DCMTK - Bug #1143

Data corruption in convertToUTF8() with oficonv in pass-through mode

2024-11-19 14:09 - Marco Eichelberg

Status:	Closed		**Start date:**	2024-11-19
Priority:	Normal		**Due date:**	
Assignee:	Marco Eichelberg		**% Done:**	100%
Category:	Library		**Estimated time:**	1:00 hour
Target version:	3.7.1+			
Module:	oficonv		**Compiler:**	
Operating System:				

Description

The oficonv module offers a pass-through mode in which the libary will simply copy input to output during a character set conversion if input and output encoding are the same. This behaviour is currently disabled by default, but was enabled by default in DCMTK 3.6.8 release. It can be enabled by compiling with the macro DCMTK_ENABLE_ICONV_PASSTHROUGH.

In this mode, the conversion of a DICOM attribute with more than 1024 characters will result in data corruption if the dataset is already encoded in UTF-8 and DcmDataset::convertToUTF8() called:
The first 1024 characters are correct, but afterwards the string is corrupted (the first 1024 characters are repeated again and again; the other characters are lost).

The attached sample program demonstrates the issue. When compiling the the current DCMTK testing or master branch, the test will pass if DCMTK is compiled without DCMTK_ENABLE_ICONV_PASSTHROUGH, and fail otherwise.

Reported 2024-11-19 by DCMTK forum user cschreib, https://forum.dcmtk.org/viewtopic.php?t=5424

History

#1 - 2024-11-19 14:09 - Marco Eichelberg

- File testprg.cc added

#2 - 2024-12-31 08:51 - Marco Eichelberg

- % Done changed from 0 to 100

- Estimated time set to 1:00 h

Closed by commit #ae6352657.

#3 - 2024-12-31 08:52 - Marco Eichelberg

- Status changed from New to Closed

Files

testprg.cc	1.95 KB	2024-11-19	Marco Eichelberg