The Wayback Machine - https://web.archive.org/web/20210608063633/http://www.firthworks.com/roger/glulx/chars.html

Roger Firth's IF pages

Inform character sets

If the print strings in your game include only symbols from the ASCII (or Basic Latin) character set -- values in the range 32..126 or $20..$7E -- then you probably don't need to read this. The information here is all about how to address the Extended Latin characters: mostly European accented letters with a few miscellaneous signs throw in for good measure. If you wish to display such characters, you have four options. Some carry disadvantages, which may affect which one you choose:

1. Insert them literally -- for example **cause célèbre** -- using whatever means your text editor provides for extending the keyboard repertoire.
   Disadvantage: Your source file isn't portable between IBM PCs and Apple Macs. The word **célèbre** created on a PC might appear as **cÈlËbre** if that file was copied to a Mac, because the two physical machines use different character encoding.

2. Use the Inform compiler's ability to embed internal character codes -- for example **cause c@@170l@@182bre** -- by typing **@@** followed by a decimal number.
   Disadvantage: Your source file isn't portable between Z-machine and Glulx. The word **c@@170l@@182bre** compiles to **célèbre** on the Z-machine, but to **cªl¶bre** when the target is Glulx, because the two virtual machines use different character encoding. (But remember: character codes in the range @@32 to @@126 are the same in Glulx and the Z-machine. Thus, @@64 is "@", @@92 is "\", @@94 is "^" and @@126 is "~" in both environments.)

3. Use the compiler's ability to embed escape sequences -- for example **cause c@'el@`ebre** -- by typing **@** followed by a character pair.

4. Use the compiler's ability to embed UniCode values -- for example **cause c@{E9}l@{E8}bre** -- by typing **@{..}** around a hexadecimal number.

So, from this wealth of possibilities, we can deduce a couple of guidelines:

- If you're not bothered about PC/Mac cross-platform compatibility -- because you work on a single platform and you've no intention of sharing your source code with anybody else -- use Option 1 because it's easiest to read.
- If you're publishing source code, you really *have* to make it cross-platform, so that means choosing one of the other options. I recommend Option 3, since an escape sequence like **@'e** is easier to recognise as **é** then either of the number-based options, and it also makes your code portable between Z-machine and Glulx.

Also, you might take a look at Wouter ten Brink's cross-VM mapchar.zip from the Archive.

These are the available Extended Latin characters.

Character	ZSCII charset	Glulx charset	Escape	UniCode
À (A grave)	@@186	@@192	@`A	@{C0}
Á (A acute)	@@175	@@193	@'A	@{C1}
Â (A circumflex)	@@196	@@194	@^A	@{C2}
Ã (A tilde)	@@208	@@195	@~A	@{C3}
Ä (A diaeresis)	@@158	@@196	@:A	@{C4}
Å (A ring)	@@202	@@197	@oA	@{C5}
à (a grave)	@@181	@@224	@`a	@{E0}
á (a acute)	@@169	@@225	@'a	@{E1}
â (a circumflex)	@@191	@@226	@^a	@{E2}
ã (a tilde)	@@205	@@227	@~a	@{E3}
ä (a diaeresis)	@@155	@@228	@:a	@{E4}
å (a ring)	@@201	@@229	@oa	@{E5}
Æ (AE ligature)	@@212	@@198	@AE	@{C6}
æ (ae ligature)	@@211	@@230	@ae	@{E6}
Ç (C cedilla)	@@214	@@199	@cC	@{C7}
ç (c cedilla)	@@213	@@231	@cc	@{E7}
È (E grave)	@@187	@@200	@`E	@{C8}
É (E acute)	@@176	@@201	@'E	@{C9}
Ê (E circumflex)	@@197	@@202	@^E	@{CA}
Ë (E diaeresis)	@@167	@@203	@:E	@{CB}
è (e grave)	@@182	@@232	@`e	@{E8}
é (e acute)	@@170	@@233	@'e	@{E9}
ê (e circumflex)	@@192	@@234	@^e	@{EA}
ë (e diaeresis)	@@164	@@235	@:e	@{EB}
Ì (I grave)	@@188	@@204	@`I	@{CC}
Í (I acute)	@@177	@@205	@'I	@{CD}
Î (I circumflex)	@@198	@@206	@^I	@{CE}
Ï (I diaeresis)	@@168	@@207	@:I	@{CF}
ì (i grave)	@@183	@@236	@`i	@{EC}
í (i acute)	@@171	@@237	@'i	@{ED}
î (i circumflex)	@@193	@@238	@^i	@{EE}
ï (i diaeresis)	@@165	@@239	@:i	@{EF}
Ñ (N tilde)	@@209	@@209	@~N	@{D1}
ñ (n tilde)	@@206	@@241	@~n	@{F1}

Ò (O grave)	@@189	@@210	@`O	@{D2}
Ó (O acute)	@@178	@@211	@'O	@{D3}
Ô (O circumflex)	@@199	@@212	@^O	@{D4}
Õ (O tilde)	@@210	@@213	@~O	@{D5}
Ö (O diaeresis)	@@159	@@214	@:O	@{D6}
Ø (O slash)	@@204	@@216	@/O	@{D8}
ò (o grave)	@@184	@@242	@`o	@{F2}
ó (o acute)	@@172	@@243	@'o	@{F3}
ô (o circumflex)	@@194	@@244	@^o	@{F4}
õ (o tilde)	@@207	@@245	@~o	@{F5}
ö (o diaeresis)	@@156	@@246	@:o	@{F6}
ø (o slash)	@@203	@@248	@/o	@{F8}
OE (OE ligature)	@@221	-	@OE	-
oe (oe ligature)	@@220	-	@oe	-
Ù (U grave)	@@190	@@217	@`U	@{D9}
Ú (U acute)	@@179	@@218	@'U	@{DA}
Û (U circumflex)	@@200	@@219	@^U	@{DB}
Ü (U diaeresis)	@@160	@@220	@:U	@{DC}
ù (u grave)	@@185	@@249	@`u	@{F9}
ú (u acute)	@@173	@@250	@'u	@{FA}
û (u circumflex)	@@195	@@251	@^u	@{FB}
ü (u diaeresis)	@@157	@@252	@:u	@{FC}
Ý (Y acute)	@@180	@@221	@'Y	@{DD}
ý (y acute)	@@174	@@253	@'y	@{FD}
ÿ (y diaeresis)	@@166	@@255	@:y	@{FF}
Ð (Eth)	@@218	@@208	@Et	@{D0}
ð (eth)	@@216	@@240	@et	@{F0}
Þ (Thorn)	@@217	@@222	@Th	@{DE}
þ (thorn)	@@215	@@254	@th	@{FE}
ß (ss)	@@161	@@223	@ss	@{DF}
¡ (inverted !)	@@222	@@161	@!!	@{A1}
¿ (inverted ?)	@@223	@@191	@??	@{BF}
« (left guillemet)	@@163	@@171	@<<	@{AB}
» (right guillemet)	@@162	@@187	@>>	@{BB}
¢ (cent)	-	@@162	-	@{A2}
£ (pound)	@@219	@@163	@LL	@{A3}
¤ (currency)	-	@@164	-	@{A4}
¥ (yen)	-	@@165	-	@{A5}
§ (section)	-	@@167	-	@{A7}
¶ (paragraph/pilcrow)	-	@@182	-	@{B6}
© (copyright)	-	@@169	-	@{A9}
® (registered)	-	@@174	-	@{AE}
¦ (broken bar)	-	@@166	-	@{A6}
- (soft hyphen)	-	@@173	-	@{AD}
¬ (not)	-	@@172	-	@{AC}
¯ (macron/overline)	-	@@175	-	@{AF}
° (degree)	-	@@176	-	@{B0}
µ (micro/mu)	-	@@181	-	@{B5}
· (middle dot)	-	@@183	-	@{B7}
± (plus-or-minus)	-	@@177	-	@{B1}
× (multiplication)	-	@@215	-	@{D7}
÷ (division)	-	@@247	-	@{F7}
¼ (one-quarter)	-	@@188	-	@{BC}
½ (one-half)	-	@@189	-	@{BD}
¾ (three-quarters)	-	@@190	-	@{BE}
ª (feminine ordinal)	-	@@170	-	@{AA}
º (masculine ordinal)	-	@@186	-	@{BA}
¹ (superscript 1)	-	@@185	-	@{B9}
² (superscript 2)	-	@@178	-	@{B2}
³ (superscript 3)	-	@@179	-	@{B3}
¨ (diaeresis)	-	@@168	-	@{A8}
´ (acute)	-	@@180	-	@{B4}
¸ (cedilla)	-	@@184	-	@{B8}