EFFECTIVE APPLICATION OF METADATA IN SOUTH AFRICAN HEI WEBSITES TO ENHANCE VISIBILITY TO SEARCH ENGINES

M. Weideman, Faculty of Business Informatics, Cape Technikon.

It has been proven that, in general, it is difficult to find relevant information on the Internet. Many programs exist to assist the user in this process, but users have trouble in using them effectively, and in interpreting their results correctly.

A variety of metadata systems exist, which could enhance the electronic visibility of any website. These include HTML metatags and the Dublin Core standard. Some search engines recognize metadata in webpage coding, and use it to properly categorize the webpage in its index. This should produce high quality answers when a user does a search.

It was assumed that the websites of South African higher education institutes are considered important enough to warrant maximal exposure. An empirical experiment was done to determine the degree of effective usage of metadata systems on these websites.

Three of the 36 websites could not be accessed after repeated attempts. None of the remaining 33 made any use of the Dublin Core standard in any way. In terms of the use of standard HTML metatags: 1 (3%) used no metatags at all, 19 (58%) used only the basic "TITLE" metatag, 4 (12%) made basic use of the relevant metatags while the remaining 9 (27%) made reasonable use of the relevant metatags. It is clear that electronic visibility has been neglected during the design of most of these websites.

1 INTRODUCTION

The human race has been involved in storing and retrieving data and information for centuries. A common problem experienced in this area has always been the effective retrieval of relevant information in a short time. The purpose of this paper is to report on an empirical experiment on coding schemes which attempt to improve categorization and subsequent retrieval of relevant information from the Internet.

Work on storage and retrieval of information started approximately at the beginning of the third millennium BC. The Sumerians are credited as being the first people to store and classify written materials into “library collections”, with the purpose of allowing various social groups to function better (Cover 2001). Everyday activities and literature were recorded on clay tablets, which were stored in special areas, with only a label bearing the opening words of the document as sole method of indexing (Greek Libraries 2001). The physical creation of these clay labels could be viewed as one of the first implementations of technology towards establishing indexing. In the absence of any advanced technological tools to make IR possible, these “libraries” were little more than marked collections of “documents”.

During the Middle Ages, indexing and simple classification of manuscripts were done. The indexers involved in this task were surrounded by an aura of mysticism, resulting from the coding schemes and alphabetical keys (Ingwersen 1992:61). Work on cataloguing started after this period, using written, guardbook and card catalogues, followed by the period of mechanization.

Lancaster broadly defined the development of methods for the physical implementation of IR systems as spanning a number of decade-long periods, from the pre-1940s (pre-co-ordinate systems, manual, printed book/card form) to the 1990s (completely paperless systems) (Lancaster 1978:15).

The high-powered computer era we currently find ourselves in provided the much-needed technology to empower this development, where document matching is made through inverted indices, string and positional searches. This increasing power of the technology has removed the economic constraint on the searching mechanism: any characteristic of the document could now be matched to a search query. In fact, there was now no technical constraint to prevent an index from including every single term of a given textual document in the index. This has been done some time ago, in the form of typical Bible concordances such as *Strong’s Exhaustive Concordance*, first published in 1890. Since all the matching features so far have been based on text only, there remain some areas where matching has not yet been perfected: special notations such as music files, photographs and other graphical information.

2 DEFINITION OF CONCEPTS

2.1 THE USER

A skill which eludes many average Internet users, is the finding of relevant information on the Internet in a short time. However, many Internet users rely on search engines to find relevant data for a variety of purposes on a daily basis (Wallace and others, 2000).

Lancaster predicted a change in the searching process, as it was known decades ago, when he described a future scenario in 1978:

"The scientist of the year 2000 will use a terminal in many different ways: to receive, transmit, compose and search for text, seek answers to factual questions, build files and converse with colleagues. The terminal provides a single entry to a wide range of capabilities that will substitute, wholly or in part, for many activities now handled in different ways" (Wallace and others, 2000).

Lancaster's "scientist" can be viewed as today's average PC user, his "terminal" has become the desktop computer, "conversing" has evolved into e-mail and his "different ways" could be the shift from using an intermediate searching expert to doing the search oneself.

However, some studies have shown that early information users were not all eager to become involved in the extraction of knowledge from an electronic source. Over three decades ago Jackson found that engineers are reluctant to use information sources. The United Engineering Information Service failed to elicit financial support from the engineering profession to establish its services (Jackson, 1971).

To further complicate matters, Large warns that information seekers must not be treated as a homogenous group – they differ in many aspects, where their information retrieval experience level has a large differentiating effect (Large and others, 1999). Other authors warn that there is a difference in the way that WWW searchers and traditional OPAC searchers work when looking for information (Jansen and Pooch, 2001). In contrast to this finding, a remarkable similarity was found between the basic methodology of early online searching and modern Internet searching. Lancaster listed a four point sequence of steps to be followed for online retrieval: logging on, negotiating the search process, manipulating the results and logging off (Wallace and others, 2000).

How does the human searcher approach the task of finding information in a collection of information? How does he/she convert the perceived information need into a search strategy? What effect does experience have on the user's perception of information usefulness?

These questions about the user cannot be answered here, but have been addressed in great detail in other research literature. **For the purpose of this project, an Internet user is defined as a person who uses search engines to find relevant information on the Internet.**

2.2 THE SEARCH ENGINE

Search engines provide the average Internet user with a (mostly) free, apparently easy way to find general information on the Internet. They are programs which offer interaction with the Internet through a front end, where the user can type in a search term, or make successive selections of relevant directories. The search engine software then compares the search term against an index file, which contains information about many websites. Matches found are returned to the user via the front end. The index is updated regularly either by human editors or by automated programs (called spiders, robots or crawlers). Both humans and spiders simply collect information of new websites by visiting as many websites as possible, and then building them into the index (Weideman, 2002).

The three components of a typical search engine (front end, index file and information collectors) have close parallels in the components of a typical information retrieval system, as defined by Lancaster many years before the Internet and search engines became freely accessible. A search engine front-end maps to the ”user system interface”, the index file maps to the “indexing subsystem” and the information collectors map to the “document selection subsystem” (Wallace and others, 2000).

However, one author claims that search engines are complex, trusted without being understood and that users simply deal with their answers without understanding why they receive those answers (Lynch, 2001).

Many authors from both the popular press and other sources have done evaluations, comparisons, measurements and a variety of other tests on a large number of search engines (Brewer, 2001, Oppenheim and others, 2000, Taubes, 1995, and a number of others).

In these studies and other works, the following search engines were mentioned:
AltaVista, Excite, Galaxy, Harvest, Hotbot, Infoseek, Looksmart, Lycos, Magellan, MetaCrawler, Northern Light, Open Text, PlanetSearch, Savvysearch, Search.com, UKPlus, Webcrawler, WWW Worm and Yahoo!. A list of commonly used search engines and meta search engines with their URLs and a short evaluation, is available on the WWW (Weideman, 2001).

However, in the studies mentioned above, a variety of methods were being used, and it is impossible to compare the results of the different evaluations with each other. Oppenheim did a study on this issue, and identified the difficulty of establishing recall on the fast-changing Internet scene as being one of the major reasons why different evaluations were difficult to compare. The creation of a standard set of tools will solve this problem and will also allow the tracking of the variation in performance of a search engine over time (Oppenheim and others, 2000). The problem of measuring recall could remain unless the search engines are compared on closed data bases where prior evaluations of pertinence of the contents have already been made.

Another study confirmed that users make little use of advanced search features. For example, a test on 1 025 910 queries submitted to the Excite search engine revealed that the following features were only used by small percentages of users, as indicated (Spink and others, 2001):

-Inclusion (+)	2%
-Exclusion (–)	0.001%
-Boolean operators (NOT / AND NOT)	0.0003%

Yet another study indicated that searchers spent a relatively short amount of time searching for one topic: the average search session seemed to last between five and 10 minutes only (Cooper, 2001). This could possibly be a reason for the lack of use of advanced operators.

From the work already done on search engines, it has become clear that these programs play an important role in the lives of Internet users. For the purposes of this study, **a search engine is defined as a program that is accessible by any average user, capable of accepting user input which defines the information it produces as output to this user.**

2.3 THE WEBPAGE

A large number of authors claim that the relationship between users and webpages with relevant information, stored on the Internet, is not always a positive one:

- "Currently, search is simply bad" (Sherman, 1999).
- "…some respondents seemed confused about what they were to report when asked to list query terms for their search" (Spink, Bateman and Jansen, 1999).
- "…the user's ability to specify good search terms and create complex search queries to clearly and precisely capture relevant retrieval seems rather low." (Spink, Bateman and Jansen, 1999).
- "As the World Wide Web grows exponentially, discovery and retrieval of useful educational material grows more problematic" (Sutton, 1999).
- "I find it difficult to search information on the Internet…" (Sutton, 1999).
- "…information seeking is a complex and difficult process for these students, who seek to reduce the task to finding an obvious answer or finding a good website…" (Taubes, 1995).
- "Only 33% of the Internet users agree or strongly agree with the statement 'It is easy to perform subject searches on the Internet.'" (Voorbij, 1999).
- "… both novice and experienced searchers were overconfident in their performance" (Wolfram and Dimitroff, 1997).

In summary it is clear that webpages also play an important part in the life of the Internet user. For the purposes of this study, **a webpage is defined as an entity, stored on the Internet, accessible by any average user, which contains some information.**

2.4 METADATA

During the Middle Ages, indexing and simple classification of manuscripts were done. The indexers involved in this task were surrounded by an aura of mysticism, resulting from the coding schemes and alphabetical keys (Ingwersen, 1992).

Work on cataloguing started after this period, using written, guardbook and card catalogues, followed by the period of mechanization. This phase saw various card systems being developed, mostly in the form of punched cards: optical coincidence, the Zatocoding system of Mooers, edge-notched coding and 80-column punched cards.

The high-powered computer era we currently find ourselves in provided the much-needed technology to empower this development, where document matching is made through inverted indices, string and positional searches. This increasing power of the technology has removed the economic constraint on the searching mechanism: any characteristic of the document could now be matched to a search query. In fact, there was now no technical constraint to prevent an index from including every single term of a given textual document in the index. This has been done some time ago, in the form of typical Bible concordances such as *Strong's Exhaustive Concordance*, first published in 1890. Since all the matching features so far have been based on text only, there remain some areas where matching has not yet been perfected: special notations such as music files, photographs and other graphical information.

The most ideal document representation is simply to include the whole document as the index, but the initial absence and later the limitations of technological tools (such as storage space) made this ideal impossible to achieve.

In the absence of computer-based full-text retrieval systems, two main lines of thought were initially followed to provide a summary of the document: assigned and derived indexing. Assigned indexing depended on an indexer's perception of the document contents, the matching of that perception to a term list or classification and the selection of appropriate descriptors to be assigned to the document. Derived indexing depended on the indexer (later the computer program) to extract terms as used in the document – keyword in context is the most notable early example.

During the late 1950s and 1960s, landmark work was done by noted authors in this area of document content presentation. The controversial Uniterm system sparked interest in the UK and the USA, which led to the Cranfield tests by Cleverdon and Keen (Cleverdon, Mills and Keen, 1966, Spink, Bateman and Jansen, 1999).

Documents were indexed via a single term in the Uniterm system (hence the name), having been extracted from the document title or abstract. After some structured tests, Uniterm results were compared to those using more traditional indexing methods. The test apparently broke down owing to the disagreement over relevance judgement, and the results were inconclusive. One group of testers claimed that the Uniterm system worked well, while the other claimed the exact opposite (Ellis, 1996).

The actual Cranfield series of tests was done at the College of Aeronautics, Cranfield, UK. It involved another comparison of performance, this time between the Uniterm system and a modified version of the Universal Decimal Classification (UDC) system. A subset of 200 documents on aeronautics was extracted and used as the master collection of documents. A further extract of 40 documents from this set was made, and these documents were used to generate a set of 40 artificial requests. The assumption was that if query number one was submitted, document number one should be returned as being the most relevant document from the collection of 200 documents.

Although certain inherent limitations of the study were evident, it did prove the effectiveness of the Uniterm system above the UDC classification (Ellis, 1996).

Modern day versions of metadata include HTML metatags and the DCMI (Dublin Core Metadata Initiative). More recent work indicates that the use of these metadata elements plays an important part in making webpages more visible. A large number of authors make a case to urge website authors to make use of HTML and the DCMI metatags respectively. The HTML TITLE, KEYWORD and DESCRIPTION metatags have been singled out as being most relevant to website visibility (Altavista, 2002, Washington University, 2002).

Again it is clear that metadata plays an important part in making webpages more visible to the user. For the purposes of this study, **metadata is defined as data about a webpage, contained in the webpage, used to describe the contents and other features of it.**

2.5 OPERATORS

The syntax of search engines differs widely, and a wide variety of operators exists to allow focusing of a search query. Lancaster referred to natural language searching, where no operators are used, in an early reference (Lancaster 1978:279). Some of the more commonly used operators include phrases (using quotes), inclusion operator (+) and exclusion operator (-). Other, more advanced and less often used ones include Boolean operators (AND, OR , NOT), proximity operators (NEAR), stemming (*) and field limiters (DOMAIN, TITLE). These operators are discussed by a variety of authors (Boulton 2001).

Frants and other authors claim that Boolean searching features in most retrieval systems (Frants and others 1999:86). These authors have shown that criticism of Boolean systems is mostly directed at the methodology employed rather than the Boolean principle itself. However, they regard Boolean and other search principles as being of equal importance (Frants and others 1999:94). A full description of the usage of Boolean operators, both in general databases and Internet search engines, is given in a book on electronic information seeking (Large, Tedd and Hartley 1999:144-152).

In summary, a recent study claimed to have found that "… most people use few search terms, few modified queries, view few Web pages, and rarely use advanced search features" (Spink and others 2001:226). This contributes to the problem of easily finding relevant information in a short time on the Internet. The author is of the belief that a contribution can be made to the solution of this problem by webpage authors.

2.6 QUERY FORMULATION

In general, the best way to construct a search query during an Internet search is not obvious to the average searcher. A common approach of these searchers is to simply type in one or more consecutive words into the search box. Where some search engines would add an implied OR operator between words (Infoseek), others will insert an AND (AltaVista and Google in their basic searching modes), and some might even treat it as a phrase (Looksmart). These three widely differing approaches would produce very different results, which would confuse rather than enlighten an average searcher. As an example, an Internet search using five common words with AND operators between them will produce x answers from a given database. The same search using the same five terms and the same database, but using OR operators between the words, will produce y answers, where $x \ll y$.

It became clear from the literature that many users found the formulation of a search query difficult. This was also evident even in searching via an intermediary on non-web electronic database sources in 1993 (Su 1993:97). Hirsh found in 1999 that elementary school children "… have difficulty formulating and modifying search queries … and have difficulties with spelling and Boolean logic" (Hirsh 1999:1266).

During a survey of 316 Excite users, Spink, Bateman and Jansen stated that few users employed logical operators and even fewer used the syntax correctly. They also had problems with search phrases and with the construction of "good search terms" and "complex search queries" (Spink, Bateman and Jansen 1999:123,125). Wallace, Kupperman and Krajcik had similar experiences in a study where sixth-graders had ample time to search for information – they found it problematic to convert their question into a query (Wallace and others, 2000:87). Bilal also confirmed this phenomenon by referring to the "… naïve Web navigational skills… " of a group of seventh grade science students who used the Yahooligans! web search engine (Bilal 2001:134).

It is clear from the literature that many humans experience problems when faced with query formulation for information extraction from the Internet. The coder of a website can play a role in making this process easier.

3 SEARCHING SUCCESS RATE

Lancaster discussed some factors which influence success in online searching many years before the advent of Internet search engines (Lancaster 1978:176). However, lately many authors agree that learners should know how to retrieve information from electronic sources (Cronje and Clarke 1999:214, De Jager and Sayed 1998:197, Edling 2000:11).

A large number of authors claim that it is difficult to find relevant information on the Internet:

- "Currently, search is simply bad" (Sherman, 1999).
- "…some respondents seemed confused about what they were to report when asked to list query terms for their search" (Spink and others, 1999:122).
- "…the user's ability to specify good search terms and create complex search queries to clearly and precisely capture relevant retrieval seems rather low." (Spink and others, 1999:125).
- "As the World Wide Web grows exponentially, discovery and retrieval of useful educational material grows more problematic" (Sutton, 1999).
- "Only 33% of the Internet users agree or strongly agree with the statement 'It is easy to perform subject searches on the Internet.'" (Voorbij, 1999,604).
- "I find it difficult to search information on the Internet…" (Voorbij, 1999,605).
- "…information seeking is a complex and difficult process for these students, who seek to reduce the task to finding an obvious answer or finding a good website…" (Wallace and others, 2000).

A large body of evidence points to the fact that most Internet searchers find it difficult to locate and retrieve relevant data on the Internet.

4 SOUTH AFRICAN HIGHER EDUCATION INSTITUTES

At the time of writing, there were a total of 36 Higher Education Institutes in South Africa: 21 Universities and 15 Technikons. Table 1 lists their names and website addresses.

NUMBER	HEI	URL
1	Medical University of South Africa	www.medunsa.ac.za
2	University of Durban-Westville	www.udw.ac.za
3	University of Fort Hare	www.ufh.ac.za
4	University of Northwest	www.uniwest.ac.za
5	University of Transkei	www.utr.ac.za
6	University of the North	www.unorth.ac.za
7	University of the Western Cape	www.uwc.ac.za
8	University of Zululand	www.uzulu.ac.za
9	Vista University	www.vista.ac.za
10	University of Venda	www.univen.ac.za
11	Potchefstroom University for Christian Higher Education	www.puk.ac.za
12	Rand Afrikaans University	www.rau.ac.za
13	Rhodes University	www.ru.ac.za
14	University of Cape Town	www.uct.ac.za
15	University of Natal	www.und.ac.za
16	University of Port Elizabeth	www.upe.ac.za
17	University of Pretoria	www.up.ac.za
18	University of Stellenbosch	www.sun.ac.za
19	University of the Free State	www.uovs.ac.za
20	University of the Witwatersrand	www.wits.ac.za
21	Border Technikon	www.bortech.ac.za
22	Cape Technikon	www.ctech.ac.za
23	Eastern Cape Technikon	www.tktech.ac.za
24	M.L. Sultan Technikon	www.mlsultan.ac.za
25	MangosuthuTechnikon	www.mantec.ac.za
26	Northwest Technikon	www.tnw.ac.za
27	Peninsula Technikon	www.pentech.ac.za
28	Port Elizabeth Technikon	www.petech.ac.za
29	Technikon Free State	www.tofs.ac.za
30	Technikon Natal	www.ntech.ac.za
32	Technikon Northern Gauteng	www.tng.ac.za
33	Technikon Pretoria	www.techpta.ac.za
34	Technikon SA	www.tsa.ac.za
35	Technikon Witwatersrand	www.twr.ac.za
36	Vaal Triangle Technikon	www.tritek.ac.za

TABLE 1 - South African HEI's

The first 10 entries in Table 1 have been classified as "Historically Disadvantaged Universities", while numbers 11 to 20 have been called "Historically Advantaged Universities". The reasons for this distinction lie rooted deeply in the history of the country, and a discussion of this classification is outside the scope of this paper.

As shown in Table 1, each HEI has a website. In most cases, these websites advertise the courses offered by the HEI, supplies mission statements, contact details, management structures, news events and even some opportunities to register online for certain courses. In all cases, it appeared as if the website played a central part in the marketing effort and online presence of the HEI.

It was assumed that each website was important enough to each HEI to warrant expenses on expert design, natural computer-human interfaces and the use of the latest technologies in animation and graphical presentation. Many websites indeed score highly on these criteria, and present the user with a pleasant and easily navigable online experience.

However, it is not immediately obvious to what extent the design of each website has been tailored to make the website "search engine friendly". In other words, will a roving search engine spider find the website easy to catalogue, and will a user in search of a HEI to enrol at find any of these websites with ease? The purpose of this research project was to establish to what extent website coding has incorporated some of the basic features available to coders to achieve the goal of making a website electronically visible.

5 METHODOLOGY

The website of every HEI in South Africa was visited, and the HTML coding of its homepage was recorded and inspected. Specific attention was given to the presence or absence of metadata, and if present, to the way it was used. The usage of relevant HTML metatags (TITLE, KEYWORD and DESCRIPTION) and Dublin Core coding was inspected.

After attempting to visit the 36 URL's listed above, 3 continuously gave time-out errors, even though repeated attempts over two days were made to access these sites. As a result, the remaining 33 websites formed the sample. Close inspection of the HTML coding revealed that a wide variety of combinations of metatags existed, which made it necessary to create categories into which these websites could be classified in terms of coding. Table 2 lists these categories and their meanings.

CATEGORY	DESCRIPTION
A	Extensive use of HTML metatags
B	Basic use of HTML metatags
C	Only TITLE metatag used
D	No metatags used

TABLE 2 - HTML Metatag Categories

Examples of coding meeting the criteria to deserve a place in the first two of these four categories are given below. Category C and D are self-explanatory.

Category A:

<meta name="keywords" content="mangosuthu technikon,durban,south africa,education,students,skills, experience,knowledge,curriculum,qualifications,graduates,training,education,students,skills,experience,know ledge,curriculum,qualifications,graduates,training">

<META NAME="Description" CONTENT="The University of the North strives to be a quality institution of higher learning and critical reflection, which is innovative responsive to change, is rooted in the issues of the society it serves.">

Category B:

<meta name="Keywords" content="University of Stellenbosch, Stellenbosch University, Maties, Stellenbosch, university">

<meta name="keywords" content="University of Natal, South Africa - Web Site">

Furthermore, a scoring system was implemented to enable a score to be allocated to each webpage. A webpage would earn a score of 0 for each one of these metatags if it were not present at all. A score of 1 would be allocated if the tag was present, but not used effectively (eg. title totally irrelevant, too many keywords, no text inside tag, text contains excessive forms of spam, spelling mistakes in important keywords, etc). A score of 2 was allocated if the element was present, and if it was used effectively. The webpage score therefore ranges between 0 and a maximum of (3 X 2) = 6.

The author would use the judgement of Sullivan plus personal experience based on many years of exposure in the field, for the subjective issues of effective use of metatags just described, including Categories A and B (Sullivan, 2002).

6 RESULTS AND ANALYSIS

Inspection of the 33 HEI websites' HTML coding produced not one instance of the usage of any one of the DCMI metatags. However, a number of HTML tags were found. The header section of every valid URL in the sample was inspected, a judgement was made on the effective usage of each one of the three metatag elements, the scoring system was applied, and the results summarized.

6.1 CATEGORIES

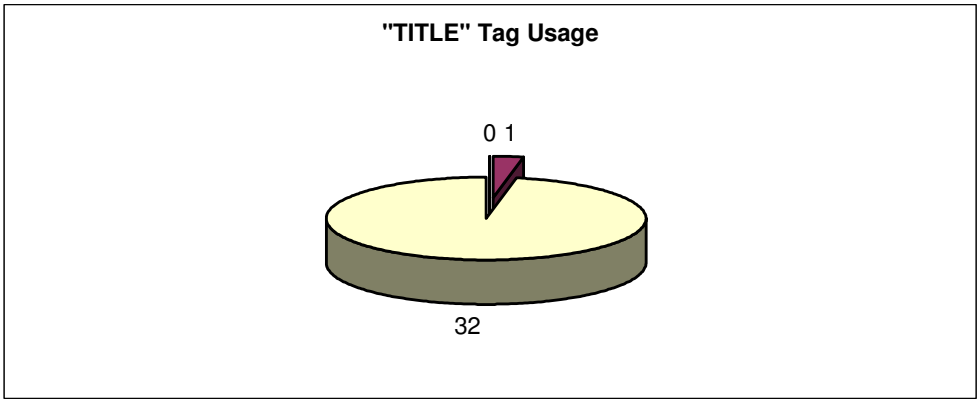
A summary of the scores achieved is given in Table 3.

CATEGORY	SCORE (/33)	SCORE (%)
A	9	27,3
B	4	12,1
C	19	57,6
D	1	3

TABLE 3 - Overall Scores achieved

More than 60% of the websites fall in the lower two categories, indicating neglect of the most basic elements of search engine friendly coding. Only 27,3% of the websites in question made effective use of the relevant HTML metatags to enhance their visibility.

6.2 TITLE metatag

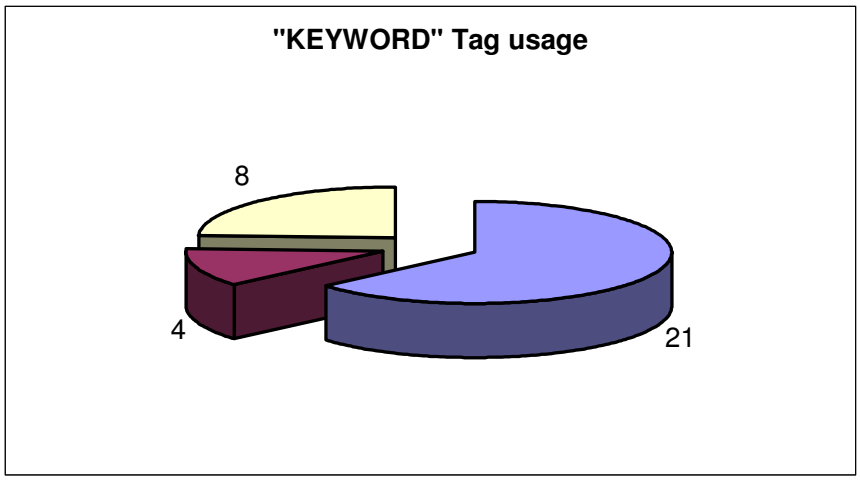


GRAPH 1 - Usage of TITLE Metatag

From Graph 1 it can be seen that 32/33 (97%) of webpages made effective use of the TITLE metatag and only 1/33 (3%) had this tag in the coding of the webpage, but did not use it correctly/effectively.

In summary, a negligible fraction of the sample did not use the TITLE metatag effectively. Since this feature is used to identify a webpage in the result listing produced by most search engines, users would at least be assured that the title of the webpage, as described by the coder, would be represented accurately in 97% of all cases. However, the value of having a correct title displayed, could be diminished by the lack of other, more useful indicators like the KEYWORD and DESCRIPTION metatags.

6.3 KEYWORD metatag

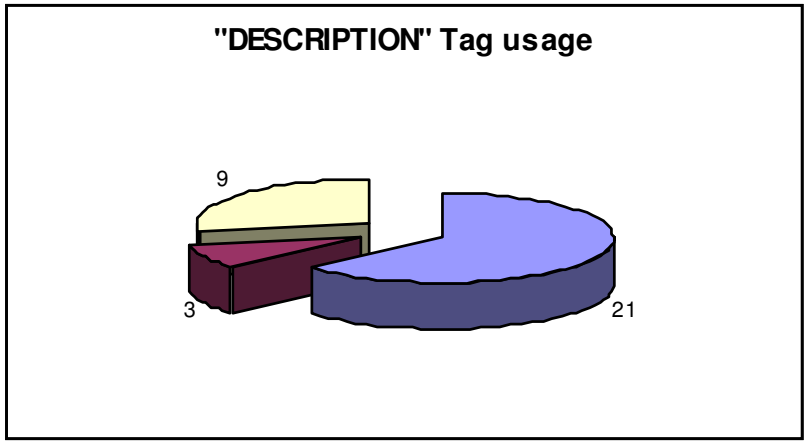


GRAPH 2 - Usage of KEYWORD Metatag

From Graph 2 it can be seen that 8/33 (24,2%) of webpages in the sample made effective use of the KEYWORD metatag and 4/33 (12,1%) had this tag in the coding of the webpage, but did not use it correctly/effectively. A total of 21/33 (63,7%) did not use it at all.

In summary, less than ¼ of the sample of websites made effective use of keywords to draw visitors to their site. More than ¾ of the sample made no or ineffective use of this metatag. This figure is lower than the equivalent figure for the TITLE metatag, indicating apparent indifference by webpage authors towards the use of metatags in webpage coding.

6.4 DESCRIPTION metatag



GRAPH 3 - Usage of DESCRIPTION Metatag

From Graph 3 it is clear that ONLY 9/33 (27,3%) of webpages in the sample made effective use of the DESCRIPTION metatag and 3/33 (9%) had this tag in the coding of the webpage, but did not use it correctly/effectively. A total of 21/33 (63,7%) did not use it at all. Once again there appears to be a low amount of effective usage of this metatag, compared to the previous two. Only about ¼ of the sampled webpages used it effectively, leaving the other ¾ without it, or using it ineffectively.

6.5 Overall Score

The average score for the 33 webpages was 3,21 out of 6, or 53,5%. This figure is a rough indication of the weighting a webpage has in an attempt to be spidered by a search engine crawler, and listed efficiently in the search engine database. A remarkable figure of 20 out of the 33 webpages used only the TITLE metatag, which guarantees only the most basic meaningful listing on a search engine result page.

7 CONCLUSION AND RECOMMENDATIONS

The figures above indicate a lack of application of a basic but potentially effective methodology to increase webpage visibility. Application of the TITLE metatag involves typically a single line of coding, containing the most basic form of description of a webpage, for example:

<title>Welcome to the University of Cape Town</title>

Secondly, use of the KEYWORD metatag involves some consultation with management, to extract relevant and meaningful keywords, which then have to be coded into the webpage. An example of a descriptive set of keywords is given:

<META NAME="KeyWords" CONTENT="Witwatersrand, university, Wits, Wits University, University of the Witwatersrand, Johannesburg, student, students, teaching, degree, lectures, lecture, lecturing, courses, college, colleges, Gauteng, South Africa, graduate, postgraduate, undergraduate, degrees, research, prospectus, faculty, faculties, learning, learn, education, tertiary, higher education, alumni">

Finally, the DESCRIPTION metatag is also relatively easy to code and include in a webpage. An example:

<meta name="DESCRIPTION" content="RAU strives for international recognition as a dynamic and socially-sensitive leader in higher education in the economic heartland of South Africa.">

It is recommended that HEI management insist on the implementation of these metatags on at least the homepages of their Institutes' websites. If this visibility enhancing mechanism is ignored, potential income opportunities are neglected.

Finally, the total absence of Dublin Core metatags indicates either ignorance on the part of designers, or lack of usefulness of these tags. Further research could shed light on this topic.

8 REFERENCES

Altavista, 2002. Search Help. [Online]. Available WWW: http://help.altavista.com/adv_search/ast_haw_metatags. Accessed: 1 July 2002.

Bilal, D. 2001. Children's use of the Yahooligans! web search engine: II. Cognitive and physical behaviors on research tasks. Journal of the American Society for Information Science and Technology. 52(2): 118-136.

Boulton, R. 2001. An introduction to information retrieval. [Online]. Available WWW: http://www.omsee.com/developer/docs/intro_ir.html. Accessed: 10 April 2001.

Brewer, E.A. 2001. When everything is searchable. Communications of the ACM 44(3): 54.

Cleverdon, C.W., Mills, J. and Keen, E.M. 1966. Factors determining the performance of indexing systems. Vol 1. Design, Part 1. Cranfield: Aslib Cranfield Research Project.

Cooper, M.D. 2001. Usage patterns of a web-based library catalog. Journal of the American Society for Information Science and Technology. 52(2): 137.

Cover, R. 2001. Electronic text corpus of Sumerian literature. [Online]. Available WWW: http://www.oasis-open.org/cover/etcsl.html. Accessed: 10 April 2001.

Cronje, J.C. and Clarke, P.A. 1999. Teaching "teaching on the Internet" on the Internet. South African Journal of Higher Education. 13(1): 213-226.

De Jager, K. and Sayed, Y. 1998. Aspects of information literacy at five institutions of higher education in the Western Cape. South African Journal of Higher Education. 12(2): 197-203.

Edling, R.J. 2000. Information technology in the classroom: experiences and recommendations. Campus-Wide Information Systems. 17(1): 10-15.

Ellis, D. 1996. Progress and problems in information retrieval. London: Library Association Publishing.

Frants, V.I., Shapiro, J., Taksa, I. and Voiskunskii, V.G. 1999. Boolean search: current state and perspectives. Journal of the American Society for Information Science. 50(1): 86-95.

Greek Libraries. 2001. Libraries. [Online]. Available WWW: http://www.libraries.gr/en_messop.htm. Accessed: 11 May 2001.

Hirsh, S.G. 1999. Children's relevance criteria and information seeking on electronic resources. Journal of the American Society for Information Science. 50(14): 1265-1283.

Ingwersen, P. 1992. Information retrieval interaction. London: Taylor Graham.

Jackson, E.B. 1971. The engineer as reluctant information user – a remedial plan. Proceedings of the International Conference on Training for Information Work. Rome. November 15-19: 430.

Jansen, B.J., and Pooch, U. 2001. A review of web searching studies and a framework for future research. Journal of the American Society for Information Science and Technology. 52(3): 244.

Lancaster, F.W. 1978. Information retrieval systems: characteristics, testing and evaluation. New York, NY: John Wiley.

Large, A., Tedd, L.A. and Hartley, R.J. 1999. Information seeking in the online age: principles and practice. London: Bowker-Saur.

Lynch, C.A. 2001. When documents deceive: trust and provenance as new factors for information retrieval in a tangled web. Journal of the American Society for Information Science and Technology. 52(1): 17.

Oppenheim, C., Morris, A., McKnight, C. and Lowley, S. 2000. The evaluation of www search engines. Journal of Documentation. 56(2): 190.

Sherman, C. 1999. The future of web search. Online. May/June: 54.

Spink, A. and others. 2001. Searching the web: the public and their queries. Journal of the American Society for Information Science and Technology. 52(3): 229.

Spink, A., Bateman, J. and Jansen, B.J. 1999. Searching the web: a survey of Excite users. Internet Research: Electronic Networking Applications and Policy. 9(2): 122, 125.

Su, L.T. 1993. Is relevance an adequate criterion for retrieval system evaluation: an empirical inquiry into the user's evaluation. Proceedings of the 56th Annual Meeting of the American Society for Information Science. 24-28 October. Columbus, OH: 93-103.

Sullivan, D. More About Meta Tags. [Online]. Available:http://searchenginewatch.com/subscribers/more/metatags.html. [Site visited on 01/07/2002].

Sutton, S.A. 1999. Conceptual design and development of a metadata framework for educational resources on the Internet. Journal of the American Society for Information Science. 50(13): 1191.

Taubes, G. 1995. Indexing the Internet. Science. 269: 1354-1356.

Voorbij, H.J. 1999. Searching scientific information on the Internet: a Dutch academic user survey. Journal of the American Society for Information Science 50(7): 604-605.

Wallace, R.M., Kupperman, J. and Krajcik, J. 2000. Science on the web: students online in a sixth-grade classroom. The Journal of the Learning Sciences. 9(1): 75.

Washington University. [Online]. Available WWW:http://depts.washington.edu/trio/comp/howto/site/design/metatags.shtml. Accessed: 1 July 2002.

Weideman, M. 2001. Search engines. [Online]. Available WWW: http://www.mwe.co.za/home/seaengin.htm. Accessed: 23 August 2001.

Weideman, M. 2002. Newspaper Articles. [Online]. Available WWW: http://www.mwe.co.za/home/seaarticles.htm. Accessed: 20 June 2002.

Wolfram, D. and Dimitroff, A. 1997. Preliminary findings on searcher performance and perceptions of performance in a hypertext bibliographic retrieval system. Journal of the American Society for Information Science. 48(12): 1145.