

Отчет: конкордансы

Ход выполнения: составитель конкордансов был выполнен на языке программирования Java, с использованием открытого морфологического словаря OpenCorpora, который используется для лемматизации токенов.

Описание программы: составитель конкордансов принимает на вход текст большого размера, разбивает его на токены, каждый токен лемматизируется. Также программа принимает на вход фразу, которая является основой конкорданса (фраза также лемматизируется)

Следующий этап работы программы: с помощью скользящего окна находим все индексы токенов, которые являются началом вхождения исходной фразы в текст. Затем обходим все эти индексы вхождений и для каждого такого вхождения формируем список конкордансов. Добавляем их в Map, которая содержит все конкордансы. Если такой конкорданс уже присутствует в Map, то увеличиваем счетчик на 1 (тем самым подсчитываем частоту вхождений)

Результат:

```
All occurrences found
13
Left contexts:
<по который он поднял рука > 1
<который он поднял рука > 1
<высок поднял рука > 1
<он поднял рука > 1
<не поднял рука > 2
<который поднял рука > 2
<когда поднял рука > 2
<Я поднял рука > 3
<поднял рука > 13
```

Right contexts:

<поднял рука свой на царь > 1
<поднял рука мой на господин > 1
<поднял рука Мою как народ > 1
<поднял рука Мою с клятва > 1
<поднял рука Мою с > 1
<поднял рука свой и > 1
<поднял рука мой > 1
<поднял рука мой на > 1
<поднял рука свой на господин > 1
<поднял рука свой на Сион > 1
<поднял рука Мою как > 1
<поднял рука Мою на > 1
<поднял рука с > 1
<поднял рука с пращным камень > 1
<поднял рука Мою на они > 1
<поднял рука с пращным > 1
<поднял рука свой и простер > 1
<поднял рука на > 2
<поднял рука на царь > 2
<поднял рука свой на > 3
<поднял рука Мою > 3
<поднял рука свой > 6
<поднял рука > 13