Hi all,

I added my comments to the discussion below marked with EV:

On 2011-09-12, at 4:42 AM, Kerstin Probiesch wrote:

 \* Requirements:

R01: Technical conformance to existing Web Accessibility Initiative

 (WAI) Recommendations and Techniques documents.

Comment (RW): I do not think we need the word technical. We should stick with WCAG as agreed when we discussed \*A01. The recommendations and techniques are not relevant here as our priority is the Guidelines.

It is possible for someone to comply with a particular guideline without using any of the recommended techniques. What we are after is methodology. I therefore suggest a suitable alternative as follows:

\*R01 Define methods for evaluating compliance with the accessibility guidelines (WCAG)

Comment (KP): As I understood R01 it stresses the formal level. If the formulation would be "R01: Technical conformance to existing Web Accessibility Initiative (WAI) Recommendations and Techniques" I would agree. Because we have in the WCAG sub-documents like "understanding", "glossary" and so on. For that "documents" for me is ok. Because of other WAI documents like e.g. ATAG I would agree with.

As long as the formal level of the documents itself and not the techniques which are in the documents is meant.

Comment (DB): I believe we need to stay on a macro level, as we're talking general methodology here. We'll have plenty of time to delve right in eventually. Right now, our main focus should be compliance with the WCAG as a whole, not to each and every technique that may or may not exist at the time of this writing. Just to build up on Richard's proposal, I would therefore suggest:

\*R01 Defining methods for evaluating WCAG 2.0 compliance

WCAG will most already have been defined in this document, so there's no need to repeat it each and every time.

Comments (VC): agree

Comment (KP): Please would someone (Shadi, Eric) give a short statement, weather we speak about form or content?

EV: Agree with R01 defining methods for evaluating WCAG 2.0 compliance. This idea behind this requirement is maybe better worded in R06 (The methodology points to the existing WCAG documents and does not reproduce them.) The idea is not that we make our own tests in this Methodology. Although in a later stage and when testing, we could ask WCAG WG to add or change tests or make changes to the format of the tests, but that is not a requirement of the Methodology.

R02: Tool and browser independent

Comment (RW): The principle is good but sometimes it may be necessary to use a particular tool such as a text-only browser. So I would prefer:

\*R02: Where possible the evaluation process should be tool and browser independent.

Comment (KP): I partly agree with "possible". When we use "possible" we should then describe/define what "possible" exactly means.

Comment (DB): Right. That works for me too. But I'd rather keep them short. So I'd vote for:

Comments (VC): Agree also

\*R02: Tool and browser independent (where possible)

Comment (KP): I'm ok with that we will find a definition of "possible".

EV: The requirement does not mean that you cannot use tools or browsers, but that you have a choice. If we limit this, then we should indeed define 'where possible'. I will use the extension 'where possible' in the next version.

R03: Unique interpretation

Comment (RW): I think this means that it should be unambiguous, that means it is not open to different interpretations. I am pretty sure that the W3C has a standard clause it uses to cover this point when building standards etc. Hopefully Shadi can find it <Grin>. This also implies use of standard terminology which we should be looking at as soon as possible so that terms like “atomic testing” do not creep into our procedures without clear /agreed definitions.

Comment (KP): Using standard terminology is an important point also for me. And I suggest that we should also regard the standard terminology used I testing theory. The advantage would be that we are using established terms which will help to avoid misunderstandings.

Comment (DB): Using standard terminology is of outmost importance to me as well. However, I personally do not believe in a single interpretation for any success criteria. And I certainly do not believe in the possibility of everyone under the same interpretation. What we should focus on is achieved results (compliance), not how people actually got there (technique used). There are multiple ways to interpret the guidelines and our methodology should reflect this. Instead of striving for "unique interpretation" I would much rather go for "agreed interpretation", even if this means actually building a document where we would document what those divergent interpretations mean. So, I suggest going with:

\*R03: Agreed interpretations

Comment (KP): I understand the Denis' arguments. The more I think about this: neither "unique interpretation" nor "agreed interpretation" work very well. I would like to suggest "Objective". Because of the following reason:

It would be one of Criteria for the quality of tests and includes Execution objectivity, Analysis objectivity and Interpretation objectivity. If we will have in some cases 100% percent fine, if not we can discuss the "tolerance".

I would suggest:

(VC) I'm still contemplating this one. I can see both arguments as plausible. I'm okay with 'objectivity' but think it needs more explanation i.e. who defines how objective it is?

\*R03: Objectivity

EV: This requirement is meant for the Methodology and not so much for the reports. The Methodology itself should be uniquely interpretable. It does not say that WCAG or any other documents used in evaluations are uniquely interpretable. We can comment on that later during the writing and testing of the Methodology if necessary. My proposal would be to keep this one as it is: R03 Unique Interpretation.

R04: Replicability: different Web accessibility evaluators who perform the same tests on the same site should get the same results within a given tolerance.

Comment (RW): The first part is good, but I am not happy with introducing “tolerance” at this stage. I think we should be clear that we are after consistent, replicable tests. I think we should add separate requirement later for such things as “partial compliance” and “tolerance. See R14 below.

\*R04: Replicability: different Web accessibility evaluators who perform the same tests on the same site should get the same results.

Comment (KP): I strongly agree with Richard. Except "Replicability" and would suggest:

R04: Reliability: different Web accessibility evaluators who perform the same tests on the same site should get the same results.

Comment (DB): As long as we take into consideration that there can be different ways/tools to run those tests, then yes, reliability and replicability are important. Getting to different results usually means evaluators do not interpret the rules the same way. This does not always mean that one is wrong and the other is right. So again, to keep those short, I would simply go with

\*R04: Reliable and replicable

Comment (KP): I'm happy with that, as long as it will not include a decision for or against any specific evaluation methodology which one of the participants in this TF uses.

The explanation that follows could then reflect the idea that different evaluators performing the same tests on the same site should get the same results.

(VC) +1 for Kerstin's i.e. reliable and replicable: understanding that different evaluators using the same methods on the same sites should obtain the same results

EV: As opposed to R03, this one is about the reporting. Agree with the change: R04: Reliable and replicable. Different evaluators performing the same tests on the same site should obtain the same results.

R05: Translatable

Comment (RW): As in translatable into different languages – Yes -agree

Comment (KP): I agree and I see especially translatable in the context of using standard terminology which would be helpful for translating.

Comment (DB): +1.

Comments (VC): +1

EV: +1

##  R06: The methodology points to the existing tests in the techniques documents and does not reproduce them.

Comment (KP): I agree.

Comment (RW): yes – but I would like it a bit clearer that it is WCAG techniques. I would also like the option to introduce a new technique if it becomes available. So I suggest

\*R06 Where possible the methodology should point to existing tests and techniques in the WCAG documentation.

Comments (DB): I agree with the general idea here as well, but it needs to be shorter. We can always reflect the intention in the description that follows.

Comments (VC): +1

\*R06 Pointing to existing tests and techniques (where possible).

EV: I would propose to add WCAG and leave out 'where possible'. We do not want to point to other tests and techniques. New wording would then be:

R06 Points to existing WCAG tests and techniques.

R07: Support for both manual and automated evaluation.

Comment (RW): Not all Guidelines can be tested automatically and it is not viable to test some others manually. This needs to be clearer that the most appropriate methods will be used, whether manual or automatic. Where both options are available they must deliver the same result.

\*R07: Use the most appropriate manual or automatic evaluation.

Where either could be used then both must deliver the same result.

Comment (KP): I see "support" as just support and the important point "deliver the same result" in the context of R04 "Replicability" or as I suggest "Reliability".

Comments (DB): I agree with the general idea here as well, but again, it needs to be shorter. We can always reflect the importance of using the most appropriate approach in the document itself.

\*R07: Reliable evaluation support (manual or automated).

Comment (KP): I'm still not sure about what "support" means.

(VC) I like:\*R07: Use the most appropriate manual or automatic evaluation.

 Where either could be used then both must deliver the same result.

EV: Not completely agree. With this requirement I mean that all three ways of evaluation would be supported by the Methodology. So the Methodology would support:

- automated evaluations only

- manual evaluations only and

- automated plus manual evaluations.

I think this is important because some people will only use automated evaluations and want to claim things about them. We should give that a place. This is the reason for the saying: R07 Support for both manual and automated evaluation. I would propose to keep the text as it is..?

R08: Users include (see target audience)

Comment (RW): Whilst user testing is essential for confirming accessibility it is not needed/essential for checking compliance with WCAG. If we feel that user testing is needed then we must specify what users, what skill level, what tasks etc…so that evaluators all use the same type of user and get the same type of result. I would prefer not to include users here as a requirement.

Comment (KP): A tricky R. - especially in the context of the above mentioned "It is possible for someone to comply with a particular guideline without using any of the recommended techniques." The question would be: How a tester can find out if an SC is met when the recommended techniques are not used? Wouldn't that mean that a tester needs deep knowledge in using for example Screenreaders as well as Magnifiers and ... We discussed this also in an another mail thread. I prefer to include users here but we have to describe what users according to Richard’s consideration in the above paragraph.

Comments (DB): Testing with "real uses" should be encouraged, but in no way should it be made mandatory. The only requirement should be to run tests with a skilled screen reader user, following a specific evaluation methodology. All the better if this evaluator happens to be a real user. So:

\*R08: Users include (see target audience)

(VC)+1. However I'm not sure about putting the screenreader as mandatory. Not everyone is proficient with screen-readers and in my experience using one poorly often gives a very unreliable result.

EV: In my opinion (disabled) users of websites as such would not be target audience for the Methodology. They could off course be evaluators, but then they have a different role. This requirement does not include user testing. My proposal would be not to include user testing into the Methodology besides stressing the importance of doing so.

R09: Support for different contexts (i.e. self-assessment, third-party evaluation of small or larger websites).

Comment (RW): Agreed.

Comment (KP): Agree

Comments (DB): +1.

Comments (VC): +1

EV: ok +1

## R10: Includes recommendations for sampling web pages and for expressing the scope of a conformance claim

Comment (RW): I agree. This is probably going to be the most difficult issue, but it is essential if our methodology is going to be useable in the real world as illustrated by discussions already taking place. Should it include tolerance metrics (R14)?

Comment (KP): I also think it’s the most difficult issue. Because of the ongoing discussion about different approaches I want to abstain for the moment.

Comments (DB): While I seem to be a little more optimistic than you two, it is an important issue. I wish that we can draw from everybody's experience and come up with something new and improved, compared to our respective approaches.

Comment (VC): This is going to be difficult. However I think we will be able to come up with something workable when we all discuss how we handle this. I think we'll all learn some new techniques here.

\*R10: Web pages sampling recommendations.

Comment (KP): I still want to abstain before making a decision. But I think we need the "scope of conformance claim". In find this a very important issue for the validity of the evaluation methodology.

We can always reflect the importance of expressing the scope of conformance claim in the document itself.

EV: Yes, this will be the most difficult thing to write into the Methodology. Also in relation to accessibility supported.

For sampling and scope I think we will not need tolerance metrics. But we will have to discuss how to determine a good sample. There are many different approaches to that in the World. I would keep the text as it is.

R11: Describes critical path analyses,

Comment (RW): I assume this is the CPA of the evaluation process (ie define website, test this, test that, write report etc.). In which case agreed

Comment (KP): I'm not sure what is meant by this R. Because of that no vote from me now.

Comments (DB): Agreed as well. This is something we never officially did ourselves at AccessibiltéWeb, but it does look like a great idea.

Comments (VC): not sure on this one

EV: The idea is that there we include Critical Paths on a website. In our call it was also called Key Scenario's of a website. It means that you take all pages necessary to fulfill a proces on a website. Like on a shopping website, you should be able to buy a product. This would mean that when sampling pages, all pages in this Scenario would have to be part of the sample.

The Methodology as a whole will describe how to go through the process of testing. But we did not describe that as a seperate requirement. Should we?

## R12: Covers computer assisted content selection and manual content selection

Comment (RW): I do not know what this means – can Eric explain?

Comment (KP): I also don't have a exactly idea what this R. could mean.

Comments (DB): Isn't this directly related to page sampling determination and critical paths analyses? I get the manual content selection part, but I can't understand how this could be computer generated in any way... right now, I don't see why this couldn't just be a part of R11.

Comments (VC) this does sound like R11, I think we need some clarification here.

EV: Sorry for the blur here. The idea is that we could also include saying something about selecting samples by computer only or manually only and combinations. Some of us use tools to select pages from a website for evaluation.

R13: Includes integration and aggregation of the evaluation results and related conformance statements.

Comment (RW): I think this means “write a nice report” in which case I agree.

Comment (KP): I agree.

Comments (DB): Lol, reports are crucial indeed and every report should be technically-biaised, with a good executive summary for the faint-hearted. But this R is definitely too complicated as is.

\*R13: Evaluation reports and related conformance statements.

Comments (VC): agree

EV: +1 Aggregation could include guidelines on how to aggregate the results into a final report including statistical metrics.

R14: Includes tolerance metrics.

Comment (RW): Agreed – but maybe combine with R10

Comment (KP): The tolerance metrics will depend on the testing procedure itself. Because of that for me I'm happy with that and suggest not to combine with any other R.

Comments (DB): I can see why it could be integrated with R10, but don't really mind if it's not. I think the wording is appropriate.

Comments (VC): agree

EV: The idea is that we state when a failure is really such that the Success criterium is regarded as failed. We could say that there is a tolerance of 5% error: so if 4% of all descriptions of images is a fail, the SC would still be regarded as ok in the report... (just brainstorming here) except when .... (there is a difference between decorative and navigational images). We could introduce different sorts of fails. More discussion would be necessary.

R15: The Methodology includes recommendations for harmonized(machine-readable) reporting.

Comment (RW): I am not sure that methodologies recommend things. Doyou mean

\*R15: Reports must be machine readable.

Comment (KP): As I understood R15 this means e.g. structures in documents but also recommendations for the content structure. If so, I agree with R15.

Comments (DB): Shouldn't this be a part of R13 as well?

\*R15: Recommendations for harmonized (machine-readable) reporting

Comment (KP): I agree.

Comments (VC): isn't this what EARL is all about? I'm not sure on this one.

EV: Yes, the idea is that we also provide a format for machine readable reports. We could provide different formats for reporting including manual reporting etc. But minimally a format for machine readable reports using EARL

Eric Velleman