Ranking websites through prioritized web accessibility barriers

Giorgio Brajnik
Dip. di Matematica e Informatica
Università di Udine, Italy
www.dimi.uniud.it/giorgio

September 29, 2006

1 Introduction

While accessibility can be tested with respect to guidelines (like WCAG 1.0, or Section 508) through a *standards review* method, other methods can be employed, like user testing (DRC, 2004) or usability inspection methods (Preece et al., 2002; Gray and Salzman, 1998; Nielsen, 1993) or those suggested by Henry and Grossnickle (2004).

To be really useful, evaluation methods should constrain the way in which the evaluator identifies problems and how they are graded in terms of importance. Only when these two kinds of decisions can be standardized, then the results produced can be used to rank web sites and to prioritize their bugs. Prioritization of defects is of paramount importance as any web developer required to fix them works always in a *scarce resource* mode.

This is not easy to achieve. As discussed by Brajnik (2006b) several definitions of accessibility exist: some refer to usability, some to effectiveness, and some to other abstract properties like perceivability, understandability and operability. The problem is that depending on the definition one chooses, different methods have to be used to investigate the website. Running a user test doesn't make sense if we want to determine conformance to guidelines, and conversely conformance testing cannot be used to determine usability of the web site with respect to disabled users.

Even simple methods like *conformance testing* (called also *standards review*) are not problem–free. As discussed by Kelly et al. (2005), WCAG 1.0 guidelines suffer from their theoretical nature, dependency on other guidelines, ambiguity, complexity, their closed nature and by some logical flaws. Most of these comments apply as well to the current WCAG 2.0 draft, to Section 508 and to the Italian official technical requirements. Guidelines are intentionally defined in terms that are independent from the technology used in implementing and in visiting web sites. As a consequence, guidelines are often too abstract to be directly applicable to a web site, creating a gap that has to be filled by the evaluator. In addition, they don't help the evaluator in distinguishing important problems from trivial ones. For example, few of the images that lack an appropriate alternative text are a true barrier: most of the images are used for emotional purposes, which in textual alternatives would almost never be achieved anyway. But an important function asked to an evaluator is to tell what the consequences of defects are: this, however, can be done only if appropriate use scenarios are considered. Priority levels attached to checkpoints (and hence to violations) are not an adequate solution. They may not reflect the impact that the defect has on users (like levels of success criteria in WCAG 2.0), or when they do it (like priority levels in WCAG 1.0) they do not depend on specific user goals.

This paper briefly describes the *barrier walkthrough* method, it highlights some of the experimental results regarding the method's usefulness and validity, and it shows how the method can be coupled with automatic testing tools to compare and rank websites according to their accessibility.

2 The barrier walkthrough method

To apply the barrier walkthrough method (Brajnik, 2006b,c) an evaluator has to consider a number of predefined barriers which are interpretations and extensions of well known accessibility principles; they are linked to user characteristics, user activities, and situation patterns so that appropriate conclusions about user effectiveness, productivity, satisfaction and safety can be drawn, and appropriate severity scores can be consequently derived.

Application of the barrier walkthrough method assumes that accessibility is defined as: "*... web sites are accessible when individuals with disabilities can access and use them as effectively as people who don't have disabilities*" (Slatin and Rush, 2003, p. 3).

The method is rooted on usability investigation practices which are based on context of use of the website. Context comprises certain user categories (like blind persons), usage scenarios (like using a given screen reader), and user goals (corresponding to *use cases*, like submitting an IRS form).

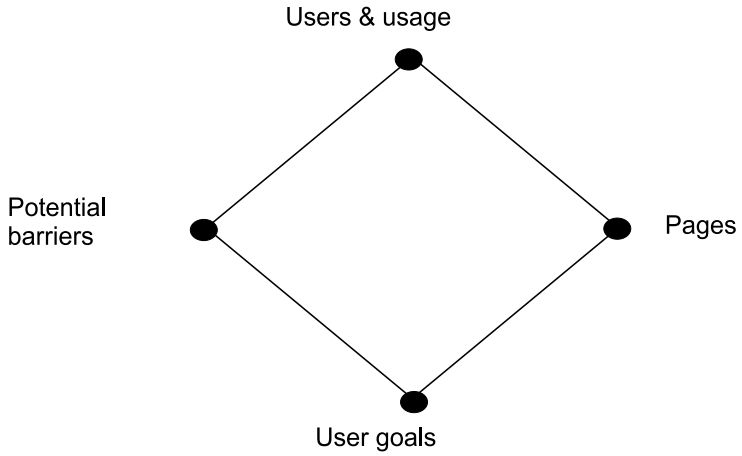

Figure 1: The ingredients of the barrier walkthrough method

An *accessibility barrier* is any condition that makes it difficult for people to achieve a goal when using the web site through specified assistive technology (see figure 2 for an example). A barrier is a failure mode of the web site, described in terms of (i) the user category involved, (ii) the type of assistive technology being used, (iii) the goal that is being hindered, (iv) the features of the pages that raise the barrier, and (v) further effects of the barrier.

Notice that several barriers can depend on the same cause: *e.g.* for a missing *skip–links* link (defect) a barrier for a blind user of a screen reader is that s/he cannot get quickly to the relevant content of the page; the barrier for a keyboard user is that s/he cannot move the focus directly to the relevant controls in the page; the barrier for a low vision person is that s/he cannot click on a link to move directly the field of vision on the relevant content.

Severity of a barrier depends on the context of the analysis (type of user, usage scenario, user goal). The barrier walkthrough method prescribes that severity is graded on a 1–2–3 scale, and is a function of *impact* (the degree to which the user goal cannot be achieved within the considered context) and

frequency (the number of times the barrier shows up while a user is trying to achieve that goal). Therefore the same barrier may have different severities in different contexts; for example, the missing *skip–links* link may turn out to be a nuisance for a blind user reading a page that has few preliminary stuff, while the same defect may show a high severity within a page that does a server refresh whenever the user interacts with links or select boxes.

Potential barriers to be considered are derived by interpretation of relevant guidelines and principles (DRC, 2004; W3C/WAI, 2004). A complete list can be found in (Brajnik, 2006c).

barrier	users cannot perceive nor understand the information conveyed by an information rich image (*e.g.* a diagram, a histogram)
defect	an image that does not have accompanying text (as an ALT attribute, content of the OBJECT tag, as running text close to the picture or as a linked separate page)
users affected	blind users of screen readers, users of small devices
consequences	users try to look around for more explanations, they spend substantial time and effort; effectiveness, productivity, satisfaction are severely affected

Figure 2: Example of barrier

Since barriers are closely related to well known guidelines, it is fairly easy to derive a conformance claim from a barrier walkthrough report. In addition, as discussed below, it is also fairly easy to deploy automated tools (like LIFT, Bobby, WebExact, ...), whose output is structured on checkpoints/guidelines, and use the output to determine where potential barriers occur.

3 Validity of the method

Evaluation methods should be valid, reliable, useful and efficient. *Validity* is "the extent to which the problems detected during an evaluation are also those that show up during real–world use of the system" whereas *reliability* is "the extent to which independent evaluations produce the same result" (Gray and Salzman, 1998; Hertzum and Jacobsen, 2001). *Usefulness* is the

effectiveness and usability of the results produced (with respect to users that have to assess, or to fix, or otherwise to manage accessibility of a web site). Finally, *efficiency* is given by the resources being utilized during an evaluation (in terms of time, persons, skill level, facilities, money) related to some level of validity, reliability and/or usefulness to be achieved.

For example, automated tests score high in reliability, high in efficiency, might be good in usefulness (depending on the quality of the data and of their presentation), but they score low in validity.

In (Brajnik, 2006b) I analyze validity and usefulness of the barrier walkthrough method compared to conformance testing. A number of accessibility reports produced by students (beginner accessibility evaluators) were analyzed, judged and compared, producing data on how many true accessibility defects were highlighted, and their severity scores.

The conclusion is that barrier walkthrough is more useful and more valid than conformance testing because in almost all of the cases barrier walkthrough is capable of highlighting a significantly larger proportion of true severe defects.

4 Ranking websites using barriers

Deployment of automated accessibility testing tools for ranking websites according to the their degree of accessibility is still full of open questions (Brajnik, 2006a). One of the hurdles to be overcome is the ability to compare two websites (or two versions of the same website, or two pages) in a meaningful way. Drawing simple statistics from numbers of *potentially valid* checkpoint violations falls short for three reasons.

First, it mixes violations of different checkpoints so that the resulting score in no way can tell us anything about how difficult it will be for a disabled person to use the website. For example, if website A shows 32 violations/page of AA WCAG checkpoints while B shows 25 violations/page, such numbers cannot be used to draw any conclusions regarding a the user experience of a blind person since not all the violations will have similar negative impact on users.

Second, those numbers don't reflect the context of usage and user goals.

Third, several of the checkpoints reported to be violated may turn out to be *false positives* (*i.e.* not real problems, due to shortcomings of the testing tools). But the error rate of these tools depends highly on tested web pages

website	pages	pot/page	valid	barriers/page
website A	50	34	[.55, .62]	[19, 21]
website B	55	28	[.36, .53]	[10, 14]

website	pages	pot/page	valid	barriers/page
website A	50	14	[.35, .58]	[5, 8]
website B	55	18	[.33, .61]	[6, 11]

Table 1: Comparison of two websites with respect to estimated barriers for blind persons (top) and motor disabled persons (bottom). *pot/page* is the number of checkpoint violations per page that the tool found; *valid* is the 5% confidence interval of the proportion of true barriers with severity 2 or 3 among the violations found by the tool; and finally *barriers/page* is the confidence interval of the estimated number of true barriers per page.

and on tested checkpoints. Therefore 32 vs. 25 does not mean anything unless we can characterize the error rate.

The following procedure can be followed to measure accessibility on a website. A testing tool can be used to collect checkpoint violations on a website. Each violation can then be mapped to one or few potential barriers (as specified by barrier walkthrough). By sampling a subset of the violations, and then asking a judge to determine severity of the barriers associated to the violated checkpoints, we can draw some quantitative and statistically well founded conclusion on the error rate *for that barrier*. At this point, by simply aggregating such data, we can draw some conclusion on whether website A is more accessible than B with respect to blind persons, or elderly people, or any other meaningful grouping of barriers we can think of.

A limited experiment performed with LIFT on two Italian websites lead to results shown in table 1 for barriers related to blind persons and motor disabled ones.

Results show that B is better than A with respect to blind persons (with 5% chances of being wrong, the number of true barriers per page in A is between 19 and 21, whereas in B it is between 10 and 14. We could say that with respect to blind users of screen readers, A is significantly worse than B (at least 40% worse than B).

However, for the motor disabled barriers, although apparently the two

averages (14 and 18) suggest that A is better than B, after considering the error rate of the tool, no difference appears to be meaningful (since the two intervals overlap, more than 5% of the times the two numbers of true barriers/page will change their relative ordering).

5 Conclusions

A careful coupling of the barrier walkthrough method with accessibility testing tools and human judging of a sample of issues can yield useful results to be used to rank, or at least to compare, two websites in a meaningful way.

This is a consequence of:

- the concrete way in which barriers are defined in the barrier walkthrough method;
- the context (user type, usage scenario, user goals) that is considered by the evaluator;
- the way in which impact and frequency of the barrier are determined and summarized in the barrier severity.

References

G. Brajnik. Automatic monitoring of web accessibility: the validity issues. In M. Snaprud, editor, *Workshop on Web Accessibility and Metamodelling*, Agder University, Grimstad, Norway, June 2006a. European Internet Accessibility Observatory. `http://www.eiao.net/wwam06`.

G. Brajnik. Web accessibility testing: When the method is the culprit. In K. Miesenberger, J. Klaus, W. Zagler, and A. Karshmer, editors, *ICCHP 2006, 10th International Conference on Computers Helping People with Special Needs*, Lecture Notes in Computer Science 4061, pages 156–163, Linz, Austria, July 2006b. Springer Verlag. `http://www.dimi.uniud.it/giorgio/publications.html#linz06`.

Giorgio Brajnik. Web accessibility testing with barriers walkthrough. `www.dimi.uniud.it/giorgio/projects/bw`, March 2006c. Visited Mar. 2006.

DRC. Formal investigation report: web accessibility. Disability Rights Commission, www.drc-gb.org/publicationsandreports/report.asp, April 2004. Visited Jan. 2006.

W.D. Gray and M.C. Salzman. Damaged merchandise: a review of experiments that compare usability evaluation methods. *Human–Computer Interaction*, 13(3):203–261, 1998.

S.L. Henry and M. Grossnickle. *Accessibility in the User–Centered Design Process*. Georgia Tech Research Corporation, Atlanta, Georgia, USA, 2004. On–line book, www.UIAccess.com/AccessUCD.

M. Hertzum and N.E. Jacobsen. The evaluator effect: a chilling fact about usability evaluation methods. *Int. Journal of Human–Computer Interaction*, 1(4):421–443, 2001.

B. Kelly, D. Sloan, L. Phipps, H. Petrie, and F. Hamilton. Forcing standardization or accomodating diversity? A framework for applying the WCAG in the real world. In S. Harper, Y. Yesilada, and C. Goble, editors, *Int. Cross–Disciplinary Workshop on Web Accessibility, W4A*, pages 46–54, Chiba, Japan, April 2005. ACM.

Jakob Nielsen. *Usability engineering*. Academic Press, Boston, MA, 1993.

J. Preece, Y. Rogers, and H. Sharp. *Interaction design*. John Wiley and Sons, 2002.

John Slatin and Sharron Rush. *Maximum Accessibility: Making Your Web Site More Usable for Everyone*. Addison-Wesley, 2003.

W3C/WAI. How people with disabilities use the web. World Wide Web Consortium — Web Accessibility Initiative, w3.org/WAI/EO/Drafts/PWD-Use-Web/20040302.html, March 2004.