

International Cataloging: Use Non-Latin Scripts

Last updated: July 2016



OCCL Online Computer Library
Center, Inc.
6565 Kilgour Place
Dublin, OH 43017-3395

Contents

- 1. Connexion client international cataloging 3
 - Introduction 3
 - 066 script codes 3
 - Right to left control characters 3
 - Romanized data 5
 - Character display 5
 - Install RLIN21 Arabic, Cyrillic, Hebrew or Latin keyboards 5
 - Install a Windows keyboard 8
 - Use CJK E-Dictionary 9
- 2. Working with records 16
 - Searching 16
 - Create records 19
 - Edit records 20
 - Use constant data 25
 - Save records and search save files 26
 - Report errors in non-Latin scripts 26
 - Export records 27
 - Import records 28

1. Connexion client international cataloging

Introduction

Connexion client supports all Unicode characters for non-Latin cataloging. Unicode is the universal character encoding scheme for written characters and text. It defines a consistent way of encoding multi-script text that enables the exchange of text data internationally. MARC-8 encoding for scripts such as Arabic, Chinese, Cyrillic, Greek, Hebrew, Japanese and Korean is a subset of Unicode. The scripts defined in MARC-8 are now expanded in OCLC interfaces to include the non-MARC-8 characters that are part of these scripts.

Connexion client version 2.63 supports scripts through Unicode version 8.0.0. Information about Unicode 8.0.0 can be found [here](#). For more information on MARC-8 characters, see the [MARC-8 Code Tables](#).

066 script codes

OCLC supports script codes for all scripts in [066 fields](#). If you want to search for any of these codes, you can use the Character Sets Present index using either of the following labels: **vp:** (index) or **vp=** (phrase).

Right to left control characters

Align script data for display and print

By default, the client displays (and prints) scripts with data aligned to the right. To toggle between displaying these scripts right-to-left or left-to-right:

Action
<p>Toggle alignment for all script data in the current record: Click View > Align Right, or press <Alt><V><I>. Default: Data aligns to the right for display and printing.</p> <p>Result: The Align Right icon next to the command on the View menu is active (highlighted) if Align Right is selected. The icon is inactive (grayed out) if Align Right is cleared.</p> <p>Or</p> <p>Toggle data alignment in the current field: Right-click a field, and on the pop-up menu, click Right-to-Left Reading Order.</p> <p>Result: The client changes alignment of the script data only in the current field.</p>

Use Unicode formatting characters to control bidirectional data

Enter Unicode formatting characters in records to correctly display left-to-right multiple-digit numbers and punctuation, including brackets, hyphens, internal spaces, etc., within a field of right-to-left script data.

- **Export/import using UTF-8 Unicode character set.** Unicode formatting control characters are retained as is in records exported or imported using the UTF-8 Unicode character set, along with other non-MARC-8 Unicode characters.
- **Export/import using MARC-8 character set.** The Unicode formatting characters are retained in Numeric Character Reference (NCR) format in records exported or imported using the MARC-8 character set, along with other non-MARC-8 characters.

Action	
1	Click to locate the cursor in the position where you want to insert a formatting control number.
2	Right-click in the field, and on the pop-up menu click Insert Unicode Control Character . Or Right-click to open the pop-up menu and then press the keystrokes shown in step 3.
3	Click one of the following characters (or press the keystroke shortcuts, shown in parentheses); <ul style="list-style-type: none"> • LRM Left-to-Right Mark (<Alt><R><L>) • RLM Right-to-Left Mark (<Alt><R><R>) • ZWJ Zero Width Joiner (<Alt><R><J>) • ZWNJ Zero Width Non-Joiner (<Alt><R><N>) • LRE Start of Left-to-Right Embedding (<Alt><R><S>) • RLE Start of Right-to-Left Embedding (<Alt><R><T>) • LRO Start of Left-to-Right Override (<Alt><R><A>) • RLO Start of Right-to-Left Override (<Alt><R><F>) • PDF Pop Directional Formatting (<Alt><R><P>) <p>Tip for one-step entry: Create a text string using Tools > Text Strings; click Add and enter one of the characters listed above using the right-click menu. Then use the Text Strings quick tool on the toolbar to enter the character.</p> <p>Or</p> <p>Assign the text string to a keystroke shortcut. Enter the character by pressing the keystroke.</p>

Example: To control the display of the data *742[1981 or 1982]* that you enter in field 260 Ꞥc, and that is preceded and to be followed by script data:

1. Click to locate the cursor in field 260 Ꞥc.
2. Right-click in the field, and in the pop-up menu click **Insert Unicode Control Character**. Then click **LRE Start of Left-to-Right Embedding**.
3. Enter the data string, **742[1981 or 1982]**, immediately following the character.
4. Without moving the cursor, right-click in the field again. In the pop-up menu click **Insert Unicode Control Character**. Then click **PDF Pop Directional Formatting**.

More information:

- For details, see the [Unicode Bidirectional Algorithm](#) report.
- See more about selecting a character set for exporting and importing bibliographic records in the client in [Cataloging: Export or Import Bibliographic Records](#).

Romanized data

When cataloging in English, romanize according to the [ALA-LC Romanization Tables](#) when they exist for a particular script/language.

Character display

- Valid Unicode characters can appear as empty boxes if they are not supported in your default font.
- The Arial Unicode MS font that OCLC recommends will not support all scripts. You will have to search for fonts to cover some of the new scripts OCLC supports. Fonts that support these scripts vary in their ability to display characters outside of the script they were designed to support. You may have to experiment to see which font works best for your needs for a particular script. The font selected as the default in the client is applied to the entire record. You may have to move between two fonts to see all of the characters in a record.

Install RLIN21 Arabic, Cyrillic, Hebrew or Latin keyboards

About keyboards for entering script data in the Connexion client

- **RLIN21 Arabic, Cyrillic, and Hebrew keyboards:**
 - **RLIN21 keyboards.** Connexion client supports keyboards that were previously created by Research Libraries Group for use with RLIN21.
 - **Windows input keyboards.** You may use Microsoft Windows input keyboards or install and use RLIN21 keyboards to enter these scripts.
 - **Difference:** RLIN21 keyboards include characters specific to each script. Microsoft keyboards include characters specific to a language written in a particular script.

Example: Using the single RLIN21 Arabic keyboard, generate Arabic script characters for Arabic, Persian, and Urdu instead of switching among separate Microsoft keyboards for each language to enter script data.

Note: For Chinese, Japanese, and Korean (CJK), you must use Windows Input Method Editors (IMEs). For scripts outside the MARC-8 characters, use Windows keyboards. See also, "Cataloging, International, Input methods for languages that use non-Latin scripts."

- **RLIN21 Latin keyboard:**
 - The client supports two methods of entering diacritic and special characters.
 - **Client function.** Use the functionality in the client for entering a diacritic or special character, click **Edit > Enter Diacritics** and select the image or name for a character, or use assigned client keystrokes (see a complete list of client assigned keys in View > Assigned Keys).
 - **RLIN21 Latin keyboard.** Use RLIN21 methods and equivalent keystrokes to enter diacritics and special characters.

Install an RLIN21 keyboard

Notes:

- You must have local administrator rights for your workstation to download and install the RLIN21 keyboards.
- After installing, you must make an installed keyboard available on your workstation—that is, select it under the **English (United States)** input locale for your workstation—before you can use the keyboard.

To install a keyboard:

	Action
1	Go to Cataloging Software Downloads .
2	Type your OCLC authorization and password and click Enter .
3	Review the Terms and Conditions by clicking the OCLC Software License Terms and Conditions heading to open the document.
4	Click the check box to accept the Terms and Conditions. This step is required. Links to software and instructions are not active until the user indicates he accepts the terms and conditions.
5	Click one of the following: <ul style="list-style-type: none">• RLIN21 Arabic keyboard• RLIN21 Cyrillic keyboard• RLIN21 Hebrew keyboard• RLIN21 Latin keyboard
6	Save the software to your hard drive.
7	For instructions about how to install the software, click the link accompanying instructions.

Make installed RLIN21 keyboard available for use

After installing a keyboard, you must make it available for use in Windows. Instructions differ somewhat, depending on your version of Windows. Follow instructions on the Microsoft Web site for your version. See:

- [How to use the United States-International keyboard layout in Windows 7 and Windows Vista](#)
- [Add a keyboard layout or input method for a language](#)
- How to select the United States-International keyboard layout in Windows 10

1. Go to the Control Panel and select **Change Input Methods**.
2. Select **Options**.
3. Select **Add an input method**.
4. Enter **international** in the top right search box. Select **United States-International** and click **Add**.
5. Click **Save**.
6. On your taskbar, you will see ENG:US. Click on the button and select **ENG:INTL** to select the United States-International keyboard.

Caution: You must make RLIN21 keyboards available for use under the **English (United States)** input locale when you follow the Microsoft instructions. Do not install under any other language.

Select an RLIN21 keyboard to input script data

When you install languages and/or input keyboards or IMEs on your workstation, a language indicator appears in the Windows system tray (bottom right corner of the desktop). If you have multiple languages installed, clicking the language indicator expands a list of them. If not, the indicator contains only English (United States).

When you install keyboards under a language and the language is selected in the language indicator, a keyboard indicator also appears in the system tray.

To select an RLIN21 keyboard for entering data:

Action	
1	<p>Make sure that EN is selected in the language indicator if necessary (that is, if you have multiple languages installed on your workstation).</p> <p>Tip: If you have multiple languages installed, use the Windows default keystroke shortcut <Left Alt><Shift> to toggle through the languages if needed.</p>
2	<p>Click the keyboard indicator and select the name of the RLIN21 keyboard you want to use.</p>

Caution:

- While using the RLIN21 keyboards, you cannot use Connexion client default keystroke shortcuts mapped to characters that include the **Alt**, **Ctrl+Alt**, or **Ctrl+Alt+Shift** keys.
- These key combinations are used with the RLIN21 keyboards and override the Connexion client keystrokes when you are using an RLIN21 keyboard.
- However, Connexion client function keys and keystrokes assigned to characters that include the **Ctrl**, **Alt+Shift** and **Ctrl+Shift** keys remain available.

Examples:

The default keystroke for **Action > Delete Record**, mapped to **<Ctrl><Alt><D>**, conflicts with the RLIN21 keyboards (contains **Ctrl+Alt**) and cannot be used. You must map **Delete Record** to another keyboard shortcut.

However, you can continue to use the default keystroke for the delimiter character (<Ctrl><D>) and the default keystroke for **Action > Export** (<F5>), since they do not conflict with RLIN21 keyboards.

Uninstall an RLIN21 keyboard

Notes:

- You must have local administrator rights to uninstall the RLIN21 keyboards.
- You must first remove the keyboard from regional options before uninstalling (next procedure). If you do not, when you uninstall, you receive an error message: Fatal error during installation.

First, remove the keyboard from regional options:

Action	
1	1. Click the Windows Start menu, and then click Settings > Control Panel > Keyboard . 2. Click the Input Locales tab.
2	In the list of language services available, under English (United States) , select the name of the RLIN21 keyboard you want to remove.
3	Click Apply and then click OK .

Then, uninstall the keyboard program:

Action	
1	Click the Windows Start menu, and then click Settings > Control Panel > Add/Remove Programs .
2	In the alphabetical list of installed programs, scroll to the name of the RLIN21 keyboard that you want to uninstall and select it.
3	Click Remove . Click Yes to confirm.

Install a Windows keyboard

If the default language of your workstation is not the language you need for entering supported non-Latin scripts for cataloging or for variant name headings in authority records, or if you do not already have an input method for the language set up on your workstation, install input languages and methods in Windows.

Action	
1	Go to Start > Settings > Control Panel > Regional and Language Options .
2	Click Languages .
3	Optional . In the Languages tab, click a check box to select "Install files for complex script and right-to-left languages (including Thai)."
4	Required for creating Chinese, Japanese, and Korean text: Click a check box to select "Install files for East Asian languages."
5	Click Details . In the Settings tab, click Add .
6	Select a language to install and click to select the Keyboard layout/IME check box to also install the keyboard.

When English is your default language and you install one or more other input languages, Windows automatically displays a language bar on the desktop for toggling through your input languages. You can also use the default Windows keystroke shortcut <Left Alt><Shift> to toggle through installed languages.

For more information and help on setting up input languages and keyboards specifically for your version of Windows:

- On the Windows **Start** menu, click **Help**, and in the Help Search box, enter “input languages.” Or
- Go to the [Microsoft Web site](#) and search the site for “input languages vista” or “input languages windows 7,” for example.

Use CJK E-Dictionary

What is the CJK E-Dictionary?

The CJK E-Dictionary (electronic dictionary for Chinese, Japanese, and Korean characters):

- Allows you to search or browse to retrieve information about a CJK character, group of related characters, homophones matching a phonetic input code, or a large set of characters in sequence by EACC or Unicode value.
- Includes all CJK characters represented in the East Asian Character Code (EACC) and supported in the Connexion client.
- For each character, provides the following types of character representation (if applicable):
 - EACC bitmap
 - EACC 3-byte code
 - Unicode font representation
 - Unicode
 - Tsang-chieh input code
 - Wade-Giles input code (if applicable)
 - Pinyin input code (if applicable)
 - McCune-Reischauer input code (if applicable)
 - Modified Hepburn input code (if applicable)

CJK phonetic input codes. See the *ALA-LC Romanization Tables* on the Library of Congress Web site for CJK phonetic input codes:

- [Chinese](#)
- [Japanese](#)
- [Korean](#)

Why search or browse the CJK dictionary?

Searching allows you to:

- Find a single character by looking up its EACC code
- Find related characters by searching with a partial EACC code
- Get input codes for a Unicode character that you paste into the dictionary search window

- Find characters that match a Tsang-chieh (TC) code or search using a partial TC code to list characters with one or more common graphic elements
- Find characters that match a phonetic input code

Browsing allows you to:

- View information about all characters encoded in an EACC layer
- View a range of EACC codes within a layer (specify the beginning EACC value to display in the layer)
- Print E-Dictionary content for an EACC layer or part of a layer

How to search or browse the CJK E-Dictionary

To ...	Do this ...
Open the E-Dictionary	On the Tools menu click CJK E-Dictionary , or press <Alt><T><E>. (Opens the E-Dictionary Search window.)
Search	
Search by EACC	<ol style="list-style-type: none"> 1. In the E-Dictionary Search window, under Input Method, select EACC. 2. In the Search by Input Code box: Find a character: Type the 3-byte (6-character) EACC code. Or Find related characters: Type two asterisks (**) in place of the first byte (2-character code for the 'Plane' in the EACC code system). 3. Click Find.
Search by Tsang-chieh (TC)	<ol style="list-style-type: none"> 1. In the E-Dictionary Search window, under Input Method, select Tsang-chieh (TC). 2. In the Search by Input Code box: Type the complete TC code for a character. Enter 1 to 5 letters that represent graphic elements of Chinese characters. Or Type a partial TC code by substituting an asterisk (*) for each letter of the code that you do not know. See "What is the Tsang-chieh input code?" below for more information. 3. Click Find.
Search by Unicode character	<ol style="list-style-type: none"> 1. Before opening the CJK E-Dictionary, copy a Unicode character from a record-editing window. 2. After opening the E-Dictionary, in the E-Dictionary Search window, under Input Method, paste the Unicode character: Press <Ctrl><V> or right-click and on the popup menu click Paste. 3. Click Find.

To ...	Do this ...
Search by phonetic input code	<ol style="list-style-type: none"> 1. In the E-Dictionary Search window, under Input Method, select Wade-Giles, Pinyin, Modified Hepburn, or McCune-Reischauer. 2. In the Search by Input Code box, type an input code of the selected type. <ul style="list-style-type: none"> • Wade-Giles and Pinyin (Chinese): Enter 1 to 8 characters. May follow with a tone qualifier and/or TC qualifier.* • Modified Hepburn (Japanese): Enter 1 to 14 characters. May follow with a TC qualifier.* • McCune-Reischauer (Korean): Enter 1 to 8 characters. May follow with a TC qualifier.* 3. Optional. Include diacritics if necessary: <ul style="list-style-type: none"> • Place the cursor before the character associated with the diacritic. • Click ALA Entry and select a diacritic; then click Insert and Close. 4. Click Find. <p>*For tone and Tsang-chieh (TC) qualifiers, enter:</p> <ul style="list-style-type: none"> • Chinese tone qualifiers 0 (all tones) (flat) (rising) (falling and then rising) 4 (falling) 5 none • Japanese and Korean TC qualifiers Tone qualifier 0 and then TC qualifier • TC qualifiers Limit retrieval to Chinese-derived characters. Enter the first non-X letter of the TC input code.
Browse	
Browse an EACC layer	<ol style="list-style-type: none"> 1. In the E-Dictionary Search window, click Browse. 2. In the Browse by EACC Layer window, under Index, select a layer. 3. Optional. In the Goto EACC Value box, type a 3-byte (6-character) EACC value for the character with which to start retrieving E-Dictionary entries. 4. Click OK. <p>Caution: Browsing layer 1 can take up to several minutes; it contains entries for almost 10,000 characters. Layers 3, 8, and 14 are also extensive. See "EACC layers: details" below for the number of characters in each layer.</p>
Search and browse results	
View or print list of matching entries	See "Open and use a dictionary entry" below.

To ...	Do this ...
Open a list entry to copy and paste values into records, workforms, constant data, or text strings	See "Open and use a dictionary entry" below.

EACC layers: details

For basics, see "What is EACC?" below.

- EACC divides the codespace into 16 layers of 6 planes each.
- Layers 10-12 and 15-16 are not currently allocated to characters.
- Within layers, not all planes are in use.
- Layers 1-12 are structured to show relationships between traditional Chinese characters and characters derived from traditional forms (layers 10-12 not in use).
- Layer 1 has traditional characters.
- Layer 2 has simplified forms of characters in layer 1.
- Layers 3 to 9 have other variants of characters in layer 1.

Linked characters and indexing:

- [What is an IME \(Input Method Editor\) and how do I use it?](#)
- [Input Language: Keyboards and IMEs](#)
- The EACC code for a variant form has the same values in the second and third bytes as the code for its corresponding traditional character, but a different value for the first byte.
- Characters with shared second and third byte values are linked.
- In WorldCat, this EACC-code linking is used to index all variants of a character as equivalents.
- A search term in Connexion that begins with a character related by EACC coding to one or more variant forms retrieves records in which the indexed field begins with any of the related forms of that character.

The following table shows the number of characters, EACC planes, and a description for each EACC layer:

Layer	Number of characters	EACC planes	Description
1	9,900	21, 22, 23	Traditional forms of Chinese characters
2	2,100	27, 28, 29	Simplified forms of Chinese traditional characters; linked to characters in layer 1
3	600	2D, 2E, 2F	Variant forms of Chinese traditional characters; linked to characters in layer 1
4	230	33, 34, 35	Other variant forms of Chinese traditional characters; linked to characters in layer 1
5	90	39, 3A, 3B	
6	30	3F	
7	55	45, 46, 47	
8	480	4B, 4C, 4D	Variant forms of Chinese traditional characters; linked to characters in layer 1; not included in Chinese Character Code for Information Interchange (CCCII)
9	18	51, 52	

Layer	Number of characters	EACC planes	Description
10 – 12	Not used		
13	298	69	Japanese characters (hiragana, katakana, and Japanese-invented kokuji) with no Chinese counterparts
14	2,073 (6F) 45 (70)	6F 70	6F - Korean characters (hangul) with no Chinese counterparts 70 - Chinese characters created in the People's Republic of China that have no related traditional forms
15 - 16	Not used		




Use the search results list

The client displays search and browse results in a list of entries with the following type of representation for each entry (character), if applicable (columns listed in default order).

Character representation	Description
EACC	3-byte East Asian Character Code (EACC) value for the character
Unicode Font (UFont) (graphic)	Character as represented in the default client font Default: Arial Unicode MS or Windows default Note: If you change the default client font, the new font must be Unicode-compliant.
Unicode	2-byte code for the character assigned in the Unicode standard
Tsang-Chieh (TC)	One to five letters that identify the character in the Tsang-chieh graphical coding scheme
Wade-Giles (WG)	Wade-Giles romanization of Chinese pronunciation(s) for the character, including any tone qualifier
Pinyin (PY)	Pinyin romanization of Chinese pronunciation(s) for the character; includes any tone qualifier
McCune-Reischauer (MC)	McCune-Reischauer romanization of Korean pronunciation(s) of the character
Modified Hepburn (HP)	Modified Hepburn romanization of Japanese pronunciation(s) of the character

To copy and paste a value or graphic from a dictionary entry, you must open the entry. You cannot copy from the list.

To temporarily customize the list for display and print:

Action
<p>Sort order: To re-sort by another column, click the column heading. The list re-sorts by values in the column, in ascending order (alphabetic or numeric) with entries that have no value appearing first. An arrow identifies the sorting column; for example:</p> 
<p>Sort direction: To reverse sort order, click the column heading for the column that sorts the list. The arrow reverses the order to indicate descending order: </p>
<p>Column size: To change the column size, point to the right border of the column heading you want to re-size. When the pointer becomes , click and hold to drag the border.</p>
<p>Column order: To change the order of a column, click, hold, and drag the column heading to a new position.</p>
<p>Note: The client does not retain changes you make to the list once you close the list.</p>

To print the list or print specified pages of the list:

Action
<p>To print an entire results list, click Print. In the Print window, click OK.</p>
<p>To print a specified number of pages in the list: When the Print window opens, click the Pages from...to button and enter the number of pages to print. Then click OK.</p>
<p>Caution: If you want to print results for an entire layer of EACC values, be aware that some layers contain a large number of entries:</p> <ul style="list-style-type: none">• Layer 1 contains entries for almost 10,000 characters. Printing all of the entries requires over 200 sheets of paper.• Layers 2, 3, 8, and 14 are also extensive.

Open and use a dictionary entry

To open a dictionary entry from the E-Dictionary Search Results window:

Action
<p>Double-click an entry, or select an entry and press <Enter>.</p>
<p>The E-Dictionary Entry window opens showing the same information that was in the results list, but in boxes for copying and pasting.</p>

To copy and paste information from a dictionary entry into records, workforms, constant data, or text strings:

	Action
1	In the E-Dictionary Entry window, select a value or graphic.
2	To copy, press <Ctrl><C>, or right-click the selected text or graphic, and on the popup shortcut menu, click Copy .
3	Close the E-Dictionary.
4	To paste, place the cursor where you want paste the text or graphic in a record or text string and press <Ctrl><V>, or right-click where you want to paste, and on the shortcut menu, click Paste .

2. Working with records

Searching

Although Connexion client 2.63 supports the additional characters, OCLC has not completed indexing them. You will not be able to search using the new characters until OCLC completes indexing. OCLC will make an announcement later in 2016 when you can search using these characters. Make a note of OCLC control numbers for new or replaced WorldCat records so you can find them before indexing is complete. Script characters OCLC supported before continue to be searchable.

Entering searches

- Search for records containing non-Latin script data using either script search terms or romanized (Latin-script equivalent) search terms.
- Both interactive and batch searching support non-Latin script search terms (**Cataloging > Search > WorldCat and Batch > Enter Bibliographic Search Keys**).
- Alternatively, copy and paste non-Latin script data into client searches from sources external to the client.
 - Non-Latin script search terms must be based on Unicode. However, only Unicode characters that can be converted to MARC-8-equivalent characters are valid in WorldCat.
 - If Unicode characters that are not convertible are in the search term, you may find no matching records.
- About using search indexes for non-Latin script search terms:
 - Use the same indexes (labels and punctuation) for non-Latin script searches that you use for Latin script searches. Enter the index labels and punctuation using Latin script.
 - Do not use derived searching for non-Latin scripts.
 - Add the same qualifiers to both Latin and non-Latin script searches. Enter them using Latin script.
 - Browsing to scan indexes for a match is available for any supported script. Browsing scans for the exact data string followed by any other data, providing automatic truncation. Enter only as many characters as are unique enough to retrieve matching record(s).
 - For **Arabic, Armenian, Bengali, Cyrillic, Devanagari, Ethiopic, Greek, Hebrew, Syriac, and Tamil** script searches, use word or phrase search indexes and word or phrase browse indexes.
 - For **CJK** script searches, the system indexes both single characters and immediately adjacent characters in a field. Use the following search strategies:
 - Word search – Enter an index label and a colon (for example, **ti:**) followed by a character string with no spaces to find a single word, or followed by more than one character string separated by a space to find multiple words, anywhere in an indexed field.
 - Phrase search – Enter an index label and an equal sign (for example, **ti=**) followed by a character string to find exact occurrences, starting with the first character in an indexed field and including each succeeding character. Truncate the character string to find the string followed by any other data without having to enter the entire data string as it appears in a field or subfield.

Note: Enter a minimum of three CJK characters if you truncate a search.

- Phrase browse – Enter the Scan command, an index label, and an equal sign (for example, **sca ti=**) followed by a character string. Phrase browsing scans an index for occurrences of the browse string at the beginning of indexed fields, followed by any other data (automatic truncation).

Note: Since all MARC-8 CJK characters are indexed singly, if you browsed for a word, the system would scan for the first character only, and results would not be significant.

- For **Thai** script searches, the system treats the entire data string you enter as both a word and a phrase, since Thai text has no spaces between words. Search for Thai terms using word or phrase search indexes and word or phrase browse indexes.

For phrase searching, you must use truncation.

To truncate, enter an asterisk (*) at the end of the search string. Enter a minimum of three Thai characters before truncating.

Note on word and phrase searching and browsing in general: Word searches find the data string you enter anywhere in the indexed field. Phrase searches find the data string starting with the first character in a field or subfield and including every character in exact order. Browsing scans an index for the closest match to the character string followed by any other data.

- If you want to retrieve all records or see sample records containing a particular script, use the “character sets present” WorldCat search index (label **vp:**) with the assigned code for a script. A list of all possible script codes that could be searched is included [here](#).

To enter one of the searches above to retrieve all records that contain a specified script, use the command line in the Search WorldCat window (**Cataloging > Search > WorldCat**).

Note: If a search for a particular script alone retrieves too many WorldCat records (limit 1,500 records), you must limit the search and try again. (See more about how the client displays WorldCat search results in [Search WorldCat](#), “Use WorldCat search results.”)

Examples

vp:ara/1991-2 (search for Arabic script records limited to those published in 1991 and 1992)

vp:ara and la:per (search for Arabic script records limited to those describing Persian language items)

See more about word and phrase searching and search methods in general in [Search WorldCat](#):

- “Search WorldCat interactively”
- “Keyword, numeric, and derived search syntax” (derived search is unavailable for non-Latin script data)

- “Browse WorldCat”
- “Customize WorldCat search and browse interfaces”
- “Enter WorldCat searches for batch processing”

Sort order of search results

You can select how the results of non-Latin script WorldCat searches are sorted:

- Alphabetically by the Latin script data
Or
- In Unicode order by the non-Latin script data

To check or change the option for sort order for WorldCat search results:

Action	
1	On the Tools menu, click Options (or press <Alt><T><O>), and then click the International tab .
2	Click the Primary Sort by Latin Script check box to select or clear the option to sort search results in alphabetical order by the Latin script data. Default: Check box selected. Search results sort alphabetically by Latin script data. Result: <ul style="list-style-type: none"> • If you clear the check box, search results are sorted in Unicode order by the non-Latin script data. • The sort order selected also determines the sort order of local bibliographic save file and local constant data search results. • Tamil Unicode 4.0 codes are not in collating order. The default, alphabetical sorting by Latin script, is recommended if romanized (Latin-equivalent) data is included in the record with Tamil script data.
3	When finished, click Close , or press <Enter> to apply the settings and close the Options window. Or Click Apply to apply the settings without closing the window.

Create records

Use as many non-Latin scripts as you need anywhere in a record, including within the same field.

Workforms

For creating bibliographic records and/or constant data with non-Latin scripts using workforms (**Cataloging > Create > Single Record > [format]** or **Constant Data**), you can set an option to display the workforms with paired fields.

Action	
1	On the Tools menu, click Options (or press <Alt><T><O>), and then click the International tab .
2	Click the Include paired fields in workforms for multi-script data check box . Default: Check box cleared. Workforms contain single fields. Results: <ul style="list-style-type: none"> The workform opens with paired fields 1XX, 245, 246, 250, 260, 300, 4XX, 5XX, 6XX, 8XX for non-Latin script entry (X = any valid tagnumber). Each pair has identical tags. The first of a paired field is for non-Latin scriptdata.
3	When finished, click Close , or press <Enter> to apply the settings and close the Options window. Or Click Apply to apply the settings without closing the window.

Setting the option for paired fields is not required. You can enter non-Latin script data only or romanized (Latin-script equivalent) data only. Or create your own paired fields and enter both:

Action	
1	In a record, add a field and enter the same tag as the corresponding existing field.
2	Enter the non-Latin script data in the first field of the pair, and optionally, enter the romanized data in the second field.
3	To link the fields, on the Edit menu, click Linking Fields > Link Field , or press <Alt><E><K><L>. Or Let the client link them automatically when you validate, reformat, save, or take a final OCLC action on the record.

Caution: If Latin script and non-Latin script parallel fields are not linked, display of records downloaded to your local system may be affected. You can set an option to get a warning before the client exports records with parallel unlinked non-Latin script fields in **Tools > Options > Export**.

Derive records

When you use **Edit > Derive > New Master Record**, **New Institution Record**, or **New Constant Data** to create records from existing records that have linked fields for non-Latin scripts, the client transfers the linked fields as pairs for each field selected in **Tools > Options > Derive Record**.

Note: Although the 066 field cannot be transferred, the client adds the field automatically to indicate the presence and type of non-Latin script when you validate, reformat, save, or take a final OCLC action on the record.

See also general information about how to create bibliographic records in Cataloging/Create Bibliographic Records, "About creating bibliographic records"

Edit records

Editing functions supported

- Find and Replace (**Edit** menu). Enter non-Latin script in both the **Find What** box and the **Replace With** boxes of the Find/Replace window.
- Cut, copy, and paste (**Edit** menu). Cut, copy, and paste non-Latin script data.
- When you move one of a paired field, the other field moves automatically (**Edit > Move Field > Up** or **Down**).
- Validate records and characters (**Edit** menu)
 - Validates all characters in record as well as validating MARC structure and tags.

If the client finds an invalid character, an error message lists the tag and position of the character to help you find it, along with other errors found in the record.

The error message may give up to three positions per field for invalid characters.

After correcting characters, you may want to validate the record again.

- Also validates all Unicode characters including the original set of scripts supported.
- Validate characters only. Validate characters separately from record validation (available for MARC-8 characters only; unavailable for all other Unicode characters including those in scripts that also contain MARC-8 characters). Use **Edit > MARC-8 Characters > Verify**.
- Reformat (**Edit** menu). Rearranges fields in MARC tag order, including:
 - Displaying paired fields together, with the non-Latin script field on top.
 - Automatically adding field 066, if not already added, with text that indicates the presence and type of script.
- Text strings (**Tools** menu). Use non-Latin scripts to create or edit text strings. See more about text strings in Basics, Set Options and Customize, "Create custom text strings."

Controlling headings unsupported

- Controlled headings in bibliographic records are linked directly to the controlling authority record so that if the authority record changes, the heading is automatically updated in the bibliographic record also.
- For non-Latin script records, however, only a heading in the Latin script field can be controlled (and subsequently updated if the authority record changes). If the controlled heading is updated, you may need to update the corresponding non-Latin script field manually to match the update.

Verify MARC-8 characters

To check the validity of characters separately from the validate records function (**Edit > Validate** or <Shift><F5>):

Action	
1	<p>On the Edit menu, click MARC-8 Characters > Verify, or press <Alt><E><8><V>.</p> <p>Results:</p> <ul style="list-style-type: none"> The client changes the color of invalid characters to red by default (ora color you specify in Tools > Options > Record Display). If no invalid characters are present, you get a message that verification is completed. Click OK or press <Enter> to close. <p>Note: MARC-8 verification is inappropriate for all scripts that are outside of the MARC-8 character set.</p> <p>Tip: If you are trying to catalog CJK characters to stay within MARC-8 characters, you can use an automatic converter to convert them to equivalent MARC-8 compliant characters. (Edit > MARC-8 Characters > Convert to MARC-8 CJK)</p>
2	<p>To remove invalid character display (display all text in the default text color or the color you selected in Tools > Options > Record Display), click MARC-8 Characters > Clear, or press <Alt><E><8><C>.</p> <p>See Tools > Options > Record Display for color options. Default color for invalid characters: red.</p>

Convert invalid CJK characters to equivalent MARC-8 characters

When you verify CJK characters as MARC-8-compliant (**Edit > MARC-8 Characters > Verify**), and the client identifies invalid character(s), you can automatically convert the character(s) in the record to MARC-8-equivalent CJK characters:

Action
<p>Click Edit > MARC-8 Characters > Convert to MARC-8 CJK, or press <Alt><E><8><J>.</p> <p>Result: The client converts the characters and changes the color of converted characters to green (by default) or to a color you specify in Tools > Options > Record Display.</p>

Tip: If you already know that a record contains invalid CJK characters, you can use the **Edit > MARC-8 Characters > Convert to MARC-8** command without first using the **Edit > MARC-8 Characters > Verify** command.

Note: The Library of Congress also has a [CJK Compatibility Database on the Cataloging Policy and Support Office \(CPSO\)](#) homepage to help with MARC-8 compliant or missing characters.

Link/unlink paired non-Latin/Latin script fields

The client automatically links two non-Latin script/Latin script fields that have the same tag number when you validate, reformat, save, or take an action on the record. The client always treats a non-Latin script field as the first of a corresponding pair.

You can link or unlink two non-Latin/Latin script fields with the same tag number.

Action
<p>To link fields: Place the cursor in the first field of a set of paired fields (the non-Latin data field), and on the Edit menu, click Linking Fields > Link Fields, or press <Alt><E><K><L> to link all paired fields. Or Right-click, and on the popup shortcut menu, click Link Fields to link the two fields where the cursor is located.</p> <p>To unlink fields: Click Linking Fields > Unlink Fields, or press <Alt><E><K><U> (unlinks all linked fields). Or Right-click, and on the popup shortcut menu, click Unlink Fields (unlinks the pair of fields where the cursor is located).</p>

When you link fields:

- The client uses a bracket to display linked fields, as in the following example showing part of a CJK record:

110	2	国立国会図書館 (Japan)
110	2	Kokuritsu Kokkai Toshokan (Japan)
245	1 0	国立国会図書館所蔵中国語・朝鮮語雑誌新聞目録 : #b 平成 3年末現在 / #c [編集国立国会図書館専門資料部].
245	1 0	Kokuritsu Kokkai Toshokan shozō Chūgokugo, Chōsengo zasshi shinbun mokuroku : #b Heisei 3-nenmatsu genzai / #c [henshū Kokuritsu Kokkai Toshokan Senmon Shiryōbu].
260		東京 : #b 国立国会図書館, #c 平成 5 [1993]
260		Tōkyō : #b Kokuritsu Kokkai Toshokan, #c Heisei 5 [1993]

- Printouts of records retain the brackets to indicate linked fields.
- If you modify the tag of one of the linked fields, the tag for the other field changes, too.
- If the cursor is in a linked field when you add a new field, the new field is added above or below the set of linked fields. Linked fields cannot be separated.
- Moving a linked field moves the set of linked fields.
- If you delete one field in a linked field set, the client keeps the other field and removes the link indicator (bracket).

Caution: If Latin script and non-Latin script parallel fields are not linked, display of the non-Latin script in records downloaded to your local system may be affected. You can set an option to get

a warning before the client exports records with unlinked non-Latin script fields in **Tools > Options > Export**.

Align right-to-left script data

By default, the client displays (and prints) right-to-left script data (e.g., Arabic, Hebrew, and Syriac) in records aligned to the right. Toggle alignment for these scripts using **View > Align Right**.

Use Unicode formatting characters for bidirectional data

Valid left-to-right character strings (multiple digit numbers and punctuation) appear mixed in with right-to-left script data in right-to-left scripts. To ensure that this bidirectional data displays correctly, use Unicode formatting control characters.

The formatting control characters distinguish how to display mixed left-to-right and right-to-left data in right-to-left scripts. To insert a control character, right-click in a field, and on the pop-up menu click **Insert Unicode Control Character**, then click a character.

For details, see the [Bidirectional Algorithm](#) report on the Unicode Web site.

Use CJK E-Dictionary

Use the CJK E-Dictionary (electronic dictionary) on the Tools menu to search or browse for information about:

- A single CJK character
- A group of related characters
- Homophones matching a phonetic input code
- A large set of characters in sequence by East Asian Character Code (EACC) or by Unicode value

See the [CJK E-Dictionary](#) for more information.

Transliterate Arabic

For existing Arabic records that contain only romanized data (Latin-script-equivalent representation of the Arabic script), the client provides two ways to automatically convert and add the equivalent romanized data:

- Use **Edit > Transliterate > Arabic [or Persian]** to transliterate romanized data in selected fields of a displayed record.
- Set an option (for bibliographic records only) in **Tools > Options > International** to auto-transliterate romanized data in all Arabic records retrieved interactively from WorldCat (records with language code **ara** or **per** and no field 066). Also select the fields to auto-transliterate.

Transliterate selected fields in a record

Action	
1	Display a bibliographic, authority, or constant data record containing romanized data that describes Arabic language materials.
2	Place the cursor in the field containing romanized data that you want to transliterate. Or Select multiple fields containing romanized data. If you select parts of fields, the client transliterates the entire field(s).
3	Click Edit > Transliterate > Arabic or [Persian] , or press <Alt><E><T><A> or <Alt><E><T><P>, or right-click and on the pop-up menu, click Transliterate > Arabic [or Persian] . See "Results of transliteration" below.

Note: Although you can transliterate into Arabic while working offline (you do not need to be logged on to the OCLC system), your workstation must have an Internet connection.

Auto-transliterate WorldCat records retrieved interactively (for bibliographic records only)

Alternatively, select an option to auto-transliterate romanized data in all WorldCat records you retrieve interactively when the records have the language code **ARA** but no field 066

Action	
1	Click Tools > Options (or press <Alt><T><O>), and then click International .
2	Click to select the Auto-transliterate Arabic records check box and/or the Auto-transliterate Persian records check box. Note: This option works for records that have the language code ara or per but no field 066.
3	Optional. Select fields to auto-transliterate: <ol style="list-style-type: none"> 1. In the International window, select the Auto-transliterate Arabic fields check box and/or the Auto-transliterate Persian fields check box. 2. Click Choose Fields to open the Choose Fields to Auto-Transliterate window. 3. Click to select or clear check boxes next to fields 1XX through 8XX (X = any valid tag number). Default: Fields 1XX through 8XX are selected. 4. Click OK to save your settings or Cancel to cancel changes. You are returned to the International window.
4	When finished, click Close , or press <Enter> to apply the settings and close the Options window. Or Click Apply to apply the settings without closing the window. Or Click Cancel to cancel changes. See "Results of transliteration" below. Note: When you retrieve and display WorldCat records, the client marks any auto-transliterated fields with the symbol ▶

Results of transliteration and auto-transliteration

The client:

- Transliterates the romanized data word by word, independently of context.
Note: If context other than that of letters within a word is a factor in the appearance of the Arabic text, you may need to edit the Arabic transliteration. See also the **caution** below.
- Creates an identical field with the same tag number (for example, two 245 tags) to contain the transliterated Arabic script.
- Places the Arabic script field above the associated romanized data field.
- Links the pair of associated fields with a bracket.
- If auto-transliterated (option for bibliographic records only, selected in **Tools > Options > International**), marks transliterated fields with the symbol ▶.

Caution: Transliteration handles the following characters incorrectly. Revise the characters manually.

- The final character taa' marbuta preceded by hamza transliterates incorrectly ashaa'.
- When 'alif maksura is followed by a period, the transliteration omits 'alif maksura.
- 'Alif laam followed by 'alif madda transliterates incorrectly as 'alif laam 'alif.
- Hyphens are incorrectly deleted in transliterated text.
- When laam kasra is followed by siin or jiim, the transliteration omits siin or jiim.
- Laam kasra followed by Haa' transliterates incorrectly as haa'.
- When two laams are followed by capital A (where the first laam is a preposition), the transliteration omits 'alif hamza. However, laam followed by lowercase "a" transliterates correctly as laam 'alif.
- When laam hyphen is followed by damma, the transliteration omits 'alif hamza.

Basis of transliteration

The client transliterates romanized data based on the rules for Arabic given in [ALA-LC Romanization Tables](#).

Use constant data

- Use non-Latin scripts in workforms to create bibliographic constant data records **Cataloging > Create > Constant Data**) or derive new constant data records from existing records (**Edit > Derive > New Constant Data**) in the same way as described above for creating and deriving bibliographic records.
- For constant data in the local file only, you can use non-Latin scripts to name constant data records and search for them by non-Latin script name.
- Also for constant data records in the local file only, you can use non-Latin scripts to specify My Status for records (**Action > Set Status**) and search for non-Latin script My Statuses. (My Status is an optional free-text status that you add to records to help distinguish them.)
- You can change the sort order for results of searching the local constant datafile to Unicode sorting by non-Latin script (default sort order: alphabetical sorting by Latin script). The same setting also determines sort order for WorldCat search results and local save file search results.

- When you apply constant data containing non-Latin script, the client:
 - Replaces a non-repeatable field in the bibliographic record with the paired non-Latin script 880 and equivalent romanized field from the constant data.
 - Or
 - Adds the paired fields from the constant data below existing repeatable fields.
- See general procedures for creating, applying, finding, and using bibliographic constant data in Cataloging/Use Bibliographic Constant Data.

Save records and search save files

- Using the Connexion client only:
 - Save and display non-Latin scripts in records for editing in the online or local save file.
 - The non-Latin data also displays in local save file lists but does not display in online save file lists.
- Note:** If you open an online save file record containing non-Latin script in the browser, the non-Latin script data displays in 880 fields at the end of the record. The record opens in display mode only. You cannot edit or take actions on it.
- You can use non-Latin scripts to specify My Status for records in the online or local save file (**Action > Set Status**) and search for non-Latin script My Statuses. (My Status is an optional free-text status that you add to records to help distinguish them.)
 - **Local file indexes for non-Latin script data.** Use non-Latin scripts to search the following indexes for records in the local save file:
 - **Name**
Index includes non-Latin script fields associated with fields 100, 111, 130, 700, 710, 711, 730 (all subfields).
 - **Title**
Index includes non-Latin script fields associated with fields/subfields:
 - 245 a b f g k n p
 - 246 a b f g n p
 - **My Status**
 - Change the sort order for results of searching the local save file to Unicode sorting by non-Latin script (**default** sort order: alphabetical by Latin script). The same setting also determines sort order for WorldCat search results and local constant data search results.
 - When you save a record online or locally, the client automatically adds field 066 indicating the presence and type of non-Latin script.

See general procedures for saving, finding, and using bibliographic save file records in Cataloging/Save Bibliographic Records.

Report errors in non-Latin scripts

You can use non-Latin script text in the message box of the Report Error window to report errors in non-Latin script records (**Action > Report Error**).

Export records

Selection options for exported data

Select the type and location of Latin script versus non-Latin script data in exported records:

Action	
1	On the Tools menu, click Options (or press <Alt><T><O>), and then click the International tab .
2	Under Export , click one of the following check boxes: <ul style="list-style-type: none"> • Include all data, with other scripts in 880 fields (default) • Include all data, with Latin script in 880 fields • Include Latin script only (deletes field 066 in record) • Include other scripts only
3	When finished, click Close , or press <Enter> to apply the settings and close the Options window. Or Click Apply to apply the settings without closing the window.

Select character set

- Select either MARC-8 (**default**) or UTF-8 Unicode to export records with Arabic, CJK, Cyrillic (within the MARC-8 character set), Greek, or Hebrew scripts. These MARC-8 character sets are subsets of UTF-8 characters, so selecting either character set covers them.
- Use only UTF-8 Unicode to export records with scripts outside of the MARC-8 set (e.g., Armenian and Bengali).
- If any non-MARC-8 characters are exported in MARC-8 data format, they are saved in Numeric Character Reference (NCR) format (see [Numeric character reference](#) for more information).

To select a character set for exporting records:

Action	
1	On the Tools menu, click Options (or press <Alt><T><O>), and then click the Export tab .
2	Click Record Characteristics .
3	Under Bibliographic Records , select one of the following from the Character Set list: <ul style="list-style-type: none"> • MARC-8 (default) • UTF-8 Unicode
4	Click OK , or press <Enter> to save your settings and close the window, or click Cancel to cancel changes. You are returned to the Export page in the Options window. When finished, click Close to close the Options window.

Select fields to delete from exported records

Action	
1	In the Export Options tab (Tools > Options > Export), click Field Export Options .
2	Under Fields to Delete , in the Bibliographic Records text box, enter tag numbers for fields you want to delete in exported bibliographic records. Separate numbers by a comma and a space, or use a hyphen to show a range. Example: <i>920, 938-999</i>
3	Repeat step 2 in the Authority Records text box if needed for exporting authority records.
4	When finished, click OK , or press <Enter> to save your settings and close the window, or click Cancel to cancel changes. You are returned to the Export page in the Options window. When ready, click Close to close the Options window.

Set warning before exporting records with parallel unlinked non-Latin script fields

If Latin script and non-Latin script parallel fields are not linked, display of the non-Latin script in records downloaded to your local system may be affected.

Set an option to get a warning when exporting before the client exports records with unlinked non-Latin script fields in **Tools > Options > Export**.

See general instructions for exporting records in **Cataloging/Export Bibliographic Records**.

Import records

- Select either **MARC-8 (default)** or **UTF-8 Unicode** to import records with Arabic, CJK, Cyrillic (within the MARC-8 character set), Greek, or Hebrew scripts. These MARC-8 character sets are subsets of UTF-8 characters, so selecting either character set covers them.
- Use only **UTF-8 Unicode** to import records with scripts outside of the MARC-8 set (e.g., Armenian and Bengali).

To select a character set for importing records:

Action	
1	On the File menu, click Import Records (or press <Alt><F><I>).
2	In the Import Records window, specify the import file and destination, and then click Record Characteristics .
3	Under Bibliographic Records , select one of the following from the Character Sets list: <ul style="list-style-type: none"> • MARC-8 (default) • UTF-8 Unicode <p>Note: Use only UTF-8 Unicode to import records with scripts outside of the MARC-8 set (e.g., Armenian and Bengali).</p>

	Action
4	Click OK or press <Enter> to save your settings and close the window, or click Cancel to cancel changes. You are returned to the Import Records window.

See general procedures in [Cataloging: Export or Import Bibliographic Records](#).

©, 2016 OCLC Online Computer Library Center, Inc. The following OCLC product and service names are trademarks or service marks of OCLC Online Computer Library Center, Inc.: OCLC, Connexion, WorldCat, and “The world’s libraries. Connected.” The WorldCat symbol and OCLC symbol are registered trademarks of OCLC. Third-party product and service names are trademarks or service marks of their respective companies. OCLC grants permission to photocopy this publication is needed.