

QExtend - Bogue #671

Verifier que les .h et .cpp sont en UTF8 et non en AINSI ou UTF8(sans BOM)

04/19/2010 01:58 PM - yan

Status:	Fermé	Start date:	04/19/2010
Priority:	Haut	Due date:	
Assignee:		% Done:	0%
Category:	Compilation	Estimated time:	0.00 hour
Target version:	Version 0.0.1	Spent time:	0.00 hour
Description			
Plusieurs fichier sont en AINSI ou UTF8			
J'ai plusieurs fois dû changer l'encodage de mes propres sources. Mais ça reviens parfois en arrière. Je ne sais pas si cela viens de svn ou visual.			

History

#1 - 04/19/2010 08:52 PM - dourouc05

Pour changer l'encodage dans VS10 : ouvrir le fichier, menu File, option Save as, cliquer sur la flèche à côté de Save, choisir Save with Encoding. Dans la fenêtre qui suit, sélectionner Unicode (UTF-8 without signature), puis Ok. Fini ! Prochaine sauvegarde, ça restera de l'UTF8. Ça doit être possible de définir un paramètre solution-wide, mais je n'ai pas encore cherché.

NB : Ainsi est un adverbe en français, ANSI est un institut américain de normalisation, dont un encodage, dit ANSI tout simplement.

#2 - 04/19/2010 08:59 PM - yan

On as dit que l'on utilisé de l'UTF8 et non de l'UTF-8 without signature...

#3 - 04/19/2010 09:01 PM - dourouc05

C'est son petit nom, à UTF-8 sans BOM. Et puis, tu n'as que ça en UTF-8 :aie:. Ou alors, de l'UTF-7. En tout cas, je ne changerai pas d'EDI pour une affaire d'encodage, quitte à le faire manuellement après.

#4 - 04/19/2010 09:26 PM - yan

VC support le fichier UTF8.

Sous VC2008, je doit les créer en UTF8 puis les utiliser avec visual. Et la il reste en UTF8.

Si l'utf8 sans bom est préférable alors partons sur cela. Faut juste se mettre d'accord ;)

#5 - 04/19/2010 09:29 PM - yan

with sous VC 2010 et 2008 express j'ai unicode (UTF8 **with** signature). Tu peut vérifier chez toi?

#6 - 04/19/2010 09:33 PM - dourouc05

J'ai vérifié, je n'ai que le sans signature... Peut-être à l'autre bout de la liste ?

#7 - 04/19/2010 09:35 PM - dourouc05

Voilà, le bon Unicode était en tout début de liste : pourquoi commencer par U ?

#8 - 04/19/2010 09:37 PM - yan

lol

affaire réglé pour visual.
Et sous les autres ide, vous avez des problèmes???

#9 - 04/19/2010 09:41 PM - johnlamerica

Oui moi sous Qt Creator ça merde aussi, que je mette le projet en UTF8 ou pas, j'ai des fichiers .cpp qui veulent pas bouger de l'ANSI...

#10 - 04/19/2010 09:47 PM - yan

Si le fichier est déjà en UTF8, y as pas de problème non plus avec QtCreator. non?

#11 - 04/20/2010 09:30 AM - gbdivers

Chez moi (Ubuntu 9.10 64b, avec Qt Creator ou en ligne de commande), j'ai systématiquement un problème d'encodage avec les erreurs :

```
../../../../include/QExtend/Core/ExtendPtr:1: erreur: stray '\357' in program
../../../../include/QExtend/Core/ExtendPtr:1: erreur: stray '\273' in program
../../../../include/QExtend/Core/ExtendPtr:1: erreur: stray '\277' in program
../../../../include/QExtend/Core/ExtendPtr:1: erreur: stray '#' in program
```

Je dois systématiquement, pour chaque fichier, faire Edition->changer l'encodage->UTF-8->Recharger avec l'encodage puis Edition->changer l'encodage->UTF-8->Sauver avec l'encodage pour pouvoir compiler.

Les caractères invisibles sont présent si j'ouvre avec Qt Creator et gedit (ne reconnait pas le format UTF-8 ?) mais correctement ouvert avec OpenOffice en UTF-8.

#12 - 04/20/2010 09:39 AM - LeGars

Je ne sais pas pourquoi, mais c'est le BOM qui plombe sous linux... J'ai le même problème sous Gentoo.

#13 - 04/20/2010 09:46 AM - gbdivers

Puisse qu'on est dans les joyeustés, voici les résultats obtenus avec la commande file de quelques fichiers (Ubuntu 9.10 64b) :

```
guillaume@guillaume-desktop:~/qextend/branches/dev/src/core$ file *
connector.cpp:      UTF-8 Unicode (with BOM) C program text
connector.h:        UTF-8 Unicode (with BOM) C program text
extendptr.h:        UTF-8 Unicode (with BOM) C program text
logger.cpp:         ASCII C program text, with CRLF line terminators
logger.h:           UTF-8 Unicode C program text, with CRLF line terminators
objectextendptrpolicy.h: UTF-8 Unicode (with BOM) C program text
objectptr.h:        UTF-8 Unicode (with BOM) C program text, with CRLF line terminators
qextend_global.h:   ASCII C program text, with CRLF line terminators
templatetools.h:    UTF-8 Unicode (with BOM) C program text
```

```
guillaume@guillaume-desktop:~/qextend/branches/dev/src/gui$ file *
recentmenu.cpp:     ASCII C program text, with CRLF line terminators
recentmenu.h:       UTF-8 Unicode C program text, with CRLF line terminators
```

Après ouverture avec Qt Creator puis ré-encodage en UTF-8 (par exemple de connector.h)...

```
guillaume@guillaume-desktop:~/qextend/branches/dev/src/core$ file connector.cpp
connector.cpp: ASCII C program text
```

Bref, ce n'est pas correctement reconnu lors de la compilation ou lors de l'ouverture avec Qt Creator mais par contre, c'est reconnu avec la commande file.

#14 - 04/20/2010 09:50 AM - gbdivers

Pour Qt et Qt Creator, la commande file me donne "ASCII English text" ou "ASCII C program text"

Il faudrait peut être repasser en ASCII

#15 - 04/20/2010 09:51 AM - yan

Je savais bien que j'avais eu un problème avec le UTF8 + bom.
En cherchant, il semble que depuis gcc 4.3 c'est supporté. Quelle version de gcc as tu?
Utf8 sans bom semble poser problème sous visual... à tester.

#16 - 04/20/2010 09:59 AM - LeGars

- File cleanBOM.sh added

Je sentais bien que l'encodage était une source de noeuds dans le cerveau...
En pièce jointe, un script qui permet de virer le BOM sous linux.

#17 - 04/20/2010 10:05 AM - yan

Je viens de vérifier, pour le défis, j'ai utilisé de l'utf8 sans BOM et ca marche sous windows et linux.
Donc je propose que l'on passe à UTF 8 sans BOM.

#18 - 04/20/2010 10:41 AM - gbdivers

Ton script fonctionne. Par contre, on se retrouve quand même avec un mélange de UTF-8 et d'ASCII

@Yan : J'utilise gcc 4.4 (boulot) et 4.3 (chez moi)

#19 - 04/20/2010 11:13 AM - LeGars

Au pire, bien que radicale comme solution, il faudrait virer les commentaires en français. Comme cela l'ascii de base sera compris par tous...

#20 - 04/20/2010 11:38 AM - yan

gbdivers a écrit:

Ton script fonctionne. Par contre, on se retrouve quand même avec un mélange de UTF-8 et d'ASCII

@Yan : J'utilise gcc 4.4 (boulot) et 4.3 (chez moi)

c'est à dire?
les source en UTF8 et les .pro,.pri,.. en ainsi?
c'est pas très grave, non?

#21 - 04/20/2010 11:40 AM - yan

LeGars a écrit:

Au pire, bien que radicale comme solution, il faudrait virer les commentaires en français. Comme cela l'ascii de base sera compris par tous...

Suivant les plateformes, ce n'est pas le même ainsi... c'est un peu le bordel en faite.
L'UTF8 permet d'écrire des string dans le code avec plus de caractère.

Puis je trouve cela pas mal de faire une doc anglaise/fançaise, non?

#22 - 04/20/2010 11:50 AM - johnlamericaïn

Oui je pense qu'il faut garder une doc anglaise/française (je dis pas ça parce que j'ai déjà commencé ;))

Il faut trouver un moyen d'avoir tout en UTF-8 et que nos IDE ne convertisse pas en ANSI à chaque enregistrement...

#23 - 04/20/2010 11:54 AM - yan

Je propose que l'on passe les sources en UTF8 sans BOM et que l'on voi si on as encore des problèmes. Il n'y as pas encore beaucoup de sources donc autant essayer. Ca prendra pas de temps.

Si quelqu'un peut faire la modif avec le script et mettre sous svn?

Chaqu'un pourra faire un update et tester

#24 - 04/20/2010 12:04 PM - LeGars

je veux bien mettre à jour le svn, par contre je n'ai pas de windows sous la main...

#25 - 04/20/2010 12:09 PM - yan

LeGars a écrit:

je veux bien mettre à jour le svn, par contre je n'ai pas de windows sous la main...

Ok merci.

Une fois mis en utf8 sans bom, si personne n'à de problème, c'est résolue :D

sinon, ben... je sais pas:(

On est bien d'accord ce ne sont que les fichiers de code que l'on passe en UTF8 sans BOM.

#26 - 04/20/2010 12:12 PM - LeGars

*.h et *.cpp ?

#27 - 04/20/2010 12:17 PM - yan

LeGars a écrit:

*.h et *.cpp ?

oui. Y aurais d'autres fichiers?

#28 - 04/20/2010 12:18 PM - yan

yan a écrit:

LeGars a écrit:

*.h et *.cpp ?

oui. Y aurais d'autres fichiers?

#29 - 04/20/2010 12:21 PM - LeGars

*.hpp et les includes sans extensions

#30 - 04/20/2010 12:26 PM - yan

LeGars a écrit:

*.hpp et les includes sans extensions

On as des .hpp? pourquoi?

sinon oui les sans extension aussi je pense

#31 - 04/20/2010 12:35 PM - LeGars

conversion et commit ok.

#32 - 04/20/2010 12:43 PM - johnlamerica

Perso mon logger.cpp qui était en ANSI n'a pas l'air d'avoir évoluer...

#33 - 04/20/2010 02:02 PM - johnlAmericain

En faite c'est peut être bon, a priori Notepad++ considère un fichier en ANSI si il détecte aucun char UNICODE dans le fichier. Qt Creator me dit bien que c'est de l'UTF-8. Une application sous windows permet-elle de connaitre l'encodage d'un ensemble de fichiers ? Merci

#34 - 04/20/2010 07:31 PM - dourouc05

Je suis plus, là... On fait quoi, au final ? UTF-8 avec ou sans BOM ? ANSI ? UTF-7, UTF-16, UTF-32, UTF-DTC ?

#35 - 04/20/2010 07:41 PM - yan

dourouc05 a écrit:

Je suis plus, là... On fait quoi, au final ? UTF-8 avec ou sans BOM ? ANSI ? UTF-7, UTF-16, UTF-32, UTF-DTC ?

Utf8 sans bom. Sinon ca pose problème sous linux.

Pour info, comme il n'y as pas de BOM, si le fichier ne contiens que des caractères ainsi-7bits, les éditeurs ne savent pas si c'est du ainsi ou de l'utf8. Vue que l'utf8 est une sur-couche de l'ainsi-7 bits.

Dans la partie licence, on pourrais ajouter un marker (éèà ou autre) pour que les éditeurs édite avec le bon encodage

#36 - 04/20/2010 07:58 PM - dourouc05

OK. Le wiki est à jour ?

PS : merci Alex d'avoir bien tout changé !

#37 - 04/20/2010 08:47 PM - yan

dourouc05 a écrit:

OK. Le wiki est à jour ?

PS : merci Alex d'avoir bien tout changé !

wiki mise à jour.

Attention il faut remplacer (C) par © dans le copyright. Comme cela, ca sert de marker utf8 ;)

Je met en résolué. Sans problème remonté

#38 - 04/20/2010 08:47 PM - yan

- Status changed from Nouveau to Fermé

Files

cleanBOM.sh	1.22 KB	04/20/2010	LeGars
-------------	---------	------------	--------