

File created: 26-Mar-2024 11:00:37 {WMEDLEY}<sources>XCCS.;70

edit by: rmk

changes to: (FNS \XCCSFORMATBYTESTREAM)

previous date: 19-Mar-2024 16:02:36 {WMEDLEY}<sources>XCCS.;68

Read Table: INTERLISP

Package: INTERLISP

Format: XCCS

(RPAQQ **XCCSCOMS**

```
[ (FNS \XCCSINCCODE \XCCSPEEKCCODE \XCCSOUTCHAR \XCCSBACKCCODE \XCCSFORMATBYTESTREAM \XCCSCHARSETFN)
  (FNS \CREATE.XCCS.EXTERNALFORMAT)
  (FNS \NSIN.24BITENCODING.ERROR)
  (INITVARS (*SIGNAL-24BIT-NSENCODING-ERROR*))
  (DECLARE%: EVAL@COMPILE DONTCOPY (EXPORT (CONSTANTS (\NORUNCODE 255)
                                                    (NSCHARSETSHIFT 255))
                                           (ALISTS (CHARACTERNAMES \NORUNCODE NSCHARSETSHIFT))
                                           (MACROS \RUNCODED)))
    (DECLARE%: DONTEVAL@LOAD DOCOPY (P (\CREATE.XCCS.EXTERNALFORMAT])
```

(DEFINEQ

(\XCCSINCCODE

[LAMBDA (STREAM COUNTP)

; Edited 8-Dec-2023 15:28 by rmk

; Edited 6-Aug-2021 15:57 by rmk:

;;; Returns a 16 bit character code. SHIFTEDCSET is STREAM's char set left shifted 8.

;;; If COUNTP is non-NIL, the variable *BYTECOUNTER* is set freely to the number of bytes read.

;;; This doesn't do EOL conversion, \INCHAR does that

```
(DECLARE (USEDFREE *BYTECOUNTER*))
(\DTEST STREAM 'STREAM)
(LET (NUMBYTES (CSET (ffetch (STREAM CHARSET) of STREAM))
      (CHAR (\BIN STREAM)))
```

; Error on EOF unless ENDOFSTREAMOP does something
; else.

;; NUMBYTES tracks the number of \BINS.

```
(IF (EQ CHAR NSCHARSETSHIFT)
    THEN
```

; Shifting character sets, toss CHAR

```
(SETQ CSET (\BIN STREAM))
(IF (NEQ NSCHARSETSHIFT CSET)
    THEN
```

; Shift to new runcode CSET: SH CS CH

```
(SETQ CHAR (\BIN STREAM))
(SETQ NUMBYTES 3)
(freplace (STREAM CHARSET) of STREAM with CSET)
```

```
ELSEIF (EQ 0 (\BIN STREAM))
    THEN
```

; SH SH CSH CS CH where CSH is 0

;; The high-order character set byte must be 0, because we don't support obese characters (24 bit)

```
(SETQ CSET (\BIN STREAM))
(SETQ CHAR (\BIN STREAM))
(SETQ NUMBYTES 5)
```

; To align with below

```
(freplace (STREAM CHARSET) of STREAM with \NORUNCODE)
ELSE (\NSIN.24BITENCODING.ERROR STREAM)
```

;; The stream now knows the new character set, runcoded or not.

```
ELSEIF (EQ CSET \NORUNCODE)
    THEN
```

; 2-bytes

```
(SETQ CSET CHAR)
(SETQ CHAR (\BIN STREAM))
(SETQ NUMBYTES 2)
```

```
ELSE ;; Runcoded CSET and CHAR
```

```
(SETQ NUMBYTES 1))
```

```
(CL:WHEN COUNTP (SETQ *BYTECOUNTER* NUMBYTES))
```

```
(CL:WHEN CHAR
```

; Typically NIL if ENDOFSTREAMOP returned NIL at EOF

```
(LOGOR (UNFOLD CSET 256)
        CHAR))])
```

(\XCCSPEEKCCODE

[LAMBDA (STREAM NOERROR)

; Edited 8-Dec-2023 15:32 by rmk

; Edited 21-Jun-2021 23:44 by rmk:

;; Modeled on \XCCSINCCODE, but peeks at the last byte in the sequence, leaves the stream unchanged

```
(\DTEST STREAM 'STREAM)
(LET ((CSET (ffetch (STREAM CHARSET) of STREAM))
      (CHAR (\PEEKBIN STREAM NOERROR)))
```

;; Returns a 16 bit character code. Doesn't do EOL conversion--\PEEKCCODE does that.

;; We don't change the charset in the stream, put the file ptr back the way it was.

```
(CL:WHEN CHAR
  (IF (EQ CHAR NSCHARSETSHIFT)
    THEN (\BIN STREAM) ; Read the peeked shifting byte
         (SETQ CSET (\BIN STREAM)) ; Consume the char shift byte
    (IF (NEQ CSET NSCHARSETSHIFT)
      THEN ;; Shift to new runcode CSET: SH CS CH. We have to BIN what we peeked, BIN, and peek again
           (SETQ CHAR (\PEEKBIN STREAM NOERROR))
           (\BACKFILEPTR STREAM)
           (\BACKFILEPTR STREAM)
      ELSEIF (EQ 0 (\BIN STREAM)) ; SH SH CSH CS CH where CSH is 0
            THEN ; Note: no eof error check on this \BIN -- an eof in the middle of a charset shift is an error
                 (SETQ CSET (\BIN STREAM))
                 (SETQ CHAR (\PEEKBIN STREAM NOERROR))
                 (\BACKFILEPTR STREAM)
                 (\BACKFILEPTR STREAM)
                 (\BACKFILEPTR STREAM)
                 (\BACKFILEPTR STREAM)
            ELSE (\INSIN.24BITENCODING.ERROR STREAM)
          ELSEIF (EQ CSET \NORUNCODE) ; 2 byte runs, BIN/PEEK/BACK
            THEN
              (SETQ CSET CHAR)
              (\BIN STREAM)
              (SETQ CHAR (\PEEKBIN STREAM NOERROR)) ; One BACKFILEPTR seems OK
              (\BACKFILEPTR STREAM)
            ;; No need to back up for the runcoded case
            (CL:WHEN CHAR
              (LOGOR (UNFOLD CSET 256)
                CHAR))))))
```

(\XCCSOUTCHAR

[LAMBDA (STREAM CHARCODE) ; Edited 13-Aug-2021 10:24 by rmk:

;; Closed function for the :XCCS external format, also called when :XCCS is the default

```
(COND
  ((EQ CHARCODE (CHARCODE EOL))
   (FREPLACE (STREAM CHARPOSITION) OF STREAM WITH 0)
   [COND
     [(NOT (\RUNCODED STREAM)) ; Charset is a constant 0, we put out the high-order byte.
      (\BOUT STREAM (\CHARSET (CHARCODE EOL))
        ((EQ (\CHARSET (CHARCODE EOL))
              (ffetch (STREAM CHARSET) of STREAM)))
      (T ; We are runcoded, and not in character set 0, have to shift.
        (\BOUT STREAM NSCHARSETSHIFT)
        (\BOUT STREAM (freplace (STREAM CHARSET) of STREAM with (\CHARSET (CHARCODE EOL))
          ;; We are now in the right charset (0) for the first EOL byte. For CRLF, the CR is immediately followed by the LF byte, without the prefix 0
          ;; byte even if not runcoded, i.e. the 2 bytes are thought of as a composite. The stream is left in CSET0 (the freplace above), read for another
          ;; shift according to the next shift in a runcoded file.
          (\BOUTEOL STREAM))
        (T (CHANGE (FFETCH (STREAM CHARPOSITION) OF STREAM)
                  (IPLUS16 1 DATUM))
          (COND
            ((NOT (\RUNCODED STREAM))
             (\BOUT STREAM (\CHARSET CHARCODE))
             (\BOUT STREAM (\CHAR8CODE CHARCODE)))
            (EQ (\CHARSET CHARCODE)
              (ffetch (STREAM CHARSET) of STREAM))
            (\BOUT STREAM (\CHAR8CODE CHARCODE)))
            (T (\BOUT STREAM NSCHARSETSHIFT)
              (\BOUT STREAM (freplace (STREAM CHARSET) of STREAM with (\CHARSET CHARCODE)))
              (\BOUT STREAM (\CHAR8CODE CHARCODE))
```

(\XCCSBACKCCODE

[LAMBDA (STREAM COUNTP) ; Edited 8-Dec-2023 15:34 by rmk
; Edited 19-Jul-2022 17:12 by rmk
; Edited 13-Aug-2021 14:08 by rmk:

```
(DECLARE (USEDFREE *BYTECOUNTER*))
(LET ((BYTE (AND (\BACKFILEPTR STREAM)
                 (\PEEKBIN STREAM)))
      (CSET (fetch (STREAM CHARSET) of STREAM)))
  (CL:WHEN BYTE
    ;; The immediately preceding byte must be a character byte. If it is a byte in a runcode, then we are done, even if the byte before is part
    ;; of a shift sequence.
    ;; But if we are currently in a nonruncoded file, we have to go back one more to get the character set byte.
    ;; If we can't back up, we are already at the beginning.
    (IF (EQ \NORUNCODE CSET)
      THEN (IF (\BACKFILEPTR STREAM)
```

```

THEN (CL:WHEN COUNTP (SETQ *BYTECOUNTER* -2))
      (LOGOR (UNFOLD (\PEEKBIN STREAM
                    256)
              BYTE)
            ELSE (CL:WHEN COUNTP (SETQ *BYTECOUNTER* -1))
              NIL)
ELSE (CL:WHEN COUNTP (SETQ *BYTECOUNTER* -1))
      (LOGOR (UNFOLD CSET 256)
            BYTE)))])

```

(\XCCSFORMATBYTESTREAM

```

[LAMBDA (STREAM BYTESTREAM) ; Edited 26-Mar-2024 11:00 by rmk
; Edited 19-Mar-2024 16:02 by rmk

```

```

(\EXTERNALFORMAT BYTESTREAM (\EXTERNALFORMAT STREAM))
;; This stream may be read as a continuation of STREAM (TTYIN, LAFITE?), and we want to make sure that the bytes are encoded properly. So
;; let's assert (and possibly mark) that that's its current situation.
(XCCSCHARSETFN BYTESTREAM (fetch (STREAM CHARSET) of STREAM)
BYTESTREAM])

```

(\XCCSCHARSETFN

```

[LAMBDA (STREAM CHARSET DONTMARKSTREAM) ; Edited 9-Dec-2023 11:18 by rmk

```

```

;; This differs from \GENERIC.CHARSET in that it actually writes the shifting bytes into an output stream, unless DONTMARKSTREAM. It will do
;; write the shifts, even if it just replicates the situation that is already there (presumably CHARSET = the old CHARSET). The client should test
;; and avoid calling if useless shifts are not desired.

```

```

(LET [(CSET (fetch (STREAM CHARSET) of (\DTEST STREAM 'STREAM)
                  (CL:WHEN CHARSET
                    (CL:WHEN (EQ CHARSET T)
                              (SETQ CHARSET \NORUNCODE))
                    (CL:UNLESS (EQ CHARSET CSET)
                              (replace (STREAM CHARSET) of STREAM with CHARSET)
                              (CL:UNLESS DONTMARKSTREAM
                                        (CL:WHEN (\IOMODEP STREAM 'OUTPUT T)
                                                  (\BOUT STREAM NSCHARSETSHIFT)
                                                  (if (EQ CHARSET \NORUNCODE)
                                                      then (\BOUT STREAM \NORUNCODE)
                                                      else (\BOUT STREAM 0)
                                                  )))
                              CSET]))

```

(DEFINEQ

(\CREATE.XCCS.EXTERNALFORMAT

```

[LAMBDA (NAME EOL) ; Edited 7-Dec-2023 23:03 by rmk
; Edited 30-Jun-2022 18:08 by rmk
; Edited 10-Sep-2021 19:49 by rmk:

```

::: Create the :XCCS external format. Stream's EOL overrides the (vacuous) default here

```

(MAKE-EXTERNALFORMAT (OR NAME :XCCS)
  (FUNCTION \XCCSINCCODE)
  (FUNCTION \XCCSPEEKCCODE)
  (FUNCTION \XCCSBACKCCODE)
  (FUNCTION \XCCSOUTCHAR)
  (FUNCTION \XCCSFORMATBYTESTREAM)
  (OR EOL 'LF)
  T NIL NIL (FUNCTION \XCCSCHARSETFN))

```

(DEFINEQ

(\NSIN.24BITENCODING.ERROR

```

[LAMBDA (STREAM) (* bvm%: "12-Mar-86 15:35")
  (DECLARE (USEDFREE *SIGNAL-24BIT-NSENCODING-ERROR*))

```

::: Called if we see the sequence shift,shift on STREAM -- means shift to 24-bit character set, which we don't support. Usually this just means we're erroneously reading a binary file as text. If this function returns, its value is taken as a character set to shift to

```

(COND
  (*SIGNAL-24BIT-NSENCODING-ERROR* ; Only cause error if user/reader cares
  (ERROR "24-bit NS encoding not supported" STREAM))
  ; Return charset zero
0])

```

(RPAQ? *SIGNAL-24BIT-NSENCODING-ERROR*)

(DECLARE%: EVAL@COMPILE DONTCOPY

::: FOLLOWING DEFINITIONS EXPORTED

(DECLARE%: EVAL@COMPILE

(RPAQQ \NORUNCODE 255)

(RPAQQ NSCHARSETSHIFT 255)

(CONSTANTS (\NORUNCODE 255)
(NSCHARSETSHIFT 255))
)

(ADDTOVAR CHARACTER NAMES (\NORUNCODE 255)
(NSCHARSETSHIFT 255))

(DECLARE%: EVAL@COMPILE

(PUTPROPS \RUNCODED MACRO (OPENLAMBDA (STREAM)

;; returns NIL is the stream is not runcoded, that is, if the stream has 16 bit bytes explicitly represented
; note that neq is ok since charsets are known to be SMALLP's

(NEQ (fetch CHARSET of STREAM)
\NORUNCODE))
)

:: END EXPORTED DEFINITIONS

(DECLARE%: DONTVAL@LOAD DOCOPY

(\CREATE.XCCS.EXTERNALFORMAT)

)

FUNCTION INDEX

\CREATE.XCCS.EXTERNALFORMAT	3	\XCCSCHARSETFN	3	\XCCSOUTCHAR	2
\NSIN.24BITENCODING.ERROR	3	\XCCSFORMATBYTESTREAM	3	\XCCSPEEKCCODE	1
\XCCSBACKCCODE	2	\XCCSINCCODE	1		

VARIABLE INDEX

SIGNAL-24BIT-NSENCODING-ERROR . . .	3	CHARACTERNAMES	4
---------------------------------------	---	--------------------------	---

CONSTANT INDEX

NSCHARSETSHIFT	4	\NORUNCODE	4
--------------------------	---	----------------------	---

MACRO INDEX

\RUNCODED	4
---------------------	---
