|  |
| --- |
|  |
| **Interface Specification**  KSP**-**12746 |
| **CPR Query Service (Solr)** |
| Prepared for: |
| KOMBIT  Halfdansgade 8  2300 København S |
|  |
|  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Versionskontrol udført af KOMBIT efter dette dokument er blevet frigivet fra SSE**   |  |  |  | | --- | --- | --- | | **Dato** | **Kommentar** | **Initialer** | | 2021.03.19 | Rettet ”historiskkommunecvr” så det optræder med 2xm” I stavningen. | KRR | | |
| Interface Specification | |
| Project: | $KSP**-**12746 |
| Revision: | $Revision: 1.23$ $Date: 23 Jun 2016$ |
| Document: | $SSE/12746/IFS/0004 |
| Copyright (c) 2016 by Systematic Group. It shall not be copied, reproduced, disclosed or otherwise made available to third party without previous consent from Systematic Group | |

Table of Contents

[1 Introduction 4](#_Toc454451681)

[2 Identification of the service 4](#_Toc454451682)

[3 Description 4](#_Toc454451683)

[3.1 callCPRPersonList operation 4](#_Toc454451684)

[3.1.1 searchParameter 4](#_Toc454451685)

[3.1.2 Other parameters 6](#_Toc454451686)

[3.1.3 Parameters that must not be used 6](#_Toc454451687)

[3.1.4 Nested queries 6](#_Toc454451688)

[3.1.5 Indexed fields 6](#_Toc454451689)

[3.1.6 Other rules 6](#_Toc454451690)

[4 Error handling 6](#_Toc454451691)

[4.1 List of faultstrings 7](#_Toc454451692)

[4.1.1 Fixed faultstrings 7](#_Toc454451693)

[5 Query examples 8](#_Toc454451694)

[5.1 Wildcard query on a partial name 8](#_Toc454451695)

[5.2 Query with AND and OR. 8](#_Toc454451696)

[5.3 Query for an interval of husnummer 8](#_Toc454451697)

[5.4 Query, boosting fields importance in the query 8](#_Toc454451698)

[5.5 Query for by:Aalborg 9](#_Toc454451699)

[5.6 Query for previous adresses 9](#_Toc454451700)

[5.7 Query for distrikt 9](#_Toc454451701)

[6 Encoding of special characters 9](#_Toc454451702)

[6.1 '=' and '&' 9](#_Toc454451703)

[6.2 Solr special characters 9](#_Toc454451704)

[6.3 Mapping of Æ, Ø and Å 10](#_Toc454451705)

[Appendix A. Appendix 11](#_Toc454451706)

[A.1 Solr schema 11](#_Toc454451707)

[6.4 Fields 11](#_Toc454451708)

[A.2 Solr type definitions from the schema 12](#_Toc454451709)

[A.3 Character mapping 13](#_Toc454451710)

# Introduction

In this document a short overview of the CPR Query Service version 1 will be given. The target readers are developers of systems needing to search the CPR replica in Service Platformen.

The service is a Soap web service using the standard token based security model.

# Identification of the service

CPRQueryService 1.0

ServiceUUID: **3aed2c45-5ec0-4c85-9435-9ef71e7d7ee1**

End point: https://<base-url>/service/CPR/CPRQueryService/1

where <base-url> is the called environment, e.g. “exttest.serviceplatformen.dk”.

WSDL: https://<base-url>/service/CPR/CPRQueryService/1?wsdl

Security: Token based security.

# Description

The service provides one operation – callCPRPersonList.

## callCPRPersonList operation

<ns2:callCPRPersonListRequest xmlns:ns2= **"http://serviceplatformen.dk/xml/wsdl/soap11/CPRQueryService/1/"** xmlns:ns1=**"http://serviceplatformen.dk/xml/schemas/CallContext/1/"**>

<!--Optional:-->

<ns1:CallContext>

<!--You may enter the optional following 3 items in any order-->

<ns1:OnBehalfOfUser>**?**</ns1:OnBehalfOfUser><ns1:CallersServiceCallIdentifier>**?**</ns1:CallersServiceCallIdentifier>

<ns1:AccountingInfo>**?**</ns1:AccountingInfo>

</ns1:CallContext>

<ns2:searchParameter>**q=\*:\***&rows=**250**&defType=**edismax**</ns2:searchParameter>

</ns2:callCPRPersonListRequest>

### searchParameter

The searchParameter element must contain a standard solr query string. This string will be validated by this webservice. Then it will be forwarded to a running instance of Solr where it will be processed by Solr's Extended DisMax (eDisMax) query parser.

The format of the search string is in general:

parameter1=value1&parameter2=value2& ... &parameter\_n=value\_n

Example of a search string:

q=fornavn:Ben\*&fq=efternavn:Hansen&start=10&rows=50&defType=edismax&qf=fornavn^2&qf=efternavn^3&stopwords=true&lowercaseOperators=true

#### q

The q parameter defines the main "query" constituting the essence of the search. The parameter supports raw input strings provided by users with no special escaping. The + and - characters are treated as "mandatory" and "prohibited" modifiers for terms. Text wrapped in balanced quote characters (for example, "San Jose") is treated as a phrase. Any query containing an odd number of quote characters is evaluated as if there were no quote characters at all. The q parameter must be provided, for solr to be able to find anything.

#### fq

The fq parameter defines the “filter query”, and follows the same syntax rules as q. The fq parameter can be specified multiple times. Using the fq parameter can improve efficiency by using the filter cache.

#### qf

The qf parameter defines the “query fields”. This parameter can be used to increase or decrease particular field’s importance in the query

#### defType

Must either be set to edismax or not set. If not set, it will default to edismax. If it is set to other values, an error is returned.

#### start

Specifies an offset into the responses at which Solr should begin displaying content. If the parameter is not set, the default value 0 will be used. The maximum value is 5000. If a larger value is used, an error is returned.

#### rows

Controls how many rows of responses are displayed at a time. If the parameter is not set, the default value 10 will be used. The maximum value is 250. If a larger value is used, an error is returned.

|  |  |
| --- | --- |
| timeAllowed |  |

Defines the time allowed for the query to be processed. If the time elapses before the query response is complete, partial information may be returned.

If the parameter is not set, the default value 1800 will be used. The maximum value is 1800. If a larger value is used, an error is returned.

The value 0 and negative values are also not allowed. If they are used, an error is returned.

### Other parameters

The possible parameters are all the parameters defined for the eDisMax including the DisMax parser and the common query parameters. These parameters will not be listed here. Instead see the online documentation:

https://cwiki.apache.org/confluence/display/solr/The+Extended+DisMax+Query+Parser

https://cwiki.apache.org/confluence/display/solr/The+DisMax+Query+Parser

https://cwiki.apache.org/confluence/display/solr/Common+Query+Parameters

These links where working at the time when this document was written. It can not be guaranteed that the information will not move to somewhere else in the future.

### Parameters that must not be used

fl, because this is set to the precomputed result.

wt, because it is set to javabin.

debug, because the service will not return debug information.

Id, a field for solr

If any of there are used, an error is returned.

### Nested queries

Nested queries are not allowed. This means that any use of the pseudo-field \_query\_ will return an error.

### Indexed fields

These fields are searchable: cprnummer, fornavn, mellemnavn, efternavn, fuldenavn, foedselsdato, koen, doed (boolean),

vejnavn, husnummer, husnummerlige (boolean), etage, sidedoer, standardadresse, postnummer, by,distrikt, kommunecvr, vejkode,

historiskhusnummer, historiskvejnavn, historisketage, historisksidedoer, historiskkommunecvr, historiskvejkode.

### Other rules

The field id is used in Solr but it is not allowed to use in queries. If it is used, an error will be returned.

The field distrikt must not be used unless it is used together with the field kommunecvr. If distrikt is used alone, an error will be returned.

# Error handling

CPR Query service operations may return errors to signal that input could not be validated or that processing of a request failed. Errors are returned as SOAPFaultExceptions.

The following is an example of a failed query to callCPRPersonListRequest:

<env:Envelope xmlns:env=**"http://schemas.xmlsoap.org/soap/envelope/"**>

<env:Header/>

<env:Body>

<env:Fault>

<faultcode xmlns:codeNS=**"http://schemas.xmlsoap.org/soap/envelope/"**>**codeNS:Client**</faultcode>

<faultstring>**Only defType allowed is edismax**</faultstring>

</env:Fault>

</env:Body>

</env:Envelope>

## List of faultstrings

### Fixed faultstrings

The following faultstrings can be returned by the service:

|  |  |  |
| --- | --- | --- |
| Faultcode | Faultstring | Meaning |
| Client | It is not allowed to search on distrikt, without kommunecvr (q=distrikt:"someValue" AND kommunecvr:"someCvr" | The field distrikt was used without being combined with kommunecvr in the q parameter. |
| distrikt not allowed in qf | The field distrikt was used in the qf parameter. |
| distrikt not allowed in fq without immediately followed by kommunecvr | The field distrikt was used without being followed by kommunecvr in the fq parameter. |
| id is not allowed as search parameter | The field id was used in the q parameter. |
| id is not allowed as qf parameter | The field id was used in the qf parameter. |
| id is not allowed as fq parameter | The field id was used in the fq parameter. |
| Only defType allowed is edismax | A value different from edismax was used in the defType parameter. |
| timeAllowed too high (max 1800) | A value > 1800 was used in the timeAllowed parameter. |
| timeAllowed too low (min 1) | A value < 1 was used in the timeAllowed parameter. |
| start too high (max 5000) | A value > 5000 was used in the start parameter. |
| rows too high (max 250) | A value > 250 was used in the rows parameter. |
| Not allowed to set fl | The fl parameter was used in a query. |
| Not allowed to set wt | The wt parameter was used in a query. |
| Not allowed to set debug | The debug parameter was used in a query. |
| Server | Unexpected Error | The server can not fullful the service request at the moment. |
| InvalidTokenServiceToken | Secure Token is not valid: XXX | The supplied token is not valid. |
| TokenServiceCertificateRevoked | The certificate of the Secure Token Service has been revoked! | The certificate of the Secure Token Service has been revoked! |
| MessageTooLarge | The request was larger than 10 mb and therefore rejected | The content length of the request was larger than 10 MiB. |
| ServiceSuspended | Service with UUID XXX is suspended | The service has been suspended. |

# Query examples

## Wildcard query on a partial name

q= fornavn:I\*

Finds (for example): Ivar, Ida and Ingolf.

## Query with AND and OR.

q=(fornavn:Alexandra OR fornavn:Henriette) AND efternavn:U\*

Finds Alexandra Urmani and Henriette Udsen.

## Query for an interval of husnummer

q=husnummer:[004 TO 012]

## Query, boosting fields importance in the query

q=fornavn:Arne AND efternavn:\*sen &qf=fornavn^2 efternavn^1.5

Default boost of fields is 1.

The distrikt field cannot be boosted.

## Query for by:Aalborg

q=by:"Aalborg SØ"

Matches only Aalborg SØ.

q=by:"Aalborg"

Matches both Aalborg and Aalborg SØ

## Query for previous adresses

q=husnummer:077A OR historiskhusnummer:077A

## Query for distrikt

Queries that contain '=' or '&' must use url encoding for these characters (see section 6), and search strings with “=” or “&” must be wrapped in quotations(“ ”).

To query for distrikt, distrikt must be combined(AND) with kommunecvr

This query searches for Social=73 in a municipality with cvr 12345678:

q=distrikt:("Social%3D73\*" AND \*vej\*) AND kommunecvr:12345678

q=kommunecvr:12345678 AND distrikt:("Social%3D73\*" AND \*vej\*)

Query for persons in social district 73, in municipality with cvr 12345678, with a fritekst string containing *vej* in a distrikt field.

Possible distrikter(*Byfornyelse, Diverse, Evakuer, Kirke, Skole, Befolk, Social, Sogne, valg, varme*)

# Encoding of special characters

## '=' and '&'

The characters '=' and '&' have special meanings as sepatators in the query string.

Searches for strings that contain '=' or '&' is possible by url encoding these characters.

Replace the character with the encoded string in the query and wrap the entire string in quotations(“ ”).

|  |  |
| --- | --- |
| Character | Encoding |
| & | %26 |
| = | %3D |

## Solr special characters

The special characters in solr are: + - && || ! ( ) { } [ ] ^ " ~ \* ? : \

To use them for searches they must be escaped with \ before the character. For example to search for (1+1):2 use the query: \(1\+1\)\:2 .

## Mapping of Æ, Ø and Å

The danish characters æ, ø and å (and their upper case versions Æ, Ø and Å) are mapped to ASCII characters in Solr.

The mappings are 'Æ' -> 'AE', 'Ø' -> 'O' and 'Å' -> 'A'.

This means that for example a query for efternavn:Jo\* will find both people named Johansen and Jørgensen. The query with efternavn:Jø\* will find the same people.

The mapping of all characters can be found in the appendix.

1. Appendix
   1. Solr schema

The fields and type definitions from the schema can be used if you want to perform experiments with your own Solr installation.

## Fields

Here are the indexed field names from the schema.

<field name="cprnummer" type="string" indexed="true" stored="true" required="false" multiValued="false" />

<field name="fornavn" type="text\_general" indexed="true" stored="true" required="false" multiValued="false" />

<field name="mellemnavn" type="text\_general" indexed="true" stored="true" required="false" multiValued="false" />

<field name="efternavn" type="text\_general" indexed="true" stored="true" required="false" multiValued="false" />

<field name="fuldenavn" type="text\_general" indexed="true" stored="true" required="false" multiValued="false" />

<field name="foedselsdato" type="string" indexed="true" stored="true" required="false" multiValued="false" />

<field name="koen" type="text\_general" indexed="true" stored="true" required="false" multiValued="false" />

<field name="doed" type="boolean" indexed="true" stored="true" required="false" multiValued="false" />

<field name="vejnavn" type="text\_general" indexed="true" stored="true" required="false" multiValued="false" />

<field name="husnummer" type="text\_general" indexed="true" stored="true" required="false" multiValued="false" />

<field name="husnummerlige" type="boolean" indexed="true" stored="true" required="false" multiValued="false" />

<field name="etage" type="string" indexed="true" stored="true" required="false" multiValued="false" />

<field name="sidedoer" type="text\_general" indexed="true" stored="true" required="false" multiValued="false" />

<field name="standardadresse" type="text\_general" indexed="true" stored="true" required="false" multiValued="false" />

<field name="postnummer" type="string" indexed="true" stored="true" required="false" multiValued="false" />

<field name="by" type="text\_general" indexed="true" stored="true" required="false" multiValued="false" />

<field name="distrikt" type="text\_general" indexed="true" stored="true" required="false" multiValued="true" />

<field name="kommunecvr type="text\_general" indexed="true" stored="true" required="false" multiValued="false" />

<field name="vejkode" type="string" indexed="true" stored="true" required="false" multiValued="false" />

<field name="historiskhusnummer" type="text\_general" indexed="true" stored="true" required="false" multiValued="true" />

<field name="historiskvejnavn" type="text\_general" indexed="true" stored="true" required="false" multiValued="true" />

<field name="historisketage" type="text\_general" indexed="true" stored="true" required="false" multiValued="true" />

<field name="historisksidedoer" type="text\_general" indexed="true" stored="true" required="false" multiValued="true" />

<field name="historiskkommunecvr" type="text\_general" indexed="true" stored="true" required="false" multiValued="true" />

<field name="historiskvejkode" type="string" indexed="true" stored="true" required="false" multiValued="true" />

* 1. Solr type definitions from the schema

<!-- The StrField type is not analyzed, but indexed/stored verbatim.

It supports doc values but in that case the field needs to be

single-valued and either required or have a default value.

-->

<fieldType name="string" class="solr.StrField" sortMissingLast="true" />

<!-- boolean type: "true" or "false" -->

<fieldType name="boolean" class="solr.BoolField" sortMissingLast="true"/>

<!-- A general text field that has reasonable, generic

cross-language defaults: it tokenizes with StandardTokenizer,

removes stop words from case-insensitive "stopwords.txt"

(empty by default), and down cases. At query time only, it

also applies synonyms. -->

<fieldType name="text\_general" class="solr.TextField" positionIncrementGap="100">

<analyzer type="index">

<tokenizer class="solr.StandardTokenizerFactory"/>

<filter class="solr.StopFilterFactory" ignoreCase="true" words="stopwords.txt" />

<filter class="solr.LowerCaseFilterFactory"/>

<filter class="solr.ASCIIFoldingFilterFactory"/>

</analyzer>

<analyzer type="query">

<tokenizer class="solr.StandardTokenizerFactory"/>

<filter class="solr.StopFilterFactory" ignoreCase="true" words="stopwords.txt" />

<filter class="solr.SynonymFilterFactory" synonyms="synonyms.txt" ignoreCase="true" expand="true"/>

<filter class="solr.LowerCaseFilterFactory"/>

<filter class="solr.ASCIIFoldingFilterFactory"/>

</analyzer>

</fieldType>

* 1. Character mapping

This is the contents of the file mapping-FoldToASCII.txt

# The ASF licenses this file to You under the Apache License, Version 2.0

# (the "License"); you may not use this file except in compliance with

# the License. You may obtain a copy of the License at

#

# http://www.apache.org/licenses/LICENSE-2.0

#

# Unless required by applicable law or agreed to in writing, software

# distributed under the License is distributed on an "AS IS" BASIS,

# WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.

# See the License for the specific language governing permissions and

# limitations under the License.

# This map converts alphabetic, numeric, and symbolic Unicode characters

# which are not in the first 127 ASCII characters (the "Basic Latin" Unicode

# block) into their ASCII equivalents, if one exists.

#

# Characters from the following Unicode blocks are converted; however, only

# those characters with reasonable ASCII alternatives are converted:

#

# - C1 Controls and Latin-1 Supplement: http://www.unicode.org/charts/PDF/U0080.pdf

# - Latin Extended-A: http://www.unicode.org/charts/PDF/U0100.pdf

# - Latin Extended-B: http://www.unicode.org/charts/PDF/U0180.pdf

# - Latin Extended Additional: http://www.unicode.org/charts/PDF/U1E00.pdf

# - Latin Extended-C: http://www.unicode.org/charts/PDF/U2C60.pdf

# - Latin Extended-D: http://www.unicode.org/charts/PDF/UA720.pdf

# - IPA Extensions: http://www.unicode.org/charts/PDF/U0250.pdf

# - Phonetic Extensions: http://www.unicode.org/charts/PDF/U1D00.pdf

# - Phonetic Extensions Supplement: http://www.unicode.org/charts/PDF/U1D80.pdf

# - General Punctuation: http://www.unicode.org/charts/PDF/U2000.pdf

# - Superscripts and Subscripts: http://www.unicode.org/charts/PDF/U2070.pdf

# - Enclosed Alphanumerics: http://www.unicode.org/charts/PDF/U2460.pdf

# - Dingbats: http://www.unicode.org/charts/PDF/U2700.pdf

# - Supplemental Punctuation: http://www.unicode.org/charts/PDF/U2E00.pdf

# - Alphabetic Presentation Forms: http://www.unicode.org/charts/PDF/UFB00.pdf

# - Halfwidth and Fullwidth Forms: http://www.unicode.org/charts/PDF/UFF00.pdf

#

# See: http://en.wikipedia.org/wiki/Latin\_characters\_in\_Unicode

#

# The set of character conversions supported by this map is a superset of

# those supported by the map represented by mapping-ISOLatin1Accent.txt.

#

# See the bottom of this file for the Perl script used to generate the contents

# of this file (without this header) from ASCIIFoldingFilter.java.

# Syntax:

# "source" => "target"

# "source".length() > 0 (source cannot be empty.)

# "target".length() >= 0 (target can be empty.)

# À [LATIN CAPITAL LETTER A WITH GRAVE]

"\u00C0" => "A"

# Á [LATIN CAPITAL LETTER A WITH ACUTE]

"\u00C1" => "A"

# Â [LATIN CAPITAL LETTER A WITH CIRCUMFLEX]

"\u00C2" => "A"

# Ã [LATIN CAPITAL LETTER A WITH TILDE]

"\u00C3" => "A"

# Ä [LATIN CAPITAL LETTER A WITH DIAERESIS]

"\u00C4" => "A"

# Å [LATIN CAPITAL LETTER A WITH RING ABOVE]

"\u00C5" => "A"

# Ā [LATIN CAPITAL LETTER A WITH MACRON]

"\u0100" => "A"

# Ă [LATIN CAPITAL LETTER A WITH BREVE]

"\u0102" => "A"

# Ą [LATIN CAPITAL LETTER A WITH OGONEK]

"\u0104" => "A"

# Ə http://en.wikipedia.org/wiki/Schwa [LATIN CAPITAL LETTER SCHWA]

"\u018F" => "A"

# Ǎ [LATIN CAPITAL LETTER A WITH CARON]

"\u01CD" => "A"

# Ǟ [LATIN CAPITAL LETTER A WITH DIAERESIS AND MACRON]

"\u01DE" => "A"

# Ǡ [LATIN CAPITAL LETTER A WITH DOT ABOVE AND MACRON]

"\u01E0" => "A"

# Ǻ [LATIN CAPITAL LETTER A WITH RING ABOVE AND ACUTE]

"\u01FA" => "A"

# Ȁ [LATIN CAPITAL LETTER A WITH DOUBLE GRAVE]

"\u0200" => "A"

# Ȃ [LATIN CAPITAL LETTER A WITH INVERTED BREVE]

"\u0202" => "A"

# Ȧ [LATIN CAPITAL LETTER A WITH DOT ABOVE]

"\u0226" => "A"

# Ⱥ [LATIN CAPITAL LETTER A WITH STROKE]

"\u023A" => "A"

# ᴀ [LATIN LETTER SMALL CAPITAL A]

"\u1D00" => "A"

# Ḁ [LATIN CAPITAL LETTER A WITH RING BELOW]

"\u1E00" => "A"

# Ạ [LATIN CAPITAL LETTER A WITH DOT BELOW]

"\u1EA0" => "A"

# Ả [LATIN CAPITAL LETTER A WITH HOOK ABOVE]

"\u1EA2" => "A"

# Ấ [LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND ACUTE]

"\u1EA4" => "A"

# Ầ [LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND GRAVE]

"\u1EA6" => "A"

# Ẩ [LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND HOOK ABOVE]

"\u1EA8" => "A"

# Ẫ [LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND TILDE]

"\u1EAA" => "A"

# Ậ [LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND DOT BELOW]

"\u1EAC" => "A"

# Ắ [LATIN CAPITAL LETTER A WITH BREVE AND ACUTE]

"\u1EAE" => "A"

# Ằ [LATIN CAPITAL LETTER A WITH BREVE AND GRAVE]

"\u1EB0" => "A"

# Ẳ [LATIN CAPITAL LETTER A WITH BREVE AND HOOK ABOVE]

"\u1EB2" => "A"

# Ẵ [LATIN CAPITAL LETTER A WITH BREVE AND TILDE]

"\u1EB4" => "A"

# Ặ [LATIN CAPITAL LETTER A WITH BREVE AND DOT BELOW]

"\u1EB6" => "A"

# Ⓐ [CIRCLED LATIN CAPITAL LETTER A]

"\u24B6" => "A"

# Ａ [FULLWIDTH LATIN CAPITAL LETTER A]

"\uFF21" => "A"

# à [LATIN SMALL LETTER A WITH GRAVE]

"\u00E0" => "a"

# á [LATIN SMALL LETTER A WITH ACUTE]

"\u00E1" => "a"

# â [LATIN SMALL LETTER A WITH CIRCUMFLEX]

"\u00E2" => "a"

# ã [LATIN SMALL LETTER A WITH TILDE]

"\u00E3" => "a"

# ä [LATIN SMALL LETTER A WITH DIAERESIS]

"\u00E4" => "a"

# å [LATIN SMALL LETTER A WITH RING ABOVE]

"\u00E5" => "a"

# ā [LATIN SMALL LETTER A WITH MACRON]

"\u0101" => "a"

# ă [LATIN SMALL LETTER A WITH BREVE]

"\u0103" => "a"

# ą [LATIN SMALL LETTER A WITH OGONEK]

"\u0105" => "a"

# ǎ [LATIN SMALL LETTER A WITH CARON]

"\u01CE" => "a"

# ǟ [LATIN SMALL LETTER A WITH DIAERESIS AND MACRON]

"\u01DF" => "a"

# ǡ [LATIN SMALL LETTER A WITH DOT ABOVE AND MACRON]

"\u01E1" => "a"

# ǻ [LATIN SMALL LETTER A WITH RING ABOVE AND ACUTE]

"\u01FB" => "a"

# ȁ [LATIN SMALL LETTER A WITH DOUBLE GRAVE]

"\u0201" => "a"

# ȃ [LATIN SMALL LETTER A WITH INVERTED BREVE]

"\u0203" => "a"

# ȧ [LATIN SMALL LETTER A WITH DOT ABOVE]

"\u0227" => "a"

# ɐ [LATIN SMALL LETTER TURNED A]

"\u0250" => "a"

# ə [LATIN SMALL LETTER SCHWA]

"\u0259" => "a"

# ɚ [LATIN SMALL LETTER SCHWA WITH HOOK]

"\u025A" => "a"

# ᶏ [LATIN SMALL LETTER A WITH RETROFLEX HOOK]

"\u1D8F" => "a"

# ᶕ [LATIN SMALL LETTER SCHWA WITH RETROFLEX HOOK]

"\u1D95" => "a"

# ạ [LATIN SMALL LETTER A WITH RING BELOW]

"\u1E01" => "a"

# ả [LATIN SMALL LETTER A WITH RIGHT HALF RING]

"\u1E9A" => "a"

# ạ [LATIN SMALL LETTER A WITH DOT BELOW]

"\u1EA1" => "a"

# ả [LATIN SMALL LETTER A WITH HOOK ABOVE]

"\u1EA3" => "a"

# ấ [LATIN SMALL LETTER A WITH CIRCUMFLEX AND ACUTE]

"\u1EA5" => "a"

# ầ [LATIN SMALL LETTER A WITH CIRCUMFLEX AND GRAVE]

"\u1EA7" => "a"

# ẩ [LATIN SMALL LETTER A WITH CIRCUMFLEX AND HOOK ABOVE]

"\u1EA9" => "a"

# ẫ [LATIN SMALL LETTER A WITH CIRCUMFLEX AND TILDE]

"\u1EAB" => "a"

# ậ [LATIN SMALL LETTER A WITH CIRCUMFLEX AND DOT BELOW]

"\u1EAD" => "a"

# ắ [LATIN SMALL LETTER A WITH BREVE AND ACUTE]

"\u1EAF" => "a"

# ằ [LATIN SMALL LETTER A WITH BREVE AND GRAVE]

"\u1EB1" => "a"

# ẳ [LATIN SMALL LETTER A WITH BREVE AND HOOK ABOVE]

"\u1EB3" => "a"

# ẵ [LATIN SMALL LETTER A WITH BREVE AND TILDE]

"\u1EB5" => "a"

# ặ [LATIN SMALL LETTER A WITH BREVE AND DOT BELOW]

"\u1EB7" => "a"

# ₐ [LATIN SUBSCRIPT SMALL LETTER A]

"\u2090" => "a"

# ₔ [LATIN SUBSCRIPT SMALL LETTER SCHWA]

"\u2094" => "a"

# ⓐ [CIRCLED LATIN SMALL LETTER A]

"\u24D0" => "a"

# ⱥ [LATIN SMALL LETTER A WITH STROKE]

"\u2C65" => "a"

# Ɐ [LATIN CAPITAL LETTER TURNED A]

"\u2C6F" => "a"

# ａ [FULLWIDTH LATIN SMALL LETTER A]

"\uFF41" => "a"

# Ꜳ [LATIN CAPITAL LETTER AA]

"\uA732" => "AA"

# Æ [LATIN CAPITAL LETTER AE]

"\u00C6" => "AE"

# Ǣ [LATIN CAPITAL LETTER AE WITH MACRON]

"\u01E2" => "AE"

# Ǽ [LATIN CAPITAL LETTER AE WITH ACUTE]

"\u01FC" => "AE"

# ᴁ [LATIN LETTER SMALL CAPITAL AE]

"\u1D01" => "AE"

# Ꜵ [LATIN CAPITAL LETTER AO]

"\uA734" => "AO"

# Ꜷ [LATIN CAPITAL LETTER AU]

"\uA736" => "AU"

# Ꜹ [LATIN CAPITAL LETTER AV]

"\uA738" => "AV"

# Ꜻ [LATIN CAPITAL LETTER AV WITH HORIZONTAL BAR]

"\uA73A" => "AV"

# Ꜽ [LATIN CAPITAL LETTER AY]

"\uA73C" => "AY"

# ⒜ [PARENTHESIZED LATIN SMALL LETTER A]

"\u249C" => "(a)"

# ꜳ [LATIN SMALL LETTER AA]

"\uA733" => "aa"

# æ [LATIN SMALL LETTER AE]

"\u00E6" => "ae"

# ǣ [LATIN SMALL LETTER AE WITH MACRON]

"\u01E3" => "ae"

# ǽ [LATIN SMALL LETTER AE WITH ACUTE]

"\u01FD" => "ae"

# ᴂ [LATIN SMALL LETTER TURNED AE]

"\u1D02" => "ae"

# ꜵ [LATIN SMALL LETTER AO]

"\uA735" => "ao"

# ꜷ [LATIN SMALL LETTER AU]

"\uA737" => "au"

# ꜹ [LATIN SMALL LETTER AV]

"\uA739" => "av"

# ꜻ [LATIN SMALL LETTER AV WITH HORIZONTAL BAR]

"\uA73B" => "av"

# ꜽ [LATIN SMALL LETTER AY]

"\uA73D" => "ay"

# Ɓ [LATIN CAPITAL LETTER B WITH HOOK]

"\u0181" => "B"

# Ƃ [LATIN CAPITAL LETTER B WITH TOPBAR]

"\u0182" => "B"

# Ƀ [LATIN CAPITAL LETTER B WITH STROKE]

"\u0243" => "B"

# ʙ [LATIN LETTER SMALL CAPITAL B]

"\u0299" => "B"

# ᴃ [LATIN LETTER SMALL CAPITAL BARRED B]

"\u1D03" => "B"

# Ḃ [LATIN CAPITAL LETTER B WITH DOT ABOVE]

"\u1E02" => "B"

# Ḅ [LATIN CAPITAL LETTER B WITH DOT BELOW]

"\u1E04" => "B"

# Ḇ [LATIN CAPITAL LETTER B WITH LINE BELOW]

"\u1E06" => "B"

# Ⓑ [CIRCLED LATIN CAPITAL LETTER B]

"\u24B7" => "B"

# Ｂ [FULLWIDTH LATIN CAPITAL LETTER B]

"\uFF22" => "B"

# ƀ [LATIN SMALL LETTER B WITH STROKE]

"\u0180" => "b"

# ƃ [LATIN SMALL LETTER B WITH TOPBAR]

"\u0183" => "b"

# ɓ [LATIN SMALL LETTER B WITH HOOK]

"\u0253" => "b"

# ᵬ [LATIN SMALL LETTER B WITH MIDDLE TILDE]

"\u1D6C" => "b"

# ᶀ [LATIN SMALL LETTER B WITH PALATAL HOOK]

"\u1D80" => "b"

# ḃ [LATIN SMALL LETTER B WITH DOT ABOVE]

"\u1E03" => "b"

# ḅ [LATIN SMALL LETTER B WITH DOT BELOW]

"\u1E05" => "b"

# ḇ [LATIN SMALL LETTER B WITH LINE BELOW]

"\u1E07" => "b"

# ⓑ [CIRCLED LATIN SMALL LETTER B]

"\u24D1" => "b"

# ｂ [FULLWIDTH LATIN SMALL LETTER B]

"\uFF42" => "b"

# ⒝ [PARENTHESIZED LATIN SMALL LETTER B]

"\u249D" => "(b)"

# Ç [LATIN CAPITAL LETTER C WITH CEDILLA]

"\u00C7" => "C"

# Ć [LATIN CAPITAL LETTER C WITH ACUTE]

"\u0106" => "C"

# Ĉ [LATIN CAPITAL LETTER C WITH CIRCUMFLEX]

"\u0108" => "C"

# Ċ [LATIN CAPITAL LETTER C WITH DOT ABOVE]

"\u010A" => "C"

# Č [LATIN CAPITAL LETTER C WITH CARON]

"\u010C" => "C"

# Ƈ [LATIN CAPITAL LETTER C WITH HOOK]

"\u0187" => "C"

# Ȼ [LATIN CAPITAL LETTER C WITH STROKE]

"\u023B" => "C"

# ʗ [LATIN LETTER STRETCHED C]

"\u0297" => "C"

# ᴄ [LATIN LETTER SMALL CAPITAL C]

"\u1D04" => "C"

# Ḉ [LATIN CAPITAL LETTER C WITH CEDILLA AND ACUTE]

"\u1E08" => "C"

# Ⓒ [CIRCLED LATIN CAPITAL LETTER C]

"\u24B8" => "C"

# Ｃ [FULLWIDTH LATIN CAPITAL LETTER C]

"\uFF23" => "C"

# ç [LATIN SMALL LETTER C WITH CEDILLA]

"\u00E7" => "c"

# ć [LATIN SMALL LETTER C WITH ACUTE]

"\u0107" => "c"

# ĉ [LATIN SMALL LETTER C WITH CIRCUMFLEX]

"\u0109" => "c"

# ċ [LATIN SMALL LETTER C WITH DOT ABOVE]

"\u010B" => "c"

# č [LATIN SMALL LETTER C WITH CARON]

"\u010D" => "c"

# ƈ [LATIN SMALL LETTER C WITH HOOK]

"\u0188" => "c"

# ȼ [LATIN SMALL LETTER C WITH STROKE]

"\u023C" => "c"

# ɕ [LATIN SMALL LETTER C WITH CURL]

"\u0255" => "c"

# ḉ [LATIN SMALL LETTER C WITH CEDILLA AND ACUTE]

"\u1E09" => "c"

# ↄ [LATIN SMALL LETTER REVERSED C]

"\u2184" => "c"

# ⓒ [CIRCLED LATIN SMALL LETTER C]

"\u24D2" => "c"

# Ꜿ [LATIN CAPITAL LETTER REVERSED C WITH DOT]

"\uA73E" => "c"

# ꜿ [LATIN SMALL LETTER REVERSED C WITH DOT]

"\uA73F" => "c"

# ｃ [FULLWIDTH LATIN SMALL LETTER C]

"\uFF43" => "c"

# ⒞ [PARENTHESIZED LATIN SMALL LETTER C]

"\u249E" => "(c)"

# Ð [LATIN CAPITAL LETTER ETH]

"\u00D0" => "D"

# Ď [LATIN CAPITAL LETTER D WITH CARON]

"\u010E" => "D"

# Đ [LATIN CAPITAL LETTER D WITH STROKE]

"\u0110" => "D"

# Ɖ [LATIN CAPITAL LETTER AFRICAN D]

"\u0189" => "D"

# Ɗ [LATIN CAPITAL LETTER D WITH HOOK]

"\u018A" => "D"

# Ƌ [LATIN CAPITAL LETTER D WITH TOPBAR]

"\u018B" => "D"

# ᴅ [LATIN LETTER SMALL CAPITAL D]

"\u1D05" => "D"

# ᴆ [LATIN LETTER SMALL CAPITAL ETH]

"\u1D06" => "D"

# Ḋ [LATIN CAPITAL LETTER D WITH DOT ABOVE]

"\u1E0A" => "D"

# Ḍ [LATIN CAPITAL LETTER D WITH DOT BELOW]

"\u1E0C" => "D"

# Ḏ [LATIN CAPITAL LETTER D WITH LINE BELOW]

"\u1E0E" => "D"

# Ḑ [LATIN CAPITAL LETTER D WITH CEDILLA]

"\u1E10" => "D"

# Ḓ [LATIN CAPITAL LETTER D WITH CIRCUMFLEX BELOW]

"\u1E12" => "D"

# Ⓓ [CIRCLED LATIN CAPITAL LETTER D]

"\u24B9" => "D"

# Ꝺ [LATIN CAPITAL LETTER INSULAR D]

"\uA779" => "D"

# Ｄ [FULLWIDTH LATIN CAPITAL LETTER D]

"\uFF24" => "D"

# ð [LATIN SMALL LETTER ETH]

"\u00F0" => "d"

# ď [LATIN SMALL LETTER D WITH CARON]

"\u010F" => "d"

# đ [LATIN SMALL LETTER D WITH STROKE]

"\u0111" => "d"

# ƌ [LATIN SMALL LETTER D WITH TOPBAR]

"\u018C" => "d"

# ȡ [LATIN SMALL LETTER D WITH CURL]

"\u0221" => "d"

# ɖ [LATIN SMALL LETTER D WITH TAIL]

"\u0256" => "d"

# ɗ [LATIN SMALL LETTER D WITH HOOK]

"\u0257" => "d"

# ᵭ [LATIN SMALL LETTER D WITH MIDDLE TILDE]

"\u1D6D" => "d"

# ᶁ [LATIN SMALL LETTER D WITH PALATAL HOOK]

"\u1D81" => "d"

# ᶑ [LATIN SMALL LETTER D WITH HOOK AND TAIL]

"\u1D91" => "d"

# ḋ [LATIN SMALL LETTER D WITH DOT ABOVE]

"\u1E0B" => "d"

# ḍ [LATIN SMALL LETTER D WITH DOT BELOW]

"\u1E0D" => "d"

# ḏ [LATIN SMALL LETTER D WITH LINE BELOW]

"\u1E0F" => "d"

# ḑ [LATIN SMALL LETTER D WITH CEDILLA]

"\u1E11" => "d"

# ḓ [LATIN SMALL LETTER D WITH CIRCUMFLEX BELOW]

"\u1E13" => "d"

# ⓓ [CIRCLED LATIN SMALL LETTER D]

"\u24D3" => "d"

# ꝺ [LATIN SMALL LETTER INSULAR D]

"\uA77A" => "d"

# ｄ [FULLWIDTH LATIN SMALL LETTER D]

"\uFF44" => "d"

# Ǆ [LATIN CAPITAL LETTER DZ WITH CARON]

"\u01C4" => "DZ"

# Ǳ [LATIN CAPITAL LETTER DZ]

"\u01F1" => "DZ"

# ǅ [LATIN CAPITAL LETTER D WITH SMALL LETTER Z WITH CARON]

"\u01C5" => "Dz"

# ǲ [LATIN CAPITAL LETTER D WITH SMALL LETTER Z]

"\u01F2" => "Dz"

# ⒟ [PARENTHESIZED LATIN SMALL LETTER D]

"\u249F" => "(d)"

# ȸ [LATIN SMALL LETTER DB DIGRAPH]

"\u0238" => "db"

# ǆ [LATIN SMALL LETTER DZ WITH CARON]

"\u01C6" => "dz"

# ǳ [LATIN SMALL LETTER DZ]

"\u01F3" => "dz"

# ʣ [LATIN SMALL LETTER DZ DIGRAPH]

"\u02A3" => "dz"

# ʥ [LATIN SMALL LETTER DZ DIGRAPH WITH CURL]

"\u02A5" => "dz"

# È [LATIN CAPITAL LETTER E WITH GRAVE]

"\u00C8" => "E"

# É [LATIN CAPITAL LETTER E WITH ACUTE]

"\u00C9" => "E"

# Ê [LATIN CAPITAL LETTER E WITH CIRCUMFLEX]

"\u00CA" => "E"

# Ë [LATIN CAPITAL LETTER E WITH DIAERESIS]

"\u00CB" => "E"

# Ē [LATIN CAPITAL LETTER E WITH MACRON]

"\u0112" => "E"

# Ĕ [LATIN CAPITAL LETTER E WITH BREVE]

"\u0114" => "E"

# Ė [LATIN CAPITAL LETTER E WITH DOT ABOVE]

"\u0116" => "E"

# Ę [LATIN CAPITAL LETTER E WITH OGONEK]

"\u0118" => "E"

# Ě [LATIN CAPITAL LETTER E WITH CARON]

"\u011A" => "E"

# Ǝ [LATIN CAPITAL LETTER REVERSED E]

"\u018E" => "E"

# Ɛ [LATIN CAPITAL LETTER OPEN E]

"\u0190" => "E"

# Ȅ [LATIN CAPITAL LETTER E WITH DOUBLE GRAVE]

"\u0204" => "E"

# Ȇ [LATIN CAPITAL LETTER E WITH INVERTED BREVE]

"\u0206" => "E"

# Ȩ [LATIN CAPITAL LETTER E WITH CEDILLA]

"\u0228" => "E"

# Ɇ [LATIN CAPITAL LETTER E WITH STROKE]

"\u0246" => "E"

# ᴇ [LATIN LETTER SMALL CAPITAL E]

"\u1D07" => "E"

# Ḕ [LATIN CAPITAL LETTER E WITH MACRON AND GRAVE]

"\u1E14" => "E"

# Ḗ [LATIN CAPITAL LETTER E WITH MACRON AND ACUTE]

"\u1E16" => "E"

# Ḙ [LATIN CAPITAL LETTER E WITH CIRCUMFLEX BELOW]

"\u1E18" => "E"

# Ḛ [LATIN CAPITAL LETTER E WITH TILDE BELOW]

"\u1E1A" => "E"

# Ḝ [LATIN CAPITAL LETTER E WITH CEDILLA AND BREVE]

"\u1E1C" => "E"

# Ẹ [LATIN CAPITAL LETTER E WITH DOT BELOW]

"\u1EB8" => "E"

# Ẻ [LATIN CAPITAL LETTER E WITH HOOK ABOVE]

"\u1EBA" => "E"

# Ẽ [LATIN CAPITAL LETTER E WITH TILDE]

"\u1EBC" => "E"

# Ế [LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND ACUTE]

"\u1EBE" => "E"

# Ề [LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND GRAVE]

"\u1EC0" => "E"

# Ể [LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND HOOK ABOVE]

"\u1EC2" => "E"

# Ễ [LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND TILDE]

"\u1EC4" => "E"

# Ệ [LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND DOT BELOW]

"\u1EC6" => "E"

# Ⓔ [CIRCLED LATIN CAPITAL LETTER E]

"\u24BA" => "E"

# ⱻ [LATIN LETTER SMALL CAPITAL TURNED E]

"\u2C7B" => "E"

# Ｅ [FULLWIDTH LATIN CAPITAL LETTER E]

"\uFF25" => "E"

# è [LATIN SMALL LETTER E WITH GRAVE]

"\u00E8" => "e"

# é [LATIN SMALL LETTER E WITH ACUTE]

"\u00E9" => "e"

# ê [LATIN SMALL LETTER E WITH CIRCUMFLEX]

"\u00EA" => "e"

# ë [LATIN SMALL LETTER E WITH DIAERESIS]

"\u00EB" => "e"

# ē [LATIN SMALL LETTER E WITH MACRON]

"\u0113" => "e"

# ĕ [LATIN SMALL LETTER E WITH BREVE]

"\u0115" => "e"

# ė [LATIN SMALL LETTER E WITH DOT ABOVE]

"\u0117" => "e"

# ę [LATIN SMALL LETTER E WITH OGONEK]

"\u0119" => "e"

# ě [LATIN SMALL LETTER E WITH CARON]

"\u011B" => "e"

# ǝ [LATIN SMALL LETTER TURNED E]

"\u01DD" => "e"

# ȅ [LATIN SMALL LETTER E WITH DOUBLE GRAVE]

"\u0205" => "e"

# ȇ [LATIN SMALL LETTER E WITH INVERTED BREVE]

"\u0207" => "e"

# ȩ [LATIN SMALL LETTER E WITH CEDILLA]

"\u0229" => "e"

# ɇ [LATIN SMALL LETTER E WITH STROKE]

"\u0247" => "e"

# ɘ [LATIN SMALL LETTER REVERSED E]

"\u0258" => "e"

# ɛ [LATIN SMALL LETTER OPEN E]

"\u025B" => "e"

# ɜ [LATIN SMALL LETTER REVERSED OPEN E]

"\u025C" => "e"

# ɝ [LATIN SMALL LETTER REVERSED OPEN E WITH HOOK]

"\u025D" => "e"

# ɞ [LATIN SMALL LETTER CLOSED REVERSED OPEN E]

"\u025E" => "e"

# ʚ [LATIN SMALL LETTER CLOSED OPEN E]

"\u029A" => "e"

# ᴈ [LATIN SMALL LETTER TURNED OPEN E]

"\u1D08" => "e"

# ᶒ [LATIN SMALL LETTER E WITH RETROFLEX HOOK]

"\u1D92" => "e"

# ᶓ [LATIN SMALL LETTER OPEN E WITH RETROFLEX HOOK]

"\u1D93" => "e"

# ᶔ [LATIN SMALL LETTER REVERSED OPEN E WITH RETROFLEX HOOK]

"\u1D94" => "e"

# ḕ [LATIN SMALL LETTER E WITH MACRON AND GRAVE]

"\u1E15" => "e"

# ḗ [LATIN SMALL LETTER E WITH MACRON AND ACUTE]

"\u1E17" => "e"

# ḙ [LATIN SMALL LETTER E WITH CIRCUMFLEX BELOW]

"\u1E19" => "e"

# ḛ [LATIN SMALL LETTER E WITH TILDE BELOW]

"\u1E1B" => "e"

# ḝ [LATIN SMALL LETTER E WITH CEDILLA AND BREVE]

"\u1E1D" => "e"

# ẹ [LATIN SMALL LETTER E WITH DOT BELOW]

"\u1EB9" => "e"

# ẻ [LATIN SMALL LETTER E WITH HOOK ABOVE]

"\u1EBB" => "e"

# ẽ [LATIN SMALL LETTER E WITH TILDE]

"\u1EBD" => "e"

# ế [LATIN SMALL LETTER E WITH CIRCUMFLEX AND ACUTE]

"\u1EBF" => "e"

# ề [LATIN SMALL LETTER E WITH CIRCUMFLEX AND GRAVE]

"\u1EC1" => "e"

# ể [LATIN SMALL LETTER E WITH CIRCUMFLEX AND HOOK ABOVE]

"\u1EC3" => "e"

# ễ [LATIN SMALL LETTER E WITH CIRCUMFLEX AND TILDE]

"\u1EC5" => "e"

# ệ [LATIN SMALL LETTER E WITH CIRCUMFLEX AND DOT BELOW]

"\u1EC7" => "e"

# ₑ [LATIN SUBSCRIPT SMALL LETTER E]

"\u2091" => "e"

# ⓔ [CIRCLED LATIN SMALL LETTER E]

"\u24D4" => "e"

# ⱸ [LATIN SMALL LETTER E WITH NOTCH]

"\u2C78" => "e"

# ｅ [FULLWIDTH LATIN SMALL LETTER E]

"\uFF45" => "e"

# ⒠ [PARENTHESIZED LATIN SMALL LETTER E]

"\u24A0" => "(e)"

# Ƒ [LATIN CAPITAL LETTER F WITH HOOK]

"\u0191" => "F"

# Ḟ [LATIN CAPITAL LETTER F WITH DOT ABOVE]

"\u1E1E" => "F"

# Ⓕ [CIRCLED LATIN CAPITAL LETTER F]

"\u24BB" => "F"

# ꜰ [LATIN LETTER SMALL CAPITAL F]

"\uA730" => "F"

# Ꝼ [LATIN CAPITAL LETTER INSULAR F]

"\uA77B" => "F"

# ꟻ [LATIN EPIGRAPHIC LETTER REVERSED F]

"\uA7FB" => "F"

# Ｆ [FULLWIDTH LATIN CAPITAL LETTER F]

"\uFF26" => "F"

# ƒ [LATIN SMALL LETTER F WITH HOOK]

"\u0192" => "f"

# ᵮ [LATIN SMALL LETTER F WITH MIDDLE TILDE]

"\u1D6E" => "f"

# ᶂ [LATIN SMALL LETTER F WITH PALATAL HOOK]

"\u1D82" => "f"

# ḟ [LATIN SMALL LETTER F WITH DOT ABOVE]

"\u1E1F" => "f"

# ẛ [LATIN SMALL LETTER LONG S WITH DOT ABOVE]

"\u1E9B" => "f"

# ⓕ [CIRCLED LATIN SMALL LETTER F]

"\u24D5" => "f"

# ꝼ [LATIN SMALL LETTER INSULAR F]

"\uA77C" => "f"

# ｆ [FULLWIDTH LATIN SMALL LETTER F]

"\uFF46" => "f"

# ⒡ [PARENTHESIZED LATIN SMALL LETTER F]

"\u24A1" => "(f)"

# ﬀ [LATIN SMALL LIGATURE FF]

"\uFB00" => "ff"

# ﬃ [LATIN SMALL LIGATURE FFI]

"\uFB03" => "ffi"

# ﬄ [LATIN SMALL LIGATURE FFL]

"\uFB04" => "ffl"

# ﬁ [LATIN SMALL LIGATURE FI]

"\uFB01" => "fi"

# ﬂ [LATIN SMALL LIGATURE FL]

"\uFB02" => "fl"

# Ĝ [LATIN CAPITAL LETTER G WITH CIRCUMFLEX]

"\u011C" => "G"

# Ğ [LATIN CAPITAL LETTER G WITH BREVE]

"\u011E" => "G"

# Ġ [LATIN CAPITAL LETTER G WITH DOT ABOVE]

"\u0120" => "G"

# Ģ [LATIN CAPITAL LETTER G WITH CEDILLA]

"\u0122" => "G"

# Ɠ [LATIN CAPITAL LETTER G WITH HOOK]

"\u0193" => "G"

# Ǥ [LATIN CAPITAL LETTER G WITH STROKE]

"\u01E4" => "G"

# ǥ [LATIN SMALL LETTER G WITH STROKE]

"\u01E5" => "G"

# Ǧ [LATIN CAPITAL LETTER G WITH CARON]

"\u01E6" => "G"

# ǧ [LATIN SMALL LETTER G WITH CARON]

"\u01E7" => "G"

# Ǵ [LATIN CAPITAL LETTER G WITH ACUTE]

"\u01F4" => "G"

# ɢ [LATIN LETTER SMALL CAPITAL G]

"\u0262" => "G"

# ʛ [LATIN LETTER SMALL CAPITAL G WITH HOOK]

"\u029B" => "G"

# Ḡ [LATIN CAPITAL LETTER G WITH MACRON]

"\u1E20" => "G"

# Ⓖ [CIRCLED LATIN CAPITAL LETTER G]

"\u24BC" => "G"

# Ᵹ [LATIN CAPITAL LETTER INSULAR G]

"\uA77D" => "G"

# Ꝿ [LATIN CAPITAL LETTER TURNED INSULAR G]

"\uA77E" => "G"

# Ｇ [FULLWIDTH LATIN CAPITAL LETTER G]

"\uFF27" => "G"

# ĝ [LATIN SMALL LETTER G WITH CIRCUMFLEX]

"\u011D" => "g"

# ğ [LATIN SMALL LETTER G WITH BREVE]

"\u011F" => "g"

# ġ [LATIN SMALL LETTER G WITH DOT ABOVE]

"\u0121" => "g"

# ģ [LATIN SMALL LETTER G WITH CEDILLA]

"\u0123" => "g"

# ǵ [LATIN SMALL LETTER G WITH ACUTE]

"\u01F5" => "g"

# ɠ [LATIN SMALL LETTER G WITH HOOK]

"\u0260" => "g"

# ɡ [LATIN SMALL LETTER SCRIPT G]

"\u0261" => "g"

# ᵷ [LATIN SMALL LETTER TURNED G]

"\u1D77" => "g"

# ᵹ [LATIN SMALL LETTER INSULAR G]

"\u1D79" => "g"

# ᶃ [LATIN SMALL LETTER G WITH PALATAL HOOK]

"\u1D83" => "g"

# ḡ [LATIN SMALL LETTER G WITH MACRON]

"\u1E21" => "g"

# ⓖ [CIRCLED LATIN SMALL LETTER G]

"\u24D6" => "g"

# ꝿ [LATIN SMALL LETTER TURNED INSULAR G]

"\uA77F" => "g"

# ｇ [FULLWIDTH LATIN SMALL LETTER G]

"\uFF47" => "g"

# ⒢ [PARENTHESIZED LATIN SMALL LETTER G]

"\u24A2" => "(g)"

# Ĥ [LATIN CAPITAL LETTER H WITH CIRCUMFLEX]

"\u0124" => "H"

# Ħ [LATIN CAPITAL LETTER H WITH STROKE]

"\u0126" => "H"

# Ȟ [LATIN CAPITAL LETTER H WITH CARON]

"\u021E" => "H"

# ʜ [LATIN LETTER SMALL CAPITAL H]

"\u029C" => "H"

# Ḣ [LATIN CAPITAL LETTER H WITH DOT ABOVE]

"\u1E22" => "H"

# Ḥ [LATIN CAPITAL LETTER H WITH DOT BELOW]

"\u1E24" => "H"

# Ḧ [LATIN CAPITAL LETTER H WITH DIAERESIS]

"\u1E26" => "H"

# Ḩ [LATIN CAPITAL LETTER H WITH CEDILLA]

"\u1E28" => "H"

# Ḫ [LATIN CAPITAL LETTER H WITH BREVE BELOW]

"\u1E2A" => "H"

# Ⓗ [CIRCLED LATIN CAPITAL LETTER H]

"\u24BD" => "H"

# Ⱨ [LATIN CAPITAL LETTER H WITH DESCENDER]

"\u2C67" => "H"

# Ⱶ [LATIN CAPITAL LETTER HALF H]

"\u2C75" => "H"

# Ｈ [FULLWIDTH LATIN CAPITAL LETTER H]

"\uFF28" => "H"

# ĥ [LATIN SMALL LETTER H WITH CIRCUMFLEX]

"\u0125" => "h"

# ħ [LATIN SMALL LETTER H WITH STROKE]

"\u0127" => "h"

# ȟ [LATIN SMALL LETTER H WITH CARON]

"\u021F" => "h"

# ɥ [LATIN SMALL LETTER TURNED H]

"\u0265" => "h"

# ɦ [LATIN SMALL LETTER H WITH HOOK]

"\u0266" => "h"

# ʮ [LATIN SMALL LETTER TURNED H WITH FISHHOOK]

"\u02AE" => "h"

# ʯ [LATIN SMALL LETTER TURNED H WITH FISHHOOK AND TAIL]

"\u02AF" => "h"

# ḣ [LATIN SMALL LETTER H WITH DOT ABOVE]

"\u1E23" => "h"

# ḥ [LATIN SMALL LETTER H WITH DOT BELOW]

"\u1E25" => "h"

# ḧ [LATIN SMALL LETTER H WITH DIAERESIS]

"\u1E27" => "h"

# ḩ [LATIN SMALL LETTER H WITH CEDILLA]

"\u1E29" => "h"

# ḫ [LATIN SMALL LETTER H WITH BREVE BELOW]

"\u1E2B" => "h"

# ẖ [LATIN SMALL LETTER H WITH LINE BELOW]

"\u1E96" => "h"

# ⓗ [CIRCLED LATIN SMALL LETTER H]

"\u24D7" => "h"

# ⱨ [LATIN SMALL LETTER H WITH DESCENDER]

"\u2C68" => "h"

# ⱶ [LATIN SMALL LETTER HALF H]

"\u2C76" => "h"

# ｈ [FULLWIDTH LATIN SMALL LETTER H]

"\uFF48" => "h"

# Ƕ http://en.wikipedia.org/wiki/Hwair [LATIN CAPITAL LETTER HWAIR]

"\u01F6" => "HV"

# ⒣ [PARENTHESIZED LATIN SMALL LETTER H]

"\u24A3" => "(h)"

# ƕ [LATIN SMALL LETTER HV]

"\u0195" => "hv"

# Ì [LATIN CAPITAL LETTER I WITH GRAVE]

"\u00CC" => "I"

# Í [LATIN CAPITAL LETTER I WITH ACUTE]

"\u00CD" => "I"

# Î [LATIN CAPITAL LETTER I WITH CIRCUMFLEX]

"\u00CE" => "I"

# Ï [LATIN CAPITAL LETTER I WITH DIAERESIS]

"\u00CF" => "I"

# Ĩ [LATIN CAPITAL LETTER I WITH TILDE]

"\u0128" => "I"

# Ī [LATIN CAPITAL LETTER I WITH MACRON]

"\u012A" => "I"

# Ĭ [LATIN CAPITAL LETTER I WITH BREVE]

"\u012C" => "I"

# Į [LATIN CAPITAL LETTER I WITH OGONEK]

"\u012E" => "I"

# İ [LATIN CAPITAL LETTER I WITH DOT ABOVE]

"\u0130" => "I"

# Ɩ [LATIN CAPITAL LETTER IOTA]

"\u0196" => "I"

# Ɨ [LATIN CAPITAL LETTER I WITH STROKE]

"\u0197" => "I"

# Ǐ [LATIN CAPITAL LETTER I WITH CARON]

"\u01CF" => "I"

# Ȉ [LATIN CAPITAL LETTER I WITH DOUBLE GRAVE]

"\u0208" => "I"

# Ȋ [LATIN CAPITAL LETTER I WITH INVERTED BREVE]

"\u020A" => "I"

# ɪ [LATIN LETTER SMALL CAPITAL I]

"\u026A" => "I"

# ᵻ [LATIN SMALL CAPITAL LETTER I WITH STROKE]

"\u1D7B" => "I"

# Ḭ [LATIN CAPITAL LETTER I WITH TILDE BELOW]

"\u1E2C" => "I"

# Ḯ [LATIN CAPITAL LETTER I WITH DIAERESIS AND ACUTE]

"\u1E2E" => "I"

# Ỉ [LATIN CAPITAL LETTER I WITH HOOK ABOVE]

"\u1EC8" => "I"

# Ị [LATIN CAPITAL LETTER I WITH DOT BELOW]

"\u1ECA" => "I"

# Ⓘ [CIRCLED LATIN CAPITAL LETTER I]

"\u24BE" => "I"

# ꟾ [LATIN EPIGRAPHIC LETTER I LONGA]

"\uA7FE" => "I"

# Ｉ [FULLWIDTH LATIN CAPITAL LETTER I]

"\uFF29" => "I"

# ì [LATIN SMALL LETTER I WITH GRAVE]

"\u00EC" => "i"

# í [LATIN SMALL LETTER I WITH ACUTE]

"\u00ED" => "i"

# î [LATIN SMALL LETTER I WITH CIRCUMFLEX]

"\u00EE" => "i"

# ï [LATIN SMALL LETTER I WITH DIAERESIS]

"\u00EF" => "i"

# ĩ [LATIN SMALL LETTER I WITH TILDE]

"\u0129" => "i"

# ī [LATIN SMALL LETTER I WITH MACRON]

"\u012B" => "i"

# ĭ [LATIN SMALL LETTER I WITH BREVE]

"\u012D" => "i"

# į [LATIN SMALL LETTER I WITH OGONEK]

"\u012F" => "i"

# ı [LATIN SMALL LETTER DOTLESS I]

"\u0131" => "i"

# ǐ [LATIN SMALL LETTER I WITH CARON]

"\u01D0" => "i"

# ȉ [LATIN SMALL LETTER I WITH DOUBLE GRAVE]

"\u0209" => "i"

# ȋ [LATIN SMALL LETTER I WITH INVERTED BREVE]

"\u020B" => "i"

# ɨ [LATIN SMALL LETTER I WITH STROKE]

"\u0268" => "i"

# ᴉ [LATIN SMALL LETTER TURNED I]

"\u1D09" => "i"

# ᵢ [LATIN SUBSCRIPT SMALL LETTER I]

"\u1D62" => "i"

# ᵼ [LATIN SMALL LETTER IOTA WITH STROKE]

"\u1D7C" => "i"

# ᶖ [LATIN SMALL LETTER I WITH RETROFLEX HOOK]

"\u1D96" => "i"

# ḭ [LATIN SMALL LETTER I WITH TILDE BELOW]

"\u1E2D" => "i"

# ḯ [LATIN SMALL LETTER I WITH DIAERESIS AND ACUTE]

"\u1E2F" => "i"

# ỉ [LATIN SMALL LETTER I WITH HOOK ABOVE]

"\u1EC9" => "i"

# ị [LATIN SMALL LETTER I WITH DOT BELOW]

"\u1ECB" => "i"

# ⁱ [SUPERSCRIPT LATIN SMALL LETTER I]

"\u2071" => "i"

# ⓘ [CIRCLED LATIN SMALL LETTER I]

"\u24D8" => "i"

# ｉ [FULLWIDTH LATIN SMALL LETTER I]

"\uFF49" => "i"

# Ĳ [LATIN CAPITAL LIGATURE IJ]

"\u0132" => "IJ"

# ⒤ [PARENTHESIZED LATIN SMALL LETTER I]

"\u24A4" => "(i)"

# ĳ [LATIN SMALL LIGATURE IJ]

"\u0133" => "ij"

# Ĵ [LATIN CAPITAL LETTER J WITH CIRCUMFLEX]

"\u0134" => "J"

# Ɉ [LATIN CAPITAL LETTER J WITH STROKE]

"\u0248" => "J"

# ᴊ [LATIN LETTER SMALL CAPITAL J]

"\u1D0A" => "J"

# Ⓙ [CIRCLED LATIN CAPITAL LETTER J]

"\u24BF" => "J"

# Ｊ [FULLWIDTH LATIN CAPITAL LETTER J]

"\uFF2A" => "J"

# ĵ [LATIN SMALL LETTER J WITH CIRCUMFLEX]

"\u0135" => "j"

# ǰ [LATIN SMALL LETTER J WITH CARON]

"\u01F0" => "j"

# ȷ [LATIN SMALL LETTER DOTLESS J]

"\u0237" => "j"

# ɉ [LATIN SMALL LETTER J WITH STROKE]

"\u0249" => "j"

# ɟ [LATIN SMALL LETTER DOTLESS J WITH STROKE]

"\u025F" => "j"

# ʄ [LATIN SMALL LETTER DOTLESS J WITH STROKE AND HOOK]

"\u0284" => "j"

# ʝ [LATIN SMALL LETTER J WITH CROSSED-TAIL]

"\u029D" => "j"

# ⓙ [CIRCLED LATIN SMALL LETTER J]

"\u24D9" => "j"

# ⱼ [LATIN SUBSCRIPT SMALL LETTER J]

"\u2C7C" => "j"

# ｊ [FULLWIDTH LATIN SMALL LETTER J]

"\uFF4A" => "j"

# ⒥ [PARENTHESIZED LATIN SMALL LETTER J]

"\u24A5" => "(j)"

# Ķ [LATIN CAPITAL LETTER K WITH CEDILLA]

"\u0136" => "K"

# Ƙ [LATIN CAPITAL LETTER K WITH HOOK]

"\u0198" => "K"

# Ǩ [LATIN CAPITAL LETTER K WITH CARON]

"\u01E8" => "K"

# ᴋ [LATIN LETTER SMALL CAPITAL K]

"\u1D0B" => "K"

# Ḱ [LATIN CAPITAL LETTER K WITH ACUTE]

"\u1E30" => "K"

# Ḳ [LATIN CAPITAL LETTER K WITH DOT BELOW]

"\u1E32" => "K"

# Ḵ [LATIN CAPITAL LETTER K WITH LINE BELOW]

"\u1E34" => "K"

# Ⓚ [CIRCLED LATIN CAPITAL LETTER K]

"\u24C0" => "K"

# Ⱪ [LATIN CAPITAL LETTER K WITH DESCENDER]

"\u2C69" => "K"

# Ꝁ [LATIN CAPITAL LETTER K WITH STROKE]

"\uA740" => "K"

# Ꝃ [LATIN CAPITAL LETTER K WITH DIAGONAL STROKE]

"\uA742" => "K"

# Ꝅ [LATIN CAPITAL LETTER K WITH STROKE AND DIAGONAL STROKE]

"\uA744" => "K"

# Ｋ [FULLWIDTH LATIN CAPITAL LETTER K]

"\uFF2B" => "K"

# ķ [LATIN SMALL LETTER K WITH CEDILLA]

"\u0137" => "k"

# ƙ [LATIN SMALL LETTER K WITH HOOK]

"\u0199" => "k"

# ǩ [LATIN SMALL LETTER K WITH CARON]

"\u01E9" => "k"

# ʞ [LATIN SMALL LETTER TURNED K]

"\u029E" => "k"

# ᶄ [LATIN SMALL LETTER K WITH PALATAL HOOK]

"\u1D84" => "k"

# ḱ [LATIN SMALL LETTER K WITH ACUTE]

"\u1E31" => "k"

# ḳ [LATIN SMALL LETTER K WITH DOT BELOW]

"\u1E33" => "k"

# ḵ [LATIN SMALL LETTER K WITH LINE BELOW]

"\u1E35" => "k"

# ⓚ [CIRCLED LATIN SMALL LETTER K]

"\u24DA" => "k"

# ⱪ [LATIN SMALL LETTER K WITH DESCENDER]

"\u2C6A" => "k"

# ꝁ [LATIN SMALL LETTER K WITH STROKE]

"\uA741" => "k"

# ꝃ [LATIN SMALL LETTER K WITH DIAGONAL STROKE]

"\uA743" => "k"

# ꝅ [LATIN SMALL LETTER K WITH STROKE AND DIAGONAL STROKE]

"\uA745" => "k"

# ｋ [FULLWIDTH LATIN SMALL LETTER K]

"\uFF4B" => "k"

# ⒦ [PARENTHESIZED LATIN SMALL LETTER K]

"\u24A6" => "(k)"

# Ĺ [LATIN CAPITAL LETTER L WITH ACUTE]

"\u0139" => "L"

# Ļ [LATIN CAPITAL LETTER L WITH CEDILLA]

"\u013B" => "L"

# Ľ [LATIN CAPITAL LETTER L WITH CARON]

"\u013D" => "L"

# Ŀ [LATIN CAPITAL LETTER L WITH MIDDLE DOT]

"\u013F" => "L"

# Ł [LATIN CAPITAL LETTER L WITH STROKE]

"\u0141" => "L"

# Ƚ [LATIN CAPITAL LETTER L WITH BAR]

"\u023D" => "L"

# ʟ [LATIN LETTER SMALL CAPITAL L]

"\u029F" => "L"

# ᴌ [LATIN LETTER SMALL CAPITAL L WITH STROKE]

"\u1D0C" => "L"

# Ḷ [LATIN CAPITAL LETTER L WITH DOT BELOW]

"\u1E36" => "L"

# Ḹ [LATIN CAPITAL LETTER L WITH DOT BELOW AND MACRON]

"\u1E38" => "L"

# Ḻ [LATIN CAPITAL LETTER L WITH LINE BELOW]

"\u1E3A" => "L"

# Ḽ [LATIN CAPITAL LETTER L WITH CIRCUMFLEX BELOW]

"\u1E3C" => "L"

# Ⓛ [CIRCLED LATIN CAPITAL LETTER L]

"\u24C1" => "L"

# Ⱡ [LATIN CAPITAL LETTER L WITH DOUBLE BAR]

"\u2C60" => "L"

# Ɫ [LATIN CAPITAL LETTER L WITH MIDDLE TILDE]

"\u2C62" => "L"

# Ꝇ [LATIN CAPITAL LETTER BROKEN L]

"\uA746" => "L"

# Ꝉ [LATIN CAPITAL LETTER L WITH HIGH STROKE]

"\uA748" => "L"

# Ꞁ [LATIN CAPITAL LETTER TURNED L]

"\uA780" => "L"

# Ｌ [FULLWIDTH LATIN CAPITAL LETTER L]

"\uFF2C" => "L"

# ĺ [LATIN SMALL LETTER L WITH ACUTE]

"\u013A" => "l"

# ļ [LATIN SMALL LETTER L WITH CEDILLA]

"\u013C" => "l"

# ľ [LATIN SMALL LETTER L WITH CARON]

"\u013E" => "l"

# ŀ [LATIN SMALL LETTER L WITH MIDDLE DOT]

"\u0140" => "l"

# ł [LATIN SMALL LETTER L WITH STROKE]

"\u0142" => "l"

# ƚ [LATIN SMALL LETTER L WITH BAR]

"\u019A" => "l"

# ȴ [LATIN SMALL LETTER L WITH CURL]

"\u0234" => "l"

# ɫ [LATIN SMALL LETTER L WITH MIDDLE TILDE]

"\u026B" => "l"

# ɬ [LATIN SMALL LETTER L WITH BELT]

"\u026C" => "l"

# ɭ [LATIN SMALL LETTER L WITH RETROFLEX HOOK]

"\u026D" => "l"

# ᶅ [LATIN SMALL LETTER L WITH PALATAL HOOK]

"\u1D85" => "l"

# ḷ [LATIN SMALL LETTER L WITH DOT BELOW]

"\u1E37" => "l"

# ḹ [LATIN SMALL LETTER L WITH DOT BELOW AND MACRON]

"\u1E39" => "l"

# ḻ [LATIN SMALL LETTER L WITH LINE BELOW]

"\u1E3B" => "l"

# ḽ [LATIN SMALL LETTER L WITH CIRCUMFLEX BELOW]

"\u1E3D" => "l"

# ⓛ [CIRCLED LATIN SMALL LETTER L]

"\u24DB" => "l"

# ⱡ [LATIN SMALL LETTER L WITH DOUBLE BAR]

"\u2C61" => "l"

# ꝇ [LATIN SMALL LETTER BROKEN L]

"\uA747" => "l"

# ꝉ [LATIN SMALL LETTER L WITH HIGH STROKE]

"\uA749" => "l"

# ꞁ [LATIN SMALL LETTER TURNED L]

"\uA781" => "l"

# ｌ [FULLWIDTH LATIN SMALL LETTER L]

"\uFF4C" => "l"

# Ǉ [LATIN CAPITAL LETTER LJ]

"\u01C7" => "LJ"

# Ỻ [LATIN CAPITAL LETTER MIDDLE-WELSH LL]

"\u1EFA" => "LL"

# ǈ [LATIN CAPITAL LETTER L WITH SMALL LETTER J]

"\u01C8" => "Lj"

# ⒧ [PARENTHESIZED LATIN SMALL LETTER L]

"\u24A7" => "(l)"

# ǉ [LATIN SMALL LETTER LJ]

"\u01C9" => "lj"

# ỻ [LATIN SMALL LETTER MIDDLE-WELSH LL]

"\u1EFB" => "ll"

# ʪ [LATIN SMALL LETTER LS DIGRAPH]

"\u02AA" => "ls"

# ʫ [LATIN SMALL LETTER LZ DIGRAPH]

"\u02AB" => "lz"

# Ɯ [LATIN CAPITAL LETTER TURNED M]

"\u019C" => "M"

# ᴍ [LATIN LETTER SMALL CAPITAL M]

"\u1D0D" => "M"

# Ḿ [LATIN CAPITAL LETTER M WITH ACUTE]

"\u1E3E" => "M"

# Ṁ [LATIN CAPITAL LETTER M WITH DOT ABOVE]

"\u1E40" => "M"

# Ṃ [LATIN CAPITAL LETTER M WITH DOT BELOW]

"\u1E42" => "M"

# Ⓜ [CIRCLED LATIN CAPITAL LETTER M]

"\u24C2" => "M"

# Ɱ [LATIN CAPITAL LETTER M WITH HOOK]

"\u2C6E" => "M"

# ꟽ [LATIN EPIGRAPHIC LETTER INVERTED M]

"\uA7FD" => "M"

# ꟿ [LATIN EPIGRAPHIC LETTER ARCHAIC M]

"\uA7FF" => "M"

# Ｍ [FULLWIDTH LATIN CAPITAL LETTER M]

"\uFF2D" => "M"

# ɯ [LATIN SMALL LETTER TURNED M]

"\u026F" => "m"

# ɰ [LATIN SMALL LETTER TURNED M WITH LONG LEG]

"\u0270" => "m"

# ɱ [LATIN SMALL LETTER M WITH HOOK]

"\u0271" => "m"

# ᵯ [LATIN SMALL LETTER M WITH MIDDLE TILDE]

"\u1D6F" => "m"

# ᶆ [LATIN SMALL LETTER M WITH PALATAL HOOK]

"\u1D86" => "m"

# ḿ [LATIN SMALL LETTER M WITH ACUTE]

"\u1E3F" => "m"

# ṁ [LATIN SMALL LETTER M WITH DOT ABOVE]

"\u1E41" => "m"

# ṃ [LATIN SMALL LETTER M WITH DOT BELOW]

"\u1E43" => "m"

# ⓜ [CIRCLED LATIN SMALL LETTER M]

"\u24DC" => "m"

# ｍ [FULLWIDTH LATIN SMALL LETTER M]

"\uFF4D" => "m"

# ⒨ [PARENTHESIZED LATIN SMALL LETTER M]

"\u24A8" => "(m)"

# Ñ [LATIN CAPITAL LETTER N WITH TILDE]

"\u00D1" => "N"

# Ń [LATIN CAPITAL LETTER N WITH ACUTE]

"\u0143" => "N"

# Ņ [LATIN CAPITAL LETTER N WITH CEDILLA]

"\u0145" => "N"

# Ň [LATIN CAPITAL LETTER N WITH CARON]

"\u0147" => "N"

# Ŋ http://en.wikipedia.org/wiki/Eng\_(letter) [LATIN CAPITAL LETTER ENG]

"\u014A" => "N"

# Ɲ [LATIN CAPITAL LETTER N WITH LEFT HOOK]

"\u019D" => "N"

# Ǹ [LATIN CAPITAL LETTER N WITH GRAVE]

"\u01F8" => "N"

# Ƞ [LATIN CAPITAL LETTER N WITH LONG RIGHT LEG]

"\u0220" => "N"

# ɴ [LATIN LETTER SMALL CAPITAL N]

"\u0274" => "N"

# ᴎ [LATIN LETTER SMALL CAPITAL REVERSED N]

"\u1D0E" => "N"

# Ṅ [LATIN CAPITAL LETTER N WITH DOT ABOVE]

"\u1E44" => "N"

# Ṇ [LATIN CAPITAL LETTER N WITH DOT BELOW]

"\u1E46" => "N"

# Ṉ [LATIN CAPITAL LETTER N WITH LINE BELOW]

"\u1E48" => "N"

# Ṋ [LATIN CAPITAL LETTER N WITH CIRCUMFLEX BELOW]

"\u1E4A" => "N"

# Ⓝ [CIRCLED LATIN CAPITAL LETTER N]

"\u24C3" => "N"

# Ｎ [FULLWIDTH LATIN CAPITAL LETTER N]

"\uFF2E" => "N"

# ñ [LATIN SMALL LETTER N WITH TILDE]

"\u00F1" => "n"

# ń [LATIN SMALL LETTER N WITH ACUTE]

"\u0144" => "n"

# ņ [LATIN SMALL LETTER N WITH CEDILLA]

"\u0146" => "n"

# ň [LATIN SMALL LETTER N WITH CARON]

"\u0148" => "n"

# ŉ [LATIN SMALL LETTER N PRECEDED BY APOSTROPHE]

"\u0149" => "n"

# ŋ http://en.wikipedia.org/wiki/Eng\_(letter) [LATIN SMALL LETTER ENG]

"\u014B" => "n"

# ƞ [LATIN SMALL LETTER N WITH LONG RIGHT LEG]

"\u019E" => "n"

# ǹ [LATIN SMALL LETTER N WITH GRAVE]

"\u01F9" => "n"

# ȵ [LATIN SMALL LETTER N WITH CURL]

"\u0235" => "n"

# ɲ [LATIN SMALL LETTER N WITH LEFT HOOK]

"\u0272" => "n"

# ɳ [LATIN SMALL LETTER N WITH RETROFLEX HOOK]

"\u0273" => "n"

# ᵰ [LATIN SMALL LETTER N WITH MIDDLE TILDE]

"\u1D70" => "n"

# ᶇ [LATIN SMALL LETTER N WITH PALATAL HOOK]

"\u1D87" => "n"

# ṅ [LATIN SMALL LETTER N WITH DOT ABOVE]

"\u1E45" => "n"

# ṇ [LATIN SMALL LETTER N WITH DOT BELOW]

"\u1E47" => "n"

# ṉ [LATIN SMALL LETTER N WITH LINE BELOW]

"\u1E49" => "n"

# ṋ [LATIN SMALL LETTER N WITH CIRCUMFLEX BELOW]

"\u1E4B" => "n"

# ⁿ [SUPERSCRIPT LATIN SMALL LETTER N]

"\u207F" => "n"

# ⓝ [CIRCLED LATIN SMALL LETTER N]

"\u24DD" => "n"

# ｎ [FULLWIDTH LATIN SMALL LETTER N]

"\uFF4E" => "n"

# Ǌ [LATIN CAPITAL LETTER NJ]

"\u01CA" => "NJ"

# ǋ [LATIN CAPITAL LETTER N WITH SMALL LETTER J]

"\u01CB" => "Nj"

# ⒩ [PARENTHESIZED LATIN SMALL LETTER N]

"\u24A9" => "(n)"

# ǌ [LATIN SMALL LETTER NJ]

"\u01CC" => "nj"

# Ò [LATIN CAPITAL LETTER O WITH GRAVE]

"\u00D2" => "O"

# Ó [LATIN CAPITAL LETTER O WITH ACUTE]

"\u00D3" => "O"

# Ô [LATIN CAPITAL LETTER O WITH CIRCUMFLEX]

"\u00D4" => "O"

# Õ [LATIN CAPITAL LETTER O WITH TILDE]

"\u00D5" => "O"

# Ö [LATIN CAPITAL LETTER O WITH DIAERESIS]

"\u00D6" => "O"

# Ø [LATIN CAPITAL LETTER O WITH STROKE]

"\u00D8" => "O"

# Ō [LATIN CAPITAL LETTER O WITH MACRON]

"\u014C" => "O"

# Ŏ [LATIN CAPITAL LETTER O WITH BREVE]

"\u014E" => "O"

# Ő [LATIN CAPITAL LETTER O WITH DOUBLE ACUTE]

"\u0150" => "O"

# Ɔ [LATIN CAPITAL LETTER OPEN O]

"\u0186" => "O"

# Ɵ [LATIN CAPITAL LETTER O WITH MIDDLE TILDE]

"\u019F" => "O"

# Ơ [LATIN CAPITAL LETTER O WITH HORN]

"\u01A0" => "O"

# Ǒ [LATIN CAPITAL LETTER O WITH CARON]

"\u01D1" => "O"

# Ǫ [LATIN CAPITAL LETTER O WITH OGONEK]

"\u01EA" => "O"

# Ǭ [LATIN CAPITAL LETTER O WITH OGONEK AND MACRON]

"\u01EC" => "O"

# Ǿ [LATIN CAPITAL LETTER O WITH STROKE AND ACUTE]

"\u01FE" => "O"

# Ȍ [LATIN CAPITAL LETTER O WITH DOUBLE GRAVE]

"\u020C" => "O"

# Ȏ [LATIN CAPITAL LETTER O WITH INVERTED BREVE]

"\u020E" => "O"

# Ȫ [LATIN CAPITAL LETTER O WITH DIAERESIS AND MACRON]

"\u022A" => "O"

# Ȭ [LATIN CAPITAL LETTER O WITH TILDE AND MACRON]

"\u022C" => "O"

# Ȯ [LATIN CAPITAL LETTER O WITH DOT ABOVE]

"\u022E" => "O"

# Ȱ [LATIN CAPITAL LETTER O WITH DOT ABOVE AND MACRON]

"\u0230" => "O"

# ᴏ [LATIN LETTER SMALL CAPITAL O]

"\u1D0F" => "O"

# ᴐ [LATIN LETTER SMALL CAPITAL OPEN O]

"\u1D10" => "O"

# Ṍ [LATIN CAPITAL LETTER O WITH TILDE AND ACUTE]

"\u1E4C" => "O"

# Ṏ [LATIN CAPITAL LETTER O WITH TILDE AND DIAERESIS]

"\u1E4E" => "O"

# Ṑ [LATIN CAPITAL LETTER O WITH MACRON AND GRAVE]

"\u1E50" => "O"

# Ṓ [LATIN CAPITAL LETTER O WITH MACRON AND ACUTE]

"\u1E52" => "O"

# Ọ [LATIN CAPITAL LETTER O WITH DOT BELOW]

"\u1ECC" => "O"

# Ỏ [LATIN CAPITAL LETTER O WITH HOOK ABOVE]

"\u1ECE" => "O"

# Ố [LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND ACUTE]

"\u1ED0" => "O"

# Ồ [LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND GRAVE]

"\u1ED2" => "O"

# Ổ [LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND HOOK ABOVE]

"\u1ED4" => "O"

# Ỗ [LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND TILDE]

"\u1ED6" => "O"

# Ộ [LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND DOT BELOW]

"\u1ED8" => "O"

# Ớ [LATIN CAPITAL LETTER O WITH HORN AND ACUTE]

"\u1EDA" => "O"

# Ờ [LATIN CAPITAL LETTER O WITH HORN AND GRAVE]

"\u1EDC" => "O"

# Ở [LATIN CAPITAL LETTER O WITH HORN AND HOOK ABOVE]

"\u1EDE" => "O"

# Ỡ [LATIN CAPITAL LETTER O WITH HORN AND TILDE]

"\u1EE0" => "O"

# Ợ [LATIN CAPITAL LETTER O WITH HORN AND DOT BELOW]

"\u1EE2" => "O"

# Ⓞ [CIRCLED LATIN CAPITAL LETTER O]

"\u24C4" => "O"

# Ꝋ [LATIN CAPITAL LETTER O WITH LONG STROKE OVERLAY]

"\uA74A" => "O"

# Ꝍ [LATIN CAPITAL LETTER O WITH LOOP]

"\uA74C" => "O"

# Ｏ [FULLWIDTH LATIN CAPITAL LETTER O]

"\uFF2F" => "O"

# ò [LATIN SMALL LETTER O WITH GRAVE]

"\u00F2" => "o"

# ó [LATIN SMALL LETTER O WITH ACUTE]

"\u00F3" => "o"

# ô [LATIN SMALL LETTER O WITH CIRCUMFLEX]

"\u00F4" => "o"

# õ [LATIN SMALL LETTER O WITH TILDE]

"\u00F5" => "o"

# ö [LATIN SMALL LETTER O WITH DIAERESIS]

"\u00F6" => "o"

# ø [LATIN SMALL LETTER O WITH STROKE]

"\u00F8" => "o"

# ō [LATIN SMALL LETTER O WITH MACRON]

"\u014D" => "o"

# ŏ [LATIN SMALL LETTER O WITH BREVE]

"\u014F" => "o"

# ő [LATIN SMALL LETTER O WITH DOUBLE ACUTE]

"\u0151" => "o"

# ơ [LATIN SMALL LETTER O WITH HORN]

"\u01A1" => "o"

# ǒ [LATIN SMALL LETTER O WITH CARON]

"\u01D2" => "o"

# ǫ [LATIN SMALL LETTER O WITH OGONEK]

"\u01EB" => "o"

# ǭ [LATIN SMALL LETTER O WITH OGONEK AND MACRON]

"\u01ED" => "o"

# ǿ [LATIN SMALL LETTER O WITH STROKE AND ACUTE]

"\u01FF" => "o"

# ȍ [LATIN SMALL LETTER O WITH DOUBLE GRAVE]

"\u020D" => "o"

# ȏ [LATIN SMALL LETTER O WITH INVERTED BREVE]

"\u020F" => "o"

# ȫ [LATIN SMALL LETTER O WITH DIAERESIS AND MACRON]

"\u022B" => "o"

# ȭ [LATIN SMALL LETTER O WITH TILDE AND MACRON]

"\u022D" => "o"

# ȯ [LATIN SMALL LETTER O WITH DOT ABOVE]

"\u022F" => "o"

# ȱ [LATIN SMALL LETTER O WITH DOT ABOVE AND MACRON]

"\u0231" => "o"

# ɔ [LATIN SMALL LETTER OPEN O]

"\u0254" => "o"

# ɵ [LATIN SMALL LETTER BARRED O]

"\u0275" => "o"

# ᴖ [LATIN SMALL LETTER TOP HALF O]

"\u1D16" => "o"

# ᴗ [LATIN SMALL LETTER BOTTOM HALF O]

"\u1D17" => "o"

# ᶗ [LATIN SMALL LETTER OPEN O WITH RETROFLEX HOOK]

"\u1D97" => "o"

# ṍ [LATIN SMALL LETTER O WITH TILDE AND ACUTE]

"\u1E4D" => "o"

# ṏ [LATIN SMALL LETTER O WITH TILDE AND DIAERESIS]

"\u1E4F" => "o"

# ṑ [LATIN SMALL LETTER O WITH MACRON AND GRAVE]

"\u1E51" => "o"

# ṓ [LATIN SMALL LETTER O WITH MACRON AND ACUTE]

"\u1E53" => "o"

# ọ [LATIN SMALL LETTER O WITH DOT BELOW]

"\u1ECD" => "o"

# ỏ [LATIN SMALL LETTER O WITH HOOK ABOVE]

"\u1ECF" => "o"

# ố [LATIN SMALL LETTER O WITH CIRCUMFLEX AND ACUTE]

"\u1ED1" => "o"

# ồ [LATIN SMALL LETTER O WITH CIRCUMFLEX AND GRAVE]

"\u1ED3" => "o"

# ổ [LATIN SMALL LETTER O WITH CIRCUMFLEX AND HOOK ABOVE]

"\u1ED5" => "o"

# ỗ [LATIN SMALL LETTER O WITH CIRCUMFLEX AND TILDE]

"\u1ED7" => "o"

# ộ [LATIN SMALL LETTER O WITH CIRCUMFLEX AND DOT BELOW]

"\u1ED9" => "o"

# ớ [LATIN SMALL LETTER O WITH HORN AND ACUTE]

"\u1EDB" => "o"

# ờ [LATIN SMALL LETTER O WITH HORN AND GRAVE]

"\u1EDD" => "o"

# ở [LATIN SMALL LETTER O WITH HORN AND HOOK ABOVE]

"\u1EDF" => "o"

# ỡ [LATIN SMALL LETTER O WITH HORN AND TILDE]

"\u1EE1" => "o"

# ợ [LATIN SMALL LETTER O WITH HORN AND DOT BELOW]

"\u1EE3" => "o"

# ₒ [LATIN SUBSCRIPT SMALL LETTER O]

"\u2092" => "o"

# ⓞ [CIRCLED LATIN SMALL LETTER O]

"\u24DE" => "o"

# ⱺ [LATIN SMALL LETTER O WITH LOW RING INSIDE]

"\u2C7A" => "o"

# ꝋ [LATIN SMALL LETTER O WITH LONG STROKE OVERLAY]

"\uA74B" => "o"

# ꝍ [LATIN SMALL LETTER O WITH LOOP]

"\uA74D" => "o"

# ｏ [FULLWIDTH LATIN SMALL LETTER O]

"\uFF4F" => "o"

# Œ [LATIN CAPITAL LIGATURE OE]

"\u0152" => "OE"

# ɶ [LATIN LETTER SMALL CAPITAL OE]

"\u0276" => "OE"

# Ꝏ [LATIN CAPITAL LETTER OO]

"\uA74E" => "OO"

# Ȣ http://en.wikipedia.org/wiki/OU [LATIN CAPITAL LETTER OU]

"\u0222" => "OU"

# ᴕ [LATIN LETTER SMALL CAPITAL OU]

"\u1D15" => "OU"

# ⒪ [PARENTHESIZED LATIN SMALL LETTER O]

"\u24AA" => "(o)"

# œ [LATIN SMALL LIGATURE OE]

"\u0153" => "oe"

# ᴔ [LATIN SMALL LETTER TURNED OE]

"\u1D14" => "oe"

# ꝏ [LATIN SMALL LETTER OO]

"\uA74F" => "oo"

# ȣ http://en.wikipedia.org/wiki/OU [LATIN SMALL LETTER OU]

"\u0223" => "ou"

# Ƥ [LATIN CAPITAL LETTER P WITH HOOK]

"\u01A4" => "P"

# ᴘ [LATIN LETTER SMALL CAPITAL P]

"\u1D18" => "P"

# Ṕ [LATIN CAPITAL LETTER P WITH ACUTE]

"\u1E54" => "P"

# Ṗ [LATIN CAPITAL LETTER P WITH DOT ABOVE]

"\u1E56" => "P"

# Ⓟ [CIRCLED LATIN CAPITAL LETTER P]

"\u24C5" => "P"

# Ᵽ [LATIN CAPITAL LETTER P WITH STROKE]

"\u2C63" => "P"

# Ꝑ [LATIN CAPITAL LETTER P WITH STROKE THROUGH DESCENDER]

"\uA750" => "P"

# Ꝓ [LATIN CAPITAL LETTER P WITH FLOURISH]

"\uA752" => "P"

# Ꝕ [LATIN CAPITAL LETTER P WITH SQUIRREL TAIL]

"\uA754" => "P"

# Ｐ [FULLWIDTH LATIN CAPITAL LETTER P]

"\uFF30" => "P"

# ƥ [LATIN SMALL LETTER P WITH HOOK]

"\u01A5" => "p"

# ᵱ [LATIN SMALL LETTER P WITH MIDDLE TILDE]

"\u1D71" => "p"

# ᵽ [LATIN SMALL LETTER P WITH STROKE]

"\u1D7D" => "p"

# ᶈ [LATIN SMALL LETTER P WITH PALATAL HOOK]

"\u1D88" => "p"

# ṕ [LATIN SMALL LETTER P WITH ACUTE]

"\u1E55" => "p"

# ṗ [LATIN SMALL LETTER P WITH DOT ABOVE]

"\u1E57" => "p"

# ⓟ [CIRCLED LATIN SMALL LETTER P]

"\u24DF" => "p"

# ꝑ [LATIN SMALL LETTER P WITH STROKE THROUGH DESCENDER]

"\uA751" => "p"

# ꝓ [LATIN SMALL LETTER P WITH FLOURISH]

"\uA753" => "p"

# ꝕ [LATIN SMALL LETTER P WITH SQUIRREL TAIL]

"\uA755" => "p"

# ꟼ [LATIN EPIGRAPHIC LETTER REVERSED P]

"\uA7FC" => "p"

# ｐ [FULLWIDTH LATIN SMALL LETTER P]

"\uFF50" => "p"

# ⒫ [PARENTHESIZED LATIN SMALL LETTER P]

"\u24AB" => "(p)"

# Ɋ [LATIN CAPITAL LETTER SMALL Q WITH HOOK TAIL]

"\u024A" => "Q"

# Ⓠ [CIRCLED LATIN CAPITAL LETTER Q]

"\u24C6" => "Q"

# Ꝗ [LATIN CAPITAL LETTER Q WITH STROKE THROUGH DESCENDER]

"\uA756" => "Q"

# Ꝙ [LATIN CAPITAL LETTER Q WITH DIAGONAL STROKE]

"\uA758" => "Q"

# Ｑ [FULLWIDTH LATIN CAPITAL LETTER Q]

"\uFF31" => "Q"

# ĸ http://en.wikipedia.org/wiki/Kra\_(letter) [LATIN SMALL LETTER KRA]

"\u0138" => "q"

# ɋ [LATIN SMALL LETTER Q WITH HOOK TAIL]

"\u024B" => "q"

# ʠ [LATIN SMALL LETTER Q WITH HOOK]

"\u02A0" => "q"

# ⓠ [CIRCLED LATIN SMALL LETTER Q]

"\u24E0" => "q"

# ꝗ [LATIN SMALL LETTER Q WITH STROKE THROUGH DESCENDER]

"\uA757" => "q"

# ꝙ [LATIN SMALL LETTER Q WITH DIAGONAL STROKE]

"\uA759" => "q"

# ｑ [FULLWIDTH LATIN SMALL LETTER Q]

"\uFF51" => "q"

# ⒬ [PARENTHESIZED LATIN SMALL LETTER Q]

"\u24AC" => "(q)"

# ȹ [LATIN SMALL LETTER QP DIGRAPH]

"\u0239" => "qp"

# Ŕ [LATIN CAPITAL LETTER R WITH ACUTE]

"\u0154" => "R"

# Ŗ [LATIN CAPITAL LETTER R WITH CEDILLA]

"\u0156" => "R"

# Ř [LATIN CAPITAL LETTER R WITH CARON]

"\u0158" => "R"

# Ȓ [LATIN CAPITAL LETTER R WITH DOUBLE GRAVE]

"\u0210" => "R"

# Ȓ [LATIN CAPITAL LETTER R WITH INVERTED BREVE]

"\u0212" => "R"

# Ɍ [LATIN CAPITAL LETTER R WITH STROKE]

"\u024C" => "R"

# ʀ [LATIN LETTER SMALL CAPITAL R]

"\u0280" => "R"

# ʁ [LATIN LETTER SMALL CAPITAL INVERTED R]

"\u0281" => "R"

# ᴙ [LATIN LETTER SMALL CAPITAL REVERSED R]

"\u1D19" => "R"

# ᴚ [LATIN LETTER SMALL CAPITAL TURNED R]

"\u1D1A" => "R"

# Ṙ [LATIN CAPITAL LETTER R WITH DOT ABOVE]

"\u1E58" => "R"

# Ṛ [LATIN CAPITAL LETTER R WITH DOT BELOW]

"\u1E5A" => "R"

# Ṝ [LATIN CAPITAL LETTER R WITH DOT BELOW AND MACRON]

"\u1E5C" => "R"

# Ṟ [LATIN CAPITAL LETTER R WITH LINE BELOW]

"\u1E5E" => "R"

# Ⓡ [CIRCLED LATIN CAPITAL LETTER R]

"\u24C7" => "R"

# Ɽ [LATIN CAPITAL LETTER R WITH TAIL]

"\u2C64" => "R"

# Ꝛ [LATIN CAPITAL LETTER R ROTUNDA]

"\uA75A" => "R"

# Ꞃ [LATIN CAPITAL LETTER INSULAR R]

"\uA782" => "R"

# Ｒ [FULLWIDTH LATIN CAPITAL LETTER R]

"\uFF32" => "R"

# ŕ [LATIN SMALL LETTER R WITH ACUTE]

"\u0155" => "r"

# ŗ [LATIN SMALL LETTER R WITH CEDILLA]

"\u0157" => "r"

# ř [LATIN SMALL LETTER R WITH CARON]

"\u0159" => "r"

# ȑ [LATIN SMALL LETTER R WITH DOUBLE GRAVE]

"\u0211" => "r"

# ȓ [LATIN SMALL LETTER R WITH INVERTED BREVE]

"\u0213" => "r"

# ɍ [LATIN SMALL LETTER R WITH STROKE]

"\u024D" => "r"

# ɼ [LATIN SMALL LETTER R WITH LONG LEG]

"\u027C" => "r"

# ɽ [LATIN SMALL LETTER R WITH TAIL]

"\u027D" => "r"

# ɾ [LATIN SMALL LETTER R WITH FISHHOOK]

"\u027E" => "r"

# ɿ [LATIN SMALL LETTER REVERSED R WITH FISHHOOK]

"\u027F" => "r"

# ᵣ [LATIN SUBSCRIPT SMALL LETTER R]

"\u1D63" => "r"

# ᵲ [LATIN SMALL LETTER R WITH MIDDLE TILDE]

"\u1D72" => "r"

# ᵳ [LATIN SMALL LETTER R WITH FISHHOOK AND MIDDLE TILDE]

"\u1D73" => "r"

# ᶉ [LATIN SMALL LETTER R WITH PALATAL HOOK]

"\u1D89" => "r"

# ṙ [LATIN SMALL LETTER R WITH DOT ABOVE]

"\u1E59" => "r"

# ṛ [LATIN SMALL LETTER R WITH DOT BELOW]

"\u1E5B" => "r"

# ṝ [LATIN SMALL LETTER R WITH DOT BELOW AND MACRON]

"\u1E5D" => "r"

# ṟ [LATIN SMALL LETTER R WITH LINE BELOW]

"\u1E5F" => "r"

# ⓡ [CIRCLED LATIN SMALL LETTER R]

"\u24E1" => "r"

# ꝛ [LATIN SMALL LETTER R ROTUNDA]

"\uA75B" => "r"

# ꞃ [LATIN SMALL LETTER INSULAR R]

"\uA783" => "r"

# ｒ [FULLWIDTH LATIN SMALL LETTER R]

"\uFF52" => "r"

# ⒭ [PARENTHESIZED LATIN SMALL LETTER R]

"\u24AD" => "(r)"

# Ś [LATIN CAPITAL LETTER S WITH ACUTE]

"\u015A" => "S"

# Ŝ [LATIN CAPITAL LETTER S WITH CIRCUMFLEX]

"\u015C" => "S"

# Ş [LATIN CAPITAL LETTER S WITH CEDILLA]

"\u015E" => "S"

# Š [LATIN CAPITAL LETTER S WITH CARON]

"\u0160" => "S"

# Ș [LATIN CAPITAL LETTER S WITH COMMA BELOW]

"\u0218" => "S"

# Ṡ [LATIN CAPITAL LETTER S WITH DOT ABOVE]

"\u1E60" => "S"

# Ṣ [LATIN CAPITAL LETTER S WITH DOT BELOW]

"\u1E62" => "S"

# Ṥ [LATIN CAPITAL LETTER S WITH ACUTE AND DOT ABOVE]

"\u1E64" => "S"

# Ṧ [LATIN CAPITAL LETTER S WITH CARON AND DOT ABOVE]

"\u1E66" => "S"

# Ṩ [LATIN CAPITAL LETTER S WITH DOT BELOW AND DOT ABOVE]

"\u1E68" => "S"

# Ⓢ [CIRCLED LATIN CAPITAL LETTER S]

"\u24C8" => "S"

# ꜱ [LATIN LETTER SMALL CAPITAL S]

"\uA731" => "S"

# ꞅ [LATIN SMALL LETTER INSULAR S]

"\uA785" => "S"

# Ｓ [FULLWIDTH LATIN CAPITAL LETTER S]

"\uFF33" => "S"

# ś [LATIN SMALL LETTER S WITH ACUTE]

"\u015B" => "s"

# ŝ [LATIN SMALL LETTER S WITH CIRCUMFLEX]

"\u015D" => "s"

# ş [LATIN SMALL LETTER S WITH CEDILLA]

"\u015F" => "s"

# š [LATIN SMALL LETTER S WITH CARON]

"\u0161" => "s"

# ſ http://en.wikipedia.org/wiki/Long\_S [LATIN SMALL LETTER LONG S]

"\u017F" => "s"

# ș [LATIN SMALL LETTER S WITH COMMA BELOW]

"\u0219" => "s"

# ȿ [LATIN SMALL LETTER S WITH SWASH TAIL]

"\u023F" => "s"

# ʂ [LATIN SMALL LETTER S WITH HOOK]

"\u0282" => "s"

# ᵴ [LATIN SMALL LETTER S WITH MIDDLE TILDE]

"\u1D74" => "s"

# ᶊ [LATIN SMALL LETTER S WITH PALATAL HOOK]

"\u1D8A" => "s"

# ṡ [LATIN SMALL LETTER S WITH DOT ABOVE]

"\u1E61" => "s"

# ṣ [LATIN SMALL LETTER S WITH DOT BELOW]

"\u1E63" => "s"

# ṥ [LATIN SMALL LETTER S WITH ACUTE AND DOT ABOVE]

"\u1E65" => "s"

# ṧ [LATIN SMALL LETTER S WITH CARON AND DOT ABOVE]

"\u1E67" => "s"

# ṩ [LATIN SMALL LETTER S WITH DOT BELOW AND DOT ABOVE]

"\u1E69" => "s"

# ẜ [LATIN SMALL LETTER LONG S WITH DIAGONAL STROKE]

"\u1E9C" => "s"

# ẝ [LATIN SMALL LETTER LONG S WITH HIGH STROKE]

"\u1E9D" => "s"

# ⓢ [CIRCLED LATIN SMALL LETTER S]

"\u24E2" => "s"

# Ꞅ [LATIN CAPITAL LETTER INSULAR S]

"\uA784" => "s"

# ｓ [FULLWIDTH LATIN SMALL LETTER S]

"\uFF53" => "s"

# ẞ [LATIN CAPITAL LETTER SHARP S]

"\u1E9E" => "SS"

# ⒮ [PARENTHESIZED LATIN SMALL LETTER S]

"\u24AE" => "(s)"

# ß [LATIN SMALL LETTER SHARP S]

"\u00DF" => "ss"

# ﬆ [LATIN SMALL LIGATURE ST]

"\uFB06" => "st"

# Ţ [LATIN CAPITAL LETTER T WITH CEDILLA]

"\u0162" => "T"

# Ť [LATIN CAPITAL LETTER T WITH CARON]

"\u0164" => "T"

# Ŧ [LATIN CAPITAL LETTER T WITH STROKE]

"\u0166" => "T"

# Ƭ [LATIN CAPITAL LETTER T WITH HOOK]

"\u01AC" => "T"

# Ʈ [LATIN CAPITAL LETTER T WITH RETROFLEX HOOK]

"\u01AE" => "T"

# Ț [LATIN CAPITAL LETTER T WITH COMMA BELOW]

"\u021A" => "T"

# Ⱦ [LATIN CAPITAL LETTER T WITH DIAGONAL STROKE]

"\u023E" => "T"

# ᴛ [LATIN LETTER SMALL CAPITAL T]

"\u1D1B" => "T"

# Ṫ [LATIN CAPITAL LETTER T WITH DOT ABOVE]

"\u1E6A" => "T"

# Ṭ [LATIN CAPITAL LETTER T WITH DOT BELOW]

"\u1E6C" => "T"

# Ṯ [LATIN CAPITAL LETTER T WITH LINE BELOW]

"\u1E6E" => "T"

# Ṱ [LATIN CAPITAL LETTER T WITH CIRCUMFLEX BELOW]

"\u1E70" => "T"

# Ⓣ [CIRCLED LATIN CAPITAL LETTER T]

"\u24C9" => "T"

# Ꞇ [LATIN CAPITAL LETTER INSULAR T]

"\uA786" => "T"

# Ｔ [FULLWIDTH LATIN CAPITAL LETTER T]

"\uFF34" => "T"

# ţ [LATIN SMALL LETTER T WITH CEDILLA]

"\u0163" => "t"

# ť [LATIN SMALL LETTER T WITH CARON]

"\u0165" => "t"

# ŧ [LATIN SMALL LETTER T WITH STROKE]

"\u0167" => "t"

# ƫ [LATIN SMALL LETTER T WITH PALATAL HOOK]

"\u01AB" => "t"

# ƭ [LATIN SMALL LETTER T WITH HOOK]

"\u01AD" => "t"

# ț [LATIN SMALL LETTER T WITH COMMA BELOW]

"\u021B" => "t"

# ȶ [LATIN SMALL LETTER T WITH CURL]

"\u0236" => "t"

# ʇ [LATIN SMALL LETTER TURNED T]

"\u0287" => "t"

# ʈ [LATIN SMALL LETTER T WITH RETROFLEX HOOK]

"\u0288" => "t"

# ᵵ [LATIN SMALL LETTER T WITH MIDDLE TILDE]

"\u1D75" => "t"

# ṫ [LATIN SMALL LETTER T WITH DOT ABOVE]

"\u1E6B" => "t"

# ṭ [LATIN SMALL LETTER T WITH DOT BELOW]

"\u1E6D" => "t"

# ṯ [LATIN SMALL LETTER T WITH LINE BELOW]

"\u1E6F" => "t"

# ṱ [LATIN SMALL LETTER T WITH CIRCUMFLEX BELOW]

"\u1E71" => "t"

# ẗ [LATIN SMALL LETTER T WITH DIAERESIS]

"\u1E97" => "t"

# ⓣ [CIRCLED LATIN SMALL LETTER T]

"\u24E3" => "t"

# ⱦ [LATIN SMALL LETTER T WITH DIAGONAL STROKE]

"\u2C66" => "t"

# ｔ [FULLWIDTH LATIN SMALL LETTER T]

"\uFF54" => "t"

# Þ [LATIN CAPITAL LETTER THORN]

"\u00DE" => "TH"

# Ꝧ [LATIN CAPITAL LETTER THORN WITH STROKE THROUGH DESCENDER]

"\uA766" => "TH"

# Ꜩ [LATIN CAPITAL LETTER TZ]

"\uA728" => "TZ"

# ⒯ [PARENTHESIZED LATIN SMALL LETTER T]

"\u24AF" => "(t)"

# ʨ [LATIN SMALL LETTER TC DIGRAPH WITH CURL]

"\u02A8" => "tc"

# þ [LATIN SMALL LETTER THORN]

"\u00FE" => "th"

# ᵺ [LATIN SMALL LETTER TH WITH STRIKETHROUGH]

"\u1D7A" => "th"

# ꝧ [LATIN SMALL LETTER THORN WITH STROKE THROUGH DESCENDER]

"\uA767" => "th"

# ʦ [LATIN SMALL LETTER TS DIGRAPH]

"\u02A6" => "ts"

# ꜩ [LATIN SMALL LETTER TZ]

"\uA729" => "tz"

# Ù [LATIN CAPITAL LETTER U WITH GRAVE]

"\u00D9" => "U"

# Ú [LATIN CAPITAL LETTER U WITH ACUTE]

"\u00DA" => "U"

# Û [LATIN CAPITAL LETTER U WITH CIRCUMFLEX]

"\u00DB" => "U"

# Ü [LATIN CAPITAL LETTER U WITH DIAERESIS]

"\u00DC" => "U"

# Ũ [LATIN CAPITAL LETTER U WITH TILDE]

"\u0168" => "U"

# Ū [LATIN CAPITAL LETTER U WITH MACRON]

"\u016A" => "U"

# Ŭ [LATIN CAPITAL LETTER U WITH BREVE]

"\u016C" => "U"

# Ů [LATIN CAPITAL LETTER U WITH RING ABOVE]

"\u016E" => "U"

# Ű [LATIN CAPITAL LETTER U WITH DOUBLE ACUTE]

"\u0170" => "U"

# Ų [LATIN CAPITAL LETTER U WITH OGONEK]

"\u0172" => "U"

# Ư [LATIN CAPITAL LETTER U WITH HORN]

"\u01AF" => "U"

# Ǔ [LATIN CAPITAL LETTER U WITH CARON]

"\u01D3" => "U"

# Ǖ [LATIN CAPITAL LETTER U WITH DIAERESIS AND MACRON]

"\u01D5" => "U"

# Ǘ [LATIN CAPITAL LETTER U WITH DIAERESIS AND ACUTE]

"\u01D7" => "U"

# Ǚ [LATIN CAPITAL LETTER U WITH DIAERESIS AND CARON]

"\u01D9" => "U"

# Ǜ [LATIN CAPITAL LETTER U WITH DIAERESIS AND GRAVE]

"\u01DB" => "U"

# Ȕ [LATIN CAPITAL LETTER U WITH DOUBLE GRAVE]

"\u0214" => "U"

# Ȗ [LATIN CAPITAL LETTER U WITH INVERTED BREVE]

"\u0216" => "U"

# Ʉ [LATIN CAPITAL LETTER U BAR]

"\u0244" => "U"

# ᴜ [LATIN LETTER SMALL CAPITAL U]

"\u1D1C" => "U"

# ᵾ [LATIN SMALL CAPITAL LETTER U WITH STROKE]

"\u1D7E" => "U"

# Ṳ [LATIN CAPITAL LETTER U WITH DIAERESIS BELOW]

"\u1E72" => "U"

# Ṵ [LATIN CAPITAL LETTER U WITH TILDE BELOW]

"\u1E74" => "U"

# Ṷ [LATIN CAPITAL LETTER U WITH CIRCUMFLEX BELOW]

"\u1E76" => "U"

# Ṹ [LATIN CAPITAL LETTER U WITH TILDE AND ACUTE]

"\u1E78" => "U"

# Ṻ [LATIN CAPITAL LETTER U WITH MACRON AND DIAERESIS]

"\u1E7A" => "U"

# Ụ [LATIN CAPITAL LETTER U WITH DOT BELOW]

"\u1EE4" => "U"

# Ủ [LATIN CAPITAL LETTER U WITH HOOK ABOVE]

"\u1EE6" => "U"

# Ứ [LATIN CAPITAL LETTER U WITH HORN AND ACUTE]

"\u1EE8" => "U"

# Ừ [LATIN CAPITAL LETTER U WITH HORN AND GRAVE]

"\u1EEA" => "U"

# Ử [LATIN CAPITAL LETTER U WITH HORN AND HOOK ABOVE]

"\u1EEC" => "U"

# Ữ [LATIN CAPITAL LETTER U WITH HORN AND TILDE]

"\u1EEE" => "U"

# Ự [LATIN CAPITAL LETTER U WITH HORN AND DOT BELOW]

"\u1EF0" => "U"

# Ⓤ [CIRCLED LATIN CAPITAL LETTER U]

"\u24CA" => "U"

# Ｕ [FULLWIDTH LATIN CAPITAL LETTER U]

"\uFF35" => "U"

# ù [LATIN SMALL LETTER U WITH GRAVE]

"\u00F9" => "u"

# ú [LATIN SMALL LETTER U WITH ACUTE]

"\u00FA" => "u"

# û [LATIN SMALL LETTER U WITH CIRCUMFLEX]

"\u00FB" => "u"

# ü [LATIN SMALL LETTER U WITH DIAERESIS]

"\u00FC" => "u"

# ũ [LATIN SMALL LETTER U WITH TILDE]

"\u0169" => "u"

# ū [LATIN SMALL LETTER U WITH MACRON]

"\u016B" => "u"

# ŭ [LATIN SMALL LETTER U WITH BREVE]

"\u016D" => "u"

# ů [LATIN SMALL LETTER U WITH RING ABOVE]

"\u016F" => "u"

# ű [LATIN SMALL LETTER U WITH DOUBLE ACUTE]

"\u0171" => "u"

# ų [LATIN SMALL LETTER U WITH OGONEK]

"\u0173" => "u"

# ư [LATIN SMALL LETTER U WITH HORN]

"\u01B0" => "u"

# ǔ [LATIN SMALL LETTER U WITH CARON]

"\u01D4" => "u"

# ǖ [LATIN SMALL LETTER U WITH DIAERESIS AND MACRON]

"\u01D6" => "u"

# ǘ [LATIN SMALL LETTER U WITH DIAERESIS AND ACUTE]

"\u01D8" => "u"

# ǚ [LATIN SMALL LETTER U WITH DIAERESIS AND CARON]

"\u01DA" => "u"

# ǜ [LATIN SMALL LETTER U WITH DIAERESIS AND GRAVE]

"\u01DC" => "u"

# ȕ [LATIN SMALL LETTER U WITH DOUBLE GRAVE]

"\u0215" => "u"

# ȗ [LATIN SMALL LETTER U WITH INVERTED BREVE]

"\u0217" => "u"

# ʉ [LATIN SMALL LETTER U BAR]

"\u0289" => "u"

# ᵤ [LATIN SUBSCRIPT SMALL LETTER U]

"\u1D64" => "u"

# ᶙ [LATIN SMALL LETTER U WITH RETROFLEX HOOK]

"\u1D99" => "u"

# ṳ [LATIN SMALL LETTER U WITH DIAERESIS BELOW]

"\u1E73" => "u"

# ṵ [LATIN SMALL LETTER U WITH TILDE BELOW]

"\u1E75" => "u"

# ṷ [LATIN SMALL LETTER U WITH CIRCUMFLEX BELOW]

"\u1E77" => "u"

# ṹ [LATIN SMALL LETTER U WITH TILDE AND ACUTE]

"\u1E79" => "u"

# ṻ [LATIN SMALL LETTER U WITH MACRON AND DIAERESIS]

"\u1E7B" => "u"

# ụ [LATIN SMALL LETTER U WITH DOT BELOW]

"\u1EE5" => "u"

# ủ [LATIN SMALL LETTER U WITH HOOK ABOVE]

"\u1EE7" => "u"

# ứ [LATIN SMALL LETTER U WITH HORN AND ACUTE]

"\u1EE9" => "u"

# ừ [LATIN SMALL LETTER U WITH HORN AND GRAVE]

"\u1EEB" => "u"

# ử [LATIN SMALL LETTER U WITH HORN AND HOOK ABOVE]

"\u1EED" => "u"

# ữ [LATIN SMALL LETTER U WITH HORN AND TILDE]

"\u1EEF" => "u"

# ự [LATIN SMALL LETTER U WITH HORN AND DOT BELOW]

"\u1EF1" => "u"

# ⓤ [CIRCLED LATIN SMALL LETTER U]

"\u24E4" => "u"

# ｕ [FULLWIDTH LATIN SMALL LETTER U]

"\uFF55" => "u"

# ⒰ [PARENTHESIZED LATIN SMALL LETTER U]

"\u24B0" => "(u)"

# ᵫ [LATIN SMALL LETTER UE]

"\u1D6B" => "ue"

# Ʋ [LATIN CAPITAL LETTER V WITH HOOK]

"\u01B2" => "V"

# Ʌ [LATIN CAPITAL LETTER TURNED V]

"\u0245" => "V"

# ᴠ [LATIN LETTER SMALL CAPITAL V]

"\u1D20" => "V"

# Ṽ [LATIN CAPITAL LETTER V WITH TILDE]

"\u1E7C" => "V"

# Ṿ [LATIN CAPITAL LETTER V WITH DOT BELOW]

"\u1E7E" => "V"

# Ỽ [LATIN CAPITAL LETTER MIDDLE-WELSH V]

"\u1EFC" => "V"

# Ⓥ [CIRCLED LATIN CAPITAL LETTER V]

"\u24CB" => "V"

# Ꝟ [LATIN CAPITAL LETTER V WITH DIAGONAL STROKE]

"\uA75E" => "V"

# Ꝩ [LATIN CAPITAL LETTER VEND]

"\uA768" => "V"

# Ｖ [FULLWIDTH LATIN CAPITAL LETTER V]

"\uFF36" => "V"

# ʋ [LATIN SMALL LETTER V WITH HOOK]

"\u028B" => "v"

# ʌ [LATIN SMALL LETTER TURNED V]

"\u028C" => "v"

# ᵥ [LATIN SUBSCRIPT SMALL LETTER V]

"\u1D65" => "v"

# ᶌ [LATIN SMALL LETTER V WITH PALATAL HOOK]

"\u1D8C" => "v"

# ṽ [LATIN SMALL LETTER V WITH TILDE]

"\u1E7D" => "v"

# ṿ [LATIN SMALL LETTER V WITH DOT BELOW]

"\u1E7F" => "v"

# ⓥ [CIRCLED LATIN SMALL LETTER V]

"\u24E5" => "v"

# ⱱ [LATIN SMALL LETTER V WITH RIGHT HOOK]

"\u2C71" => "v"

# ⱴ [LATIN SMALL LETTER V WITH CURL]

"\u2C74" => "v"

# ꝟ [LATIN SMALL LETTER V WITH DIAGONAL STROKE]

"\uA75F" => "v"

# ｖ [FULLWIDTH LATIN SMALL LETTER V]

"\uFF56" => "v"

# Ꝡ [LATIN CAPITAL LETTER VY]

"\uA760" => "VY"

# ⒱ [PARENTHESIZED LATIN SMALL LETTER V]

"\u24B1" => "(v)"

# ꝡ [LATIN SMALL LETTER VY]

"\uA761" => "vy"

# Ŵ [LATIN CAPITAL LETTER W WITH CIRCUMFLEX]

"\u0174" => "W"

# Ƿ http://en.wikipedia.org/wiki/Wynn [LATIN CAPITAL LETTER WYNN]

"\u01F7" => "W"

# ᴡ [LATIN LETTER SMALL CAPITAL W]

"\u1D21" => "W"

# Ẁ [LATIN CAPITAL LETTER W WITH GRAVE]

"\u1E80" => "W"

# Ẃ [LATIN CAPITAL LETTER W WITH ACUTE]

"\u1E82" => "W"

# Ẅ [LATIN CAPITAL LETTER W WITH DIAERESIS]

"\u1E84" => "W"

# Ẇ [LATIN CAPITAL LETTER W WITH DOT ABOVE]

"\u1E86" => "W"

# Ẉ [LATIN CAPITAL LETTER W WITH DOT BELOW]

"\u1E88" => "W"

# Ⓦ [CIRCLED LATIN CAPITAL LETTER W]

"\u24CC" => "W"

# Ⱳ [LATIN CAPITAL LETTER W WITH HOOK]

"\u2C72" => "W"

# Ｗ [FULLWIDTH LATIN CAPITAL LETTER W]

"\uFF37" => "W"

# ŵ [LATIN SMALL LETTER W WITH CIRCUMFLEX]

"\u0175" => "w"

# ƿ http://en.wikipedia.org/wiki/Wynn [LATIN LETTER WYNN]

"\u01BF" => "w"

# ʍ [LATIN SMALL LETTER TURNED W]

"\u028D" => "w"

# ẁ [LATIN SMALL LETTER W WITH GRAVE]

"\u1E81" => "w"

# ẃ [LATIN SMALL LETTER W WITH ACUTE]

"\u1E83" => "w"

# ẅ [LATIN SMALL LETTER W WITH DIAERESIS]

"\u1E85" => "w"

# ẇ [LATIN SMALL LETTER W WITH DOT ABOVE]

"\u1E87" => "w"

# ẉ [LATIN SMALL LETTER W WITH DOT BELOW]

"\u1E89" => "w"

# ẘ [LATIN SMALL LETTER W WITH RING ABOVE]

"\u1E98" => "w"

# ⓦ [CIRCLED LATIN SMALL LETTER W]

"\u24E6" => "w"

# ⱳ [LATIN SMALL LETTER W WITH HOOK]

"\u2C73" => "w"

# ｗ [FULLWIDTH LATIN SMALL LETTER W]

"\uFF57" => "w"

# ⒲ [PARENTHESIZED LATIN SMALL LETTER W]

"\u24B2" => "(w)"

# Ẋ [LATIN CAPITAL LETTER X WITH DOT ABOVE]

"\u1E8A" => "X"

# Ẍ [LATIN CAPITAL LETTER X WITH DIAERESIS]

"\u1E8C" => "X"

# Ⓧ [CIRCLED LATIN CAPITAL LETTER X]

"\u24CD" => "X"

# Ｘ [FULLWIDTH LATIN CAPITAL LETTER X]

"\uFF38" => "X"

# ᶍ [LATIN SMALL LETTER X WITH PALATAL HOOK]

"\u1D8D" => "x"

# ẋ [LATIN SMALL LETTER X WITH DOT ABOVE]

"\u1E8B" => "x"

# ẍ [LATIN SMALL LETTER X WITH DIAERESIS]

"\u1E8D" => "x"

# ₓ [LATIN SUBSCRIPT SMALL LETTER X]

"\u2093" => "x"

# ⓧ [CIRCLED LATIN SMALL LETTER X]

"\u24E7" => "x"

# ｘ [FULLWIDTH LATIN SMALL LETTER X]

"\uFF58" => "x"

# ⒳ [PARENTHESIZED LATIN SMALL LETTER X]

"\u24B3" => "(x)"

# Ý [LATIN CAPITAL LETTER Y WITH ACUTE]

"\u00DD" => "Y"

# Ŷ [LATIN CAPITAL LETTER Y WITH CIRCUMFLEX]

"\u0176" => "Y"

# Ÿ [LATIN CAPITAL LETTER Y WITH DIAERESIS]

"\u0178" => "Y"

# Ƴ [LATIN CAPITAL LETTER Y WITH HOOK]

"\u01B3" => "Y"

# Ȳ [LATIN CAPITAL LETTER Y WITH MACRON]

"\u0232" => "Y"

# Ɏ [LATIN CAPITAL LETTER Y WITH STROKE]

"\u024E" => "Y"

# ʏ [LATIN LETTER SMALL CAPITAL Y]

"\u028F" => "Y"

# Ẏ [LATIN CAPITAL LETTER Y WITH DOT ABOVE]

"\u1E8E" => "Y"

# Ỳ [LATIN CAPITAL LETTER Y WITH GRAVE]

"\u1EF2" => "Y"

# Ỵ [LATIN CAPITAL LETTER Y WITH DOT BELOW]

"\u1EF4" => "Y"

# Ỷ [LATIN CAPITAL LETTER Y WITH HOOK ABOVE]

"\u1EF6" => "Y"

# Ỹ [LATIN CAPITAL LETTER Y WITH TILDE]

"\u1EF8" => "Y"

# Ỿ [LATIN CAPITAL LETTER Y WITH LOOP]

"\u1EFE" => "Y"

# Ⓨ [CIRCLED LATIN CAPITAL LETTER Y]

"\u24CE" => "Y"

# Ｙ [FULLWIDTH LATIN CAPITAL LETTER Y]

"\uFF39" => "Y"

# ý [LATIN SMALL LETTER Y WITH ACUTE]

"\u00FD" => "y"

# ÿ [LATIN SMALL LETTER Y WITH DIAERESIS]

"\u00FF" => "y"

# ŷ [LATIN SMALL LETTER Y WITH CIRCUMFLEX]

"\u0177" => "y"

# ƴ [LATIN SMALL LETTER Y WITH HOOK]

"\u01B4" => "y"

# ȳ [LATIN SMALL LETTER Y WITH MACRON]

"\u0233" => "y"

# ɏ [LATIN SMALL LETTER Y WITH STROKE]

"\u024F" => "y"

# ʎ [LATIN SMALL LETTER TURNED Y]

"\u028E" => "y"

# ẏ [LATIN SMALL LETTER Y WITH DOT ABOVE]

"\u1E8F" => "y"

# ẙ [LATIN SMALL LETTER Y WITH RING ABOVE]

"\u1E99" => "y"

# ỳ [LATIN SMALL LETTER Y WITH GRAVE]

"\u1EF3" => "y"

# ỵ [LATIN SMALL LETTER Y WITH DOT BELOW]

"\u1EF5" => "y"

# ỷ [LATIN SMALL LETTER Y WITH HOOK ABOVE]

"\u1EF7" => "y"

# ỹ [LATIN SMALL LETTER Y WITH TILDE]

"\u1EF9" => "y"

# ỿ [LATIN SMALL LETTER Y WITH LOOP]

"\u1EFF" => "y"

# ⓨ [CIRCLED LATIN SMALL LETTER Y]

"\u24E8" => "y"

# ｙ [FULLWIDTH LATIN SMALL LETTER Y]

"\uFF59" => "y"

# ⒴ [PARENTHESIZED LATIN SMALL LETTER Y]

"\u24B4" => "(y)"

# Ź [LATIN CAPITAL LETTER Z WITH ACUTE]

"\u0179" => "Z"

# Ż [LATIN CAPITAL LETTER Z WITH DOT ABOVE]

"\u017B" => "Z"

# Ž [LATIN CAPITAL LETTER Z WITH CARON]

"\u017D" => "Z"

# Ƶ [LATIN CAPITAL LETTER Z WITH STROKE]

"\u01B5" => "Z"

# Ȝ http://en.wikipedia.org/wiki/Yogh [LATIN CAPITAL LETTER YOGH]

"\u021C" => "Z"

# Ȥ [LATIN CAPITAL LETTER Z WITH HOOK]

"\u0224" => "Z"

# ᴢ [LATIN LETTER SMALL CAPITAL Z]

"\u1D22" => "Z"

# Ẑ [LATIN CAPITAL LETTER Z WITH CIRCUMFLEX]

"\u1E90" => "Z"

# Ẓ [LATIN CAPITAL LETTER Z WITH DOT BELOW]

"\u1E92" => "Z"

# Ẕ [LATIN CAPITAL LETTER Z WITH LINE BELOW]

"\u1E94" => "Z"

# Ⓩ [CIRCLED LATIN CAPITAL LETTER Z]

"\u24CF" => "Z"

# Ⱬ [LATIN CAPITAL LETTER Z WITH DESCENDER]

"\u2C6B" => "Z"

# Ꝣ [LATIN CAPITAL LETTER VISIGOTHIC Z]

"\uA762" => "Z"

# Ｚ [FULLWIDTH LATIN CAPITAL LETTER Z]

"\uFF3A" => "Z"

# ź [LATIN SMALL LETTER Z WITH ACUTE]

"\u017A" => "z"

# ż [LATIN SMALL LETTER Z WITH DOT ABOVE]

"\u017C" => "z"

# ž [LATIN SMALL LETTER Z WITH CARON]

"\u017E" => "z"

# ƶ [LATIN SMALL LETTER Z WITH STROKE]

"\u01B6" => "z"

# ȝ http://en.wikipedia.org/wiki/Yogh [LATIN SMALL LETTER YOGH]

"\u021D" => "z"

# ȥ [LATIN SMALL LETTER Z WITH HOOK]

"\u0225" => "z"

# ɀ [LATIN SMALL LETTER Z WITH SWASH TAIL]

"\u0240" => "z"

# ʐ [LATIN SMALL LETTER Z WITH RETROFLEX HOOK]

"\u0290" => "z"

# ʑ [LATIN SMALL LETTER Z WITH CURL]

"\u0291" => "z"

# ᵶ [LATIN SMALL LETTER Z WITH MIDDLE TILDE]

"\u1D76" => "z"

# ᶎ [LATIN SMALL LETTER Z WITH PALATAL HOOK]

"\u1D8E" => "z"

# ẑ [LATIN SMALL LETTER Z WITH CIRCUMFLEX]

"\u1E91" => "z"

# ẓ [LATIN SMALL LETTER Z WITH DOT BELOW]

"\u1E93" => "z"

# ẕ [LATIN SMALL LETTER Z WITH LINE BELOW]

"\u1E95" => "z"

# ⓩ [CIRCLED LATIN SMALL LETTER Z]

"\u24E9" => "z"

# ⱬ [LATIN SMALL LETTER Z WITH DESCENDER]

"\u2C6C" => "z"

# ꝣ [LATIN SMALL LETTER VISIGOTHIC Z]

"\uA763" => "z"

# ｚ [FULLWIDTH LATIN SMALL LETTER Z]

"\uFF5A" => "z"

# ⒵ [PARENTHESIZED LATIN SMALL LETTER Z]

"\u24B5" => "(z)"

# ⁰ [SUPERSCRIPT ZERO]

"\u2070" => "0"

# ₀ [SUBSCRIPT ZERO]

"\u2080" => "0"

# ⓪ [CIRCLED DIGIT ZERO]

"\u24EA" => "0"

# ⓿ [NEGATIVE CIRCLED DIGIT ZERO]

"\u24FF" => "0"

# ０ [FULLWIDTH DIGIT ZERO]

"\uFF10" => "0"

# ¹ [SUPERSCRIPT ONE]

"\u00B9" => "1"

# ₁ [SUBSCRIPT ONE]

"\u2081" => "1"

# ① [CIRCLED DIGIT ONE]

"\u2460" => "1"

# ⓵ [DOUBLE CIRCLED DIGIT ONE]

"\u24F5" => "1"

# ❶ [DINGBAT NEGATIVE CIRCLED DIGIT ONE]

"\u2776" => "1"

# ➀ [DINGBAT CIRCLED SANS-SERIF DIGIT ONE]

"\u2780" => "1"

# ➊ [DINGBAT NEGATIVE CIRCLED SANS-SERIF DIGIT ONE]

"\u278A" => "1"

# １ [FULLWIDTH DIGIT ONE]

"\uFF11" => "1"

# ⒈ [DIGIT ONE FULL STOP]

"\u2488" => "1."

# ⑴ [PARENTHESIZED DIGIT ONE]

"\u2474" => "(1)"

# ² [SUPERSCRIPT TWO]

"\u00B2" => "2"

# ₂ [SUBSCRIPT TWO]

"\u2082" => "2"

# ② [CIRCLED DIGIT TWO]

"\u2461" => "2"

# ⓶ [DOUBLE CIRCLED DIGIT TWO]

"\u24F6" => "2"

# ❷ [DINGBAT NEGATIVE CIRCLED DIGIT TWO]

"\u2777" => "2"

# ➁ [DINGBAT CIRCLED SANS-SERIF DIGIT TWO]

"\u2781" => "2"

# ➋ [DINGBAT NEGATIVE CIRCLED SANS-SERIF DIGIT TWO]

"\u278B" => "2"

# ２ [FULLWIDTH DIGIT TWO]

"\uFF12" => "2"

# ⒉ [DIGIT TWO FULL STOP]

"\u2489" => "2."

# ⑵ [PARENTHESIZED DIGIT TWO]

"\u2475" => "(2)"

# ³ [SUPERSCRIPT THREE]

"\u00B3" => "3"

# ₃ [SUBSCRIPT THREE]

"\u2083" => "3"

# ③ [CIRCLED DIGIT THREE]

"\u2462" => "3"

# ⓷ [DOUBLE CIRCLED DIGIT THREE]

"\u24F7" => "3"

# ❸ [DINGBAT NEGATIVE CIRCLED DIGIT THREE]

"\u2778" => "3"

# ➂ [DINGBAT CIRCLED SANS-SERIF DIGIT THREE]

"\u2782" => "3"

# ➌ [DINGBAT NEGATIVE CIRCLED SANS-SERIF DIGIT THREE]

"\u278C" => "3"

# ３ [FULLWIDTH DIGIT THREE]

"\uFF13" => "3"

# ⒊ [DIGIT THREE FULL STOP]

"\u248A" => "3."

# ⑶ [PARENTHESIZED DIGIT THREE]

"\u2476" => "(3)"

# ⁴ [SUPERSCRIPT FOUR]

"\u2074" => "4"

# ₄ [SUBSCRIPT FOUR]

"\u2084" => "4"

# ④ [CIRCLED DIGIT FOUR]

"\u2463" => "4"

# ⓸ [DOUBLE CIRCLED DIGIT FOUR]

"\u24F8" => "4"

# ❹ [DINGBAT NEGATIVE CIRCLED DIGIT FOUR]

"\u2779" => "4"

# ➃ [DINGBAT CIRCLED SANS-SERIF DIGIT FOUR]

"\u2783" => "4"

# ➍ [DINGBAT NEGATIVE CIRCLED SANS-SERIF DIGIT FOUR]

"\u278D" => "4"

# ４ [FULLWIDTH DIGIT FOUR]

"\uFF14" => "4"

# ⒋ [DIGIT FOUR FULL STOP]

"\u248B" => "4."

# ⑷ [PARENTHESIZED DIGIT FOUR]

"\u2477" => "(4)"

# ⁵ [SUPERSCRIPT FIVE]

"\u2075" => "5"

# ₅ [SUBSCRIPT FIVE]

"\u2085" => "5"

# ⑤ [CIRCLED DIGIT FIVE]

"\u2464" => "5"

# ⓹ [DOUBLE CIRCLED DIGIT FIVE]

"\u24F9" => "5"

# ❺ [DINGBAT NEGATIVE CIRCLED DIGIT FIVE]

"\u277A" => "5"

# ➄ [DINGBAT CIRCLED SANS-SERIF DIGIT FIVE]

"\u2784" => "5"

# ➎ [DINGBAT NEGATIVE CIRCLED SANS-SERIF DIGIT FIVE]

"\u278E" => "5"

# ５ [FULLWIDTH DIGIT FIVE]

"\uFF15" => "5"

# ⒌ [DIGIT FIVE FULL STOP]

"\u248C" => "5."

# ⑸ [PARENTHESIZED DIGIT FIVE]

"\u2478" => "(5)"

# ⁶ [SUPERSCRIPT SIX]

"\u2076" => "6"

# ₆ [SUBSCRIPT SIX]

"\u2086" => "6"

# ⑥ [CIRCLED DIGIT SIX]

"\u2465" => "6"

# ⓺ [DOUBLE CIRCLED DIGIT SIX]

"\u24FA" => "6"

# ❻ [DINGBAT NEGATIVE CIRCLED DIGIT SIX]

"\u277B" => "6"

# ➅ [DINGBAT CIRCLED SANS-SERIF DIGIT SIX]

"\u2785" => "6"

# ➏ [DINGBAT NEGATIVE CIRCLED SANS-SERIF DIGIT SIX]

"\u278F" => "6"

# ６ [FULLWIDTH DIGIT SIX]

"\uFF16" => "6"

# ⒍ [DIGIT SIX FULL STOP]

"\u248D" => "6."

# ⑹ [PARENTHESIZED DIGIT SIX]

"\u2479" => "(6)"

# ⁷ [SUPERSCRIPT SEVEN]

"\u2077" => "7"

# ₇ [SUBSCRIPT SEVEN]

"\u2087" => "7"

# ⑦ [CIRCLED DIGIT SEVEN]

"\u2466" => "7"

# ⓻ [DOUBLE CIRCLED DIGIT SEVEN]

"\u24FB" => "7"

# ❼ [DINGBAT NEGATIVE CIRCLED DIGIT SEVEN]

"\u277C" => "7"

# ➆ [DINGBAT CIRCLED SANS-SERIF DIGIT SEVEN]

"\u2786" => "7"

# ➐ [DINGBAT NEGATIVE CIRCLED SANS-SERIF DIGIT SEVEN]

"\u2790" => "7"

# ７ [FULLWIDTH DIGIT SEVEN]

"\uFF17" => "7"

# ⒎ [DIGIT SEVEN FULL STOP]

"\u248E" => "7."

# ⑺ [PARENTHESIZED DIGIT SEVEN]

"\u247A" => "(7)"

# ⁸ [SUPERSCRIPT EIGHT]

"\u2078" => "8"

# ₈ [SUBSCRIPT EIGHT]

"\u2088" => "8"

# ⑧ [CIRCLED DIGIT EIGHT]

"\u2467" => "8"

# ⓼ [DOUBLE CIRCLED DIGIT EIGHT]

"\u24FC" => "8"

# ❽ [DINGBAT NEGATIVE CIRCLED DIGIT EIGHT]

"\u277D" => "8"

# ➇ [DINGBAT CIRCLED SANS-SERIF DIGIT EIGHT]

"\u2787" => "8"

# ➑ [DINGBAT NEGATIVE CIRCLED SANS-SERIF DIGIT EIGHT]

"\u2791" => "8"

# ８ [FULLWIDTH DIGIT EIGHT]

"\uFF18" => "8"

# ⒏ [DIGIT EIGHT FULL STOP]

"\u248F" => "8."

# ⑻ [PARENTHESIZED DIGIT EIGHT]

"\u247B" => "(8)"

# ⁹ [SUPERSCRIPT NINE]

"\u2079" => "9"

# ₉ [SUBSCRIPT NINE]

"\u2089" => "9"

# ⑨ [CIRCLED DIGIT NINE]

"\u2468" => "9"

# ⓽ [DOUBLE CIRCLED DIGIT NINE]

"\u24FD" => "9"

# ❾ [DINGBAT NEGATIVE CIRCLED DIGIT NINE]

"\u277E" => "9"

# ➈ [DINGBAT CIRCLED SANS-SERIF DIGIT NINE]

"\u2788" => "9"

# ➒ [DINGBAT NEGATIVE CIRCLED SANS-SERIF DIGIT NINE]

"\u2792" => "9"

# ９ [FULLWIDTH DIGIT NINE]

"\uFF19" => "9"

# ⒐ [DIGIT NINE FULL STOP]

"\u2490" => "9."

# ⑼ [PARENTHESIZED DIGIT NINE]

"\u247C" => "(9)"

# ⑩ [CIRCLED NUMBER TEN]

"\u2469" => "10"

# ⓾ [DOUBLE CIRCLED NUMBER TEN]

"\u24FE" => "10"

# ❿ [DINGBAT NEGATIVE CIRCLED NUMBER TEN]

"\u277F" => "10"

# ➉ [DINGBAT CIRCLED SANS-SERIF NUMBER TEN]

"\u2789" => "10"

# ➓ [DINGBAT NEGATIVE CIRCLED SANS-SERIF NUMBER TEN]

"\u2793" => "10"

# ⒑ [NUMBER TEN FULL STOP]

"\u2491" => "10."

# ⑽ [PARENTHESIZED NUMBER TEN]

"\u247D" => "(10)"

# ⑪ [CIRCLED NUMBER ELEVEN]

"\u246A" => "11"

# ⓫ [NEGATIVE CIRCLED NUMBER ELEVEN]

"\u24EB" => "11"

# ⒒ [NUMBER ELEVEN FULL STOP]

"\u2492" => "11."

# ⑾ [PARENTHESIZED NUMBER ELEVEN]

"\u247E" => "(11)"

# ⑫ [CIRCLED NUMBER TWELVE]

"\u246B" => "12"

# ⓬ [NEGATIVE CIRCLED NUMBER TWELVE]

"\u24EC" => "12"

# ⒓ [NUMBER TWELVE FULL STOP]

"\u2493" => "12."

# ⑿ [PARENTHESIZED NUMBER TWELVE]

"\u247F" => "(12)"

# ⑬ [CIRCLED NUMBER THIRTEEN]

"\u246C" => "13"

# ⓭ [NEGATIVE CIRCLED NUMBER THIRTEEN]

"\u24ED" => "13"

# ⒔ [NUMBER THIRTEEN FULL STOP]

"\u2494" => "13."

# ⒀ [PARENTHESIZED NUMBER THIRTEEN]

"\u2480" => "(13)"

# ⑭ [CIRCLED NUMBER FOURTEEN]

"\u246D" => "14"

# ⓮ [NEGATIVE CIRCLED NUMBER FOURTEEN]

"\u24EE" => "14"

# ⒕ [NUMBER FOURTEEN FULL STOP]

"\u2495" => "14."

# ⒁ [PARENTHESIZED NUMBER FOURTEEN]

"\u2481" => "(14)"

# ⑮ [CIRCLED NUMBER FIFTEEN]

"\u246E" => "15"

# ⓯ [NEGATIVE CIRCLED NUMBER FIFTEEN]

"\u24EF" => "15"

# ⒖ [NUMBER FIFTEEN FULL STOP]

"\u2496" => "15."

# ⒂ [PARENTHESIZED NUMBER FIFTEEN]

"\u2482" => "(15)"

# ⑯ [CIRCLED NUMBER SIXTEEN]

"\u246F" => "16"

# ⓰ [NEGATIVE CIRCLED NUMBER SIXTEEN]

"\u24F0" => "16"

# ⒗ [NUMBER SIXTEEN FULL STOP]

"\u2497" => "16."

# ⒃ [PARENTHESIZED NUMBER SIXTEEN]

"\u2483" => "(16)"

# ⑰ [CIRCLED NUMBER SEVENTEEN]

"\u2470" => "17"

# ⓱ [NEGATIVE CIRCLED NUMBER SEVENTEEN]

"\u24F1" => "17"

# ⒘ [NUMBER SEVENTEEN FULL STOP]

"\u2498" => "17."

# ⒄ [PARENTHESIZED NUMBER SEVENTEEN]

"\u2484" => "(17)"

# ⑱ [CIRCLED NUMBER EIGHTEEN]

"\u2471" => "18"

# ⓲ [NEGATIVE CIRCLED NUMBER EIGHTEEN]

"\u24F2" => "18"

# ⒙ [NUMBER EIGHTEEN FULL STOP]

"\u2499" => "18."

# ⒅ [PARENTHESIZED NUMBER EIGHTEEN]

"\u2485" => "(18)"

# ⑲ [CIRCLED NUMBER NINETEEN]

"\u2472" => "19"

# ⓳ [NEGATIVE CIRCLED NUMBER NINETEEN]

"\u24F3" => "19"

# ⒚ [NUMBER NINETEEN FULL STOP]

"\u249A" => "19."

# ⒆ [PARENTHESIZED NUMBER NINETEEN]

"\u2486" => "(19)"

# ⑳ [CIRCLED NUMBER TWENTY]

"\u2473" => "20"

# ⓴ [NEGATIVE CIRCLED NUMBER TWENTY]

"\u24F4" => "20"

# ⒛ [NUMBER TWENTY FULL STOP]

"\u249B" => "20."

# ⒇ [PARENTHESIZED NUMBER TWENTY]

"\u2487" => "(20)"

# « [LEFT-POINTING DOUBLE ANGLE QUOTATION MARK]

"\u00AB" => "\""

# » [RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK]

"\u00BB" => "\""

# “ [LEFT DOUBLE QUOTATION MARK]

"\u201C" => "\""

# ” [RIGHT DOUBLE QUOTATION MARK]

"\u201D" => "\""

# „ [DOUBLE LOW-9 QUOTATION MARK]

"\u201E" => "\""

# ″ [DOUBLE PRIME]

"\u2033" => "\""

# ‶ [REVERSED DOUBLE PRIME]

"\u2036" => "\""

# ❝ [HEAVY DOUBLE TURNED COMMA QUOTATION MARK ORNAMENT]

"\u275D" => "\""

# ❞ [HEAVY DOUBLE COMMA QUOTATION MARK ORNAMENT]

"\u275E" => "\""

# ❮ [HEAVY LEFT-POINTING ANGLE QUOTATION MARK ORNAMENT]

"\u276E" => "\""

# ❯ [HEAVY RIGHT-POINTING ANGLE QUOTATION MARK ORNAMENT]

"\u276F" => "\""

# ＂ [FULLWIDTH QUOTATION MARK]

"\uFF02" => "\""

# ‘ [LEFT SINGLE QUOTATION MARK]

"\u2018" => "\'"

# ’ [RIGHT SINGLE QUOTATION MARK]

"\u2019" => "\'"

# ‚ [SINGLE LOW-9 QUOTATION MARK]

"\u201A" => "\'"

# ‛ [SINGLE HIGH-REVERSED-9 QUOTATION MARK]

"\u201B" => "\'"

# ′ [PRIME]

"\u2032" => "\'"

# ‵ [REVERSED PRIME]

"\u2035" => "\'"

# ‹ [SINGLE LEFT-POINTING ANGLE QUOTATION MARK]

"\u2039" => "\'"

# › [SINGLE RIGHT-POINTING ANGLE QUOTATION MARK]

"\u203A" => "\'"

# ❛ [HEAVY SINGLE TURNED COMMA QUOTATION MARK ORNAMENT]

"\u275B" => "\'"

# ❜ [HEAVY SINGLE COMMA QUOTATION MARK ORNAMENT]

"\u275C" => "\'"

# ＇ [FULLWIDTH APOSTROPHE]

"\uFF07" => "\'"

# ‐ [HYPHEN]

"\u2010" => "-"

# ‑ [NON-BREAKING HYPHEN]

"\u2011" => "-"

# ‒ [FIGURE DASH]

"\u2012" => "-"

# – [EN DASH]

"\u2013" => "-"

# — [EM DASH]

"\u2014" => "-"

# ⁻ [SUPERSCRIPT MINUS]

"\u207B" => "-"

# ₋ [SUBSCRIPT MINUS]

"\u208B" => "-"

# － [FULLWIDTH HYPHEN-MINUS]

"\uFF0D" => "-"

# ⁅ [LEFT SQUARE BRACKET WITH QUILL]

"\u2045" => "["

# ❲ [LIGHT LEFT TORTOISE SHELL BRACKET ORNAMENT]

"\u2772" => "["

# ［ [FULLWIDTH LEFT SQUARE BRACKET]

"\uFF3B" => "["

# ⁆ [RIGHT SQUARE BRACKET WITH QUILL]

"\u2046" => "]"

# ❳ [LIGHT RIGHT TORTOISE SHELL BRACKET ORNAMENT]

"\u2773" => "]"

# ］ [FULLWIDTH RIGHT SQUARE BRACKET]

"\uFF3D" => "]"

# ⁽ [SUPERSCRIPT LEFT PARENTHESIS]

"\u207D" => "("

# ₍ [SUBSCRIPT LEFT PARENTHESIS]

"\u208D" => "("

# ❨ [MEDIUM LEFT PARENTHESIS ORNAMENT]

"\u2768" => "("

# ❪ [MEDIUM FLATTENED LEFT PARENTHESIS ORNAMENT]

"\u276A" => "("

# （ [FULLWIDTH LEFT PARENTHESIS]

"\uFF08" => "("

# ⸨ [LEFT DOUBLE PARENTHESIS]

"\u2E28" => "(("

# ⁾ [SUPERSCRIPT RIGHT PARENTHESIS]

"\u207E" => ")"

# ₎ [SUBSCRIPT RIGHT PARENTHESIS]

"\u208E" => ")"

# ❩ [MEDIUM RIGHT PARENTHESIS ORNAMENT]

"\u2769" => ")"

# ❫ [MEDIUM FLATTENED RIGHT PARENTHESIS ORNAMENT]

"\u276B" => ")"

# ） [FULLWIDTH RIGHT PARENTHESIS]

"\uFF09" => ")"

# ⸩ [RIGHT DOUBLE PARENTHESIS]

"\u2E29" => "))"

# ❬ [MEDIUM LEFT-POINTING ANGLE BRACKET ORNAMENT]

"\u276C" => "<"

# ❰ [HEAVY LEFT-POINTING ANGLE BRACKET ORNAMENT]

"\u2770" => "<"

# ＜ [FULLWIDTH LESS-THAN SIGN]

"\uFF1C" => "<"

# ❭ [MEDIUM RIGHT-POINTING ANGLE BRACKET ORNAMENT]

"\u276D" => ">"

# ❱ [HEAVY RIGHT-POINTING ANGLE BRACKET ORNAMENT]

"\u2771" => ">"

# ＞ [FULLWIDTH GREATER-THAN SIGN]

"\uFF1E" => ">"

# ❴ [MEDIUM LEFT CURLY BRACKET ORNAMENT]

"\u2774" => "{"

# ｛ [FULLWIDTH LEFT CURLY BRACKET]

"\uFF5B" => "{"

# ❵ [MEDIUM RIGHT CURLY BRACKET ORNAMENT]

"\u2775" => "}"

# ｝ [FULLWIDTH RIGHT CURLY BRACKET]

"\uFF5D" => "}"

# ⁺ [SUPERSCRIPT PLUS SIGN]

"\u207A" => "+"

# ₊ [SUBSCRIPT PLUS SIGN]

"\u208A" => "+"

# ＋ [FULLWIDTH PLUS SIGN]

"\uFF0B" => "+"

# ⁼ [SUPERSCRIPT EQUALS SIGN]

"\u207C" => "="

# ₌ [SUBSCRIPT EQUALS SIGN]

"\u208C" => "="

# ＝ [FULLWIDTH EQUALS SIGN]

"\uFF1D" => "="

# ！ [FULLWIDTH EXCLAMATION MARK]

"\uFF01" => "!"

# ‼ [DOUBLE EXCLAMATION MARK]

"\u203C" => "!!"

# ⁉ [EXCLAMATION QUESTION MARK]

"\u2049" => "!?"

# ＃ [FULLWIDTH NUMBER SIGN]

"\uFF03" => "#"

# ＄ [FULLWIDTH DOLLAR SIGN]

"\uFF04" => "$"

# ⁒ [COMMERCIAL MINUS SIGN]

"\u2052" => "%"

# ％ [FULLWIDTH PERCENT SIGN]

"\uFF05" => "%"

# ＆ [FULLWIDTH AMPERSAND]

"\uFF06" => "&"

# ⁎ [LOW ASTERISK]

"\u204E" => "\*"

# ＊ [FULLWIDTH ASTERISK]

"\uFF0A" => "\*"

# ， [FULLWIDTH COMMA]

"\uFF0C" => ","

# ． [FULLWIDTH FULL STOP]

"\uFF0E" => "."

# ⁄ [FRACTION SLASH]

"\u2044" => "/"

# ／ [FULLWIDTH SOLIDUS]

"\uFF0F" => "/"

# ： [FULLWIDTH COLON]

"\uFF1A" => ":"

# ⁏ [REVERSED SEMICOLON]

"\u204F" => ";"

# ； [FULLWIDTH SEMICOLON]

"\uFF1B" => ";"

# ？ [FULLWIDTH QUESTION MARK]

"\uFF1F" => "?"

# ⁇ [DOUBLE QUESTION MARK]

"\u2047" => "??"

# ⁈ [QUESTION EXCLAMATION MARK]

"\u2048" => "?!"

# ＠ [FULLWIDTH COMMERCIAL AT]

"\uFF20" => "@"

# ＼ [FULLWIDTH REVERSE SOLIDUS]

"\uFF3C" => "\\"

# ‸ [CARET]

"\u2038" => "^"

# ＾ [FULLWIDTH CIRCUMFLEX ACCENT]

"\uFF3E" => "^"

# ＿ [FULLWIDTH LOW LINE]

"\uFF3F" => "\_"

# ⁓ [SWUNG DASH]

"\u2053" => "~"

# ～ [FULLWIDTH TILDE]

"\uFF5E" => "~"

################################################################

# Below is the Perl script used to generate the above mappings #

# from ASCIIFoldingFilter.java: #

################################################################

#

# #!/usr/bin/perl

#

# use warnings;

# use strict;

#

# my @source\_chars = ();

# my @source\_char\_descriptions = ();

# my $target = '';

#

# while (<>) {

# if (/case\s+'(\\u[A-F0-9]+)':\s\*\/\/\s\*(.\*)/i) {

# push @source\_chars, $1;

# push @source\_char\_descriptions, $2;

# next;

# }

# if (/output\[[^\]]+\]\s\*=\s\*'(\\'|\\\\|.)'/) {

# $target .= $1;

# next;

# }

# if (/break;/) {

# $target = "\\\"" if ($target eq '"');

# for my $source\_char\_num (0..$#source\_chars) {

# print "# $source\_char\_descriptions[$source\_char\_num]\n";

# print "\"$source\_chars[$source\_char\_num]\" => \"$target\"\n\n";

# }

# @source\_chars = ();

# @source\_char\_descriptions = ();

# $target = '';

# }

# }

|  |  |  |  |
| --- | --- | --- | --- |
| sse_logo_corpblue_rgb |  |  |  |
| **Denmark**  Søren Frichs Vej 39  8000 Aarhus C, DK  Tel.: +45 8943 2000  more.info@systematic.com | **Australia**  Tower A, Level 5,7 London Circuit  Canberra ACT 2600, AU  Tel.: +61 (0)2 6169 4088  more.info.au@systematic.com | **New Zealand**  15 Level, 171 Featherston Street  Wellington 6011, NZ  Tel.: +64 04 894 8571  more.info.nz@systematic.com | **UAE**  World Trade Centre, Level 17, Suite 56  Abu Dhabi, UAE  Tel.: +971 2 654 4675  Fax: +45 8943 2020  more.info.uae@systematic.com |
| Landgreven 3, 2.sal  1301 Copenhagen K, DK | **Finland**  Finlaysoninkuja 19  33210 Tampere, FI  Tel.: +358 207 463 870  more.info.fi@systematic.com | **Singapore**  15 Hoe Chiang Road  #12-02 Tower Fifteen  Singapore 089316  Tel.: +65 6653 7492  more.info.sg@systematic.com | **United Kingdom**  Meadow Gate, Farnborough Airport Farnborough, Hampshire GU14 6XA, UK  Tel.: +44 1276 675533  more.info.uk@systematic.com |
|  | **France**  5 Place de la Bastille  75004 Paris, FR  Tel.: +45 8943 2000  (HQ in Denmark)  more.info.fr@systematic.com | **Sweden**  Ostermalmstorg 1, 4th Floor  Stockholm 114 42, SE  Tel.: +46 770 770109  more.info.se@systematic.com | **United States of America**  5875 Trinity Parkway, Suite 130  Centreville, Virginia 20120-1971, USA  Tel.: +1 703 385 7522  more.info.us@systematic.com |
|  | **Germany**  Im Zollhafen 24  50678 Köln, DE  Tel.: +49 221 650 783 71  more.info.de@systematic.com |  |  |
|  |  |  |  |