

Extrait du Easter-eggs - Spécialiste GNU/Linux

<http://www.easter-eggs.com>

Validation d'un fichier CSV à l'aide d'un schéma XML

- Études de cas - Développement applicatif -

Date de mise en ligne : jeudi 26 juin 2008

Easter-eggs - Spécialiste GNU/Linux

Le format de fichier CSV présente certains avantages :

- il est facilement manipulable pour les utilisateurs à l'aide d'un tableur comme Calc, Gnumeric ou bien Excel ;
- il permet d'écrire rapidement et facilement des routines d'import / export ;
- il permet d'échanger les données entre des bases et des systèmes hétérogènes.

Mais ce format de fichier ne dispose pas d'un système de validation. Dans la plupart des cas ce sont les routines d'import / export qui prennent en charge la validation des données et de leur mise en forme.

Ce travail est pénible, répétitif et ennuyeux mais nécessaire si vous ne voulez pas qu'un fichier mal formaté ne corrompe les données de votre base.

L'objectif de cet article est de proposer une solution simple permettant la validation d'un fichier CSV en utilisant les fonctionnalités offertes par les schémas XML.

Les exemples fournis sont écrits à l'aide du langage de programmation PHP5.

Transformation CSV vers XML

Votre fichier doit contenir sur la 1re ligne, le nom de chaque colonne :

```
"CHANNEL" ;"COMPANY" ;"BVD_ID_Groupe_13" ;"BVD_ID_Groupe_23" ;"GENRE" ;"URL" ;"OTHER_URL"
;"COMPETENT_AUTHORITY" "Canal+ Film 13" ;"C MORE ENTERTAINMENT AB" ;"NL332685953
;"DE83302617943" ;"CIN" ;"http://www.canalplus.se" ;"http://www.prosiebensat1.de/" ;"... "Canal+ Film 23" ;"C
MORE ENTERTAINMENT AB" ;"NL332685953" ;"DE83302617943" ;"CIN" ;"http://www.canalplus.se" ;"http://www.prosiebensat1.de/" ;"...
http://www.prosiebensat1.de/" ;"... "Canal+ Film 2 Sport Weekend" ;"C MORE ENTERTAINMENT AB"
;"NL332685953" ;"DE83302617943" ;"CIN" ;"http://www.canalplus.se" ;"http://www.prosiebensat1.de/" ;"...
"Canal+ Film 33" ;"C MORE ENTERTAINMENT AB" ;"NL332685953" ;"DE83302617943" ;"CIN" ;"
http://www.canalplus.se" ;"http://www.prosiebensat1.de/" ;"... "Canal+ Film HD" ;"C MORE ENTERTAINMENT
AB" ;"NL332685953" ;"DE83302617943" ;"CIN" ;"http://www.canalplus.se" ;"http://www.prosiebensat1.de/" ;"..."
```

Pour transformer le fichier CSV en XML, il suffit de lire le fichier CSV avec la fonction `fgetcsv` et mettre les données dans des balises XML : [csv_to_xml.php](#)

Un fichier CSV converti en XML ressemble à cela : [channel.xml](#)

Validation du fichier XML

Il ne reste plus qu'à écrire le schéma permettant de valider le fichier XML :

[channel.xsd](#)

Vous pouvez tester vos schéma en ligne de commande en utilisant la commande `xmllint` issue du paquet `libxml2-utils` Debian :

```
$ xmllint &mdash;noout &mdash;schema channel.xsd channel.xml    channel.xml validates
```

Validation en PHP

Une fois que votre schéma est correct et permet de valider le contenu du fichier XML, il faut l'intégrer directement dans votre application : [validate.php](#)

Conclusion

Cette méthode apporte les avantages suivants :

- Les règles de validation sont décrites directement via le schéma et celui-ci dispose de nombreuses fonctionnalités tels les types de bases, les expression rationnelles, les séquences, etc.
- Les messages d'erreur sont compréhensibles pour l'utilisateur même s'il faut parfois les traduire (en convertissant les numéros de lignes et certaines notions) ;
- La simplicité.

Cette technique a été mise en oeuvre dans le cadre du projet [MAVISE](#) pour valider plus d'une vingtaine de formats de fichiers différents.