

## Bib-1 configuration guideline for Japanese Z39.50 library application

This is the Bib-1 configuration guideline for the Z39.50 target in Japanese library systems, and is used as a complement to the Z39.50 library application specifications. The Bath Profile and the U.S. National Z39.50 Profile are the two most highly used and comprehensive Z39.50 library application specifications. The Bib-1 configuration guideline was created to handle Japanese MARC records and characters, since the library application specifications mentioned previously cannot process them without additional standards.

### *Purpose and scope, and area of application:*

The Library System Study Group of Japan created the Bib-1 configuration guideline for Japanese Z39.50 library application in order to facilitate access to Japanese bibliographic records in the global arena. The Library System Study Group of Japan is an independent organizational body consisting of librarians and systems engineers from academic libraries, research institutions, and system vendors.

The guidelines suggest the minimum requirements for searching library catalogs using Z39.50, and are based on the reasonable assumption that the systems should support MARC21 and ISO10646 as the de facto standards among Japanese library systems. Appendix 2 accompanies the guidelines to have better assumption on index of Japanese library systems.

### *Background information on bibliographic records in Japanese library systems:*

Although ISO10646 seemed to resolve the large problem of handling multilingual characters — including the Japanese characters set — in library systems, there still have been difficulties with the implementation of Z39.50 within library systems in Japan. One of the largest obstacles is the difference between the various MARC formats such as MARC21, Japan/MARC, and other commercial MARC. In spite of some limitations, the various MARC formats have been developed and used in Japanese library systems for a variety of reasons, a discussion of which would be beyond the scope of this paper. Suffice it to say that MARC21 is considered a “single” MARC format in the Bib-1 guideline and is used in place of a domestically developed format because of its widespread use and status as an international standard. However, other problems with the format include the lack of standards for word division and the lack of standard Romanization of words using the Japanese character set. The value setting of “right truncation” in Search Type-1 may ease both problems.

### *Revision history:*

The first version was released on Dec 5, 2003

*Functional requirements:*

Basic bibliographic search and retrieval with a primary focus on library catalogs.

*Stakeholders in the community or the application:*

- Any individuals and institutions that need access to Japanese bibliographic records
- Library systems developers who handle Japanese MARC records with Japanese characters

*Maintenance agency:*

Keio University in Tokyo, Japan serves as the editor, while the Library System Study Group of Japan is the advisor.

*Mechanisms for public review and comment:*

A secretariat of the group receives public comments via email at [libsys-jim@lib.keio.ac.jp](mailto:libsys-jim@lib.keio.ac.jp) and posts them to the mailing list for discussion among the group members. The board members of the Library Systems Study Group moderate the mailing list. Once a consensus is reached on an issue, the decision will appear on the web site of the group at <http://libsys.lib.keio.ac.jp/index.html>

*Acknowledgement:*

The Library System Study Group would like to acknowledge the contributions of participants of the meetings and mailing list, whose discussions have helped develop a consensus for the Bib-1 guidelines.

-----

*Z39.50 Specifications*

Protocol version supported: version 2 is required; version 3 is recommended

Z39.50 Objects:

The followings are objects used in the guideline:

<b>Object</b>	<b>OID</b>
Bib-1 attribute set	1.2.840.10003.3.1
MARC21 record syntax	1.2.840.10003.5.10
XML record syntax	1.2.840.10003.5.109.10
Simple unstructured records syntax (SUTRS)	1.2.840.10003.5.101

Z39.50 Services supported:

- Init
- Search

The guideline requires Z-clients and Z-servers to use Z39.50 Type 1 queries (i.e., general purpose Boolean query structures)

- Present

Sort service shall be supported on Z-clients if needed.

Syntax supported:

- MARC21
- SUTRS
- XML

Character en-code supported:

Requires UTF-8

Note) Z39.50 Maintenance Agency recommends that the Character Set and Language Negotiation should be used to negotiate Unicode with UTF-8 encoding (*see* the response to a question from the ZIG at: <http://lcweb.loc.gov/z3950/agencyh/wisdom/unicode.html>); however, this guideline doesn't require the negotiation, since it has been hardly seen in the Japanese clients and servers. Z39.50 implementers based on this guideline would still keep in mind that the servers may be asked the negotiation and may need to return a UTF-8 character set identifier to the clients.

Bib-1 Attribute Set types and values required:

Attribute Type	Attribute Values	Attribute Names	Default Values
<b>Use (1)</b>	4	title	1016
	7	ISBN	
	8	ISSN	
	21	subject heading,	
	1003	author	
	1016	any	
<b>Relation (2)</b>	3	equal	3
<b>Position (3)</b>	1	first in field	3
	3	any position in field	
<b>Structure (4)</b>	1	Phrase	1
	2	word	
<b>Truncation (5)</b>	1	right truncation	100
	100	do not truncate	
<b>Completeness (6)</b>	1	incomplete subfield	1
	3	complete field	

\* Set of default attribute values and an error message for unsupported values:

When an attribute value is not specified, the server processes a query with operands as the default attribute values, as shown in the above tables. The guidelines strongly recommend that the server return an error message if the server doesn't support the attribute values. This mechanism is necessary to improve the precision and performance of distributed searching.

\* Use / Date of Publication:

The Date of Publication Use Attribute must be used as a search limiter in conjunction with other operands. Z-servers may reject a query that only brings the Date of Publication Use Attribute.

\* Structure:

"Phrase" should be supported as the default, since there are many generic terms in bibliographic records in scientific fields, which may cause a large result set: a search for "computer" as a word results in 80,000 records hit against RLG's Z39.50 server; "science" brings 310,000 records; even "computer AND science" still retrieves 22,000, while "computer science" as a phrase results in 95 records.

\* Truncate:

In addition to “no truncate” as the default, “right truncation” is required to ease the discrepancy of search results caused by the inconsistent word division in Japanese language.

\* Completeness:

Although the default value is “incomplete subfield”, complete field is useful for searches using generic terms such as “nature” as a journal title.

Combinations of recommended attribute values for each specified search:

Author-search

Attribute Type	Attribute Values	Attribute Names
Use(1)	1003	author
Relation(2)	3	equal
Position(3)	3	any position in field
Structure(4)	2	word
Truncation(5)	1	right truncation
Completeness(6)	1	incomplete subfield

Title-search

Attribute Type	Attribute Values	Attribute Names
Use(1)	4	Title
Relation(2)	3	equal
Position(3)	3	any position in field
Structure(4)	1	phrase
Truncation(5)	1	right truncation
Completeness(6)	1	incomplete subfield

ISBN-search

Attribute Type	Attribute Values	Attribute Names
Use(1)	7	ISBN
Relation(2)	3	equal
Position(3)	1	first in field
Structure(4)	1	phrase
Truncation(5)	100	do not truncate
Completeness(6)	1	incomplete subfield

ISSN-search

Attribute Type	Attribute Values	Attribute Names
Use(1)	8	ISSN
Relation(2)	3	equal
Position(3)	1	first in field
Structure(4)	1	phrase
Truncation(5)	100	do not truncate
Completeness(6)	1	incomplete subfield

Subject-search

Attribute Type	Attribute Values	Attribute Names
Use(1)	21	Subject heading
Relation(2)	3	equal
Position(3)	1	first in field
Structure(4)	1	phrase
Truncation(5)	1	right truncation
Completeness(6)	1	incomplete subfield

**Appendix 1:**

Mapping between bib-1 use attributes and MARC21 tags

4	Title	a word, phrase, character, or group of characters, normally appearing in an item, that names the item or the work contained in it.	130, 21X-24X, 440, 490, 730, 740, 830, 840, subfield \$t in the following: 600, 610, 611, 700, 710, 711, 800, 810, 811
7	ISBN		020
8	ISSN		022, 4XX\$x, 7XX\$x
21	Subject	The primary topic on which a work is focused	600,610,611,630, 650, 651, 653, 654, 655, 656,657,69X
1003	Author-name	A personal or corporate author, or a conference or meeting name.	100, 110, 111, 700, 710, 711, 800, 810, 811

**Appendix 2:**

The recommended indexes

While the guideline was created for Z39.50 server and client developers, this appendix was written for Japanese library system developers for an effective application of the guideline.

The recommended indexes to a local bibliographic database for better searching for words using the Japanese character set are:

Word division index in Kanji

Word division index in Kana (either Katakana or Hiragana)

Romanization index

To improve the recall of a search using a Japanese language database, it is important to cover all the notation systems (i.e., Kanji, Kana, and Romanization) in the index and word division index in each notation. There will likely be only one word division index in Katakana unlike the word division index in Kanji. Although the Z39.50 target supports “phrase” search, there is a certain desire to search each word in Kanji. The Romanization index is necessary to provide a retrieval function for organizations without the tools to handle Japanese characters.