Министерство образования Республики Беларусь

Учреждение образования

«Брестский государственный технический университет»

Кафедра ИИТ

Лабораторная работа №3

По дисциплине: «Естественно-языковой интерфейс ИС»

# Тема: «Реализация алгоритмов автоматического морфологического и лексико-грамматического анализа текста естественного языка»

Выполнил:

Студент 3 курса

Группы ИИ-21

Литвинюк Т. В.

Проверила:

Якимук А. В.

Брест 2024

**Цель:** закрепить знания морфологического и лексико-грамматического анализа текста.

**Ход работы:**

|  |  |  |
| --- | --- | --- |
| **Номер**  **варианта** | **Язык текста** | **Формат документа** |
| 7 | Английский | docx |

**Задание 1.** Создать программную реализацию решения задачи о редакционном расстоянии методом динамического программирования между словами, подаваемыми на вход системы и словами текста.

**Задание 2.** Спроектируйте внешний интерфейс автоматизированной системы осуществляющей решения задачи о редакционном расстоянии между словами, подаваемыми на вход системы и словами текста.

**Задание 3.** Разработайте алгоритмы сервисных функций, предоставляемых Вашей системой.

**Задание 4.** Реализуйте разработанные алгоритмы в виде программ на языке программирования.

**Задание 5.** Протестируйте разработанную систему, используя различные входные тексты.

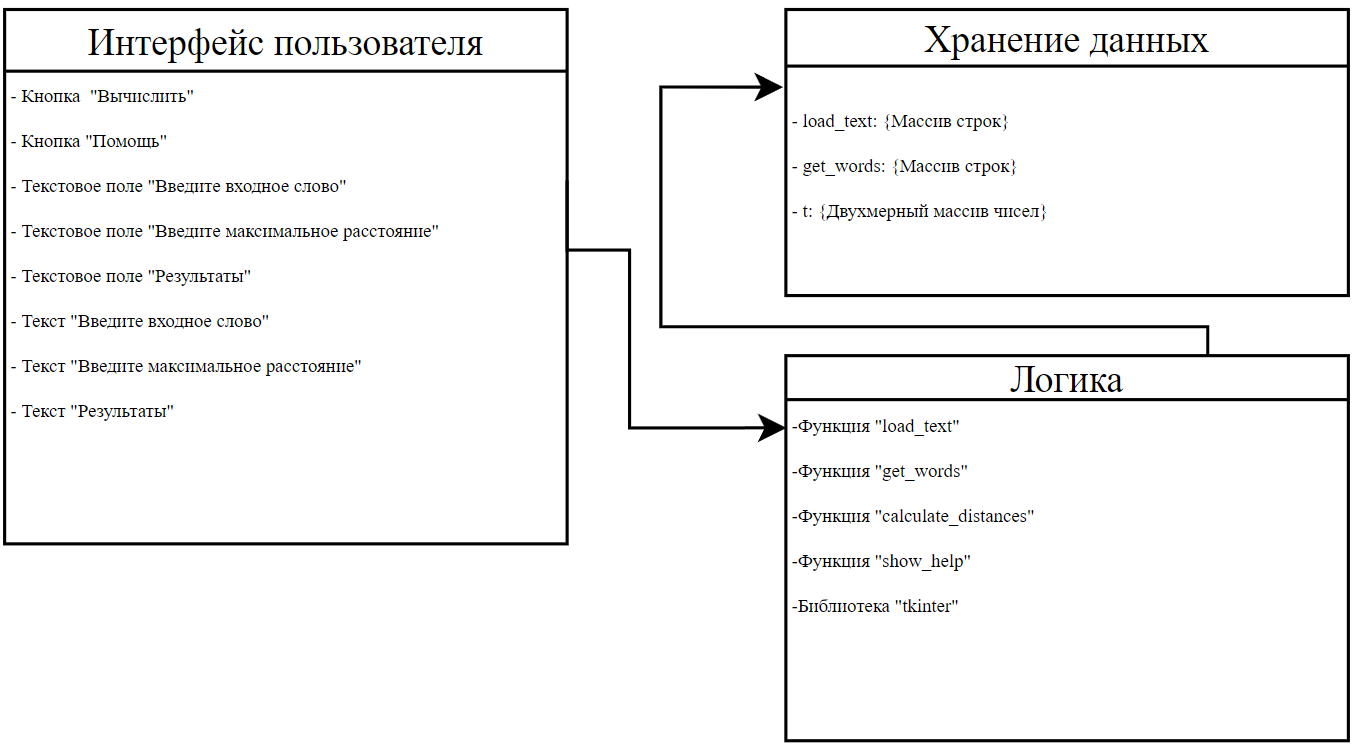
**Задание 6.** Разработайте и внедрите в систему средства помощи пользователю.

**Задание 7.** Протестируйте систему, включающую средства помощи пользователю.

**Методические указания:**

* на входе – слова естественно-языкового текста содержащие орфографические ошибки;
* на выходе – перечень слов, расстояние до которых от входного слова не превышает некоторого заданного Вами значения, упорядоченный по убыванию величины расстояния редактирования;
* интерфейс системы должен быть предельно простым и доступным для пользователей любого уровня, содержать понятный набор инструментов и средств, а также help-средства;
* наличие средств распечатки полученной на выходе информации.

**Структурно-функциональная схема:**



**Листинг:**

import tkinter as tk

import docx2txt

# Функция вычисления редакционного расстояния

def edit\_distance(str1, str2):

m = len(str1) + 1

n = len(str2) + 1

t = [[i + j for j in range(n)] for i in range(m)]

for i in range(1, m):

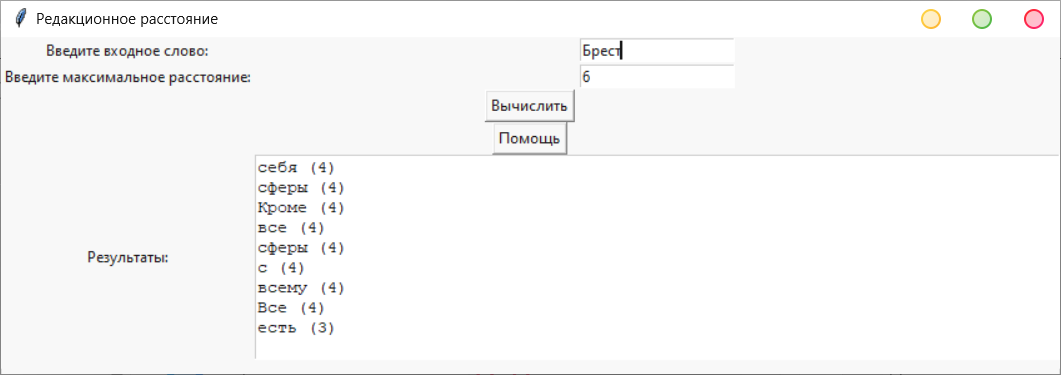
c = i - 1

for j in range(1, n):

d = j - 1

t[i][j] = min(t[c][j] + 1, t[i][d] + 1, t[c][d] + (str1[c] != str2[d]))

return t[m - 1][n - 1]



root = tk.Tk()

root.title("Редакционное расстояние")

def load\_text(filename):

text = docx2txt.process(filename)

return text

def get\_words(text):

words = text.split()

return words

text = load\_text('EAI/LAB3/text.docx')

words = get\_words(text)

def calculate\_distances():

input\_word = input\_entry.get()

max\_distance = int(max\_distance\_entry.get())

results = []

for word in words:

distance = edit\_distance(input\_word, word)

if distance <= max\_distance:

results.append((word, distance))

results.sort(key=lambda x: x[1], reverse=True)

output\_text.delete("1.0", tk.END)

for word, distance in results:

output\_text.insert(tk.END, f"{word} ({distance})\n")

# Интерфейс

input\_label = tk.Label(root, text="Введите входное слово:")

input\_entry = tk.Entry(root, width=100)

max\_distance\_label = tk.Label(root, text="Введите максимальное расстояние:")

max\_distance\_entry = tk.Entry(root, width=100)

calculate\_button = tk.Button(root, text="Вычислить", command=calculate\_distances)

output\_label = tk.Label(root, text="Результаты:")

output\_text = tk.Text(root, height=10)

input\_label.pack()

input\_entry.pack()

max\_distance\_label.pack()

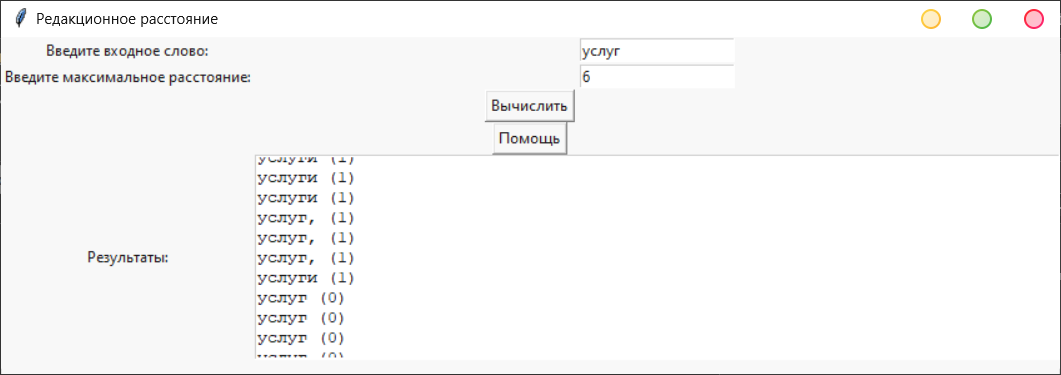
max\_distance\_entry.pack()

calculate\_button.pack()

output\_label.pack()

output\_text.pack()

root.mainloop()



**Вывод:** в ходе выполнения лабораторной работы закрепил знания морфологического и лексико-грамматического анализа текста.