Automatic Benchmarking and Presentation of the First Results from the European Internet Accessibility Observatory

Mikael SNAPRUD[1], Morten GOODWIN OLSEN[1], and Finn ASLAKSEN[2]
[1]*Faculty of Engineering, Agder University College, NO-4876, Norway*
Tel: +47 37253000, Fax: 47 37253001, Email: mikael.snaprud@hia.no
[2]*Vista Utredning AS, C.J. Hambros plass 5, 0164 Oslo, Norway,*
Tel: +47 22479880, Fax: 47 22479890 Email: finn.aslaksen@vistautredning.no

Abstract: Access to web content is the key to participate in the Information society. Barriers preventing access can often easily be removed or even avoided by raising the awareness among those who are building and purchasing web applications. This paper outlines an Open Source tool that will support large scale evaluation of public web content, to monitor the development, and to raise the awareness. Some early results from the automatic evaluations are provided to show the nature of the benchmarking.

1Introduction

Web accessibility benchmarking is carried out in many European countries to assess the accessibility status and to increase the general awareness. Different countries have different benchmarking methods and carry out the assessment with different frequency. Even if the evaluations are based on the same guidelines, accessibility evaluations are in practice often carried out in different ways. This is preventing international comparisons and systematic monitoring of this basic requirement for a democratic development of the information society.

A tool enabling frequent and automatic evaluations at a low cost could allow policy makers to monitor the development more closely to identify good practices, allow for regional comparisons, and assess the impact of policy measures.

The European Internet Accessibility Observatory[1] is a project designed to provide a prototype of such tool, namely an automatic large scale web evaluation service producing data on the accessibility status and development focusing on public content. The final version of the prototype Observatory will publish monthly updated measurements from 10.000 web sites. The results will be available online from a data warehouse to support flexible analysis, provide a basis for policy-making, research and actions to improve the accessibility to Internet content.

[1] The project is co-funded by the European Commission DG Information Society and Media, under the contract IST-004526.

The project was launched in September 2004 and has a duration of three years. The EIAO project is carried out in co-operation and partnership with industry, Open Source developers and users. It brings together the following institutions from across Europe:

Agder University College, Norway (Co-ordinating partner)
Vista Utredning AS, Norway
FTB-Volmarstein, Germany
Manchester Metropolitan University, UK
Nettkroken AS, Norway
University of Tromsø, Norway (withdrew from the project 12.05)
FBL, Italy
Technical University of Warsaw, Poland
Aalborg University, Denmark
Intermedium AS, Norway

The Web Content Accessibility Guidelines, developed by W3C have been adopted for public content by many national governments. Based on those guidelines three projects including 23 partners in a cluster[2] have developed the Unified Web Evaluation Methodology (UWEM). The' objective of UWEM is to provide means for ensuring that large scale monitoring and local evaluation are compatible and coherent among themselves and with the Web Content Accessibility Guidelines from W3C. While small scale, local evaluations can by performed manually, large scale evaluations require support of automated tools. UWEM should also support certification of web sites.

2Objectives of this paper

This paper will outline the properties and perspectives of large scale benchmarking based on the first release of the European Internet Accessibility Observatory. The outline will cover both the basic design of the software and a brief discussion of the nature and potential use of the benchmarking results.

A practical objective of the paper is to encourage national take-up measures of large scale accessibility benchmarking to support the efforts to improve access to public content. The paper gives a snapshot from cutting-edge research including preliminary results.

3Methodology used

The UWEM is based on the Web Content Accessibility Guidelines 1.0 (WCAG 1.0) from the World Wide Web Consortium (W3C) Web Accessibility Initiative (WAI). The methodology will be synchronised with the foreseen migration from WCAG 1.0 to WCAG 2.0 in the near future. Thus, UWEM will enhance the support for evaluating, certifying, and benchmarking web content. UWEM has been developed in a collaboration among 24 European organisations, called the WAB Cluster [2].

The Observatory developed within the EIAO project is implemented according to UWEM based on Open Source Software. The Open Source is essential to make sure that measurements of access to information – a prerequisite for any sound e-Government benchmarking – are carried out in a transparent and democratic way. Thus, allowing

[2] The EIAO project is carried out as part of the Web Accessibility Benchmarking cluster together with the projects Support-EAM and BenToWeb. See also http://www.wabcluster.org/

inspection of how the measurements are actually implemented, and encouraging a collaborative approach to improve them.

4Technology Description

The main elements of the first release of the Observatory are briefly discussed below. They consist of a crawler, a data warehouse, Web Accessibility Metrics, and a web user interface.

4.1Crawler

The key component in this architecture is the crawler, which is based on the Open Source web crawler HarvestMan [4]. HarvestMan was chosen because it is a good and mature web crawler that was easy to adapt to the system. Moreover, the HarvestMan developers were interested in actively supporting the EIAO project. HarvestMan has been integrated with the other components of the Observatory and is essential for the EIAO Observatory.

4.2Web Accessibility Metrics (WAMs)

A Web Accessibility Metric (WAM) is a formal rule specifying how to make a statement about accessibility barriers of a given Web resource.

A set of WAMs has been specified based on the tests described in UWEM0.5, checking for deviations from WCAG 1.0 checkpoints. The first set of WAMs is based on Relaxed/Schematron [6]. Furthermore, these address simpler forms of rules that can be expressed as Schematron rules. More complex rules may be implemented in Schematron (or in some other language) in a later release.

4.3Data warehouse

A data warehouse is used in the architecture to support storage of large amounts of data and support flexible analysis from an on-line user interface. We also expect the hierarchical design of the databased schema in a data warehouse to facilitate analysis of the data.

PostgreSQL was chosen since it is the only Open Source database with extensibility, table partitioning and bit-mapped indexes. Materialised view support is being built by the BizGres project. PostgreSQL performs very well for complex queries on large databases.

4.4Web user interface

The measurements stored in the data warehouse will be made available on the web. The data warehouse will contain the data from monthly measurements. In this way, trend analysis can be carried out, for example to assess the impact of some change in national implemented public procurement policy or accessibility legislation.

The EIAO Report Web Site will support reports for single web sites and NUTS regions (Nomenclature of territorial units for statistics) and NACE branches (Classification of Economic Activities in the European Community). These reports will be available for a given point-in-time as well as for a time period. In this way, comparisons of the status and development in e.g. local municipalities in different countries or regions will be supported. Furthermore accessibility scorecards are computed based on a set of criteria derived from the WCAG checkpoints.

Subsequent evaluations of the EIAO Report Web Site will guide future development of the functionality related to the detailed reporting. Based on feedback from end users (e.g. policy makers, associations of disabled people etc.), the tools for collecting, assessing and

disseminating data will be continuously improved. User testing will allow to continuously improve the relevance of the automatically collected data.

5Results from measurements

In this chapter we present some preliminary results from the first evaluation of five European Prime Minister web sites. The chosen sites represent the ministries in Germany, Norway, Spain, the Netherlands, and in France.

The measurement machinery is not yet optimised for speed and we currently need some more time to produce a more comprehensive measurement. The evaluation of one web site includes 75 tests on some two hundred web pages, and currently takes about 1-2 hours. Note that we currently only perform automatic testing, and thus only present results from tests that can be automatised. Although this sample is not large scale, the results should illustrate the kind of information that can be produced by the Observatory. This evaluation was carried out in May 2006.

Please note that one important limitation with this approach is that the Observatory will only report on tests that can be carried out automatically. The number of tests is a large improvement of some of the earlier work [7] which only focused on counting number of HTML deviations.

5.1Barriers detected

Table 1 shows the number of barriers, average number of barriers per page, and the percentage of evaluated elements with barriers present. The results are here presented as numbers where we count the number of violations of WCAG checkpoints.

Table 1: Number of barriers detected on five European Prime Minister web sites.

Web Site	***Number of barriers***	***Average number of barriers per page***	***Percentage of elements with barrier***
www.bundeskanzlerin.de	990	10	6 %
www.dep.no	24.324	118	27 %
www.la-moncloa.es	2.390	28	13 %
www.minaz.nl	2.752	29	42 %
www.premier-ministre.gouv.fr	3.098	17	5 %

We see from table 1 that www.dep.no has the highest number of barriers per page, while the www.minaz.nl has the highest percentage of elements with barrier. One reason for this may be that the pages seem to contain relatively many elements per page.

5.2Common barriers

Table 2 shows the top 10 barriers for the evaluated web sites along with the corresponding WCAG checkpoint priority and the description as provided from W3C.

Table 2: The most common barriers detected on five European Prime Minister web sites.

Number of barriers	*WCAG Checkpoint*	*WCAG level*	*Description*	*HTML Example*
19579	11.2	2	Avoid deprecated features of W3C technologies.	This uses deprecated elements.
6791	3.3	2	Use style sheets to control layout and presentation.	<body bgcolor="red"> This does not use stylesheets. </body>
2178	12.3	2	Divide large blocks of information into more manageable groups where natural and appropriate.	<table border="1" summary="table2"> <tr> <th>Table without caption</th> <th>A2</th> </tr> </table>
1635	10.1	2	Until user agents allow users to turn off spawned windows, do not cause pop-ups or other windows to appear and do not change the current window without informing the user.	Popup window.
502	1.1	1	Provide a text equivalent for every non-text element (e.g., via "alt", "longdesc", or in element content). This includes: images, graphical representations of text (including symbols), image map regions, animations (e.g., animated GIFs), applets and programmatic objects, ASCII art, frames, scripts, images used as list bullets, spacers, graphical buttons, sounds (played with or without user interaction), stand-alone audio files, audio tracks of video, and video.	<image src="./images/no_alternative_text_provided.gif"/>
305	4.3	3	Identify the primary natural language of a document.	<html xmlns="http://www.w3.org/1999/xhtml" lang="en">
292	10.4	3	Until user agents handle empty controls correctly, include default, place-holding characters in edit boxes and text areas.	No default value present in <input name="second" type="text" size="30" maxlength="40"/>
291	9.2	2	Ensure that any element that has its own interface can be operated in a device-independent manner.	You need a mouse to click here.
266	12.4	2	Associate labels explicitly with their controls.	<input type="text" name="test" value="No control present"/>

Number of barriers	*WCAG Checkpoint*	*WCAG level*	*Description*	*HTML Example*
248	3.4	2	Use relative rather than absolute units in markup language attribute values and style sheet property values.	<frameset cols="50%,50%" rows="100pt"> <noframes> Absolute values to define number of rows in a frame. </noframes> </frameset>

The two most common barriers detected in the evaluated web sites are the use of deprecated features and not using CSS for layout. The deprecated features from W3C also includes issues referring to changes where text attributes have been moved from HTML tags to CSS such as the use the font tag in HTML for specifying font size. Using the font tag in HTML may result in difficulties in overriding the font size in some browsers. This is clearly a barrier when users prefer/need a large font in the presented web pages. Furthermore, not using style sheets for layout may cause additional difficulties for users who prefer/need to have custom layout. This could be a challenge when users need a high colour contrast for reading web pages when the background colour is specified with HTML such as <body bgcolor=”red”>.

Misleading use of structural elements, which is common in the evaluated web sites, may result in a barrier for users who require a special browser such as a screen reader. For these web browser it is hard to distinguish between tables used for design and tables used for presenting data and such misleading use of tables may result in confusion. Additional problems occurs for users with screen readers when alternative text to non-text elements are missing, which is detected more than 500 times in the evaluated web sites.

Not providing a device independent interface, e.g. requiring a mouse to visit some links, makes it difficult for those users who have challenges with operating a mouse. This barrier has been detected almost 300 times in the evaluated web sites, clearly providing a barrier for users.

5.3Barriers with priority

Table 3 indicates the distribution of detected barriers based on the WCAG1.0 priorities.

Table 3: Distribution of barriers according to WCAG priority. of five European Prime Minister web sites.

Web Site	*Priority 1*	*Priority 2*	*Priority 3*
www.dep.no	507	23.409	408
www.minaz.nl	315	2.181	256
www.premier-ministre.gouv.fr	156	2.941	1
www.bundeskanzlerin.de	91	646	253
www.la-moncloa.es	54	2.189	147

We note that the majority of the barriers detected for all web sites are deviations from WCAG level 2 guidelines. We still need to analyse in more detail what this means for users and how to best improve the sites.

1.1Use of the results to encourage improvements of web sites accessibility

The results from the evaluations could be used to generate ranking lists of public web sites. Such lists seem to attract media attention and could therefore encourage the site owners responsible for poorly performing web sites to improve, and to learn from those who are scoring better. Awareness about accessibility issues is important both for procuring and maintaining public web sites. Using the data in this manor can both raise the general awareness and speed up the dissemination of best practice solutions.

6Current development challenges

In order to scale the Observatory to handle at least 10 000 monthly have several issues we need to address.

First of all, we need to collect URLs to10 000 public European web sites. These sites needs to be representative of each country and include prime minister sites, national banks, national libraries, sites from federal organisations such as president/monarchy, ministries, parliaments and national statistics agency. To decrease the amount of manual work, we plan to create a semi-automatic mechanism that automatically retrieves web site URLs from online national surveys etc.

Additionally, the performance of the Observatory itself needs to be drastically increased. We are currently using between 1-2 hours for downloading, evaluating and storing the results of each monitored web site. In order to reach the goal of evaluating 10 000 sites each month, we must in average use 3 minutes for the complete evaluation of each site, which is a drastically improvement.

Furthermore, we need to increase the precision of the results from our evaluations. Our Observatory uses a sampling strategy extracting random samples from each web site. Currently, we extract 100 samples from each evaluated web each site. For some sites, this strategy gives us a good statistical significance. However, for sites with quite diverse structure, the results does not have a good enough statistical significance. In order to solve this issue, we plan to implement a more adaptive sampling strategy. This means that we will extract samples from each site until a predefined statistical significance is reached in contrast to the current strategy of sampling until a predefined number of samples is reached. This would ensure that the results from all all sampled sites have a good precision compared to the results presented in this publication.

7Conclusions and Summary Recommendations

In this paper the first large scale web evaluation framework based on UWEM is outlined. The design is based on Open Source to allow inspection and continuous improvement of the evaluation software modules. This open design is essential to make sure that a machinery deployed for evaluation of core components of the e-Government implementation is indeed transparent. Choosing the Open Source model facilitates the sharing of results and skills, including external contributors and the sharing of resources on a larger scale. The project lead time and cost are significantly reduced by reuse of Open Source components.

Our early experiments show that the measurements can uncover information about the more frequent barriers encountered, indications about possible regional differences etc.

The access to the data will be free, similar to the access to popular search engines. In this way, we hope that the EIAO benchmarking can fuel a competition for improved

accessibility between countries, organisations, developers etc. It may also be used to provide background for political action.

For the near future, we will prepare a user-interface to allow online access to the data warehouse and address the bottlenecks to improve the performance of the Observatory software, to allow more efficient collection and analysis of larger amounts of data.

8Acknowledgements

Please note that this paper presents the results of a team effort with substantial contributions from 10 organisations. You find more details on the team the organisations and their contributions on the project web site at http://www.eiao.net [3].

9References

[1] Unified Web Evaluation Methodology version 0.5 http://www.wabcluster.org/uwem05/
[2] WAB Cluster, Web Accessibility Benchmarking http://www.wabcluster.org/
[3] European Internet Accessibility Observatory, 2006 - http://www.eiao.net
[4] Anand B. Pillai HarvestMan, 2006 - http://harvestman.freezope.org/
[5] Web Content Accessibility Guidelines version 1.0 - http://www.w3.org/TR/WAI-WEBCONTENT/
[6] Petr Nalevka Relaxed HTML validator, 2006 - http://badame.vse.cz/validator/
[7] M. Snaprud, N. Ulltveit-Moe, O-C. Granmo, M. Rafoshei-Klev, A. Wiklund, A. Sawicka. Quantitative assessment of public web sites Accessibility - Some early results - http://www.etsi.org/cce/proceedings/7_3.htm