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 Пучинский А.А.

> Проверил: Булей Е.В.

Цель: освоить на практике основные принципы машинного перевода документов.

Ход работы:

№	Направление перевода	Предметная область
9	Англо-русский	Научные статьи по медицине,
		Критика предметов
		изобразительного искусства

- ✓ 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. №3 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.
 №1 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.

Код программы:

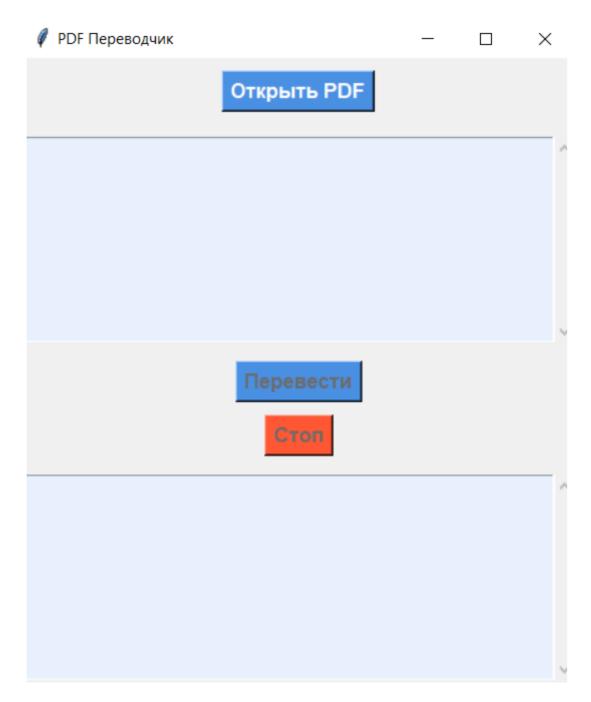
```
import tkinter as tk
from tkinter import filedialog, scrolledtext, messagebox
import fitz # PyMuPDF для обработки PDF
from deep_translator import GoogleTranslator # Используем deep_translator
import tempfile
import os
import threading
class PDFTranslator:
   def init (self, window):
       self.window = window
       self.window.title("PDF Переводчик")
       self.window.configure(bg="#F0F0F0")
        # Кнопка для открытия PDF
       self.open_button = tk.Button(window, text="Открыть PDF", command=self.load_pdf,
bg="#4A90E2", fg="white", font=("Arial", 12, "bold"))
        self.open button.pack(pady=10)
       # Область для отображения текста PDF
```

```
self.text area = scrolledtext.ScrolledText(window, wrap=tk.WORD, width=60,
height=10, bg="#E8F0FE", fg="#333333", font=("Arial", 10))
        self.text area.pack(pady=10)
        # Кнопка для начала перевода
        self.translate_button = tk.Button(window, text="Перевести",
command=self.start_translation, bg="#4A90E2", fg="white", font=("Arial", 12, "bold"))
        self.translate_button.pack(pady=5)
        self.translate button.config(state=tk.DISABLED)
        # Кнопка для остановки перевода
        self.stop button = tk.Button(window, text="Cτοπ", command=self.stop translation,
state=tk.DISABLED, bg="#FF5733", fg="white", font=("Arial", 12, "bold"))
        self.stop_button.pack(pady=5)
        # Область для отображения переведённого текста
        self.translated_area = scrolledtext.ScrolledText(window, wrap=tk.WORD, width=60,
height=10, bg="#E8F0FE", fg="#333333", font=("Arial", 10))
        self.translated_area.pack(pady=10)
        # Инициализация переводчика с deep_translator
        self.translator = GoogleTranslator(source='en', target='ru')
        self.temp_file = None
        self.stop_translation_flag = False
   def load pdf(self):
        file_path = filedialog.askopenfilename(filetypes=[("PDF Files", "*.pdf")])
       if file_path:
           try:
                document = fitz.open(file_path)
                self.temp_file = tempfile.NamedTemporaryFile(delete=False, mode='w+',
encoding='utf-8')
                for page in document:
                    page_text = page.get_text("text")
                    self.temp_file.write(page_text + '\n\n')
                self.temp_file.flush()
                self.temp_file.seek(0)
                # Очистка поля и отображение текста PDF
                self.text_area.delete(1.0, tk.END)
                self.text_area.insert(tk.END, self.temp_file.read())
                self.translated_area.delete(1.0, tk.END)
                self.translate_button.config(state=tk.NORMAL)
            except Exception as e:
                messagebox.showerror("Ошибка", f"He удалось открыть PDF: {e}")
    def start_translation(self):
        """Запуск перевода в отдельном потоке."""
        self.stop_translation_flag = False
        translation_thread = threading.Thread(target=self.perform_translation)
        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):
        """Остановка перевода."""
        self.stop translation flag = True
        self.stop_button.config(state=tk.DISABLED)
    def perform translation(self):
        """Процесс перевода текста PDF."""
        if self.temp_file:
            try:
                self.temp_file.seek(0)
                self.translated_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():
                        try:
                            translated_line = self.translator.translate(line.strip()) #
перевод с английского на русский
                            self.translated_area.insert(tk.END, translated_line + '\n\n')
                        except Exception as e:
                            self.translated_area.insert(tk.END, f"Ошибка перевода: {e}\n\n")
                        self.translated_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 error:
                messagebox.showerror("Ошибка", f"He удалось перевести текст: {error}")
                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):
        """Удаление временного файла при завершении программы."""
        if self.temp file:
           try:
                os.remove(self.temp_file.name)
            except Exception as e:
                print(f"He удалось удалить временный файл: {e}")
if __name__ == "__main__":
   main_window = tk.Tk()
    app_instance = PDFTranslator(main_window)
    main_window.mainloop()
```

Результат:

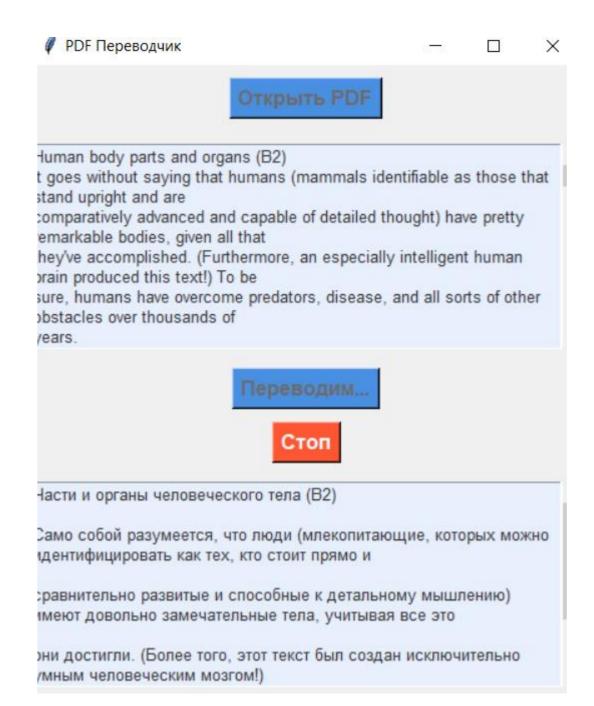
Исходное главное меню:



Выбран файл:

	_		×
Открыть PDF			
Topic: Applying machine learning techniques to an Abstract	nalyze big	data	-
n recent decades, machine learning has become big data analysis. This paper discusses the key machine learning techniques us			
analyze big data and their applications in various fields such as medicine, ecsciences. The main algorithms			
ncluding neural networks, decision trees and supp described, and the problems	oort vector	method ar	e ,
Перевести			
Стоп			
			,

Перевод:



Вывод: в ходе выполнения лабораторной работы освоилна практике основные принципы машинного перевода документов.