

File created: 22-Oct-2025 23:28:42 {WMEDLEY}<library>UNICODE-TABLES.;4

edit by: rmk

changes to: (VARS UNICODE-TABLESCOMS)

previous date: 16-Oct-2025 16:47:54 {WMEDLEY}<library>UNICODE-TABLES.;3

Read Table: INTERLISP

Package: INTERLISP

Format: MCCS

(RPAQQ UNICODE-TABLESCOMS

;; Read Unicode mapping tables. A separate file before UNICODE in the loadup, because the tables must be loaded while UTF8TOMCODE
;; and MCODETOUTF8 are still equivalenced to EVQ. This file has to come before UNICODE in the loadup sequence.

(COMS ; Read Unicode mapping files

(INITVARS (UNICODEDIRECTORIES NIL))
(GLOBALVARS UNICODEDIRECTORIES)
(VARS XCCS-CHARSETS)
(FNS READ-UNICODE-MAPPING-FILENAMES READ-UNICODE-MAPPING))

(COMS ; Make translation tables for UTF external formats

(FNS MAKE-UNICODE-TRANSLATION-TABLES XCCSTOMCCS-MAPPING MERGE-UNICODE-TRANSLATION-TABLES
UNICODE.UNMAPPED UNICODE-EXTEND-TRANSLATION?)
(FNS ALL-UNICODE-MAPPINGS XCCSJAPANESECHARSETS)
(INITVARS (*MCCSTOUNICODE*)
(*UNICODETOMCCS*)
(*MCCS-LOADED-CHARSETS*)
(*UNICODE-LOADED-CHARSETS*)
(*LARGEUNICODES*))

[DECLARE%: DONTEVAL@LOAD DOCOPY (P (MAKE-UNICODE-TRANSLATION-TABLES 'ALL]
(DECLARE%: EVAL@COMPILE DONTCOPY (FILES (LOADCOMP
UNICODE-EXPORTS))

;; Read Unicode mapping tables. A separate file before UNICODE in the loadup, because the tables must be loaded while UTF8TOMCODE and
;; MCODETOUTF8 are still equivalenced to EVQ. This file has to come before UNICODE in the loadup sequence.

;; Read Unicode mapping files

(RPAQ? UNICODEDIRECTORIES NIL)

(DECLARE%: DOEVAL@COMPILE DONTCOPY

(GLOBALVARS UNICODEDIRECTORIES)
)

(RPAQQ XCCS-CHARSETS

((LATIN "0")
(JAPANESE-SYMBOLS1 "41")
(JAPANESE-SYMBOLS2 "42")
(EXTENDED-LATIN "43")
(HIRAGANA "44")
(KATAKANA "45")
(GREEK "46")
(CYRILLIC "47")
(FORMS "50")
(RUNIC-GOTHIC "51")
(MORE-CYRILLIC "52")
(UNKNOWN1 "56")
(UNKNOWN2 "57")
(JIS "60-166")
(ARABIC "340")
(HEBREW "341")
(IPA "342")
(HANGUL "343")
(GEORGIAN-ARMENIAN "344")
(DEVANAGRI "345")
(BENGALI "346")
(GURMUKHI "347")
(THAI-LAO "350")
(SYMBOLS3 "353")
(EXTENDED-ITC-DINGBATS "354")
(ITC-DINGBATS1 "355")
(SYMBOLS2 "356")
(SYMBOLS1 "357")
(LIGATURES "360")
(ACCENTED-LATIN1 "361")
(ACCENTED-LATIN2 "362")
(ACCENTED-GREEK1 "363")
(ACCENTED-GREEK2 "364")
(MORE-ARABIC "365")
(GRAPHIC-VARIANTS "375")
(DEFAULT LATIN ACCENTED-LATIN1 EXTENDED-LATIN SYMBOLS1 SYMBOLS2 FORMS JAPANESE-SYMBOLS1
JAPANESE-SYMBOLS2)
(JAPANESE HIRAGANA KATAKANA JIS)))

(DEFINEQ

(READ-UNICODE-MAPPING-FILENAMES
[LAMBDA (FILESPEC)

; Edited 16-Oct-2025 16:43 by rmk
 ; Edited 4-Sep-2025 00:11 by rmk
 ; Edited 27-Jan-2025 16:46 by rmk
 ; Edited 21-Jan-2025 22:51 by rmk
 ; Edited 19-Jan-2025 12:21 by rmk
 ; Edited 3-Feb-2024 11:00 by rmk
 ; Edited 30-Jan-2024 08:45 by rmk
 ; Edited 26-Jan-2024 14:02 by mth
 ; Edited 5-Aug-2020 15:59 by kaplan
 ; Edited 4-Aug-2020 17:31 by rmk:

;; FILESPEC can be a file name, character-set name, the name of a collection of character sets, an XCCS character code, or a list of those. Maps
 ;; those into the names of files that contain the indicated Unicode mappings.

(CL:REMOVE-DUPPLICATES [for F X CSI inside (if (EQ FILESPEC 'ALL)

then
 ;; Perhaps should figure out which files in the directories and subdirectories are
 ;; relevant?

(for N in XCCS-CHARSETS collect (CAR N))
 else FILESPEC)

join ;; Last case hopes to pick up all the tables that are grouped together in a subdirectory (e.g. if F is JIS)

(OR (CL:WHEN (CHARCODEP F) ; An XCCS code can retrieve its character set
 (for D FN (FOCTAL ← (OCTALSTRING (LRSH F 8))) inside

UNICODEDIRECTORIES
 when (SETQ FN (FILDIR (PACKFILENAME 'DIRECTORY D 'BODY (CONCAT
 'XCCS-
 FOCTAL
 '=*)

'EXTENSION
 'TXT
 'VERSION "")))

do (RETURN FN)))

(MKLIST (FINDFILE (PACKFILENAME 'BODY F 'EXTENSION 'TXT 'VERSION ""))
 T UNICODEDIRECTORIES))

(for D inside UNICODEDIRECTORIES when [SETQ \$\$VAL
 (OR (FILDIR (PACKFILENAME 'NAME
 (CONCAT
 "XCCS-*=

" F)

'EXTENSION
 'TXT
 'VERSION ""
 'BODY D))
 (FILDIR (PACKFILENAME 'NAME
 (CONCAT "XCCS-"
 F "=*")
 'EXTENSION
 'TXT
 'VERSION ""
 'BODY D]

do (RETURN \$\$VAL))
 (AND (SETQ CSI (ASSOC F XCCS-CHARSETS))
 (READ-UNICODE-MAPPING-FILENAMES (CDR CSI)))
 (for D inside UNICODEDIRECTORIES when (DIRECTORYNAMEP (SETQ D
 (CONCAT D ">" F
 ">"))))

join (FILDIR (CONCAT D ">*.TXT;"]

:TEST
 (FUNCTION STRING.EQUAL])

(READ-UNICODE-MAPPING
[LAMBDA (FILESPEC PRINT NOERROR)

; Edited 16-Oct-2025 11:25 by rmk
 ; Edited 11-Oct-2025 12:08 by rmk
 ; Edited 4-Sep-2025 00:17 by rmk
 ; Edited 24-Apr-2025 15:32 by rmk
 ; Edited 31-Jan-2025 17:43 by rmk
 ; Edited 17-Jan-2025 16:41 by rmk

; Edited 3-Feb-2024 00:21 by rmk
 ; Edited 5-Jan-2024 12:26 by rmk
 ; Edited 3-Jul-2021 13:37 by rmk:

;; Combines the char-mapping tables from FILES coded in the Unicode-CDROM format. Comments prefixed by # and

;; Column 1: XCCS input hex code in the format 0XXXXX
 ;; Column 2: Corresponding Unicode code-sequence in the format
 ;; 0XXXXX ... 0YYYYY
 ;; Column 3: (after #) Character name in some mapping files, utf-8 character
 ;; for XCCS mapping files
 ;;

(RESETLST

(for FILE STREAM [SEPBITTABLE ← (MAKEBITTABLE (CHARCODE (TAB SPACE) in (READ-UNICODE-MAPPING-FILENAMES
 FILESPEC)

join ;; External format :THROUGH means read as bytes, so the Unicode UTF-8 comments cannot cause reading problems.

[RESETSAVE [SETQ STREAM (OPENSTREAM FILE 'INPUT NIL '((FORMAT :THROUGH)
 (EOLCONVENTION LF)
 '(PROGN (CLOSEF? OLDVALUE]
 (bind LINE NAME CHARSET START first (CL:UNLESS (FILEPOS "Name:" STREAM NIL NIL NIL T)
 (ERROR "NOT A UNICODE MAPPING FILE" (FULLNAME
 STREAM)))
 (SETQ NAME (CL:STRING-TRIM " " (CL:READ-LINE STREAM NIL
 NIL)))
 (SETQ CHARSET (CL:IF (FILEPOS "XCCS charset:" STREAM
 NIL NIL NIL T)
 (CL:STRING-TRIM " " (CL:READ-LINE
 STREAM
 NIL NIL))
 ""))
 (CL:WHEN PRINT
 ; Strip off XCCS in front of name
 (PRINTOUT T T CHARSET " " [SUBSTRING
 NAME
 (CONSTANT (ADD1 (NCHARS
 "XCCS

"]

T))
 while (SETQ LINE (CL:READ-LINE STREAM NIL NIL)) when (SETQ START (STRPOSL SEPBITTABLE LINE
 1 T))
 unless (EQ (CHARCODE %#)
 (NTHCHARCODE LINE START))
 collect [bind END CODES while [SETQ END (OR (STRPOSL SEPBITTABLE LINE START)
 (ADD1 (NCHARS LINE)
 collect [CHARCODE.DECODE (SUBSTRING LINE START (SUB1 END)
 (CONSTANT (CONCAT]
 repeatwhile (AND (SETQ START (STRPOSL SEPBITTABLE LINE END T))
 (NEQ (CHARCODE %#)
 (NTHCHARCODE LINE START)))
 finally (CL:WHEN (CDDR \$\$VAL) ; Combiners go into a CADR list
 (RPLACD \$\$VAL (CONS (CDR \$\$VAL))))]
 finally (CLOSEF? STREAM))))]

)

;; Make translation tables for UTF external formats

(DEFINEQ

(MAKE-UNICODE-TRANSLATION-TABLES
 [LAMBDA (MAPPING REINSTALL)

; Edited 11-Oct-2025 11:54 by rmk
 ; Edited 4-Sep-2025 00:30 by rmk
 ; Edited 24-Apr-2025 15:47 by rmk
 ; Edited 31-Jan-2025 17:46 by rmk
 ; Edited 26-Jan-2025 19:36 by rmk

; Edited 22-Jan-2025 14:22 by rmk
 ; Edited 19-Jan-2025 15:08 by rmk
 ; Edited 18-Jan-2025 11:52 by rmk
 ; Edited 3-Feb-2024 00:24 by rmk
 ; Edited 30-Jan-2024 09:54 by rmk
 ; Edited 21-Aug-2021 13:12 by rmk:

;; MAPPING is the list of numeric code correspondence pairs constructed by applying READ-UNICODE-MAPPING to XCCS-to-Unicode mapping
 ;; files. This applies the XCCS-to-MCCS translations, and then updates or produces two recoding arrays, one maps left-side codes into right-side
 ;; codes (e.g. MCCS or ISO8859-1 to Unicode), for printing, the other maps right-side (Unicode) codes to corresponding right-side codes (e.g.
 ;; MCCS).
 ; Edited 17-Aug-2020 08:46 by rmk:

```
(CL:UNLESS [AND (LISTP MAPPING)
  (FOR PAIR R IN MAPPING AS I TO 10 ALWAYS (AND (LISTP PAIR)
    (CHARCODEP (CAR PAIR))
    [FIXP (SETQ R (CAR (MKLIST (CADR PAIR]
    (CHARCODEP (IABS R])
```

;; Seems like the argument is not already a list of mapping pairs (perhaps with a combiner), presumably a list of charsets to be read.

```
(SETQ MAPPING (READ-UNICODE-MAPPING MAPPING)))
(SETQ MAPPING (XCCSTOMCCS-MAPPING MAPPING))
```

;; This updates or produces two recoding arrays, one maps left-side codes into right-side codes (e.g. XCCS or ISO8859-1 to Unicode), for printing,
 ;; the other maps right-side (Unicode) codes to corresponding right-side codes (e.g. XCCS).

;;

;; If REINSTALL is T, the new mapping vectors replace the current maps in the *XCCSTOUNICODE* and *UNICODETOXCCS* global variables.
 ;; Otherwise we create new tables (mostly for comparison and debugging).

;;

```
(if REINSTALL
  then (SETQ *MCCS-LOADED-CHARSETS* (SETQ *UNICODE-LOADED-CHARSETS* NIL))
    (SETQ *NEXT-PRIVATE-MCCSCODE* FIRST-PRIVATE-MCCSCODE)
    (SETQ *NEXT-PRIVATE-UNICODE* FIRST-PRIVATE-UNICODE)
    (LET [(TABLE (HASHARRAY (LENGTH MAPPING)))
      (INVERSETABLE (HASHARRAY (LENGTH MAPPING)
        (MERGE-UNICODE-TRANSLATION-TABLES NIL MAPPING TABLE INVERSETABLE)
        (SETQ *MCCSTOUNICODE* TABLE)
        (SETQ *UNICODETOXCCS* INVERSETABLE)
        (LIST *MCCSTOUNICODE* *UNICODETOXCCS*))
      else (CL:UNLESS (BOUNDP *NEXT-PRIVATE-MCCSCODE*)
        (SETQ *NEXT-PRIVATE-MCCSCODE* FIRST-PRIVATE-MCCSCODE)
        (SETQ *NEXT-PRIVATE-UNICODE* FIRST-PRIVATE-UNICODE)
        (MERGE-UNICODE-TRANSLATION-TABLES NIL MAPPING])
```

```
(XCCSTOMCCS-MAPPING
[LAMBDA (XTOMAPPING)
```

; Edited 11-Oct-2025 12:57 by rmk

;; This translates the pairs that map XCCS to Unicode into pairs that translate MCCS to Unicode.

;; We grab the affected pairs before we make any changes so that we don't get into ordering issues.

```
(LET* ([XTOMCODES (CHARCODE ((Currency Dollar)
  (Dollar Currency)
  (Uparrow Circumflex)
  (Circumflex Uparrow)
  (Leftarrow Lowline)
  (Lowline Leftarrow]
  (AFFECTED (for MP in XTOMAPPING when (thereis XP in XTOMCODES suchthat (EQ (CAR MP)
    (CAR XP))))
    collect MP)))
  (for AP in AFFECTED do (RPLACA AP (CADR (ASSOC (CAR AP)
    XTOMCODES))))
  finally (push XTOMAPPING (CHARCODE (DEL DEL)))
  (RETURN XTOMAPPING))
```

```
(MERGE-UNICODE-TRANSLATION-TABLES
[LAMBDA (INVERSE MAPPING TABLE INVERSETABLE)
```

; Edited 11-Oct-2025 10:24 by rmk
 ; Edited 24-Apr-2025 15:28 by rmk
 ; Edited 1-Feb-2025 21:42 by rmk
 ; Edited 26-Jan-2025 12:58 by rmk
 ; Edited 22-Jan-2025 08:20 by rmk

; Edited 19-Jan-2025 15:58 by rmk
 ; Edited 18-Jan-2025 11:49 by rmk
 ; Edited 27-Mar-2024 12:10 by rmk
 ; Edited 3-Feb-2024 12:46 by rmk
 ; Edited 31-Jan-2024 10:06 by rmk

;; MAPPINGS is a list of pairs that map domain codes to range codes. TABLE and INVERSETABLE default to *MCCSTOUNICODE*
 ;; *UNICODETOMCCS* respectively.

(CL:UNLESS TABLE

[SETQ TABLE (OR *MCCSTOUNICODE* (SETQ *MCCSTOUNICODE* (HASHARRAY (LENGTH MAPPING))

(CL:UNLESS INVERSETABLE

[SETQ INVERSETABLE (OR *UNICODETOMCCS* (SETQ *UNICODETOMCCS* (HASHARRAY (LENGTH MAPPING))
(for M D R OLDR in MAPPING first (CL:IF INVERSE (swap TABLE INVERSETABLE)) everytime (SETQ D (CAR M))
(SETQ R (CADR M))

;; We don't do combinators, but we are
 ;; allowing non-SMALLP's

unless (OR (LISTP D)
(LISTP R))

do ;; The (CONS R OLDR) deals with alternatives: (U X1) (U X2) => (U (X1 X2)), lowest code first. Those are only possible in the U-to-X
 ;; direction when the tables contain (X1 U) and (X2 U). There are no duplicates/alternative table entries in the X-to-U direction.

(SETQ OLDR (GETHASH D TABLE))

(CL:UNLESS (MEMB R OLDR)

(PUTHASH D (SORT (CONS R OLDR))
TABLE))

(swap D R)

(SETQ OLDR (GETHASH D INVERSETABLE))

(CL:UNLESS (MEMB R OLDR)

(PUTHASH D (SORT (CONS R OLDR))
INVERSETABLE)))

(LIST TABLE INVERSETABLE)]

(UNICODE.UNMAPPED

[LAMBDA (CODE TABLE DONTFAKE)

; Edited 24-Apr-2025 15:48 by rmk
 ; Edited 22-Jan-2025 08:19 by rmk
 ; Edited 19-Jan-2025 22:02 by rmk
 ; Edited 18-Jan-2025 12:02 by rmk
 ; Edited 2-Feb-2024 23:52 by rmk
 ; Edited 31-Jan-2024 10:07 by rmk
 ; Edited 11-Aug-2020 20:23 by rmk:

;; This is the slow fall-out when UNICODE.TRANSLATE determines that CODE has no fast mapping in TRANSLATION-TABLE.

;;

;; If we have not already installed the mapping segment for that code, we try to retrieve it from the numeric file. If that segment mapping doesn't
 ;; exist or doesn't have an entry for CODE, we fake up a mapping with a negative range in both directions. One way or the other, there will be an
 ;; entry for that segment in both mapping vectors.

;;

(PROG ((INVERSE (EQ TABLE *UNICODETOMCCS*))
RANGE HASH)

;; If we already looked up CODE's character set in a file, then we have already filled in its information in the translation table. If it didn't have a code
 ;; for a particular character, then we fake it here. Faked codes are negative, so we can detect them easily, and interpret them with IABS.

(CL:WHEN (AND (UNICODE-EXTEND-TRANSLATION? CODE TABLE)
(SETQ RANGE (GETHASH CODE TABLE)))

;; We might have gotten the segment that didn't have an entry for CODE.

(RETURN RANGE))

;;

(CL:UNLESS DONTFAKE

;; Our attempt at extending the known tables did not provide a mapping for CODE. So we fake it up with the next unused private
 ;; code in the code space.

;; The number of possible faked mappings is determined by the number of private-use Unicodes, since the MCCS character space is
 ;; pretty sparse. The codes don't have to come from the same part of the code space, and the NEXTCODEs are saved in global
 ;; variables. The last available codes are constants.

(CL:WHEN (IEQP *NEXT-PRIVATE-MCCSCODE* LAST-PRIVATE-MCCSCODE)

; Same number of available codes both ways

```

      (ERROR "EXHAUSTED RANGE FOR UNMAPPED CODES"))
    (if INVERSE
      then (SETQ RANGE *NEXT-PRIVATE-MCCSCODE*)
            (add *NEXT-PRIVATE-MCCSCODE* 1)
      else (SETQ RANGE *NEXT-PRIVATE-UNICODE*)
            (add *NEXT-PRIVATE-UNICODE* 1))
    (MERGE-UNICODE-TRANSLATION-TABLES INVERSE (CONS (LIST CODE RANGE)))
    ;; CONS because of LIST convention so we can eventually distinguish combiners.
    (RETURN (CONS RANGE))))

```

(UNICODE-EXTEND-TRANSLATION?
[LAMBDA (CODE TABLE)

```

; Edited 11-Oct-2025 09:49 by rmk
; Edited 4-Sep-2025 00:34 by rmk
; Edited 29-Jun-2025 16:44 by rmk
; Edited 24-Apr-2025 15:49 by rmk
; Edited 26-Jan-2025 11:26 by rmk
; Edited 21-Jan-2025 22:31 by rmk
; Edited 18-Jan-2025 12:40 by rmk
; Edited 13-Jan-2025 23:50 by rmk
; Edited 26-Aug-2024 16:49 by rmk
; Edited 27-Mar-2024 23:02 by rmk
; Edited 5-Feb-2024 13:48 by rmk
; Edited 3-Feb-2024 12:40 by rmk

```

;; There is currently no mapping for CODE in TABLE, hopefully just because the relevant character-set mapping has not been installed. We infer
;; from TABLE whether CODE is an MCCS or UNICODE code and look for the proper mapping table (forward or inverted) for its character set.

;; We record which character sets we have already expanded so we don't do them again.

```

(LET ((CHARSET (CHARSET CODE))
      (INVERSE (EQ TABLE *UNICODETO-MCCS*)
                (EQ TABLE *MCCS-TO-UNICODE*)))
  (MAPPING FILE)

```

;; If we already looked for CHARSET in the file and found anything, it has already been merged. Otherwise, it would just fail again

```

(CL:UNLESS (MEMB CHARSET (CL:IF INVERSE
                                *UNICODE-LOADED-CHARSETS*
                                *MCCS-LOADED-CHARSETS*)))

```

;; Don't try this charset again.

```

(CL:IF INVERSE
  (push *UNICODE-LOADED-CHARSETS* CHARSET)
  (push *MCCS-LOADED-CHARSETS* CHARSET))
(SETQ FILE (FINDFILE (CL:IF INVERSE
                            'UNICODE-TO-MCCS-MAPPINGS
                            'MCCS-TO-UNICODE-MAPPINGS)
                     T UNICODEDIRECTORIES))

```

;; The mappings files are indexed by CHARSET.

```

(CL:WHEN [AND FILE (SETQ MAPPING (CL:WITH-OPEN-FILE (STREAM FILE :INPUT)
                                                    (CL:WHEN (FILEPOS (CONCAT "[" CHARSET " ")
                                                                (STREAM NIL NIL NIL T)
                                                                (READ STREAM))

```

;; Merge MAPPING into both tables, respecting the direction indicated by TABLE.

```

(MERGE-UNICODE-TRANSLATION-TABLES INVERSE MAPPING)
T)))

```

)

(DEFINEQ

(ALL-UNICODE-MAPPINGS
[LAMBDA (INVERTED FILE)

```

; Edited 24-Apr-2025 15:51 by rmk
; Edited 31-Jan-2025 17:46 by rmk
; Edited 26-Jan-2025 13:40 by rmk
; Edited 22-Jan-2025 14:07 by rmk
; Edited 19-Jan-2025 12:20 by rmk
; Edited 17-Jan-2025 22:32 by rmk
; Edited 15-Jan-2025 09:49 by rmk
; Edited 27-Mar-2024 14:48 by rmk

```

; Edited 5-Feb-2024 13:14 by rmk

; Edited 3-Feb-2024 09:16 by rmk

;; Reads all the XCCS-to-UNICODE mapping files that we know about, and produces a 2-level index that maps between MCCS codes and
 ;; UNICODE codes, depending on INVERTED.

;; The first index level segments all the domain codes according to their character sets. The segments are sorted by character set, the pairs within
 ;; each segment are sorted by their domain codes.

;; E.g. if INVERTED=NIL and given a XCCS code, the lookup for the corresponding Unicode(s) is

;; (CADR (ASSOC MCCSCODE (\CHARSET MCCSCODE) INDEX))).

;; If FILE is not NIL, the result is written to a file. If FILE is T, the file is either MCCS-TO-UNICODE-MAPPINGS.TXT or
 ;; UNICODE-TO-MCCS-MAPPINGS.TXT, depending on INVERTED.

(LET (INDEX)

(for PAIR DOMAIN RANGE CHARSET in (READ-UNICODE-MAPPING 'ALL) eachtime (SETQ DOMAIN (CAR PAIR))
 (SETQ RANGE (CADR PAIR))

;; (LISTP RANGE) is a combiner, ignored
 ;; for now.

unless (LISTP RANGE) do (CL:WHEN INVERTED (SWAP DOMAIN RANGE))

;; One segment for each high-byte character set. This aligns with UNICODE-EXTEND.TRANSLATION?

[SETQ CHARSET (OR (ASSOC (\CHARSET DOMAIN)
 INDEX)

(CAR (push INDEX (CONS (\CHARSET DOMAIN]

;; For alternative mappings (in the U-to-M direction) we end up with (D R1 R2 ...). (CADR is the first (and
 ;; almost always) the only one.

(pushnew [CDR (OR (ASSOC DOMAIN (CDR CHARSET))
 (CAR (push (CDR CHARSET)
 (CONS DOMAIN]

RANGE))

;; Push the charset mappings down an extra CONS, so that a subsequent READ will get them all after a FILEPOS search for super-paren [

[for CS in INDEX do (for M in (CDR CS) when (CDDR M) do ; Sort the range alternatives, if any

(change (CDR M)
 (SORT DATUM)))

;; Sort by domain codes and push down a level

(change (CDR CS)
 (CONS (SORT DATUM T]

(SETQ INDEX (SORT INDEX T)) ; Sort character sets
 (if FILE

then (SETQ FILE (PACKFILENAME 'BODY (if (NEQ FILE T)
 then FILE
 elseif INVERTED
 then 'UNICODE-TO-MCCS-MAPPINGS
 else 'MCCS-TO-UNICODE-MAPPINGS)

'DIRECTORY
 (CAR (MKLIST UNICODEDIRECTORIES))
 'EXTENSION
 'TXT))

(CL:WITH-OPEN-FILE (STREAM FILE :DIRECTION :OUTPUT :IF-EXISTS :NEW-VERSION)

;; We can FILEPOS for "[nnn " then READ for each segment. Or just READFILE to get them all.

(for I in INDEX do (PRINTOUT STREAM "[" (CAR I)

" " "
 (CADR I)
 "]" T T))

(PRINTOUT STREAM "STOP" T)
 (FULLNAME STREAM))

else INDEX])

(XCCSJAPANESECHARSETS
 [LAMBDA (OCTAL FILE)

; Edited 11-Jun-2025 23:00 by rmk

;; Returns the list of numbers for the Japanese character sets.

(for F POS CS in (READ-UNICODE-MAPPING-FILENAMES "JIS") when (SETQ POS (STRPOS "XCCS-" F 1 NIL NIL T))
 collect [SETQ CS (SUBSTRING F POS (SUB1 (STRPOS '=JIS F POS]

```
(CL:IF OCTAL
  CS
  (MKATOM (CONCAT CS "Q")))
finally (SORT $$VAL)
(CL:WHEN FILE
  (RETURN (CL:WITH-OPEN-FILE (STREAM (PACKFILENAME 'BODY (CL:IF (EQ FILE T)
    "JAPANESECHARSETS"
    FILE)
    'DIRECTORY
    (CAR (MKLIST UNICODEDIRECTORIES))
    'EXTENSION
    'TXT)
    :DIRECTION :OUTPUT :IF-EXISTS :NEW-VERSION)
    (PRINT $$VAL STREAM)
    (FULLNAME STREAM))))))
)

(RPAQ? *MCCSTOUNICODE* )

(RPAQ? *UNICODETOMCCS* )

(RPAQ? *MCCS-LOADED-CHARSETS* )

(RPAQ? *UNICODE-LOADED-CHARSETS* )

(RPAQ? *LARGEUNICODES* )

(DECLARE%: DONTEVAL@LOAD DOCOPY

(MAKE-UNICODE-TRANSLATION-TABLES 'ALL)
)

(DECLARE%: EVAL@COMPILE DONTCOPY

(FILESLoad (LOADCOMP
  UNICODE-EXPORTS)
)
)
```

FUNCTION INDEX

ALL-UNICODE-MAPPINGS	6	UNICODE-EXTEND-TRANSLATION?	6
MAKE-UNICODE-TRANSLATION-TABLES	3	UNICODE.UNMAPPED	5
MERGE-UNICODE-TRANSLATION-TABLES	4	XCCSJAPANESECHARSETS	7
READ-UNICODE-MAPPING	2	XCCSTOMCCS-MAPPING	4
READ-UNICODE-MAPPING-FILENAMES	2		

VARIABLE INDEX

LARGEUNICODES	8	*UNICODE-LOADED-CHARSETS* ..	8	XCCS-CHARSETS	1
MCCS-LOADED-CHARSETS	8	*UNICODETOMCCS*	8		
MCCSTOUNICODE	8	UNICODEDIRECTORIES	1		
