

From new version 16 “Indexing” Web Guide:

In 16.1 we have a new WEB Guide, named "indexing", which pulls together information that was contained in OPAC and in CAT in one place. There are the relevant sections on normalization; I cannot bring up the web page, so here are the sections from the guide files on ram19. I have removed most of the html formatting to make it more readable.

3.1 FILING OF HEADINGS

Headings are filed (organized) in the browse list according to the "filing text" of the heading.

The filing text is built in three steps, based on the field text:

- Display text (usually data taken from the bibliographic record without the final punctuation.
- Normalization of the display text.
- Normalized text to filing text.

Conversion takes place according to the procedures defined in the tab_filing table of the library's tab directory, and in the tab_character_conversion_line table of the alephe/unicode directory.

3.1.1 Normalization

Normalization refers to the process whereby diacritics, most punctuation marks, special characters, and case differences are stripped from access fields (headings) for the purpose of determining the uniqueness of headings.

The normalized form of the headings is built according to the rules defined in the library's tab_filing table:

To view the whole list of filing routines and an explanation of normalization routines refer to the Sorting and Word-breaking (Sorting Headings and Indexes) section.

The aims of normalization are:

- To treat the same headings alike.
- To ensure that each unique heading is stored only once in the headings
- To distinguish headings that are different by means of unique identifiers.

=====

8.1 SORTING HEADINGS AND INDEXES</p>

Each Heading (Z01), Index (Z11) and Sort key (Z101) record has what are termed "filing keys" or "filing text". This is the form of the heading or index term for filing (sorting) purposes. The rules that govern the values of the filing keys are set in the library's tab_filing table.

Headings have an additional related feature, called normalization. Normalization refers to the process whereby diacritics, most punctuation, special characters, and case differences are stripped from headings. This is in order to neutralize slight differences. This is important in the headings index where each unique heading is stored only once.

Normalization routines are defined in the tab_filing table together with filing routines and display text routines.

When a new heading is added to the database, if there is already another heading with the same normalized text (Z01-NORMALIZED-TEXT),

The filing key of the headings is built in two stages:

- Display text (Z01-DISPLAY-TEXT) to normalized text (Z01-NORMALIZED-TEXT).
- Normalized text (Z01-NORMALIZED-TEXT) to filing text (Z01-REC-KEY).

For this reason, when creating the filing form of the heading, it is not necessary to perform routines that have already been performed in order to create the normalized text of the heading.

The library's tab/tab_filing table defines the normalization and filing routines that are used. The order of the subroutines within a routine is important; for example, you cannot relate to a subfield code if you have

previously set "del subfield code". The table includes the following columns:

Column 1 - Identifier

Contains the two-digit identifier of the filing routine. This identifier is used in column 5 of the tab00.lng table (for headings and direct indexes) and column 3 of the tab01.lng table (for sort keys).

Column 2 - Routine Usage

This column is relevant only for headings (Z01). It is not relevant for direct indexes (Z11) or for sort keys (Z101). It defines the usage of the routine. The following are the available options:

D = Display text procedures

N = Normalized text procedures

F = Filing text procedures

Note that the routines defined in the tab_filing table require an 'F' section, so even if this section is not needed (because the lines for 'N' suffice), an 'F' line still needs to be present with "no" as the filing procedure.

Column 3 - Procedure

This column contains the name of the filing or normalization procedure.

Column 4 - Parameters

Parameters for the filing/normalization procedure (when relevant). If the parameters are characters, and a character is out of the ASCII range, type this character in Unicode notation.

The following are the available filing and normalization procedures:

etc., etc.

-----Original Message-----

From: Jerry Specht [mailto:jerry.specht@exlibris-usa.com]

Sent: Monday, May 19, 2003 6:18 AM

To: Judy Levi

Subject: 14.2 vs. 15.2 headings

Judy, I was wondering if you know of a document describing the differences

in the Headings between 15.2 compared to 14.2 -- specifically:

- * normalization (the role of Z01-NORMALIZED-TEXT)

- * z0101

I found some 15.2 rep_ver's but I was wondering what else there might be.