Министерство образования Республики Беларусь

Учреждение образования

«Брестский государственный технический университет»

Кафедра ИИТ

Лабораторная работа №1

За шестой семестр

По дисциплине “ЕЯИИС”

Тема: «Разработка автоматизированной системы формирования словаря естественного языка»

Выполнил:

Студент 3 курса

Группы ИИ-16

Пилипук З.А.

Проверил:

Крапивин Ю.Б.

Брест, 2021

**Цель работы** – освоить принципы разработки прикладных сервисных программ для подготовки исходных данных естественно-языковых систем.

**Вариант 1:** 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.

**Текст программы:**

import PySimpleGUI as sg

from collections import Counter

import pymorphy2

def getLexAndForms(text):

morph = pymorphy2.MorphAnalyzer()

words = text.lower().replace(',', '').replace('.', '').replace(' - ', ' ').split()

normal = []

for word in words:

normal.append(morph.parse(word)[0].normal\_form)

return sorted(list(set(words))), sorted(Counter(normal).items())

sg.theme('DarkAmber')

col = [

[sg.Multiline(size=(30, 5), key='textbox', enable\_events = True)],

[sg.Button('Analyze'), sg.Button('Analyze in.txt'), sg.Button('MORPH INFO'), sg.Button('ABOUT')]

]

layout = [

[sg.Listbox(values=[], size=(20, 12), key='-FORMS-', enable\_events=True), sg.Listbox(values=[], size=(20, 12), key='-LEX-', enable\_events=True)],

[sg.Column(col)],

]

window = sg.Window('Словоформы и лексемы', layout)

addInfo = {}

while True: # Event Loop

event, values = window.read()

if event in (sg.WIN\_CLOSED, 'Exit'):

break

if event == 'Analyze in.txt' or event == 'Analyze':

words\_form, lexems = [], []

if (event == 'Analyze in.txt'):

file = open('in.txt')

fileText = file.read()

words\_form, lexems = getLexAndForms(fileText)

else:

words\_form, lexems = getLexAndForms(values['textbox'])

window.FindElement('-FORMS-').Update(values=words\_form)

window.FindElement('-LEX-').Update(values=lexems)

addInfo = dict.fromkeys(words\_form)

if event == 'MORPH INFO':

if values['-FORMS-']:

add = [

[sg.Input()],

[sg.Button('Enter')]

]

layout2 = [

[sg.Text('Словоформа: {}'.format(values['-FORMS-'][0]))],

[sg.Text('Дополнительные сведения: {}'.format(addInfo[values['-FORMS-'][0]]), key='INF')],

[sg.Column(add)]]

win2 = sg.Window('Доп.Инфо', layout2)

while True:

event2, vals2 = win2.Read()

if event2 == 'Enter':

addInfo[values['-FORMS-'][0]] = vals2[0]

win2.FindElement('INF').Update('Дополнительные сведения: {}'.format(vals2[0]))

if event2 in (sg.WIN\_CLOSED, 'Exit'):

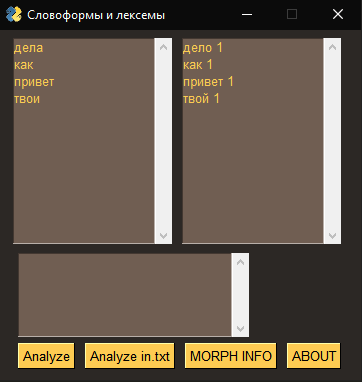
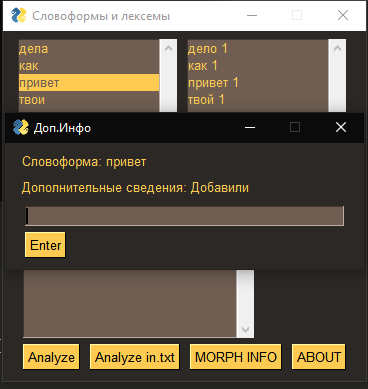
break

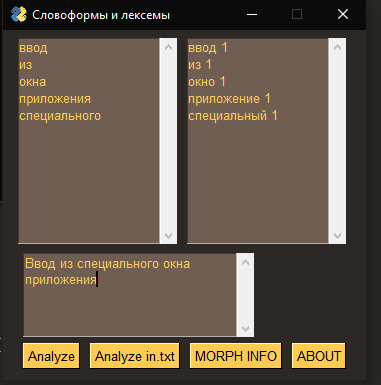
if event == 'ABOUT':

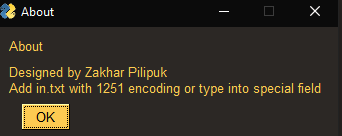
sg.Popup('About', 'Designed by Zakhar Pilipuk\nAdd in.txt with 1251 encoding or type into special field')

window.close()

**Результат:**





**Вывод:** освоили принципы разработки прикладных сервисных программ для подготовки исходных данных естественно-языковых систем.