USER DOCUMENTATION

How to Configure an ALEPH Union Catalog

Ex Libris
© Ex Libris Ltd., 2002
Release 15.2 and later
Last Update: February 20, 2002
Table of Contents

1 INTRODUCTION ............................................................................................ 3

2 ORACLE TABLES .......................................................................................... 3

3 SERVER / SHARED TABLES ........................................................................ 4

4 LIBRARY-SPECIFIC TABLES...................................................................... 5
   4.1 Defining a Library as a Union Catalog.................................................... 5
   4.2 Assigning Weights for Record Matching ................................................ 5
   4.3 List of Common Serial Titles................................................................. 6
   4.4 Selecting the Preferred Record ............................................................. 7
   4.5 Merged Display..................................................................................... 7
   4.6 Union Catalog Indexes ......................................................................... 8

5 UNION CATALOG MANAGEMENT.......................................................... 10
   5.1 Converting and Loading Records.......................................................... 10
      5.1.1 Loading Records.............................................................................. 10
      5.1.2 Building Record Equivalencies (Z120) ............................................. 10
1 Introduction

The Classic Union Catalog is a bibliographic catalog that contains the records and holdings from multiple contributing institutions. Each of these institutions has their own bibliographic record and holdings. Therefore the Union Catalog inevitably contains duplicate records for any given title. The ALEPH Union Catalog uses a sophisticated algorithm, developed in conjunction with MELVYL (California Digital Libraries) to detect duplicate records and merge them for OPAC searching and display. The process of identifying duplicate records is specific to the Union Catalog product and consists of four steps; candidate selection, duplicate detection / equivalency building, preferred record selection, and merged display. Each of these functions is controlled by a combination of programs and configuration tables. In this document we will look at the Union Catalog specific configuration tables and processes as well as the Union Catalog specific settings included in standard ALEPH configuration tables. A more detailed description of the equivalency algorithm can be found in the Equivalency Algorithm document. There are two additional documents available: How to Load Records Into a Union Catalog, which describes the process of loading records into a Union Catalog and Using the (Union Catalog) Location Table.

2 Oracle Tables

There is a single Oracle table that is specific to the Union Catalog – the Z120. The Z120 table stores record equivalencies – that is, it stores a list of system numbers of records that the equivalency algorithm determines to be equivalent. The data in this Oracle table is used to create de-duplicated search results and merged display. The Z120 table contains the following data elements: system number, total number of equivalent records, system numbers of all equivalent records, system number of preferred record, update flag. Each record in the Union Catalog has its own Z120 table, and all equivalent records share the same preferred record. Z120 records are built after the initial data load and then updated each time any one of the equivalent records is updated.

```sql
02 Z120-REC-KEY.
   03 Z120-DOC-NUMBER PICTURE 9(9).
9(9).
02 Z120-REC-KEY-1.
   03 Z120-PREFERRED-DOC-NUMBER PICTURE 9(9).
9(9).
02 Z120-UPDATE-FLAG PICTURE X(1).
X(1).
02 Z120-SAME-NO-LINES PICTURE 9(3).
9(3).
02 Z120-SAME OCCURS 100.
   03 Z120-SAME-DOC-NUMBER PICTURE 9(9).
```
3 Server / Shared Tables

ALEPH supports a number of different Union Catalog models. The union_global_param table sets the particular program and configuration table that should be used for each of the four Union Catalog functions; candidate selection, record matching, preferred record selection, merged display and normalization. The table is located in the alephe_tab directory:

```
<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>USM90 B</td>
<td>candidate_prog</td>
<td>union_candidate_cdl</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>union_match_cdl</td>
<td></td>
</tr>
<tr>
<td></td>
<td>USM90 B</td>
<td>preferred_prog</td>
<td>union_preferred_cdl</td>
</tr>
<tr>
<td></td>
<td>merge_prog</td>
<td>union_merge_aleph</td>
<td></td>
</tr>
<tr>
<td></td>
<td>normalize_prog</td>
<td>union_normalize_cdl</td>
<td></td>
</tr>
</tbody>
</table>
```

Key to the table:

- **Column 1 – Library Code:**
  Library code of the Union Catalog database (for example, USM90).

- **Column 2 – Context:**
  Values are ‘T’ and ‘B’, T for Testing and B for Batch.

- **Column 3 and 4 – Code and Program:**
  There are five codes, one for each of the four Union Catalog functions and a fifth for normalization.

  - **candidate_prog:** This parameter controls the selection of the program used for selecting candidate records. There are currently three programs; only one of the three programs should be selected.

  - **union_candidate_cdl:** A pool of candidate records is created by retrieving records that match on ISSN, ISBN, LCCN and Short title.

  - **match_prog:** This parameter controls the selection of a program/algorithm for building record equivalencies. There are six programs to choose from.

  - **union_match_cdl:** Use the CDL record merging algorithm for identifying equivalent records. The program consults three configuration tables, tab_cdl_mo_weights, tab_cdl_se_weights and tab_com_tit_cdl.

  - **preferred_prog:** This parameter is used for defining the process for selecting the preferred record from a set of equivalent records. There are two programs to choose from.
• **union_preferred_cdl** – each record in the set is assigned a record weight, the record with the highest weight is the preferred record. Weights are assigned in the table `union_preferred`.

• **merge_prog**: controls the merged display of records. There is only one program for merge, it is `union_merge_aleph` and it uses the table `tab_merge_union`.

• **normalize_prog**: controls the selection of normalization routines for the candidate and matching phases of the union catalog process. There is currently only one choice of programs, `union_normalize_cdl`.

### 4 Library-Specific Tables

The remaining tables and table settings are all specific to the union catalog library, XXX90.

#### 4.1 Defining a Library as a Union Catalog

To define an ALEPH library as a Union Catalog set `tab10` flag 51 to Yes.

```
  51 Y YN              TAB10-UNION-LIBRARY
```

#### 4.2 Assigning Weights for Record Matching

If `union_match_cdl` has been selected as the record matching program, then two configuration tables must be present and set up. They are `tab_cdl_mo_weights` and `tab_cdl_se_weights`. These tables assign a threshold value for achieving record equivalencies as well as weights for each specific match or non-match in the equivalency algorithm.

The first table, `tab_cdl_mo_weights` establishes weights for merging books/print material. The second table, `tab_cdl_se_weights` establishes weights for merging serials. Note that the text in the first column is fixed. The numeric value in the second column should always be expressed as a positive value. The text indicates which values are subtracted and which values are added to the overall weight of the record. Both tables are located in the `data_tab` directory of the Union Catalog database [usm90/tab]. Record equivalency building processes need to be rerun after changes are made to either of these tables.
4.3 List of Common Serial Titles

The list of common serial titles is consulted during the matching stage of the cdl equivalency building process. Title matches in serial records that have titles that are considered “common” receive significantly fewer points than title matches for other serial records. The table lists “common” titles in a given database or union catalog. If the contents of the table change, then p-union-02 should be rerun.

ANNUAL REPORT
ANNUAL REPORT FOR
ANNUAL REPORT FOR THE FISCAL YEAR ENDED MARCH
BIENNIAL REPORT
BULLETIN
CALENDER
CATALOGUE
CIRCULAR
4.4 Selecting the Preferred Record

If union_preferred_cdl was selected as the preferred program in union_global_param, then the union_preferred table located in the data_tab directory of the Union Catalog database is used. This table establishes the basis for selecting a preferred record from a set of equivalent record by assigning points for field presence, and/or subfield or fixed field values. After each record in a set of equivalent records gets a weight, the record with the greatest weight becomes the preferred record.

<table>
<thead>
<tr>
<th>Field Tag</th>
<th>Subfield Position and Count</th>
<th>Operator</th>
<th>Subfield or Fixed Field Value</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDR</td>
<td>F05-01</td>
<td>EQUAL</td>
<td>d</td>
<td>-10</td>
</tr>
<tr>
<td>LDR</td>
<td>F17-01</td>
<td>NOT-EQUAL</td>
<td>1, 2, 3, 4, 5, 7, 8, u, z</td>
<td>001</td>
</tr>
<tr>
<td>040</td>
<td>a</td>
<td>EQUAL</td>
<td>DLC</td>
<td>003</td>
</tr>
<tr>
<td>100##</td>
<td></td>
<td>PRESENT</td>
<td></td>
<td>001</td>
</tr>
<tr>
<td>110##</td>
<td></td>
<td>PRESENT</td>
<td></td>
<td>001</td>
</tr>
<tr>
<td>111##</td>
<td></td>
<td>PRESENT</td>
<td></td>
<td>001</td>
</tr>
<tr>
<td>130##</td>
<td></td>
<td>PRESENT</td>
<td></td>
<td>001</td>
</tr>
<tr>
<td>24###</td>
<td></td>
<td>PRESENT</td>
<td></td>
<td>001</td>
</tr>
<tr>
<td>6####</td>
<td></td>
<td>PRESENT</td>
<td></td>
<td>001</td>
</tr>
<tr>
<td>700##</td>
<td></td>
<td>PRESENT</td>
<td></td>
<td>001</td>
</tr>
<tr>
<td>710##</td>
<td></td>
<td>PRESENT</td>
<td></td>
<td>001</td>
</tr>
<tr>
<td>711##</td>
<td></td>
<td>PRESENT</td>
<td></td>
<td>001</td>
</tr>
<tr>
<td>730##</td>
<td></td>
<td>PRESENT</td>
<td></td>
<td>001</td>
</tr>
<tr>
<td>800##</td>
<td></td>
<td>PRESENT</td>
<td></td>
<td>001</td>
</tr>
<tr>
<td>810##</td>
<td></td>
<td>PRESENT</td>
<td></td>
<td>001</td>
</tr>
<tr>
<td>830##</td>
<td></td>
<td>PRESENT</td>
<td></td>
<td>001</td>
</tr>
<tr>
<td>880##</td>
<td></td>
<td>PRESENT</td>
<td></td>
<td>001</td>
</tr>
</tbody>
</table>

Key to the table:

Column 1 – field tag

Column 2 – Subfield or fixed field position and count

Column 3 – Operator (Equal, Not-equal, Present):

Column 4 – Subfield or fixed field value. If there are multiple values, then separate them with a comma.

Column 5 - Points

4.5 Merged Display

If union_merge_aleph is selected as the merge_program in the union_global_param table, then the system uses the configuration table tab_merge_union located in the data_tab directory of the union catalog library. This table controls record merging for display and allows you to set, on a tag-by-tag basis, which fields display from the preferred and non-preferred records.
### tab_merge_union

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>90 1 Y ###</td>
<td>90 1 N SID##</td>
<td>90 1 N 852##</td>
<td>90 1 N 856##</td>
</tr>
<tr>
<td>90 2 Y SID##</td>
<td>90 2 Y 852##</td>
<td>90 2 Y 856##</td>
<td></td>
</tr>
</tbody>
</table>

**Key to the table:**

- **Column 1 – Merge Set** – will always be 90 for Union Catalog merge
- **Column 2 – Merging direction** (1 refers to preferred record, 2 to each equivalent record)
- **Column 3 – Action** (Y, N, C Y – retain field, N – delete field, C – retain field if its original)
- **Column 4 – Field tag.**
- **Column 5 – Tag** (This column can also be used to define sub-field and contents to match on. Example: 01 2 Y 590##.5,*abc*)

**Note:** It is not sufficient to configure union_global_param and tab_merge_union. In addition, the program expand_doc_merge must be called in tab_expand each time a merged display is required.

For example, a Union Catalog site might want to have a merged display in the Web OPAC and Z39.50 but not in the GUI. If this is the case, tab_expand would be set up as follows:

```
WEB-FULL   expand_doc_merge
Z39_SERVER     expand_doc_merge
```

### 4.6 Union Catalog Indexes

Libraries that are Union Catalogs and that use the cdl merge algorithm must have the following indexes and filing routines defined in their tables.
tab11

11 I 010##  010
11 I 020    020 a
11 I z020  020
11 I 022##  022 a
11 I y022  022
11 I z022  022
11 I SID##  SID
11 I 245##  NTL abnp

tab11_acc

245## SE    STL abnp  2

tab00.lng

H SID    IND  01 00  0000 SID  SID
H NTL    IND  90 00  0000 NTL  Normal Title
H STL    ACC  11 00  0000 STL  Serial title

tab_filing:

90  del_subfield
90  to_lower
90  to_blank  !@$%^&*()_+-={]}":<>?,./~`
90  char_conv  FILING-KEY-01
90  compress_blank
90  first_25

92  compress        -
92  compress_blank
92  non_numeric

91  del_subfield
91  to_upper
91  suppress
91  numbers
91  compress     ,
91  to_blank     !@$%^&*()_+-={]}":<>?,./~`
91  expand_num
91  non_filing
5 Union Catalog Management

5.1 Converting and Loading Records

5.1.1 Loading Records
Initial load uses the standard ALEPH loader- p_manage_18. Records in an ALEPH Union Catalog must include an SID field (System Identifier). This field is the equivalent of the control number. It identifies both the contributing institution and the record system number in that institution's catalog. This field is used to identify records for subsequent update. The SID field should be assigned prior to running b-manage-18. The process that assigns the SID is called p_union_fix_doc and the parameters are: library, input file, output file, SID program.

After running p_union_fix, load the Union Catalog records using p-manage-18.

5.1.2 Building Record Equivalencies (Z120)
The Z120 table is the Oracle table that stores record equivalencies. Each record has its own Z120 record. The Z120 table contains the following fields: system number, system number of preferred record, update flag, number of equivalent records, system number of preferred records, system number of all equivalent records.

There are three processes for building and maintaining record equivalencies.

- **p_union_01** – run after initial conversion and data load. Builds empty Z120 records for each bibliographic record. Parameters are Database names.

- **p_union_02** – run after database has been indexed. Populates empty Z120 records. Can be run with multiple processes. Parameters are Database, start number, end number, rebuild links, number of processes, batch/test. This process locks the library.

- **p_union_04** – ongoing, rebuilds record equivalencies for records that have Z120 records flagged N for update. Z120s can be flagged N by the p_union-03 process. This process does not lock the library.