**Московский авиационный институт  
(национальный исследовательский университет)**

Факультет прикладной математики и физики  
Кафедра вычислительной математики и программирования

**Лабораторная работа №5  
по курсу «Обработка естественно-языковых текстов»**

Поиск коллокаций

|  |  |  |
| --- | --- | --- |
|  | **Студент:** | Рожлейс И.А. |
| **Группа:** | М8О-108М-18 |
| **Преподаватель:** | Калинин А.Л. |
| **Дата:** | 25.09.2019 |

Москва, 2019

**Цель работы**

Необходимо найти коллокации в имеющемся корпусе, использованного для построения поисковой системы (или его случайному подмножеству достаточного размера). Для поиска коллокаций необходимо использовать как минимум два статистических алгоритма из рассмотренных на лекциях. Сравнить выделенные коллокации между собой, пояснить различия с точки зрения использованных алгоритмов.

В отчёте должны быть приведены количество найденных коллокаций, оценка точности метода, примеры найденных коллокаций и ошибочно найденных словосочетаний.

**Ход работы**

А

**Заключение**

А

**Источники**

[1] Маннинг, К. Введение в информационный поиск // Вильямс, 2014. – 528 с.

[2] How can I sort a std::map first by value, then by key? [Электронный ресурс] // URL: <https://stackoverflow.com/questions/19842035/how-can-i-sort-a-stdmap-first-by-value-then-by-key>

[3] Список частотности по НКРЯ [Электронный ресурс] // URL: [https://‌ru.wiktionary.org/wiki/Приложение:Список\_частотности\_по\_НКРЯ](https://‌ru.wiktionary.org/‌wiki/Приложение:Список_частотности_по_НКРЯ)