<http://stackoverflow.com/questions/3801431/python-any-way-to-perform-this-hybrid-split-on-multi-lingual-e-g-chinese>

<http://stackoverflow.com/questions/3797746/how-to-do-a-python-split-on-languages-like-chinese-that-dont-use-whitespace>

<http://stackoverflow.com/questions/36640587/how-to-remove-chinese-punctuation-in-python>

<http://stackoverflow.com/questions/6910312/how-to-filter-chineseonly-chinese>

Unicode Table:

<http://jrgraphix.net/r/Unicode/2000-206F>

<http://jrgraphix.net/r/Unicode/3000-303F>

<http://www.rikai.com/library/kanjitables/kanji_codes.unicode.shtml>

Zhon :

<http://zhon.readthedocs.io/en/latest/#installation>

Write to UTF-8 :

<http://stackoverflow.com/questions/934160/write-to-utf-8-file-in-python>

Deal with punctuation :

<http://stackoverflow.com/questions/21209024/python-regex-remove-all-punctuation-except-hyphen-for-unicode-string>