



How to Normalize Headings

Release 505_12.4

Last update: 21 September, 2000

Table of Contents

General:	3
Display Text:	3
Filing Text:	4
Non-filing Characters.....	4
Filing Routines.....	4
Character Conversion.....	5
Normalization:	5
Appendix 1 – Filing Routines	6
Appendix 2 – Character Conversion	6

General:

The oracle table for headings – z01 - contains two textual fields – a filing text field and a display text field. Data for the display text field comes directly from the MARC record. Data for the filing text field, on the other hand, is processed in three ways; non-filing characters are removed, a filing routine is used and characters undergo character conversion. A difference in the contents of either of these fields will result in the creation of a new heading. Thus, when two records would generate headings that have identical values in the filing text field but different values in the display text field – there will be two headings and not one. Differences between fields in the MARC record as seemingly insignificant as sub-field value, additional spaces, or differences in punctuation will result in different display texts and thus in multiple headings. Frequently, headings that appear to be identical are not – minor differences in one of the two fields are present but easily overlooked.

Display Text:

Data for the display text field is taken directly from the MARC record [via tab11].

Example One:

MARC record
24510 L \$\$aJournal of blacks in higher education\$\$h[computer file]

Tab11
11 A 245## TTL -ch6 2

Heading
02 z01_display_text\$\$aJournal of blacks in higher education

Example Two:

MARC record
7760 L \$\$tJournal of blacks in higher education\$\$x1077-
3711\$\$w(DLC) 93664388\$\$w(OCOLC)28945968

Tab11
11 A 776## TTL ast

Heading
02 z01_display_text\$\$tJournal of blacks in higher education

Filing Text:

Data for the Filing text field is processed in the following three ways:

- 38 non-filing characters are removed [non-filing character is set in tab11 col. 9]
- 39 text goes through the appropriate filing routine [filing routine is listed in tab00.eng col. 5]
- 40 characters go through character conversion [char_conv.A]

Non-filing Characters

When non-filing characters are removed, this is done using indicator values for non-filing characters. The indexing table – tab11 – includes a column where the appropriate indicator for each field is flagged: the second indicator for 245, and the first for 130.

Example One:

MARC record
24504 L \$\$aThe Journal of blacks in higher education.

Will have a heading with the following three values

Heading
03 filing_textjournal of blacks in higher education
02 z01_non_filing_char04
02 z01_display_text\$\$aThe Journal of blacks in higher education

Filing Routines

In version 14 and higher, filing routines can be customized. In earlier versions, customers had to choose between ALEPH's eight existing routines. Filing routines manipulate text by compressing spaces, deleting sub-fields, changing case, and adding leading zeroes for numeric filing, among other things. There are routines for filing LC call numbers, Dewey call numbers, and text. One routine, AM, files Mc as Mac in certain specific cases. There is a complete list of filing routines and what they do at the end of this document.

Example One:

MARC record

Tab11
11 A 700## AUT a

Tab00.eng
H AUT ACC AM 00 0000 AUT Author

Heading

03 filing_textmaccain john 1922
02 z01_display_text\$\$aMcCain, John\$\$d1922-

Character Conversion

Character conversion tables convert specific ASCII or ANSEL values for filing purposes. Characters can be filed as themselves, or they can be ignored, converted to a space or to one or more different characters. The standard ALEPH character conversion table for headings is included at the end of this document. One example of what the character conversion table can do can be seen in the *umlaut* – ASCII value 232. Most North American libraries want *umlauts* to be ignored for purposes of filing. To achieve this, 232 would be set to file as 000. In some European libraries, an ampersand - & - files as “and”. To achieve this, the table files 038, the ANSEL value for &, as 97 + 110 + 106.

Example One:

MARC record
24510 |a Über die Mundbildung der Wirbeltiere

Char_conv.A table
38 000

Heading
03 filing_textüber die mundbildung der wirbeltiere
02 z01_display_text\$\$aÜber die Mundbildung der Wirbeltiere

Example Two:

MARC record
24510 \$\$aReligion & belief

Char_conv.A
38 097+110+100

Heading
02 z01_display_text\$\$aReligion & belief
03 filing_textreligion and belief

Normalization:

In order to achieve full normalization of headings, both the display text and the filing text have to be changed. Changing one without the other will result in correct filing but duplicate displays. A normalization routine was written for Notre Dame. This routine uses ALEPH’s expand functionality to generate a virtual field which is in uppercase only; in addition, internal punctuation and initial articles have been removed. This virtual field is indexed in place of the MARC fields.

Example One:

MARC field
24514 L \$\$aThe illustrated Faerie queene

Virtual field
TIT5 L \$\$aILLUSTRATED FAERIE QUEENE

Heading
filing_textillustrated faerie queene
02 z01_display_text\$\$aILLUSTRATED FAERIE QUEENE

Appendix 1 – Filing Routines

- A – packs spaces, ^ is changed to a blank
- AB – same as A, but replaces \$\$ of sub-fields with spaces
- A3 – same as A, but leading zeroes are added to numbers to make a seven-digit number
- AM – same as A, but Mc at beginning of text is opened to mac
- F – for filing Library of Congress call numbers
- F1 – for filing Library of Congress call numbers – used in North America
- D – deletes all hyphens; does not use character conversion table
- Z1 – for linking private routines

Appendix 2 – Character Conversion

Col. 1 has the ASCII value of the character contained in the record. Col. 2 has the ASCCI value it should be filed as.

AL 001	000	AL 016	000
AL 002	000	AL 017	000
AL 003	000	AL 018	000
AL 004	000	AL 019	000
AL 005	000	AL 020	000
AL 006	000	AL 021	000
AL 007	000	AL 022	000
AL 008	000	AL 023	000
AL 009	000	AL 024	000
AL 010	000	AL 025	000
AL 011	000	AL 026	000
AL 012	000	AL 027	000
AL 013	000	AL 028	000
AL 014	000	AL 029	000
AL 015	000	AL 030	000

AL 031	000	AL 079	111
AL 032	032	AL 080	112
AL 033	000	AL 081	113
AL 034	000	AL 082	114
AL 035	000	AL 083	115
AL 036	032	AL 084	116
AL 037	032	AL 085	117
AL 038	038	AL 086	118
AL 039	000	AL 087	119
AL 040	032	AL 088	120
AL 041	032	AL 089	121
AL 042	000	AL 090	122
AL 043	043	AL 091	032
AL 044	032	AL 092	032
AL 045	032	AL 093	032
AL 046	032	AL 094	032
AL 047	032	AL 095	032
AL 048	048	AL 096	000
AL 049	049	AL 123	032
AL 050	050	AL 124	032
AL 051	051	AL 125	032
AL 052	052	AL 126	032
AL 053	053	AL 127	000
AL 054	054	AL 128	000
AL 055	055	AL 129	000
AL 056	056	AL 130	000
AL 057	057	AL 131	000
AL 058	032	AL 132	000
AL 059	032	AL 133	000
AL 060	032	AL 134	000
AL 061	032	AL 135	000
AL 062	032	AL 136	000
AL 063	032	AL 137	000
AL 064	032	AL 138	000
AL 065	097	AL 139	000
AL 066	098	AL 140	000
AL 067	099	AL 141	000
AL 068	100	AL 142	000
AL 069	101	AL 143	000
AL 070	102	AL 144	000
AL 071	103	AL 145	000
AL 072	104	AL 146	000
AL 073	105	AL 147	000
AL 074	106	AL 148	000
AL 075	107	AL 149	000
AL 076	108	AL 150	000
AL 077	109	AL 151	000
AL 078	110	AL 152	000

AL 153	000	AL 194	000
AL 154	000	AL 195	000
AL 155	000	AL 196	000
AL 156	000	AL 197	000
AL 157	000	AL 198	000
AL 158	000	AL 222+115	115+115
AL 159	000	AL 225	000
AL 160	000	AL 226	000
AL 161	108	AL 227	000
AL 162	124	AL 228	000
AL 163	100	AL 229	000
AL 164	116	AL 230	000
AL 165	123	AL 231	000
AL 166	111+101	AL 232	000
AL 167	000	AL 233	000
AL 168	000	AL 234	000
AL 169	000	AL 234+065	125
AL 170	000	AL 234+097	125
AL 171	000	AL 235	000
AL 172	111	AL 236	000
AL 173	117	AL 237	000
AL 174	000	AL 238	000
AL 176	000	AL 239	000
AL 177	108	AL 240	000
AL 178	124	AL 241	000
AL 179	100	AL 242	000
AL 180	116	AL 243	000
AL 181	123	AL 244	000
AL 182	111+101	AL 245	000
AL 183	000	AL 246	000
AL 184	105	AL 247	000
AL 185	032	AL 248	000
AL 186	116	AL 249	000
AL 188	111	AL 250	000
AL 189	117	AL 251	000
AL 192	000	AL 254	000
AL 193	108		