

Учреждение образования
«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ
Кафедра интеллектуальных информационных технологий

Отчёт
по курсу «Естественно-языковой интерфейс интеллектуальных
систем»

Лабораторная работа №1
«Разработка автоматизированной системы формирования словаря
естественного языка»

Выполнили студенты группы 121701:	Липский Р. В. Жолнерчик И. А. Стронгин А. В.
Проверил:	Крапивин Ю.Б.

Минск 2023

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.

Интерфейс:

Text Analyzer

Selected File:

/home/rlipski/opt/bsuir/semester-6/EYAzIIS/LW

Select File

Analyze File

Word:

o

 Morphologic Info:

Text

 Occurences:

5

 <= x <=

15

Word	Morphologic Info	Occurrences
office	Text: office, Lemma: office, Pos: noun	7
of	Text: of, Lemma: of, Pos: adposition	13
for	Text: for, Lemma: for, Pos: adposition	5
into	Text: into, Lemma: into, Pos: adposition	5
to	Text: to, Lemma: to, Pos: particle	13

Contact with developers

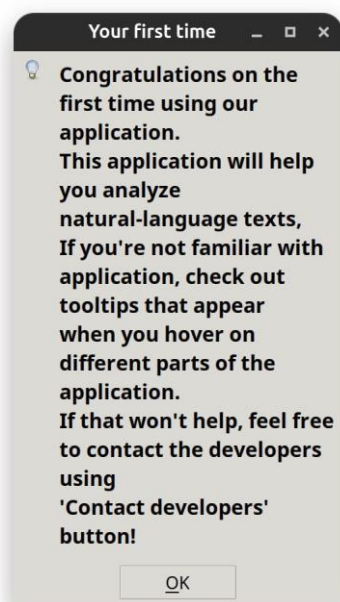
Edit Selected

into

Text: into, Lemma: into,

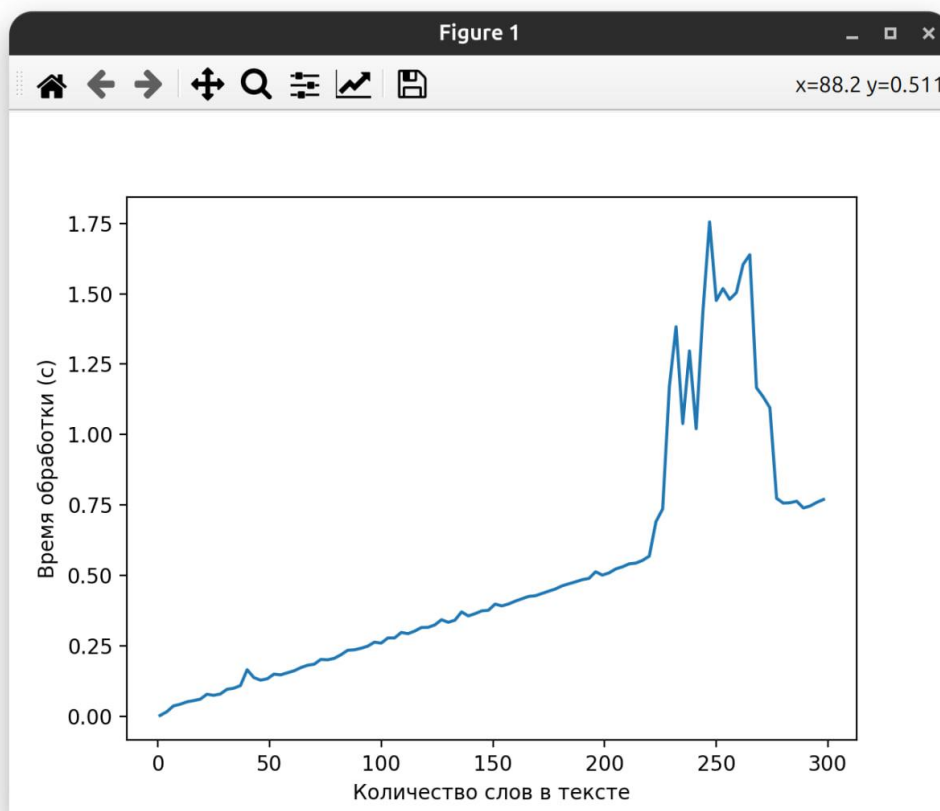
Apply Changes

Главное окно



Дополнительное окно при первом входе в систему для благоприятного начала работы пользователем

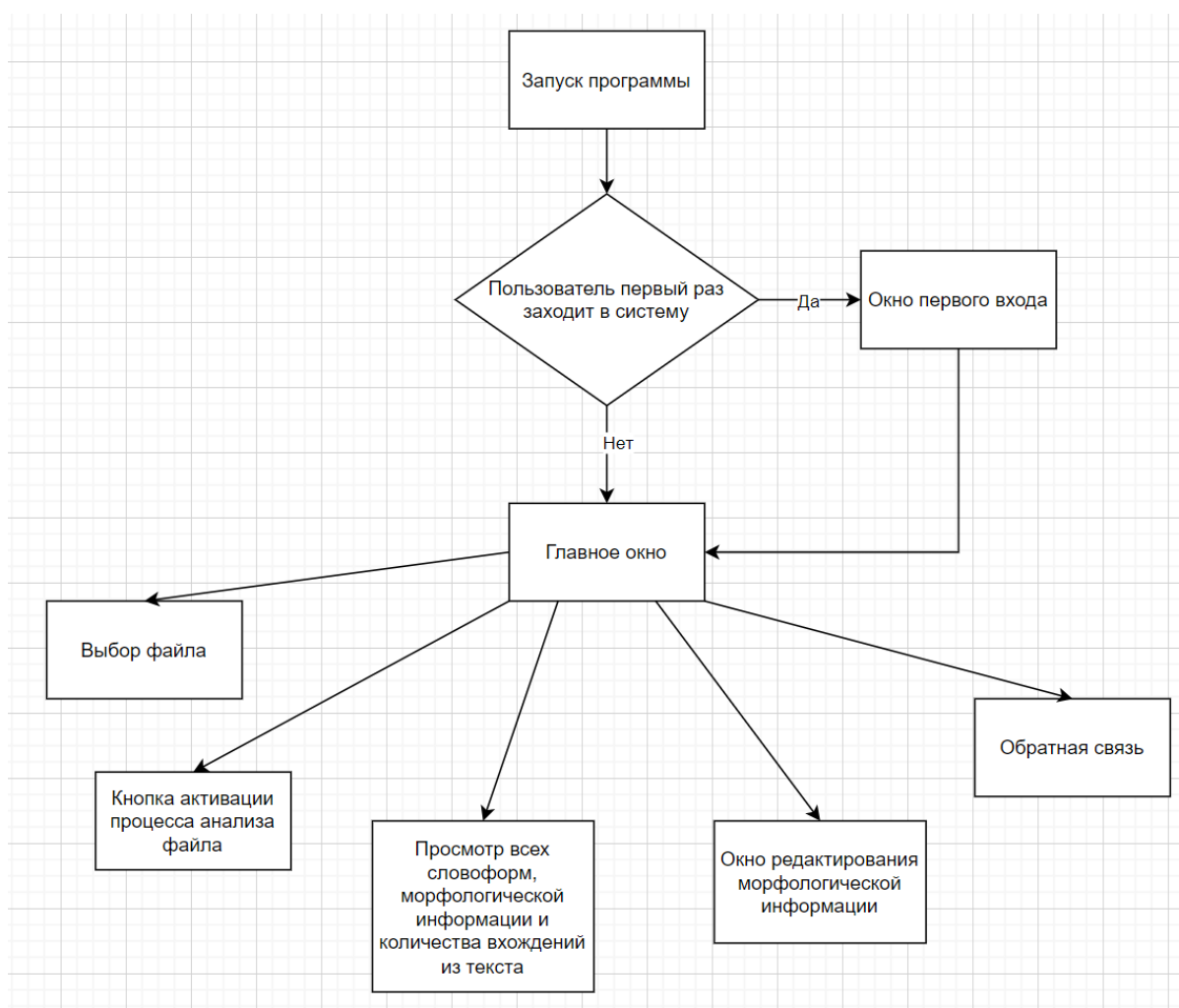
Тест производительности:



Структура хранения данных



Структурная схема приложения



Описание алгоритмов:

Поиск по подстроке: Начало – Получаем все слова из БД и сохраняем в список слов – Получаем нужную пользователю подстроку для поиска в переменную – Перебираем все слова из списка слов и ищем совпадения подстроки в словоформах – Отображаем результаты поиска в таблице – Конец

Поиск по количеству вхождений: Начало – Получаем все словоформы из БД и сохраняем в список слов – Получаем нижнее и верхнее пороговое значения для количества вхождений – Перебираем все слова из списка слов и ищем те, количество вхождений которых удовлетворяет пороговым значениям – Отображаем результаты поиска в таблице – Конец

Поиск по подстроке морфологической информации: Начало – Получаем все слова из БД и сохраняем в список слов – Получаем нужную пользователю подстроку для поиска в переменную – Перебираем все слова из списка слов и ищем совпадения подстроки в морфологической информации – Отображаем результаты поиска в таблице – Конец

Разбиение текста на словоформы и получение их морфологической информации: Начало – Получаем файл формата TXT/RTF – Считываем текст файла – Получаем список словоформ – Получаем всю морфологическую информацию о каждой – Преобразуем в неформатированную строку – Загружаем словоформы и их морфологическую информацию в БД – Конец

Вывод:

Данная программа представляет собой инструмент для создания списка словоформ, включающих как лексемы, так и словоформы с указанием частоты их встречаемости и их морфологическими характеристиками. Она также предоставляет возможность пользователю редактировать морфологическую информацию для словоформ. Данная программа имеет перспективы в области анализа текстов и книг для подведения различных статистик.