



MetaLib Version 4.2 (345-365)

Release Notes

April 2008

Confidential Information

The information herein is the property of Ex Libris Ltd. or its affiliates and any misuse or abuse will result in economic loss. DO NOT COPY UNLESS YOU HAVE BEEN GIVEN SPECIFIC WRITTEN AUTHORIZATION FROM EX LIBRIS LTD.

This document is provided for limited and restricted purposes in accordance with a binding contract with Ex Libris Ltd. or an affiliate. The information herein includes trade secrets and is confidential.

Disclaimer

The information in this document will be subject to periodic change and updating. Please confirm that you have the most current documentation. There are no warranties of any kind, express or implied, provided in this documentation, other than those expressly agreed upon in the applicable Ex Libris contract.

Any references in this document to non-Ex Libris Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this Ex Libris product and Ex Libris has no liability for materials on those Web sites.

Copyright Ex Libris Limited, 2008. All rights reserved.
Documentation produced January 2008

Document version 1.0

Web address: <http://www.exlibrisgroup.com>.

Table Contents

| | | |
|-----------|--|----------|
| 1. | Introduction | 1 |
| 1.1. | Purpose and Scope | 1 |
| 1.1.1. | <i>Version Name and Number</i> | 1 |
| 1.1.2. | <i>Version Highlights</i> | 1 |
| 1.2. | Related Documents and References | 2 |

| | | |
|-----------|---|----------|
| 2. | Service Pack Content Summary | 3 |
|-----------|---|----------|

| | | |
|-----------|--|----------|
| 3. | Functional Enhancements Detailed Description | 6 |
| 3.1. | CJK Enhancements..... | 6 |
| 3.2. | New Fields in the IRD: CJK Display Name Language and CJK Transliterated Display Name | 6 |
| 3.3. | CJK Display Name Language Added to Sort Key | 7 |
| 3.4. | CJK A to Z Lists | 8 |
| 3.5. | “Starts with” Find Database Option..... | 10 |
| 3.6. | Normalization from Traditional to Simplified Chinese..... | 10 |
| 3.7. | Japanese Support..... | 11 |
| 3.7.1. | <i>Indexing and Search</i> | 11 |
| 3.7.2. | <i>Normalization of Hiragana and Katakana</i> | 11 |

| | | |
|-----------|--|-----------|
| 4. | Detailed Implementation Instructions | 12 |
| 4.1. | Re-indexing of DAT01 | 12 |
| 4.2. | HTML changes and Text to translate (for languages other than English) | 12 |
| 4.3. | Complementary Table to tab_character_conversion_line – tab_character_codes13 | |

1. Introduction

1.1. Purpose and Scope

This document describes the enhancements and the functional and technical changes that are included in this minor version release and provides implementation instructions where appropriate.

Note: This document details the changes included in this minor version release only.

1.1.1. Version Name and Number

MetaLib version 4.2.0 (345-365), April 2008.

1.1.2. Version Highlights

This minor version release further enhances CJK support in MetaLib version 4, providing new enhancements as well as bug fixes. There are no items requiring mandatory implementation actions. However, MetaLib customers that provide a non-English user interface will need to translate some new labels. In addition, if you made a change to the A-Z list page, you will need to restore the changes since this Service Pack replaces this file (see below). The delivery mechanism for this minor version release is the standard MetaLib Service Pack (SP) mechanism, which automatically applies all software changes related to these items where possible.

MetaLib 4.2.0 is the second of three planned minor version releases for the current development cycle through 2008. The purpose of these minor version releases is to continue reducing MetaLib's total cost of ownership (TCO) for our customers by streamlining product implementation and maintenance workflows, incorporating customer-requested enhancements, and addressing competitive influences in the market.

MetaLib 4.2.0 includes the following items to enhance CJK support:

- Correct CJK segmentation in A-Z database list – enhances locating CJK databases in the Find Database and MetaSearch modules
- New CJK display – enhanced browsing in the Find Database module
- Enhanced CJK clustering – language-specific stop word lists referenced by the embedded Vivisimo clustering engine enhance topic clusters generated for CJK and other non-English search results

As mentioned above, there are no items requiring mandatory implementation actions, however, note the following changes:

- Due to CJK related enhancements, the Service Pack includes automatic re-indexing of the DAT01 database. For more details, refer to section [4.1](#).
- Because of the changes to the Find Databases module, this service pack replaces the HTML file for the A-Z database list (`./insNN/www_v_eng/find-db-1-body-title`). If you made any local changes to this file, you will need to restore them. The minor version release copies the original file into the `./insNN/www_v_eng/find-db-1-body-title.sp.365` file, where 365 is the SP level.
- As mentioned previously, some new labels must be translated if you provide non-English user interfaces for MetaLib. For more details, refer to section [4.2](#).
- Item number 355 includes a new character conversion table. Customers who previously changed the character set on `tab_character_conversion_line` should refer to section [4.3](#).

In addition, MetaLib 4.2.0 includes the following corrections related to CJK support:

- **E-mails** – some problems with non-Latin text in the body emails were corrected.
- **Saving a record to EndNote** – some problems when records were saved to EndNote using character set UTF-8 were corrected.
- **Reading Lists** – problems in the way some character sets were saved in the Reading List format have been fixed.

Note: The minor version release package can be applied on any MetaLib 4.x.x version, irrespective of the MetaLib 4.x.x service pack level installed. Customers who have not yet applied previous MetaLib Version 4 service packs should refer to the related documents listed below.

1.2. Related Documents and References

- How to apply the Service Pack:
MetaLib Version 4.1.x Service Pack
- Release Notes for previous MetaLib Service Packs :
METALIB-4.10-ServicePackItemContent-234-293
METALIB-4.1.1-ServicePackItemContent-294-344

2. Service Pack Content Summary

[Table 1](#) contains a list of corrections and enhancements that are included in this Service Pack. Where appropriate, this table references other sections for detailed information, such as implementation instructions.

Table 1: Enhancements and Corrections

| # | Fix Number | Description | Module | Fix Type | Support Incident |
|---|------------|---|-------------|-------------|------------------|
| 1 | 345-348 | CJK Related Functionality enhancements. For details, refer to section 3.1 . Note that due to changes in the appearance of the A to Z list, the HTML file for this page has been changed and some text has been added. For implementation details for the HTML change and text translation, refer to sections 4.1 and 4.2 . | /V | Enhancement | |
| 2 | 349 | Util S 4-3-1 did not recognize that Apache was running. | Maintenance | Correction | 8192-78000 |
| 3 | 350 | The Vivisimo based clustering did not recognize "stop words" (e.g. "and", "have", "with" etc.) on resources with languages other than English. Clustering now works with Vivisimo's stop word files for Norwegian, Finnish, Swedish, Russian and Turkish in addition to English. | /V | Correction | 8192-101261 |
| 4 | 351 | On "Find e-Journal", the "Help" icon did not link to the Help window. | /V | Correction | 16384-5011 |
| 5 | 352 | Adding a search to "My Space" did not work when JavaScript was disabled on the Internet browser. | /V | Correction | 8192-102946 |
| 6 | 353 | The "Duplicate Item" function failed when there were over 20 entries in the database/s. | /V | Correction | 8192-97300 |
| 7 | 354 | Under "Find a Resource", the "Resource Type" drop-down list returned to the default definitions even if it was customized. | /M | Correction | 8192-96566 |

| # | Fix Number | Description | Module | Fix Type | Support Incident |
|----|------------|--|--------|------------|----------------------------|
| 8 | 355 | <p>The following problems relating to character sets were fixed:</p> <ul style="list-style-type: none"> • Characters were not always correctly saved in Reading lists. Note that the save as Reading List option should be used only with the UTF character set. • Emails would not open correctly with every character set. This problem was resulting in problems for some character sets in the header and body of the email. <p>In order to support these changes, a new table was added to the system. This table specifies the official international character set used.</p> <p>For details, see section 4.3.</p> | /V | Correction | 8192-75459 and 8192-100048 |
| 9 | 356 | Fields that the customer defined as linkable in the IRD would not always work. | /V | Correction | 16384-4652 |
| 10 | 357 | When records were saved to EndNote using character set UTF-8, sometimes there was a problem opening the file. | /V | Correction | |
| 11 | 358 | The indexing process (p_index_database) had problems indexing very large amounts of IRD records (e.g. more than 45000 records). | | Correction | 8192-101966 |
| 12 | 359 | The self-registration form displayed the error message "You have missed a mandatory field. Please fill in the form again" even if all the fields were filled-in correctly. (The user record was created regardless of the message). | /V | Correction | 16384-8232 |
| 13 | 360 | Access permissions to databases were not checked when MetaLib was searched via deep links. | /V | Correction | 8192-74109 |
| 14 | 361 | The option to sort the results list by the database will now sort the database groups alphabetically for customers who have upgraded to version 4. | /V | Correction | 8192-94612 |
| 15 | 362 | Lateral links functionality would not always work on the complete value (e.g. it would display only the 'W' for the author John W Smith when the author link was clicked). | /V | Correction | 16384-11209 |

| # | Fix Number | Description | Module | Fix Type | Support Incident |
|----|------------|--|--------|-------------|------------------|
| 16 | 363 | <p>The "filter by subset" option, that gives the option to search a resource by sub fields, was added to external and webconfig resources (in addition to the Z39 method that was already working). A filter is a fixed value that is added to the user's query to limit the search to specific sub-sets of the target database.</p> <p>The option is reachable from the Management interface, Find a resource, presentation, Primary.</p> | /M | Enhancement | |
| 17 | 365 | <p>The Java converter did not start as expected. The harvesting in MetaIndex went into a loop.</p> | /M | Correction | 8192-98870 |

3. Functional Enhancements Detailed Description

3.1. CJK Enhancements

The following CJK related enhancements are available with MetaLib 4.2.x:

- Support for CJK A-Z Lists.
- Improved sorting of CJK records in results lists - CJK languages are sorted separately.
- Normalization of Traditional to Simplified Chinese.
- Enhanced support for Japanese (bringing it to the level of Chinese and Korean).
- Improved ranking of CJK results by taking segmentation into account. Note that this enhancement was already functioning in sites where the `CJK_LANG` parameter is set to `K`.

The following sections describe the main changes in the system and provide a more detailed description of these enhancements.

3.2. New Fields in the IRD: CJK Display Name Language and CJK Transliterated Display Name

CJK Display Name Language

This field indicates if the display name of the resource is one of the CJK languages: Chinese, Japanese or Korean. This is required due to some special functionality for CJK languages, especially sorting of the resource by display name. The following values are valid:

- Chinese
- Korean
- Japanese
- Not CJK

The system defaults the field based on the character set entered in the Display Name field on the Subscription tab. If Chinese characters are entered, the language is determined by the `CJK_LANG` parameter defined in the `metalib_start` file as follows:

- C, T → Chinese
- K → Korean
- J → Japanese
- N → Not CJK

CJK Transliterated Display Name

This field allows you to enter a transliterated name for the CJK display name, which MetaLib uses for grouping and sorting purposes.

This field is mandatory if the CJK Language is Chinese and the `CJK_LANG` parameter in `metalib_start` is C or Japanese.

If the CJK Display Name Language is Chinese and the `CJK_LANG` parameter in `metalib_start` is T (Taiwan) or the CJK Display Name Language is Korean, no transliteration is required because the name is normalized for grouping and sorting purposes automatically by the system (based on keystroke for Chinese in Taiwan and Hangul for Korean).

If the display name field includes CJK characters, the system will create a transliterated title automatically as follows. The operator can modify this transliteration.

- Chinese – Pinyin based on a transliteration table. If the CJK Language parameter in `metalib_start` is “C” (i.e. in China), the Pinyin will be preceded by the Chinese ideogram.
- Japanese – since automatic translation is not possible the system enters the text `***Missing transliteration`. The Yomigana should be entered manually.

3.3. CJK Display Name Language Added to Sort Key

To ensure that results are grouped by languages (that is Chinese, Japanese and Korean records are not inter-filed), the sort key allows you to specify the CJK language. This ensures that Chinese, Japanese and Korean records sort separately. The sorting of the language groups is determined by the `CJK_LANG` parameter from the `metalib_start` file so that in China and Taiwan, Chinese records are sorted first; in Japan, Japanese records are sorted first; and in Korea, Korean records are sorted first. If the `CJK_LANG` is N (the primary language is not a CJK language), records are sorted as follows: Chinese, Korean, and Japanese.

3.4. CJK A to Z Lists

The A to Z List functionality differs based on the `CJK_LANG` parameter in `metalib_start`. If the value is `C`, `J`, `K`, or `T`, in addition to the default Latin A to Z List, there will be an additional A to Z list in the primary language CJK language as defined by the following `CJK_LANG` values:

- **C** (A to Z list of Chinese titles):
 - Items are grouped according to the Pinyin transliteration. The headers are the Latin characters A-Z.

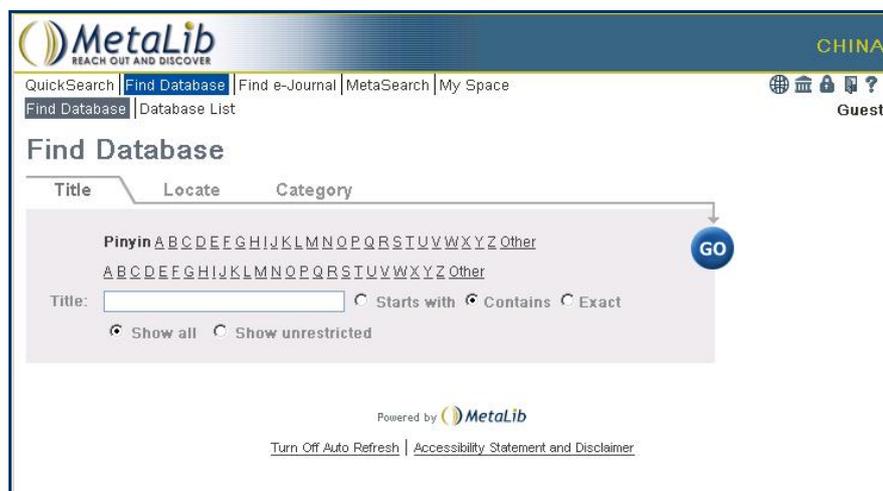


Figure 1: A-Z List of Chinese Titles- Pinyin Grouping

- **T** (A to Z list of Chinese titles – Two A-Z lists):
 - Items are grouped by stroke count of the first Chinese character (based on a table). The headers are numbers 1-25 and 26+.
 - Items are grouped by the Bopomofo of the first Chinese character (based on a table). The headers are the Bopomofo characters.



Figure 2: A-Z List of Japanese Title

Titles in one of the CJK languages other than the primary language are added to the “Other” group of the Latin A to Z list. These titles are grouped by language. In other words, in Mainland China and Taiwan, the “Other” group of the Latin A-Z list includes the active Korean and Japanese titles; in Korea, the “Other” group of the Latin A-Z list includes the active Chinese and Japanese titles; in Japan, the “Other” group of the Latin A-Z list includes the active Chinese and Korean titles.

If the CJK_LANG is N, any titles in CJK are added to the “Other” group of the default A to Z list.

3.5. “Starts with” Find Database Option

Until version 4.2, the **Starts with** option searches the browse (headings) indexes and displays a list of headings. However, this is problematic for CJK titles because of the need to sort based on the transliterated title that may be manipulated manually. If the CJK_LANG parameter is C, J, K, or T, the **Starts with** option behaves as a find and creates a set of results that match the query instead of a browse list.

3.6. Normalization from Traditional to Simplified Chinese

If the CJK_LANG parameter in metalib_start is C or T then all resources with CJK Display Name Language=Chinese are indexed as simplified Chinese (using a Traditional Chinese to Simplified normalization table. Any query in CJK characters is also normalized to simplified Chinese (using the same table).

The display of records and the query is the original form.

3.7. Japanese Support

3.7.1. Indexing and Search

Indexing of all three Japanese, writing systems (Hanja, Katakana and Hiragana) are based on 1-gram segmentation. In addition, MetaLib uses 1-gram segmentation for the query input by the user.

3.7.2. Normalization of Hiragana and Katakana

For indexing and searching, the following rules apply to the title and the search terms:

1. Compose all decomposed characters.

There are two Japanese Diacritics (Voiced Sound Marks); each one has two forms; the same diacritics are used for Katakana and Hiragana; the diacritic is composed with the preceding character to a composed character.

2. Remove CJK and standard punctuation.
3. Replace CJK space by standard space.
4. Half width Katakana >> Full width Katakana.
5. Small Hiragana “ << Big ”Hiragana letter.
6. Full width Latin >> Half Width Latin.
7. Uppercase/Lowercase Latin >> Lowercase/Uppercase Latin.

For normalization of the manually entered Yomigana (CJK transliterated title) apply the following rules also:

8. Katakana >> Hiragana
9. Normalize (remove) Japanese Diacritics (Voiced Sound Marks)
10. Replace “Prolong Sound Mark” by the appropriate “Japanese Vowel”

Note: The sorting based on the transliterated title will be the Unicode (UTF-8) value.

4. Detailed Implementation Instructions

This section contains the procedures needed to implement some of the fixes in this minor version release.

Where possible, the Service Pack mechanism automatically makes changes to relevant files. Each item in this section indicates whether a change was made automatically or a manual change is needed. In general, there are no manual changes needed in this minor version release, except for customers who use languages other than English. Such sites will need to translate manually some texts for the User interface.

4.1. Re-indexing of DAT01

As a result of fixes 345 - 346, the system must re-index DAT01 automatically after the service pack has completed. The following message appears at the end of the service pack installation:

```
As part of MetaLib 4.2 Minor-Release, indexing of the DAT01
database is necessary and will be running in the background.
Notice that this procedure can run between 30 minutes to
several hours (depending on the size of your database). While
this procedure continues, please do not start the MetaLib
server.
```

```
Press any key to continue.
```

After clicking any key, the process runs. The SP Util option appears. If you choose `y` for the Util request, you can choose option `7` to receive the status of the indexing.

To make sure that the process has succeeded (optional), you can view the following log file:

```
$TMPDIR/p_index_dat01.log
```

Note: If the CKB update is scheduled for the same time that the system is re-indexing DAT01, the update from CKB will be stopped. The customer should either make the update manually or update it on the next monthly update.

4.2. HTML changes and Text to translate (for languages other than English)

Due to the changes made in the A-Z list page, some new text has been added and the HTML file has been changed.

To translate these texts, perform the following steps:

1. Modify the `www_const.lng` tab library for all of your instances (e.g. `ins03`).
2. Translate the following texts to the relevant language:

A to Z List
Pinyin
Stroke Sequence

Note that former "123" button was transferred to "Other". The term "Other" has appeared before and exists on the translations table.

Note: The A-Z list HTML file (`./insNN/www_v_eng/find-db-1-body-title`) is replaced by the service pack. If you made any local changes to this file, you will need to restore them. As with previous Service Packs, this minor version release saves the original files. The name of the saved file is `./insNN/www_v_eng/find-db-1-body-title.sp.365`, where 365 is the SP level.

4.3. Complementary Table to `tab_character_conversion_line` – `tab_character_codes`

The `tab_character_conversion_line` table usually converts one code to another. For example, the table gets `UTF_TO_WEB_MAIL`, uses `line_utf2line_sb` program for conversion, and changes it to `unicode_to_8859_1`.

The table looks like this:

```
UTF_TO_WEB_MAIL      ##### # line_utf2line_sb          unicode_to_8859_1
```

The `tab_character_conversion_line` table does not include the official character set name, which is required to be able to work with e-mails. The new `tab_characters_codes` table has been added to convert the character set defined in the `tab_character_conversion_line` table to the official character set code or name.

The table looks like this:

```
!!!!!!!!!!!!!!!!!!!!-!!!!!!!!!!!!-!!!!
UTF_TO_WEB_MAIL      iso-8859-1          28591
UTF_TO_WEB_MAIL_ASCII us-ascii           20127
```

The second column (**yellow lines**) indicates the official character set name. The first column is the description key as it appears in `tab_character_conversion_line` and the last column indicates the character set code.

If you have added additional lines and/or encodings to `tab_character_conversion_line`, you should make sure that `tab_character_codes` has a matching definition.

Note that the upgrade will automatically create lines in `tab_character_codes` for all out of the box `tab_character_conversion_line` definitions. To check if the upgrade succeeded, you can check the `$TMPDIR/rc_355.log` file.