

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»

Выполнил
Пашин Н.А.
студент группы
121701

Проверил Крапивин Ю. Б.

Минск 2024

Цель работы:

Разработать текстовый корпус и его менеджер.

Задача:

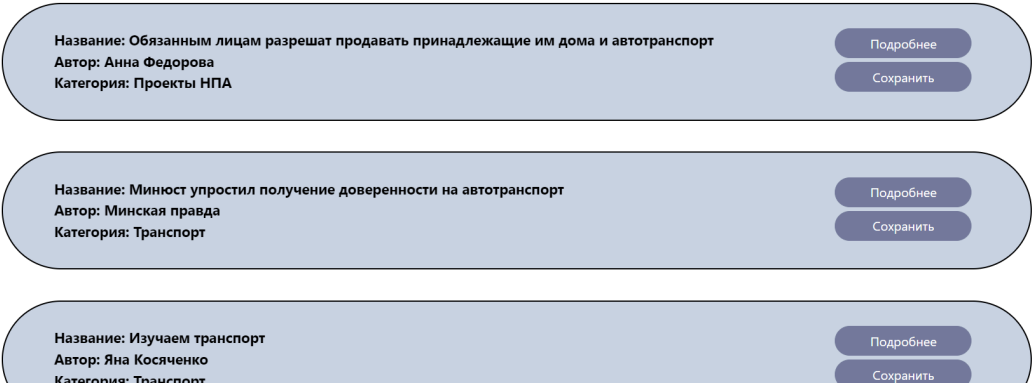
Предметная область: “Транспорт”.
Язык: русский.

Интерфейс

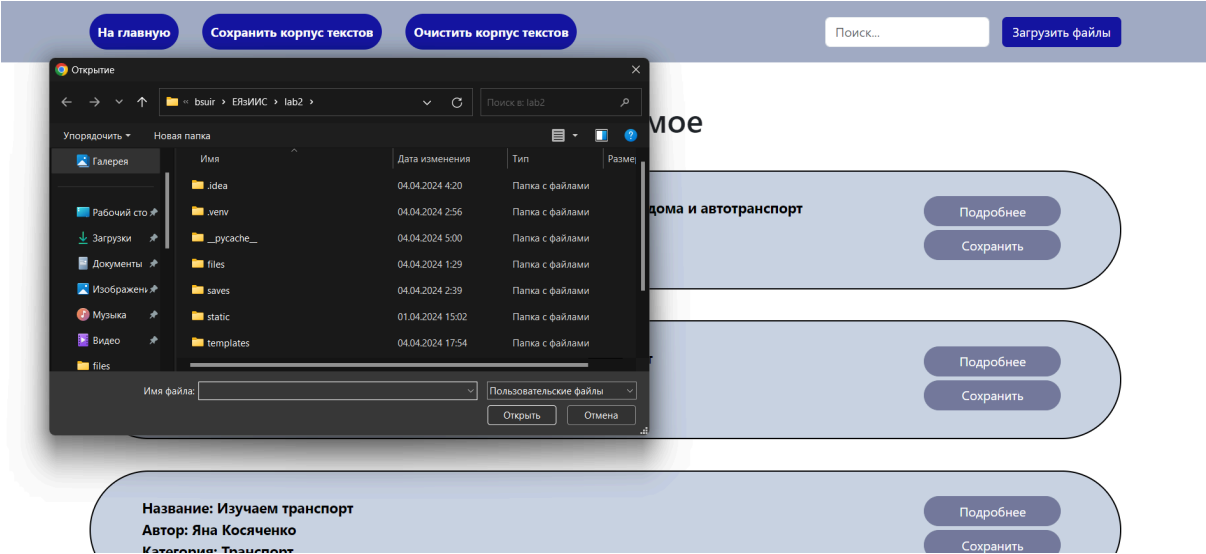
Главная страница



Содержимое



Окно загрузки файла



Страница текста

На главную

Сохранить корпус текстов

Очистить корпус текстов

Поиск...

Загрузить файлы

Название: Обязанным лицам разрешат продавать принадлежащие им дома и автотранспорт

Автор: Анна Федорова

Категория: Проекты НПА

Показать текст

Показать разметку

Сохранить разметку

Просмотр слов

Поправки в Кодекс о браке и семье предусматривают возможность отчуждения принадлежащих обязанным лицам недвижимости и транспорта, уточняют ряд норм, связанных с порядком изъятия детей из неблагополучных семей и взысканием с родителей расходов на ребенка, находящегося на гособеспечении.

Законопроект «Об изменении законов по вопросам регулирования брачных и семейных отношений» 29.11.2023 внесен Совмином в Палату представителей. Документом предлагается корректировка нескольких нормативных правовых актов, включая Кодекс о браке и семье (далее – Кодекс).

В числе планируемых новшеств – установление исключения из общего правила о запрете отчуждения имущества, принадлежащего обязанному лицу. Напомним, что обязанными считаются лица, в отношении которых вступило в законную силу решение суда о взыскании расходов, затраченных государством на содержание их детей, находящихся (находившихся) на государственном обеспечении. Сегодня таким лицам запрещается отчуждать принадлежащее им недвижимое имущество, подлежащее госрегистрации, и транспортные средства (п. 12 Декрета от 24.11.2006 № 18 «О дополнительных мерах по государственной защите детей в неблагополучных семьях»).

Законодатели предлагают включить в Кодекс новую ст. 933, предусматривающую изъятие из этого правила. Речь идет о недвижимом имуществе, подлежащем госрегистрации, включая недвижимое имущество, являющееся общей совместной собственностью супругов, а также доли в праве собственности на недвижимое имущество (если оно находится в общей долевой собственности) и транспортное средство. Согласно ст. 933 обязанным

На главную

Сохранить корпус текстов

Очистить корпус текстов

Поиск...

Загрузить файлы

Название: Обязанным лицам разрешат продавать принадлежащие им дома и автотранспорт

Автор: Анна Федорова

Категория: Проекты НПА

Показать текст

Показать разметку

Сохранить разметку

Просмотр слов

```
<?xml version="1.0" encoding="utf-8"?><text>
<p>
<s>
<w>Поправки <ana lemma="поправка" pos="СУЩ" грам="неод,жр мн,им" /></w>
<w>в <ana lemma="в" pos="пр" грам="" /></w>
<w>Кодекс <ana lemma="кодекс" pos="СУЩ" грам="неод,мр ед,вн" /></w>
<w>о <ana lemma="о" pos="пр" грам="" /></w>
<w>браке <ana lemma="брак" pos="СУЩ" грам="неод,мр ед,пр" /></w>
<w>и <ana lemma="и" pos="СОЮЗ" грам="" /></w>
<w>семье <ana lemma="семья" pos="СУЩ" грам="неод,жр ед,пр" /></w>
<w>предусматривают <ana lemma="предусматривать" pos="ГЛ" грам="несов,перех мн,3л,наст,изъяв" /></w>
<w>возможность <ana lemma="возможность" pos="СУЩ" грам="неод,жр ед,вн" /></w>
</s>
</p>
</text>
```

Окно поиска

На главнуюСохранить корпус текстовОчистить корпус текстов

Поиск...Загрузить файлы

Название: Обязанным лицам разрешат продавать принадлежащие им дома и а

Автор: Анна Федорова

Категория: Проекты НПА

Показать текстПоказать разметкуСохранить разметкуПросмотр слов

<

```
<?xml version="1.0" encoding="utf-8"?><text>
<p>
<s>
<w>Исправки <ana lemma="поправка" pos="СУЩ" gram="неод,жр мн,им" /></w>
<w>в <ana lemma="в" pos="ПР" gram="" /></w>
<w>Кодекс <ana lemma="кодекс" pos="СУЩ" gram="неод,мр ед,вн" /></w>
<w>о <ana lemma="о" pos="ПР" gram="" /></w>
<w>браке <ana lemma="брак" pos="СУЩ" gram="неод,мр ед,пр" /></w>
<w>и <ana lemma="и" pos="СОЮЗ" gram="" /></w>
<w>семье <ana lemma="семья" pos="СУЩ" gram="неод,жр ед,пр" /></w>
<w>предусматривают <ana lemma="предусматривать" pos="ГЛ" gram="несов,перех мн,зл,наст,изъяв" /></w>
<w>возможность <ana lemma="возможность" pos="СУЩ" gram="неод,жр ед,вн" /></w>
```

Обязанным лицам разрешат продавать принадлежащие им дома и а
Минюст упростил получение доверенности на автотранспорт
Изучаем транспорт

Окно сохранения корпуса текстов

На главнуюСохранить корпус текстовОчистить корпус текстов

Поиск...Загрузить файлы

Введите название корпуса текстов

Сохранить

Главная страница слов

На главную

Сохранить корпус текстов

Очистить корпус текстов

Поиск...

Загрузить файлы

Вернуться

Все слова в тексте

а

Подробнее

администрация

Подробнее

акт

Подробнее

алименты

Подробнее

Страница одного определенного слова

На главную

Сохранить корпус текстов

Очистить корпус текстов

Поиск...

Загрузить файлы

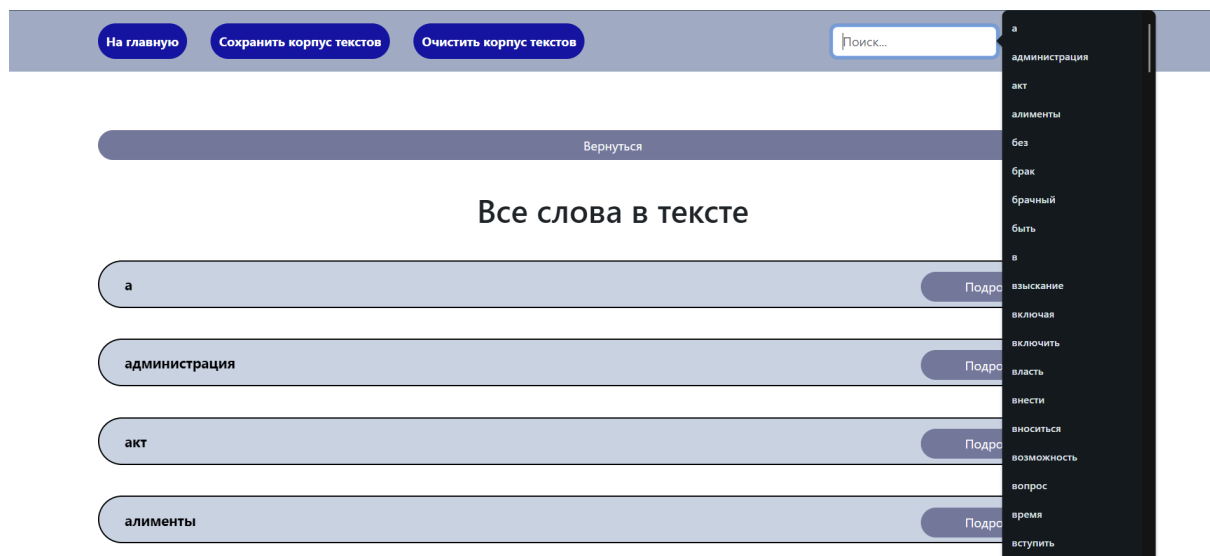
администрация

администрации

Часть речи
Имя существительное
Одушевленность
Неодушевленное
Падеж
Родительный
Род
Женский
Число
Единственное

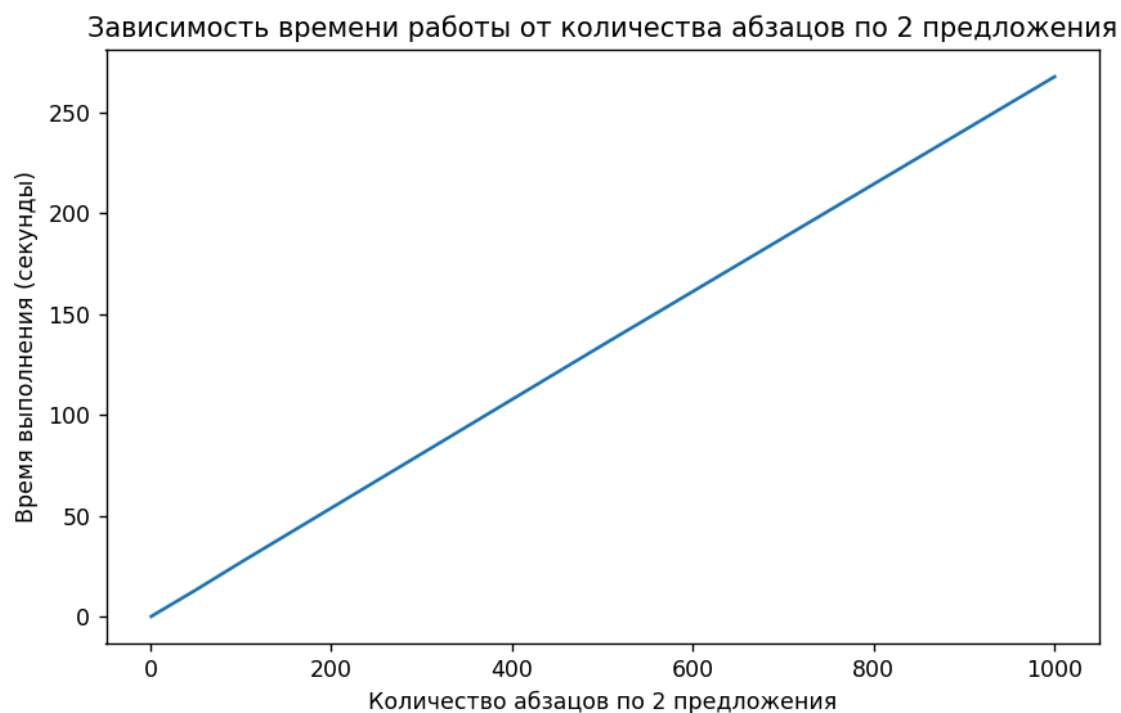
Вернуться

Окно поиска слов



Тест производительности

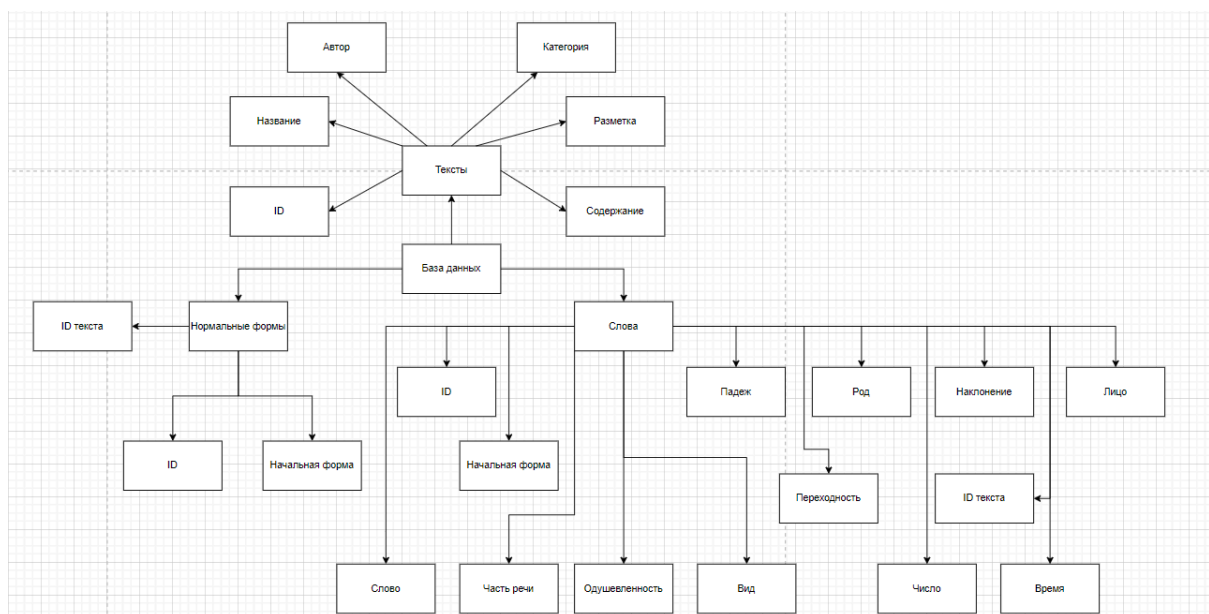
Тест производительности, приведенный на рисунке ниже, показывает время обработки текстов для разбиения, нахождения морфологических признаков и создания разметки. Абзац выглядит следующим образом “Это первое предложение. А это второе.”.



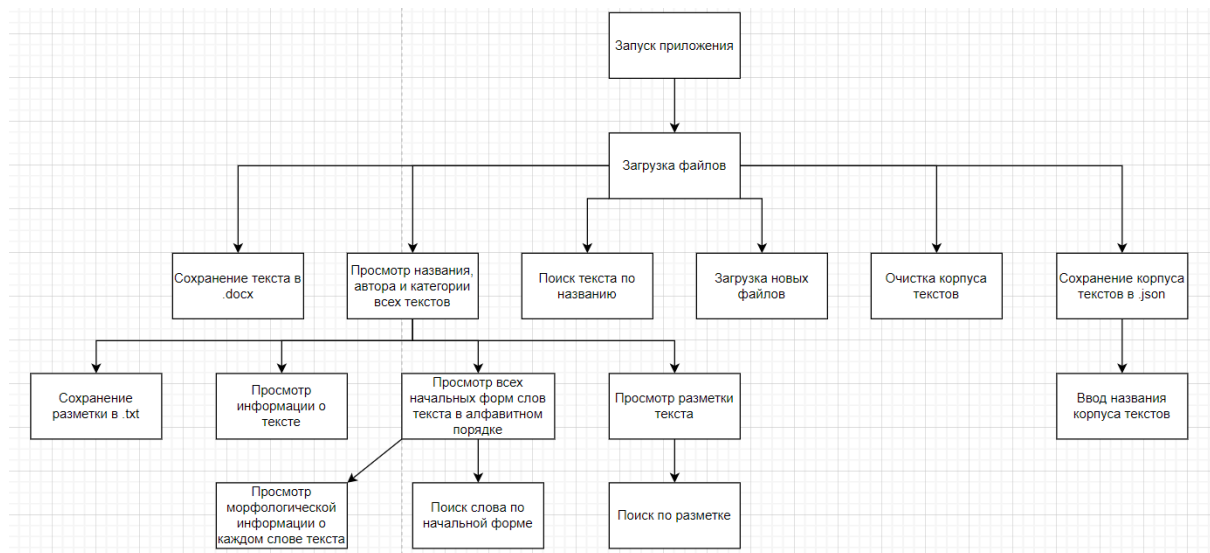
Образец разметки

```
<?xml version="1.0" encoding="windows-1251" ?> <text> <p>
<s>
<w>Звонили<ana lemma="ЗВОНИТЬ" pos="Г" gram="мн,нс,нп,дст,прш," /></w>
<w>к<ana lemma="к" pos="ПРЕДЛ" gram="" /></w>
<w>вечерне
<ana lemma="ВЕЧЕРНЯ" pos="С" gram="жр,ед,дт,пр,но," />
<ana lemma="ВЕЧЕРНИЙ" pos="П" gram="ср,ед,кр," /></w>
<pun>.</pun> </s>
<s><w>Торжественный<ana lemma="ТОРЖЕСТВЕННЫЙ" pos="П" gram="мр,ед,им,вн,"
/></w>
<w>гул<ana lemma="ГУЛ" pos="С" gram="мр,ед,им,вн,но," /></w>
<w>колоколов
<ana lemma="КОЛОКОЛ" pos="С" gram="мр,фам,ед,им,од," />
<ana lemma="КОЛОКОЛОВ" pos="С" gram="мр,мн,рд,но," /></w>
.....<pun>.</pun> </s></p></text>
```

Структура хранения данных



Структурная схема приложения



Описание алгоритмов

Добавление информации в БД : Начало — получить файлы — для каждого файла разобрать содержимое —определить название, автора, категорию и содержимое — создать разметку — добавить в таблицу texts — разобрать содержимое на слова — определить морфологические признаки — добавить слово в таблицу word_table — повторить последние 2 шага для оставшихся слов — Конец

Сохранение корпуса текстов: Начало — получить название корпуса текстов — получить все тексты из БД — создать файл .json — разобрать информацию и добавить в .json — выгрузить файл .json — Конец

Сохранение текста: Начало — получить текст с таким id из БД — создать файл .docx — собрать текст и добавить в файл .docx — выгрузить файл .docx — Конец

Сохранение текста: Начало — получить разметку текста с таким id текста из БД — создать файл .txt — добавить разметку в файл .txt — выгрузить файл .txt — Конец

Поиск по подстроке слов: Начало — получить начальную форму слова — найти слово в БД — получить информацию о всех словах с такой начальной формой — вывести пользователю информацию — Конец

Поиск по подстроке текстов: Начало — получить название текста— найти текст в БД — получить информацию о тексте с таким названием — вывести пользователю информацию — Конец

Поиск по разметке: Начало — получить искомое значение— найти в тексте — выделить при помощи `` — вывести пользователю информацию — Конец

Вывод:

В результате выполнения данной лабораторной работы мы изучили понятие текстового корпуса и его менеджера. Текстовый корпус представляет собой набор текстовых документов, собранных для анализа и обработки в рамках определенной задачи или исследования. Он является важным инструментом в области обработки естественного языка, машинного обучения и лингвистического анализа.