DCMTK - Bug #684

Check whether delimiters (such as a backslash) are treated correctly during character set conversion

2016-05-27 17:04 - Jörg Riesmeier

Status:	Closed		**Start date:**	2016-05-27
Priority:	Normal		**Due date:**	
Assignee:	Jörg Riesmeier		**% Done:**	100%
Category:	Library		**Estimated time:**	0:00 hour
Target version:	3.6.2			
Module:	dcmdata		**Compiler:**	
Operating System:				

Description

The following excerpt is from DICOM PS 3.5 Section 6.1.2.3:

> The Unicode and GB18030 standards have distinct Yen symbol, backslash, and several forms of reverse solidus. The separator for multi-valued data elements in DICOM is the character valued 05/12 regardless of what glyph is used to enter or display this character. The other reverse solidus characters that have a very similar appearance are not separators. The choice of font can affect the appearance of 05/12 significantly. Multi-byte encoding systems, such as GB18030, GBK and ISO 2022, may generate encodings that contain a byte valued 05/12. Only the character that encodes as a single byte valued 05/12 is a delimiter.

It is not sure whether the DcmSpecificCharacterSet class handles the backslash (and other delimiters) correctly for multi-byte character sets such as ISO 2022 IR 87, GB18030 and GBK.

See sample datasets from Bhuvan Bose (sent by email).

History

#1 - 2016-05-27 17:16 - Jörg Riesmeier

- Description updated

#2 - 2017-01-13 15:29 - Jörg Riesmeier

- Status changed from New to Closed

- % Done changed from 0 to 100

Fixed with commit 9e6362e.