Representing Middle English on the Web with UTF-8

St. Erkenwald Manuscript, lines 257-264:

Arial Unicode MS		CODE2000	
yᵉ bisshop bythes hȳ ȝet wᵗ bale at his hert	257	yᵉ bisshop bythes hȳ ȝet wᵗ bale at his hert	257
yaḡħ mē menskid hī so how hit myȝt worthe	258 U+0304	yaḡħ mē menskid hī so how hit myȝt worthe	258 U+0304
yaḡħ mē menskid hī so how hit myȝt worthe	258 U+0305	yaḡħ mē menskid hī so how hit myȝt worthe	258 U+0305
yᵗ his clothes wer so clene in cloutes me thynkes	259	yᵗ his clothes wer so clene in cloutes me thynkes	259
hom burde haue rotid & bene rent ī ratt long sythen	260	hom burde haue rotid & bene rent ī ratt long sythen	260
yⁱ body may be enbawmyd hit bashis me noght	261	yⁱ body may be enbawmyd hit bashis me noght	261
yᵗ hit thar ryne ne route ne no ronke wormes	262	yᵗ hit thar ryne ne route ne no ronke wormes	262
bot yⁱ colour ne yⁱ clothe I know ī no wise	263	bot yⁱ colour ne yⁱ clothe I know ī no wise	263
how hit myȝt lye by mōnes lor & last so longe	264	how hit myȝt lye by mōnes lor & last so longe	264

A Middle English alliterative poem written about 1390 by an unknown author; manuscript copy dated 1477, British Library MS Harley 2250. From J.A. Burrow and Thorlac Turville-Petre, *A Book of Middle English*, Blackwell Publishers, Oxford (1992).

This page is encoded in UTF-8 with HTML markup. The manuscript includes liberal use of overlining, mostly to denote vowels followed by "m" or "n", represented here by U+0304 Combining Macron (the intention of the bar over "gh" in line 258 is unclear). If your browser does not handle combining sequences, the macron appears right the overlined letter as a dotted circle with a bar above.

Underlining (represented here by markup) is used by the copyist to identify material that is questionable and/or glossed in the margins. Also note the struck-out letter "u" of "route" in line 262 ("`<strike>u</strike>`"), indicating a correction by the copyist.

The letter "ȝ" (yogh) represents "y" at the beginning of word or before a stressed vowel, "gh" at the end of word or before another consonant, and "w" between vowels.

The letter "y" is used for both "y" and "þ" (thorn, modern "th") and "u" is written for "v". No punctuation is used.

Superscripts (represented here by markup) are sometimes used to denote abbreviation (wᵗ = "with") and other times in common short words such as "yᵉ" or "yⁱ" (alternative spellings of "þe" = "the") or "yᵗ" ("that").

Although markup should be used for superscript letters, a couple of them (such as "i" and "n") are encoded directly in Unicode for round-trip compatibility with other character sets. Thus "yⁱ" (`y<sup>i</sup>`) can also be encoded as "y" followed by U+2071 Superscript Small Letter "i": "yⁱ" (Unicode 3.1 and later).

For reference, here are the special letters of Middle English (not all of which are used in the sample above), together with their unicode values:

Name	Capital	Small	Description
Ash	U+00C6 Æ	U+00E6 æ	As in modern English "hat"
Thorn	U+00DE Þ	U+00FE þ	Modern "th" (survives in Icelandic)
Eth	U+00D0 Ð	U+00F0 ð	Modern "th" (survives in Icelandic)
Yogh	U+021C Ȝ	U+021D ȝ	Y, w, gh

Wynn	U+01F7	ƿ	U+01BF	ƿ	W (also called Wen)

Tools Used To Make This Page: The Kermit 95 2.0 terminal emulator to a Unix host with the GNU EMACS text editor, version 21.2. In EMACS I select UTF-8 as my file, keyboard, and terminal coding system. In Kermit, I choose UTF-8 as my terminal character set and then enter any non-ASCII values that are not directly accessible on my keyboard by their 4-digit hexadecimal values in the Alt-N dialog (press Alt-N, enter 4 four hex digits), as illustrated HERE. To view obscure characters such as Yogh and Wynn in Kermit's terminal emulation screen, I use a well-populated monospace font such as Everson Mono Terminal or Agfa/Monotype Andale Mono WT J.

[UTF-8 Sampler] [Kermit Home] [Unicode Fonts]

UTF-8 Sampler / The Kermit Project / Columbia University / kermit@columbia.edu / 11 August 2002