

トキポナのトークナイザを作ろう

ミニマリズム言語のトークナイザならかんたんにできるよ

トークナイザを作りたい！

- トークナイザ
 - ▷ 機械でテキストを扱うためには, 文章を単語や記号の単位に分割しなければならない
- むずかしい！
- **かんたん**な言語でなら作れる？

トキポナとは

- みんなが知ってるミニマリズム人工言語
 - ▷ 語彙数: 120 語
 - ▷ 習得がかんたん たのしい!
 - ▷ できる人が割と多い たのしい!
- ごくわずかな知らない人のために
 - ▷ 初級トキポナ文法簡介
 - ▷ 2 分 40 秒で世界一簡単な言語を紹介して伝授する
 - ▷ トキポナレッスン1「トキポナってなに」
 - ▷ jan Pije's lessons (注)
 - ▷ 【日本語訳】toki pona li toki pona - トキポナソング

トキボナのトークナイザを作りました

- [github:nymwa/ilonimi](#)
 - ▷ Python 3 で書いた
 - ▷ pip が使えれば使えます
- この資料では, このトークナイザが何をやってるか説明します

どのような処理を行うか？

単語のリストを作る

記号と単語を分離する

固有名詞の判定をする

固定名詞のオートマトン

nm と nn を除外する規則
