# Designing a belief function-based accessibility indicator to draw disabled people to the most adapted Web pages

JEAN-CHRISTOPHE DUBOIS[1], YOLANDE LE GALL[2],
ARNAUD MARTIN[3]

The purpose of this study is to provide an accessibility measure of Web pages, in order to draw disabled users to the most adapted Web pages. Our approach is based on the bhe rd[(ad)-7(ap)-7(tee31-5(f)8( )-50(t)-H3,Tm[(b)-10(h)6(e )o7( rd)-17)-3198-17pn-4( ) 545.71 Tm[(

ommendations of standards, using automatic evaluation tools. They often give a final value, continuous or discrete, to represent content accessibility. However, the fact remains that tests on accessibility criteria are far from being trivial [4]. Evaluation reports of automatic assessors contain errors considered as certain, but also warnings or potential problems which are uncertain. Moreover there are differences between assessor evaluations, even for errors considered as certain.

This work provides a new measure of accessibility and an information fusion framework to fuse information coming from the reports of automatic assessors allowing search engines to re-rank their results according to an accessibility level, as some users would like to [10]. This accessibility indicator considers several categories of deficiencies. Our approach is based on the theory of the belief functions adapted to take into account the defects of accessibility given by several automatic assessors seen as information sources, the uncertainty of their results, as well as the possible conflicts between the sources.

In the sections 2 and 3 we will give a description of accessibility tools based on a recent standard and of data provided in their reports. In the 4th section, we will describe the principles of our indicator and develop how we implement the belief functions. In the 5th part, we will present an experiment before concluding.

## 2 Defect detection of Web page accessibility

Various accessibility standards propose recommendations for improving accessibility of web sites. The Web Content Accessibility Guidelines (WCAG 2.0) [5] proposed by the W3C normalization organism, constitutes an international reference in the field. These guidelines cover a wide range of disabilities (visual, auditory, physical, speech, cognitive, etc.) and several layers of guidance are provided:

- 4 overall principles: perception, operability, understandability & robustness;
- testable success criteria : for each guideline, testable success criteria are provided. Every criterion is associated to one of the 3 defined conformance levels (A, AA and AAA), each representing a requirement of accessibility for users.

Several automatic accessibility assessors, based on various accessibility standards, have been developed [2] for IT professionals. Their limits depend on the automatic tests. Because it is at present not possible to test some criteria about the quality of some pages, some assessor results are given with ambiguity. Consequently, the existing automatic assessors look for the criteria which are not met and give the defects according to 3 levels of validity: the number of errors, which are estimated certain, the number of likely problems (warnings) whose reality is not guaranteed and the number of potential problems (also called generic or non testable) which leads to a complete uncertainty on the tested criterion accessibility.

Finally, even though the results obtained by different assessors match for some tested common criteria, results can differ, even for errors considered as certain.

## 3 Proposed accessibility indicator

After a request, the indicator has to supply information describing to users the accessibility level of each Web page proposed by a search engine after a request. Presented jointly to these pages, the indicator‘s information cover two aspects:

- the accessibility for categories of deficiencies: as previously proposed for accessibility estimation [3] we use 4 major categories: visual, hearing, motor and cognitive deficiencies, as defined by [16]. They are called “deficiency frames”;
- the level of accessibility for each deficiency frame.

Collecting results from several assessors has allowed us to benefit from each of their performance. In addition, it strengthens accessibility evaluation for similar results and manages conflicts in case of disagreements. Automatic assessors check a set of criteria which correspond to many deficiencies. As our accessibility evaluation varies for every deficiency frame, our method consists in selecting the relevant criteria for each deficiency frame and then balancing each criterion to consider the difficulties met by users in case of failure. This weighting is based on the criterion conformance level (A, AA, AAA), which corresponds to decreasing priorities (A: most important, etc.). The errors and problems detected for every criterion of the accessibility standard affect the accessibility indicator of the Web content tested according to the deficiency frame the criterion belongs to, its weighting within the frame, the number of occurrences when it is analyzed as a defect in the Web page and the defect’s degree of certainty (error, likely or potential problem).

## 4 Defect detection and accessibility evaluation

After collecting Web page Uniform Resource Locators ($URL_p$) selected by a search engine from a request, these addresses are supplied to the accessibility assessors and successively for each page, we detect accessibility defects, then estimate accessibility level by deficiency frame for each assessor, before fusing the data by deficiency frame and taking the decision for every deficiency frame [7].

### 4.1. Assessor evaluations of selected pages

Each $URL_p$ is submitted to the accessibility evaluation tests by each assessor $i$ that tests all the criteria $k$ of the WCAG 2.0 standard, and the following data are collected by a filter that extracts the required data for each deficiency frame:

- : errors observed for a criterion $k$ by an assessor $i$;
- : correct checkpoints for a criterion $k$ by an assessor $i$.

- : tests that can induce errors for a criterion *k* by an assessor *i*;
- : likely problems detected for a criterion *k* by an assessor *i*;
- : tests that can induce likely problems for a criterion *k* by an assessor *i*;
- : potential problems suspected for a criterion *k* by an assessor *i*;
- : tests that can induce potential problems for a criterion *k* by an assessor *i*;
- : total tests by an assessor *i,* with

### *4.2. Accessibility indicator level of the pages*

To model initial information including uncertainties, the reliability of the assessors seen as information sources and their possible conflicts, we use the theory of belief functions [6] [13]. Our objective is to define if a Web page is accessible *(Ac)* or not accessible ( and to supply an indication by deficiency frame. Consequently, these questions can be handled independently for every deficiency frame . We can consider every power set ∅ .

The estimation of the accessibility for a deficiency frame *h* and a source *i* (assessor) is estimated from the number of correct tests for each of the criteria *k* occurring in this frame, and from their conformance level represented by :

The estimation of the non accessibility for a deficiency frame *h* and a source *i* is estimated from the number of errors for each of the criteria *k* occurring in this frame, and from the coefficient. A weakening coefficient is also introduced to model the degree of certainty of the error:

The estimation of the ignorance for a deficiency frame *h* and a source *i* is estimated from the number of likely and potential problem for each of the criteria *k* occurring in this frame, and from the coefficient. The weakening coefficients or are also used to model the degree of certainty of the problem:

The mass functions of the subsets of are computed from the estimations:

In addition, the source reliability can be modeled [11] with a coefficient, which constitutes a benefit when some assessors are more efficient than others:

### *4.3. Merging assessor results and decision-making*

Once the masses for each assessor have been obtained, a fusion of the results is conducted by deficiency frame, using the conjunctive rule [14], to combine them and give information in the form of a mass function. These rule properties, which strengthen common results and manage conflicts between sources, are particularly relevant in this context, to deal with divergences between assessor evaluations. To calculate the final decision for a page by deficiency frame, we use the pignistic probability [14].

There are several of presenting the accessibility indicator to users. To visualize the deficiency frames, existing specific pictograms are effective. To present the accessibility level we discretize the decision into 5 levels (very good, good, moderate, bad or very bad accessibility) using thresholds and visualized it by an arrow.

- if , the Web content accessibility is very bad (↓),
- if , the Web content accessibility is moderate (→),
- if , the Web content accessibility is very good (↑) etc.

## 5 Experiments

To validate our approach, we present here the results obtained on a set of 100 news Websites, among the most visited ones, all referenced by the OJD organism which provides certification and publication of attendance figures for websites[4]. We test their homepages, following a study [12] concluding that their usability is predictive of the whole site. We chose two open source assessors ACHECKER, (source 1, noted AC ) [9], and TAW (source 2) from which we extract automatically the accessibility test results. Weight and threshold values given in Table 1 were previously empirically defined from Webpages[5] assumed to be accessible.

The results of these sources are summarized in Figure 1 for the 3 levels of certainty defects. The box plots present how their defects are distributed: minimum

[4] OJD: http://www.ojd.com/Chiffres/Le-Numerique/Sites-Web/Sites-Web-GP

[5] Sites labeled by Accessiweb: http://www.accessiweb.org/index.php/galerie.html

| | | | |
|---|---|---|---|
| Weightings | | A, AA, AAA conformance levels | 1 ; 0.8 ; 0.6 |
| | | Certainty levels of errors or problems | 1 ; 0.5 ; 1 |
| | | AC and TAW reliabilities (sources) | 1 ; 1 |
| Thresholds | S1 ; S2 ; S3; S4 | Accessibility indicator levels | 0.6; 0.7; 0.8; 0.9 |

Table 1. Constant values for our accessibility metric

and maximum (whiskers), 1st (bottom box plot) and 3rd quartiles (top box plot) and average (horizontal line). We observe similarities between the assessors' results for the errors detected as certain, but also huge differences for the likely (warnings) and potential (non testable) problems. The number of likely problems is almost null for AC and the potential one remains always the same for TAW.

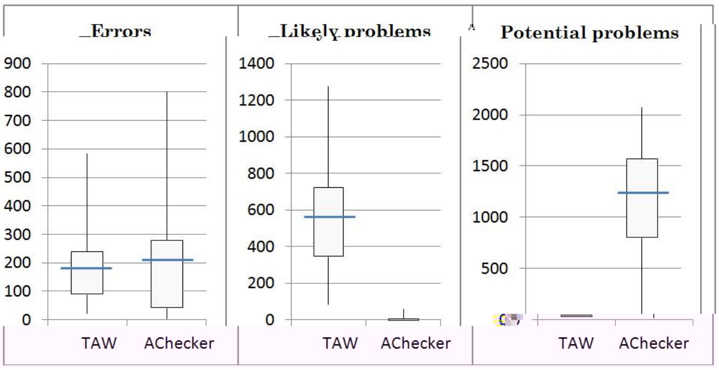

Figure 1. Results of automatic assessors

The detected defects are taken into account in our accessibility indicator results presented in Figure 2. The mass function values of accessibility for the 2 sources, TAW and AC, and the fusion result are visualized for 3 deficiency frames among the 4, and globally for all deficiencies. Firstly, we can see that is not evenly distributed between the 2 sources: their distributions of errors (Figure 1) are comparable even if there is a larger range for AC; however the mass function of accessibility is smaller for AC compared to TAW. This is due to the more numerous potential problems (non testable criteria) detected by the AC assessor, increasing substantially the denominator in the computation of *m(Ac)* (Eq. 5). By the way, the values of *E( )* and consequently of *m( ),* are more important, as the weight for potential problems is 2 times higher than for the likely problems (warnings). We can also notice that the fusion result obtained by the conjunctive rule strengthens the mass functions of the 2 assessors.

In this corpus, visual and cognitive deficiencies have a higher impact on content accessibility than the motor ones. This is logical for news websites, as their homepages include a large number of images. By the way, the motor indicator is less impacted, in particular by the lack of alternatives for images, useful for visual and cognitive deficiencies. Finally, we observe a similarity between the visual and global indicators, as around 80% of all the checkpoints concern visual deficiencies and also because these controls are properly taken into account by assessors.

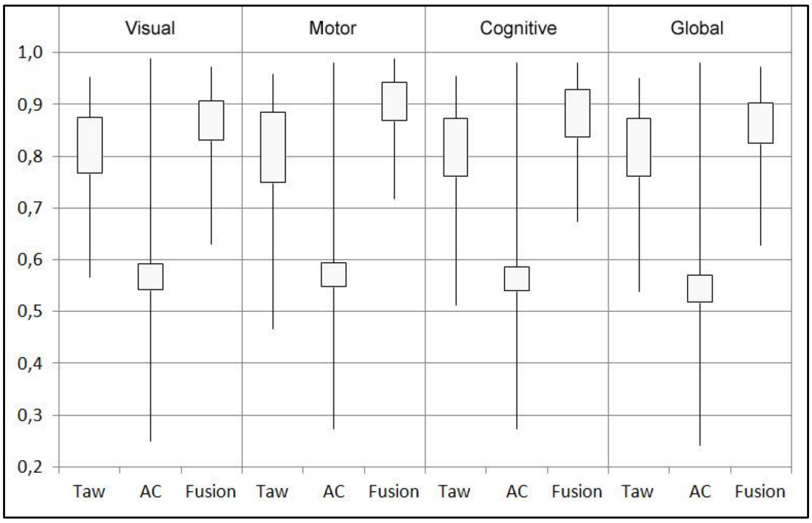

Figure 2. Accessibility indicator results

In Table 2 are presented detailed results for several sites with significant indicator result differences. For examples, LePoint.fr and Arte.tv, respectively 19th and 33th most consulted websites in France, obtain only 0,627 and 0,686 for the global result, whereas LeParisien.fr, ranked 12th, reaches 0,971. For Family.fr we observe differences between the deficiencies, nevertheless focus on accessibility generally benefits all deficiencies on the whole corpus.

| Web content | Decision | | | |
|---|---|---|---|---|
| | Visual | Motor | Cognitive | Global |
| LeParisien.fr | 0,972 ↑ | 0,989 ↑ | 0,974 ↑ | 0,971 ↑ |
| Famili.fr | 0,769 → | 0,924 ↑ | 0,838 ↗ | 0,766 → |
| Arte.tv | 0,701 → | 0,718 → | 0,717 → | 0,686 ↘ |
| LePoint.fr | 0,630 ↘ | 0,725 → | 0,673 ↘ | 0,627 ↘ |

Table 2. Examples of detailed accessibility results by deficiency frame

## 6 Conclusion

We present an indicator estimating Web page accessibility levels for distinct categories of deficiencies, in order to supply easily understandable accessibility information to users on pages proposed by a search engine. Our method based on belief function theory fuses results from several automatic assessors and considers their uncertainties. An accurate modelization of the assessor characteristics and of the impact of defect guideline criteria on accessibility is proposed. An experiment performed on a set of 100 news websites validates the method, which benefits from each of the assessor performances on specific criterion tests. Our future research will focus on the implementation of a user's personal weighting to balance the importance of criteria.

# REFERENCES

1. Julio Abascal, Myriam Arrue, Inmaculada Fajardo, Nestor Garay, and Jorge Tomhas. 2004. The use of guidelines to automatically verify web accessibility. Universal Access in the Information Society, 3(1):71–79.
2. Ben Caldwell, Michael Cooper, Loretta Guarino Reid, Gregg Vanderheiden, 2008. Web Content Accessibility Guidelines (WCAG) 2.0, W3C Recommendation, December 11th 2008. DOI: http:// http://www.w3.org/TR/WCAG20/. Accessed Jun 2010.
3. A.P. Dempster. Upper and Lower probabilities induced by a multivalued mapping. Anals of Mathematical Statistics, 38 :325-339, 1967.
4. Giorgio Brajnik, 2004. Comparing accessibility evaluation tools: a method for tool effectiveness. Universal Access in the Information Society (UAIS) 3(3-4), 252-263.
5. Giorgio Brajnik, 2006. Web Accessibility Testing: When the Method is the Culprit. In K. Miesenberger et al. (Eds.). Computers Helping People with Special Needs. ICCHP 2006 (Linz, Austria, July 12-14, 2006). LNCS 4061, 156-163.
6. Giorgio Brajnik, 2009. Validity and Reliability of Web Accessibility Guidelines. ACM SIGACCESS Conference on Computers and Accessibility, ASSETS'09, 131-138.
7. Jean-Christophe Dubois, Yolande le Gall, Arnaud Martin. Procédé de traitement de données d'accessibilité, dispositif et programme correspondant (Software Engineering System for the Analysis of Accessibility Data). Patent No. 1451562, Filed February 26th 2014.
8. European Disability Strategy 2010-2020: A Renewed Commitment to a Barrier-Free Europe. Communication from the commission to the European parliament, the Council, the European economic and social Committee and the Committee of the Regions, COM (2010) 636 final. November 15, 2010.
9. Greg Gay and Cindy Qi Li. 2010. AChecker: open, interactive, customizable, web accessibility checking. In Proceedings of the W4A International Cross Disciplinary Conference on Web Accessibility, (33) ACM, New York, NY, USA.
10. Melody Y. Ivory, Shiging Yu and Kathryn Gronemyer, 2004. Search result exploration: a preliminary study of blind and sighted users' decision making and performance. CHI Extended Abstracts (Vienna, Austria, April, 2004), 1453-1456, ACM Press, 2004.
11. Arnaud Martin, Christophe Osswald, Jean Dezert, and Florentin Smarandache, 2008. General combination rules for qualitative and quantitative beliefs, Journal of Advances in Information Fusion, Vol 3(2), pp. 67-82, December 2008.
12. Jakob Nielsen and Marie Tahir, 2001. Homepage Usability: 50 Websites Deconstructed. New Riders Publishing.
13. G. Shafer. A mathematical theory of evidence. Princeton University Press, 1976.
14. Philippe Smets, 1993. Belief Functions: the Disjunctive Rule of Combination and the Generalized Bayesian Theorem. International Journal of Approximate Reasoning, 9:1-35.
15. Harry Thornburg. 2001. Introduction to Bayesian Statistics. (March 2001). Retrieved March 2, 2005. DOI: http://ccrma.stanford.edu/~jos/bayes/bayes.html.
16. Gregg C. Vanderheiden and Katherine R. Vanderheiden, 1991. Accessible design guide I: guidelines for the design of consumer products to increase their accessibility to people with disabilities or who are aging. Madison, WI: Trace R&D Center.
17. Markel Vigo and Giorgio Brajnik, 2011. Automatic web accessibility metrics: where we are and where we can go. Interacting with Computers 23(2), 137-155. Elsevier.