Schrift Konvertierungen

Immer wieder kommt die Frage auch, wie man die verschiedenen japanischen (Um-)Schriften ineinander konvertiert. Es wäre schön, hier mal eine Übersicht zu haben, wo man die nötigen Infos zu den jeweiligen Konvertierungen findet. Bitte um Ergänzung, wenn man Ideen, zu den jeweiligen Punkten hat.

Hiragana <-> Katakana

Die Umwandlung ist in Unicode kein Problem.

Hiragana hat den Codepoint-Bereich [0x3040-0x309f](#) und Katakana den Codepoint-Bereich [0x30a0-0x30ff](#).

Eine Umrechnung kann also durch einfache Addition/Subtraktion, je nach Richtung, mit 0x60 (96) erfolgen.

Hiragana -> Romaji

Sollte es schon einige Lösungen im Internet geben, neben der Wadoku eigenen (u.a. von Ulrich und Hans Jörg), welche mittlerweile in Java geschrieben ist (Furigana.java), aber nur für Ulrichs erweiterte Hiragana-Auszeichnung zuverlässige Ergebnisse liefert. Liegt das Original nur im ungetagten Hiragana vor, dann ist die Konvertierung ohne Wörterbuch allein regelbasiert nicht richtig durchführbar (Fehlerquote 5%?).

Katakana -> Romaji

Ist trivial, wenn man Hiragana -> Romaji kann 😊

Romaji -> Hiragana/Katakana

Benötigt eine explizite Angabe der Zielschrift, also ob in Hiragana oder Katakana gewandelt wird. Man könnte aber z.B. klein geschriebene Worte immer in Hiragana und groß geschriebene immer in Katakana wandeln. Allerdings geht das schon eher in Richtung IME, da die wenigsten nur Kana und keine Kanji wollen.