

2008-08-18 (mon)

Canna 辞書の品詞：C5?

品詞記号一覧^{*1}には見当たらないんだけど、五段活用には二種類あるみたい。

「書かない」、「行かない」はいっしょ。

でも、「行ってきます」はよくても、「書ってきます」とはいえない。書くが K5 で、行くが C5 なのかな。

2008-08-16 (sat)

文字クラス

文字クラスにおける日本語^{*2}

文字クラスではメタ文字のハイフン(-)で範囲指定が可能です。

そこでひらがなとカタカナを表す文字クラスは概ね次のように記述することが出来ます。

/[ぁ-ん]/u

/[ァ-ヴ]/u

漢字を表す文字クラスは文字コードによって異なります。

/[一-龠]/u

/[亜-熙]/s

上記は文字コード表を厳密に調べたものではありません。取りこぼす漢字が含まれている可能性があります。より良い文字クラスの定義の仕方が分かりましたら更新します。

^{*1} <http://www.geocities.co.jp/SiliconValley-SanJose/4514/toku13.html#6>

^{*2} <http://www.ruby-life.jp/regexp/japan/index3.html>