Subject: compiler's output message of vc10free Chinese edition
Posted by LeiMing on Tue, 11 Oct 2011 05:40:45 GMT
View Forum Message <> Reply to Message

old title (before edit):
UTF-8 Encoded src files and vc10free Chinese edition

Hello,

I'm using ultimate++ 3211 under windows 7 chinese, the compilers are visual c++ 2003 toolkit (compiler updated with the files extracted from vc2003 sp1) w/ win2003 psdk and visual c++ 2010 express sp1 w/ winsdk 7.0.

The application I'm working on has many strings written in Chinese. It works well with visual c++ 2003 toolkit compiler (english) but when turned to the vc10 chinese compiler, errors come out. The error messages are also shown in human-unreadable forms (it should be in chinese with native encoding like gbk, but I guess it's mis-decoded).

maybe vc7.1 compiler and vc10 compiler work in different ways or vc compiler english and vc compiler chinese work in different ways to recognize the encoding of source code files.

after I tried converting the source code files into utf8 with bom from utf8 without bom, the compiler works. But when open it in theIDE, the very first character is abnormal.

So I wonder if it could be solved in next release version. now it works with source code encoded in utf8 without bom mark but vc10free chinese compiler doesn't support this (it requires the bom mark).

this is the one I wrote for test.
http://173.224.214.51/ftp/test2.7z
It works with vc7.1 english compiler but failed in vc10 chinese compiler

Subject: Re: UTF-8 Encoded src files and vc10free Chinese edition
Posted by mirek on Thu, 13 Oct 2011 18:24:55 GMT
View Forum Message <> Reply to Message

LeiMing wrote on Tue, 11 October 2011 01:40Hello,

I'm using ultimate++ 3211 under windows 7 chinese, the compilers are visual c++ 2003 toolkit (compiler updated with the files extracted from vc2003 sp1) w/ win2003 psdk and visual c++ 2010 express sp1 w/ winsdk 7.0.

The application I'm working on has many strings written in Chinese. It works well with visual c++ 2003 toolkit compiler (english) but when turned to the vc10 chinese compiler, errors come out. The error messages are also shown in human-unreadable forms (it should be in chinese with native encoding like gbk, but I guess it's mis-decoded).

maybe vc7.1 compiler and vc10 compiler work in different ways or vc compiler english and vc compiler chinese work in different ways to recognize the encoding of source code files.

after I tried converting the source code files into utf8 with bom from utf8 without bom, the compiler works. But when open it in theIDE, the very first character is abnormal.

So I wonder if it could be solved in next release version. now it works with source code encoded in utf8 without bom mark but vc10free chinese compiler doesn't support this (it requires the bom mark).

this is the one I wrote for test.
http://173.224.214.51/ftp/test2.7z
It works with vc7.1 english compiler but failed in vc10 chinese compiler

Well, this is an issue we are fighting with for some time now.

We have solved it as long as you are using i18n system of U++.

Finding that BOM fixes the issue as well is a good find and I guess we should consider this for the next release.

Anyway, one thing I am afraid about is how other compilers would react to BOM? What does GCC say about it? (We now unfortunately included Watcom builder as well - how does it react to BOM?).

Mirek

Subject: Re: UTF-8 Encoded src files and vc10free Chinese edition
Posted by mirek on Thu, 13 Oct 2011 18:26:45 GMT
View Forum Message <> Reply to Message

Related info:

http://gcc.gnu.org/bugzilla/show_bug.cgi?id=33415

Mirek

Subject: Re: UTF-8 Encoded src files and vc10free Chinese edition
Posted by Sender Ghost on Thu, 13 Oct 2011 21:10:38 GMT
View Forum Message <> Reply to Message

mirek wrote on Thu, 13 October 2011 20:24
We now unfortunately included Watcom builder as well - how does it react to BOM?

The Open Watcom compiler has following options for this case:

-zk0 double-byte char support for Kanji
-zk0u translate Kanji double-byte characters to UNICODE
-zk1 double-byte char support for Chinese/Taiwanese
-zk2 double-byte char support for Korean
-zkl double-byte char support if current code page has lead bytes
-zku=<codepage> load UNICODE translate table for specified code page

Yes, it is not very capable to build U++ applications, after some tests, but it capable to build Open Watcom sample applications.
If you think, it is unfortunate to have such a builder, you can remove it before release. I can move it to Bazaar section on the forum.

Subject: Re: UTF-8 Encoded src files and vc10free Chinese edition
Posted by LeiMing on Fri, 14 Oct 2011 02:48:37 GMT
View Forum Message <> Reply to Message

today I have a try on version 4011.
the compiler's message shown correctly.

theIDE editor is still not works with those with bom so I convert the source code manually.

this is what I have tested in 3211

vc toolkit 7.1 sp1 eng
with bom: work
without bom: compile with no error but the text in my program shown incorrectly, so I say it "not work".

vc 2008 sp1 eng
with bom: work
without bom: it doesn't compile

vc 2008 sp1 chs
with bom: work

without bom: it doesn't compile

vc 2010 eng
with bom: work
without bom: it doesn't compile

vc 2010 chs
with bom: work
without bom: it doesn't compile

tdm-gcc 4.5.2
with bom: work
without bom: work

openwatcom 1.9 doesn't compile the source code. it runs into error from the first library of ultimate++.
(in my experience, openwatcom compiler only compiles a small number of opensource libraries...orz)

mingw-gcc 3.4 doesn't compile the ultimate++ library.

by the way, when I setup the build method, I found that the vc 8 build method created by automatic setup won't have the correct sdk directory.

environment:
intel i3 2120
4G ram
1T hd (fs: ntfs)
windows 7 32bit sp1 ultimate chinese

Subject: Re: UTF-8 Encoded src files and vc10free Chinese edition
Posted by mirek on Fri, 14 Oct 2011 19:20:43 GMT
View Forum Message <> Reply to Message

OK, adding task to RM...

Mirek

Subject: Re: UTF-8 Encoded src files and vc10free Chinese edition
Posted by LeiMing on Tue, 05 Feb 2013 14:29:35 GMT
View Forum Message <> Reply to Message

Hello,
I test it with stable release ver 5485 again,

the source compiled and message shows normally.
but vc's output message is still displayed incorrectly.
like this:

Quote:NsisVPatcherCreator.cpp

Currently I'm trying replace language pack of vc with an English one to work around.
hope it would be fixed.

some possible useful information:
I run notepad++ and select menu "encoding -> encode in ansi", paste the chaos output message from the IDE, and select menu "encoding -> encode in utf8", many of the characters become readable, the left may be destroyed during copy and paste.
