CoNLL-U Format

https://universaldependencies.org/format.html

・テキストファイル
・UTF-8, LF

■ 3 種類の行

1. 単語行
    1.word/token
    2.10 種類のフィールド、タブ区切り
2. 空行
    1. 文の区切りを意味する
3. コメント行
    1. ハッシュ (#) で始める

■ 10 種類のフィールド

1.ID:　文内の単語の番号（１から）
2.FORM: 語形（句読点も）
3.LEMMA: レマ
4.UPOS: 普遍品詞タグ
5.XPOS: 言語固有の品詞タグ（ない場合は、アンダスコア）
6.FEATS: 形態的特徴
7.HEAD: その単語のヘッドになる単語の ID（それ自身の場合は０）
8.DEPREL: ヘッドとの依存関係
9.DEPS: 拡張依存グラフ
10.MISC: 備考

■注

1. 各フィールは空ではいけない。
2.FORM, LEMMA, MISC 以外は、スペースを入れてはいけない
3. 値がない場合はアンダスコア

処理プログラム

https://universaldependencies.org/tools.html

ファイル Viewer

https://universaldependencies.org/conllu_viewer.html