

AUTOMATED BENCHMARKING OF LOCAL GOVERNMENT WEB SITES: FULL PAPER INCLUDING DISCUSSION OF FINDINGS

Brian Kelly (*)

() UKOLN, University of Bath, Bath, UK
e-mail: B.Kelly@ukoln.ac.uk*

Abstract

This paper describes an automated approach to the benchmarking of Web sites. The author reviews the approaches which have been taken by UKOLN's WebWatch work to auditing and evaluating Web sites across a number of public sector communities, initially through use of locally-developed software, but more recently through use of freely-available Web services. The benchmarking surveys have provided useful information to members of the communities. However comparisons of results from use of different tools have revealed significant differences. The paper argues that there is a need for standard definitions in order to allow benchmarking results provided by different services to be comparable. The paper concludes by describing some of the technical limitations of the approach described and suggesting that a "Web Service" approach is needed.

Keywords: camera ready, instructions, conference proceedings

INTRODUCTION

Automated benchmarking of Web sites involves the use of software to analyse Web site characteristics across a community. Measurements of various aspects of the Web sites are carried out in order to establish compliance with agreed standards, guidelines and service level agreements. Benchmarking can also help to detect trends and the take-up of architectural frameworks, technologies, etc.

Web site benchmarking is likely to grow in importance as public sector organisations increase the range of services they provide and seek to maximise the coverage of their services. We are likely to see the development of compliance services which will measure compliance of public sector Web sites with national and international standards and guidelines – for example see the UK Government's e-GIF guidelines [1].

Manual benchmarking will be needed in areas in which automated techniques are not suitable. However the cost of manual benchmarking can be reduced by complementing such surveys with automated approaches. Although one might expect automated

benchmarking surveys to produce consistent and reproducible results, in practice this is not the case. This paper reviews a number of approaches to automated benchmarking and highlights areas in which inconsistencies may occur.

THE WEBWATCH PROJECT

The WebWatch project [2] was initially funded by the British Library Research and Innovation Centre (BLRIC). It was set up to:

- Develop robot software to gather information on usage of Web technologies within a number of communities within the UK.
- Use the software to collect the data.
- Develop (if appropriate) and use analysis tools to provide statistical analyses of the data.
- Produce reports explaining the analyses.
- Make recommendations to appropriate bodies on the information collected.
- Publicise reports to relevant communities.
- Make use of other tools to assist analyses in related areas, such as monitoring log files.

During its 18 month lifetime the project developed WebWatch auditing software, based on the Harvest software [3]. The software was used to analyse several communities including UK Public Library Web sites [4] and UK University and College entry points [5].

After the project funding ceased and the WebWatch software developer left it was decided that the WebWatch software should no longer be developed. However the feedback from the WebWatch reports had been very positive and it was felt that we should continue to carry out benchmarking surveys across its user communities.

WEBWATCH MARK II

By the late 1990s, a wide range of freely available Web sites were available which could provide a variety of services such as Web statistics, voting systems, etc. [6]. Several of these services provided various aspects of Web site benchmarking. It was decided to make use of these tools to carry out the benchmarking of our communities in WebWatch Mark II.

A wide range of Web-based auditing, analysis and benchmarking services are available such as **Dr HTML** [7], **NetMechanic** [8], **WebSiteGarage** [9], **Bobby** [10],

LinkPopularity [11] and **NetCraft** [12]. Typically these tools are implemented as CGI scripts with a Web interface. Many of the services are available free-of-charge, although some will require payment, typically an annual rental fee. Many of the services which charge for use will provide free access to a subset of their services.

Although use of these services was initially aimed at Webmasters and other individuals who wished to evaluate their own Web sites, many of them can be used indirectly, by using a URL string which provides a set of input parameters, such as the URL of the Web site to be analysed and control parameters for the service.

Figure 1 illustrates the Web interface to the NetMechanic service.

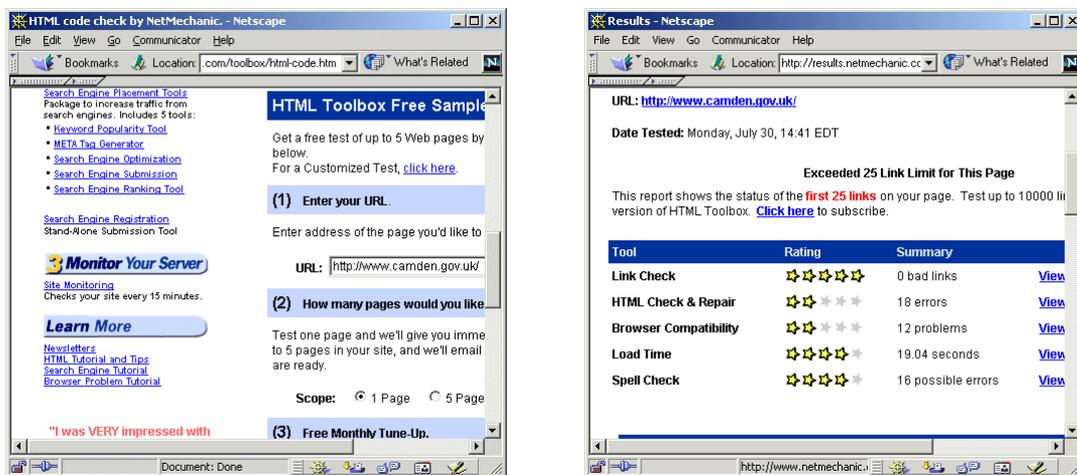


Figure 1: Use of the NetMechanic Service

The interface shown in Figure 1 is clearly intended for interactive use. However it is possible to use the input string as a HTML link which will allow the results to be reproduced without making use of NetMechanic's input screen. It is a simple task to change the URL string of the Web site to be analysed and use this approach to analyse a community of Web sites.

BENCHMARKING UK LOCAL AUTHORITY WEB SITES

The approach described above has been used to benchmark several communities, including UK University and Colleges and UK Public Library Web sites [13]. These surveys have made use of a number of Web services. It has been noticed that the results provided by different services are not always consistent. A smaller survey of a sample of UK Local Authority Web sites has been benchmarked in order to explore the reasons for the differences and as a pilot to establish if a more comprehensive survey of UK Local Authority Web sites would be useful.

The Sample

The sample used in this benchmarking study consisted of the shortlisted candidates for the 2001 SPIN-SOCITM Website Awards [14], as listed in Table 1.

Country	Local Authority	Web Site
England	L B Brent	http://www2.brent.gov.uk/
	L B Camden	http://www.camden.gov.uk/
	L B Richmond	http://www.richmond.gov.uk/
	Tameside MBC	http://www.tameside.gov.uk/
	Wokingham Council	http://www.wokingham.gov.uk/
Scotland	Dumfries & Galloway Council	http://www.dumgal.gov.uk/
	Dundee City Council	http://www.dundeeecity.gov.uk/
	East Renfrewshire Council	http://www.eastrenfrewshire.gov.uk/
	Moray Council	http://www.moray.gov.uk/
	West Lothian Council	http://www.westlothian.gov.uk/
Wales	Cardiff CC	http://www.cardiff.gov.uk/
	Ceredigion CC	http://www.ceredigion.gov.uk/
	Isle of Anglesey CC	http://www.anglesey.gov.uk/
	Wrexham CBC	http://www.wrexham.gov.uk/
Northern Ireland	Antrim BC	http://www.antrim.gov.uk/
	Armagh DC	http://www.armagh.gov.uk/
	Belfast City Council	http://www.belfastcity.gov.uk/
	Newtownabbey BC	http://www.newtownabbey.gov.uk/

Table 1: The Shortlisted Candidates For The 2001 SPIN-SOCITM Website Awards

This sample was selected as it was small enough to be manageable while covering a diverse geographical range and type of local authority. Since the sample Web sites had been selected as candidates for a national award, it was felt that the Web sites may make use of innovative features which may test the capabilities of the benchmarking tools. Finally since the shortlisted candidates would be likely to be regarded as examples of best practices which other local authority Web sites should seek to emulate, it was felt that the results of the benchmarking exercise should seek to see if this was the case.

The Tools

The Web sites listed in Table 2 were used.

Web Site	Measurements
NetMechanic	File size, link quality and HTML quality of entry points
Bobby	Accessibility, HTML quality and file size of entry points

Netcraft	Web server and platform and server uptime
LinkPopularity	Link popularity of Web sites

Table 2: Benchmarking Web Services Used

An automated visual display of the entry points, 404 error pages and `robots.txt` files was also produced.

These tools were used in order to provide the following functions:

NetMechanic:

Reports on the size of an organisation's entry point. This can be an indication of the download time. In addition examining the outliers may provide an indication of unusual approaches, such as intensive use of JavaScript, use of redirects, splash screens and proprietary plug-in technologies.

Bobby:

Reports on the accessibility of an organisation's entry point. Also reports on the file size, which can be compared with the figures provided by NetMechanic.

Netcraft:

Reports on the Web server platform and server software. This can be useful in observing trends. Significant trends may be useful in negotiations or identifying training needs. Also reports on the server uptime, which provides an indication of the reliability of the service.

LinkPopularity:

Reports on the number of links to a Web site, as measured by the AltaVista, Google and HotBot search engines. The number of links to a Web site may provide an indication of the popularity of the site.

Web Page Purifier:

Displays a Web site through a browser emulator. This can show if a Web site is not accessible to particular browsers.

Note that there are many desktop tools which can be used for benchmarking Web sites. Desktop tools, such as the **Bobby** application [15], **Xenu** link-checker [16], etc. may provide richer functionality than Web-based alternatives. However this paper focusses on Web-based services since these services can be used by everyone with access to a Web browser – software does not have to be installed locally. This ensures that the benchmarking process is more open, and allows everyone to see the tools and methodology employed, to try out the benchmarks for themselves, enabling the current results to be obtained and allowing Web sites not included in the initial benchmarking survey to be tested and the results compared. This transparency is likely to be important for public sector organisations.

REPORT ON FINDINGS

The detailed findings are available at [17]. A summary is given below.

Link Quality

Only one organisation appeared to have a broken link on its home page (although NetMechanic only analysed the first 25 links on each page).

HTML Quality

All entry points contained HTML errors, according to the NetMechanic analysis. The numbers of errors ranged from 1 to 59.

Accessibility Analysis

Of the 18 Web sites surveyed, 8 (44.44%) initially had no Priority 1 errors on the entry point according to Bobby (although one of these sites was found to have a Priority 1 error on a subsequent analysis). The highest number of Priority 1 errors was 4. All the Web sites had Priority 2 and 3 and browser compatibility errors.

File Size

The largest entry point according to the Bobby tool was 153.91 Kb. The largest according to NetMechanic was 127,578 b.

Links To Web Sites

The most linked to Web site is Dundee City Council (896 links according to AltaVista) or L B Brent (626 links according to Google). The least linked to Web site is Newtownabbey BC (88 links according to AltaVista and 53 links according to Google).

Server Analysis

Of the 18 Web sites surveyed according to Netcraft 12 (66.7%) were running on a Microsoft Windows platform, five (27.8%) were running on a Unix platform and one (5.6%) on an unknown platform.

The 12 (66.7%) organisations running on a Microsoft platform were also running a MS Windows Web server. Five (27.8%) were running an Apache Web server and one (5.6%) on a Lotus Domino Web server.

404 Error Page

Of the 18 Web sites surveyed, 12 (66.7%) provided the Web server default 404 error page, one (5.6%) provided a branded 404 error page and five (27.8%) provided a branded 404 error page which contained additional functionality, such as search facilities or a navigation bar.

What are the implications of the findings of this small survey? As might be expected, the numbers of **broken links** on the home pages is very low, with only one being found. Further analysis revealed that this link was included on a splash screen and would not

normally be noticeable by users. A recommendation from this observation would be that automated tools are needed to detect unusual examples of broken links (such as broken links on splash screens, 404 error pages) or links other than conventional hyperlink links (such as links to external style sheet files).

The failure of entry of the entry points to fully comply with **HTML standards** would appear to be worrying. Failure to comply with standards may result in the page not being usable by certain browsers or user agents. Further inspection of a number of the pages revealed that the problems were not necessarily a problem – for example, missing DTD declarations, specifying the colour of a HR element, etc. However even minor errors, which do not have a significant affect on the functionality of a page, will make it difficult to distinguish between these errors and more serious errors. Also, as XML becomes more widely used on Web sites through use of XHTML, script compliance with standards will become more important in order to enable resources to be transformed through use of the XSLT transformation language.

The **server** analysis indicates that the Windows NT is the most popular platform. Does this mean that services developed under Windows NT are more likely to be nominated for awards? Is this trend true across all local government Web sites? If so, can the community benefit from bulk purchases?

The **accessibility** analysis showed that only 8 (44.4%) of the entry points had no Priority 1 (P1) errors – and one of these sites had a P1 error the next day. If local government Web sites aim to conform with W3C WAI guidelines, these figures would indicate that much work still remains to be done.

The analysis of **links to the Web sites** does not show a great deal of variation across the sample, with the most-linked to site having 896 links to it and the least-linked to site having 88 links (according to AltaVista). An analysis of trends would be useful as this could indicate the effectiveness of an organisation publicity campaign for its Web site.

The analysis of **404 error pages** indicates that most Web sites use the server default error message. This default message fails to provide navigational or help information for users, and does not provide contain the branding of the Web site.

REPORT ON THE TOOLS

How valid are the results which have been obtained? How can differences in the results obtained from use of different services be explained?

Although all pages appeared to contain HTML errors further examination showed that this could be due to use of new HTML features, which were not defined in the assumed DTD.

An important difference between the tools was the inconsistencies over the analysis of a Web page. Some tools respected the Robot Exclusion Protocol [18] and would not download and measure files. Other tools would follow redirects whereas other would simply analyse the redirect headers or page. Similarly some tools would measure an initial “splash screen” whereas others would measure the final page. None of the Web services appeared to measure combination of pages.

This can be considered as uncertainties as to how benchmarking services should audit personalised or dynamic pages. Pages which make use of frames will normally download the content of the <NOFRAMES> tag to a robot (which, of course, may be useful content or may simply be a message suggesting that the user upgrades their browser). Similarly “cookies” can be used to deliver personalised content. A number of other techniques can be used to provide personalised content, such as the time of day, the page viewed previous (using the HTTP “referer” field), preferred language, etc. The contents of a page can also change through inclusion of news items, which may vary from day to day, or even more frequently.

It should also be noted that in general the tools do not make it easy for the numerical results to be easily processed. Bobby, for example, contains numbers of errors within a textual description and the numeric values are not highlighted from the text. NetMechanic provides an attractive display of the summary results. However additional information is sometimes hidden – such as the failure to analyse resources due to the Robots Exclusion Protocol. NetMechanic does not allow a persistent URL to be given to the detailed information, as this is stored on a temporary file: a URL can only be given to the summary information (which is reanalysed).

BENEFITS TO THE COMMUNITY

How useful is this type of benchmarking survey? A number of similar surveys have been carried out, such as Barry’s survey of Australian Library Web sites [19]. We are also finding that Web site benchmarking companies are being set up, including commercial services [20] [21] and community-based services [22] [23]. Automated benchmarking surveys would also be useful to those who are involved in judging Web sites and awarding prizes – it would be unfortunate if, for example, a prize was awarded to a government Web site which failed to comply with the government’s own guidelines!

Such benchmarking services are being set up in response to a clear need for providers and funders of Web services to obtain objective evaluations of their services. The need for such services is likely to grow as funding bodies, government agencies, etc. define Service Level Agreements which govern the level of services to be provided.

Many benchmarking surveys will include a subjective analysis. However manual analysis

of large numbers of Web sites will be very time-consuming and hence very costly. The automated approaches described in this paper can complement manual surveys, help to reduce costs and indicate areas in which more detailed manual analysis may be required.

However in order for automated surveys to be useful, and especially if they will be used to monitor compliance with standards and may affect the funding of services, it is important that standards are provided and guidance is provided to providers and users of benchmarking services.

RECOMMENDATIONS

The following recommendations are provided in order to help in the development of more rigorous and standardised benchmarking services.

A Need For Standards

There is a clear need for standard definitions to enable benchmarking surveys carried out using different applications and services to provide comparable results. For example there is a need to define the term “Web page”. The definition should cater for embedded resources (e.g. style sheets, JavaScript files, etc.), redirects, “splash screens”, user-agent negotiation, etc.

A definition should be provided by an international standards body, such as W3C [24]. It should be noted that W3C’s Web Characterisation Activity [25] is no longer active and so W3C would not appear to be an appropriate body. However W3C’s work on the development of EARL [26] - the Evaluation And Report Language (which aims to provide a framework for describing data about the evaluations of resources) may be of relevance.

Other bodies which may have a role to play in the standardisation work include the Interactive Advertising Bureau [27], ABCE [28] and JICWEBS [29] – bodies which represent the Web advertising and auditing communities.

Web Services

Once definitions have been standardised software vendors will be able to provide control over their analyses, in order to comply with the standards. The software vendors should also look into the mechanisms for providing control over their services, in order to provide modularisation, to facilitate change control and to allow results to be easily reproduced. There may be a need to move towards use of “Web Service” technologies such as SOAP [30]. The development of applications which are designed to be used by other applications will help to address some of the limitations of the approach describe in this paper, including change control and providing management control over access to services.

Commercial software vendors will, of course, make their decisions on whether to provide enhanced access to their services based on commercial considerations. However it could be argued that there is a need for freely available, open source services which will provide Web benchmarking facilities, along the lines of the W3C validation services [31] [32].

CONCLUSIONS

This paper argues that automated benchmarking of Web sites can provide a useful service for communities as well as complementing manual surveys by providing additional coverage at a low cost and identifying areas in which additional manual analysis may be needed. However use of a number of benchmarking services has shown that there are inconsistencies in the way they operate, which can result in inconsistencies and difficulties in reproducing results. The author argues that standardisation of terms is needed by the benchmarking community.

Agreement on standards is desirable before attempting a benchmarking survey across a large community, such as the Local Authority Web sites in the UK [32]. Despite these difficulties, the author would welcome collaboration in this area with interested people within the UK and overseas.

REFERENCES

- 1 *e-Government Interoperability Framework*, Office of the E-Envoy, UK Online, <<http://www.iagchampions.gov.uk/publications/frameworks/egif2/egif2.htm>>
- 2 *WebWatch*, UKOLN Web Site, <<http://www.ukoln.ac.uk/web-focus/webwatch/>>
- 3 *Harvest Web Indexing*, <<http://www.tardis.ed.ac.uk/harvest/>>
- 4 *Robot Seeks Public Library Websites*, LA Record, December 1997 Vol. 99(12) Also see <<http://www.ukoln.ac.uk/web-focus/webwatch/articles/la-record-dec1997/>>
- 5 *WebWatching UK Universities and Colleges*, B. Kelly, Ariadne, Issue 12, November 1997, <<http://www.ariadne.ac.uk/issue12/web-focus/>>
- 6 *Using Externally-Hosted Web Services*, B. Kelly, Ariadne, Issue 23, March 2000, <<http://www.ariadne.ac.uk/issue23/web-focus/>>
- 7 *Dr HTML* <<http://www.drhtml.com/>>
- 8 *NetMechanic* <<http://www.netmechanic.com/>>
- 9 *WebSiteGarage* <<http://www.websitegarage.com/>>
- 10 *Bobby*, CAST <<http://www.cast.org/bobby/>>

- 11 *LinkPopularity*, <<http://www.linkpopularity.com/>>
- 12 *Netcraft*, <<http://www.netcraft.com/>>
- 13 *WebWatch Articles*, B. Kelly, UKOLN Web Site, <<http://www.ukoln.ac.uk/web-focus/webwatch/articles/#latest>>
- 14 *EPI 2001: About The Conference: SPIN Awards*, SPIN Web Site, <<http://www.spin.org.uk/epi2001/about.html#spinawards>>
- 15 *Download Bobby*, CAST, <<http://www.cast.org/Bobby/DownloadBobby316.cfm>>
- 16 *Find broken links on web sites with Xenu's Link Sleuth*, <<http://home.snafu.de/tilman/xenulink.html>>
- 17 *Survey Of Shortlist for 2001 SPIN-SOCITM Website Awards*, UKOLN, <<http://www.ukoln.ac.uk/web-focus/events/conferences/euroweb-2001/survey/>>
- 18 *A Standard for Robot Exclusion*, <<http://info.webcrawler.com/mak/projects/robots/norobots.html>>
- 19 *The User Interface: glitz versus content*, A. Barry, October 2000 <<http://tony-barry.emu.id.au/pubs/2000/alia2000/>>
- 20 *Welcome to WebAudits*, WebAudits, <<http://www.webaudits.net/>>
- 21 *Audit Service Detail*, CleinWaterman, <http://www.cleinwaterman.com/services/cw_audit_service_detail.asp>
- 22 *Web Site Reviews*, HEIST, <<http://www.heist.co.uk/services/copywriting/websitereviews.htm>>
- 23 *Better Connected 2001*, MAPIT, SOCITM, <<http://www.socitm.gov.uk/mapp/w2001.htm>>
- 24 *The World Wide Web Consortium (W3C)*, <<http://www.w3.org/>>
- 25 *Web Characterization Activity*, W3C, <<http://www.w3.org/WCA>>
- 26 *EARL - the Evaluation And Report Language*, W3C, <<http://www.w3.org/2001/03/earl/>>
- 27 *Internet Advertising Bureau*, <<http://www.iab.net/>>
- 28 *ABC Electonic*, <<http://www.abce.org.uk/>>
- 29 *JICWEBS (The Joint Industry Committee for Web Standards in the UK and Ireland)*, <<http://www.jicwebs.org/>>
- 30 *SOAP Version 1.2*, W3C, <<http://www.w3.org/TR/soap12/>>
- 31 *HTML Validator*, W3C, <<http://validator.w3.org/>>
- 32 *W3C CSS Validation Service*, W3C, <<http://jigsaw.w3.org/css-validator/>>
- 33 *UK Local Government Websites*, Tagish, <<http://www.tag.co.uk/links/localgov.htm>>