
ADAM character conversion of Danish characters

- **Article Type:** General
 - **Product:** Aleph
 - **Product Version:** 21
-

Problem Symptoms:

During indexing of ADAM objects the Danish characters are not correctly converted when using the Danish interface. When using the English interface they are converted correctly.

Cause:

Configuration

Resolution:

Here is how the indexing process of objects works:

First, the source document (DOC or PDF) is converted to text by using the routine specified in Z403-CHARACTER-SET or ADAM-INDEX-CHAR-SET in xxx01/tab/tab100, by using Oracle Text tools. This has nothing to do with the current issue.

Then, the conversion program creates an xml out of the text file created in the previous stage.

In this process, the program performs 3 conversions of the text (according to the setup of \$alephe_unicode/tab_character_conversion_line):

- 1.The routine specified in Z403-CHARACTER-SET or, if it is empty, the routine specified by ADAM-INDEX-CHAR-SET in Tab100
- 2.WORD-FIX
- 3.WORD-LNG, e.g. WORD-ENG or WORD-DAN. This is determined according to the interface language (ENG, DAN, etc.).

Stage 3 is the reason for the difference in the final outcome when using ENG or DAN, since you have the following setup in \$alephe_unicode/tab_character_conversion_line:

```
WORD-DAN ##### # line_utf2line_utf unicode_to_word_dan
```

In \$alephe_unicode/unicode_to_word_dan you have the following conversion for ? and ?

```
00F8 007C #LATIN SMALL LETTER O WITH STROKE  
00E6 007B #LATIN SMALL LETTER AE
```

In \$alephe_unicode/unicode_to_word_gen (used in the English interface) you have the following conversion for ? and ?

```
00F8 006F #LATIN SMALL LETTER O WITH STROKE  
00E6 0061 0065 #LATIN SMALL LETTER AE
```

Category: ADAM (500)

- **Article last edited:** 10/8/2013