

Министерство образования Республики Беларусь

Учреждение образования
“Белорусский государственный университет
информатики и радиоэлектроники”

Факультет информационных технологий и управления
Кафедра интеллектуальных информационных технологий

Отчёт по лабораторной работе №1
по дисциплине «Естественно-языковой интерфейс интеллектуальных
систем»

Выполнили
Маевский В.Г
121701

Проверил Крапивин Ю. Б.

Минск 2024

Цель работы:

Освоить принципы разработки прикладных сервисных программ для решения задачи автоматического лексического и лексико-грамматического анализа текста естественного языка.

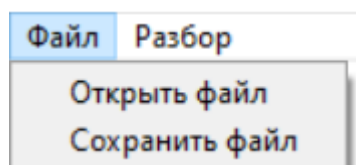
Задание:

Список слов, упорядоченный по алфавиту и включающий только лексемы с дополнительно оформленными записями о месте и роли данного слова в составе предложения. К такой информации относится описание того, каким членом предложения может быть данное слово и в какой форме (падеж, число, время и т.п.). Например, если это существительное в именительном падеже, то оно может выступать в роли подлежащего; если это существительное в родительном падеже, то оно может быть дополнением; если это прилагательное, то оно может быть определением и т.п.

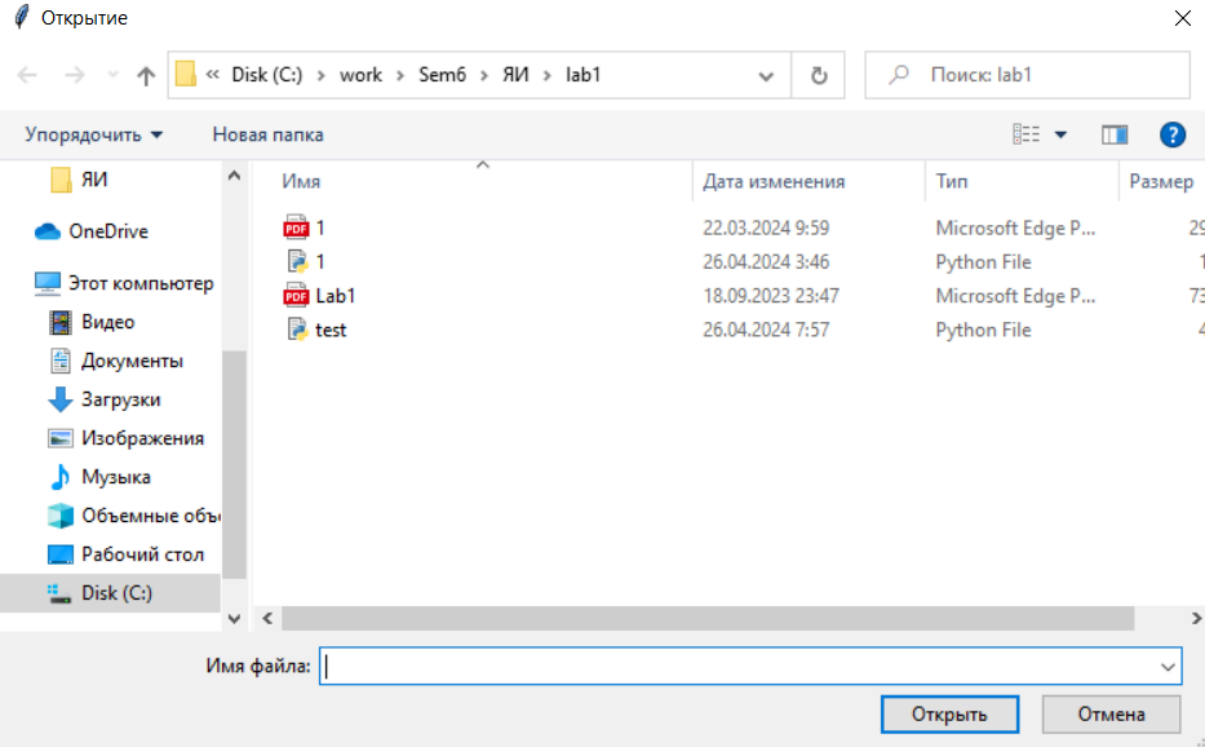
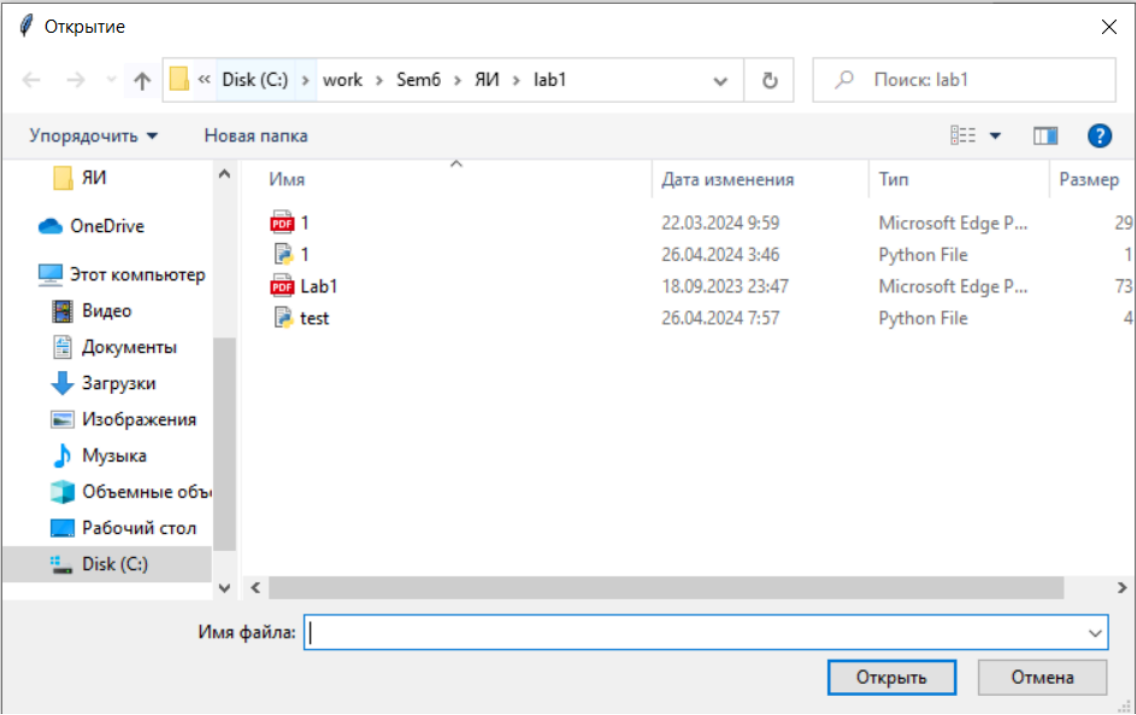
Вариант:

Русский	PDF	Задание 4
---------	-----	-----------

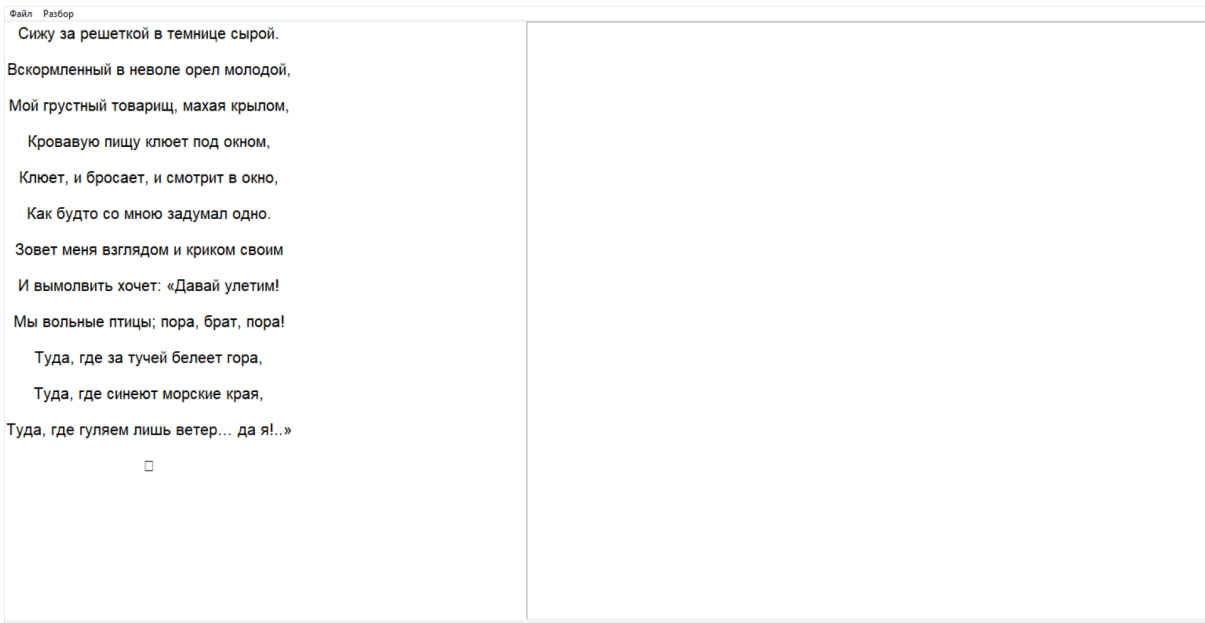
Интерфейс



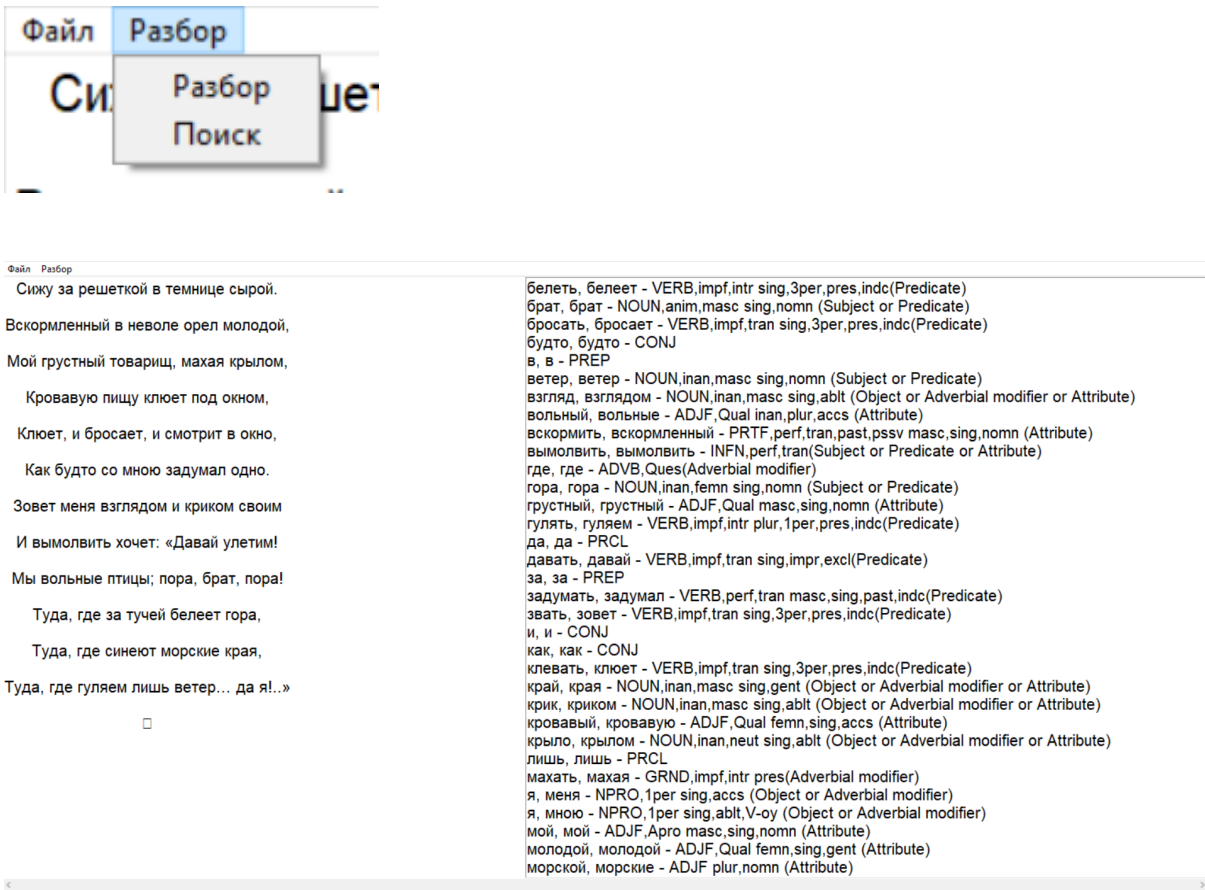
Раздел Файл отвечает за такие процессы как загрузка и сохранение файлов:



При выборе нужного файла в левой части приложения отображается искомый текст:



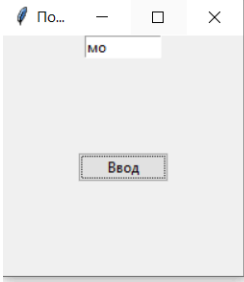
Далее нам следует обратиться к разделу Разбор:



Вариант разбор разбивает искомый текст на лексемы, сохраняя и описывая форму лексемы, а именно проводит морфологический разбор и на основе полученных данных определяет возможные роли слова в предложении.

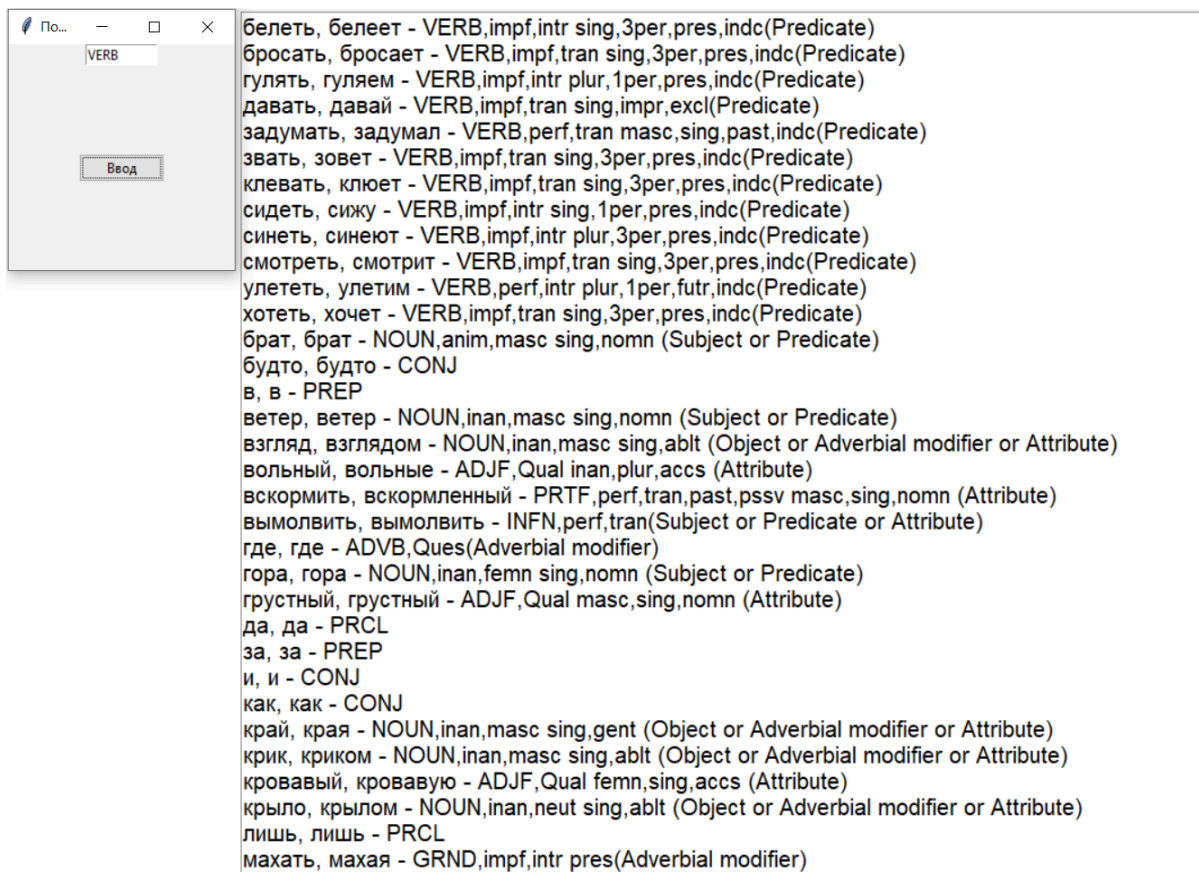
Несмотря на то, что полученный список упорядочен, поиск нужной лексемы может занимать некоторое время, поэтому для ускорения получения результата был реализован поиск нужной лексемы(группы лексем)

Допустим мы хотим найти в первую очередь лексемы, которые имеют в составе слог -мо:



смотреть, смотрит - VERB,impf,tran sing,3per,pres,indc(Predicate)
морской, морские - ADJF plur,nomn (Attribute)
молодой, молодой - ADJF,Qual femn,sing,gent (Attribute)
мой, мой - ADJF,Apro masc,sing,nomn (Attribute)
вымолвить, вымолвить - INFN,perf,tran(Subject or Predicate)
белеть, белеет - VERB,impf,intr sing,3per,pres,indc(Predicate)
брат, брат - NOUN,anim,masc sing,nomn (Subject or Predicate)
бросать, бросает - VERB,impf,tran sing,3per,pres,indc(Predicate)
будто, будто - CONJ
в, в - PREP
ветер, ветер - NOUN,inan,masc sing,nomn (Subject or Predicate)
взгляд, взглядом - NOUN,inan,masc sing,abl (Object or Adverbial modifier or Attribute)
вольный, вольные - ADJF,Qual inan,plur,accs (Attribute)
вскормить, вскормленный - PRTF,perf,tran,past,pssv masc,sing,nomn (Attribute)
где, где - ADVB,Ques(Adverbial modifier)
гора, гора - NOUN,inan,femn sing,nomn (Subject or Predicate)
грустный, грустный - ADJF,Qual masc,sing,nomn (Attribute)
гулять, гуляем - VERB,impf,intr plur,1per,pres,indc(Predicate)
да, да - PRCL
давать, давай - VERB,impf,tran sing,impr,excl(Predicate)
за, за - PREP
задумать, задумал - VERB,perf,tran masc,sing,past,indc(Predicate)
звать, зовет - VERB,impf,tran sing,3per,pres,indc(Predicate)
и, и - CONJ
как, как - CONJ
клевать, клюет - VERB,impf,tran sing,3per,pres,indc(Predicate)
край, края - NOUN,inan,masc sing,gent (Object or Adverbial modifier or Attribute)
крик, криком - NOUN,inan,masc sing,abl (Object or Adverbial modifier or Attribute)
кровавый, кровавую - ADJF,Qual femn,sing,accs (Attribute)
крыло, крылом - NOUN,inan,neut sing,abl (Object or Adverbial modifier or Attribute)
лишь, лишь - PRCL
махать, махая - GRND,impf,intr pres(Adverbial modifier)
я, меня - NPRO,1per sing,accs (Object or Adverbial modifier)

Или же мы хотим вывести в первую очередь все слова, что являются глаголами:



Описание алгоритма:

1. Основной алгоритм обработки текста

```
def transfer():
    l1.configure(state=NORMAL)
    l1.delete('1.0', END)
    text = l2['text']
    text = re.split(r'\s+|[,;?!.-><«»\s*]', text.lower())
    text = list(set(text))
    text.sort()
    text.pop(0)

    for word in text:
        morph_form = morph_work(word)
        l1.insert(END, morph_form[0] + ' - ' + word + ' - ' + morph_form[1] + '\n')
    l1.configure(state=DISABLED)
```

В этой функции удаляем из текста все ненужные символы, после чего разделяет текст на слова и для каждого слова проводит морфологический разбор с определением возможной роли слова в предложении.

```

17 def morph_work(word):
18     form = morph.parse(word)[0]
19     form2 = str(form[1])
20     if 'NOUN' in form2:
21         if 'nomn' in form2:
22             form2+=' (Subject or Predicate)'
23         else:
24             form2+=' (Object or Adverbial modifier or Attribute)'
25     elif 'ADJF' in form2 or 'PRTF' in form2:
26         form2+=' (Attribute)'
27     elif 'ADJS' in form2 or 'PRTS' in form2 or 'PRED' in form2:
28         form2+=' (Predicate or Attribute)'
29     elif 'COMP' in form2:
30         form2+=' (Adverbial modifier or Attribute)'
31     elif 'INFN' in form2:
32         form2+=' (Subject or Predicate or Attribute)'
33     elif 'VERB' in form2 :
34         form2+=' (Predicate)'
35     elif 'NPRO' in form2:
36         if 'nomn' in form2:
37             form2+=' (Subject or Predicate)'
38         else:
39             form2+=' (Object or Adverbial modifier)'
40     elif 'ADVB' in form2 or 'GRND' in form2:
41         form2+=' (Adverbial modifier)'
42     form1 = form[2]
43     return form1.form2

```

Эта функция подробно описывает процесс МР

```

def start(text):
    l1.configure(state=NORMAL)
    inputs_1 = l1.get('1.0', END)
    lines = inputs_1.splitlines()
    count = 0
    for line in lines:
        count+=1
        if text in line:
            buff = line
            lines.remove(line)
            lines.insert(0,buff)
    l1.delete('1.0',END)
    for line in lines:
        l1.insert(END,line + '\n')
    l1.configure(state=DISABLED)

```

Эта функция реализует возможность поиска нужного слова

Тест производительности

Текст из 69 слов:

время получения содержимого файла - 26 мс

время работы алгоритма - 8 мс

Текст из 2655 слов:

время получения содержимого файла - 19 мс

время работы алгоритма - 784 мс

Вывод:

Данная программа представляет собой инструмент для создания и управления списком слов, упорядоченных по алфавиту, с включением лексем, содержащих дополнительные записи. В каждой записи хранятся слова, их характеристики и возможная роль в предложении. Программа обеспечивает эффективное создание и организацию словаря, что делает ее полезным инструментом для лингвистических исследований, обработки текстов и работы с языком и в образовательных целях.