



CoNLL-U

Условие

У вас есть пять файлов формата CoNLL-U (`.conllu`) с разобранными первыми абзацами статьи из Википедии с заголовком “Язык” на разных языках:

1. русский
2. турецкий
3. арабский
4. индонезийский
5. баскский

Задание

1. Сначала посмотрите на русский язык. Составьте словарь `russian_tags`, где каждой части речи (`upos`) будет соответствовать словарь морфологических признаков (`feats`), внутри которого ключами будут признаки, а значениями — списки значений, принимаемых признакам. Например, неполный словарь может выглядеть так:

```
{ 'NOUN': { 'Number': [ 'Sing', 'Plur' ] }}
```

NB! Советуем сохранить полученный словарь в формате JSON (`.json`) и отсортировать в алфавитном порядке.

Не переживайте, если их сокращения кажутся совсем непонятными. Вот описание (к сожалению, на английском) для русских частей речи и морфологических признаков.

2. Прodelайте подобную работу и для остальных языков, получив
 - a. турецкий словарь `turkish_tags` (описание тегов)
 - b. арабский словарь `arabic_tags` (описание тегов)

с. индонезийский словарь `indonesian_tags` (описание тегов)

д. баскский словарь `basque_tags` (описание тегов)

3. Самая важная часть!

Посмотрите, какие части речи, морфологические признаки и их значения отсутствуют в русском языке, но присутствуют в других языках. Составьте список подобных отличий, где будут присутствовать все промежуточные “ключи” и отсутствующие признаки. Например, можно сделать список строк:

```
['cconj', 'NOUN Number Plur']
```

4. (Если останется время)

Попробуйте вывести слова соответствующие значениям из списка в предыдущем задании.