is important to know the page from which they have accessed our Web site. These origin pages are known as *referrers* and indicate the Web sites that link the visitors.

The identification of the main referrers of a Web site is very important, at which point some strategies must be taken into consideration (e.g., establishment of the contents, establishment of the interchange recommendations through the use of links). Obviously, sites which generate more traffic should be considered a priority in such strategies.

For the analysis of referrers, we propose taking as a basis the information which comes from the log files generated by the Web server, excluding those registers with an internal referrer or those without any referrer. Table 2 shows the measures defined for this task.

For example, in Figure 2 we present the main referrers of museothyssen.org from January to June 2005. We can highlight the importance of the traffic directly received from

TABLE 2. Measures for Task 1.1.2.

Name	Description
Referrer visits	No. of user sessions initiated from a site that refers the user to the studied site
%Referrer visits	%Site's visits initiated from an external site
Search engine visits	No. of visits that are initiated from a search engine query
%Search engine visits	%Referrer visits associated with a search engine
%Search engine visits	%Referrer visits associated with *X* search engine (Google, Yahoo, MSN, etc.)
%Site visits	%Referrer visits associated with *X* search engine

DOI: 10.1002/asi

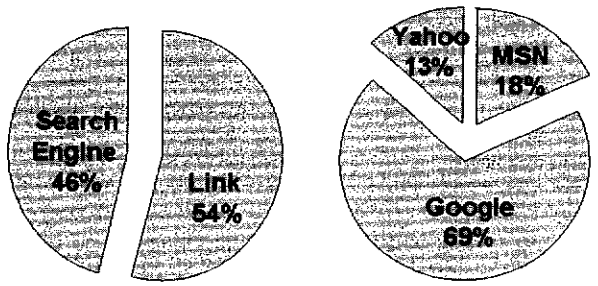

FIG. 2. Museothyssen.org referrers.

other pages (46%) as well as the importance of Google as compared to other search engines.

Task 1.1.3: Internal search engines. In the design process of a Web site, it is necessary to perform an analysis that allows us to identify words or concepts that are closely linked to the site topics.

These concepts might be, for example, the brand name or certain attributes which are associated with it (e.g., geographical location), cultural icons (e.g., painters' names in the case of museums), and so on. This work will give us a list of concepts (or even sentences) that will be used during the method phase.

Internal search engines that carry out indexing and search only within one particular site offer the possibility of attaining information that is not accessible from external search engines, such as query formulas that robots are unable to navigate. However, the analysis of the queries performed by users allows us to improve a site's usability.

From the visibility point of view, the analysis of queries allows us to identify words which are very commonly used by users and that do not obtain good results in external search engines.

Taking all this into consideration, Table 3 shows the measure defined for this task.

Task 1.1.4: Search terms. The purpose of this task is to identify the search terms being successfully used by both users and the search engines. As a result, we will be able to identify unexpected search sentences together with others that, although they are logical, do not allow the user to find a Web site in the search engines. Table 4 shows the measures defined for this task.

This kind of study allows us to extract conclusions such as:

- the topics of the Web site that users are interested in,
- which sentences or words have generated more visits, and
- general sentences that do not eventually produce as many references as expected.

TABLE 3. Measures for Task 1.1.3.

Name	Description
%Nonvisible words	%Words/sentences successfully used by users in the internal search engine and that do not generate results in the external search engines

TABLE 4. Measures for Task 1.1.4.

Name	Description
Sentence visits	Visits generated from a query with a word/sentence
Search engine sentence visits	Visits generated from a query with a word/sentence in a search engine

TABLE 5. Measure for Task 1.1.5.

Name	Description
Competitor's name	Alexa competitor's name

Task 1.1.5: Popularity due to traffic. The objective of this task is twofold: to evaluate the popularity of a Web site and the popularity of its competitors with regard to Internet users.

The best way in which to analyze this popularity is through log file analysis. These data already were analyzed in Task 1.1.1 when the traffic of a site was studied. However, the objective of this particular task is to obtain traffic data that allow us to make comparisons with other competitors or reference sites.

The identification of competitors who are of interest to similar internauts may be carried out by using the references offered by Alexa (http://www.alexa.com). Alexa offers for a Web site a list with other Web sites that also are of interest for its visitors. The identification of these recommendations is based on the recorded navigation habits of the users who have installed Alexa's navigation bar. As Alexa takes information only of its users, the information obtained must not be considered in absolute values, although they can be used for evaluating the progression of the traffic of a site. Using Alexa, the measures shown in Table 5 were obtained. For each of its Web sites, this search engine offers a list of other Web sites which also may be of interest to its visitors.

Activity 1.2: Search Engine Analysis

Regarding search engines, it is fundamental to know (a) which are the most important search engines and (b) the search sentences used by users to find our Web site. By analyzing this information, we can identify the absence of logical key sentences that are apparently not used. We also can identify those sentences which are present, but which never result, in the appearence of our Web site.

Search engines use agents called "robots" to extract information from Web sites. These agents have a programmed collecting engine which limits the time dedicated to a Web site or to the maximum exploration depth depending upon, among other variables, the importance or reputation of each site.

This activity has seven tasks that are explained in the following subsections.

Task 1.2.1: Indexed pages. It is important to discover the amount of pages in our site that the main search engines

DOI: 10.1002/asi

(or at least Google and Yahoo) have indexed. Discovering our indexed pages is critical, as all those pages which are unknown (indexed) to the search engines will not appear as a result of any query. Table 6 shows the measures used to carry out this task.

The analysis of the indexed pages of a site as compared to those of its competitors will offer the Web administrator a quantitative vision of the relative importance of the contents and services of the Web site. The analysis of indexed pages includes the study of pages in whose content we may find certain keywords (e.g., a painter's name or an artistic movement name, in the case of museums). This type of analysis (see Table 7) indicates that the most relevant artistic movement in Google is "Abstract" followed by "Pop Art" while in MSN and Yahoo it is "Impressionism" (again, in the case of museums).

Task 1.2.2: Types of files. The files published by a Web site are of different types, and the identification of these types can be made by using the file format. The content of a file in a given Web site can be supposed with a high degree of exactitude. The measures used for this task are presented in Table 8.

TABLE 6. Measures for Task 1.2.1.

Name	Description
Search engine indexed pages	No. of indexed pages by a search engine (Google, MSN, Yahoo, etc.)
Topic search engine pages	No. of pages indexed by a search engine related to a topic

TABLE 7. Top-10 painters and movements (26 analyzed domains).

Top-10 Topics in Indexed Pages		
Google	MSN	Yahoo
Van Gogh	Impressionism	Picasso
Abstract	Degas	Van Gogh
Picasso	Monet	Pop art
Matisse	Gauguin	Degas
Degas	Matisse	Impressionism
Monet	Picasso	Edgar Degas
Pop art	Van Gogh	Monet
Gauguin	Lautrec	Matisse
Lautrec	Manet	Manet
Manet	Matisse	Vincent Van Gogh

TABLE 8. Measures for Task 1.2.2.

Name	Description
N search engine files	No. of indexed files of a type in a search engine
Topic search engine files	No. of indexed files of a type in a search engine related to *Z* topic
%Pages file	%files of a type indexed by a search engine with relation to the total of indexed pages

Pdf files in university Web sites contain, to a high degree, the results of research work whereas in the field of museums, large images normally represent works of art. For example, we can consider that for museum Web sites, the graphical component is a key factor. It is thus fundamental to analyze the volume of images as well as the rate of images/indexed pages.

Task 1.2.3: Topology. Discovering and modeling Internet topology is an important research topic in the area of information recovery (Broder et al., 2000). Presently, the study of links as a measure of the importance of a page is one of the main lines of work of search engines, which orientate recovery algorithms towards models that value connectivity (Henzinger, 2007) and the importance of links in the field of a query. A typical algorithm of this type is Hyperlink-Induced Topic Search (HITS), which is characterized by its classification of pages into hubs and authorities (Chirita, Olmedilla, & Nejdl, 2003).

Search engines have algorithms which, when evaluating the importance of a document, consider not only the number of links a document has but also the text associated with it. Search engines such as Google have developed a page ranking called Page Rank (Page, Brin, Motwani, & Winograd, 1999) which considers, among many other variables, the number of links of a page. Some other proposals (Brinkmeier, 2006; Nielsen, 2004) also have tried to improve the Page Rank.

The fact that a Web site is well known and recommended by links increases its good reputation for the search engines, which take this factor into consideration when assigning their resources to information collection and indexing (Rafiei & Mendelzon, 2000, p. 1).

The purpose of this topologic analysis task is to study the relationships established between one Web site and others through the links, which are considered as quotations. Table 9 shows the measures for this task.

In this study, the calculation of the Web Impact Factor (WIF) is clearly important, and was proposed by Ingwersen (1998) as an evolution of the publishing impact factor developed by Garfield (1994). The WIF measures the existing relationship between the number of quotations of a site in the form of links and the total number of pages on the site. Dominions that have the greatest WIF are those which are most referenced or quoted and are, therefore, the best known.

The percentage of deep links that do not link to the main page is an indicator of the attractiveness of the contents of a site. Links to the main page are considered to be links of an

TABLE 9. Measures for Task 1.2.3.

Name	Description
Web Impact Factor (WIF)	No. of average links received by a page; data obtained from *X* search engine
%Deep links	%links that do not directly indicate the home page; data obtained from *X* search engine
Reciprocity links	Links from *Y* to *Z* minus links from *Z* to *Y*; data obtained from *X* search engine

DOI: 10.1002/asi

institutional type, which are more highly related to the brand, company, or to the content in general.

Another analyzed topologic variable is the relationship that exists between two sites that quote each other. This analysis values the reciprocity of links between sites. This analysis reflects how the relationship, in terms of links between, for example, nga.gov and getty.edu, is not compensated since nga.gov quotes getty.edu three times, and the latter quotes the former 105 times.

The analysis of quotations in museum sites clearly shows that it is a sector in which it is not common to recommend "competitors." In other sectors, such as that of universities, quotations as links are a common practice.

Task 1.2.4: Popularity according to references. As we cannot obtain real data from our Web visitors to quantify the real popularity of a site, we have taken the number of links received by a Web site as a reference. Thus, a Web site is considered to be more important as more pages establish links with it.

On the other hand, the number of links is obviously important in making a direct impact upon the number of visitors; the more Web sites that recommend our Web site, the more visitors the Web site will have. This information was used as a basis through which to define the measures shown in Table 10.

Task 1.2.5: Robots. From the moment in which Web documents are published on a server to the moment in which they appear as a result of a query through a search engine, these documents go through a series of processes (Brin & Page, 1998). The collecting process is carried out automatically by computer applications known as "robots" or "spiders." All pages that are not included in this collecting process cannot be queried by search engines.

If we are to analyze robot activity, we need to know which search engines are most interested in a Web site, and we also need to identify the activity of robots which are not associated with search engines (which may submit a Web server to excessive workloads). For this task, we propose the measures in Table 11.

Task 1.2.6: Invisibility. Invisible pages are those which, although they have been published on the Internet, cannot be found through search engines. A content "becomes invisible" during the collecting process that robots (Arroyo, 2004) carry out through automatic navigation on the Web due to denied accesses, interpretation problems, difficulties in finding these pages, and temporary dependencies (Crowther, Bechhofer, & Horan, 2004). The invisibility analysis aims to quantify the volume of registers which are available through search engines and the volume of those which have become invisible. Table 12 shows the measure defined for this task.

Invisibility audit requires the verification of the main search engines which have as many indexed pages as the registers that exist in the table(s) to which they have access. For example, the query page fichaobraampliada.asp of museothyssen.org gives access to 690 works of art.

Task 1.2.7: Keywords. Certain works highlight the fact that 85% of Internet users use the search engines to query services. From the latest surveys, we also can conclude that the more advanced an Internet user is, the more intensive the use he or she makes of search engines and the more complex his or her searches.

Obviously, one of the main objective of a Web site (especially if it belongs to a cultural institution) is to obtain the maximum diffusion of its works and activities. Therefore, appearing in a good position in the result of search engines for a given number of search sentences is a key factor in this diffusion. The measure for this task is shown in Table 13.

TABLE 10. Measures for Task 1.2.4.

Name	Description
Search engine quotations	Links from *Y* to the analyzed site; data obtained from *X* search engine
Term search engine quotations	Links from *Y* page with *Z* content to the analyzed site; data obtained from *X* search engine

TABLE 11. Measures for Task 1.2.5.

Name	Description
Robot No. of visits	No. of visits performed by the robot of *X* search engine
%Robot visits	%Visits of the robot of *X* search engine over the total of robots visits

TABLE 12. Measure for Task 1.2.6.

Name	Description
X search engine invisibility	%Static or dynamic pages not indexed by *X* search engine; the dynamic pages will be scored according to the number of registers to which they give access

TABLE 13. Measure for Task 1.2.7.

Name	Description
Search engine sentence	Position of the analyzed site when searching by using *Y* sentence in *X* search engine

Phase 2: Identification of Competitors

As was indicated in the Introduction, the Internet is a space of great competence, and thus, the identification of the sites with which we compete is one of the main tasks from the point of view of positioning. Two Web sites are competitors if their contents and/or services belong to the same knowledge or business area and if they have a similar target public.

DOI: 10.1002/asi

Activity 2.1: Identification of Competitors

The sole activity of this task is to discover and characterize (from the point of view of positioning) a Web site's competitors. The activity consists of four tasks.

Task 2.1.1: Identification of competitors related by content. Competitors through content are Web sites that appear as a result of a search engine query. Obviously, this kind of competitor depends upon the query; for this reason, it is advisable to select queries that clearly indicate the activity and contents of a site.

To select the Web sites which fulfill these conditions, it is possible to use the WebQL tool (www.ql2.com) of QL2 Software which allows us to query the Web by using a query language similar to SQL, which considers the whole Web as a large database. This tool considers the Internet as a huge, unstructured data warehouse, makes the query through a robot, and has the ability of going through links according to different criteria.

Task 2.1.2. Identification of competitors related by topic. Competitors which are related by topic are Web sites whose contents are semantically similar. In other words, they have the same or similar keywords.

Google offers the keyword related:URL, which lists the Web sites that are similar to a given page. Googleguide.com offers a complete list of Google search commands.

Task 2.1.3: Identification of competitors related by preferences. As in task 1.1.5, the identification of competitors that are interested in similar Internet users can be carried out by using Alexa.

Task 2.1.4: Identification of competitors related by the directory category. The final method proposed for the identification of competitors is based on directories such as DMOZ. Those Web sites which appear in the same category are considered to be competitors. DMOZ, also known as Open Directory Project (ODP; www.dmoz.org), is currently the most important directory. It has almost 60.000 volunteer editors who classify the Web sites proposed by any Internet user. Although other directories such as those of Yahoo exist, DMOZ is considered to be the best directory of those pages organized according to categories.

Phase 3: Analysis of Competitor Positioning

Once the competitors have been identified, we propose carrying out some of the Phase 1 tasks that are related to the competitors' sites. This will allow us to make comparisons of the best and worst positioned sites, as well as the practices used by such sites.

When performing a comparative competitor analysis with the aim of being able to carry out a temporary tracking, it is advisable to develop a ranking that takes the information obtained from each site into consideration. The creation of rankings has the objective of showing a reference-site's visibility, as opposed to that of its competitors, in a synthetic manner. These rankings can be carried out by using any of the indicators obtained in the previous phases.

The variables to be analyzed when creating a ranking depend on the characteristics of the Web site upon which are we carrying out the study. Possible examples of rankings might be:

- Generic Ranking: The sites analyzed are ordered in accordance with their visibility within the Internet, without taking into account the topics that they analyze. This ranking considers the relative importance of each site for each indicator. This ranking should be constructed by normalizing the absolute values obtained in each indicator with regard to the maximum value of each indicator.
- Topic Ranking: The sites are ordered in accordance with their visibility restricted to the topic analyzed.

Phase 4: Evaluation of Results and Plan of Action

Once analysis of all the previous aspects has been carried out, and we have a general idea of the position of our Web site, it is advisable to elaborate a plan of action in which the reasonable objectives, whose fulfillment will be checked in each of the analyses performed, are defined. Objectives that could be included are, for example, to increase the number of users by 10%, to increase the duration of each visit to 10 min, or to assure that Google indexes all the new pages of our Web site.

Obviously, achieving certain objectives requires certain programming tasks and Web site design improvement. Others will require the elaboration of business policies such as the interchange of links with other institutions, the sending of news, taking advantage of other institutions' publicity or preparing some kind of marketing actions through which to attract new users (thus allowing them to become familiar with the Web site, its content, and its services), and so on. These objectives must be reasonable and coherent with the importance of the institution (and the extent of its business) and the contents offered by the Internet.

One unreasonable objective would be that of attempting to attain good positions for search phrases that are related to cultural tourism, due to the simple fact of incorporating "cultural tourism" as a keyword in our Web site. The only manner in which to attain this objective would probably be to incorporate a good Web channel dedicated to this theme.

The process of analysis and improvement should be continuous; many Web sites work to improve their position and manage to attain their objective of being in the top position in a Google search one week only to lose that position the following week.

The KYV Tool

The KYV tool implements part of the proposed method, calculates the visibility of a Web site (offering a report with

DOI: 10.1002/asi

corrective actions if necessary), and elaborates rankings. The tool is available at http:///kyv.webportalquality.com, and its main screen in shown in Figure 3.

The objective of the KYV tool is to give the user information about Web positioning of Web site domains. This process cannot take place in real time because it is necessary to ask various search engines several times for different values to calculate the visibility indicators and to prepare the rankings. The application thus stores and processes the Web site throughout the period of time that is necessary. As well as giving information about visibility, the application also offers some activities that could be applied in the quest to improve that Web site's visibility.

Finally, the tool stores statistical information about the Web sites analyzed to be able to develop visibility rankings by topics.

As has previously been explained, some of the measures included in the method upon which KYV is based (see Figure 1) must be worked out by using internal information, and others can be calculated by using public information.

These last measures are those used by KYV as indicators of the visibility of a Web site and to create rankings. The KYV application is expected to perform queries in the most important search engines to calculate these measures and to analyze the results, usually in text format, to obtain the requested values.

KYV will then carry out measurements for the following variables or parameters:

- Documents or Pages Indexed (SITE variable). Related to Task 1.2.1 of the method.
- Links to the domain (LINK-SITE variable). Related to Task 1.1.2 of the method.
- Alexa Popularity/Traffic Rank (POPULARITY variable). Related to Task 1.1.5 of the method.

Two further variables also are used in the visibility analysis, but these have not been included in the production of the ranking, although they are used when reporting the results to the user. These variables, along with a justification of why KYV does not consider them when drawing up the ranking, are described in the following lines:

- Rich Files Indexed (PDF SITE variable). Related to Task 1.2.2 of the method. As rich files are applicable only to certain domain categories (e.g., Universities) and are not usual in other areas (e.g., Banks or Museums), this variable is not used for the production of the ranking.

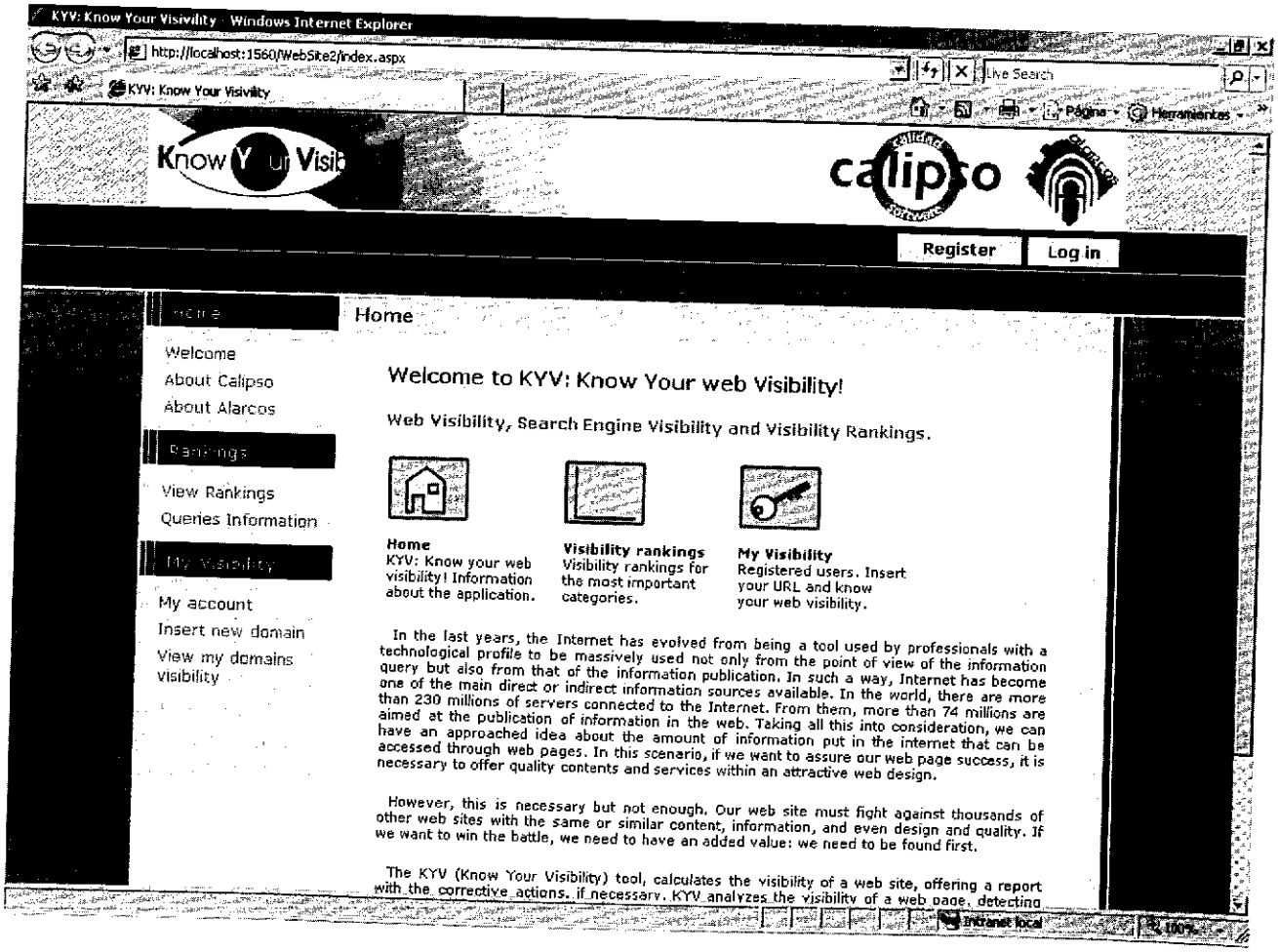

FIG. 3. KYV main screen.

DOI: 10.1002/asi

- Google Page Rank (PAGE RANK variable). Related to Task 1.2.3 of the method. Page Rank is not applicable to the creation of rankings because it usually assigns similar scores to topic-related domains. Moreover, the Page Rank is redundant with the link domain because they both use the same information as their basis (links).

In Figure 4, we show which variables are calculated in which search engines when working out visibility and constructing the rankings.

For the SITE, LINK-SITE, and PDF SITE variables, calculations are performed in the three most important search engines, Google, Yahoo, and MSN, because they concentrate 90% of user requests throughout the world. The other two variables are calculated in their respective search engines. KYV uses all the data recovered to calculate the mean for each variable and search engine (if applicable) for a given month.

As noted earlier, KYV considers only the SITE, LINK-SITE, and POPULARITY variables in the development of rankings. Each ranking (see Figure 5) will be produced by first grouping queries (Google site), then by grouping indicators (Site), and finally by giving a single evaluation to each domain and ranking.

A weighted formula that is invisible to users is applied to the values obtained. At the end of this process, an ordered classification is given in which the overall position for each domain and parameter is ascertained. Users must take into account that these rankings have a subjective part because the parameters are weighted. Therefore, different users with varied levels and types of experience can weigh the parameters in different ways. Those used by KYV have been determined by our own experience and, from our point of view, have obtained useful results in the Web sites to which they have been applied.

CONSULTA	GOOGLE	YAHOO	MSN	ALEXA
SITE	X	X	X	
LINK-SITE		X	X	
PDF SITE	X	X	X	
POPULARITY				X
PAGERANK	X			

FIG. 4. Set of variables calculated in the visibility analysis related to the search engines.

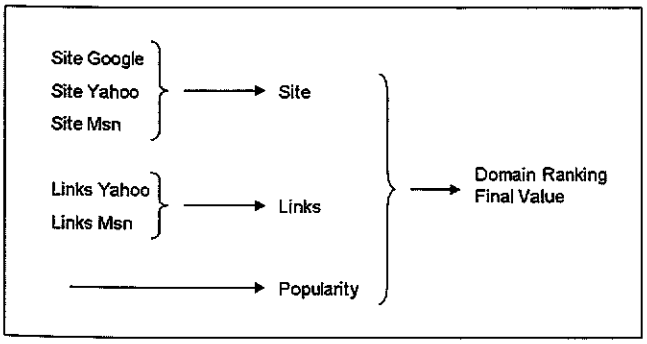

FIG. 5. Development of rankings.

The values of the variables are used to generate a report with recommendations for visibility improvement. There are two different situations: (a) The studied domain belongs to a given domain and is therefore part of a ranking, and (b) the domain is isolated (i.e., does not allow a ranking to take place).

In the first case, we consider that a variable has a good value when its value is in a good position within the ranking. If it is near the medium positions, then the value is considered to be sufficiently good. If not, then a recommendation of how to improve this aspect is sent to the user.

When the domain is not included in a ranking, then the comparison is carried out by using different aspects as a basis. For the SITE variable, the relationship between the variable value and the total number of pages on the site is used. This relationship tries to capture the percentage of the pages on the site that is visible to the search engines. If this relationship is not good enough (i.e., both values are not similar), then recommendations for the improvement of this variable are given to the user.

As the other two variables (LINK-SITE and TR) are computed by search engines, we compare the site values of these variables with the values that the variables took for all the sites that do not allow a ranking to take place. If the values are not similar, then recommendations for improvement are generated.

Case Studies

In this section, we present two case studies with which we illustrate the proposed method with figures. The first is the application of the complete method to museothyssen.org; the second is the application of the corrective actions given by the KYV tool for the visibility improvement of kyv.webportalwuality.com; that is, the site of the tool.

Application to museothyssen.org

In this section, we show data obtained through an analysis of museum Web site visibility with the following characteristics:

Analyzed words: gauguin, henckel, nolde, kirchner, matisse, chagall, modigliani, derain, manet, sisley, marc, lautrec, vangogh, cezanne, pissarro, renoir, monet, degas, picasso, cubismo, abstraction, futurismo, kandisnsky, dadaismo, mondrian, leger, delaunay, muler, braque, pechstein, postimpressionism, fauvism, symbolism, expressionism, impressionism, macke, kokoschka, pop art, juan gris
Search engines analyzed: Google, MSN, Yahoo, Exalead
Measurement period: December 2005–January 2006
Readings carried out: 39,208

DOI: 10.1002/asi

When possible, for some measures, we analyzed the following sites: artic.edu, cnacgp.fr, getty.edu, guggenheim.org, hermitagemuseum.org, lacma.org, louvre.fr, macba.es, mcasandiego.org, metmuseum.org, mfa.org, mfacmchicago.org, mfah.org, miamiartmuseum.org, moma.org, museeorsay.fr, museoprado.mcu.es, museoreinasofia.es, museothyssen.org, nationalgallery.org.uk, nga.gov, nga.gov.au, rijksmuseum.nl, tate.org.uk, uffizi.firenze.it, vangoghmuseum.nl

Phase 1: Analysis of Web Site Positioning

Task 1.1.1: Visitors. The analysis of visitors reflects how important the Thyssen-Bornemisza Museum's Web page is as a channel for the difussion of its activities, as it is the access point for over 900,000 people per year who have, for various reasons, shown an interest in its Web site (see Table 14).

Upon comparing the average number of daily sessions per year from 2001 (see Figure 6), we can observe a constant growth—with the exception of 2003, which coincided with a period in which the museum was under reconstruction and therefore did not have any tempory exhibitions, which led to a period of low offers of contents in its Web site.

The average length of a visit was of >10 min, which reflects a clear interest in the contents offered by the site (Table 15).

Task 1.1.2: Referrers. We analyzed the main referrers of museothyssen.org from January to June 2005 (see Figure 2). We can highlight the importance of the traffic directly received from other pages (46%) as well as the importance of Google as compared to other search engines.

TABLE 14. Values for Task 1.1.1.

Identification	Description
Sessions/day	2.480
Sessions/month	75.447
Sessions/year	905.365

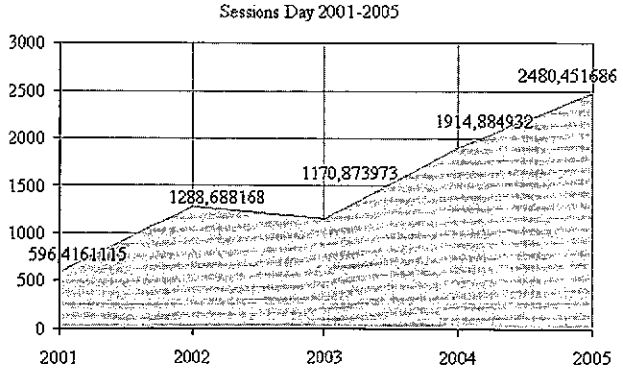

FIG. 6. Sessions/Day 2001–2005 museothyssen.org.

TABLE 15. Values for Task 1.1.2.

Identification	Description
Min/session	11 min 22 s

Task 1.1.3: Internal search engines. The analysis used to identify words or concepts that were closely linked to the museothyssen.org topics resulted in a list of words which was made up of the "brand name" (i.e., "Thyssen-Bornemisza") and its attributes as well as of the names of the painters, the paintings, and the movements associated with those paintings.

The analysis of these queries revealed that the users of museothyssen.org search for authors by name and surname, and yet the term most frequently used in searches is "Guernica," which refers to the painting by Picasso which is at present on display in the Museo Nacional Centro de Arte Reina Sofía.

Although this word is not highlighted in the pages of museothyssen.org, it *is* highlighted on the page of the Pablo Picasso biography.

Task 1.1.4: Search terms. This kind of study allows us to extract conclusions such as those listed next, which have been obtained from an analysis of the log files of the museothyssen.org Web site for the Year 2005:

- Museothyssen.org visitors are interested in "pintura impresionista" (Impressionist painting) since this search sentence generated 3,503 visits.
- Picasso is a reference painter since the sentence "cuadros picasso" (Picasso's paintings) generated 1,909 visits.
- Visitors do not find this Web site by using the sentence "museos de arte" (art museums) since this generic sentence only generated 579 visits.

Task 1.1.5: Popularity due to traffic. The study of competitors, which is based on the navigation habits registered by Alexa (Table 16) widens the spectrum of competitors to the Web sites of Latin American museums, and to other sites such as milímetros.com which are not related to museum activities.

Task 1.2.1: Indexed pages. In the analysis of indexed pages (see Figure 7), the museothyssen.org site has an important position, with more than 23,000 pages indexed by Google.

The analysis of indexed pages by Google, Yahoo, Exalead, and MSN for the 26 museum dominions analyzed shows that the number of pages managed by Google is 5.8 times that of the pages managed by Yahoo, 23 times that of Exalead, and 8 times that of MSN.

As indicated in section 1.2.1, the analysis of indexed pages includes the study of pages in whose content we can find a certain keyword (e.g., a painter's name or the name of an artistic movement in the case of the museums). This type of analysis (see Table 7) indicates that the most relevant artistic

DOI: 10.1002/asi

TABLE 16. Competitors by preference.

Competitors by preference (Alexa)
museoprado.mcu.es
www.mim.cl
www.maravillosarealidad.com
museoreinasofia.mcu.es
pymes.tsai.es/museoescultura
www.elmuseovirtual.com
www.chopo.unam.mx
www.agora.com.ar/museogp
www.mcu.es
www.milimetros.com

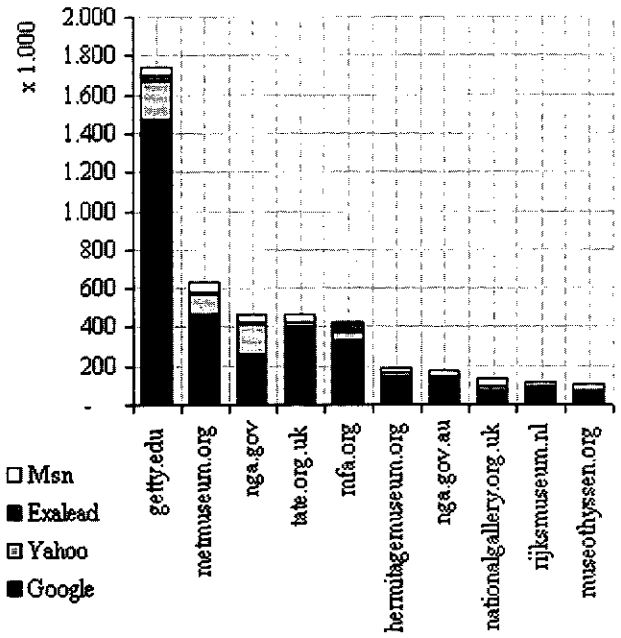

FIG. 7. Indexed pages, top-10 sites.

movement in Google is Abstract followed by Pop Art while in MSN and Yahoo it is Impressionism.

If we refer solely to the dominion of museothyssen.org, as we can see in Table 17, then the most relevant painters are not the same in each search engine, although Gaughin is in first place in MSN and Yahoo, and in second place in Google.

Task 1.2.2: Types of files. The analysis that was carried out indicates a clear leadership by the Metropolitan Museum, with more than 107,000 images. Other sites that have fewer images, such as Tate.org.uk (60,900), have a graphical richness ratio that is higher than 25%.

Task 1.2.3: Topology. The museothyssen.org WIF, according to Yahoo data, is 0.25. In other words, each page receives on average 0.25 external links, a similar value to that obtained by metmuseum.org. The museothyssen.org site has 16.3% links connected directly to its main page, as compared to dominions such as nga.gov which has only 7.8%.

TABLE 17. Top-10 painters and movements (museothyssen.org).

Top-10 Topics in muscothyssen.org indexed pages		
Google	Yahoo	MSN
Monet	Gauguin	Gauguin
Gauguin	Van gogh	Monet
Cezanne	Monet	Cezanne
Matisse	Picasso	Matisse
Van Gogh	Cezanne	Van Gogh
Pissarro	Paul Gauguin	Pissarro
Degas	Kirchner	Impressionism
Renoir	Pissarro	Paul Gauguin
Picasso	Matisse	Renoir
Kirchner	Degas	Degas

Task 1.2.4: Popularity according to references. The popularity study carried out by a Yahoo search engine reveals a clear leadership by the getty.edu and nga.gov sites, with more than 600,000 links each. On the contrary, museothyssen.org receives 25,100 links.

This analysis contemplates the study of the links received from pages related to a given topic. The analysis of received links (see Figure 8) from pages including words such as "impressionism" or "expressionism" clearly reflects that the Musée d'Orsay site is an authority, as compared to the Thyssen-Bornemisza Museum site which, in comparison, receives less links.

Task 1.2.5: Robots. When analyzing the collecting activity of the robots (see Figure 9) in museothyssen.org, one can see that in 2005, the Google robot only represents 26% of this activity while that of Yahoo has increased from 36% in 2004 to 67% in 2005. However, this increase in the collecting activity of Yahoo has not reflected a higher number of visits since from the total that have their origin in a search engine only 3.88% come from Yahoo as compared to 88.41% from Google.

Task 1.2.6: Invisibility. To audit the visibility of the cards with information about the works of museothyssen.org, a small program developed in VB.Net has been used which

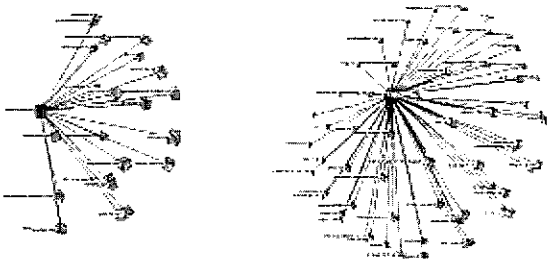

FIG. 8. Links of Impressionism, Expressionism topics to museothyssen.org and to Museemusee-orsay.fr.

DOI: 10.1002/asi

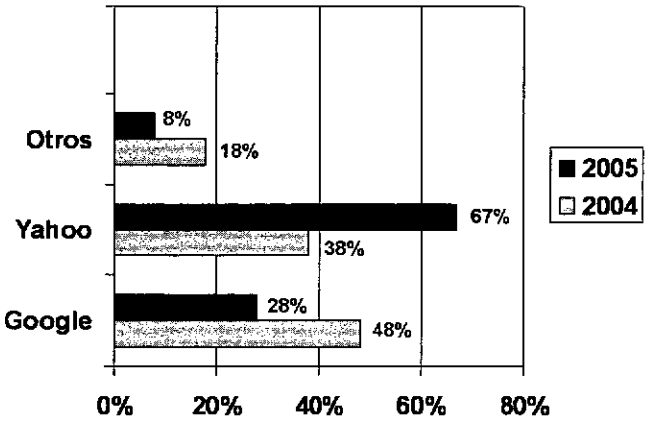

FIG. 9. Search engines spider activity.

verifies the indexing of the 690 pages of FichaObraAmpliada.asp with the corresponding code parameter. The final result indicates a degree of invisibility of 21%, mainly in the pages written in English.

Task 1.2.7: Keywords. The analysis of the results of the positioning in search engines must be performed over the same set of words with which the analysis of indexed documents was carried out (see Task 1.2.2). In the case of museum Web sites, we have considered the names of the painters and movements previously indicated.

The analysis of the proposed keywords reflects bad visibility for all the sites analyzed. Even in the case of the word "Van Gogh," the Museum of Amsterdam obtains the second position as opposed to the commercial dominion vangoghgallery.com, which obtains the first position.

Phase 2: Competitor Identification

Task 2.1.1: Competitor identification by content. To select museothyssen.org's competitors, the search terms "museo arte" and "art museum" were used (Figure 10). A restriction also was used, which consisted of selecting Web sites which contained only the words "museum" or "museo" in

```
select
        url, URL_HOST(url) as URL_SITE, source_title
from
        links
within
        'http://www.google.com/search?num=100&hl=es&rls=GGLD%2CGGLD%3A2
004-43%2CGGLD%3Aes&as_qdr=all&q=museo+arte&k='
where
        URL_HOST(url) matching 'museo' or URL_HOST(url) matching 'museum'
union
select
        url, URL_HOST(url) as URL_SITE, source_title
from
        links
within
        'http://www.google.com/search?q=museo+arte&num=100&hl=es&lr=&rls=GG
LD,GGLD:2004-43,GGLD:es&as_qdr=all&start=100&sa=N'
where
        URL_HOST(url) matching 'museo' or URL_HOST(url) matching 'museum'
```

FIG. 10. WebQL query.

TABLE 18. Competitors by content.

Competitors by content (Google.com)
nga.gov
metmuseum.org
british-museum.co.uk
museum.cornell.edu
hermitagemuseum.com
moma.org
artmuseum.ro
guggenhein.org
louvre.fr
nationallgallery.org.uk

their URL. The consultation was carried out with Google, and only the first 500 results were taken into account.

To select the Web sites which complied with these conditions, the WebQL tool was used. This allowed us to query the Web by making use of a language similar to that of SQL in that it considers the Internet as a large database. This type of tool behaves like a robot and has the capacity to navigate links according to various criteria.

The consultation resulted in 163 Web sites. Certain relevant sites, such as that of the National Gallery (nga.gov), were not included in the list because its URL did not contain the word "museum." The sites of the most relevant museums therefore have been added manually. To select the main competitors, the 10 which have the highest number of pages indexed by Google and Yahoo have been identified (Table 18).

Task 2.1.2. Identification of competitors related by topic. When using the TouchGraph tool (http://www.touchgraph.com/) to analyze the competitors related to museothyssen.org, one can observe a coherence of the obtained sites. Museothyssen.org is mainly related to museums or cultural institutions. Otherwise, we would have to analyze the content of the pages of museothyssen.org in detail to determine why Google does not consider the museum sites as being related.

Task 2.1.3. Identification of competitors related by user preferences. The study of competitors, based on the information given by Alexa.com together with Web sites of Latin American museums and other sites such as milímetros.com, gives the results shown in Table 16.

Task 2.1.4. Identification of competitors related by categories. The study of competitors related by categories (Table 19) was carried out in the Alexa and Dmoz directories. Given that the directories have multiple categories at their disposal, "Reference > Museums" was chosen as the starting point.

The results show that the museothyssen.org reference domain was not found in the categories analyzed while the domains of other museum Web sites are included in these categories.

DOI: 10.1002/asi

TABLE 19. Competitors per position in directories.

Directory: Alexa Category: Top > Reference > Museums > Arts and Entertainment > Art Museums > European	Directory:dmoz.org Category: Top: Reference: Museums: Arts and Entertainment: Art Museums: European: Spanish
www.tate.org.uk	www.caam.net
www.louvre.fr	www.bcn.fjmiro.es
www.hollandmuseums.nl	www.salvador-dali.org
www.nationalgallery.org.uk	www.paeria.es/mmorera
www.npg.org.uk	museoprado.mcu.es
www.designmuseum.org	www.guggenheim-bilbao.es
www.chateauversailles.fr	www.museoreinasofia.e
www.tate.org.uk	www.museupicasso.bcn.es
www.rijksmuseum.nl	www.museoromano.com
www.vangoghmuseum.nl	www.caam.net

Phase 3: Analysis of Competitors Positioning

The data obtained were used to create two rankings, whose objective was to present the relative importance of the sites analyzed.

According to the generic ranking (Table 20), which does not take in account the measures related to the topics analyzed, the Metropolitan Museum of New York Web site has the highest degree of visibility on the Internet. The museothyssen.org reference Web site is listed 22nd.

According to the topic ranking carried out (Table 21), the Metropolitan Museum of New York Web site has the highest degree of visibility on the Internet. The museothyssen.org reference Web site is listed 19th.

There are slight variations in the sites which occupy the top positions of this ranking and the general ranking. The results obtained show a clear domination of the Web sites of museums in the United States, which have a greater volume of contents on the Internet and are more widely known, and consequently receive a greater number of links.

This domination of the museums from the United States could be considered to be logical if we take into account the fact that it is a country in which technology plays a highly important role in the daily lives of its citizens. However, note that one of the greatest advantages of the Internet is that of the elimination of geographical barriers, and this dominant position therefore poses a threat to the Web sites of other museums because it is more likely that internauts will find the contents that they desire in sites with higher visibility.

Phase 4: Evaluation of Results and Action Plan

For our example of the museothyssen.org, a reasonable objective would be that of achieving a good position for search sentences related to cultural tourism. An easy way to do this would be to incorporate turismo cultural-cultural tourism as keywords in our Web site. However, it is probable that the only way in which to achieve this objective completely would be to incorporate a specific channel into the Web site which is dedicated to this, and to assure its indexation and quotation through links.

TABLE 20. General ranking.

Domain	Ranking site	Ranking link domain	Ranking link popularity	Ranking general
metmuseum.org	3	1	3	1
getty.edu	1	2	6	2
nga.gov	4	4	8	3
moma.org	12	3	5	4
tate.org.uk	7	5	7	5
artic.edu	2	8	10	6
louvre.fr	9	6	9	7
guggenheim.org	21	7	14	8
vam.ac.uk	5	10	11	9
thebritishmuseum.ac.uk	17	9	12	10
nationalgallery.org.uk	8	11	13	11
hermitagemuseum.org	11	12	15	12
lacma.org	19	13	17	13
rijksmuseum.nl	10	16	16	14
centrepompidou.fr	16	14	19	15
musee-orsay.fr	14	15	20	16
khm.at	6	20	26	17
museoprado.mcu.es	33	17	4	18
vangoghmuseum.nl	32	18	21	19
mv.vatican.va	22	22	2	20
gulbenkian.pt	18	19	25	21
museothyssen.org	**15**	**24**	**18**	**22**
museupicasso.bcn.es	46	33	1	23
guggenheim-bilbao.es	27	21	24	24
stedelijk.nl	34	23	33	25
kunsthaus.ch	24	25	29	26
museoreinasofia.es	13	43	23	27
pinakothek.de	20	28	27	28
modernamuseet.se	25	26	28	29
kunstmuseumbasel.ch	23	31	35	30
beyeler.com	26	27	31	31
staatsgalerie.de	35	32	30	32
ambrosiana.it	30	37	47	33
boijmans.rotterdam.nl	31	36	36	34
belvedere.at	38	30	34	35
leopoldmuseum.org	28	41	42	36
modernart.ie	29	40	43	37
gemeentemuseum.nl	36	34	39	38
nationalgallery.ie	45	29	38	39
musee-picasso.fr	41	35	37	40
mumok.at	40	39	41	41
kunstmuseumbern.ch	44	38	45	42
museopicassomalaga.org	39	42	32	43
gamtorino.it	42	44	40	44
gallerieaccademia.org	37	48	48	45
educathyssen.org	43	47	22	46
museumsinsel-berlin.de	47	46	46	47
museoegizio.org	48	45	44	48

Various visibility studies performed during several months with the method proposed for the site museothyssen.org have been used to introduce multiple improvements. We are continuously applying the method to obtain more information that allows us to improve the visibility of our Web site over time. Among the improvements since the beggining of the method application, we can highlight:

- The development of value-added contents which may be of interest to the target public.
- To include contents of the Museum Thyssen from the museothyssen.org domain.

DOI: 10.1002/asi

TABLE 21. Topic ranking.

Domain	Size ranking	Visibility ranking	Topic ranking
metmuseum.org	1	1	**1**
nga.gov	3	3	2
moma.org	2	8	3
tate.org.uk	4	4	4
artic.edu	8	2	5
getty.edu	6	6	6
guggenheim.org	5	9	7
vam.ac.uk	10	5	8
nationalgallery.org.uk	9	7	9
louvre.fr	7	18	10
thebritishmuseum.ac.uk	11	12	11
hermitagemuseum.org	13	10	12
lacma.org	12	19	13
musee-orsay.fr	14	16	14
vangoghmuseum.nl	15	20	15
museoprado.mcu.es	16	35	16
rijksmuseum.nl	17	27	17
centrepompidou.fr	18	14	18
museothyssen.org	**22**	**11**	**19**
khm.at	19	15	20
pinakothek.de	29	13	21
guggenheim-bilbao.es	21	22	22
stedelijk.nl	20	28	23
kunsthaus.ch	27	23	24
leopoldmuseum.org	41	17	25
modernamuseet.se	24	26	26
beyeler.com	28	24	27
modernart.ie	32	21	28
nationalgallery.ie	23	33	29
kunstmuseumbasel.ch	25	38	30
museupicasso.bcn.es	26	43	31
gemeentemuseum.nl	37	25	32
mv.vatican.va	30	34	33
gulbenkian.pt	34	31	34
belvedere.at	35	32	35
staatsgalerie.de	31	37	36
musee-picasso.fr	33	36	37
mumok.at	39	29	38
museopicassomalaga.org	40	30	39
boijmans.rotterdam.nl	36	40	40
kunstmuseumbern.ch	38	42	41
museoreinasofia.es	42	39	42
ambrosiana.it	44	44	43
serralves.pt	43	45	44
gamtorino.it	45	47	45
museumsinsel-berlin.de	46	41	46

- The strengthening of the external links, mainly the deep ones.
- Offering contents in pdf format.
- Verifying the presence of the most adequate categories on the different directories.
- The elimination of the flash on the home page.
- The use of nonstandard navigator scroll bars.
- The substitution of dynamic pages for static ones.

These improvements have given results such as:

- An increase in continuous traffic (see Figure 11)
- An increase in the number of pages indexed by search engines such as Google, which has gone from indexing 1,500 pages in January 2004 to 63,200 in January 2006.

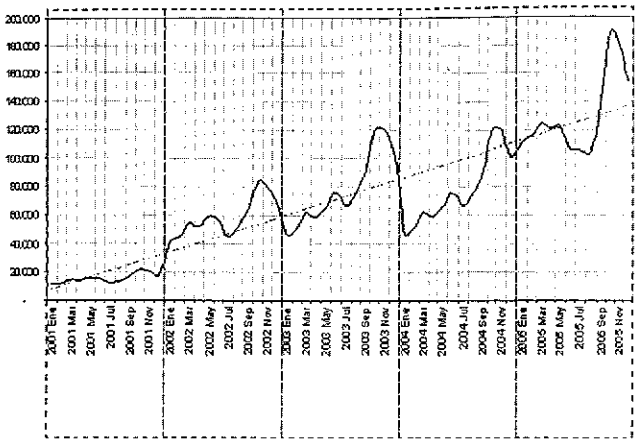

FIG. 11. Museothyssen.org visitors (January 2001–December 2005).

- Invisibility reduction, which has gone from 82% in January 2004 to 5% in January 2006.

Application to kyv.webportalquality.com

In this section, we show the results obtained from KYV for kyv.webportalquality.com (the tool's site). After an initial analysis, KYV suggested the following improvements for each parameter.

SITE variable. The best way to improve the number of indexed pages is through the elimination of barriers to the search engines, which can be achieved by taking the following advice:

- Avoid the generation of URLs by using scripting techniques. In particular, if the Web uses JavaScript for the generation of URLs, then the usage of additional URLs is recommended. This not only allows the spiders to find them but also ensures their correct visualization by devices that do not support JavaScript. If the site uses interactive multimedia contents (e.g., Macromedia Flash), then the use of alternative links also is recommended.
- Avoid the use of (a) cookies that modify the content, and (b) session variables that affect this content.
- Ensure that there are no broken links.
- Validate the correctness of the HTML code by, for example, ensuring that it follows the W3C recommendations.
- Use the robots.txt file correctly in such a way that the spider will know which directories can or cannot be traced. It is necessary to be sure that the spider will not ignore any portion of the site.

LINK-SITE variable. To improve this variable, it is necessary to use strategies through which to obtain new links to our site. As possible actions, we can enumerate:

- Registering the domain in the most important search engines (e.g., Google, Yahoo, MSN, etc.) and directories (e.g., DMOZ, etc.)
- Interchanging links with other Web administrators.
- Adding references to our domain in other places on the Web (e.g., logs, files, etc.)

DOI: 10.1002/asi

POPULARITY variable. Although the popularity of a Web site can be improved by increasing the number of links, the best way to do so is by improving its contents. A content with quality will be translated into cites, references, and of course links to the Web site, and this means an increase in its popularity.

After applying the aforementioned improvements, the following changes were obtained:

- The Page Rank: In the first 2 months of the Web site, this was 0 (i.e., not evaluated), on the third month was 3, and finally 4 in the fourth month of the Web site life.
- The indexing of the three main search engines (thanks to the register) took place.
- There was an increase in the number of entering links because the search engines found pages upon which our site was referenced.

Our tool has thus improved its position in the main search engines. For example, from not appearing on the first page of any search engine under any search term, when searching now:

- "know Web visibility" in Google and in MSN appears in second place on the first page;
- "Web visibility rankings" in Google appears in third place on the first page, and also appears on the first page in MSN, although not on the first positions; and
- "Web visibility analysis" appears on the first page of both Google and MSN, but it does not appear in the first positions.

We nevertheless believe that these data will improve in future queries, thanks to the continual improvements that we are applying to our site.

Conclusions

The correct positioning of contents and services that a site offers as Web pages, in an environment which is as highly competitive as is the Internet, has a direct impact upon the traffic it receives. Consequently, the positioning audit is a fundamental task which must involve Web administrators, designers, programmers, content generators, and experts in the marketing area.

The proposed method offers a starting point for the quantitative audit of site positioning on the Internet, looking at it from the viewpoint of search engines and the way in which they organize information. Its application does not require advanced computer knowledge or any investment in analysis tools. The method is accompanied, in most of its tasks, by several measures of the key aspects of each task. The changes included in a Web site to improve the value of the measures proposed will be rewarded by greater visibility, greater popularity, and an increase in traffic to our Web site. Compared with other similar works, such as that proposed by Petrieck, Escher, Cox, and Margetts (2006), our method is more complete and does not imply the loss of ease of use.

It is necessary to take the proposed method as a starting point that will have to be maintained and updated according to the algorithms applied by search engines. As we know, these algorithms vary to avoid the possible "frauds" commited by Web designers who wish to improve their Web visibility (i.e., by taking advantage of the policies followed by the search engine algorithms).

Part of our method lies in the implementation of the KYV tool which we also have presented in this work (see http://kyv.webportalquality.com).

We have presented the application of the complete method to museothyssen.org and to kyv.webportalquality.com, and have applied the improvements presented by the KYV tool. In both cases, visibility was considerably improved. We shall, nevertheless, continue working on the validation of this tool by applying it to other types of Web sites to refine it and to ascertain its utility.

Acknowledgments

This article was supported by the MEC Spanish Ministry projects CALIPSO (TIN20005–24055-E), ESFINGE (TIN2006–15175-C05–05), and by the DIMENSIONS project (PBC-05–012–1) financed by the Castilla-La Mancha government. The authors thank Felipe Moyano for implementation of the KYV. The authors also thank the reviewers for their valuable comments.

References

Aguillo, I.F., Prieto, J.A., Ortega, J.L., & Granadino, B. (2005, July). Ranking of world universities in the Web. Poster presented at the 10th International Conference of the International Society for Scientometrics and Informetrics, Stockolm.

Almind, T.C., & Ingwersen, P. (1997). Informetric analyses on the world wide web: Methodological approaches to "webometrics." Journal of Documentation, 53, 404–426.

Arroyo, N. (2004, September). What is the Invisible Web? A crawler perspective. Association of Internet Researchers (AoIR) International Conference. Brighton, United Kingdom. Retrieved 22 May 2008, from http://cybermetrics.wlv.ac.uk/AoIRASIST/arroyo.html

Baeza-Yates, R., & Poblete, B. (2003). Evolution of the Chilean web structure composition. Proceedings of the Latin American Web Conference (pp. 11–13). Santiago, Chile: IEEE CS Press.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. In Proceedings of the 7th WWW Conference (pp. 107–117), Brisbane, Australia.

Brinkmeier, M. (2006). PageRank revisited. ACM transactions on Internet Technology, 6(3), 282–301.

Broder, A.Z., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., & Wiener, J.L. (2000). Graph structure in the Web. In Proceedings of the 9th World Wide Web Conference (pp. 309–320). ACM: Amsterdam, The Netherlands.

Chirita, P., Olmedilla, D., & Nejdl, W. (2003). Finding related hubs and authorities. Proceedings of the First Latin American Web Congress (pp. 214–215). Santiago, Chile: IEEE Publishers.

Crowther, P., Bechhofer, S., & Horan, B. (2004). Shedding light on the hidden Web. UK: University of Manchester, Department of Computer Science Pre-Print No. CSPP-28.

Cutler, M., & Sterne J. (2000). E-metrics—Business metrics for the new economy. Tech. Report, NetGenesis Corp. Retrieved 22 May 2008, from http://www.targeting.com/emetrics.pdf

Dhyani, D., Ng, W.K., & Bhowmick, S.S. (2002). A survey of web metrics. ACM Computing Surveys, 34(4), 469–503. DOI: http://doi.acm.org/10.1145/592642.592645

European Indicators, Cyberspace and the Science-Technology-Economy System (EICSTES). (2007). Retrieved 10 September 2007, from http://www.eicstes.org

DOI: 10.1002/asi

Garfield, E. (1994). The impact factor. Current Contents, June 20, 25, 3–7.

Henzinger, M. (2007). Search technologies for the Internet. Science, 317, 468, 468–471. Retrieved 3 October 2007, from http://sciencemag.org

Ingwersen, P. (1998). The calculation of Web impact factors. Journal of Documentation, 54(2), 236–243.

Nielsen, J. (2004). Designing web usability. Markt-Technik Verlag. Munich, Germany.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The Page Rank citation ranking: Bringing order to the Web. Stanford, CA: Stanford University Press.

Petrieck, V., Escher, T., Cox, I.J., & Margetts, H. (2006, May). The web structure of e-government. Developing a methodology for quantitative evaluation. WWW 2006. ACM1-59593–323–9/06/0005. UK: Edinbourgh.

QL2 Software. Retrieved 3 October 2007, from www.ql2.com

Rafiei, D., & Mendelzon, A.O. (2000, June). What is this Page Known for? Computing Web Page Reputations. Computer Networks 33(1), 823–835 (13).

Smith, A. (1999). ANZAC webometrics: Exploring Australasian Web structures. Australasian Information Online & On Disc. Retrieved 3 October 2007, from http://www.csu.edu.au/special/online99/proceedings99/203b.htm

Touchgraph. Retrieved 3 October 2007, from http://www.touchgraph.com/

Web Indicators for Science, Technology and Innovation Research (WISER). (2007). Retrieved 10 September 2007, from http://www.wiserweb.org

Wu, B., & Davison, B.D. (2006, April). Undue influence: Eliminating the impact of link plagiarism on web search rankings. ACM SAC 2006. ACM 1–59593–108–2/06/0004. Dijon, France.

DOI: 10.1002/asi

Journal of the American Society for Information Science and Technology

JASIST

INSTRUCTIONS TO AUTHORS

The *Journal of the American Society for Information Science and Technology* (*JASIST*) is a scholarly journal devoted to the various fields of documentation and information science and serves as a forum for discussion and experimentation. All articles are subject to single-blind peer review prior to acceptance. Manuscripts for submission to the journal (in English only) should be forwarded to the Editor as follows:

1. Go to your Internet browser (e.g., Netscape, Internet Explorer).
2. Go to the URL http://mc.manuscriptcentral.com/jasisit
3. Register (if you have not done so already).
4. Go to the Author Center and follow the instructions to submit your paper.
5. For review purposes, the text of the paper can be submitted in .doc, .rtf, or .ps format. Images can be in .jpg, .tif, .gif or .pict format and should be submitted separate from the paper. Because this Journal has a completely digital workflow, if you are eventually asked for a revision, you will be asked to provide the source files for your text in .rtf or .doc format for production purposes.
6. Please avoid .pdf, .pds, .cd, .txt, and other proprietary image formats. Also, please note that the system will not accept executable files, encapsulated formats, or other potentially viral material formats, such as .exe, .com, .bat, .zip, or .sea.
7. To ensure the highest print quality, figure files must be submitted with dpi (dots per inch) of 300 or higher for regular figures and at least 1200 dpi for line art (simple bar graphs, charts, etc.).
8. Please also submit your tables separately. All files will be concatenated automatically into a single .PDF file by the system during upload. This is the file that will be used for review.
9. Authors for whom English is a second language may choose to have their manuscript professionally edited before submission or during the review process. Authors wishing to pursue a professional English-language editing service should make contact and arrange payment with the editing service of their choice. For more details regarding the recommended services, please refer to http://www.blackwellpublishing.com/bauthor/english_language.asp

Please scan your files for viruses before you send them, and keep a copy of what you send in a safe place in case any of the files need to be replaced.

JASIST does not generally publish articles in multiple parts or those exceeding 36 pages. However, authors may submit articles exceeding 36 pages along with a justification included in the cover letter for the editor and reviewers to consider.

Books, monographs, reports, and other items for review may be sent to the Associate Editor, Book Reviews: Robin Peek, Graduate School of Library and Information Science, Simmons College, 300 The Fenway, Boston, MA 02115. Tel: (617) 521-2807; E-mail: robin.peek@simmons.edu; http://web.simmons.edu/~peek/jasistreview.html

Types of manuscripts: Three types of contributions are considered for publication: full-length articles, brief communications of 1,000 words or less, and letters to the editor. Letters and brief communications can generally be published sooner than full-length articles. All materials submitted will be acknowledged on receipt and (except for letters) subject to peer review. Copies of the referees' comments will be forwarded to the author, along with the editor's decision.

Copyright: No article can be published unless accompanied by a signed publica-tion agreement, which serves as a transfer of copyright from author to publisher. A publication agreement may be obtained from the editor or the publisher. A copy of the publication agreement appears in many issues of the journal. Only original papers will be accepted and copyright in published papers will be vested in the publisher. It is the author's responsibility to obtain written permission to reproduce material that has appeared in another publication.

Only articles that have not been published previously, that have not been submitted elsewhere, and that are not under review for another publication in any medium (e.g., printed journal, conference proceeding, electronic, or optical medium) should be submitted to *JASIST*. The *JASIST* editors will assume that submission of an article to this journal implies that all the foregoing conditions are applicable.

Note to NIH Grantees: Pursuant to NIH mandate, Wiley-Blackwell will post the accepted version of contributions authored by NIH grant-holders to PubMed upon acceptance. This accepted version will be made publicly available 12 months after publication. For further information, see http://www.wiley.com/go/nihmandate

Wiley-Blackwell also offers its OnlineOpen (http://www.blackwellpublishing.com/static/onlineopen.asp) and Funded Access (http://www3.interscience.wiley.com/authorresources/funded_access.html) services. Upon payment of the OnlineOpen or Funded Access fee, we will deposit the published version of the article into PubMed Central, with public availability in PubMed Central and on the journal's website immediately upon publication.

Open Access: As of January 2008, *JASIST* will follow the "green road" open access model, whereby authors publishing in the Journal are granted self-archival rights for preprints, with linking to the the final article and to the online journal, as detailed below.

Submitted version.

The Contributor has the right to self-archive on the Contributor's personal website or in the Contributor's institution's/employer's institutional repository or archive. This right extends to both intranets and the Internet. The Contributor may not update the submission version or replace it with the published contribution. Links to the final article on Wiley-Blackwell's website are encouraged where appropriate.

Version as accepted.

The Contributor has the right to self-archive on the Contributor's personal website or in the Contributor's institution's/employer's institutional repository or archive. This right extends to both intranets and the Internet. Prior to publication, the Contributor must include the following notice on the preprint: "This is a preprint of an article accepted for publication in Journal of the American Society for Information Science and Technology copyright (C) [year] (American Society for Information Science and Technology)". After publication of the contribution by Wiley-Blackwell, the preprint notice should be amended to read as follows: "This is a preprint of an article published in [include the complete citation information for the final version fot he contribution as published in the print edition of the Journal]", and should provide an electronic link to the Journal's WWW site. The Contributor may also update the preprint with any corrections made, in which case the notice shall further be amended with the following language: "This preprint has been updated to reflect changes in the final version."

Contributors should refer to the current Copyright Transfer Agreement for information regarding use of the Final Published Version of the article.

Format of submitted material: All manuscripts must be submitted in standard format (adequate margins, single column, reasonable font - no smaller than size 10, on standard sized paper - 21 × 28 cm). All copy, including references and captions, must be typed double-spaced. The first page of the manuscript must bear the title of the paper and the full names of the authors, as well as their affiliations, full addresses, telephones and fax numbers, and e-mail addresses. In the case of multiple authors, please indicate which author is to receive correspondence and proofs. Financial support may be acknowledged in the Acknowledgment section at the end of the article. An informative abstract of 200 words or less is required for articles and brief communications.

Style: In general, the style should follow the forms given in the *Publication Manual of the American Psychological Association* (Washington, DC : 2001). The use of metric forms is required. Organization: In general, the background and purpose of the study should be stated first, followed by details of the methods, materials, procedures, and equipment used. Findings, discussion, and conclusions should follow in that order. Appendices may be employed where appropriate. The Publication Manual of the American Psychological Association should be consulted for details as needed.

Figures: Figures should be professionally prepared and submitted in a form suitable for reproduction (camera-ready copy). Computer-generated graphs are acceptable only if they have been printed with a good quality laser printer. All artwork must be identified on the back in light blue pencil with the figure number, author's name and address, and a shortened form of the title. Photographs should be submitted as unscreened glossy prints (20 × 25 cm), digitized in the paper.

Bibliography: The accuracy and completeness of the references is the responsibility of the author. References to personal letters, papers presented at meetings, and other unpublished material may be included. If such material may be of help in the evaluation of the paper, copies should be made available to the editor. Papers which are part of a series should include a citation of the previous paper. Explanatory material may be appended to the end of a citation to avoid footnotes in text. Wiley's Journal Styles Are Now in EndNote, a software product that we recommend to our journal authors to help simplify and streamline the research process. Using EndNote's bibliographic management tools, you can search bibliographic databases, build and organize your reference collection, and then instantly output your bibliography in any Wiley journal style.

Download reference style for this Journal: If you already use EndNote, you can download the reference style at http://www.interscience.wiley.com/jendnotes To learn more about EndNote, or to purchase your own copy, go to http://www.wiley.com/trackthrough?urlcode=53026752.

Technical support: If you need assistance using EndNote, e-mail: endnote@isiresearchsoft.com, or visit http://www.endnote.com/support.

The format for citations in text and for bibliographic references follows the *Publication Manual of the American Psychological Association* (5th ed., 2001). Citation of an author's work in the text should follow the author-date method of citation; the surname of the author(s) and the year of publication should appear in text.

For example:

Paisley (1993) found that . . .

Recent research has shown that . . . (Schauder, 1994).

In other work (Gordon & Lenk, 1992; Harman, 1991) . . .

Examples of citations to a journal article, a book, a chapter in a book, and published proceedings of a meeting follow:

Buckland, M., & Gey, F. (1994). The relationship between recall and precision. Journal of the American Society for Information Science, 45, 12–19.

Borgman, C. L. (Ed.). (1990). Scholarly communication and bibliometrics. London: Sage.

Bauin, S., & Rothman, H. (1992). Impact of journals as proxies for citation counts. In P. Weingart, R. Schringer, & M. Winterhager (Eds.), Representations of science and technology (pp. 225–239). Leiden: DWSO Press.

Hoppe, K., Ammersbach, K., Lutes-Schaab, B., & Zinssmeister, G. (1990). EXPRESS: An experimental interface for factual information retrieval. In J.-L. Vidick (Ed.), Proceedings of the 13th International Conference on Research and Development in Information Retrieval (ACM SIGIR '91) (pp. 63–81). Brussels: ACM.

Electronic document: The date should indicate either the date of the publication or the most recent update; if neither can be determined, then it should indicate the date of the search. The availability statement should provide sufficient information to retrieve the material.

Ford, A., Kline, C., & Wayne, B. (2002). Information science and digital archives. Libraries Journal Today, 12, 340–350. Retrieved October 13, 2002, from http://www.libraryinfo.org/articles.html

Smith, A., & Jones, B. (2002, August 19). Libraries in the digital age. Journal of Libraries, 54, 1110–1120. Retrieved October 13, 2002, from PsycINFO database.

Correspondence: All correspondence beyond that of the refereeing process should be addressed to: Journal of the American Society for Information Science and Technology, c/o Wiley-Blackwell, 350 Main Street, Malden, MA 02148 -5020.

Reprints: Reprints of articles may be ordered from the publisher when corrected proofs are returned.

For more detailed instructions on general journal manuscript submission guidelines, please visit our Website at www.wiley.com and use the Resources for Authors link. For your convenience, you can also go directly to the link at http://www.wiley.com/WileyCDA/Section/id-301854.html

Professor Donald H. Kraft, Editor. Tel: (719) 481-4843; E-mail: kraft@bit.csc.lsu.edu