

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鰲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鰲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。 $\text{W32T}_{\text{E}}\text{X}$ の Omega は、ファイルの先頭の `0x, 0xfe` (little endian UCS2), `0xfe, 0x` (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、`lambda.j` 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囀夷
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囀夷

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 0x, 0xfe (little endian UCS2), 0xfe, 0x (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鯁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鯁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 0x, 0xfe (little endian UCS2), 0xfe, 0x (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圀夷
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圀夷

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 0x, 0xfe (little endian UCS2), 0xfe, 0x (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。