Министерство образования Республики Беларусь

Учреждение образования

«Брестский государственный технический университет»

Кафедра ИИТ

Лабораторная работа №2

По дисциплине: «Естественно-языковой интерфейс ИС»

# Тема: «Построение и использование корпусов текстов естественного языка»

Выполнил:

Студент 3 курса

Группы ИИ-21

Литвинюк Т. В.

Проверила:

Якимук А. В.

Брест 2024

**Цель:** 1. изучить принципы построения корпусов текстов, виды разметки и способы аннотирования, инструменты работы с корпусами текстов, 2. построить корпус текстов и разработать корпусный менеджер.

**Ход работы:**

|  |  |  |
| --- | --- | --- |
| **Номер**  **варианта** | **Язык текста** | **Предметная область** |
| 6 | Русский | Литература |

**Методические указания:**

Требуется спроектировать и программно реализовать структуры хранения данных, алгоритмы их обработки, необходимые в рамках следующих базовых требований к разрабатываемому приложению:

* входные данные – фрагмент текста (фраза или слово) на естественном языке – запрос корпусному менеджеру;
* выходные данные – частотные характеристики словоформ, лексем, грамматических категорий, леммы, морфологические характеристики словоформ и их метаданные (библиографические, типологические), конкордансные списки, согласно согласованным с преподавателем требованиям к функциональности корпусного менеджера;
* взаимодействие с пользователем посредствам графического интерфейса (интерфейс должен быть интуитивно-понятным и дружественным пользователю);
* наличие системы средств помощи пользователю;
* обеспечение возможности построения, сохранения, просмотра, редактирования, пополнения, фильтрации и поиска по заданному условию, документирования текста и/или его фрагмента в соответствии с реализуемой функциональностью корпусного менеджера;
* поддержка различных форматов представления входных данных (TXT, RTF, PDF, DOC, DOCX).

import tkinter as tk

from customtkinter import CTk as Tk, CTkButton as Button, CTkFrame as Frame

from manager import TextManager

# Создание экземпляра корпусного менеджера

manager = TextManager()

# Создание графического интерфейса

root = Tk()

root.title("Программа")

menu\_bar = tk.Menu(root)

root.config(menu=menu\_bar)

menu\_bar.add\_command(label="Открыть файл", command=manager.load\_file)

# Создание фрейма для первой колонки кнопок

left\_frame = Frame(root)

left\_frame.pack(side="left")

word\_freq\_button = Button(left\_frame, text="Частотные характеристики словоформ", command=manager.word\_frequencies\_)

word\_freq\_button.pack(fill='x', pady=5)

lemma\_freq\_button = Button(left\_frame, text="Частотные характеристики лексем", command=manager.lemma\_frequencies)

lemma\_freq\_button.pack(fill='x', pady=5)

pos\_tag\_freq\_button = Button(left\_frame, text="Частотные характеристики грамматических категорий", command=manager.pos\_tag\_frequencies)

pos\_tag\_freq\_button.pack(fill='x', pady=5)

lemma\_button = Button(left\_frame, text="Леммы", command=manager.lemmas)

lemma\_button.pack(fill='x', pady=5)

morph\_analysis\_button = Button(left\_frame, text="Морфологический анализ", command=manager.morphological\_analysis)

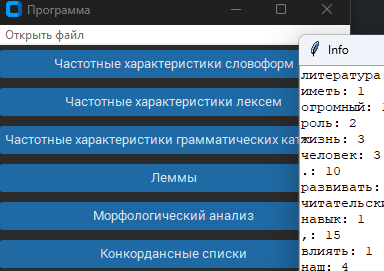
morph\_analysis\_button.pack(fill='x', pady=5)

concordance\_button = Button(left\_frame, text="Конкордансные списки", command=manager.concordance\_lists)

concordance\_button.pack(fill='x', pady=5)

root.mainloop()

# Результат:



**Вывод:** в ходе выполнения лабораторной работы освоил принципы построения корпусов текстов, виды разметки и способы аннотирования, инструменты работы с корпусами текстов.