

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 №2
по дисциплине «Естественно-языковой интерфейс интеллектуальных
систем»

Выполнили
Маевский В.Г
121701

Проверил Крапивин Ю. Б.

Минск 2024

Цель работы:

Изучить принципы построения корпусов текстов, виды разметки и способы аннотирования, инструменты работы с корпусами текстов.

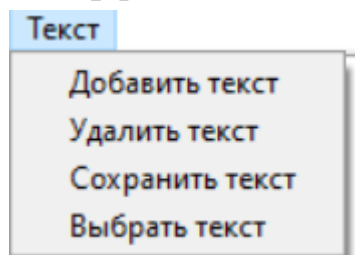
Построить корпус текстов и разработать корпусный менеджер.

Задание:

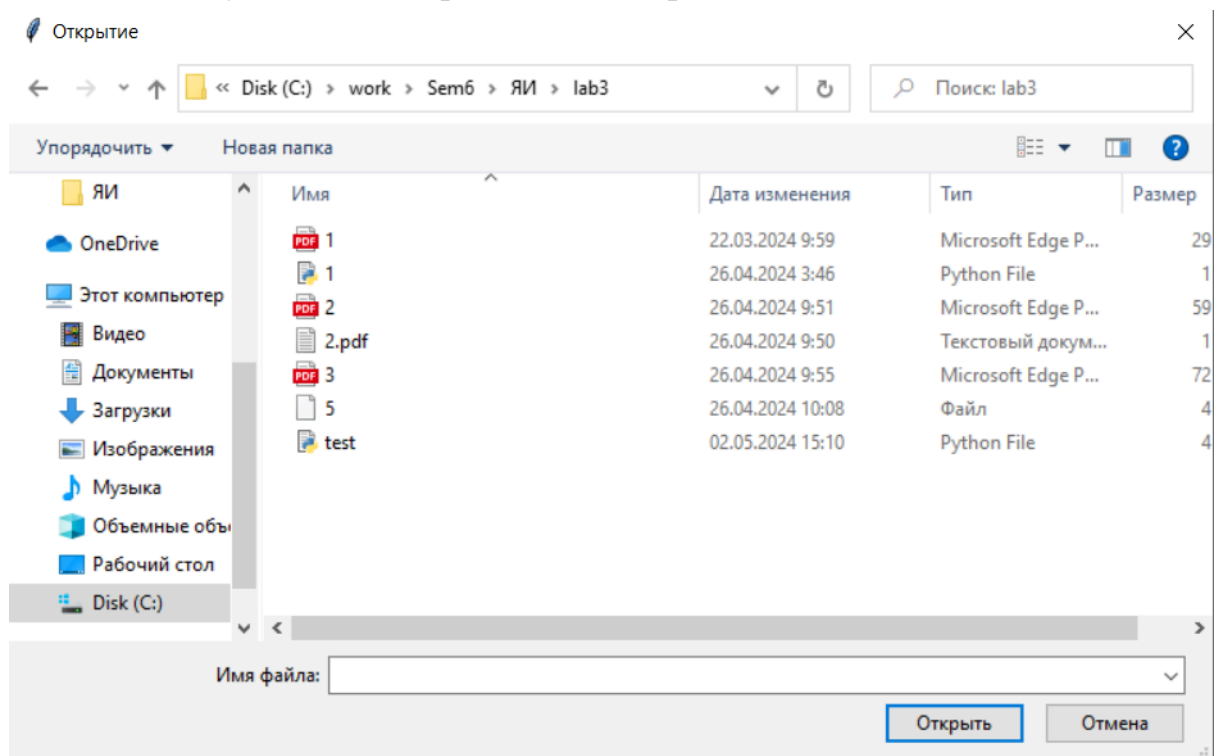
1. Сформировать электронный корпус текстов по выбранной предметной области.

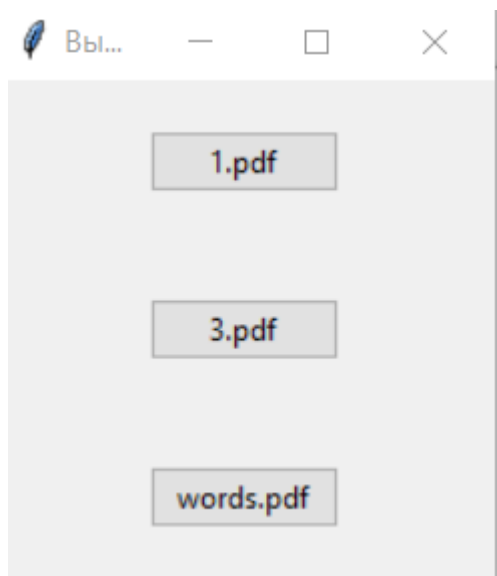
2. Используя результаты лабораторной работы №1 (возможность получения лингвистических сведений для произвольной леммы естественного языка) разработать корпусный менеджер, обеспечивающий базовую функциональность работы с созданным корпусом текстов

Интерфейс

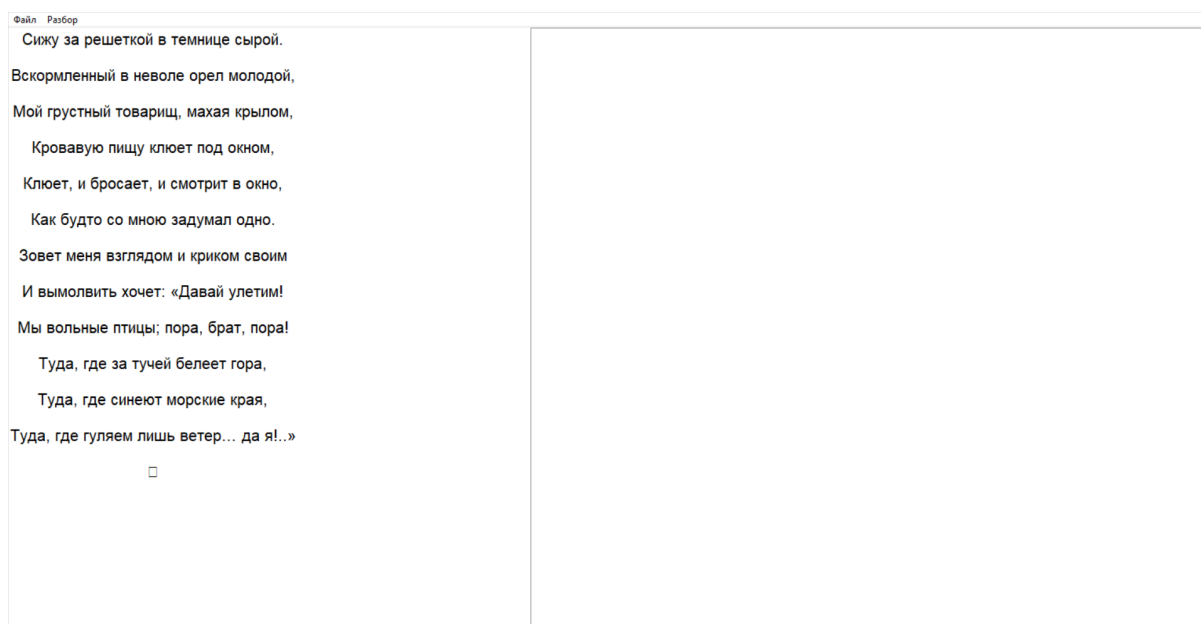


Раздел Текст отвечает за такие процессы управления текстами, такими как : добавление, удаление, сохранение, выбор

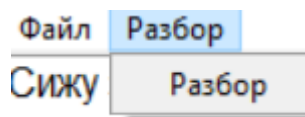




При выборе нужного текста в левой части приложения отображается его содержимое:



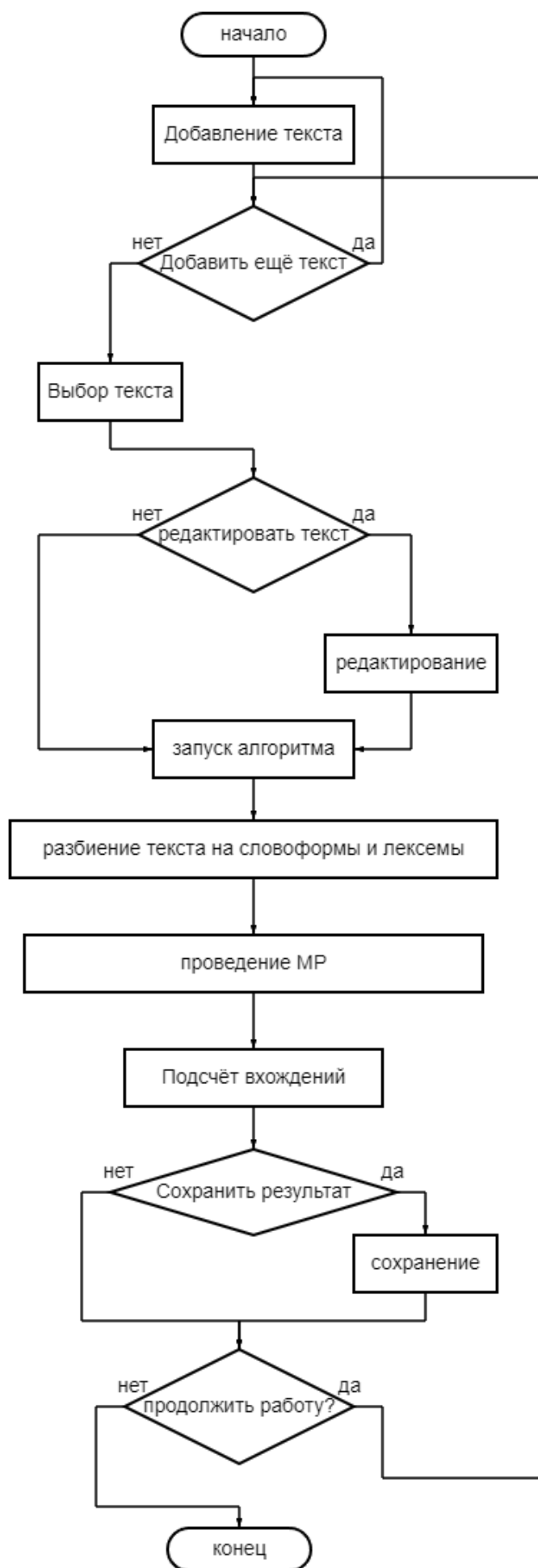
Далее нам следует обратиться к разделу Разбор:



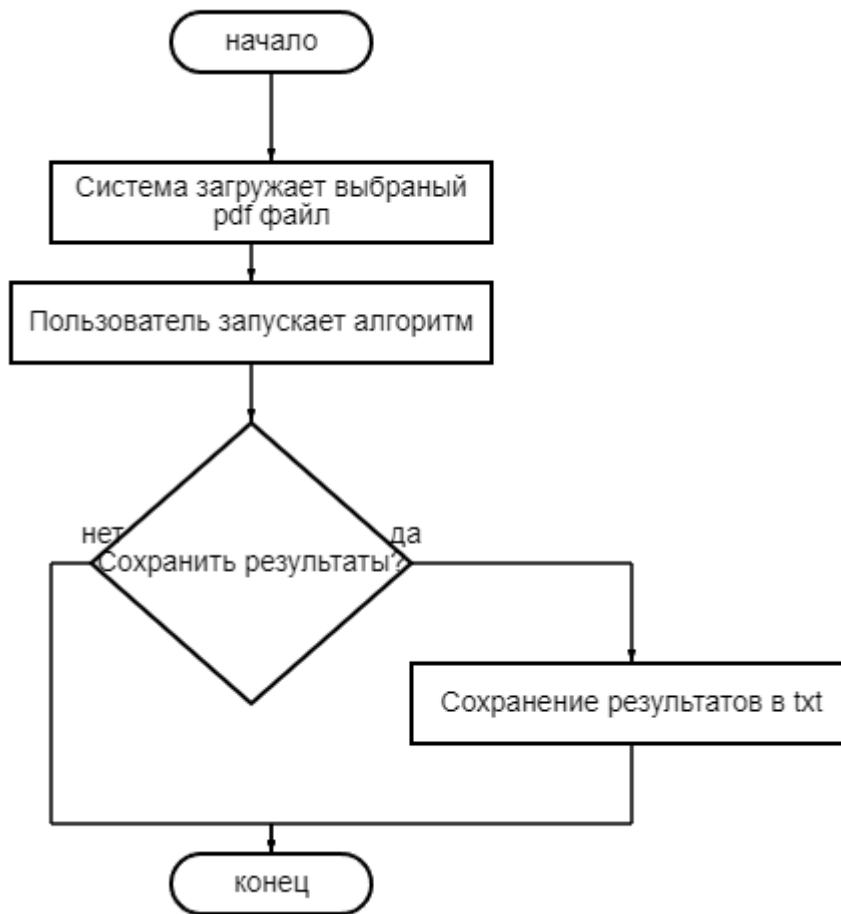
Файл	Разбор	Текст
Сижу	Разбор	
Сижу за решеткой в темнице сырой.	белеет , lex = 'белеть' , gr = VERB,impf,intr sing,3per,pres,indc , count = 1	
Вскормленный в неволе орел молодой,	брат , lex = 'брат' , gr = NOUN,anim,masc sing,nomn , count = 1	
Мой грустный товарищ, махая крылом,	бросает , lex = 'бросать' , gr = VERB,impf,tran sing,3per,pres,indc , count = 1	
Кровавую пищу клюет под окном,	будто , lex = 'будто' , gr = CONJ , count = 1	
Клюет, и бросает, и смотрит в окно,	в , lex = 'в' , gr = PREP , count = 1	
Как будто со мною задумал одно.	ветер , lex = 'ветер' , gr = NOUN,inan,masc sing,nomn , count = 1	
Зовет меня взглядом и криком своим	взглядом , lex = 'взгляд' , gr = NOUN,inan,masc sing,abl , count = 1	
И вымолвить хочет: «Давай улетим!	воплые , lex = 'воплый' , gr = ADJF,Qual inan,plur,accs , count = 1	
Мы вольные птицы; пора, брат, пора!	вскормленный , lex = 'вскормить' , gr = PRTF,perf,tran,past,pssv masc,sing,nomn , count = 1	
Туда, где за тучей белеет гора,	вымолвить , lex = 'вымолвить' , gr = INFN,perf,tran , count = 1	
Туда, где синюют морские края,	где , lex = 'где' , gr = ADVB,Ques , count = 1	
Туда, где гуляем лишь ветер... да я!..»	гора , lex = 'гора' , gr = NOUN,inan,femn sing,nomn , count = 1	
	грустный , lex = 'грустный' , gr = ADJF,Qual masc,sing,nomn , count = 1	
	гуляем , lex = 'гулять' , gr = VERB,impf,intr plur,1per,pres,indc , count = 1	
	да , lex = 'да' , gr = PRCL , count = 1	
	давай , lex = 'давать' , gr = VERB,impf,tran sing,impf,excl , count = 1	
	за , lex = 'за' , gr = PREP , count = 1	
	задумал , lex = 'задумать' , gr = VERB,perf,tran masc,sing,past,indc , count = 1	
	зовет , lex = 'звать' , gr = VERB,impf,tran sing,3per,pres,indc , count = 1	
	и , lex = 'и' , gr = CONJ , count = 1	
	как , lex = 'как' , gr = CONJ , count = 1	
	клюет , lex = 'клевать' , gr = VERB,impf,tran sing,3per,pres,indc , count = 1	
	края , lex = 'край' , gr = NOUN,inan,masc sing,gent , count = 1	
	криком , lex = 'крик' , gr = NOUN,inan,masc sing,abl , count = 1	
	кровавую , lex = 'кровавый' , gr = ADJF,Qual femn,sing,accs , count = 1	
	крылом , lex = 'крыло' , gr = NOUN,inan,neut sing,abl , count = 1	
	лишь , lex = 'лишь' , gr = PRCL , count = 1	
	махая , lex = 'махать' , gr = GRND,impf,intr pres , count = 1	
	меня , lex = 'я' , gr = NPRO,1per sing,accs , count = 3	
	мною , lex = 'я' , gr = NPRO,1per sing,abl,V-oy , count = 3	
	мой , lex = 'мой' , gr = ADJF,Apro masc,sing,nomn , count = 1	
	молодой , lex = 'молодой' , gr = ADJF,Qual femn,sing,gent , count = 1	
	морские , lex = 'морской' , gr = ADJF plur,nomn , count = 1	
	мы , lex = 'мы' , gr = NPRO,1per plur,nomn , count = 1	
	неволе , lex = 'неволя' , gr = NOUN,inan,femn sing,loct , count = 1	
	одно , lex = 'один' , gr = ADJF,Apro,Anum neut,sing,accs , count = 1	
	, lex = 'окно' , gr = NOUN,inan,neut sing,accs , count = 2	

Вариант разбор разбивает искомый текст на словоформы, находит их лексемы, делает морфологический разбор и подсчитывает кол-во встреченных лексем в тексте.

Описание алгоритма:



Структура данных для хранения входной и выходной информации



Тест производительности

Текст из 69 слов:

время получения содержимого файла - 27 мс

время работы алгоритма - 89 мс

Текст из 2655 слов:

время получения содержимого файла - 810 мс

время работы алгоритма - 2699 мс

Вывод:

Данная программа представляет собой инструмент для автоматического разбиения текста на словоформы, лексемы и их морфологического разбора. Что делает ее полезным инструментом для лингвистических исследований, обработки текстов и работы с языком и в образовательных целях.