

GaBuZoMeu, a parser for language tags

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 19 Septembre 2006. Dernière mise à jour le 25 Août 2009

<http://www.bortzmeyer.org/gabuzomeu-parsing-language-tags.html>

GaBuZoMeu is a set of programs to parse and check language tags (see the RFC 5646¹ produced by the IETF Working Group LTRU (<http://www.ietf.org/html.charters/ltru-charter.html>) - Language Tag Registry Update).

Language tags are used by several protocols (like HTTP) or languages (like XML) to indicate the human language you want or have. Examples of language tags are "fr" (French), "apa" (Apache), "se-FI" (Swedish, as practiced in Finland) or "uk-Latn" (ukrainian in the latin script).

An IANA registry stores the registered values (<http://www.iana.org/assignments/language-subtag-registry>)

A language tag can be well-formed (syntactically correct) or not. This can be tested without access to the registry. A language tag can be valid or not (all its subtags registered). This depends on the copy of the registry you use (since the registry changes).

GaBuZoMeu includes the following programs :

- `check-wf` checks the well-formedness of tag(s). Example of use : "`check-wffren-AU`".
- `check-valid` checks the validity of tag(s). Same use.

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

It complies with RFC 5646.

You can download GaBuZoMeu (en ligne sur <http://www.bortzmeyer.org/files/gabuzomeu.tar.gz>).

GaBuZoMeu is written in Haskell so you'll need an Haskell compiler such as ghc (<http://haskell.org/ghc/>).

To compile, just type "make". To check that everything is OK, you can type "maketest". You should get zero "Errors" and zero "Failures".

GaBuZoMeu is distributed as free software, under the GNU General Public Licence. Remarks, patches and bug reports are welcome.

Written and maintained by Stéphane Bortzmeyer <bortzmeyer@nic.fr>.

Thanks for AFNIC (http://www.afnic.fr/_en), the .fr registry, for making this work possible.

Some other language tag parsers can be found at the language tag Web site (<http://www.langtag.net/>).