DCMTK - Conformance #1112

oficonv accepts Shift-JIS multi-byte characters when ISO_IR 13 (JIS X0201) is active

2024-03-07 11:41 - Marco Eichelberg

Status:	Closed	**Start date:**	2024-03-07
Priority:	Normal	**Due date:**	
Assignee:	Marco Eichelberg	**% Done:**	100%
Category:	Library and Apps	**Estimated time:**	2:00 hours
Target version:			
Module:		**Compiler:**	
Operating System:			

Description

Currently, the oficonv implementation of character set conversion will silently accept two-byte Shift-JIS characters when ISO_IR 13 (JIS X0201) is active as the current input character set. Shift-JIS is true superset of JIS X0201 (which by itself is a single-byte character set). DICOM does not support Shift-JIS. It would be preferrable thus if oficonv would report an error in these cases, like the GNU libiconv implementation does.

In the attached sample file,

- attribute (0010,4000) contains the byte sequence "b1\41\5c\1b\24\42\25\22\23\41\21\40", which is a valid ISO 2022 ISO_IR 13/ISO 2022 ISO_IR 87 string.
- attribute (0020,4000) contains the byte sequence "b1\41\5c\83\41\82\60\81\5f". This generates the same Kanji characters though Shift-JIS codes, which are not allowed in DICOM.

Dump the file with dcmdump +U8 shift_jis_example.dcm to demonstrate the effect.

Reported 2024-03-04 by David Gobbi <david.gobbi@gmail.com> and Mathieu Malaterre <mathieu.malaterre@gmail.com>.

History

#1 - 2024-05-25 16:38 - Marco Eichelberg

- Status changed from New to Closed

- % Done changed from 0 to 100

- Estimated time set to 2:00 h

Closed by commit #fb948a289.

Files

shift_jis_example.dcm	5.34 KB	2024-03-07	Marco Eichelberg