

Министерство образования Республики Беларусь

Учреждение образования

“Белорусский государственный университет информатики и
радиоэлектроники”

Кафедра информационных интеллектуальных технологий

Лабораторная работа 1

«Разработка автоматизированной системы формирования словаря
естественного языка»

Выполнил

гр.121701

Мулярчик Д.С.

Лемантович Д.К.

Проверил

Крапивин Ю. Б

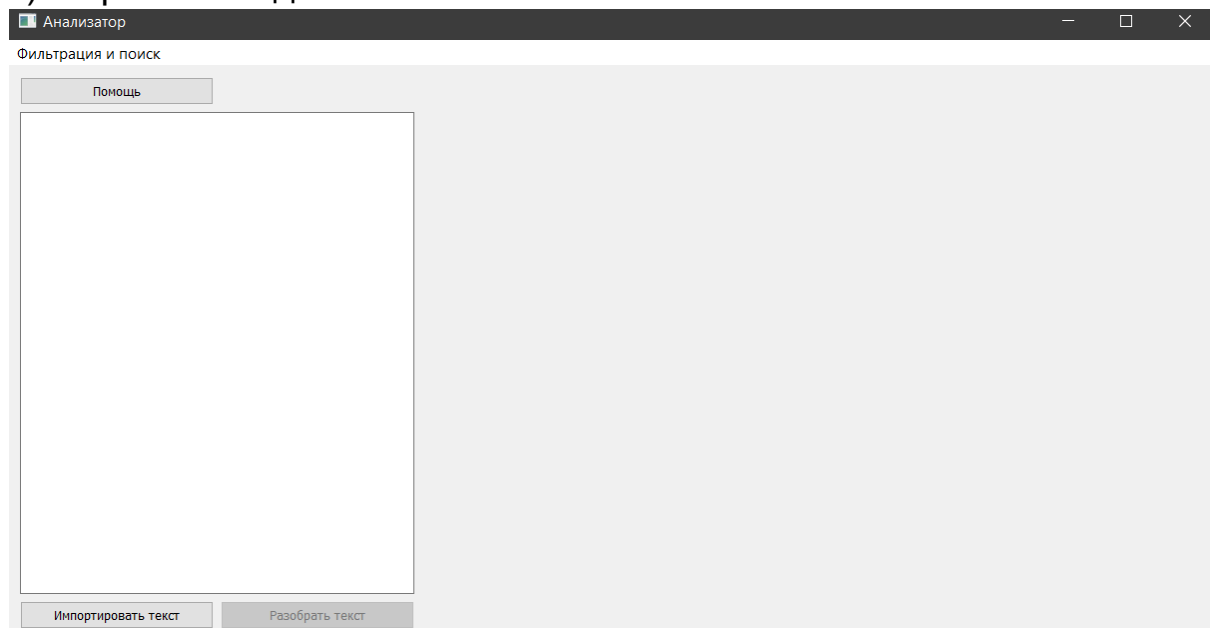
Минск 2024

Цель работы: Освоить принципы разработки прикладных сервисных программ для решения задачи автоматического лексического и лексико-грамматического анализа текста естественного языка.

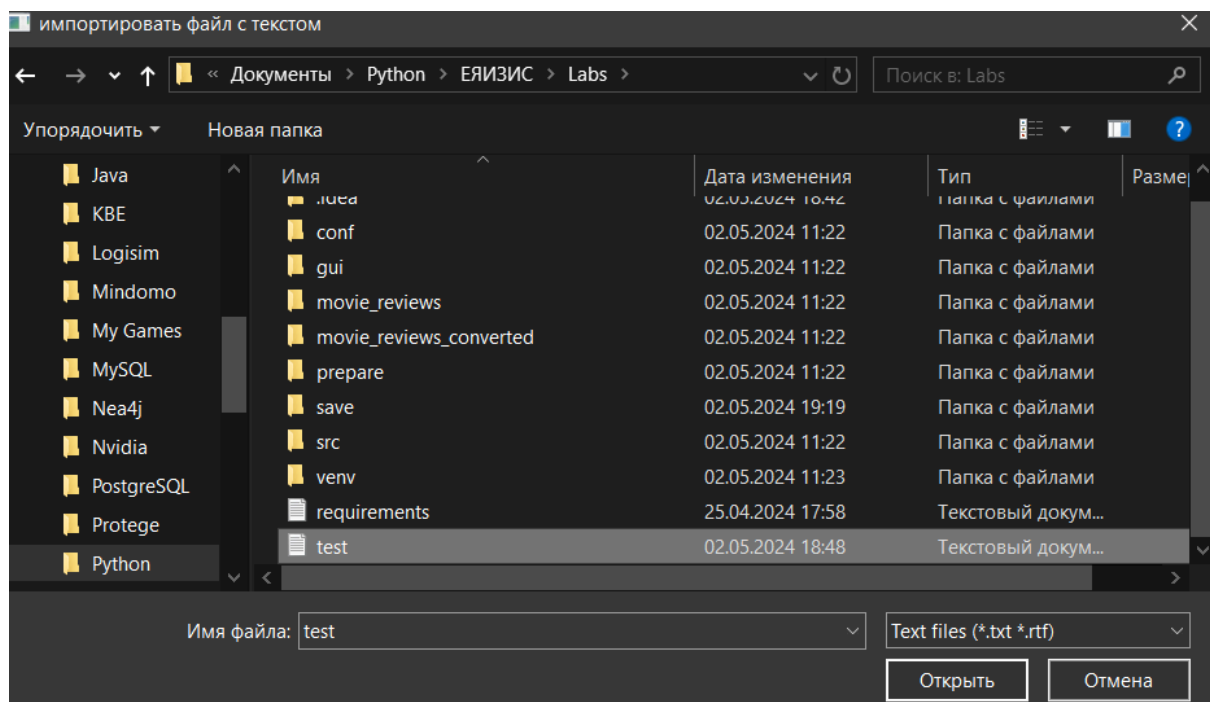
Задание: Список слов, упорядоченный по алфавиту и включающий как лексемы, так и словоформы, с указанием частоты встречаемости каждой из форм. Для словоформ пользователю должна быть предоставлена возможность вводить дополнительную морфологическую информацию, а именно, отнесение слова к соответствующей части речи, указание рода, числа, падежа и т.п. При этом морфологическая информация может быть оформлена как отдельная неформатированная запись, т.е. это просто текст, который пользователь может оформлять произвольным образом. Язык текста – русский, формат входного документа – TXT, RTF.

Алгоритм работы программы:

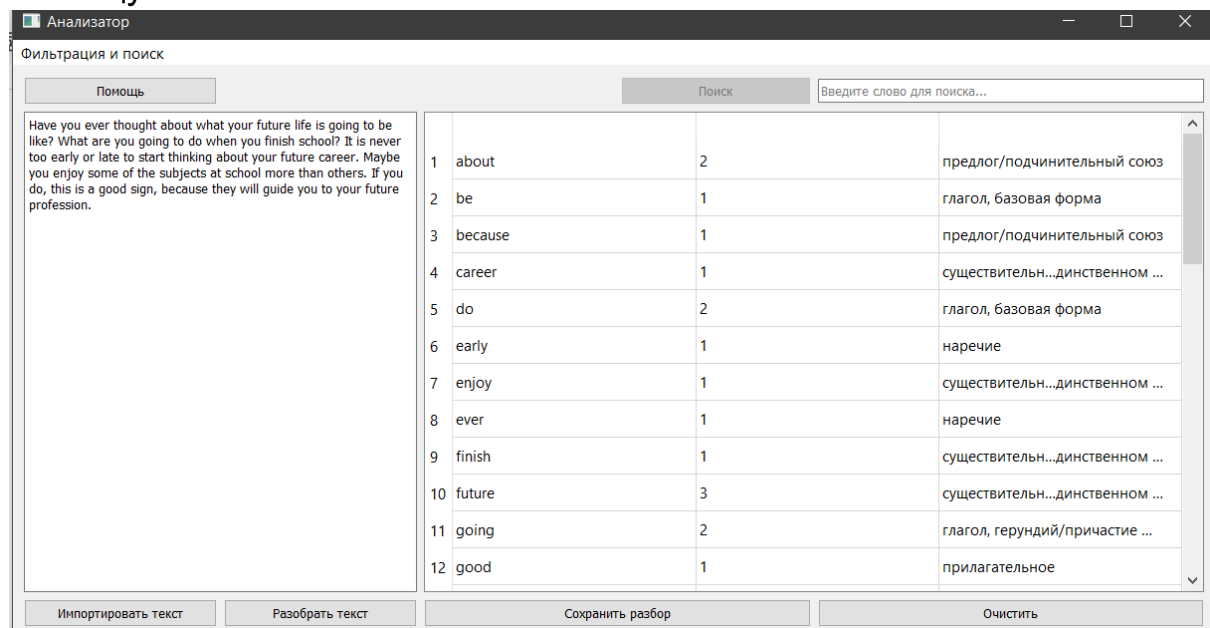
1) Открывается диалоговое окно:



2) Пользователь импортирует текст нажав на кнопку импорта.



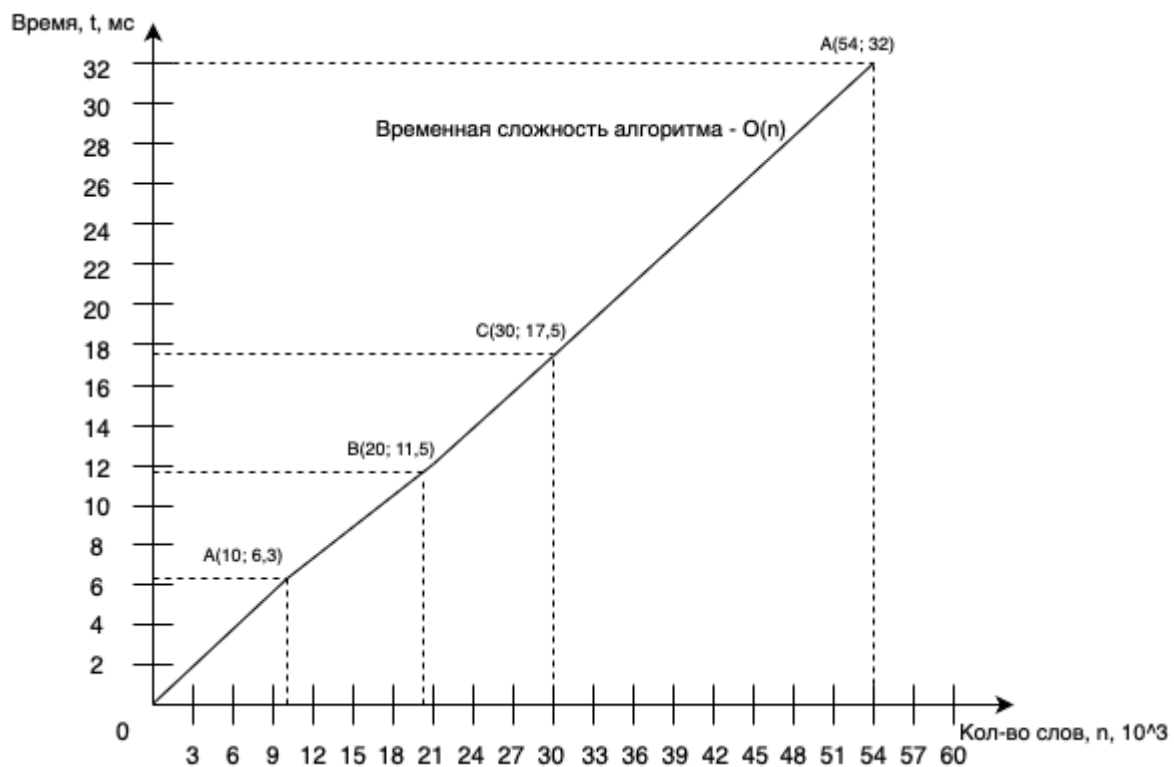
3) Пользователь выбирает разобрать текст и программа проходит по всему тексту, используя регулярное выражение, добавляет их в таблицу.



4) Пользователь может сохранить разобранный текст в json формате.

```
{
  "about": {
    "frequency": 2,
    "additional information": "предлог/подчинительный союз"
  },
  "be": {
    "frequency": 1,
    "additional information": "глагол, базовая форма"
  },
  "because": {
    "frequency": 1,
    "additional information": "предлог/подчинительный союз"
  },
  "career": {
    "frequency": 1,
    "additional information": "существительное в единственном числе"
  },
  "do": {
    "frequency": 2,
```

График зависимости времени обработки от количества слов.



Вывод: Освоены принципы разработки прикладных сервисных программ для решения задачи автоматического лексического и лексико-грамматического анализа текста естественного языка. Разработанное приложение полезно для составления статистики и расчета количества определенных слов в документах.