

Morphological Dictionary to lexc file style

IPA: It is recommended by MeCab. Last update is in 2007. Weakness in new words and proper nouns. Focusing on grammar and technical terms.

Juman: Based on Kyoto corpus <https://github.com/ku-nlp/KyotoCorpus>
Last update in 2016. 30000 basic words are analyzed by human and other words are from wikipedia. Strength in accuracy.

Unidic: Very large words data. Not very suitable in natural language processing as it focus on example search on search engine.

mecab-ipadic-NEologd: IPA dictionary + new words from online. Update twice a week. Some errors are seen in word classification.

Reference: https://zenn.dev/ymmt1089/articles/20220919_mecab_dictionary

As an accuracy is important in lexc file, I will use JUMAN dictionary. I have found the useful website, which explains the inside of JUMAN dictionary
<https://hayashibe.jp/tr/juman/dictionary/entry>