

Text with mixed encodings

いろいろなエンコーディングで $\text{T}_{\text{E}}\text{X}$ ソースを書いて、`lambda` でタイプセットをする 実験です。まず `utf8` から始めます。

明朝体 亜啞娃阿哀愛挨始逢葵茜穉惡握渥旭 葦芦鰻梓压幹扱宛姐虻飴絢綾鮎
或栗 裕安庵按暗案闇鞍杏以伊位依偉囿夷

ゴシック体 亜啞娃阿哀愛挨始逢葵茜穉惡握渥旭 葦芦鰻梓压幹扱宛姐虻飴絢
綾鮎或栗 裕安庵按暗案闇鞍杏以伊位依偉囿夷

これはラムダ(`lambda`)でタイプセットしたものです。改行や、禁則処理は `inunijapan.ocp` で実現しています。`inunijapan.ocp` は Unicode \rightarrow Unicode の translation であり、(1)日本語の改行、(2)禁則処理、(3)行末をスペースに解釈しないこと、を実現しています。予め `\usepackage{unijapan}` によって `unijapan` パッケージを読み込んでおく必要があります。いろいろな encoding から Unicode への translation には 別の専用の `ocp` を使えば、どのような場合にも改行、禁則処理などが可能になります。 $\text{W}_{32}\text{T}_{\text{E}}\text{X}$ の Ω は、ファイルの先頭の `0xff, 0xfe` (little endian UCS2), `0xfe, 0xff` (big endian UCS2), `0xef, 0xbb, 0xbf` (`utf8`) を読み飛ばすようにしてありますので、Windows application である、メモ帳や `wordpad` で保存したものを、そのままコンパイルすることができます。また、`lambdaj` でサポートしている `\UGT`, `\GT`, `\UMS` コマンドは、同じように使用できます。例えば、`森\UMS{9DD7}` 外 とすると、森鷗外 が得られます。

次に JIS コードです。

明朝体 亜啞娃阿哀愛挨始逢葵茜穉惡握渥旭 葦芦鯪梓压幹扱宛姐虻飴絢綾鮎
或栗 裕安庵按暗案闇鞍杏以伊位依偉囿夷

ゴシック体 亜啞娃阿哀愛挨始逢葵茜穉惡握渥旭 葦芦鯪梓压幹扱宛姐虻飴絢
綾鮎或栗 裕安庵按暗案闇鞍杏以伊位依偉囿夷

これはラムダ(lambda)でタイプセットしたものです。改行や、禁則処理は inunijapan.ocp で実現しています。inunijapan.ocp は Unicode → Unicode の translation であり、(1)日本語の改行、(2)禁則処理、(3)行末をスペースに解釈しないこと、を実現しています。予め \usepackage{unijapan} によって unijapan パッケージを読み込んでおく必要があります。いろいろな encoding から Unicode への translation には別の専用の ocp を使えば、どのような場合にも改行、禁則処理などが可能になります。W32T_EX の Omega は、ファイルの先頭の 0xff, 0xfe (little endian UCS2), 0xfe, 0xff (big endian UCS2), 0xef, 0xbb, 0xbf (utf8) を読み飛ばすようにしてありますので、Windows application である、メモ帳や wordpad で保存したものを、そのままコンパイルすることができます。また、lambdaj でサポートしている \UGT, \GT, \UMS コマンドは、同じように使用できます。例えば、森\UMS{9DD7}外 とすると、森鷗外 が得られます。

次に EUC-JAPAN をテストします。

明朝体 亜啞娃阿哀愛挨始逢葵茜穉惡握渥旭 葦芦鯪梓压幹扱宛姐虻飴絢綾鮎
或栗 裕安庵按暗案闇鞍杏以伊位依偉囿夷

ゴシック体 亜啞娃阿哀愛挨始逢葵茜穉惡握渥旭 葦芦鯪梓压幹扱宛姐虻飴絢
綾鮎或栗 裕安庵按暗案闇鞍杏以伊位依偉囿夷

これはラムダ(lambda)でタイプセットしたものです。改行や、禁則処理は inunijapan.ocp で実現しています。inunijapan.ocp は Unicode → Unicode の translation であり、(1)日本語の改行、(2)禁則処理、(3)行末をスペースに解釈しないこと、を実現しています。予め \usepackage{unijapan} によって unijapan パッケージを読み込んでおく必要があります。いろいろな encoding から Unicode への translation には 別の専用の ocp を使えば、どのような場合にも改行、禁則処理などが可能になります。W32T_EX の Omega は、ファイルの先頭の 0xff, 0xfe (little endian UCS2), 0xfe, 0xff (big endian UCS2), 0xef, 0xbb, 0xbf (utf8) を読み飛ばすようにしてありますので、Windows application である、メモ帳や wordpad で保存したものを、そのままコンパイルすることができます。また、lambdaj でサポートしている \UGT, \GT, \UMS コマンドは、同じように使用できます。例えば、森\UMS{9DD7}外 とすると、森鷗外 が得られます。

最後に Shift-JIS コードを使用してみます。

明朝体 亜啞娃阿哀愛挨始逢葵茜穉惡握渥旭 葦芦鰻梓压幹扱宛姐虻飴絢綾鮎
或栗 裕安庵按暗案闇鞍杏以伊位依偉囿夷

ゴシック体 亜啞娃阿哀愛挨始逢葵茜穉惡握渥旭 葦芦鰻梓压幹扱宛姐虻飴絢
綾鮎或栗 裕安庵按暗案闇鞍杏以伊位依偉囿夷

これはラムダ(lambda)でタイプセットしたものです。改行や、禁則処理は inunijapan.ocp で実現しています。inunijapan.ocp は Unicode → Unicode の translation であり、(1)日本語の改行、(2)禁則処理、(3)行末をスペースに解釈しないこと、を実現しています。予め \usepackage{unijapan} によって unijapan パッケージを読み込んでおく必要があります。いろいろな encoding から Unicode への translation には 別の専用の ocp を使えば、どのような場合にも改行、禁則処理などが可能になります。W32T_EX の Omega は、ファイルの先頭の 0xff, 0xfe (little endian UCS2), 0xfe, 0xff (big endian UCS2), 0xef, 0xbb, 0xbf (utf8) を読み飛ばすようにしてありますので、Windows application である、メモ帳や wordpad で保存したものを、そのままコンパイルすることができます。また、lambdaj でサポートしている \UGT, \GT, \UMS コマンドは、同じように使用できます。例えば、森\UMS{9DD7}外 とすると、森鷗外 が得られます。