Subject: [FIXED] ToUpper, ToLower,and ToAscii (char*, int) causes AssertFailed
Posted by omari on Fri, 03 Jun 2016 11:21:11 GMT
View Forum Message <> Reply to Message

Hi,

the following code causes Assertion fail:

```
#include <Core/Core.h>
using namespace Upp;
CONSOLE_APP_MAIN
{
 char* s = "abcd";
 ToUpper(s, 4);
}
```

the backtrace:

```
7fefd4531f3 (KERNELBASE.dll)
Upp::AssertFailed(file=13f22b238 "c:\\upp\\uppsrc\\core\\Vcont.h", line=84 'T', cond=13f22b570 "i
>= 0 && i < items") at c:\upp\uppsrc\core\util.cpp 166
Upp::Vector<Upp::CharSetData * __ptr64>::Get(i=255) at c:\upp\uppsrc\core\vcont.h 84
Upp::Vector<Upp::CharSetData * __ptr64>::operator[](i=255) at c:\upp\uppsrc\core\vcont.h 100
Upp::Array<Upp::CharSetData>::Get(i=255) at c:\upp\uppsrc\core\vcont.h 229
Upp::Array<Upp::CharSetData>::operator[](i=255) at c:\upp\uppsrc\core\vcont.h 244
Upp::AMap<Upp::String,Upp::CharSetData,Upp::Array<Upp::CharSetData> >::operator[](i=255) at
c:\upp\uppsrc\core\map.h 101
Upp::s_cset(charset=255) at c:\upp\uppsrc\core\charset.cpp 2149
Upp::ToUpper(t=13f212a50 "abcd", s=13f212a50 "abcd", len=4, charset=255) at
c:\upp\uppsrc\core\charset.cpp 2633
Upp::ToUpper(s=13f212a50 "abcd", len=4, charset=0) at c:\upp\uppsrc\core\charset.cpp 2660
ConsoleMainFn_() at c:\upp\myapps\consol\consol.cpp 7
Upp::AppExecute__(app=13f0288c3) at c:\upp\uppsrc\core\app.cpp 326

invoke_main()
__scrt_common_main_seh()
__scrt_common_main()
mainCRTStartup()
774459bd (kernel32.dll)
```

Subject: Re: [BUG] ToUpper, ToLower,and ToAscii (char*, int) causes AssertFailed
Posted by mirek on Sun, 05 Jun 2016 16:15:28 GMT
View Forum Message <> Reply to Message

There are two reasons why this crashes:

a) default encoding is UTF8 and with UTF8 conversion in general does not guarantee that the ouput number of bytes is the same as input one. Now in this case, it is, but I do not think it is even healthy to support this conversion. You can make it work by adding some 8-bit charset, like

```
 ToUpper(s, 4, CHARSET_WIN1252);
```

b) you are overwriting string literal, which is undefined behaviour. So

```
#include <Core/Core.h>

using namespace Upp;

CONSOLE_APP_MAIN
{
 char s[] = "abcd";
 ToUpper(s, 4, CHARSET_WIN1250);
}
```

Mirek

Subject: Re: [BUG] ToUpper, ToLower,and ToAscii (char*, int) causes AssertFailed
Posted by omari on Mon, 06 Jun 2016 10:56:14 GMT
View Forum Message <> Reply to Message

Yes, there is tow,

 a - the first can be avoided by specifying a charset,

  a possible solution to this is to test if charset is equal to UTF8 (like others ToUpper functions), and do not call s_cset(charset) in this case.

replace : (in file CharSet.cpp, line ~ 2660)

```
void ToUpper(char *t, const char *s, int len, byte charset)
{
 charset = ResolveCharset(charset);
 CharSetData& cs = s_cset(charset);
 const char *lim = s + len;
 while(s < lim)
  *t++ = cs.FromUnicode(UPP::ToUpper(cs.ToUnicode(*s++)));
}
```

by:

```
void ToUpper(char *t, const char *s, int len, byte charset)
{
 charset = ResolveCharset(charset);
 if(charset == CHARSET_UTF8)
 {
  String u = ToUtf8(ToUpper(FromUtf8(s, len)));
  memcpy(t, ~u, len);
  return;
 }

 CharSetData& cs = s_cset(charset);
 const char *lim = s + len;
 while(s < lim)
  *t++ = cs.FromUnicode(UPP::ToUpper(cs.ToUnicode(*s++)));
}
```

b - here a test case for the second BUG:
 the code below produce EXCEPTION_ACCESS_VIOLATION

```
CONSOLE_APP_MAIN
{
 char* s = "abcd";
 ToUpper(s, s, 4, CHARSET_WIN1250);
}
```

 but the code below work as expected:

```
CONSOLE_APP_MAIN
{
 char* s = "abcd";
 StringBuffer r(4);
 ToUpper(r, s, 4, CHARSET_WIN1250);
}
```

then, the bug can be resolved by replacing : (in file CharSet.cpp, line ~ 2630)

```
void ToUpper(char *s, int len, byte charset)
{
 ToUpper(s, s, len, charset);
}
```

by

```
void ToUpper(char *s, int len, byte charset)
{
 StringBuffer r(len);
 ToUpper(r, s, len, charset);
 memcpy(s, r, len);
}
```

Subject: Re: [BUG] ToUpper, ToLower,and ToAscii (char*, int) causes AssertFailed
Posted by mirek on Thu, 09 Jun 2016 12:27:58 GMT
View Forum Message <> Reply to Message

OK, I think I can fix it for UTF8.

However, I am afraid you have not really understood the second issue. You simply cannot write to string literals, which you do...

Mirek

Subject: Re: [BUG] ToUpper, ToLower,and ToAscii (char*, int) causes AssertFailed
Posted by omari on Fri, 10 Jun 2016 09:56:37 GMT
View Forum Message <> Reply to Message

You are right Mirek, the second issue is in my test case.

please consider to fix the UTF8 issue for ToUpper, ToLower,and ToAscii.
