

**ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ  
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ  
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»**

Факультет компьютерных наук  
Департамент программной инженерии

**СОГЛАСОВАНО**

Доцент департамента программной  
инженерии факультета компьютерных наук,  
канд. техн. наук.



Х.М. Салех

«15» 05 2023 г.

**УТВЕРЖДАЮ**

Академический руководитель  
образовательной программы  
«Программная инженерия»  
профессор департамента программной  
инженерии, канд. техн. наук

В.В. Шилов

«13» 05 2023 г.

**УЛУЧШЕНИЕ ПОЛЬЗОВАТЕЛЬСКОГО ИНТЕРФЕЙСА РУССКОГО УЧЕБНОГО  
КОРПУСА**

**Текст программы**

**ЛИСТ УТВЕРЖДЕНИЯ**

**RU.17701729.09.11-01 12 01-1-ЛУ**

Исполнитель

студент группы БПИ 217



/ А.В. Цейтин/

«13» 05 2023 г.

**Москва 2023**

Подп. и дата	
Инв. № дубл.	
Взам. инв. №	
Подп. и дата	
Инв. № подл	

УТВЕРЖДЕН  
RU.17701729.09.11-01 12 01-1-ЛУ

**УЛУЧШЕНИЕ ПОЛЬЗОВАТЕЛЬСКОГО ИНТЕРФЕЙСА РУССКОГО УЧЕБНОГО  
КОРПУСА**

**Текст программы**

**RU.17701729.09.11-01 12 01-1**

**Листов 9**

<i>Подп. и дата</i>	
<i>Инв. № дубл.</i>	
<i>Взам. инв. №</i>	
<i>Подп. и дата</i>	
<i>Инв. № подл</i>	

**Москва 2023**

### АННОТАЦИЯ

В данном документе приведены ссылки на код, выполняющий следующие задачи:

1. Получения данных из базы данных
2. Хранение данных в базе данных (Django модели)
3. Токенизация и морфологический разбор текстов с применением машинного обучения
4. Сбор, обработка, отображение и визуализация статистических данных проекта
5. Анимирование статических элементов страницы

Среда разработки – PyCharm.

Настоящий документ разработан в соответствии с требованиями:

- 1) ГОСТ 19.101-77 Виды программ и программных документов;
- 2) ГОСТ 19.103-77 Обозначения программ и программных документов;
- 3) ГОСТ 19.104-78 Основные надписи;
- 4) ГОСТ 19.105-78 Общие требования к программным документам;
- 5) ГОСТ 19.106-78 Требования к программным документам, выполненным печатным способом
- 6) ГОСТ 19.505-79 ГОСТ 19.401-78 Текст программы. Требования к содержанию и оформлению. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.

Изменения к данному Руководству оператора оформляются согласно ГОСТ 19.603-78 [8], ГОСТ 19.604-78

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.09.11-01 12				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

**СОДЕРЖАНИЕ**

<b>1. ВВЕДЕНИЕ .....</b>	<b>4</b>
<b>2. ТЕКСТ ПРОГРАММЫ.....</b>	<b>5</b>
<b>2.1. Модели.....</b>	<b>6</b>
2.1.1. content/models.py .....	6
2.1.2. corpus/models.py .....	6
<b>2.2. Many-to-many .....</b>	<b>6</b>
2.2.1. corpus/document/annotators .....	6
<b>2.3. Пользовательское руководство .....</b>	<b>6</b>
<b>2.4. Новостные статьи .....</b>	<b>7</b>
2.4.1. Фронтэнд составляющая описана в файле: create_article.html.....	7
2.4.2. content\views.py .....	7
<b>2.5. Фильтрация и контекстный поиск.....</b>	<b>7</b>
<b>3. СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ.....</b>	<b>8</b>
<b>ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ .....</b>	<b>9</b>

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.09.11-01 12				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

## 1. ВВЕДЕНИЕ

### 1.1. Наименование программы

Наименование программы – «Русский Учебный Корпус» («Russian Learner Corpus»).

Краткое наименование программы – «RLC».

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.09.11-01 12				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

## 2. ТЕКСТ ПРОГРАММЫ

Репозиторий программы «Русский учебный корпус» находится по ссылке [https://github.com/shar3nda/rlc\\_new](https://github.com/shar3nda/rlc_new). В корневой директории находятся файлы .editorconfig (конфигурация форматирования в проекте), .gitignore (файл игнорируемых файлов в системе контроля версий Git), Dockerfile и nginx.Dockerfile (файлы, описывающие контейнеры Docker), README.md (файл с информацией о проекте), docker-compose.yml (файл, описывающий связь и взаимодействие различных контейнеров Docker), manage.py (сгенерированный Django файл для взаимодействия с проектом), nginx.conf (файл конфигурации веб-сервера nginx), requirements.txt (файл со списком python-зависимостей проекта) и seed.py (скрипт для генерации тестовых данных). Также, в корневой директории находятся следующие поддиректории:

1. api ([https://github.com/shar3nda/rlc\\_new/tree/master/api](https://github.com/shar3nda/rlc_new/tree/master/api)) – Django-приложение, содержащее API сайта для связи бекэнда и фронтенда
2. auto\_annotator ([https://github.com/shar3nda/rlc\\_new/tree/master/auto\\_annotator](https://github.com/shar3nda/rlc_new/tree/master/auto_annotator)) – Python-библиотека для авторазметки текстов
3. content ([https://github.com/shar3nda/rlc\\_new/tree/master/content](https://github.com/shar3nda/rlc_new/tree/master/content)) – Django-приложение, содержащее отображения и модели для новостных статей и главной страницы
4. corpus ([https://github.com/shar3nda/rlc\\_new/tree/master/corpus](https://github.com/shar3nda/rlc_new/tree/master/corpus)) – Django-приложение, содержащее основные модели и логику приложения
5. data\_import ([https://github.com/shar3nda/rlc\\_new/tree/master/data\\_import](https://github.com/shar3nda/rlc_new/tree/master/data_import)) – Python-утилита для импорта документов из json-файла
6. locale ([https://github.com/shar3nda/rlc\\_new/tree/master/locale](https://github.com/shar3nda/rlc_new/tree/master/locale)) – папка с информацией о локали и переводах
7. rlc\_new ([https://github.com/shar3nda/rlc\\_new/tree/master/rlc\\_new](https://github.com/shar3nda/rlc_new/tree/master/rlc_new)) – настройки Django-проекта
8. static ([https://github.com/shar3nda/rlc\\_new/tree/master/static](https://github.com/shar3nda/rlc_new/tree/master/static)) – статические файлы проекта (.js, .css и изображения)
9. templates ([https://github.com/shar3nda/rlc\\_new/tree/master/templates](https://github.com/shar3nda/rlc_new/tree/master/templates)) – html-шаблоны страниц сайта

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.09.11-01 12				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

## 2.1. Модели

### 2.1.1. content/models.py

[https://github.com/shar3nda/rlc\\_new/blob/master/content/models.py](https://github.com/shar3nda/rlc_new/blob/master/content/models.py)

2.1.1.1. class Article - класс для хранения новостных статей. Содержит следующие поля:

- 2.1.1.1.1. owner – владелец статьи
- 2.1.1.1.2. date – отображаемая дата
- 2.1.1.1.3. created – время создания
- 2.1.1.1.4. updated – время обновления
- 2.1.1.1.5. text\_rus – статья на русском
- 2.1.1.1.6. text\_eng – статья на английском

### 2.1.2. corpus/models.py

[https://github.com/shar3nda/rlc\\_new/blob/master/corpus/models.py](https://github.com/shar3nda/rlc_new/blob/master/corpus/models.py)

2.1.2.1. class Annotation – класс для хранения информации об аннотации. Содержит следующие поля:

- 2.1.2.1.1. document – документ к которому относится аннотация
- 2.1.2.1.2. sentence – предложение к которому относится аннотация
- 2.1.2.1.3. user – сотрудник, который добавил
- 2.1.2.1.4. guid – уникальный идентификатор аннотации в тексте
- 2.1.2.1.5. json – объект аннотации в формате json
- 2.1.2.1.6. alt – является ли исправление альтернативным
- 2.1.2.1.7. функция serialize для экспорта

2.1.2.2. class Token - класс для хранения информации о токене (слове). Содержит следующие поля:

- 2.1.2.2.1. token – сам токен (слово)
- 2.1.2.2.2. document – номер текста, к которому относится слово
- 2.1.2.2.3. sentence – номер предложения, к которому относится слово
- 2.1.2.2.4. start – начальная позиция слова в предложении
- 2.1.2.2.5. end – конечная позиция слова в предложении
- 2.1.2.2.6. number – номер слова в предложении
- 2.1.2.2.7. pos – часть речи
- 2.1.2.2.8. feats – грамматические характеристики
- 2.1.2.2.9. lemma – начальная форма слова

## 2.2. Many-to-many

### 2.2.1. corpus/document/annotators

2.2.1.1. document\_id – индикационный номер документа

2.2.1.2. user\_id - индикационный номер аннотатора

## 2.3. Пользовательское руководство

Содержится в файле: help.html

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.09.11-01 12				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

## 2.4. Новостные статьи

### 2.4.1. Фронтэнд составляющая описана в файле: create\_article.html

### 2.4.2. content\views.py

2.4.2.1. Бэкэнд составляющая добавления статьи происходит в функции:  
article\_create\_view

2.4.2.2. Бэкэнд составляющая удаления статьи происходит в функции: delete\_article

## 2.5. Фильтрация и контекстный поиск

Реализовано в corpus\views.py в функциях documents и search

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.09.11-01 12				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата



### 3. СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

- 1) ГОСТ 19.101–77 Виды программ и программных документов. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
- 2) ГОСТ 19.102–77 Стадии разработки. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
- 3) ГОСТ 19.103–77 Обозначения программ и программных документов. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
- 4) ГОСТ 19.104–78 Основные надписи. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
- 5) ГОСТ 19.105–78 Общие требования к программным документам. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
- 6) ГОСТ 19.106–78 Требования к программным документам, выполненным печатным способом. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
- 7) ГОСТ 19.603–78 Общие правила внесения изменений. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
- 8) ГОСТ 19.604–78 Правила внесения изменений в программные документы, выполненные печатным способом. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
- 9) ГОСТ 19.401–78 Текст программы. Требования к содержанию и оформлению. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.09.11-01 12				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

[illegible]