

**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

«13» 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**

Руководство оператора

ЛИСТ УТВЕРЖДЕНИЯ

RU.17701729.09.11-01 34 01-1-ЛУ

Исполнитель

студент группы БПИ 217



/ М.Д. Пересторонин/

«13» 05 2023 г.

Москва 2023

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

УТВЕРЖДЕН
RU.17701729.09.11-01 34 01-1-ЛУ

УЛУЧШЕНИЕ ПОЛЬЗОВАТЕЛЬСКОГО ИНТЕРФЕЙСА РУССКОГО УЧЕБНОГО
КОРПУСА
Руководство оператора

RU.17701729.09.11-01 34 01-1

Листов 17

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

Москва 2023

АННОТАЦИЯ

Руководство оператора – это вид программной документации, предназначенный для предоставления помощи пользователям некоего программного продукта. Данное Руководство оператора необходимо для организации корректной работы с ПО «Русский учебный корпус», и содержит следующие секции: «Назначение программы», «Условия выполнения программы», «Выполнение программы».

Секция «Назначение программы» нужна для введения пользователя в область применения данного ПО, а также для передачи основной информации о принципе работы программы.

Раздел «Условия выполнения программы» содержит основную информацию о необходимых условиях, для корректной работы данного ПО.

Раздел «Выполнение программы» содержит основные инструкции как для сотрудника корпуса, так и для обычного пользователя.

Настоящий документ разработан в соответствии с требованиями:

- 1) ГОСТ 19.101-77 Виды программ и программных документов [1];
- 2) ГОСТ 19.102-77 Стадии разработки [2];
- 3) ГОСТ 19.103-77 Обозначения программ и программных документов [3];
- 4) ГОСТ 19.104-78 Основные надписи [4];
- 5) ГОСТ 19.105-78 Общие требования к программным документам [5];
- 6) ГОСТ 19.106-78 Требования к программным документам, выполненным печатным способом [6];
- 7) ГОСТ 19.505-79 Руководство оператора. Требования к содержанию и оформлению [7].

Изменения к данному Руководству оператора оформляются согласно ГОСТ 19.603-78 [8], ГОСТ 19.604-78 [9].

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.09.11-01 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

СОДЕРЖАНИЕ

1. НАЗНАЧЕНИЕ ПРОГРАММЫ	4
1.1. Функциональное назначение	4
1.1.1 Функциональное назначение программы «RLC»	4
1.2. Эксплуатационное назначение	4
1.2.1 Эксплуатационное назначение программы «»	4
2. УСЛОВИЯ ВЫПОЛНЕНИЯ ПРОГРАММЫ	5
2.1. Минимальный состав аппаратных средств.....	5
2.2. Минимальный состав программных средств.....	5
2.3. Требования к пользователю	5
2.3.1 Посетитель корпуса	5
2.3.2 Сотрудник корпуса	5
3. ВЫПОЛНЕНИЕ ПРОГРАММЫ	6
3.1. Доступ к сайту русского учебного корпуса	6
3.2. Описание интерфейса и работы программы «RLC».....	6
3.2.1 Регистрация и вход в аккаунт	6
3.2.2 Страница добавления текста	7
3.2.3 Список всех текстов	8
3.2.4 Статистика	9
3.2.5 Аннотирование и редактирование текста	11
4. СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ	15

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.09.11-01 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

1. НАЗНАЧЕНИЕ ПРОГРАММЫ

1.1. Функциональное назначение

1.1.1 Функциональное назначение программы «RLC»

Функциональным назначением данного приложения является загрузка текста с соответствующими метаданными в систему, аннотация загруженных текстов путем исправления ошибок и присвоения им тэгов, сбор статистики по загруженным текстам, а также выгрузка данных из приложения в формате JSON. Таким образом приложение является своего рода коллекцией большого количества текстов, написанных иностранцами на русском языке.

1.2. Эксплуатационное назначение

1.2.1 Эксплуатационное назначение программы «»

Данная программа была разработана для помощи иностранцам в изучении русского языка, а также сбора статистических данных, которые в дальнейшем могут применяться в анализе данных и машинном обучении. Сотрудники корпуса вручную размечают текста, исправляя ошибки и присваивая им советующий тег.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.09.11-01 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

2. УСЛОВИЯ ВЫПОЛНЕНИЯ ПРОГРАММЫ

2.1. Минимальный состав аппаратных средств

Для корректной работы ПО требуется следующий набор технических средств:

- 1) Процессор: 1 ГГц или более быстрый процессор.
- 2) Оперативная память (RAM): не менее 2 ГБ.
- 3) Жесткий диск: свободное место на диске не менее 500 МБ.
- 4) Видеокарта: поддержка графики с разрешением не менее 1024x768 пикселей.

2.2. Минимальный состав программных средств

Для корректной работы ПО требуется следующий набор программных средств:

- 1) Операционная система: Windows 7 или более поздняя версия, MacOS, Linux или другие поддерживаемые операционные системы.
- 2) Браузер: последняя версия любого из популярных браузеров, таких как Google Chrome, Mozilla Firefox, Apple Safari или Microsoft Edge.

2.3. Требования к пользователю

2.3.1 Посетитель корпуса

Программа не требует специальных знаний, навыков и квалификации пользователя.

2.3.2 Сотрудник корпуса

Для использования программы необходимо, чтобы пользователь разбирался в устройстве работы русского учебного корпуса, обладал всеми необходимыми знаниями в профессиональной области (лингвистика, филология).

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.09.11-01 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

3. ВЫПОЛНЕНИЕ ПРОГРАММЫ

В данном разделе описан пример работы с программой и указано назначение элементов интерфейса.

3.1. Доступ к сайту русского учебного корпуса

Для доступа к сайту русского учебного корпуса необходимо ввести следующий URL:
http://134.209.87.41:5656/

В дальнейшем URL может измениться и стать более информативным.

3.2. Описание интерфейса и работы программы «RLC»

3.2.1 Регистрация и вход в аккаунт

Сайт автоматически сохраняет сессию пользователя, поэтому нет необходимости каждый раз входить в аккаунт. Если же пользователю необходимо войти в аккаунт, то он может нажать на кнопку войти на хедере сайта (в правом верхнем углу страницы) (см. рис. 1). Для входа необходимо ввести логин в поле для логина, пароль в поле для пароля и нажать на кнопку для авторизации. Если пользователь ввел некорректные данные, то появится надпись с сообщением об ошибке, где пользователя попросят ввести корректные данные.

Т. к. аккаунт в корпусе нужен исключительно для персонала, на сайте отсутствует стандартная кнопка регистрации. Вместо обычной само регистрации, сотрудник корпуса должен создать аккаунт для нового сотрудника через админ панель. После чего новый сотрудник сможет войти на сайт обычным способом.

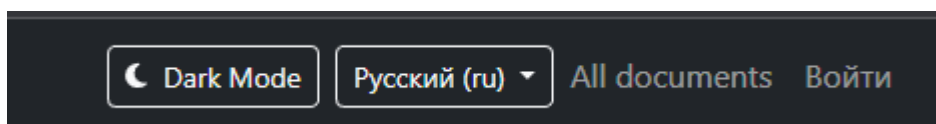


Рисунок 1. Кнопки на хедере сайта

После входа в аккаунт кнопка войти в хедере меняется на кнопку с именем пользователя. При нажатии на нее появляется выпадающее меню, где есть другие кнопки: профиль и выйти (рис 2).

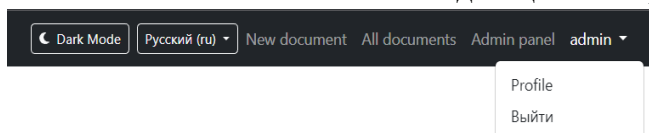


Рисунок 2. Выпадающее меню

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.09.11-01 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

При нажатии на кнопку выйти в выпадающем меню, пользователь выходит из аккаунта и теряет все привилегии зарегистрированного пользователя. На хедере вновь появляется кнопка войти (рис 1).

При нажатии на кнопку профиль, открывается страница с профилем пользователя, где указывается электронная почта, имя и фамилия пользователя.

3.2.2 Страница добавления текста

Сотрудник корпуса должен ввести непосредственно сам текст, а также данные о нем и его авторе (см. рис 3). Если сотрудник понимает, что у в корпусе еще планируется хранить текста этого же автора, то он может его сохранить. Если при добавлении текста сотрудник понимает, что автор этого текста уже был сохранен, то он может не вводить данные о нем вновь, а выбрать его имя из выпадающего списка (см. рис 4) таким образом, работнику корпуса не придется вводить один и те же данные несколько раз.

Add document

Title: Written in:

Genre: Subcorpus:

Текст:

Time limit: ☐ Oral: ☐

Author name: Gender:

Program: Language background:

Dominant language: Source:

Add to favorites: ☐

Language level:

Рисунок 3. Добавление текста в корпус

Обязательными для заполнения являются следующие поля:

Название текста, сам текст, а также полная информация об авторе (Имя, программа, пол, тип носителя, источник, доминантный язык, уровень знания языка автором).

Поля: дата написания и жанр текста могут не заполняться. Помимо вышеперечисленных полей, корпус также хранит информацию о том, какой из пользователей добавил текст.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.09.11-01 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

3.2.3 Список всех текстов

При выборе вкладки все документы (All Documents), перед пользователем открывается страница, где вверху находится фильтр и поиск, а внизу все текста, представленные в корпусе (рис 4).

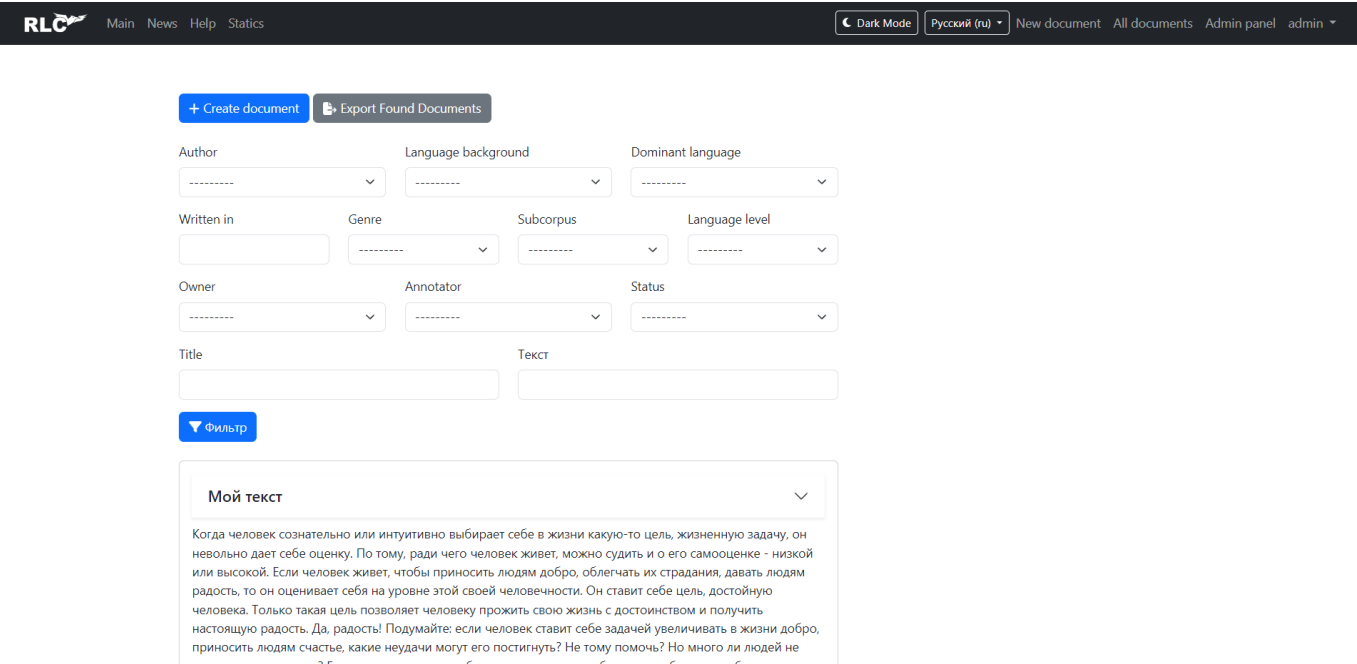


Рисунок 4. Все документы

При выборе тех или иных фильтров и нажатии на кнопку “Фильтр”, внизу будут отображаться не все документы корпуса, а только те, которые подходят под выбранный фильтр. Среди фильтров также присутствует поиск по формам (контекстный поиск). При его использовании, в списке текстов будут представлены документы, содержащие соответствующие формы. В добавок к этому эти формы будут выделены отдельным цветом, чтобы их легко можно было увидеть и различить среди прочего текста (рис 5).

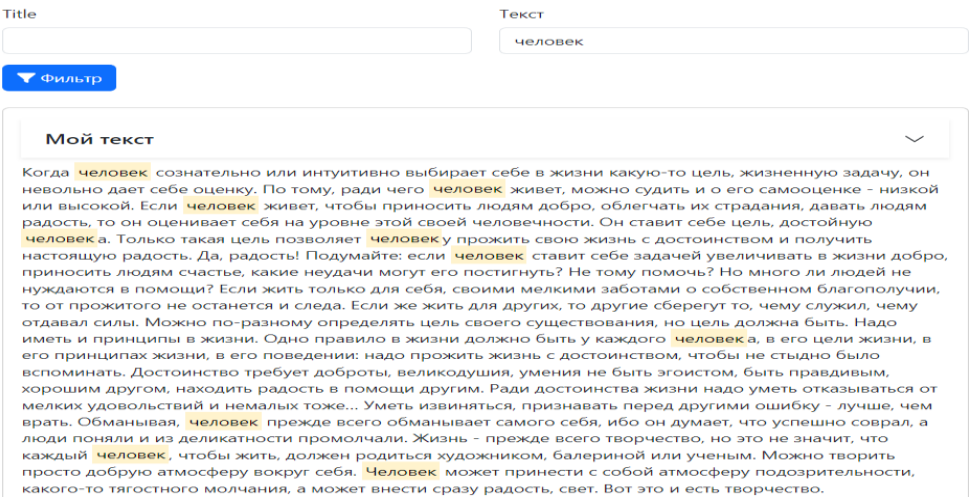
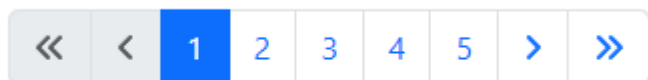


Рисунок 5. Поиск

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.09.11-01 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

Программа также поддерживает пагинацию. На странице отображаются не сразу все тексты, а только десять за раз. Внизу веб-страницы есть переключатель страниц (рис 6).



3.2.4 Статистика

При нажатии на кнопку “Статистика” в хэдере сайта, открывается страница со статистикой. Вверху страницы отображается таблица с основной информацией о корпусе (рис 6). Внизу таблицы расположены аккордеоны, каждый из которых содержит статистические графики по соответствующей тематике (рис 7).

Статистика по родным языкам

Доминантный язык	Количество текстов	Количество предложений
Английский	6	6
Бенгальский	1	1
Дари	2	8
Испанский	4	13
Курдский	4	10
Монгольский	4	12
Непальский	9	47
Норвежский	5	17
Хорватский	7	21
Шведский	10	32
Шона	6	17

Статистика по документам



Статистика по предложениям



Рисунок 6. Статистика


Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.09.11-01 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата




Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.09.11-01 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

3.2.5 Аннотирование и редактирование текста

При нажатии на кнопку “Аннотировать” на странице “Все документы” открывается страница с аннотациями. Эта страница доступна только для сотрудников корпуса. На ней весь текст разбит на предложения. Каждое предложение располагается в своем собственном блоке. Каждый такой блок содержит непосредственно само предложение, а также два аккордеона: Альтернативное исправление, морфологический разбор. Рядом с самим предложением также располагается иконка с волшебной палочкой (рис 8).


В одну плоскую тарелку высыпает сухари, в другую муку 5) разливаем подсолнечное масло в сковороду и греем на максимальной мощности, до тех пор пока при подношении руки сверху не буди горячо. 

В одну плоскую тарелку высыпает сухари, в другую муку 5) разливаем подсолнечное масло в сковороду и греем на максимальной мощности, до тех пор пока при подношении руки сверху не буди горячо.

Альтернативное исправление 

В одну плоскую тарелку высыпает сухари, в другую муку 5) разливаем подсолнечное масло в сковороду и греем на максимальной мощности, до тех пор пока при подношении руки сверху не буди горячо.

В одну плоскую тарелку высыпает сухари, в другую муку 5) разливаем подсолнечное масло в сковороду и греем на максимальной мощности, до тех пор пока при подношении руки сверху не буди горячо.

Морфологический разбор 

Lemma: сухарь POS: NOUN
Morph: {'Animacy': 'Inan',
'Case': 'Acc', 'Gender': 'Masc',
'Number': 'Sing'}

В одну плоскую тарелку высыпает сухари , в другую муку 5) разливаем подсолнечное масло в сковороду и греем на максимальной мощности , до тех пор пока при подношении руки сверху не буди горячо .

Рисунок 8. Блок с предложением

При наведении курсора на любое слово в аккордеоне “морфологический разбор”, над ним открывается черная секция, содержащая морфологический разбор данного слова. Если начать выделять курсором текст в самой верхней части блока, то появится меню добавления аннотации (рис 9)

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.09.11-01 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

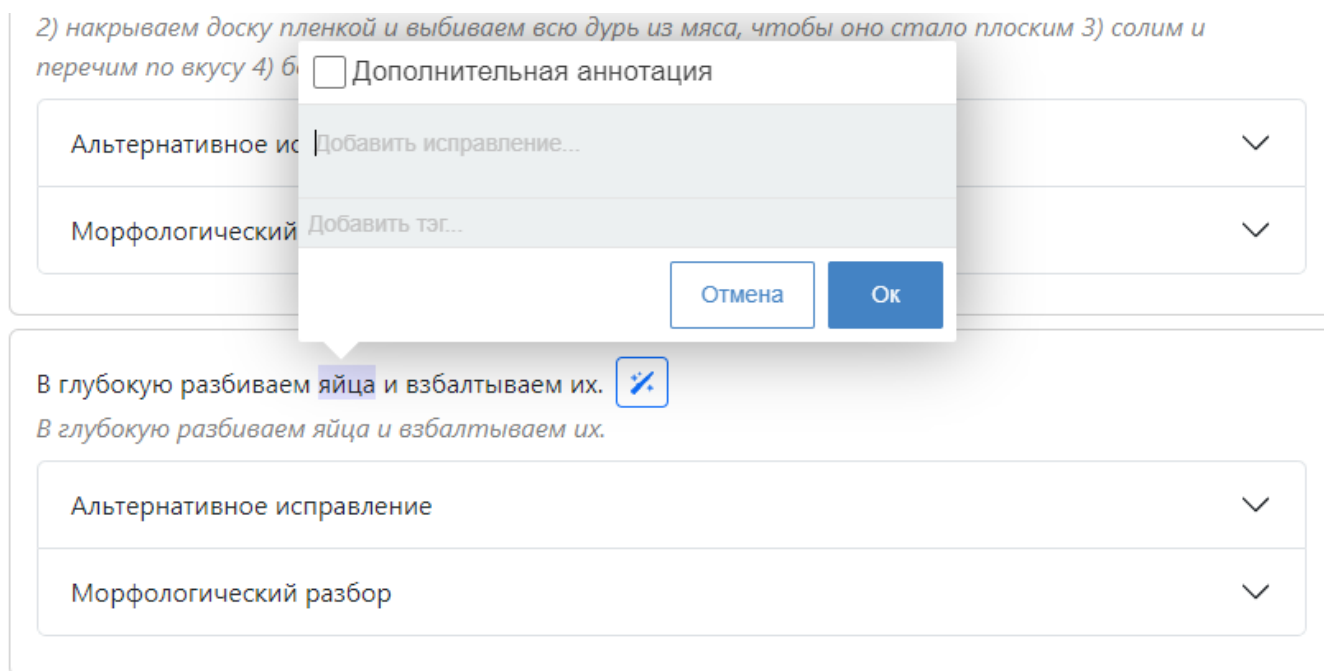


Рисунок 9. Меню добавления аннотации

В этом меню можно добавить исправление выделенного фрагмента текста, а также добавить тэг выделенной ошибки. После чего при нажатии на клавишу “Ок”, к тексту применяется аннотация. Аннотированный фрагмент предложения начинает светиться желтым и под самим предложением (на второй строке, где дублируется предложение, но только меньшим шрифтом) отображается версия предложения со всеми исправлениями (рис 10).

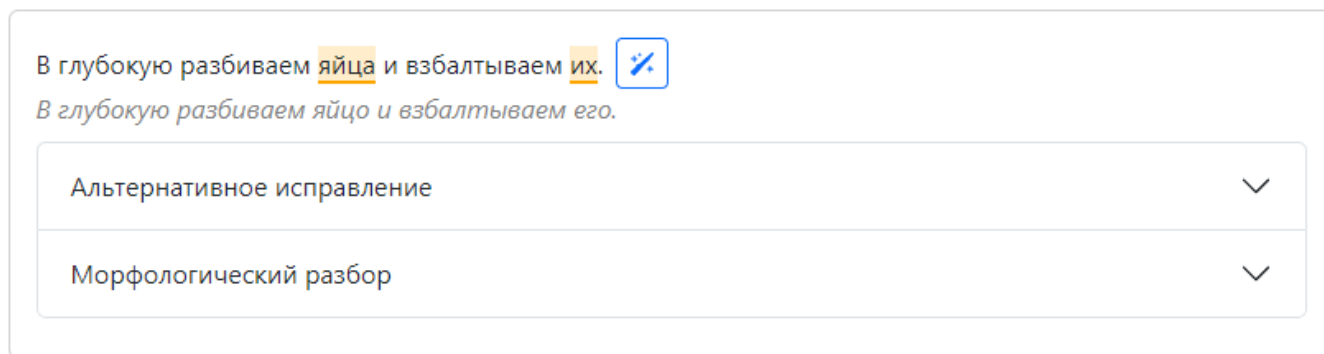


Рисунок 10. Аннотированное предложение

При добавлении аннотации также есть чекбокс “Дополнительная аннотация” (рис 9). При отметке которого добавленная аннотация не будет отображаться в исправленном тексте, находящемся под основным предложением, а в отдельной секции “Альтернативное исправление”. При нажатии на иконку волшебной палочки открывается режим автоаннотирования. Сотрудник корпуса может просто ввести исправленную версию текста, а программа сама постарается расставить аннотации. Надо учитывать, что данный инструмент не обладает идеальной точностью. При нажатии на

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.09.11-01 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

существующую аннотацию (фрагмент текста, выделенный желтым) появляется меню этой аннотации, где ее можно отредактировать, добавить тег, написать комментарий и удалить. Там также отображается, кто и когда поставил эту аннотацию (рис 11).

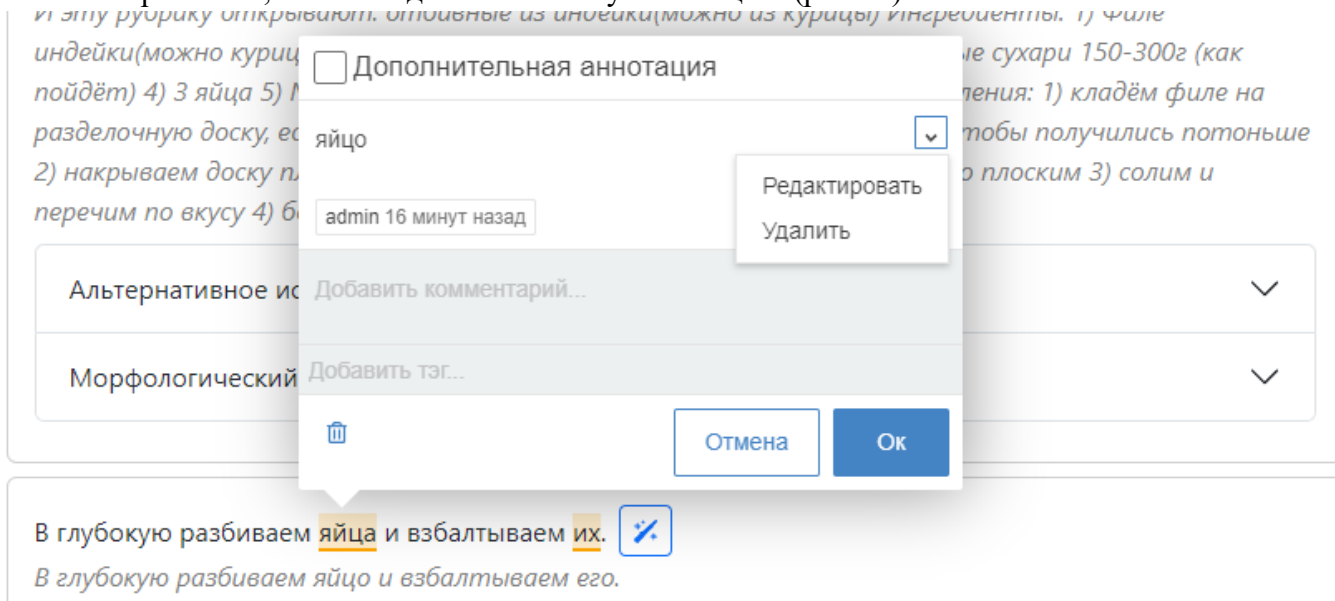


Рисунок 11. Меню добавленной аннотации

В самом верху страницы отображается название текста рядом с двумя кнопками: изменить и удалить. Само название находится на аккордеоне, при нажатии на который раскрывается общая информация о тексте. Выше названия текста есть поле с его категорией (рис 12). В корпусе тексты делятся на три основные категории: не аннотирован, аннотирован, проверен. Когда текст только добавляется в корпус, ему автоматически присваивается статус “не аннотирован”. Во время аннотирования, когда сотрудник закончил исправлять все ошибки в тексте, он меняет статус этого поля на “аннотирован”. После чего другой сотрудник просматривает аннотированные тексты и смотрит на качество работы, если его все устраивает, то он меняет статус текста на проверен. Посмотреть только аннотированные, не аннотированные и проверенные можно через фильтр на странице “Все документы”, описанной ранее.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.09.11-01 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

Статус документа
Не аннотирован

Отбивные из индейки

Добавлено: 13 мая 2023 г. 12:49
Автор: Андрей

Жанр: Блог
Пол: Мужской

Подкорпус: HSE
Родной язык: Болгарский

Уровень языка: C2
Источник: Телеграмм канал

Статус: Новый
Тип носителя: Эритажный

Создатель: admin
Программа: C++

Изменить
Удалить

И эту рубрику открывают: отбивные из индейки(можно из курицы) Ингредиенты: 1) Филе

Рисунок 12. Страница аннотирования, общая информация о тексте

При нажатии на кнопку “Изменить”, открывается меню редактирования текста, где можно изменить непосредственно сам текст, а также всю информацию о нем. Если сотрудник корпуса решил изменить информацию об авторе, то ему нужно нажать на кнопку “Редактировать автора”. При нажатии на которую открывается отдельная страница со всеми полями автора (рис 13)

Редактировать автора

Автор: Андрей
Пол: Мужской

Программа: C++
Тип носителя: Эритажный

Родной язык: Болгарский
Источник: Телеграмм канал

Добавить в сохраненные:

Сохранить

Рисунок 13. Редактирование автора

Тут можно изменить всю информацию об авторе, а также добавить его в список сохраненных авторов. После завершения редактирования информации об авторе необходимо нажать кнопку “Сохранить”.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.09.11-01 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

4. 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 19.505-79 Руководство оператора. Требования к содержанию и оформлению. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
- 8) ГОСТ 19.603–78 Общие правила внесения изменений. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
- 9) ГОСТ 19.604–78 Правила внесения изменений в программные документы, выполненные печатным способом. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
- 10) Django documentation [Электронный ресурс]. Режим доступа: <https://docs.djangoproject.com/en/4.2/>, свободный (Дата обращения 10.05.2023)
- 11) Bootstrap documentation [Электронный ресурс]. Режим доступа: <https://getbootstrap.com/docs/5.3/getting-started/introduction/>, свободный (Дата обращения 10.05.2023)
- 12) jQuery API Reference [Электронный ресурс]. Режим доступа: <https://api.jquery.com/>, свободный (Дата обращения 10.05.2023)
- 13) RecogitoJS API Reference [Электронный ресурс]. Режим доступа: <https://github.com/recogito/recogito-js/wiki/API-Reference>, свободный (Дата обращения 10.05.2023)
- 14) Docker documentation [Электронный ресурс]. Режим доступа: <https://docs.docker.com/get-started/>, свободный (Дата обращения 10.05.2023)
- 15) Web annotation data model [Электронный ресурс]. Режим доступа: <https://www.w3.org/TR/annotation-model/>, свободный (Дата обращения 10.05.2023)
- 16) RLC [Электронный ресурс]. Режим доступа: <http://web-corpora.net/RLC>, свободный (Дата обращения 10.05.2023)

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.09.11-01 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

- 17) Natasha — качественное компактное решение для извлечения именованных сущностей из новостных статей на русском языке [Электронный ресурс]. Режим доступа: <https://natasha.github.io/ner/>, свободный (Дата обращения 10.05.2023)

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.09.11-01 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ

[illegible]