

DCMTK - Bug #1113

oficonv creates illegal characters when converting from ISO\_IR 192 to ISO\_IR 101

2024-03-08 13:58 - Marco Eichelberg

Status:	Closed	Start date:	2024-03-08
Priority:	Normal	Due date:	
Assignee:	Marco Eichelberg	% Done:	100%
Category:	Library and Apps	Estimated time:	4:00 hours
Target version:	3.6.9	Compiler:	
Module:	oficonv		
Operating System:			

**Description**

When converting a string that contains greek letters from Unicode (ISO\_IR 192) to Latin-2 (ISO\_IR 101), oficonv does not report an error for characters that cannot be converted, but instead writes a weird byte sequence "bd\b4\2f\34" for each character.

The problem can be demonstrated by converting the attached sample file to Latin-2:

```
dcmconv +C "ISO_IR 101" unicode_with_greek_chars.dcm - | dcmdump - --search PatientName
```

Apparently, Latin-3 (ISO-IR 109) and Latin-4 (ISO-IR 110) are also affected, while Latin-1 (ISO\_IR 100) is not.

Reported 2024-03-07 by Fabian Günther, see <https://forum.dcmtk.org/viewtopic.php?t=5367>

History

#1 - 2024-03-08 14:16 - Marco Eichelberg

Apparently, this is caused by incorrect translation tables, in this case oficonv/datasrc/csmapper/ISO-8859/UCS%ISO-8859-2.src. This is remarkable, because these tables come from the latest FreeBSD source, without any modification.

#2 - 2024-04-07 13:10 - Marco Eichelberg

- File `check_iso8859_mapping_table.pl` added
- Status changed from *New* to *Closed*
- Assignee set to *Marco Eichelberg*
- % Done changed from *0* to *100*
- Estimated time set to *4:00 h*

The iconv mapping tables from Unicode to ISO-8859-2 and ISO-8859-3 contained many incorrect mappings where characters not available in the ISO character set were mapped to four character sequences essentially containing garbage. This has now been fixed by removing all mappings to four-byte character sequences that were not also present in the other ISO-8859 mapping tables.

These issues are also present in the original FreeBSD source from which oficonv has been ported. They have been reported to FreeBSD: [https://bugs.freebsd.org/bugzilla/show\\_bug.cgi?id=278229](https://bugs.freebsd.org/bugzilla/show_bug.cgi?id=278229)

I have written a small (primitive) Perl script for visualizing the mapping tables. This is attached to this issue and may be useful for similar reports in the future.

Closed by commit #5d7495d8c.

Files

unicode_with_greek_chars.dcm	5.25 KB	2024-03-08	Marco Eichelberg
check_iso8859_mapping_table.pl	22.4 KB	2024-04-07	Marco Eichelberg