

ISO/IEC JTC 1/SC 34

Date: 2006-11-1

ISO/IEC CD 19757-7

ISO/IEC JTC 1/SC 34/WG 1

Secretariat: Standards Council of Canada

Document Schema Definition Languages (DSDL) — Part 7: Character Repertoire Description Language

Warning

This document is not an ISO International Standard. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an International Standard.

Recipients of this document are invited to submit, with their comments, notification of any relevant patent rights of which they are aware and to provide supporting documentation.

Copyright notice

This ISO document is a Draft International Standard and is copyright-protected by ISO. Except as permitted under the applicable laws of the user's country, neither this ISO draft nor any extract from it may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, photocopying, recording or otherwise, without prior written permission being secured.

Requests for permission to reproduce should be addressed to either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
 Case postale 56 x CH-1211 Geneva 20
 Tel. + 41 22 749 01 11
 Fax + 41 22 749 09 47
 E-mail copyright@iso.ch
 Web www.iso.ch

Reproduction may be subject to royalty payments or a licensing agreement.

Violators may be prosecuted.

Contents

Page

Foreword.....	iii
Introduction.....	iv
1 Scope.....	1
2 Normative references.....	1
3 Terms and definitions.....	1
4 Notation.....	2
5 Kernel and hull.....	2
6 Unicode regular expressions.....	2
7 Syntax.....	2
8 Semantics.....	3
8.1 General.....	3
8.2 characterCollection.....	3
8.3 collection.....	3
8.4 union.....	4
8.5 intersection.....	4
8.6 difference.....	5
8.7 ref.....	5
8.8 namedCollection.....	5
9 Validation.....	5
10 Conformance.....	6
Annex A (informative)	7
A.1 8859-6.....	7
A.2 8859-15.....	7
A.3 The Japanese list of kanji characters for the first grade.....	7
A.4 The Japanese list of kanji characters for the second grade.....	9
Bibliography.....	11

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

ISO/IEC 19757-7 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information Technology*, Subcommittee SC 34, Document Description and Processing Languages.

ISO/IEC 19757 consists of the following parts, under the general title *Document Schema Definition Languages (DSDL)*:

- *Part 1: Overview*
- *Part 2: Regular-grammar-based validation — RELAX NG*
- *Part 3: Rule-based validation — Schematron*
- *Part 4: Namespace-based validation dispatching language — NVDL*
- *Part 5: Datatypes*
- *Part 6: Path-based integrity constraints*
- *Part 7: Character repertoire description language — CRDL*
- *Part 8: Declarative document manipulation*
- *Part 9: Datatype- and namespace-aware DTDs*
- *Part 10: Validation management*

Introduction

This International Standard defines a set of Document Schema Definition Languages (DSDL) that can be used to specify one or more validation processes performed against Extensible Markup Language (XML) documents. A number of validation technologies are standardized in DSDL to complement those already available as standards or from industry.

The main objective of this International Standard is to bring together different validation-related technologies to form a single extensible framework that allows technologies to work in series or in parallel to produce a single or a set of validation results. The extensibility of DSDL accommodates validation technologies not yet designed or specified.

This part of ISO/IEC 19757 provides a language for describing collections of characters defined in ISO/IEC 10646 or Unicode or default grapheme clusters defined in UAX#29. Descriptions in this language may be referenced from schemas. Furthermore, they may also be referenced from forms and stylesheets.

NOTE At present, no schema languages provides mechanisms for referencing to CRDL schemas.

Descriptions of collections need not to be exact. To provide non-exact descriptions, this part of ISO/IEC 19757 provides kernels and hulls, which provide the lower limit and upper limits, respectively.

The structure of this part of ISO/IEC 19757 is as follows. Clause 5 introduces kernels and hulls of collections. Clause 6 shows how Unicode regular expression can be used to describe permissible characters and default grapheme characters. Clause 7 describes the syntax of CRDL schemas. Clause 8 describes the semantics of a correct CRDL schema; the semantics specify when a character is contained by a collection described by a CRDL schema. Clause 9 describes modes of CRDL validators.

Document Schema Definition Languages (DSDL) — Part 7: Character Repertoire Description Language

1 Scope

This part of the International Standard specifies a Character Repertoire Description Language (CRDL). A CRDL schema describes a collection of characters defined in ISO/IEC 10646 or Unicode or default grapheme clusters defined in UAX#29.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies. However, ISO/IEC 10646 references to all versions of ISO/IEC 10646, while Unicode references to all versions of Unicode.

Each of the following documents has a unique identifier that is used to cite the document in the text. The unique identifier consists of the part of the reference up to the first comma.

RELAX NG, *ISO/IEC 19757-2, Document Schema Definition Languages (DSDL) — Part 2: Grammar-based validation — RELAX NG*

W3C XML, *Extensible Markup Language (XML) 1.0 (Third Edition)*, W3C Recommendation, 04 February 2004, available at <<http://www.w3.org/TR/2004/REC-xml-20040204/>>

W3C XML-Names, *Namespaces in XML*, W3C Recommendation, 14 January 1999, available at <<http://www.w3.org/TR/1999/REC-xml-names-19990114/>>

W3C XML Schema Part 2, *XML Schema Part 2: Datatypes Second Edition*, W3C Recommendation, 28 October 2004, available at <<http://www.w3.org/TR/2004/REC-xmldata-20041028/>>

IETF RFC 3986, *Uniform Resource Identifiers (URI): Generic Syntax*, Internet Standards Track Specification, January 2005, available at <<http://www.ietf.org/rfc/rfc3987.txt>>

IETF RFC 3987, *Internationalized Resource Identifiers (IRIs)*, Internet Standards Track Specification, January 2005, available at <<http://www.ietf.org/rfc/rfc3987.txt>>

ISO/IEC 10646, *Universal multiple-octet coded Character Set*

Unicode, *The Unicode Standard*

UAX#29, *Text Boundaries*, Unicode Standard Annex #29, 12 October 2006, available at <<http://www.unicode.org/reports/tr29/>>

3 Terms and definitions

For the purposes of this part of ISO/IEC 19757, the term "character" as defined in ISO/IEC 10646 or Unicode, the term "default grapheme cluster" as defined in UAX#29, and the following apply.

3.2 collection

a set of characters or default grapheme clusters

3.3

kernel

a set of characters and default grapheme clusters that are guaranteed to be in the collection

3.4

hull

a set of characters and default grapheme clusters that may be in the collection

4 Notation

$\text{in}(x, A)$: character or default grapheme cluster x is in collection A

$\text{notin}(x, A)$: character or default grapheme cluster x is not in collection A

$\text{unknown}(x, A)$: it is unknown whether character or default grapheme cluster x is in collection A

5 Kernel and hull

A kernel contains characters and default grapheme clusters that are guaranteed to be in the collection; the collection may contain other characters or default grapheme clusters. A hull gives an outer boundary so that characters or default grapheme clusters which are not in the hull are guaranteed not to be in the collection; some characters or default grapheme clusters in the hull may not actually be in the collection.

NOTE Because the repertoire of characters in the Universal Character Set is growing, characters may be continuously added to a set. In this case, it is impossible to specify the set exactly, but it is often possible to specify which character is absolutely included, and which character is absolutely excluded.

6 Unicode regular expressions

Unicode regular expressions defined by W3C XML Schema are used for constraining permissible characters or default grapheme clusters. Default Grapheme clusters are represented by character strings. For example, one of the representations of Latin small e with acute accent is a base character e (U+0065) followed by an acute accent (U+0301). A Unicode regular expression "é" allows this representation.

An expression e is an error when strings u and v match e and the first character of u occurs in v as a non-first character. Conformant implementations of CRDL SHOULD detect this error.

NOTE 1 This error does not happen as long as base characters (the first characters in combining character sequences) do not occur as non-first characters of default grapheme clusters. When this error does not happen, it is possible to compare a given string against a Unicode regular expression without decomposing the string into a sequence of characters or default grapheme clusters. UAX#29 provides the decomposition procedure.

NOTE 2 When a string consisting of a single LF character matches a given Unicode regular expression, it is possible to divide a multi-line string into lines and examine each line independently.

7 Syntax

An CRDL schema in the full syntax shall be an XML document valid against the following RELAX NG schema in the compact syntax.

```
default namespace = "toBeSupplied"

start = element characterCollection {
  commonAtts,
  attribute minUcsVersion {text}?,
  attribute maxUcsVersion {text}?,
```

```

coll}

coll = union | intersection | difference |
      ref | namedCollection | collection

union = element union { commonAtts, coll* }
intersection = element intersection { commonAtts, coll* }
difference = element difference { commonAtts, coll, coll }

ref = element ref { commonAtts, attribute href { xsd:anyURI }}
namedCollection = element namedCollection { commonAtts, attribute ns { xsd:NCName } }

collection =
  element collection {
    commonAtts,
    (text |
     element kernel { commonAtts, text } |
     element hull { commonAtts, text } |
     (element kernel { commonAtts, text }, element hull { commonAtts, text })))
  }

commonAtts =
  attribute xml:id {xsd:ID}

```

The character content of a collection, kernel or hull element is a regular expression as specified in Annex F of W3C XML Schema Part 2.

8 Semantics

8.1 General

Given a character or grapheme x and a collection A , either $\text{in}(x, A)$, $\text{notin}(x, A)$, or $\text{unknown}(x, A)$ holds.

8.2 characterCollection

The root element of a CRVL schema specifies `characterCollection` as the tag name.

The semantics of regular expressions occurring in a CRVL schema depend on the version of the Unicode standard. The author of a CRVL schema may specify the intended versions with the `minUcsVersion` and `maxUcsVersion`. If the CRDL processor cannot use some version between these two, it should report an error and may stop normal processing.

The semantics of `<characterCollection .../> A </characterCollection>` is defined below.

- $\text{in}(x, \text{<characterCollection .../> } A \text{ </characterCollection>})$ when $\text{in}(x, A)$.
- $\text{notin}(x, \text{<characterCollection .../> } A \text{ </characterCollection>})$ when $\text{notin}(x, A)$.
- $\text{unknown}(x, \text{<characterCollection .../> } A \text{ </characterCollection>})$ when $\text{unknown}(x, A)$.

8.3 collection

The semantics of `<collection> ... </collection>` is defined below.

- Case 1: the content of the collection element is text

It is assumed that this element has a kernel element and a hull element that specifies the content of this collection. The rest is the same as in Case 4.

- Case 2: the collection element has a kernel element but does not have a hull element.
 - $\text{in}(x, \langle \text{collection} \rangle \dots \langle / \text{collection} \rangle)$ when x matches the Unicode regular expression specified as the content of the kernel element.
 - $\text{notin}(x \langle \text{collection} \rangle \dots \langle / \text{collection} \rangle)$ never holds.
 - $\text{unknown}(x \langle \text{collection} \rangle \dots \langle / \text{collection} \rangle)$ when $\text{in}(x, \langle \text{collection} \rangle \dots \langle / \text{collection} \rangle)$ does not hold.
- Case 3: the given collection element has a hull element but does not have a kernel element.
 - $\text{in}(x, \langle \text{collection} \rangle \dots \langle / \text{collection} \rangle)$ never holds
 - $\text{notin}(x \langle \text{collection} \rangle \dots \langle / \text{collection} \rangle)$ when x does not match the Unicode regular expression specified as the content of the hull element.
 - $\text{unknown}(x \langle \text{collection} \rangle \dots \langle / \text{collection} \rangle)$ when $\text{notin}(x, \langle \text{collection} \rangle \dots \langle / \text{collection} \rangle)$ does not hold.
- Case 4: the given collection element has a hull element and a kernel element.
 - $\text{in}(x, \langle \text{collection} \rangle \dots \langle / \text{collection} \rangle)$ when x matches the Unicode regular expression specified as the content of the kernel element.
 - $\text{notin}(x \langle \text{collection} \rangle \dots \langle / \text{collection} \rangle)$ when $\text{in}(x, \langle \text{collection} \rangle \dots \langle / \text{collection} \rangle)$ does not hold and x does not match the Unicode regular expression specified as the content of the hull element.
 - $\text{unknown}(x \langle \text{collection} \rangle \dots \langle / \text{collection} \rangle)$ otherwise.

8.4 union

A character or default grapheme cluster is in $\langle \text{union} \rangle A B \dots \langle / \text{union} \rangle$ if and only if it is in A or B . It is not in the union if and only if neither it is in A nor it is in B .

- $\text{in}(x, \langle \text{union} \rangle A B \langle / \text{union} \rangle)$ when $\text{in}(x, A)$ or $\text{in}(x, B)$.
- $\text{notin}(x, \langle \text{union} \rangle A B \langle / \text{union} \rangle)$ when $\text{notin}(x, A)$ and $\text{notin}(x, B)$.
- $\text{unknown}(x, \langle \text{union} \rangle A B \langle / \text{union} \rangle)$ otherwise.

8.5 intersection

A character or default grapheme cluster is in $\langle \text{intersection} \rangle A B \dots \langle / \text{intersection} \rangle$ if and only if it is in A and B . It is not in the intersection if and only if either it is not in A or it is not in B .

- $\text{in}(x, \langle \text{intersection} \rangle A B \dots \langle / \text{intersection} \rangle)$ when $\text{in}(x, A)$ and $\text{in}(x, B)$.
- $\text{notin}(x, \langle \text{intersection} \rangle A B \dots \langle / \text{intersection} \rangle)$ when $\text{notin}(x, A)$ or $\text{notin}(x, B)$
- $\text{unknown}(x, \langle \text{intersection} \rangle A B \dots \langle / \text{intersection} \rangle)$ otherwise.

8.6 difference

A character or default grapheme cluster is in $\langle \text{difference} \rangle A B \dots \langle / \text{difference} \rangle$ if and only if it is in A and it is not in B . It is not in the difference if and only if either it is not in A or it is in B .

- $\text{in}(x, \langle \text{difference} \rangle A B \dots \langle / \text{difference} \rangle)$ when $\text{in}(x, A)$ and $\text{notin}(x, B)$
- $\text{notin}(x, \langle \text{difference} \rangle A B \dots \langle / \text{difference} \rangle)$ when $\text{notin}(x, A)$ or $\text{in}(x, B)$
- $\text{unknown}(x, \langle \text{difference} \rangle A B \dots \langle / \text{difference} \rangle)$ otherwise.

8.7 ref

Given $\langle \text{ref href}="uri"/ \rangle$, a CRDL schema (say S) shall be obtained by dereferencing uri . When dereferencing uri is not successful (e.g., network errors), the CRDL processor should report an error and should continue normal processing by assuming that "unknown" holds. If it is successful, the semantics is defined below:

- $\text{in}(x, \langle \text{ref href}="uri"/ \rangle)$ when $\text{in}(x, S)$.
- $\text{notin}(x, \langle \text{ref href}="uri"/ \rangle)$ when $\text{notin}(x, S)$.
- $\text{unknown}(x, \langle \text{ref href}="uri"/ \rangle)$ when $\text{unknown}(x, S)$.

NOTE The values of the `minUcsVersion` and `maxUcsVersion` attributes of the referenced schema are ignored.

8.8 namedCollection

$\langle \text{namedCollection name}="ncname"/ \rangle$ allows the use of named collections. When $ncname$ is registered at IANA, the semantics is provided by the registration information. When $ncname$ is an implementation-dependent encoding name, the semantics is implementation-dependent.

When the CRDL processor does not support the named collection $ncname$, it should report an error and continue normal processing by assuming that "unknown" holds.

9 Validation

A CRDL processor is a computer program that validates characters and default grapheme clusters against CRDL schemas.

A CRDL processor receives a character or default grapheme cluster and a CRDL schema. When a Unicode string

is given, a CRDL processor can be used only after decomposing the string into a sequence of characters and default grapheme clusters. UAX#29 provides the decomposition procedure.

10 Conformance

Different conformant CRDL processors may report different results. There are three reasons for such discrepancies.

The first reason is that dereferencing IRIs may fail. However, the semantics of CRDL is defined so that such failures make conformant CRDL processors err on the safe side. In other words, such failures do not lead to "in" when "notin" or "unknown" would have been reported, and do not lead to "notin" when "in" or "unknown" would have been reported.

The second reason is that the semantics of Unicode regular expressions depends on the Unicode version. Different conformant CRDL processors may behave very differently. For example, one may report "in", while another, "notin".

The third reason is that encodings are implementation dependent. Again, different conformant CRDL processors may behave very differently.

Annex A (informative)

A.1 8859-6

The charset ISO-8859-6 is described by the following CRDL schema.

```
<characterCollection xmlns="toBeSupplied">
  <collection>[&#x20;-&#x7F;]|&#xA0;|&#xA4;|&#xAD;|&#x60C;|
    &#x61B;|&#x61F;|[&#x621;-&#x63A;]|[&#x640;-&#x652;]</collection>
</characterCollection>
```

An alternative representation is shown below.

```
<characterCollection xmlns="toBeSupplied">
  <union>
    <collection>[&#x20;-&#x7F;]</collection>
    <collection>&#xA0;</collection>
    <collection>&#xA4;</collection>
    <collection>&#xAD;</collection>
    <collection>&#x60C;</collection>
    <collection>&#x61B;</collection>
    <collection>&#x61F;</collection>
    <collection>[&#x621;-&#x63A;]</collection>
    <collection>[&#x640;-&#x652;]</collection>
  </union>
</characterCollection>
```

A.2 8859-15

The charset ISO-8859-15 is described by the following CRDL schema.

```
<characterCollection xmlns="toBeSupplied">
  <collection>[&#x20;-&#x7F;]| [&#xA0;-&#xA3;]| &#xA5; | &#xA7; |
    [&#xA9;-&#xB3;]| [&#xB5;-&#xB7;]| [&#xB9;-&#xBB;]|
    [&#xBF;-&#xFF;]| [&#x152;-&#x153;]| [&#x160;-&#x161;]|
    &#x178; | [&#x17D;-&#x17E;]| &#x20AC;</collection>
</characterCollection>
```

A.3 The Japanese list of kanji characters for the first grade

The Japanese Ministry of Education maintains six lists of kanji characters. The list for the first grade is described by the following CRDL schema.

```
<characterCollection xmlns="toBeSupplied">
  <union>
    <collection>&#x4E00;</collection>
    <collection>&#x4E03;</collection>
    <collection>[&#x4E09;-&#x4E0B;]</collection>
    <collection>&#x4E2D;</collection>
    <collection>&#x4E5D;</collection>
    <collection>&#x4E8C;</collection>
    <collection>&#x4E94;</collection>
    <collection>&#x4EBA;</collection>
    <collection>&#x4F11;</collection>
    <collection>&#x5148;</collection>
    <collection>&#x5165;</collection>
    <collection>&#x516B;</collection>
  </union>
</characterCollection>
```

<collection>#x516D;</collection>
 <collection>#x5186;</collection>
 <collection>#x51FA;</collection>
 <collection>#x529B;</collection>
 <collection>#x5341;</collection>
 <collection>#x5343;</collection>
 <collection>#x53E3;</collection>
 <collection>#x53F3;</collection>
 <collection>#x540D;</collection>
 <collection>#x56DB;</collection>
 <collection>#x571F;</collection>
 <collection>#x5915;</collection>
 <collection>#x5927;</collection>
 <collection>#x5929;</collection>
 <collection>#x5973;</collection>
 <collection>#x5B50;</collection>
 <collection>#x5B57;</collection>
 <collection>#x5B66;</collection>
 <collection>#x5C0F;</collection>
 <collection>#x5C71;</collection>
 <collection>#x5DDD;</collection>
 <collection>#x5DE6;</collection>
 <collection>#x5E74;</collection>
 <collection>#x624B;</collection>
 <collection>#x6587;</collection>
 <collection>#x65E5;</collection>
 <collection>#x65E9;</collection>
 <collection>#x6708;</collection>
 <collection>#x6728;</collection>
 <collection>#x672C;</collection>
 <collection>#x6751;</collection>
 <collection>#x6797;</collection>
 <collection>#x6821;</collection>
 <collection>#x68EE;</collection>
 <collection>#x6B63;</collection>
 <collection>#x6C17;</collection>
 <collection>#x6C34;</collection>
 <collection>#x706B;</collection>
 <collection>#x72AC;</collection>
 <collection>#x7389;</collection>
 <collection>#x738B;</collection>
 <collection>#x751F;</collection>
 <collection>#x7530;</collection>
 <collection>#x7537;</collection>
 <collection>#x753A;</collection>
 <collection>[#x767D;-#x767E;]</collection>
 <collection>#x76EE;</collection>
 <collection>#x77F3;</collection>
 <collection>#x7A7A;</collection>
 <collection>#x7ACB;</collection>
 <collection>#x7AF9;</collection>
 <collection>#x7CF8;</collection>
 <collection>#x8033;</collection>
 <collection>#x82B1;</collection>
 <collection>#x8349;</collection>
 <collection>#x866B;</collection>
 <collection>#x898B;</collection>
 <collection>#x8C9D;</collection>
 <collection>#x8D64;</collection>
 <collection>#x8DB3;</collection>
 <collection>#x8ECA;</collection>
 <collection>#x91D1;</collection>
 <collection>#x96E8;</collection>
 <collection>#x9752;</collection>
 <collection>#x97F3;</collection>
 </union>

</characterCollection>

NOTE One could use a single regular expression since this list has 80 characters only. However, some other lists of kanji characters have thousands of kanji characters, which prohibit the use of a single regular expression.

A.4 The Japanese list of kanji characters for the second grade

The list for the second grade is described by the following CRDL schema.

```
<characterCollection xmlns="toBeSupplied">
  <union>
    <ref href="theURIforTheFirstGrade"/>
    <collection>&#x4E00;</collection>
    <collection>&#x4E03;</collection>
    <collection>[&#x4E09;-&#x4E0B;]</collection>
    <collection>&#x4E2D;</collection>
    <collection>&#x4E5D;</collection>
    <collection>&#x4E8C;</collection>
    <collection>&#x4E94;</collection>
    <collection>&#x4EBA;</collection>
    <collection>&#x4F11;</collection>
    <collection>&#x5148;</collection>
    <collection>&#x5165;</collection>
    <collection>&#x516B;</collection>
    <collection>&#x516D;</collection>
    <collection>&#x5186;</collection>
    <collection>&#x51FA;</collection>
    <collection>&#x529B;</collection>
    <collection>&#x5341;</collection>
    <collection>&#x5343;</collection>
    <collection>&#x53E3;</collection>
    <collection>&#x53F3;</collection>
    <collection>&#x540D;</collection>
    <collection>&#x56DB;</collection>
    <collection>&#x571F;</collection>
    <collection>&#x5915;</collection>
    <collection>&#x5927;</collection>
    <collection>&#x5929;</collection>
    <collection>&#x5973;</collection>
    <collection>&#x5B50;</collection>
    <collection>&#x5B57;</collection>
    <collection>&#x5B66;</collection>
    <collection>&#x5C0F;</collection>
    <collection>&#x5C71;</collection>
    <collection>&#x5DDD;</collection>
    <collection>&#x5DE6;</collection>
    <collection>&#x5E74;</collection>
    <collection>&#x624B;</collection>
    <collection>&#x6587;</collection>
    <collection>&#x65E5;</collection>
    <collection>&#x65E9;</collection>
    <collection>&#x6708;</collection>
    <collection>&#x6728;</collection>
    <collection>&#x672C;</collection>
    <collection>&#x6751;</collection>
    <collection>&#x6797;</collection>
    <collection>&#x6821;</collection>
    <collection>&#x68EE;</collection>
    <collection>&#x6B63;</collection>
    <collection>&#x6C17;</collection>
    <collection>&#x6C34;</collection>
    <collection>&#x706B;</collection>
    <collection>&#x72AC;</collection>
    <collection>&#x7389;</collection>
    <collection>&#x738B;</collection>
```

```
<collection>#x751F;</collection>
<collection>#x7530;</collection>
<collection>#x7537;</collection>
<collection>#x753A;</collection>
<collection>[#x767D;-&#x767E;]</collection>
<collection>#x76EE;</collection>
<collection>#x77F3;</collection>
<collection>#x7A7A;</collection>
<collection>#x7ACB;</collection>
<collection>#x7AF9;</collection>
<collection>#x7CF8;</collection>
<collection>#x8033;</collection>
<collection>#x82B1;</collection>
<collection>#x8349;</collection>
<collection>#x866B;</collection>
<collection>#x898B;</collection>
<collection>#x8C9D;</collection>
<collection>#x8D64;</collection>
<collection>#x8DB3;</collection>
<collection>#x8ECA;</collection>
<collection>#x91D1;</collection>
<collection>#x96E8;</collection>
<collection>#x9752;</collection>
<collection>#x97F3;</collection>
</union>
</characterCollection>
```

Bibliography

- [1] *A Notation for Character Collections for the WWW*, W3C Note, 14 January 2000, available at <http://www.w3.org/TR/charcol>