
Subject: Sorting of strings is broken

Posted by [Novo](#) on Thu, 04 Jul 2019 01:46:39 GMT

[View Forum Message](#) <> [Reply to Message](#)

Sorting of strings is broken, or I'm missing something ...

```
static const char* sa[] = {  
"1911",  
"1911 AL Record vs. opponents",  
"1911 AL Record vs. opponents/doc",  
"1911 American League Standings",  
"1911 American League standings",  
"1911 Big 9 football standings",  
"1911 Britannica articles needing updates progress",  
"1911 College Football Composite All-Southerns",  
"1911 College Football Consensus All-Americans",  
"1911 Essendon Bombers premiership players",  
"1911 Essendon premiership players",  
"1911 Havmanden class submarines",  
"1911 Helms Foundation NCAA Men's Basketball All-Americans",  
"1911 MLB season by team",  
"1911 Missouri Valley football standings",  
"1911 NL Record vs. opponents",  
"1911 NL Record vs. opponents/doc",  
"1911 National League Standings",  
"1911 National League standings",  
"1911 Philadelphia Athletics",  
"1911 Princeton Tigers football navbox",  
"1911 RMFAC football standings",  
"1911 SAIAA football standings",  
"1911 SIAA football standings",  
"1911 South Australia State Football Team",  
"1911 Western Conference football standings",  
"1911 college football independents records",  
"1911 college football records",  
"1911 films",  
"1911 in Asian football (AFC)",  
"1911 shipwrecks",  
"1911-12 Big Nine Conference men's basketball standings",  
"1911-12 Big Ten Conference men's basketball standings",  
"1911-12 Essendon Bombers dual premiership players",  
"1911-12 NHA season by team",  
"1911-12 NHA standings",  
"1911-12 Western Conference men's basketball standings",  
"1911-12 in English football",  
"1911-12 in European Football (UEFA)",  
"1911-12 in European football (UEFA)",  
"1911-12 in Scottish football",
```

```
"1911/12 Essendon Bombers dual premiership players",
"1911/12 Essendon dual premiership players",
"1911/doc",
"1911/sandbox",
"1911EB",
"1911s/doc",
"191112 Big Nine Conference men's basketball standings",
"191112 Big Ten Conference men's basketball standings",
"191112 Essendon Bombers dual premiership players",
"191112 NHA season by team",
"191112 NHA standings",
"191112 Western Conference men's basketball standings",
"191112 in English football",
"191112 in European Football (UEFA)",
"191112 in European football (UEFA)",
"191112 in Scottish football",
"1911s",
NULL,
};
```

```
CONSOLE_APP_MAIN
```

```
{
    if (String("1911s/doc") < String("1911s"))
        NEVER();
    Vector<String> sv;
    for (int i = 0; sa[i]; ++i)
        sv.Add(sa[i]);
    Sort(sv);
    RDUMPC(sv);
}
```

Because 1911s is less than 1911s/doc it should be put before it, IMHO.
Default charset is UTF8.

Subject: Re: Sorting of strings is broken
Posted by [mirek](#) on Thu, 04 Jul 2019 08:43:57 GMT
[View Forum Message](#) <> [Reply to Message](#)

```
* /home/cxl/upp.out/se/CLANG.Shared/SortTest 04.07.2019 10:40:39, user: cxl
```

```
sv:
[0] = 1911
[1] = 1911 AL Record vs. opponents
[2] = 1911 AL Record vs. opponents/doc
[3] = 1911 American League Standings
```

[4] = 1911 American League standings
[5] = 1911 Big 9 football standings
[6] = 1911 Britannica articles needing updates progress
[7] = 1911 College Football Composite All-Southerns
[8] = 1911 College Football Consensus All-Americans
[9] = 1911 Essendon Bombers premiership players
[10] = 1911 Essendon premiership players
[11] = 1911 Havmanden class submarines
[12] = 1911 Helms Foundation NCAA Men's Basketball All-Americans
[13] = 1911 MLB season by team
[14] = 1911 Missouri Valley football standings
[15] = 1911 NL Record vs. opponents
[16] = 1911 NL Record vs. opponents/doc
[17] = 1911 National League Standings
[18] = 1911 National League standings
[19] = 1911 Philadelphia Athletics
[20] = 1911 Princeton Tigers football navbox
[21] = 1911 RMFAC football standings
[22] = 1911 SAIAA football standings
[23] = 1911 SIAA football standings
[24] = 1911 South Australia State Football Team
[25] = 1911 Western Conference football standings
[26] = 1911 college football independents records
[27] = 1911 college football records
[28] = 1911 films
[29] = 1911 in Asian football (AFC)
[30] = 1911 shipwrecks
[31] = 1911-12 Big Nine Conference men's basketball standings
[32] = 1911-12 Big Ten Conference men's basketball standings
[33] = 1911-12 Essendon Bombers dual premiership players
[34] = 1911-12 NHA season by team
[35] = 1911-12 NHA standings
[36] = 1911-12 Western Conference men's basketball standings
[37] = 1911-12 in English football
[38] = 1911-12 in European Football (UEFA)
[39] = 1911-12 in European football (UEFA)
[40] = 1911-12 in Scottish football
[41] = 1911/12 Essendon Bombers dual premiership players
[42] = 1911/12 Essendon dual premiership players
[43] = 1911/doc
[44] = 1911/sandbox
[45] = 191112 Big Nine Conference men's basketball standings
[46] = 191112 Big Ten Conference men's basketball standings
[47] = 191112 Essendon Bombers dual premiership players
[48] = 191112 NHA season by team
[49] = 191112 NHA standings
[50] = 191112 Western Conference men's basketball standings
[51] = 191112 in English football

[52] = 191112 in European Football (UEFA)
[53] = 191112 in European football (UEFA)
[54] = 191112 in Scottish football
[55] = 1911EB
[56] = 1911s
[57] = 1911s/doc

I have tested in Win32 both mingw/msc too and I so far I do not see a reason in the code...

Can you post your .log?

Are you getting the bug consistently?

Mirek

Subject: Re: Sorting of strings is broken
Posted by [Novo](#) on Thu, 04 Jul 2019 13:17:43 GMT
[View Forum Message](#) <> [Reply to Message](#)

I'm getting it all the time.
System Ubuntu 19.04 x64.
compiler: clang version 8.0.0-3 (tags/RELEASE_800/final) Target: x86_64-pc-linux-gnu
I discovered this problem when preparing data for another algorithm which requires list of sorted strings as input.

```
$ env | grep LANG
LANGUAGE=en_US
GDM_LANG=en_US
LANG=en_US.UTF-8
```

Two strings below are sorted correctly ...

```
"1911 NL Record vs. opponents",
"1911 NL Record vs. opponents/doc",
```

File Attachments

1) [test.zip](#), downloaded 237 times

Subject: Re: Sorting of strings is broken
Posted by [mirek](#) on Thu, 04 Jul 2019 14:24:29 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you, fixed. The bug was lost when posted in [code] - it was utf-8 related. (That said, I am of course quite frightened and ashamed that something like this was in Core for quite some time...)

The problem was that this failed:

```
String a = "1911s";  
String b = "1911-12 in Scottish football";  
ASSERT(a < b);
```

'-' there is not basic ascii, but some higher utf-8 encoded unicode. So it is 3 byte sequence starting with 0xe3, hence it should be sorted after a. However, in one critical place characters were compared signed, which lead in < transitivity failure, which lead to sorting 1911s after 1911s/doc (you can check svn revision for details).

I am sorry about any inconveniences this has caused.

(Note: the '-' in code section here is 'normal', I had to fix it because otherwise it gets ignored by the forum).

Mirek

Subject: Re: Sorting of strings is broken
Posted by [Novo](#) on Thu, 04 Jul 2019 15:38:51 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thanks a lot! I guess, I always need to post whole project instead of just a code snippet.
