

Домашнее задание 2

Работаем с файлом References.txt.

Важно иметь в виду, что в файле имеются слова, содержащие буквенные символы с диакритиками (ń, Ö, é и др.). Они не входят в диапазон A-Za-z, как и буква ё, зато их находит \w.

Задание на 6 баллов

Задания не нужно выполнять с использованием Python.

На 6 баллов нужно выполнить пункты 1-3.

1) Написать регулярное выражение, которое найдет все упоминания автора(-ов) и года. Необходимо учесть и такие случаи:

Haspelmath, Martin. 2008a.

Haspelmath, Martin. 2008b.

2) Написать регулярное выражение, которое найдет все последовательности, соответствующие формату “автор и год” для работ без соавторов, т.е. с одним автором.

Формат такой:

Фамилия, Имя (после имени может быть добавлено сокращение, например, второго имени). Год

Фамилии и имя могут содержать дефисы

Подходит под формат:

Aikhenvald, Aleksandra Y. 2013

Buechel, Eugene. 1939

Koptjevskaja-Tamm, Maria. 2003

McGregor, William. 1996

Не подходит под формат:

McGregor, William, 1996

Croft William. 1996

Chappell, Hilary & McGregor, William. 1996

3) Написать регулярное выражение, которое найдет все упоминания в скобках.

Задания на 8-10 баллов

Задания нужно выполнять с использованием Python.

На 8 баллов, нужно выполнить пункты 1-3.

На 9 баллов, нужно выполнить пункты 1-4.

На 10 баллов, нужно выполнить пункты 1-5.

(8 баллов)

1) Найти все упоминания автора(-ов) и года. Необходимо учесть и такие случаи:

Haspelmath, Martin. 2008a.

Haspelmath, Martin. 2008b.

2) Написать программу, которая читает файл по строкам и проверяет, что строка начинается с упоминания автора и года для работ без соавторов, т.е. с одним автором.

Формат такой:

Фамилия, Имя (после имени может быть добавлено сокращение, например, второго имени). Год

Фамили и имя могут содержать дефисы

Подходит под формат:

Aikhenvald, Aleksandra Y. 2013

Buechel, Eugene. 1939

Koptjevskaja-Tamm, Maria. 2003

McGregor, William. 1996

Не подходит под формат:

McGregor, William, 1996

Croft William. 1996

Chappell, Hilary & McGregor, William. 1996

3) Посчитать количество упоминаний в скобках.

(9 баллов)

4) Добавить к годам скобки, т.е. в результате замены ссылки должны выглядеть так:

Aikhenvald, Aleksandra Y. (2013). Possession and ownership: A cross-linguistic perspective. In

Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), Possession and ownership: A cross-

linguistic typology, 1-64. Oxford: Oxford University Press.

Haspelmath, Martin. (2008a). Frequency vs. iconicity in explaining grammatical asymmetries.

Cognitive Linguistics 19(1). 1-33.

(10 баллов)

5) Найти публикации издательства Benjamins

Программа обрабатывает файл со списком публикаций и печатает:

Работы, опубликованные в Benjamins:

Bybee, Joan L. 1985.

Morphology: A study of the relation between meaning and form.

Corbett, Greville G., Andrew Hippisley, Dunstan Brown & Paul Marriott. 2001.

Frequency, regularity and the paradigm: A perspective from Russian on a complex relation. In Joan

L. Bybee & Paul Hopper (eds.), Frequency and the emergence of linguistic structure,

201-226.

...

Подсказка: помогут группирующие скобки (...), скобочные группы (?: ...) и re.finditer()

