

Tester si son site Web est techniquement internationalisé

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 7 octobre 2010

<https://www.bortzmeyer.org/i18n-checker.html>

Il y avait déjà des validateurs de l'HTML <<http://validator.w3.org/>>, des validateurs de flux de syndication <<http://validator.w3.org/feed/>>, et des validateurs de feuille de style CSS <<http://jigsaw.w3.org/css-validator/>> (sans compter ceux pour l'accessibilité). Désormais, le webmestre consciencieux a en plus un outil pour vérifier que son site est techniquement internationalisé : <<http://qa-dev.w3.org/i18n-checker/>>.

Je dis bien « techniquement internationalisé » car, de même qu'un site Web peut être entièrement en HTML valide et quand même inutilisable, un site Web peut être « techniquement internationalisé » et quand même peu adapté à une audience internationale. Un outil automatique ne peut en effet pas détecter toutes les erreurs classiques d'internationalisation (comme de tenir pour acquis que tout le monde sur Terre connaît les sigles nationaux, ou bien les films qui passent en ce moment dans tel pays). L'outil ne dispense donc pas d'un examen manuel, mais il permet de s'assurer que la base technique est correcte.

Pour l'anecdote (et pour donner des exemples concrets) mon blog affichait plusieurs problèmes (dont certains ont été corrigés) : pas d'attribut lang, seulement `xml:lang` ("A tag uses an `xml:lang` attribute without an associated `lang` attribute.") parce que j'avais oublié que tout le monde n'analyse pas le XML comme du XML mais parfois comme du HTML traditionnel. C'est très bien d'étiqueter les pages en indiquant la langue <<http://www.w3.org/TR/i18n-html-tech-lang/>> mais il faut le faire correctement. (Voir le RFC 5646¹ pour plus de détails sur l'étiquetage.)

- Éléments de présentation utilisés sans précautions ("*jb* tags found with no class attribute"). En effet, les règles de style typographiques peuvent varier selon les pays et utiliser aveuglément `` et `<i>` peut mener à des ennuis (cf. "*Using `jb` and `ji` elements*" <<http://www.w3.org/International/questions/qa-b-and-i-tags.en.php>>). Ce n'est pas trop grave dans mon cas, car l'HTML est produit automatiquement, donc un `` peut être changé en `<i>` facilement mais, bon, il faudrait quand même le faire plus proprement en indiquant systématiquement une classe pour ces éléments ou mieux, en n'utilisant plus d'éléments de présentation du tout.

Un autre problème n'avait pas été détecté par le service de validation : des textes auxiliaires (comme l'avertissement en bas de chaque page) étaient systématiquement en français même lorsque l'article était en anglais. Ce genre d'erreurs n'est pas détectable automatiquement. Désormais, j'ai corrigé celle-ci (enfin, pas à 100 %), donc on a tout en français dans les articles en français <<https://www.bortzmeyer.org/rapport-fdi-langues.html>> comme dans ceux en anglais <<https://www.bortzmeyer.org/canonicalize-zones.html>>.

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc5646.txt>