

DCMTK - Feature #1077

Improve handling of extended character sets in "unexpected" places

2023-06-07 13:28 - Marco Eichelberg

Status:	New	Start date:	2023-06-07
Priority:	Normal	Due date:	
Assignee:		% Done:	0%
Category:	Library and Apps	Estimated time:	0:00 hour
Target version:		Compiler:	
Module:	dcmdata		
Operating System:			
Description Some devices seem to permit free-text user input into fields that are stored as DICOM Code Strings. An example has surfaced where an accented ISO 8859/1 character is present in a CS attribute. Currently, DCMTK will silently ignore such fields when performing character set conversion, and dcm2xml will produce an XML file that is technically invalid because it contains characters that are not valid UTF-8. Although the reason is clearly an invalid DICOM file, it might be possible to improve the handling of such errors in DCMTK: <ul style="list-style-type: none">• Attributes that must not use a specific character set should cause a warning during character set conversion if any extended characters (byte values > 127) are present• We should introduce an option that causes character set conversion to fail in this case• We should introduce a command line option in dcm2xml (based on the option mentioned in the previous bullet point) that causes the XML generation to fail in this case Opinions in the team are split whether or not we should offer an option that would cause character set conversion also in such VRs, which might be useful for dcm2xml, but possibly harmful elsewhere.			