A Framework for Quality Assurance for Data from Large Scale Accessibility Evaluations

Annika Nietzio[1], Nils Ulltveit-Moe[2], Terje Gjøsæter[2], Morten Goodwin Olsen[2], and Mikael Snaprud[2]

[1]Forschungsinstitut Technologie and Behinderung (FTB) der Evangelischen Stiftung Volmarstein, D-58300 Wetter (Ruhr), Germany
eiao@ftb-volmarstein.de
[2]Faculty of Engineering, Agder University College, NO-4876 Grimstad, Norway
{nils.ulltveit-moe, terje.gjosater, morten.g.olsen, mikael.snaprud}@hia.no

Abstract. In this extended abstract, we describe quality assurance issues for for large scale evaluation of accessibility to web sites. We present initial quality requirements in different areas of the development and give some examples of how they are to be addressed.

1 Introduction

In the Information Society, where a lot of information is made available on the web, it is essential to make this content accessible to all people, including people with special needs. To get an overview of the accessibility status of a large number of sites, manual evaluation by experts or disabled users will produce the most reliable results but this often turns out to be time-consuming and expensive. An automatic assessment of web accessibility is an alternative, even though it can not provide a complete accessibility overview. However, it can measure certain features that can be utilised as indicators for accessibility and allows the monitoring of a large number of web sites.

The EIAO project[1] has established the technical basis for a European Internet Accessibility Observatory (EIAO). An Internet crawler for automatic and frequent collection of data on web accessibility has been developed. The evaluation of these data is performed by a set of web accessibility metrics reporting accessibility problems and deviations from web standards according to the Unified Web Evaluation Methodology (UWEM) version 1.0,[2] based on WCAG 1.0.[3] A data warehouse provides on-line access to collected accessibility data via a graphical user interface (GUI). To ensure that the results produced by the Observatory are reliable, quality assurance has a high priority.

In this extended abstract, we discuss UWEM quality requirements for large scale automatic evaluations, and give some concrete examples of how they are addressed. The following section deals with the quality assurance of the underlying

---

[2] http://www.wabcluster.org/uwem1/
[3] http://www.w3.org/TR/WAI-WEBCONTENT/

methodology. Section 3 describes the development of requirements, and finally we have concluding remarks in section 4.

2 UWEM Quality Assurance

The Unified Web Evaluation Methodology (UWEM) is the basis for the EIAO Observatory. This methodology is the result of a joint harmonisation effort by 23 European organisations in three European projects combined in a cluster called the Web Accessibility Benchmarking Cluster (WAB cluster).[4]

This section provides some insights on how UWEM is developed. One central point is the issue tracking system of UWEM where all the problems and comments are collected. It contains the issues identified by external reviewers of the project, during implementations (e.g. by EIAO), by expert evaluators using UWEM, and from UWEM document review.

When talking about quality assurance, it is always useful to keep in mind the interests of the target audience, because different people often have different priorities regarding quality issues. In the case of UWEM, the main interest groups are policy makers and stakeholders, people with special needs and of course web developers.

The main quality demand of the policy maker and stakeholder group is that the data produced by the application of the methodology is reliable and that the resulting comparisons and rankings make sense. To meet this the UWEM sets forth the requirement of *unique interpretation*, i.e. the evaluation process must be unambiguous.

The evaluation is carried out to assess the conformance to the widely known and recognised criteria (the WCAG 1.0 guidelines developed by W3C/WAI).[5]

The credibility of UWEM accessibility evaluations to all target groups (especially to web developers) depends on the *repeatability* requirement. We now describe the quality assurance of the UWEM methodology exemplified by actions addressing the three requirements mentioned above.

2.1 Unique Interpretation

The unique interpretation of the test descriptions and results is crucial to ensure the usefulness of the methodology. If the objective or procedure of a test were ambiguous, it would be challenging to obtain the same results from different evaluations/implementations. Therefore it is required that the evaluation steps shall have only one way of being interpreted. The test procedures were developed in collaboration by several people. From a series of discussions, a common understanding of the objective of each test was reached. Furthermore, the tests were reviewed by independent web accessibility evaluators who were asked to apply the methodology and report back their findings about ambiguities or other issues preventing efficient use. These findings were one source of input to the issue tracking process.

---

[4] http://www.wabcluster.org
[5] http://www.w3.org/TR/WAI-WEBCONTENT/

2.2 Based on WAI Guidelines

The main goal of UWEM is to provide a harmonised European interpretation of the WCAG guidelines developed by WAI/W3C. This is important because in that way the methodology can replace the slightly diverging national implementations of the WAI guidelines and allow comparisons. This ambitious goal may be reachable, since UWEM is being developed as a joint effort between 23 organisations from across Europe collaborating in the WAB cluster.

2.3 Repeatability

The repeatability of the tests means that different evaluations of the same pages/sites should yield the same results (within a given tolerance). This requirement is verified by the detailed comparison of the evaluation results produced by several evaluations.[6] Any findings from this analysis are handled by the UWEM issue tracking process where open issues are brought together and addressed.

3 Development of requirements

When performing and presenting measurements of web accessibility, especially when the measurements are done automatically, it is essential that the methods for producing the data are statistically sound and thus the resulting data is reliable. With a special focus on this, we set up a series of possible requirements for the Observatory.[7] These requirements are to a large extent focused on the core part of the Observatory; sampling, aggregation and presentation. The requirements were gathered and carefully selected in order to get a well-grounded basis for our implementation. This was done in collaboration with both statistics and web accessibility experts.

After the initial collection of potential requirements, a detailed analysis of dependencies and possible incompatibilities was carried out, in order to get a good overview of this, mind maps where used. See Figure 1 for an example of a mind map. Thereafter, we carefully reviewed all requirements, selecting the most important ones. During this process, we took into the account what would be most desirable for the Observatory, in addition to keeping in mind what would be viable from an implementation point of view including performance and duration of the actual implementation.

The requirements provide input to both the EIAO implementation and the UWEM methodology.

[6] This comparison includes both expert and tool based evaluations.
[7] This was done during our second implementation in order to eliminate flaws in the Observatory.
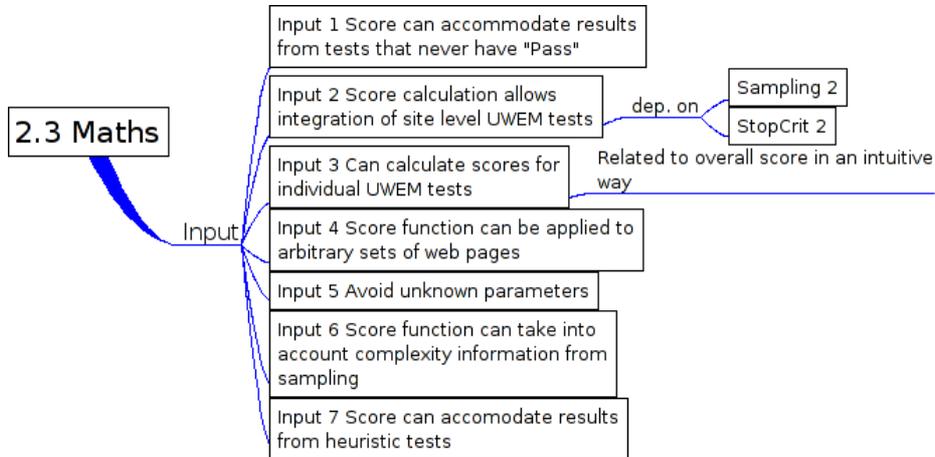
Figure 1: A mind map for mathematical requirements

After the selection of the requirements was stabilised, algorithms and functions for the different parts of the Observatory were designed. Later on, these will be the foundation for the actual implementations in the Observatory.

Even though we have carefully selected the requirements and algorithms, we need to quality assure all of these when the implementation is finalised. This is the case particularly for some of the requirements describing statistical properties of the result data. These requirements can only be verified on the basis of real data collected with a working implementation. It might be that the implementation leads us to discover that some parts of the designed algorithms show undesirable behaviour. In this case there will be a need to go back and adjust the algorithms, and maybe even the selection of requirements.
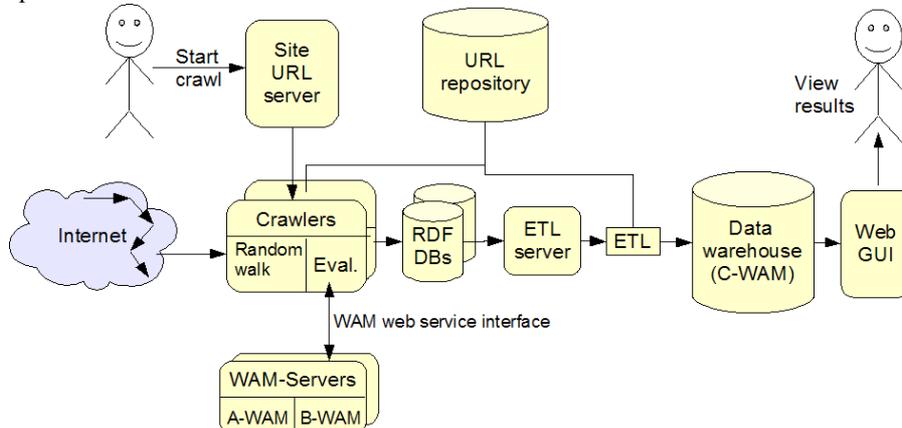
Figure 2: Observatory Architecture

The Observatory architecture, shown in Figure 2, is designed for easy substitution or modification of the components defining its functionality. For the crawling, it includes a customised version of the Open source web crawler HarvestMan.[8]

---

[8] http://harvestman.freezope.org/

HarvestMan has a modular design that supports easy development of plug-ins for different aspects of its functionality, allowing us to try out different sampling algorithms, such as the *random walk* algorithm used in the first releases of the EIAO Observatory. Sampling a set of web pages from each web site might be necessary since exhaustive crawling of each web site site and evaluation of all web pages might not be viable from a performance perspective.

The aggregation function and stopping rule for the sampling is defined in the *adaptive sampling* module. We can for example define that sampling from a site will stop when a certain error margin of the calculated web site score within a given confidence interval is reached. The required error margin and confidence interval can easily be adjusted. Furthermore, changing the aggregation function will only affect the adaptive sampling module and data warehouse.

It is noticeable from the illustration that in order to test e.g. a different sampling algorithm, only a small part of the Observatory needs to be updated. Components for handling URLs (URL repository, site URL server), storing data (RDF repositories, data warehouse), execution of the UWEM test (WAMs) and presentation in the GUI need no or very minor adjustments to support new sampling algorithms.

This framework allows us to experiment with different sampling algorithms, stopping rules and and aggregation functions, without having to rewrite large parts of the Observatory software. In that way we can find algorithms that are suitable to our purposes when it comes to performance and statistical properties. We may select the most promising algorithm candidates from the design phase, and try them out in order to analyse and compare the statistical properties of the produced data, and also the performance, before making the final decision.

4 Conclusion

In this extended abstract, we have discussed the EIAO approach to quality assurance, and highlighted some quality assurance issues related to large scale evaluation of web accessibility. We have outlined a strategy for making the resulting data from the Observatory reliable through a process of creating statistically sound methods used for the production of results.

We have further outlined the framework of the Observatory allowing us to experiment with different sampling / aggregation algorithms.

We believe that this approach will contribute to making the results from the EIAO Observatory reliable and representative.