
file-20: Errors for non-ascii characters

- **Article Type:** General
- **Product:** Aleph
- **Product Version:** 20

Problem Symptoms:

Errors when non-ascii characters (for foreign student names/addresses) are present when attempting to load plif (p_file_20) UTF-8-character-set files in Aleph Sequential format.

Cause:

The Aleph Sequential format is dealing with a limited field length; a foreign character may require more than one byte per character, and the field length is fixed.

Resolution:

1. Fix the input file for every special character. In this case: if you move the positions 5 characters to the left, the input is possible again
2. Change the input file to XML structure, thus avoiding fixed field length.

The second is the recommended solution.

Additional Information

Error messages:

```
Error (b_file_20_chk_edit_record) : Errors in Z303. Record No 000001
Can not insert record. Match found.
Error (b_file_20_chk_edit_record) : Errors in Z304. Record No 000001
Can not insert. Match found.
Error (b_file_20_chk_edit_record) : Errors in Z304. Record No 000001
Can not insert. Match found.
Error (b_file_20_chk_edit_record) : Errors in Z305. Record No 000001
Unsupported action
```

From the file-20 Help (as seen in \$aleph_root/pc_b_help_eng/p-file-20.html):

```
<li><b>Character Conversion</b>
```

<p> If the Input file is created in non UTF-8 use 'Character Conversion' parameter to convert to UTF-8.

<p>

Enter here the character conversion routine identifier to be used. The routine identifier is defined in

$alephe_unicode/tab_character_conversion_line - col.1. Examples:

8859_1_TO_UTF, MAB_TO_UTF.

Note: Character Conversion is not applicable when the input file is in XML format.

To handle XML input file with non UTF-8 characters, the input file should include the encoding type declaration at the top of the file (this should match to the encoding that the file was saved in).

Here are examples of encoding declarations:

<li type=circle>in the case of West European characters such as tilde and umlaut - the input file should include the encoding type as follows: <?xml version= "1.0" encoding= "ISO-8859-1"?>.

<li type=circle>in the case of Arabic characters - the input file should include the encoding type as follows: <?xml version= "1.0" encoding= "ISO-8859-6"?>.

<li type=circle>in the case of Hebrew characters - the input file should include the encoding type as follows: <?xml version= "1.0" encoding= "ISO-8859-8"?>.

-
- **Article last edited:** 30-May-2016