Министерство образования Республики Беларусь

Учреждение образования

«Брестский государственный технический университет»

Кафедра ИИТ

Лабораторная работа №3

По дисциплине: «Естественно-языковой интерфейс ИС»

# Тема: «Разработка системы машинного перевода документов»

Выполнил:

Студент 4 курса

Группы ИИ-21

Ливтинюк Т. В.

Проверила:

Якимук А. В.

Брест 2024

**Цель:** освоить на практике основные принципы машинного перевода документов. 

**Ход работы:**

на входе – естественно-языковой текст на входном языке, подлежащий процедуре

машинного перевода;

• подсчитать количество слов во входном тексте, количество переведенных слов,

определить грамматическую информацию (теги частей речи и их расшифровка, для

этого следует использовать функциональность лр. No3 весенний семестр).

• на выходе: – перевод входного текста на выходной язык; – упорядоченный по частоте

встречаемости в тексте список слов и их переводов на

выходной язык с грамматической информацией (вкладка 1) (можно использовать

функциональность лр. No1 весеннего семестра);

– построенное дерево синтаксического разбора выбранного предложения

(вкладка 2).

• обеспечить наличие утилиты автоматического пополнения/корректировки

полученного словаря (таблицы БД).

• обеспечить сохранение и распечатку результатов перевода и упорядоченных по

частоте встречаемости в тексте списков слов и их переводов на выходной язык с

грамматической информацией в файл формата txt кодировки Unicode.

• интерфейс системы должен быть простым и доступным для пользователей любого

уровня, содержать понятный набор инструментов.

import tkinter as tk

from tkinter import filedialog, scrolledtext, messagebox

import fitz

from googletrans import Translator

import tempfile

import os

import threading

class PDFTranslatorApp:

def \_\_init\_\_(self, root):

self.root = root

self.root.title("Перакладчык")

self.open\_button = tk.Button(root, text="Открыть PDF", command=self.open\_pdf)

self.open\_button.pack(pady=10)

self.text\_area = scrolledtext.ScrolledText(root, wrap=tk.WORD, width=60, height=20)

self.text\_area.pack(pady=10)

self.translate\_button = tk.Button(root, text="Перевести", command=self.start\_translation\_thread)

self.translate\_button.pack(pady=10)

self.stop\_button = tk.Button(root, text="Стоп", command=self.stop\_translation, state=tk.DISABLED)

self.stop\_button.pack(pady=10)

self.translated\_text\_area = scrolledtext.ScrolledText(root, wrap=tk.WORD, width=60, height=20)

self.translated\_text\_area.pack(pady=10)

self.translator = Translator()

self.temp\_file = None

self.stop\_translation\_flag = False

def open\_pdf(self):

file\_path = filedialog.askopenfilename(filetypes=[("PDF Files", "\*.pdf")])

if file\_path:

try:

doc = fitz.open(file\_path)

self.temp\_file = tempfile.NamedTemporaryFile(delete=False, mode='w+', encoding='utf-8')

for page\_num in range(doc.page\_count):

page = doc.load\_page(page\_num)

page\_text = page.get\_text("text")

self.temp\_file.write(page\_text + '\n\n')

self.temp\_file.flush()

self.temp\_file.seek(0)

self.text\_area.delete(1.0, tk.END)

self.text\_area.insert(tk.INSERT, self.temp\_file.read())

self.translated\_text\_area.delete(1.0, tk.END)

self.translate\_button.config(state=tk.NORMAL)

except Exception as e:

messagebox.showerror("Ошибка", f"Не удалось открыть PDF: {e}")

def start\_translation\_thread(self):

"""Запуск перевода в отдельном потоке"""

self.stop\_translation\_flag = False

translation\_thread = threading.Thread(target=self.translate\_text)

translation\_thread.start()

self.stop\_button.config(state=tk.NORMAL)

self.translate\_button.config(state=tk.DISABLED)

self.open\_button.config(state=tk.DISABLED)

def stop\_translation(self):

self.stop\_translation\_flag = True

self.stop\_button.config(state=tk.DISABLED)

def translate\_text(self):

if self.temp\_file:

try:

self.temp\_file.seek(0)

self.translated\_text\_area.delete(1.0, tk.END)

self.translate\_button.config(text="Переводим...")

for line in self.temp\_file:

if self.stop\_translation\_flag:

break

if line.strip():

translated = self.translator.translate(line.strip(), src='en', dest='fr').text # перевод с английского на немецкий

self.translated\_text\_area.insert(tk.END, translated + '\n\n')

self.translated\_text\_area.update\_idletasks()

self.translate\_button.config(state=tk.NORMAL, text="Перевести")

self.open\_button.config(state=tk.NORMAL)

self.stop\_button.config(state=tk.DISABLED)

except Exception as e:

messagebox.showerror("Ошибка", f"Не удалось перевести текст: {e}")

self.translate\_button.config(state=tk.NORMAL)

self.stop\_button.config(state=tk.DISABLED)

self.open\_button.config(state=tk.NORMAL)

else:

messagebox.showwarning("Внимание", "Сначала откройте PDF файл.")

def \_\_del\_\_(self):

"""Функция для удаления временного файла при завершении программы"""

if self.temp\_file:

try:

os.remove(self.temp\_file.name)

except Exception as e:

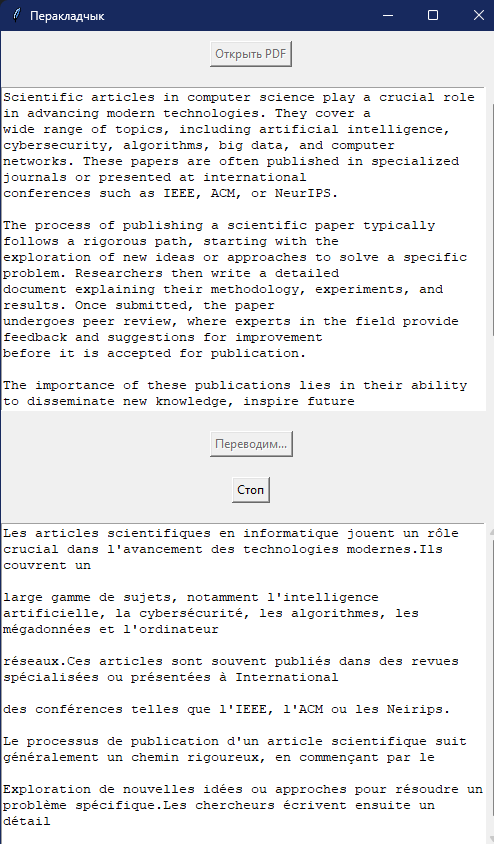
print(f"Не удалось удалить временный файл: {e}")

if \_\_name\_\_ == "\_\_main\_\_":

root = tk.Tk()

app = PDFTranslatorApp(root)

root.mainloop()

****

**Вывод:** в ходе выполнения лабораторной работы освоил принципы машинного перевода документов.