Subject: Speller dictionaries...
Posted by mirek on Mon, 16 Nov 2009 09:46:30 GMT
View Forum Message <> Reply to Message

With the new format of speller dictionaries, I am looking for volunteer that would convert OO files into new ".udc" and uploaded them to sf.net for as many languages as possible... (plus original sources as well into GPL section - LGPL).

Mirek

Subject: Re: Speller dictionaries...
Posted by koldo on Mon, 16 Nov 2009 13:04:37 GMT
View Forum Message <> Reply to Message

Hello Mirek

I can help, at least for latin alphabet languages.
I may begin with Catalan, Basque, Galician and Spanish (different varieties).

Best regards
Koldo

Subject: Re: Speller dictionaries...
Posted by mirek on Mon, 16 Nov 2009 16:48:49 GMT
View Forum Message <> Reply to Message

Perfect.

I guess OO->list of files, the conversion utility will need some checking of correct input encoding.

Thanks.

Mirek

Subject: Re: Speller dictionaries...
Posted by cbpporter on Tue, 17 Nov 2009 08:28:13 GMT
View Forum Message <> Reply to Message

I can test it for my native language at least. But where is the conversion utility? What can this .udc do? And why can't we import original OO file? It is just a text file with each word on a line followed optionally by plurals, gender info an synonyms. There is even some very poor code on the OO site which loads a such a file.

PS: I've been meaning for a while now to translate TheIDE in my native language. Maybe now is the time.

Subject: Re: Speller dictionaries...
Posted by mirek on Tue, 17 Nov 2009 09:23:55 GMT
View Forum Message <> Reply to Message

cbpporter wrote on Tue, 17 November 2009 03:28I can test it for my native language at least. But where is the conversion utility? What can this .udc do? And why can't we import original OO file? It is just a text file with each word on a line followed optionally by plurals, gender info an synonyms. There is even some very poor code on the OO site which loads a such a file.

Sorry, some missing details:

in uppbox, there are two new package:

ConvertDic - this is capable of converting 'older' openoffice .dic format into plain list of words. It is at the moment a halfbaked code - especially, there is no check for source codepage. Any improvements are welcome.

MakeSpellScd (note that the name is missleading, we are now creating '.udc') - this takes such list of words and compresses it into .udc file (e.g. en-us.udc). RichEdit then searches for .udc files (on its directory, all parent directories, PATH and LIB) for given language (it also searches for older .scd files if .udc is not found).

Note: We are not using openoffice files directly in RichEdit simply because I do not know how to make spellchecker efficient with them... .udc format is simple, but surprisingly effective. E.g. it compresses 400MB basque dictionary into 6MB file

Quote:
PS: I've been meaning for a while now to translate TheIDE in my native language. Maybe now is the time.

TheIDE? Why?

Do you think there are many good C++ programmers in the world not knowing English at the level required to use theide?

Subject: Re: Speller dictionaries...
Posted by cbpporter on Tue, 17 Nov 2009 12:15:25 GMT
View Forum Message <> Reply to Message

luzr wrote on Tue, 17 November 2009 11:23

ConvertDic - this is capable of converting 'older' openoffice .dic format into plain list of words. It is at the moment a halfbaked code - especially, there is no check for source codepage. Any improvements are welcome.

I don't have .dic, I have .dat. Actually I have zip files.

Quote:
Note: We are not using openoffice files directly in RichEdit simply because I do not know how to make spellchecker efficient with them... .udc format is simple, but surprisingly effective. E.g. it compresses 400MB basque dictionary into 6MB file

Efficient as in size?

Quote:
TheIDE? Why?

Do you think there are many good C++ programmers in the world not knowing English at the level required to use theide?

No, but I can say the same for Notepad or other applications, and people still want them internationalized. Especialy if every single application is in a language, the one odd English one sure stands out.

Subject: Re: Speller dictionaries...
Posted by koldo on Tue, 17 Nov 2009 15:45:06 GMT

Hello Mirek

Quote:TheIDE? Why?

Do you think there are many good C++ programmers in the world not knowing English at the level required to use theide?

I agree with cbpporter:

Perhaps TheIde and Upp is for everybody as in fact it is more productive but also simpler to program here than using other libraries.

So to an unskilled C++ programmer, I would propose him/her Upp over any other option.

And many people knows a few of English so they prefer to use their software in their language.

So I think that adding the language option in TheIde would be good for having more users.

Best regards

Subject: Re: Speller dictionaries...
Posted by mirek on Wed, 18 Nov 2009 07:36:34 GMT
View Forum Message <> Reply to Message

Quote:
cbpporter wrote on Tue, 17 November 2009 07:15
Do you think there are many good C++ programmers in the world not knowing English at the level required to use theide?

No, but I can say the same for Notepad or other applications

Really? Well, I know a lot of regular computer with no english knowledge users using localised Notepad or Word. There i18n definitely makes a sense.

I guess development tools and developers is different group.

Also consider manual texts - do you think that at current docs progress pace, there will be many translations soon? And frankly, library docs are much more important than a dozen of those labels in theide...

However, both theide translation and manuals definitely CAN be done. So if you keep insisting, I will undergo that boring process of turning all text literals to t_'s

P.S.: BTW, Visual Studion stands untranslated for my language, unlike OpenOffice or Word....

Subject: Re: Speller dictionaries...
Posted by koldo on Wed, 18 Nov 2009 08:50:28 GMT
View Forum Message <> Reply to Message

Hello Mirek

TheIde translation is not for me. It is easier to handle only one language. But there are many people (more than we think) that only knows one language well. I think it happens mainly in not english speaking relatively big countries (Spain is an example).

So perhaps TheIde internationalization is not a priority, but it would have to be done. We can help you!

And remember that, for example, GCC is translated. This is a problem in case of submitting issues to international forums but, many many people really fears using english just because they

know only a few words.

Best regards
Koldo

Subject: Re: Speller dictionaries...
Posted by koldo on Wed, 18 Nov 2009 11:18:37 GMT
View Forum Message <> Reply to Message

Hello Mirek

Some fix in ConvertDic. The command line takes only one dictionary file name without extension, but the text says "Usage: ConvertDic <file.dic file.aff>\n"

MakeSpellScd creates a info.txt file. What is it for ?.

Best regards
Koldo

Subject: Re: Speller dictionaries...
Posted by mirek on Wed, 18 Nov 2009 11:30:50 GMT
View Forum Message <> Reply to Message

koldo wrote on Wed, 18 November 2009 06:18
MakeSpellScd creates a info.txt file. What is it for ?.

It is sort of debugging info - list of all produced segments.

You can switch it off via #define at the start of file, I believe....

Mirek

Subject: Re: Speller dictionaries...
Posted by koldo on Wed, 18 Nov 2009 12:41:24 GMT
View Forum Message <> Reply to Message

Hello all

An advise: The final udc file has to be with '-' like ca-es.udc, not ca_es.udc. This is not got by TheIde speller.

Quote:I may begin with Catalan, Basque, Galician and Spanish (different varieties).

Catalan, Basque, Galician and Spanish (Spain and Mexico) are ready

Best regards
Koldo