
Reading diacritics in MARC input files; Textpad

- **Article Type:** General
 - **Product:** Aleph
 - **Product Version:** 14.2
-

Description:

You want to see if a MARC input file has the correct representations for diacritics; "vi" just shows garbage for diacritics.

Resolution:

The standard MARC representations for characters can be seen at <http://lcweb2.loc.gov/cocoon/codetables/1.html> .

These are in hex. So what you need to do is to get the hex values for the diacritic characters in the input file.

To isolate just one problem record from a large input file do:

```
grep 000123456 xxxxx > rec000123456
```

where 000123456 is the ALEPH record# and xxxxx is the input file. A file rec000123456 with just the 000123456 record in it will be created.

This unix command will copy the input file as a hex file:

```
od -x rec000123456 > 000123456odx
```

But it's hard to read this. What I have learned to do as an additional step is:.

```
od -c rec000123456 > 000123456odc
```

This produces a file in which regular letters and numbers display normally, but the diacritics and other non-displayables have octal (or escape) values. Most importantly, the line numbers in this display are the same as those in the "od -x" output. So once you find the character in the "000123456odc", you can easily go to the same line in the "000123456odx" and find the hex representation.

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