

Le Web et la version originale des documents

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 23 novembre 2008

<https://www.bortzmeyer.org/web-et-version-originale.html>

Aujourd'hui, beaucoup de sites Web sont disponibles en plusieurs langues. Par exemple le site de Barcelone, <<http://www.barcelona.cat/>>, est en catalan, en espagnol et en anglais. Celui de Katmandou, <<http://www.kathmandu.gov.np/>>, est en népalais et en anglais, etc. Mais comment choisit-on sa version ? Le problème est plus compliqué qu'il n'y paraît.

La norme du protocole HTTP, le RFC 2616¹, fournit une solution simple, la négociation de contenu (en-tête `Accept-Language`, RFC 2616, section 14.4). Le navigateur Web est configuré par son utilisateur avec une liste de langues préférées (par exemple, je préfère le breton, sinon le français, sinon l'anglais), il indique cette liste au serveur et le serveur envoie la page dans la langue préférée dont il dispose. Un logiciel comme Apache a ce qu'il faut pour cela <<http://httpd.apache.org/docs/2.2/content-negotiation.html>>. C'est très simple pour l'utilisateur, qui n'a pas à répéter son choix à chaque fois. S'il utilise un navigateur comme Konqueror, il n'aura même rien à faire, le navigateur utilisant les préférences globales de l'environnement KDE. Pour les autres navigateurs, on peut suivre l'excellente documentation de Debian <<http://www.debian.org/intro/cn>>.

Mais cette méthode est très peu utilisée. C'est dû en grande partie à la profonde ignorance de la plupart des webmasters : très rares sont ceux qui ont lu le RFC 2616. Mais il y a d'autres raisons : avec Apache, mettre en action cette option interfère avec d'autres options comme la sélection des formats des pages. Enfin, cette technique ne permet pas au lecteur de changer, par exemple s'il veut voir à quoi ressemblent les versions dans d'autres langues. Il faut donc toujours au moins fournir un échappement manuel. Par exemple, le site de Debian, <<http://www.debian.org/>>, traduit en des dizaines de langues, y compris le finnois et le coréen, fonctionne automatiquement avec la négociation de langue mais permet également, par des liens en bas de chaque page, de modifier ce choix.

La grande majorité des sites Web multilingues ignorent donc les préférences de langue exprimées via le navigateur. À la place, ils fournissent un mécanisme manuel où l'utilisateur doit, pour chaque site,

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc2616.txt>

trouver le lien permettant de sélectionner la langue, trouver sa langue (un problème difficile en soi, surtout avec les sites qui s'obstinent à utiliser de petits drapeaux <<http://www.useit.com/alertbox/flagproblem.html>>) et la sélectionner.

Ce système ne tire pas profit des possibilités de HTTP et est pénible pour l'utilisateur. Souvent, en outre, le code qui le met en œuvre est bogué, comme c'est le cas sur le site de Katmandou cité plus haut. Mais il est le seul à donner un contrôle complet à l'utilisateur, sur tous les aspects du choix de langue.

En effet, le mécanisme du RFC 2616 a une grosse limite : il ne traite pas le cas où les textes sont de qualité différentes selon les langues. Ce cas est fréquent (il suffit de comparer <<http://www.microsoft.com/en/us/>> et <<http://www.microsoft.com/fr/fr/>>) : bien que ma langue maternelle soit le français, si je demande systématiquement la version française, je me retrouve en général avec des mauvaises traductions, incomplètes et/ou dépassées. Parfois, la langue d'origine est le français et le problème est le même : la version anglaise est loin d'être équivalente.

Maintenir plusieurs versions du même site, en différentes langues, et les garder synchrones et de qualité équivalente à travers les changements et les réorganisations qu'affectent les services de communication est en effet une tâche herculéenne. Très peu d'organisations le font réellement.

Au cinéma, je ne cherche pas systématiquement des films en français, ni d'ailleurs systématiquement des films en anglais. Je préfère des films en **version originale**, la langue dans laquelle ils ont été faits. Pour les romans, où il n'y a pas de sous-titres, je préfère des versions originales si c'est une langue que je comprends, puis une traduction française et, s'il n'y a pas d'autre choix, une traduction anglaise.

Malheureusement, HTTP n'a pas cette notion de version originale et donc, en pratique, la sélection automatique de langue n'a que peu d'intérêt. J'ai donc cessé de l'utiliser (alors que je trouvais cette idée très astucieuse).

Voir aussi sur ce sujet l'article "*International Web Usability*" <<http://www.useit.com/alertbox/9608.html>>. L'entièreté du problème de choix de la langue est couvert très en détail (configuration d'Apache, pièges, moyens de les éviter) dans la section 10.5 du livre de Patrick Andries « Unicode en pratique » <<https://www.bortzmeyer.org/unicode-en-pratique.html>> ».

Le même problème se pose évidemment pour le courrier électronique, pour lequel une solution a été normalisée, dans le RFC 8255.