

БИБЛИОТЕКИ ПОДПРОГРАММ (SDK)  
ДЛЯ ТЕКСТОПОНИМАНИЯ И ТЕКСТОГЕНЕРАЦИИ  
НА ОСНОВЕ ТЕХНОЛОГИЙ МАШИННОГО ОБУЧЕНИЯ

**РУКОВОДСТВО ПОЛЬЗОВАТЕЛЯ**

Самара, 2022

## **ВВЕДЕНИЕ**

Настоящий документ ориентирован, в первую очередь, на самостоятельное изучение материала и приобретение практических навыков, требуемых для эффективной работы в Системе с применением библиотек подпрограмм (SDK) для текстопонимания и текстогенерации на основе технологий машинного обучения, далее «Подсистемы».

Руководство состоит из трех основных разделов.

Первый раздел содержит сведения о работе с Подсистемой «Модуль обработки корпоративной документации».

Во втором разделе описана работа с Подсистемой «Конфигуратор извлекаемых данных».

В третьем разделе описана работа с Подсистемой «Классификатор корпоративной документации».

# Оглавление

<b>Перечень принятых сокращений.....</b>	<b>5</b>
<b>1. Модуль обработки корпоративной документации.....</b>	<b>6</b>
1.1. Краткое описание Системы .....	6
1.2. Запуск Системы.....	6
1.3. Интерфейс главной страницы Системы.....	7
1.4. Работа с документами .....	8
1.5. Интерфейс раздела «Документы» .....	8
1.6. Просмотр документов .....	9
1.7. Загрузка документов в Систему .....	11
1.8. Распознавание документов и извлечение единиц информации .....	13
1.8.1. Редактирование извлеченных единиц информации .....	13
1.9. Создание проектов документов .....	14
1.9.1. Создание проектов документов по шаблону.....	14
1.9.2. Редактирование проектов документов .....	15
1.9.3. Сравнение документов.....	17
1.10. Удаление документов .....	20
1.11. Поиск документов .....	21
1.12. Работа с шаблонами документов.....	21
1.13. Создание шаблонов.....	21
1.14. Редактирование шаблонов .....	25
1.15. Удаление шаблонов.....	28
1.16. Поиск шаблона .....	28
<b>2. Конфигуратор извлекаемых данных.....</b>	<b>30</b>
2.1. Краткое описание Подсистемы .....	30
2.2. Запуск Подсистемы.....	31
2.3. Интерфейс Подсистемы.....	31
2.4. Просмотр RDF-графов .....	32
2.5. Создание RDF-графов.....	35
2.6. Редактирование RDF-графов .....	35
2.6.1. Добавление узла .....	35
2.6.2. Добавление литерала.....	36
2.6.3. Удаление элементов RDF-графа.....	37
2.7. Удаление RDF-графа.....	37

2.8. Импорт RDF-графа.....	37
2.9. Экспорт RDF-графа.....	38
2.10. Создание RDF-графов на основе аналогов.....	38
<b>3. Классификатор корпоративной документации .....</b>	<b>39</b>
3.1. Краткое описание Подсистемы .....	39
3.2. Запуск Подсистемы.....	39
3.3. Интерфейс Подсистемы.....	39
3.4. Работа с подсистемой.....	40

## Перечень принятых сокращений

- JSON – JavaScript Object Notation, формат для хранения и обмена информацией, доступной для чтения человеком
- HDD – накопитель на жёстких магнитных дисках
- HTML – HyperText Markup Language, язык гипертекстовой разметки
- HTTP – HyperText Transfer Protocol, протокол прикладного уровня передачи данных
- RDF – Resource Description Framework, модель для представления данных
- БД – база данных
- ЛКМ – левая кнопка мыши
- ОС – операционная система
- ПО – программное обеспечение
- ПК – персональный компьютер
- ПМИ – программа и методика испытаний
- СУБД – система управления базами данных
- ФИО – фамилия, имя, отчество

# 1. Модуль обработки корпоративной документации

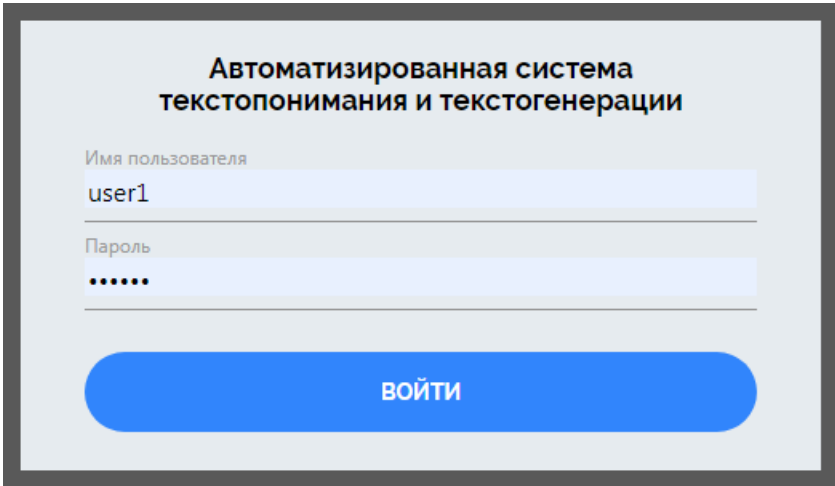
## 1.1. Краткое описание Системы

Системы Библиотеки подпрограмм (SDK) для текстопонимания и текстогенерации на основе технологий машинного обучения (далее – библиотеки подпрограмм (SDK) для текстопонимания и текстогенерации) предназначены для автоматизированного анализа, экспертизы, генерации корпоративной документации и могут применяться в государственном управлении и при решении отраслевых задач: статистика, финансовый мониторинг, медицина, безопасность человека и окружающей среды, образование, разработка технологических процессов и др.

Библиотеки подпрограмм (SDK) для текстопонимания и текстогенерации подключаются в качестве сервиса к системам и обеспечивают распознавание загружаемых документов, классификацию, хранение, а также «интеллектуальную» генерацию документов.

## 1.2. Запуск Системы

Работа с Системой осуществляется в веб-интерфейсе по протоколу HTTP. Введите в адресной строке браузера соответствующий адрес расположения сервера. После этого будет открыта форма для входа в Подсистему, где пользователь должен ввести логин в поле «Имя пользователя» - user1 и пароль в поле «Пароль»- qwe123. Далее нажмите кнопку «Войти» (Рисунок 1).



The image shows a login form with the following elements:

- Title: **Автоматизированная система текстопонимания и текстогенерации**
- Input field for 'Имя пользователя' (Username) containing the text 'user1'.
- Input field for 'Пароль' (Password) containing six dots '.....'.
- A blue button labeled 'ВОЙТИ' (Log In).

Рисунок 1

При корректно введенных значениях логина и пароля откроется главная страница Системы (Рисунок 2).

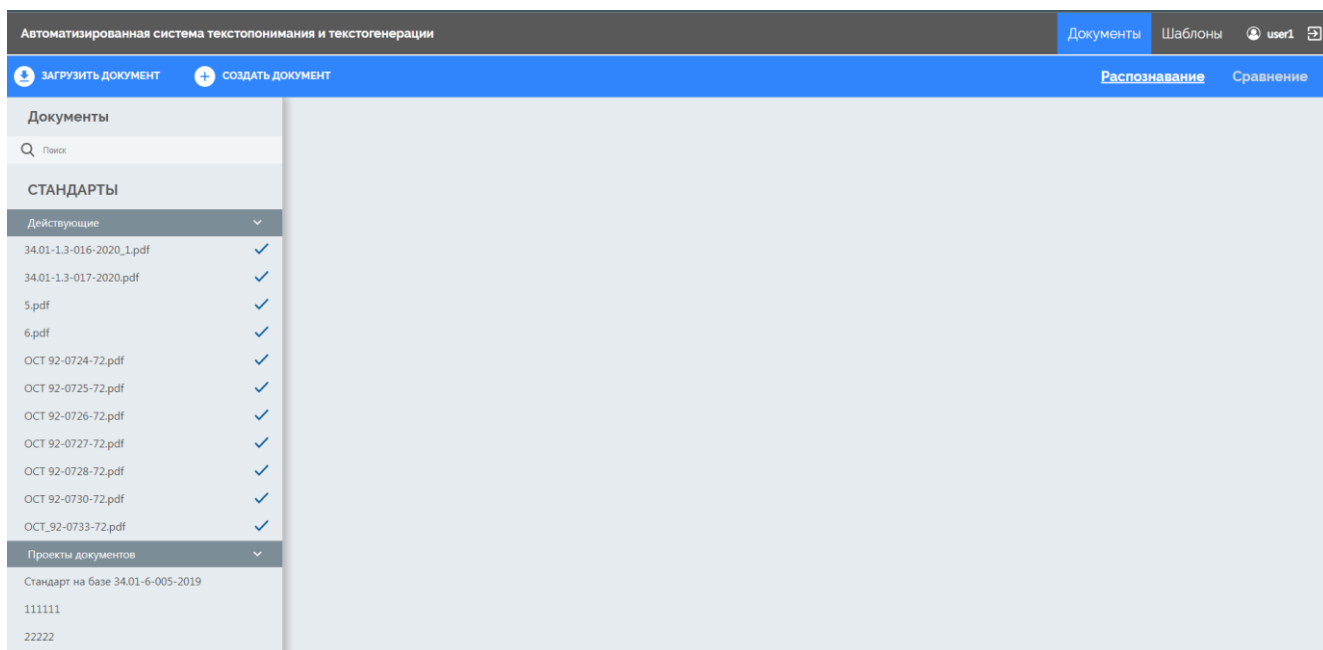


Рисунок 2

Если логин или пароль введены неверно, то Система выведет соответствующее сообщение (Рисунок 3).

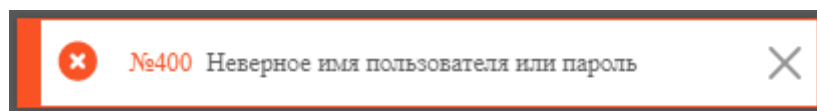



Рисунок 3

При некорректном вводе логина/пароля проверьте правильность языка ввода, а также нажатие клавиши «Caps Lock» на клавиатуре, и затем повторно введите логин/пароль.

### 1.3. Интерфейс главной страницы Системы

В правой верхней части главной страницы (Рисунок 2) отображается ФИО пользователя, а также кнопка «» (Выйти), при нажатии на которую выполняется завершение работы под учетной записью пользователя и возвращение к форме входа в Систему.

Слева от ФИО пользователя располагается меню с основными разделами Системы: работа с Документами, работа с Шаблонами (Рисунок 4).



Рисунок 4

По умолчанию, при входе в Систему автоматически открывается раздел «Документы».

## 1.4. Работа с документами

Система работает с документами типа «Стандарт организации». Стандарты организации могут быть введены в действие, а могут создаваться пользователями Системы (проекты документов).

Раздел «Документы» предназначен для:

- загрузки документов в Систему;
- распознавания документов;
- извлечения информации из документов;
- создания и редактирования проектов документов.

## 1.5. Интерфейс раздела «Документы»

Страница раздела «Документы» разделена на две части (Рисунок 5).

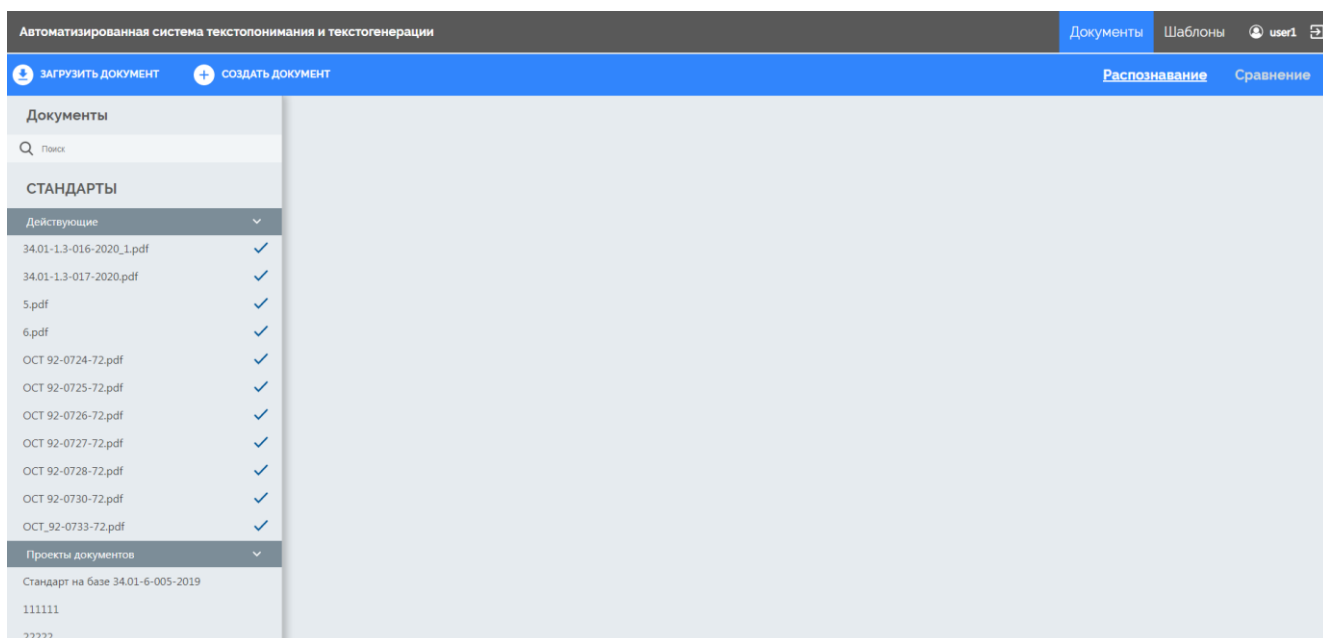


Рисунок 5

В левой части страницы располагаются вкладки со списками действующих стандартов и с проектами документов (Рисунок).



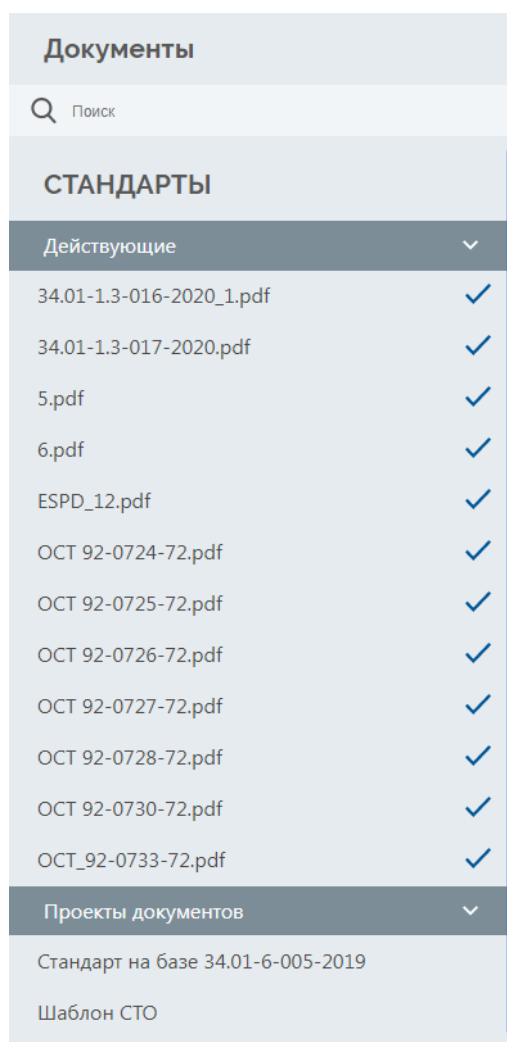


Рисунок 6

В правой части страницы располагается область для просмотра и редактирования документов.

## 1.6. Просмотр документов

Для просмотра документа выберите его из списка, нажав один раз ЛКМ на его названии. Название выбранного документа будет подсвечиваться синим цветом в списке документов (Рисунок 6, поз.1).

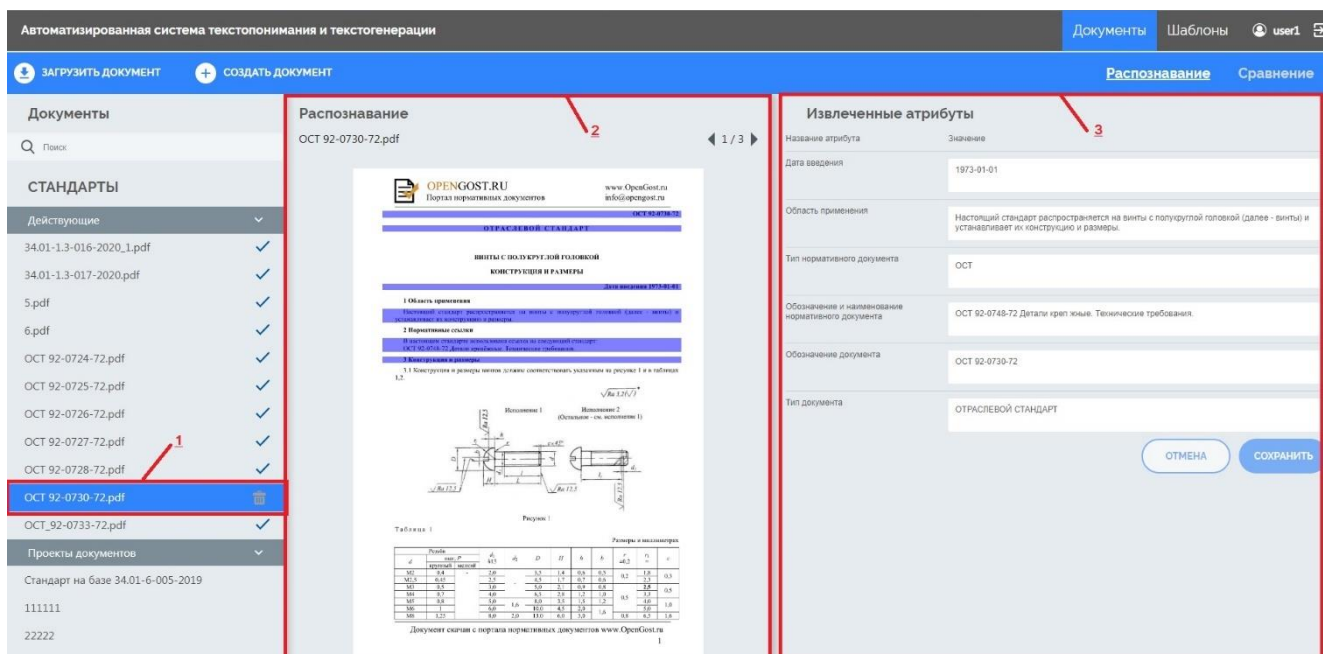




Рисунок 6

После сделанного выбора, справа от списка будет открыта страница документа (Рисунок 6, поз.2, поз.3).

Страница документа содержит два окна: «Распознавание» и «Извлеченные атрибуты».

В области «Распознавание» отображается наименование и содержимое документа (Рисунок 6, поз.2). Справа от названия документа находится служебная информация (текущая страница, общее количество страниц в документе) а также элементы управления отображением:

-  – позволяет отобразить предыдущую страницу документа;
-  – позволяет отобразить следующую страницу документа.

В окне «Извлеченные атрибуты» отображается список наименований извлеченных из документа атрибутов, в порядке их появления в тексте, и их значения (Рисунок 7).

## Извлеченные атрибуты


Название атрибута	Значение
Дата введения	1973-01-01
Область применения	Настоящий стандарт распространяется на винты с полукруглой головкой (далее - винты) и устанавливает их конструкцию и размеры.
Тип нормативного документа	ОСТ
Обозначение и наименование нормативного документа	ОСТ 92-0748-72 Детали креп жные. Технические требования.
Обозначение документа	ОСТ 92-0730-72
Тип документа	ОТРАСЛЕВОЙ СТАНДАРТ

ОТМЕНА
СОХРАНИТЬ

Рисунок 7

### 1.7. Загрузка документов в Систему

Для загрузки документа в Систему необходимо выполнить следующий порядок действий:

1. Нажать на кнопку « ЗАГРУЗИТЬ ДОКУМЕНТ» (Загрузить документ).
2. В стандартном окне открытия документа выбрать загружаемый документ и нажать на кнопку «Открыть» (Рисунок 9).

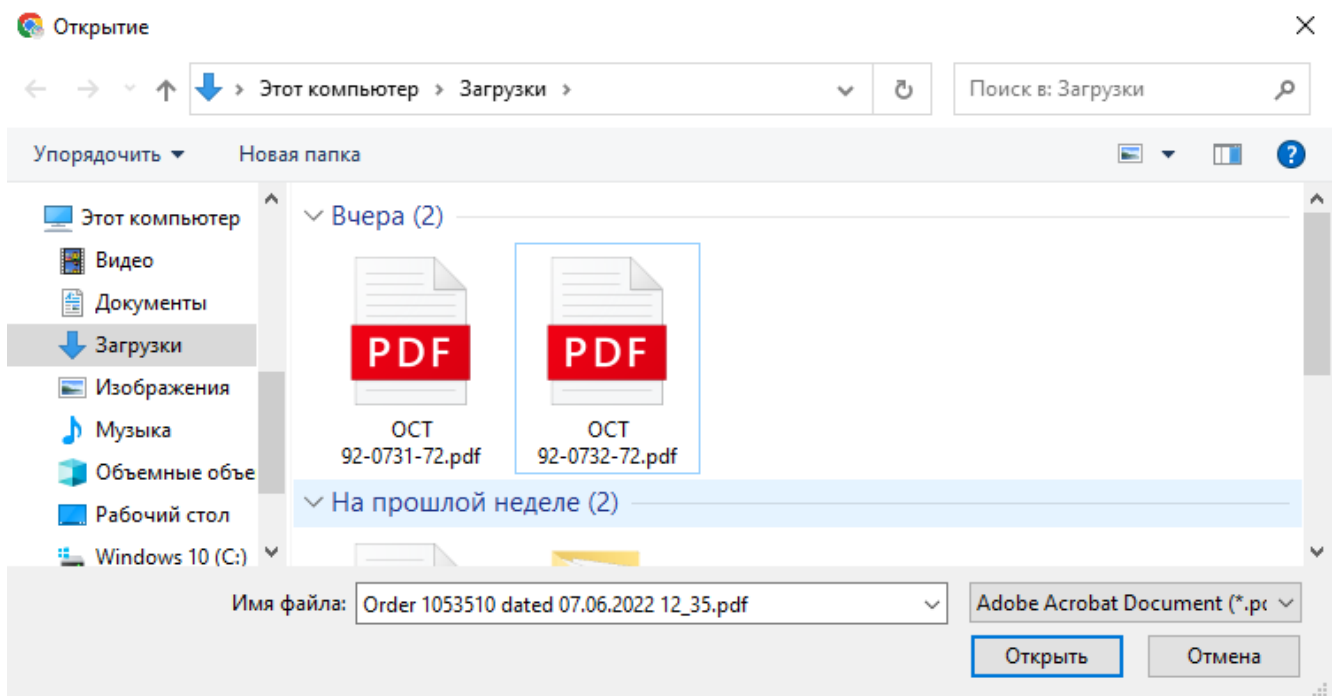
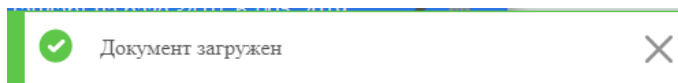


Рисунок 9

После успешной загрузки документа будет отображаться информирующее окно




, а сам документ отобразится в списке действующих документов (Рисунок 10).

Документы	
🔍 Поиск	
СТАНДАРТЫ	
Действующие ▾	
34.01-1.3-016-2020_1.pdf	✓
34.01-1.3-017-2020.pdf	✓
5.pdf	✓
6.pdf	✓
ОСТ 92-0724-72.pdf	✓
ОСТ 92-0725-72.pdf	✓
ОСТ 92-0726-72.pdf	✓
ОСТ 92-0727-72.pdf	✓
ОСТ 92-0728-72.pdf	✓
ОСТ 92-0730-72.pdf	✓
ОСТ 92-0731-72.pdf	✓
ОСТ_92-0733-72.pdf	✓

Рисунок 10

## 1.8. Распознавание документов и извлечение единиц информации

Распознавание документа и извлечение единиц информации из документа начинается автоматически сразу после загрузки документа в Систему. Данный процесс визуально показан прогресс-индикатором «», находящимся справа от названия файла загруженного документа (Рисунок 11).

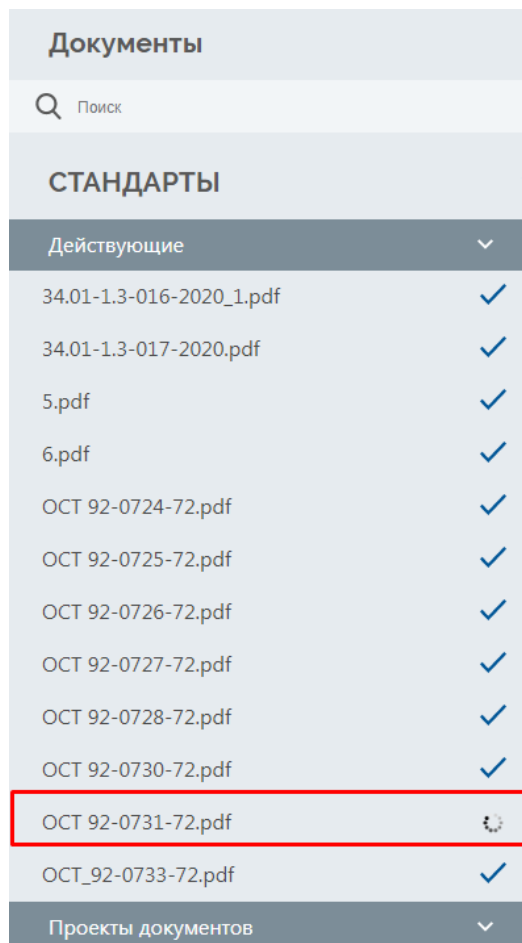



Рисунок 11

После завершения процесса распознавания и извлечения единиц информации в списке документов будет отображаться наименование файла документа.

Распознанный документ в списке отмечается знаком «» справа от его наименования.

Примечание: качество распознавания текста напрямую зависит от качества исходного изображения документа. Рекомендуется загружать документы с разрешением не менее 300dpi, высотой символов – 20 пикселей, отсутствием шумов, искажений и артефактов печати.

### 1.8.1. Редактирование извлеченных единиц информации

Для редактирования извлеченных единиц информации необходимо:

1. Выбрать необходимый документ в списке действующих стандартов, щелкнув

один раз ЛКМ на его названии (Рисунок 6, поз.1).

2. В окне «Извлеченные атрибуты» выбрать атрибут для редактирования, щелкнув один раз ЛКМ в его поле (Рисунок 12) внести новые данные.

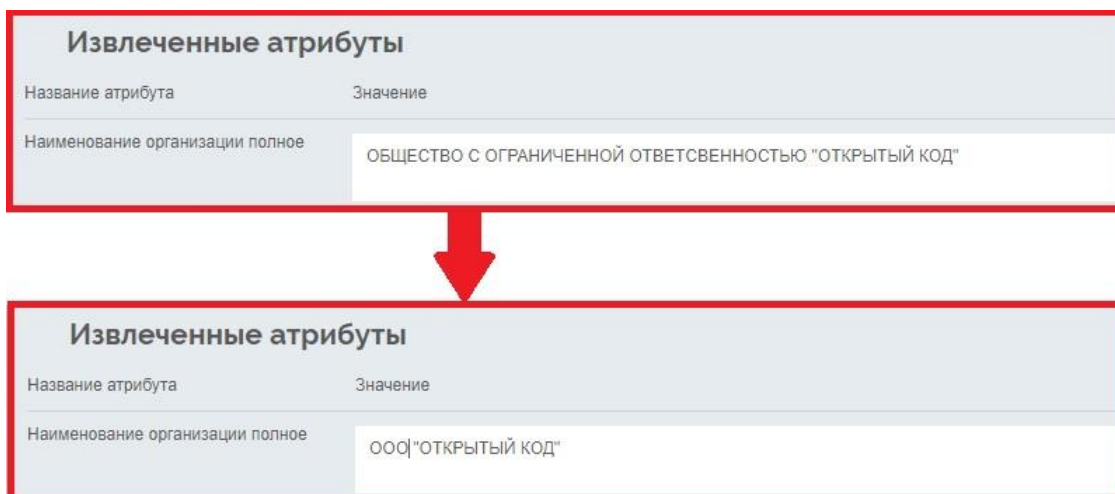
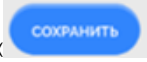



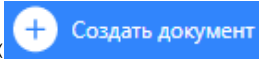
Рисунок 12

3. Нажать на кнопку «  » (Сохранить) для применения изменений, или кнопку «  » (Отмена) для отмены редактирования и возврата к предыдущему сохраненному результату.

## 1.9. Создание проектов документов

### 1.9.1. Создание проектов документов по шаблону

Для создания проекта документа по шаблону необходимо выполнить следующий порядок действий:

1. Нажать на кнопку «  » (Создать документ).
2. В окне создания документа выбрать тип документа, шаблон, указать наименование документа (Рисунок 13).

✕

Создание нового документа

Тип документа  
Отраслевой стандарт ✕

Шаблон  
Шаблон ОСТ ✕

Название документа  
ОСТ|92

---

ОТМЕНА
СОЗДАТЬ

Рисунок 13

3. Нажать на кнопку «Создать» (Рисунок 14).



Рисунок 14

Созданный проект документа будет отображаться в списке Документов.

### 1.9.2. Редактирование проектов документов

Для редактирования проекта документа необходимо:

1. Выбрать проект документа в списке проектов документов, щелкнув один раз на его названии ЛКМ. Выбранный проект документа будет подсвечиваться синим цветом (**Error! Reference source not found.**).

2. Нажать на знак «✎» справа от названия проекта документа в списке. Откроется страница для редактирования (Рисунок 15).

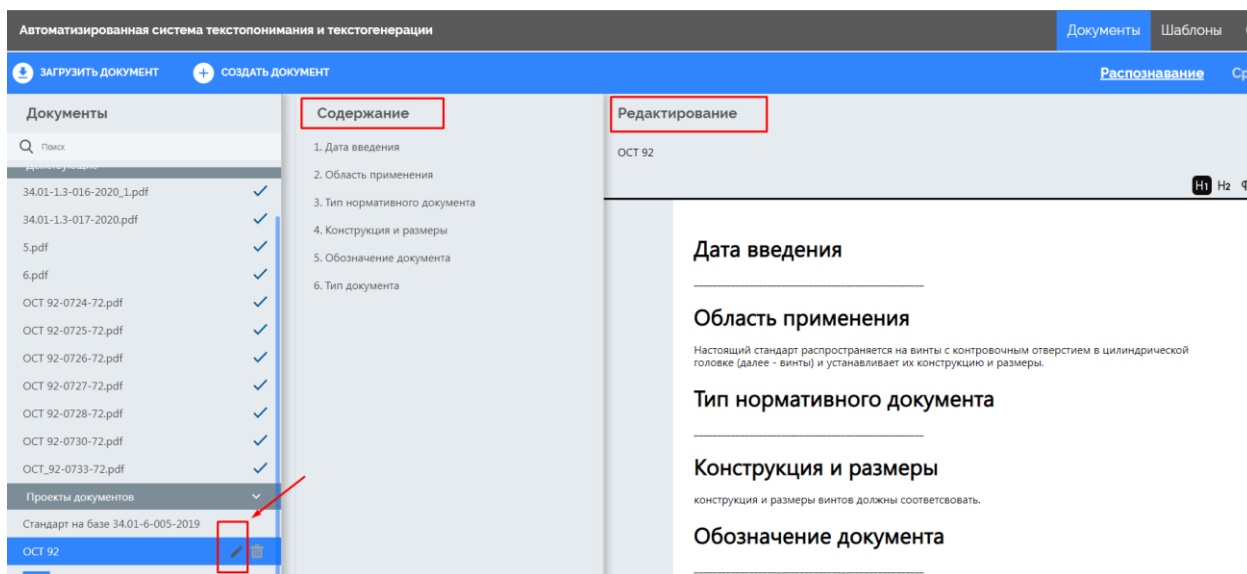


Рисунок 15

Страница для редактирования разделена на две части (Рисунок 15) «Содержание» и «Редактирование».

В области «Содержание» отображается оглавление проекта документа (Рисунок 16).

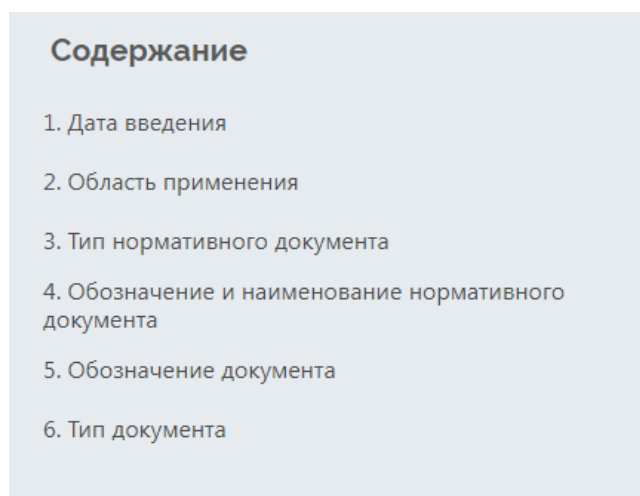


Рисунок 16

3. В области «Содержание» выбрать раздел для редактирования, щелкнув на него один раз ЛКМ. Выбранный раздел в области «Редактирование» будет находиться в верхней части (Рисунок 17).

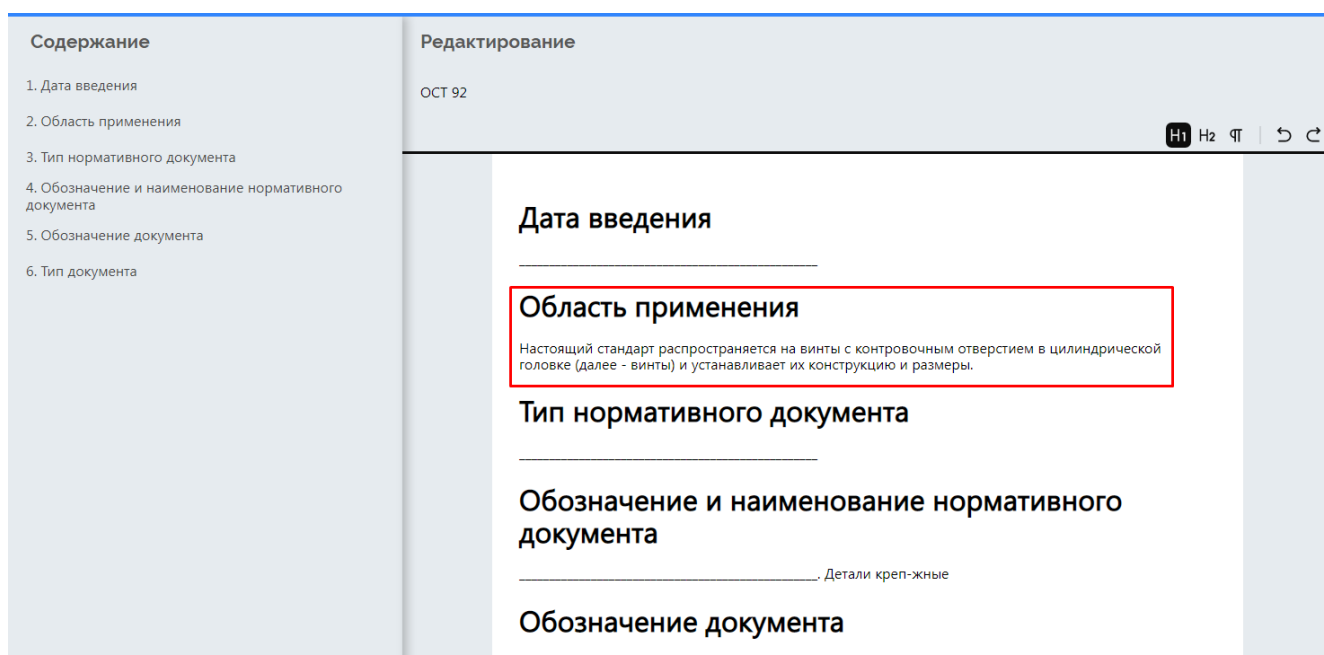
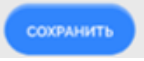



Рисунок 17

4. В области «Редактирование» отредактировать необходимую информацию (Рисунок 18) и нажать на кнопку «» (Сохранить) для применения изменений, или кнопку «» (Отмена) для отмены внесенных изменений и возврата к предыдущему сохраненному результату.



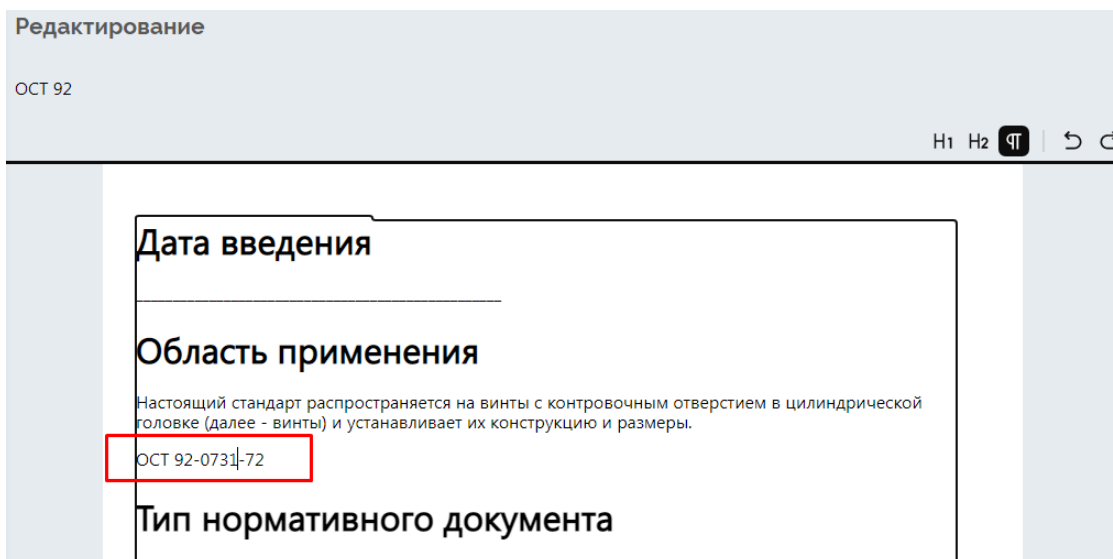


Рисунок 18

### 1.9.3. Сравнение документов

В Системе существует возможность редактирования проектов документов с помощью сравнения с действующими документами.

Для проведения процедуры сравнения необходимо открыть подраздел «Сравнение» в меню раздела «Документы» (Рисунок 19).

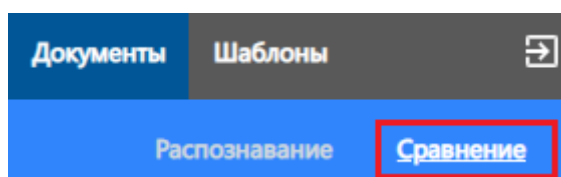


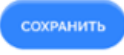

Рисунок 19

Страница для сравнения документов (Рисунок 20) состоит из следующих частей:

- 1) Список проектов документов подлежащих сравнению (Рисунок 20, поз. 1);
- 2) Область отображения и редактирования документа, который подлежит сравнению (Рисунок 20, поз. 2). В данной области отображается проект документа в редактируемом виде.

Текст можно редактировать:

- добавлять текст;
- редактировать текст;
- удалять часть текстовой информации.

Отредактированный проект документа можно сохранить, нажав на кнопку «» (Сохранить). Редактирование проекта документа можно отменить, нажав на кнопку «» (Отмена).

3) Область отображения документа, с которым происходит сравнение (Рисунок 20, поз. 3);

4) Область отображения похожих документов (Рисунок 20, поз. 4). В данной области в виде предпросмотра отображаются документы, на которые похож сравниваемый проект документа.

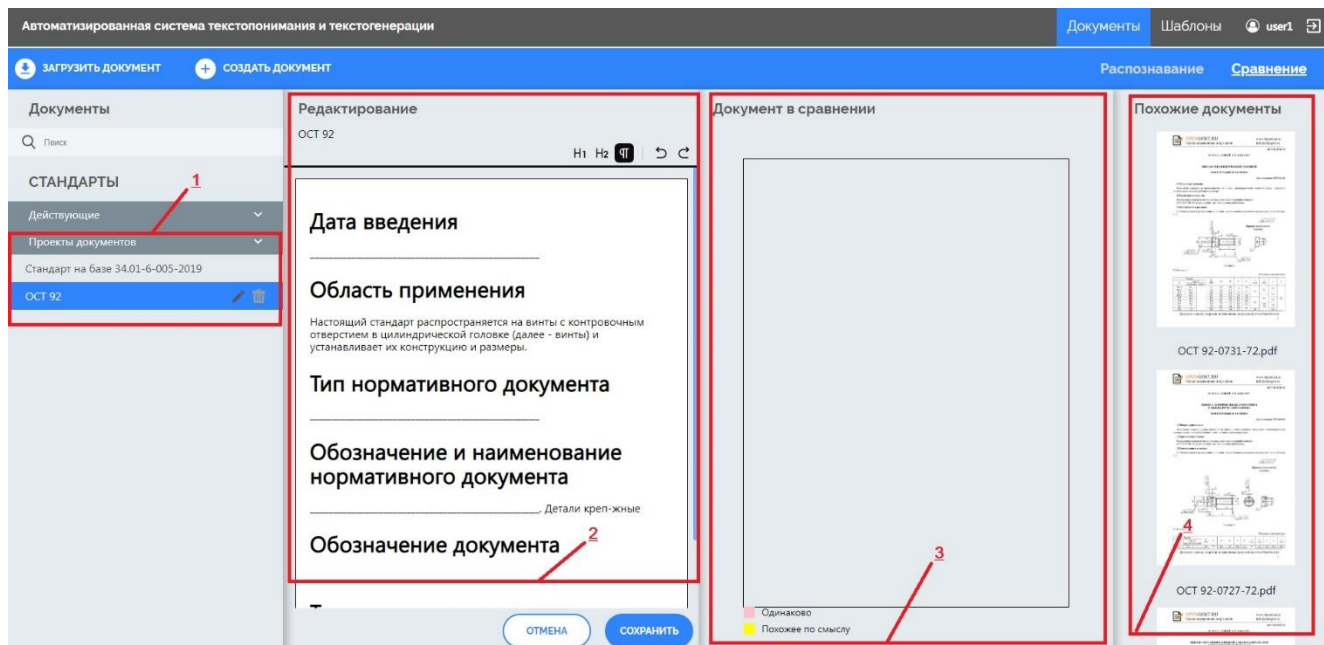


Рисунок 20

Для выполнения сравнения необходимо выбрать проект документа в списке «Проекты документов», один раз щелкнув ЛКМ на его названии. Содержимое выбранного проекта документа будет отображаться в окне «Редактирование» (Рисунок 20, поз. 2).

В области отображения похожих документов следует выбрать документ для сравнения, один раз щелкнув ЛКМ на его изображении. Содержимое документа откроется в области «Документ в сравнении» (Рисунок 21).

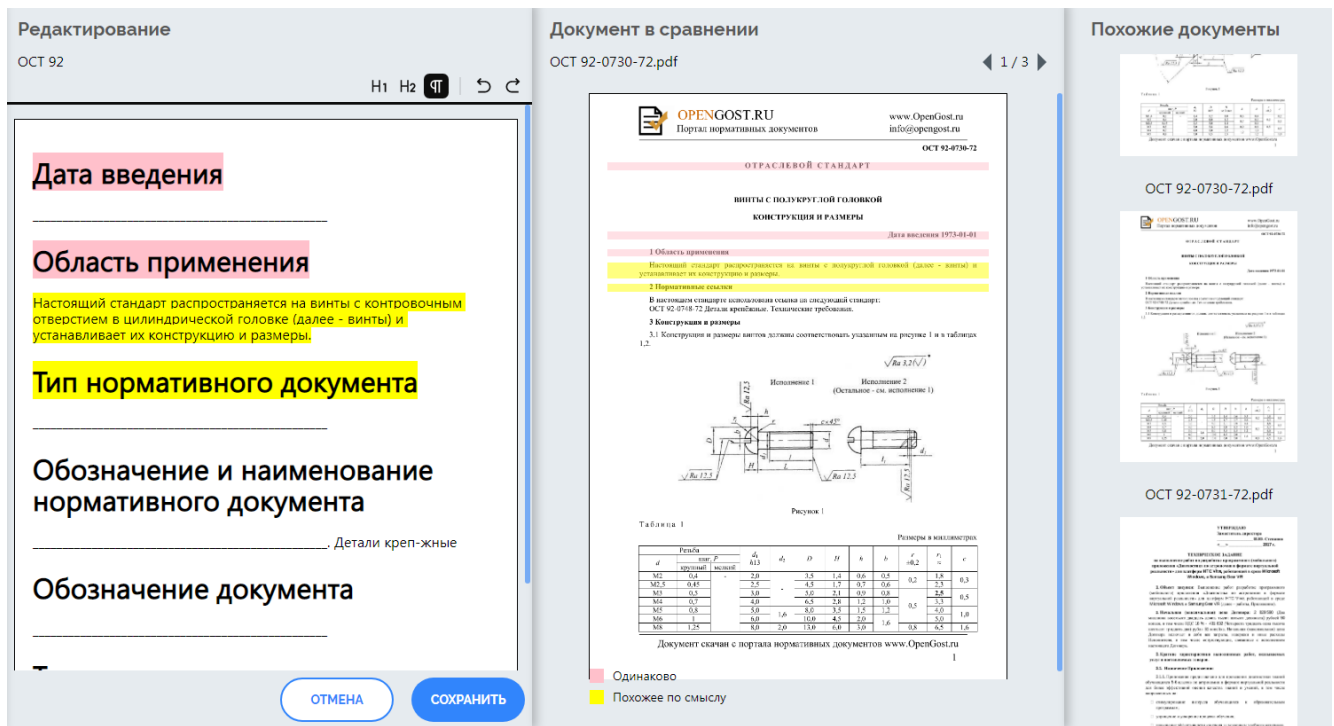


Рисунок 21

Сравнение проекта документа с действующим документом выполняется автоматически после выбора документа для сравнения.

Во время выполнения процесса сравнения окно Системы становится недоступным для пользователя и появляется сообщение об ожидании завершения процесса (Рисунок 22).

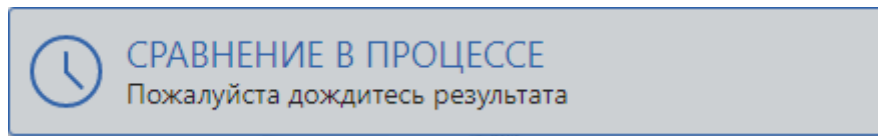


Рисунок 22

После выполненного сравнения в тексте в области «Редактирование» и «Документ в сравнении» цветом выделяются найденные похожие части текста: красным отображаются одинаковые части, желтым отображаются семантически похожие части (Рисунок 21).

После редактирования теста проекта документа с последующим сохранением изменений, в области «Редактирование» автоматически запускается повторное сравнение проекта документа и действующего документа (Рисунок 23).

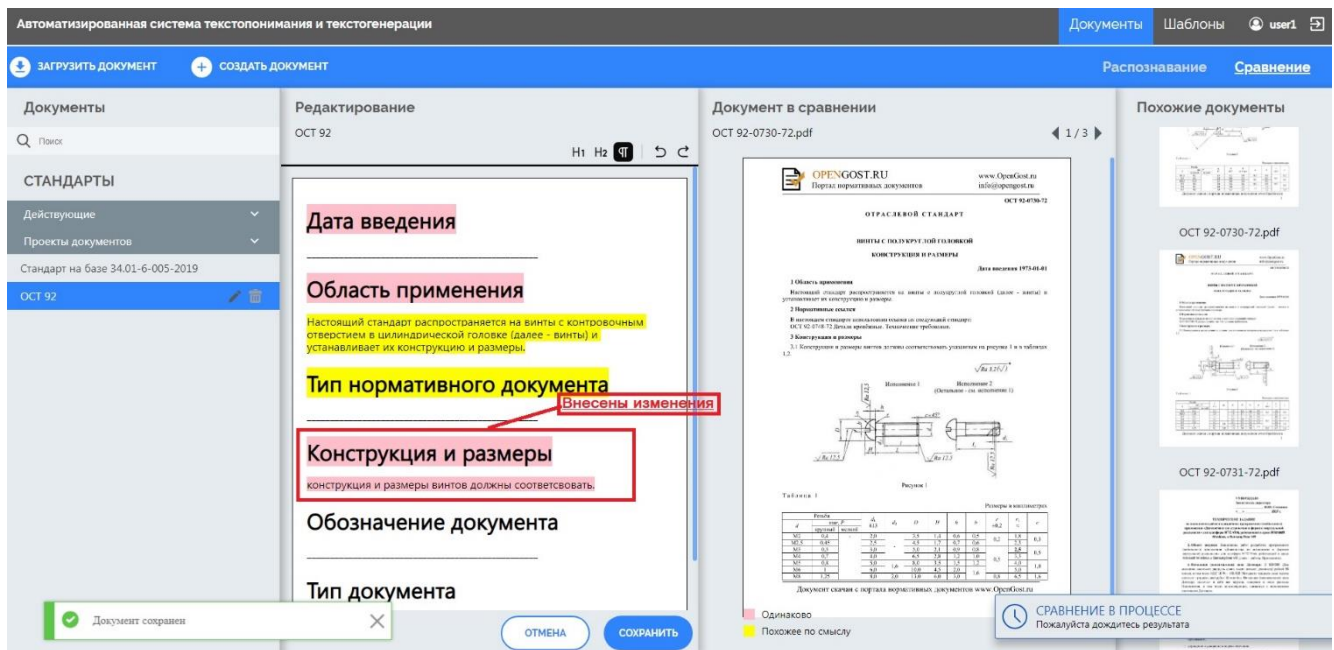


Рисунок 23

После выполненного повторного сравнения в тексте в области «Редактирование» и «Документ в сравнении» цветом выделяются найденные похожие части текста: красным отображаются одинаковые части, желтым отображаются семантически похожие части.

## 1.10. Удаление документов

Для удаления документа из Системы необходимо:

1. Выбрать документ в списке документов, щелкнув один раз на его названии ЛКМ. Выбранный документ будет подсвечиваться синим цветом (Рисунок 24).

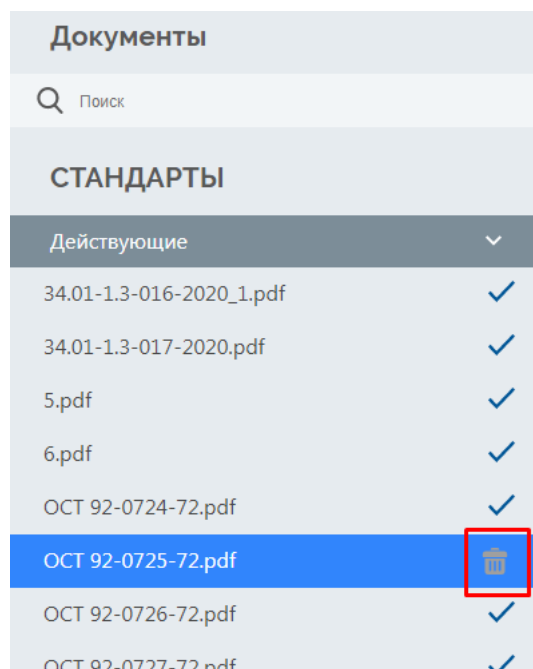



Рисунок 24

2. Нажать на значок «» справа от наименования файла документа.

## 1.11. Поиск документов

Для поиска документа в списке необходимо в поле «Поиск» ввести название файла документа (Рисунок 25).

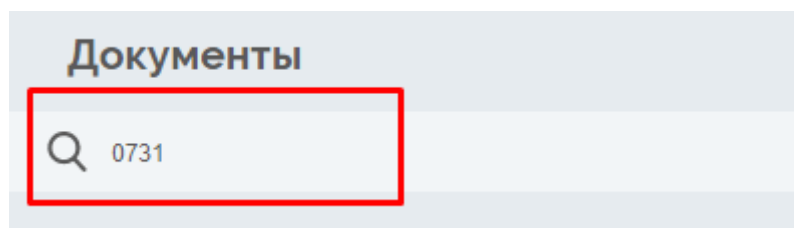


Рисунок 25

При вводе символов в строку поиска отфильтровываются документы, содержащие указанную последовательность символов в названии.

## 1.12. Работа с шаблонами документов

Раздел «Шаблоны» предназначен для:

- загрузки готовых шаблонов в Систему;
- создания, редактирования шаблонов.

Для открытия раздела необходимо на главной странице Системы нажать на название раздела «Шаблоны» (Рисунок 26).

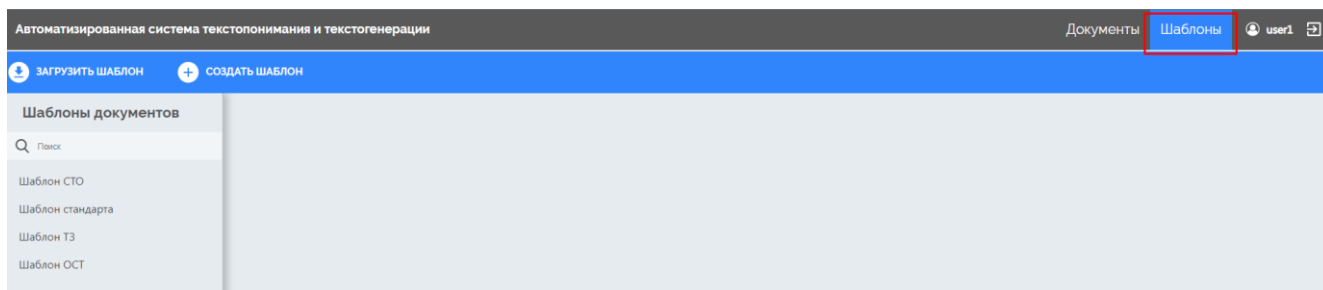



Рисунок 26

## 1.13. Создание шаблонов

Для создания шаблона необходимо нажать на кнопку « Создать шаблон» (Создать шаблон). Созданный шаблон будет отображаться последним в списке шаблонов документов (Рисунок 27).

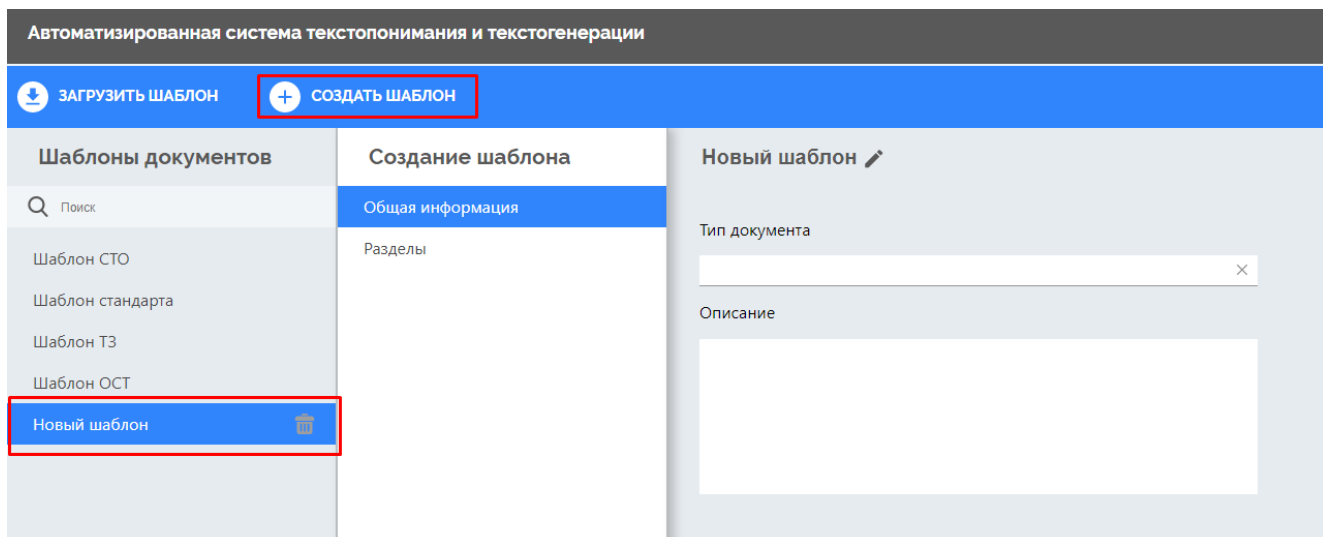


Рисунок 27

При необходимости можно отредактировать наименование шаблона документа, нажав на знак «✎» справа от его наименования (Рисунок 28).

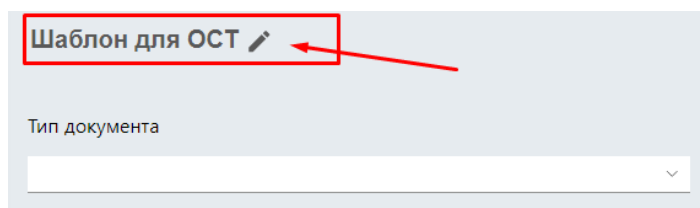


Рисунок 28

Далее заполните обязательные поля в разделе «Общая информация»:

- Нажмите на поле «тип документа» из выпадающего списка выберите нужный тип документа (Рисунок 29);

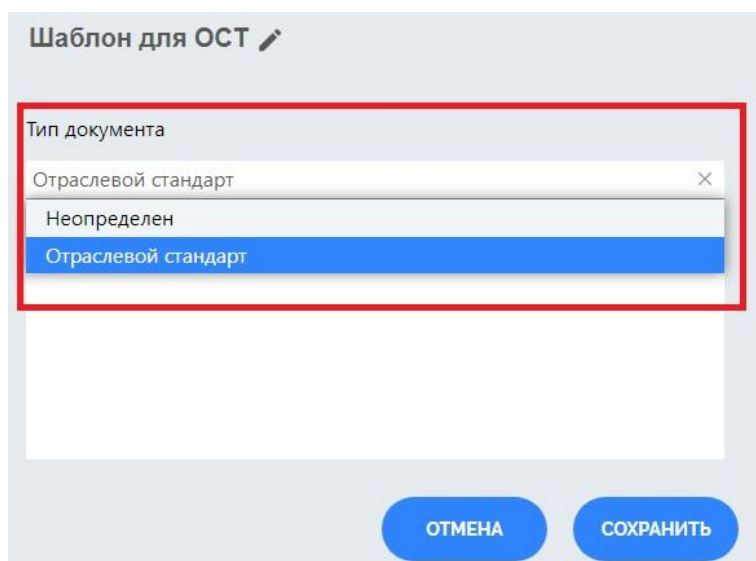


Рисунок 29

- Заполните поле «Описание» (Рисунок 30).

Шаблон для ОСТ ✎

Тип документа  
Отраслевой стандарт ✕

Описание  
Шаблон для ОСТ

ОТМЕНА СОХРАНИТЬ

Рисунок 30

Для сохранения внесенных изменений нажмите на кнопку «Сохранить» (Рисунок 30). При успешном сохранении шаблона система отобразит информирующее окно «Шаблон создан», и внесет в шаблон информацию о создателе шаблона (Рисунок 31).

Автоматизированная система текстовопонимания и текстогенерации | Документы | Шаблоны | user1

ЗАГРУЗИТЬ ШАБЛОН + СОЗДАТЬ ШАБЛОН

Шаблоны документов  
Поиск  
Шаблон СТО  
Шаблон стандарта  
Шаблон ТЗ  
Шаблон ОСТ  
Шаблон для ОСТ

Создание шаблона  
Общая информация  
Разделы

Шаблон для ОСТ ✎

Тип документа  
Отраслевой стандарт ✕

Описание  
Шаблон для ОСТ

Создал  
Петров Никита

ОТМЕНА СОХРАНИТЬ

Шаблон создан ✕

Рисунок 31

Для дальнейшего создания шаблона перейдите в раздел «Разделы» (Рисунок 32).

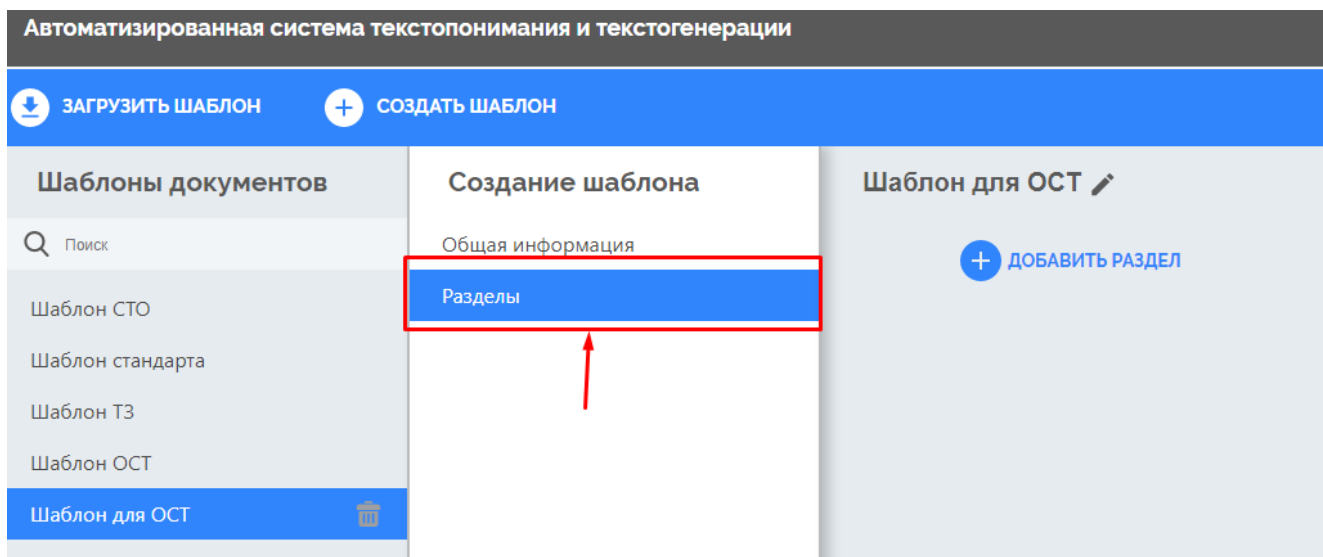



Рисунок 32

Для наполнения раздела информацией нажмите на кнопку  «Добавить раздел». В открывшемся поле укажите все важные данные из документа в шаблон для правильного прохождения сравнения и распознавания текста системой (Рисунок 33).

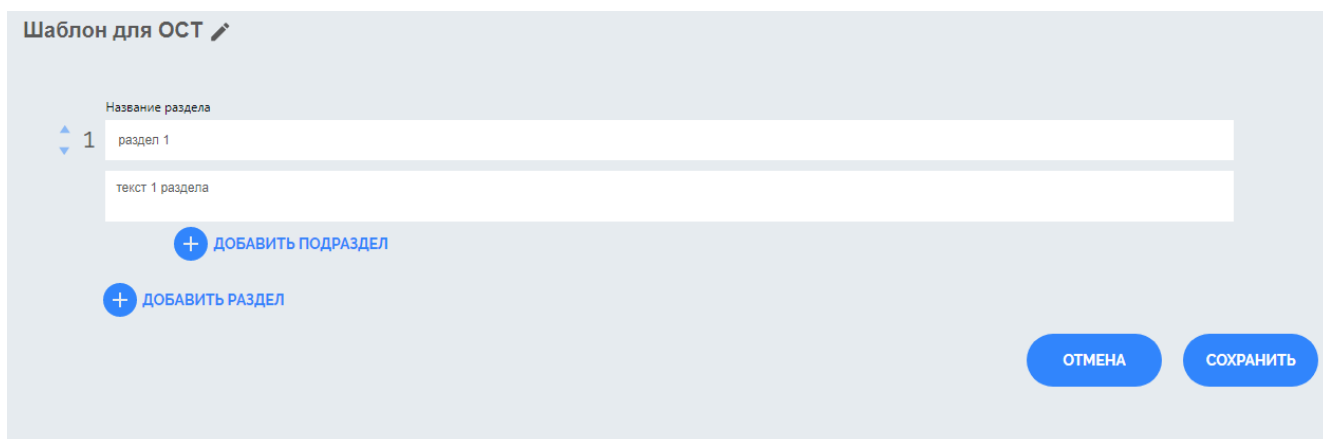


Рисунок 33

Добавляйте и заполняйте последовательно разделы и подразделы. Укажите название раздела и текст раздела в соответствующих полях. Для применения выполненных изменений следует нажать на кнопку «Сохранить» (Рисунок 34).



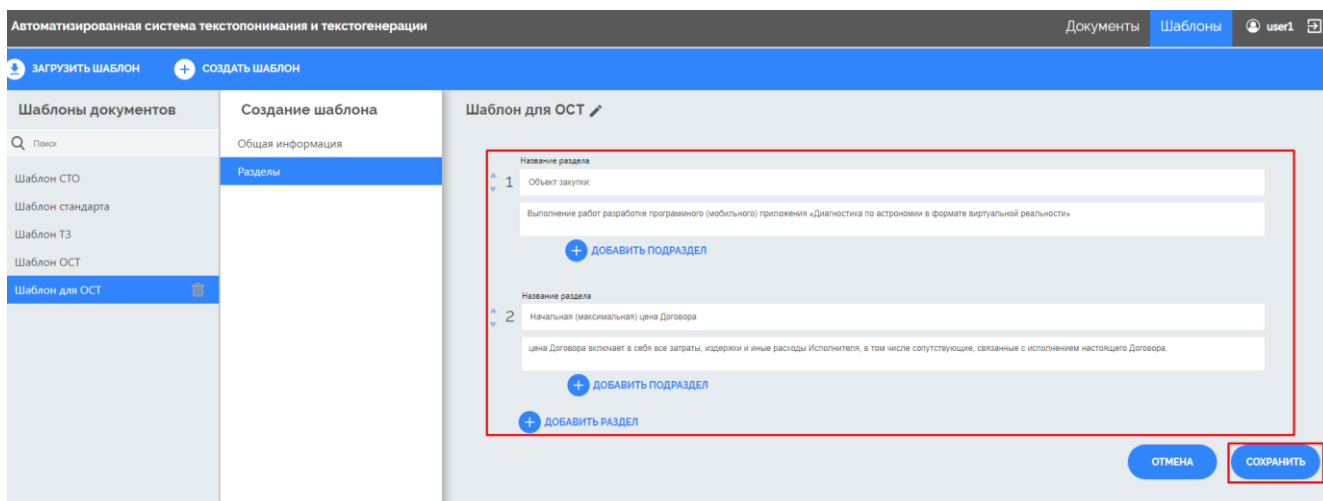


Рисунок 34

## 1.14. Редактирование шаблонов

Для редактирования шаблона необходимо:

1. Выбрать шаблон в списке шаблонов, щелкнув один раз на его названии ЛКМ.

Выбранный шаблон будет подсвечиваться синим цветом (Рисунок 35);

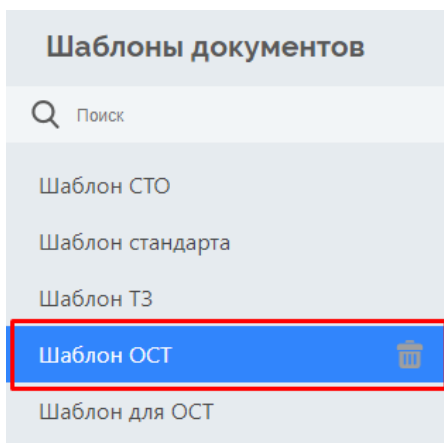


Рисунок 35

2. На странице редактирования шаблона в окне «Создание шаблона» выбрать нужный раздел, щелкнув один раз ЛКМ на его названии (Рисунок 36);

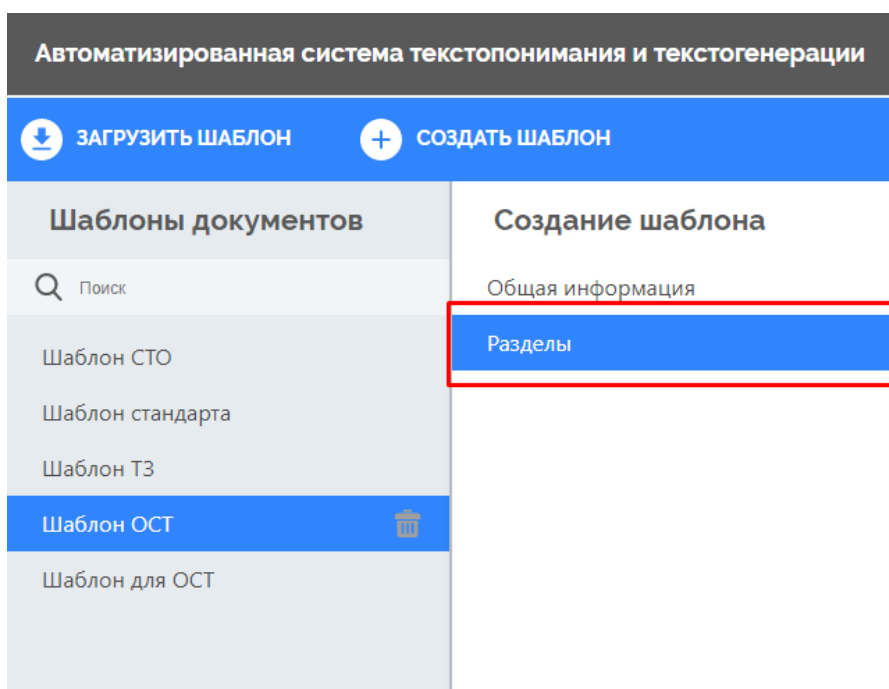


Рисунок 36

3. В разделе «Общая информация» вы можете изменить сведения о типе шаблона документа (в поле «Тип документа»), а так же изменить его краткое описание в поле «Описание» (Рисунок 37);

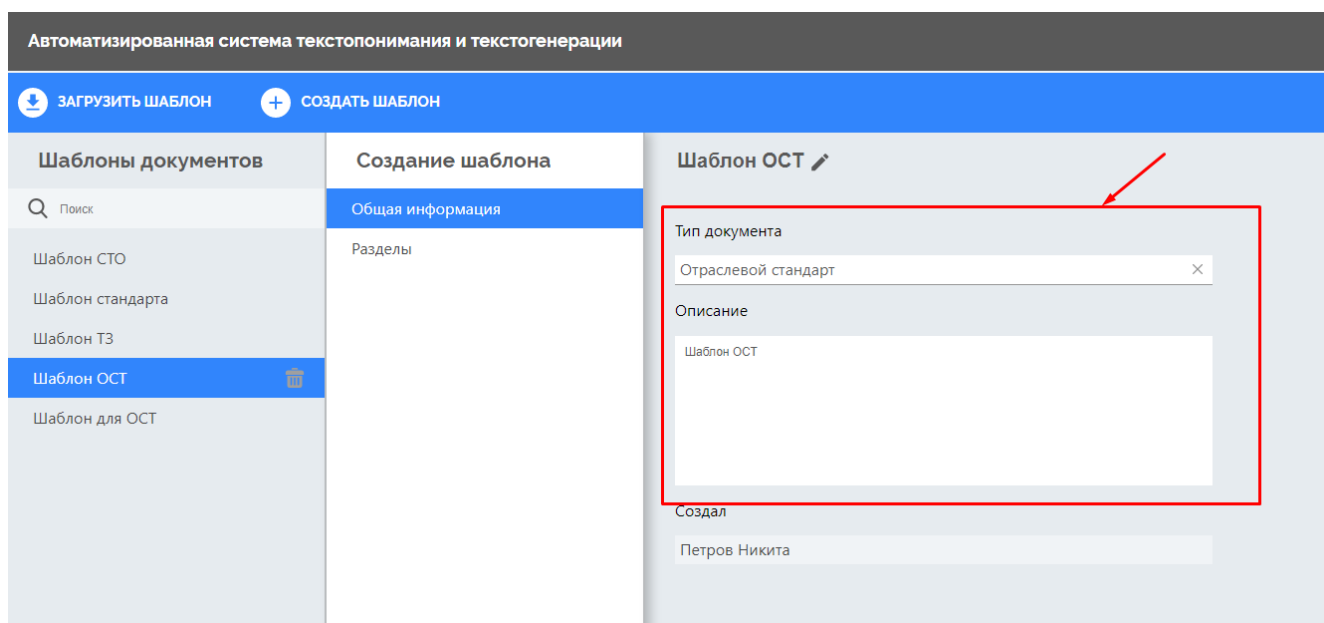


Рисунок 37

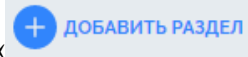
4. Для редактирования разделов шаблона документа, нажмите один раз ЛКМ на раздел (Рисунок 38);

## Создание шаблона

Общая информация

Разделы

Рисунок 38

4.1 Для добавления новых разделов нажмите кнопку «  » (Добавить раздел). А для редактирования разделов нажмите на поле с нужным разделом и укажите название раздела и текст раздела в соответствующих полях (Рисунок 39);

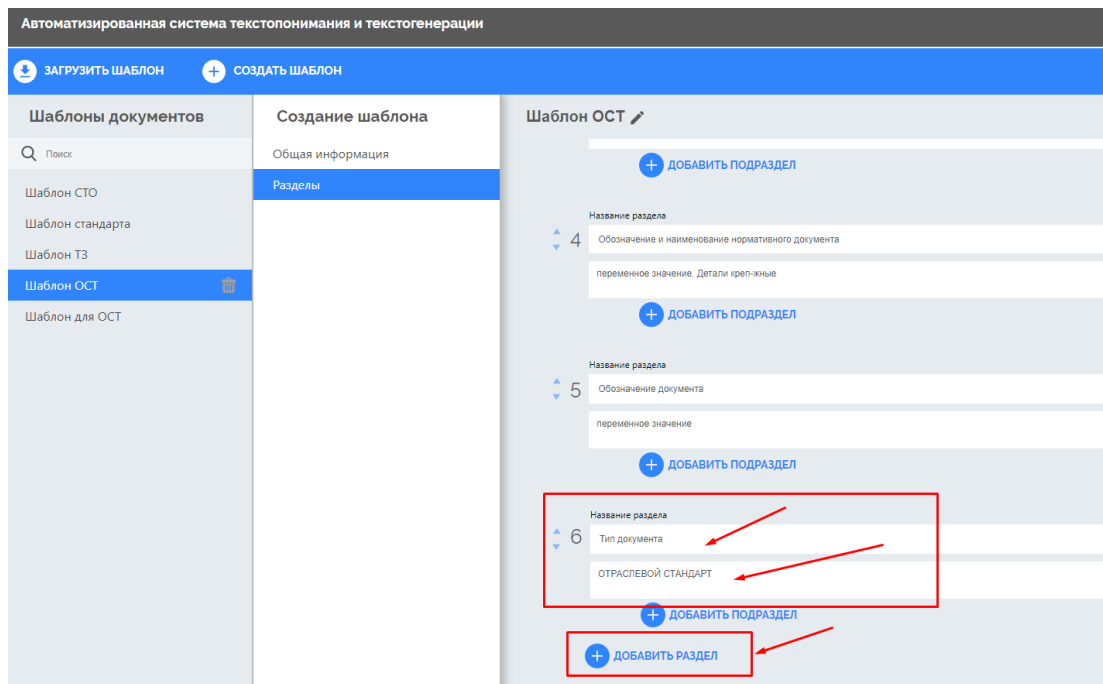



Рисунок 39

4.3 Для добавления новых подразделов нажмите кнопку «  » (Добавить подраздел). А для редактирования подразделов нажмите на поле с нужным подразделом и укажите название подраздела и текст подраздела в соответствующих полях (Рисунок 40);

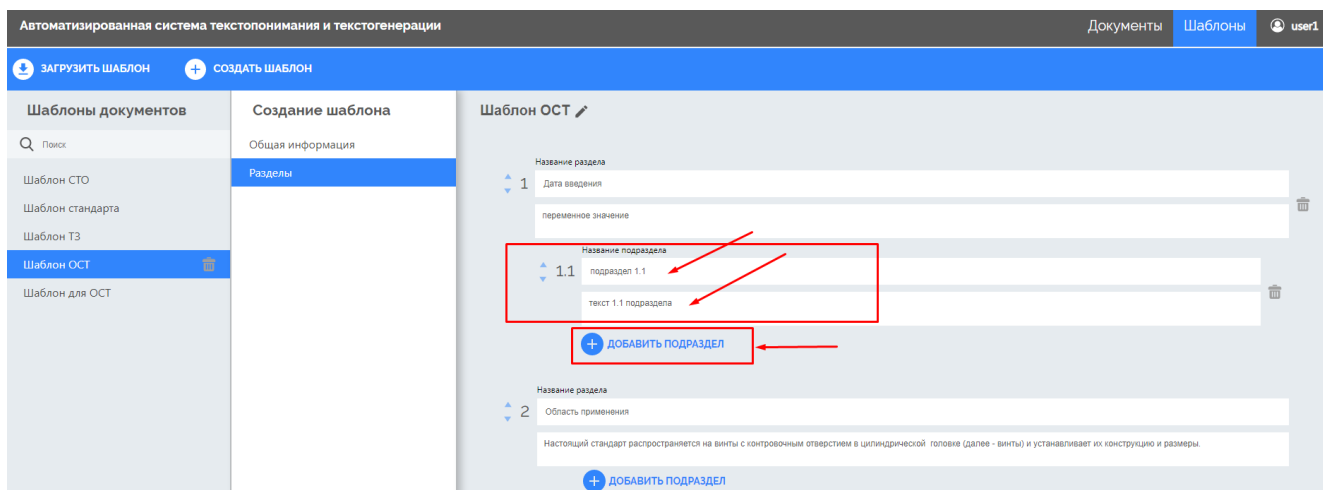


Рисунок 40

Для применения выполненных изменений нажмите кнопку «Сохранить»

СОХРАНИТЬ

### 1.15. Удаление шаблонов

Для удаления шаблона документа из Системы выполните:

1. Выберите документ в списке документов, щелкнув один раз на его названии ЛКМ. Выбранный шаблон будет подсвечиваться синим цветом (Рисунок 41);

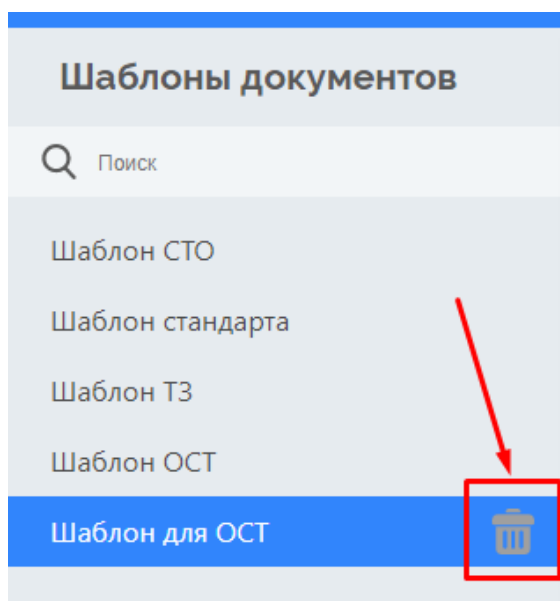



Рисунок 41

2. Для удаления шаблона из системы нажмите на значок  «корзина», справа от наименования шаблона документа.

### 1.16. Поиск шаблона

Для поиска документа в списке необходимо в поле «Поиск» ввести название шаблона документа (Рисунок 42).

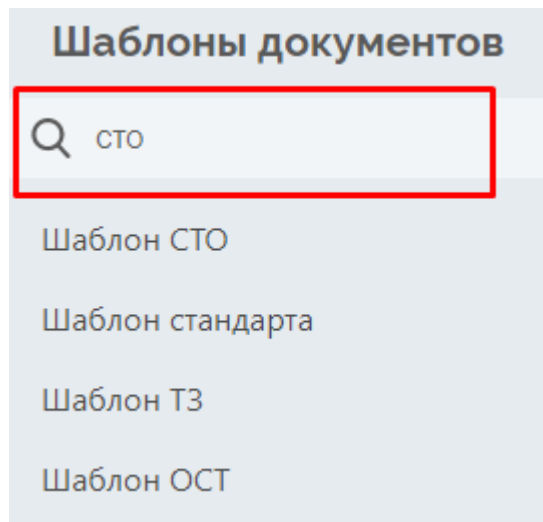


Рисунок 42

При вводе символов в строку поиска отфильтровываются шаблоны документов, содержащие указанную последовательность символов в названии.

## 2. Конфигуратор извлекаемых данных

### 2.1. Краткое описание Подсистемы

Работа с RDF-графами выполняется в подсистеме семантического анализа документов в редакторе RDF-графов (далее – Подсистема). Данная Подсистема предназначена для работы с онтологиями в графическом редакторе, описывающими входящие документы, и представляет собой графический интерфейс для обеспечения создания и управления предметно-ориентированной онтологией для выполнения следующих задач:

- создание предметно-ориентированной онтологии;
- редактирование предметно-ориентированной онтологии;
- удаление предметно-ориентированной онтологии;
- хранение данных;
- прием/передача данных;
- получение массива извлекаемых данных;
- обеспечение доступа к предметно-ориентированной онтологии;
- визуализация предметно-ориентированной онтологии.

RDF-граф получает текст документа (file) и информацию о типе документа (classified\_type) поступившие от классификатора корпоративной документации. После проведенного импорта запускается поиск узлов в нейросети для поиска онтологии, соответствующей указанному типу документа. После выполненного поиска в соответствии с типом документа (например «PROTOCOL\_B») выгружается граф, содержащий:

- атрибуты, подлежащие извлечению (AttrName);
- регулярные выражения для поиска предложений, вероятно содержащих искомый атрибут (RegExpFilter);
- yargy-правила для извлечения атрибута из массива найденных предложений (YargyRule, YargyRegExpFilter);
- регулярные правила для извлечения атрибутов (RegExpRule).

Полученный ассоциативный массив данных, где ключами являются названия атрибутов, а значениями – правила для применения (yargy-правила и регулярные правил) и документ передаются в производственный текстопониматель для дальнейшей обработки.

## 2.2. Запуск Подсистемы

Работа с Подсистемой осуществляется в веб-интерфейсе по протоколу HTTP. Пользователь вводит в адресной строке браузера соответствующий адрес расположения сервера и после этого по умолчанию будет открыта страница редактора RDF-графов.

## 2.3. Интерфейс Подсистемы

Интерфейс Подсистемы состоит из нескольких частей (Рисунок 43).

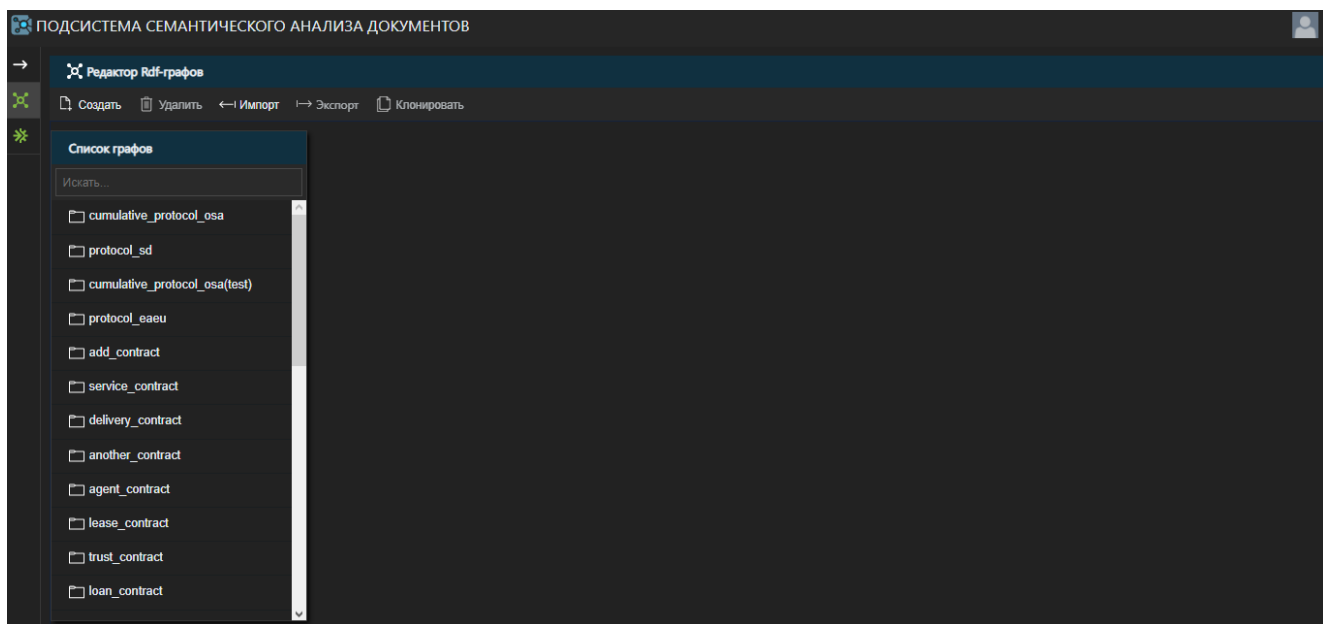


Рисунок 43

В верхней левой части страницы располагается полное название Подсистемы и меню для редактирования RDF-графа с кнопками:

- «Создать». Позволяет создать новый RDF-граф.
- «Удалить». Позволяет удалить выбранный RDF-граф.
- «Импорт». Позволяет выгрузить выбранный RDF-граф в файл в формате HDD.
- «Экспорт». Позволяет загрузить в Подсистему файл в формате JSON с описанием RDF-графа.
- «Клонировать». Позволяет создать копию существующего RDF-графа.

В левой части страницы находится список RDF-графов (Рисунок 44).

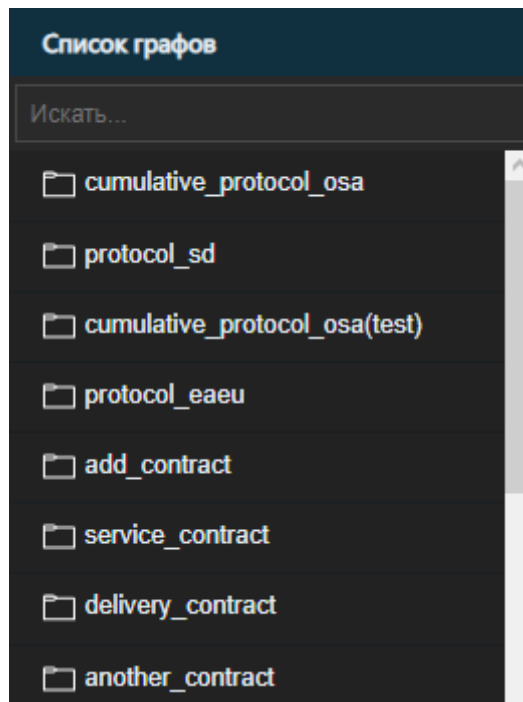


Рисунок 44

В центральной части экрана расположена область отображения RDF-графов

## 2.4. Просмотр RDF-графов

Для просмотра RDF-графа необходимо выбрать его в списке, щелкнув один раз ЛКМ на его названии. Название выбранного RDF-графа в списке будет подсвечено серым цветом (Рисунок 45)

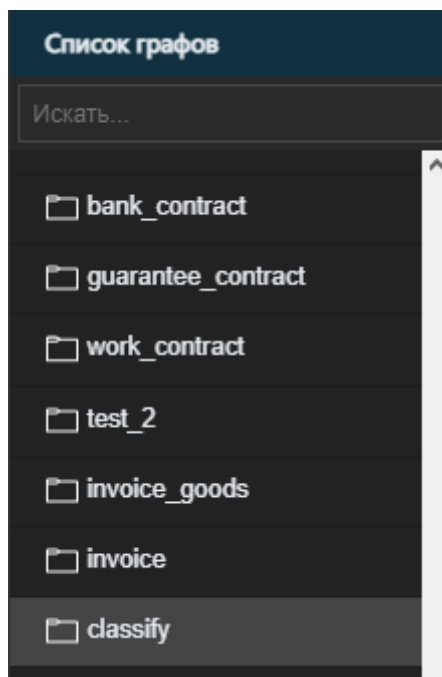


Рисунок 45



Выбранный RDF-граф будет отображаться в области справа от списка RDF-графов (Рисунок 46).

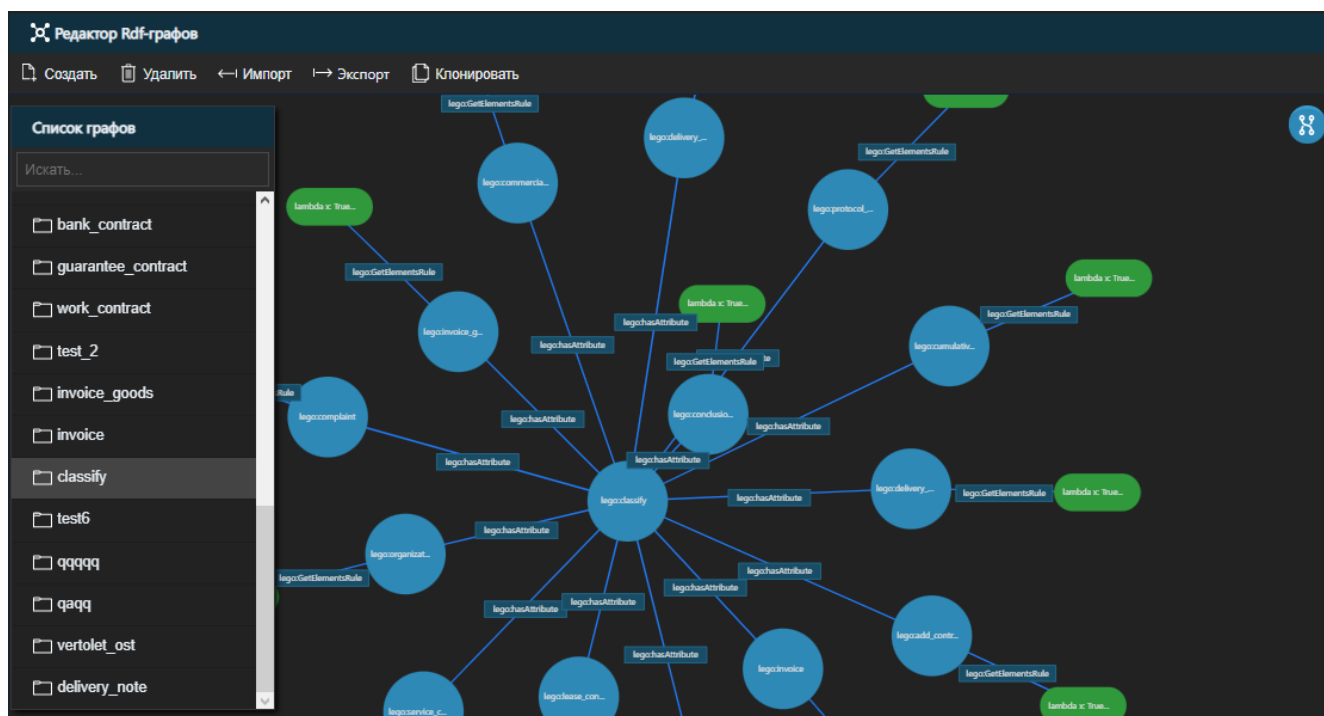


Рисунок 46

В данной области отображаются узлы, литералы и отношения RDF-графов. Узлы RDF-графа визуально отображаются в виде окружностей, а литералы – в виде овалов. Также выполнено цветовое выделение элементов графа: узлы графа выделены синим цветом, литералы – зеленым.

Для управления отображением RDF-графа созданы элементы интерфейса, позволяющие:

- скрывать атрибуты предметно-ориентированной онтологии;
- визуализировать узлы предметно-ориентированной онтологии;
- просматривать свойства элемента предметно-ориентированной онтологии:

наименование атрибутов, уагу-правила и регулярные выражения.




При нажатии на кнопку «» можно скрыть все элементы RDF-графа кроме узлов. Результат отображения в таком режиме показан на рисунке 47.



Рисунок 47

Для просмотра свойств элемента RDF-графа следует щелкнуть один раз ЛКМ на нужном элементе. Свойства выбранного элемента будут отображаться в окне «Свойства» справа от области отображения графов (Рисунок 48).

**cumulative\_protocol\_osa: Свойства**

**ВЫРАЖЕНИЕ 1**

Общее количество голосов [.\n\W\daA-яя\s]\*? Общества ((.\n)\*?)\.

**ВЫРАЖЕНИЕ 2**

(Число\sgолосов,\skоторыми\собладали [.\WaA-яя\s\n\d]\*?)собрании

**ВЫРАЖЕНИЕ 3**

(Число\sgолосов,\skоторыми\собладали [.\WaA-яя\s\n\d]\*?собрании, для голосования .\*)

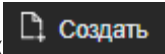
**ВЫРАЖЕНИЕ 4**

(([Ч]исло| количество)\sgолосов,\skоторыми\sobлада[лиют]{2}[.\WaA-яя\s\n\d]\*?собрании[aA-яя\s\n\W]\*?, [aA-яя\d\s\W\n]\*?голос)

Рисунок 48

Для атрибутов свойствами является название атрибута, для литералов узлов свойствами являются уагу-правила и регулярные выражения.

## 2.5. Создание RDF-графов

Для создания RDF-графа необходимо в меню нажать на кнопку « Создать» (Создать).

В результате создания в Подсистему добавляется RDF-граф с двумя узлами и связью между ними (Рисунок 49).

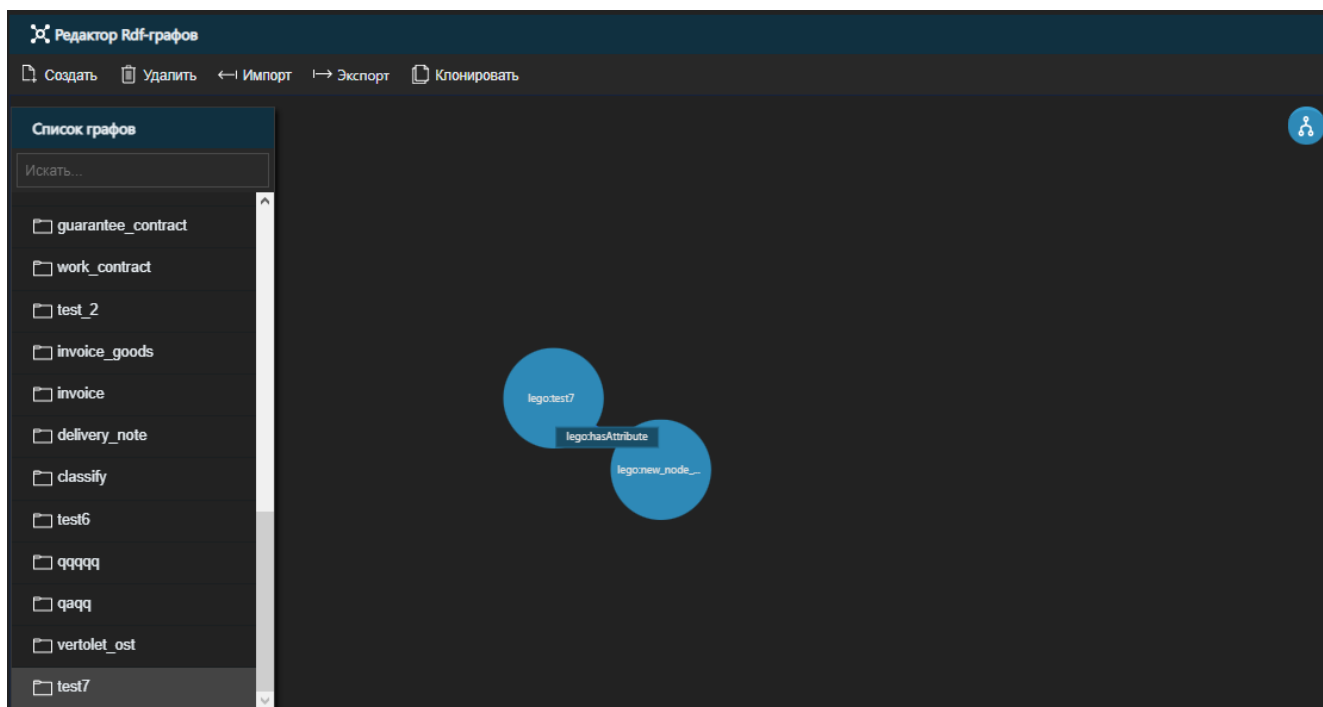


Рисунок 49

## 2.6. Редактирование RDF-графов

### 2.6.1. Добавление узла

Для добавления узла следует выполнить следующий порядок действий:

1. В области отображения графа выбрать базовый узел, щелкнув один раз на нем ЛКМ. Выбранный узел будет подсвечиваться синим цветом (Рисунок 50).

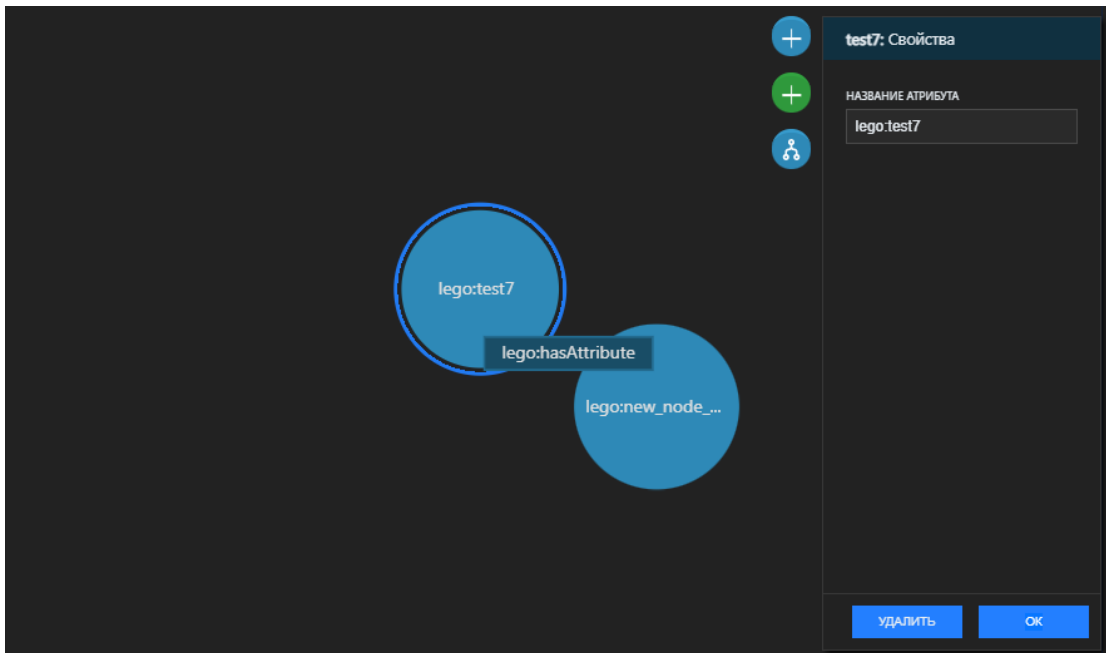



Рисунок 50

2. В поле редактора графа нажать кнопку добавления узла «».
3. В открывшемся окне в поле «Название» ввести название нового узла, в поле «Тип связи» выбрать связь (Рисунок 51).

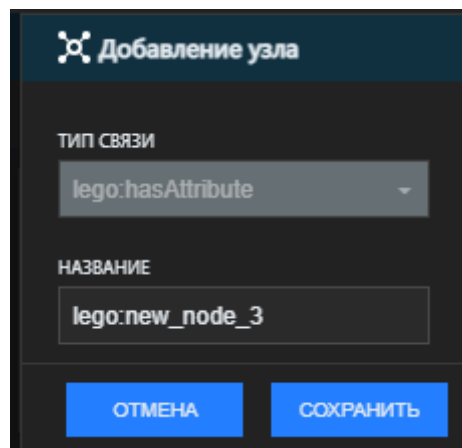



Рисунок 51

4. Нажать на кнопку «Сохранить».

### 2.6.2. Добавление литерала

Для добавления литерала следует выполнить следующий порядок действий:

1. В области отображения графа выбрать узел, щелкнув один раз на нем ЛКМ. Выбранный узел будет подсвечиваться синим цветом.
2. В поле редактора графа нажать кнопку добавления литерала «».
3. В окне «Добавление литерала» в поле «Название» ввести название нового атрибута, в поле «Тип связи» выбрать соответствующий тип связи (Рисунок 52).

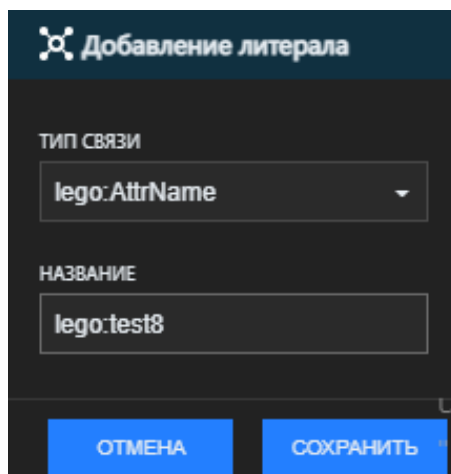
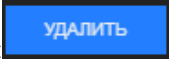


Рисунок 52

4. Нажать на кнопку «Сохранить».

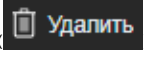
### 2.6.3. Удаление элементов RDF-графа

Для удаления элемента RDF-графа следует выполнить следующий порядок действий:

1. В области отображения RDF-графа выбрать базовый узел, щелкнув один раз на нем ЛКМ. Выбранный узел будет подсвечиваться синим цветом (Рисунок 50)
2. В поле редактора графа нажать кнопку «» (Удалить).
3. В окне подтверждения удаления нажать на кнопку «ОК» для удаления элемента.


### 2.7. Удаление RDF-графа

Для удаления RDF-графа следует выполнить следующий порядок действий:

1. В списке RDF-графов выбрать необходимый RDF-граф, щелкнув один раз ЛКМ на его названии.
2. В меню нажать на кнопку «» (Удалить).
3. В окне подтверждения удаления нажать на кнопку «ОК» для удаления RDF-графа.

### 2.8. Импорт RDF-графа

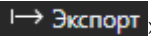
Импорт RDF-графов предназначен для загрузки графов в Подсистему из файла формата JSON.

Для импорта RDF-графа необходимо в меню нажать на кнопку «» (Импорт).

## 2.9. Экспорт RDF-графа

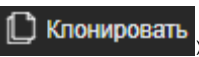
Экспорт RDF-графов предназначен для выгрузки графов из Подсистемы в файл формата JSON.

Для экспорта RDF-графа необходимо выполнить следующий порядок действий:

1. В списке RDF-графов выбрать необходимый RDF-граф, щелкнув один раз ЛКМ на его названии.
2. В меню нажать на кнопку « Экспорт» (Экспорт).
3. Сохранить полученный файл.

## 2.10. Создание RDF-графов на основе аналогов

В Подсистеме можно создавать RDF-графы на основе существующих. Для создания RDF-графа на основе аналога необходимо выполнить следующий порядок действий:

1. В списке RDF-графов выбрать необходимый RDF-граф, щелкнув один раз ЛКМ на его названии.
2. В меню нажать на кнопку « Клонировать» (Клонировать).
3. В окне «Клонирование графа» указать имя RDF-графа (Рисунок 53).

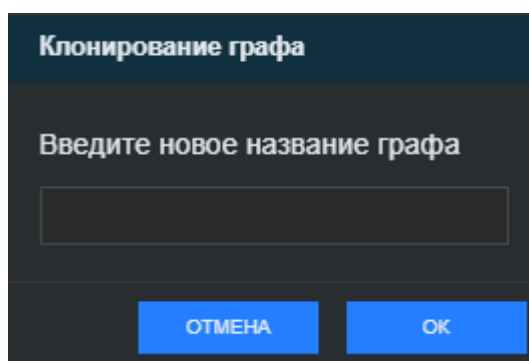


Рисунок 53

4. Нажать на кнопку «ОК».

## 3. Классификатор корпоративной документации

### 3.1. Краткое описание Подсистемы

Подсистема «Классификатор корпоративной документации» обеспечивает возможность классификации различных видов корпоративных документов с возможностью обучения моделей машинного обучения, а именно подбора, обучения и стекинга моделей для классификации поступающих в систему видов корпоративной документации и решает следующие задачи:

- определение набора классификационных признаков в тексте документа;
- машинное обучение моделей;
- построение классификатора корпоративной документации;
- обучение классификатора корпоративной документации;


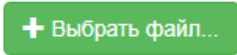
### 3.2. Запуск Подсистемы

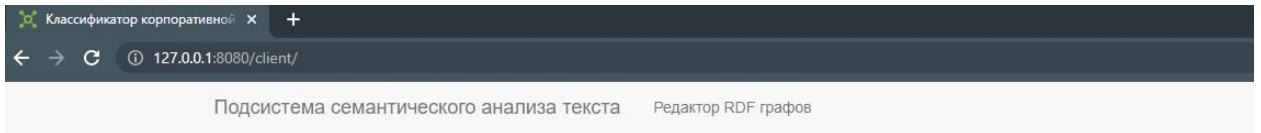
Работа с Подсистемой осуществляется в веб-интерфейсе по протоколу HTTP. Пользователь вводит в адресной строке браузера соответствующий адрес расположения сервера и после этого по умолчанию будет открыта страница

### 3.3. Интерфейс Подсистемы

Интерфейс Подсистемы изображен на Рисунке 54, описание активных кнопок Подсистемы находится в таблице ниже.

<i>Кнопка в подсистеме</i>	<i>Описание кнопки</i>
	При нажатие на кнопку Вы перейдете в подсистему «Модуль обработки корпоративной документации»
	При нажатие на кнопку Вы перейдете в подсистему «Конфигуратор извлекаемых данных»
	При нажатие на кнопку подсистема откроет выпадающий список с типами документов

	При нажатие на кнопку подсистема запустит загрузку выбранного документа
	При нажатие на кнопку подсистема откроет окно с поиском документов на вашем ПК



## Классификатор корпоративной документации


Файлы могут быть загружены или переташены на параметры

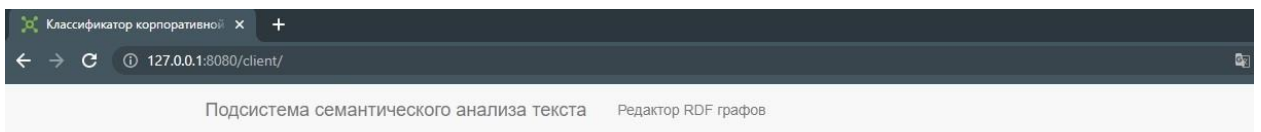
Протокол или решение




Рисунок 54

### 3.4. Работа с подсистемой

Для автоматизированного обучения по распознаванию данных в документе, выберите тип документа. Нажмите на кнопку  и из выпадающего списка выберите нужный тип документа для обучения (Рисунок 55).



## Классификатор корпоративной документации

Файлы могут быть загружены или переташены на параметры

Протокол или решение

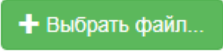
Протокол или решение

Договор




Рисунок 55



После выбора типа документа выберите на Вашем ПК документ на котором будет обучаться подсистема, для этого нажмите на кнопку «Выбрать файл» , как показано на рисунке 56.

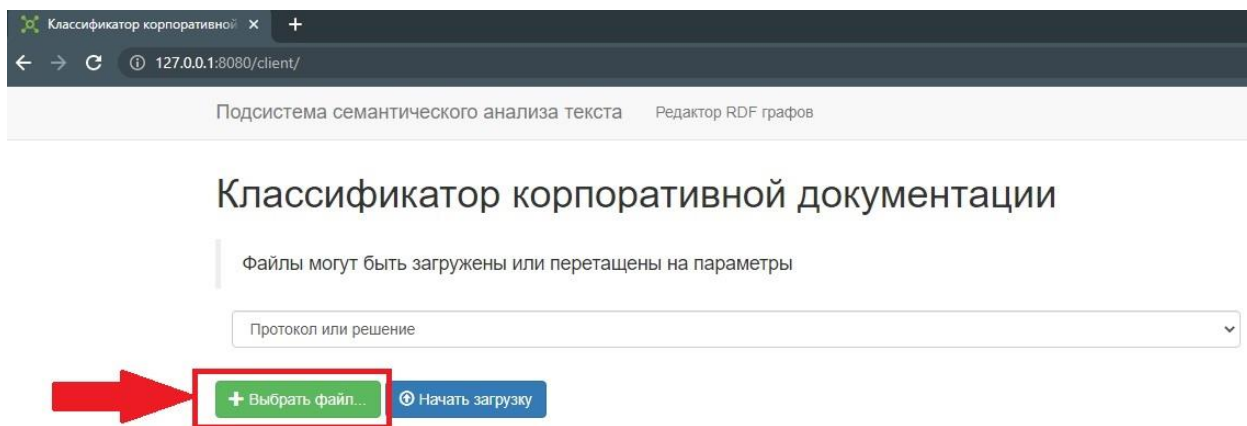


Рисунок 56

В открывшемся окне выберите нужный документ и загрузите его в подсистему, нажав на кнопку «Открыть», как показано на рисунке 57.

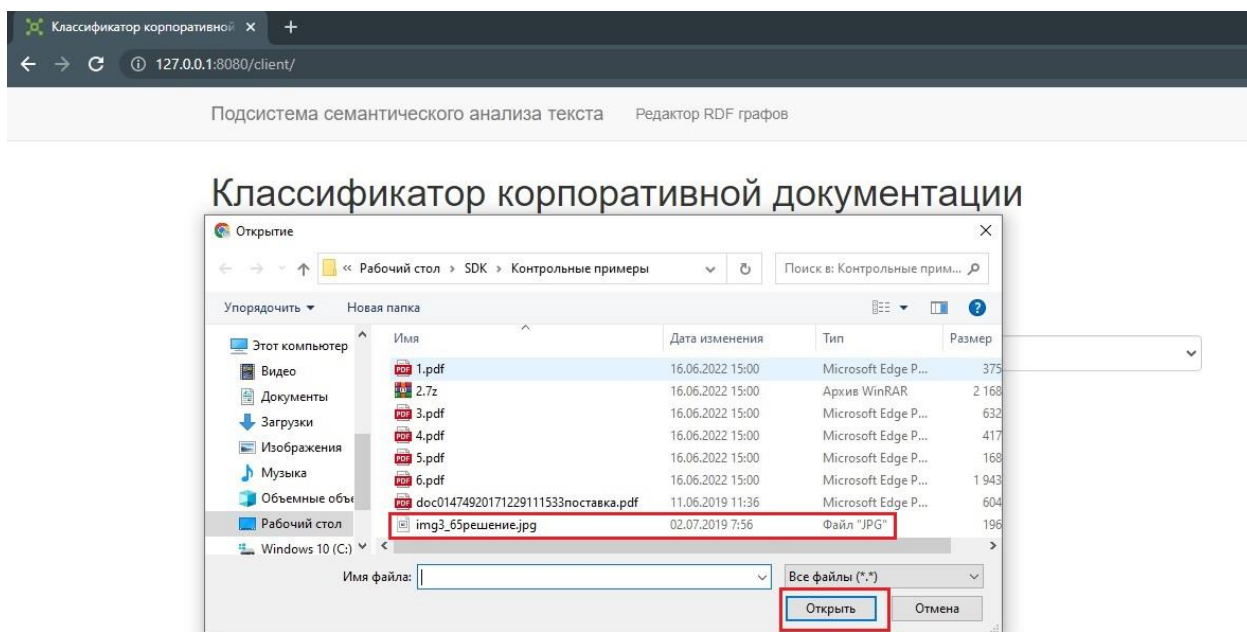



Рисунок 57

После загрузки документа в подсистему пользователю станет доступна возможность начала обучения. Для этого нажмите на кнопку «Начать» , как показано на рисунке 58.

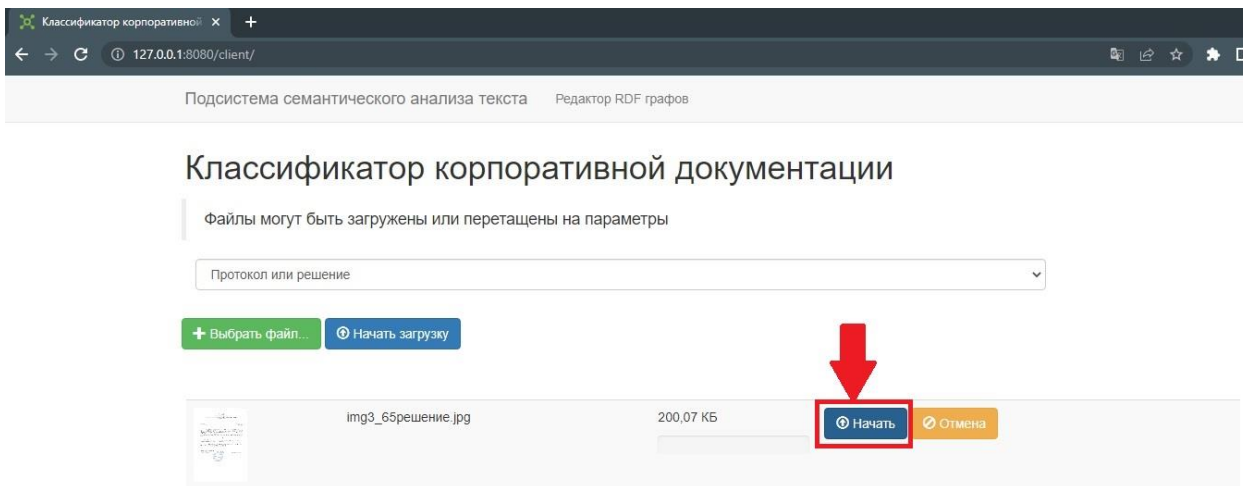



Рисунок 58

После нажатия на кнопку «Начать»  Подсистема запустит автоматическое обучение, а на экране появится индикатор выполнения обучения, как показано на рисунке 59. Обучение займет некоторое время, не закрывайте страницу пока не закончится обучение.

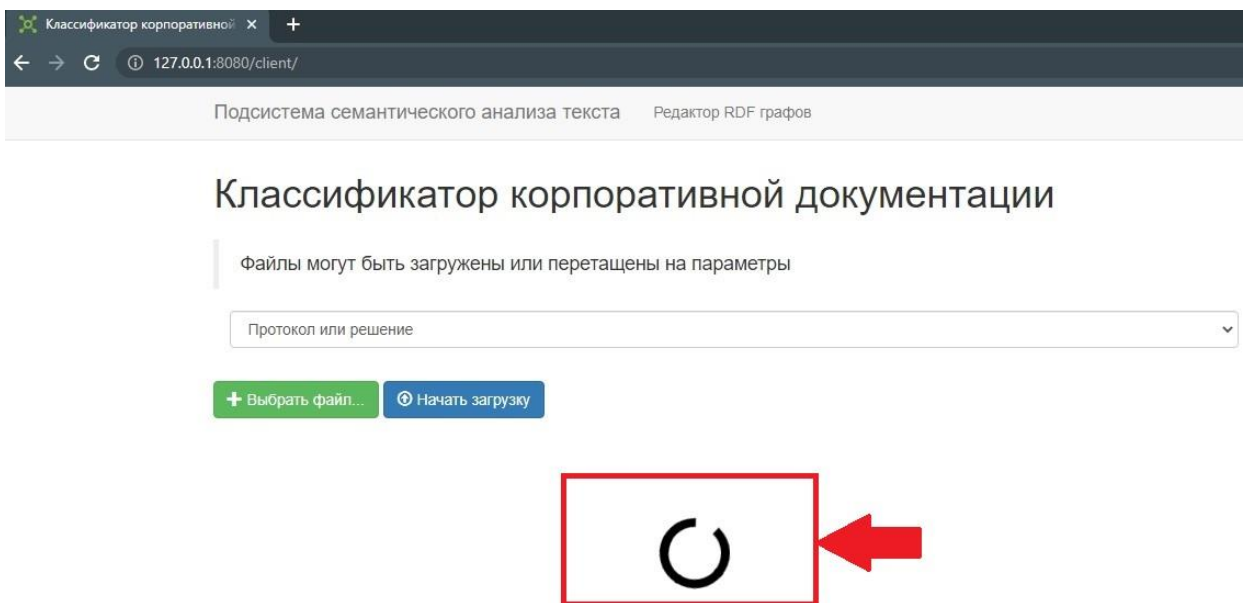


Рисунок 59

После завершения обучения Подсистема сформирует автоматическое сообщение о выполненном процессе обучения, как показано на рисунке 60.

Для нового обучения выполните в подсистеме последовательные шаги которые описаны в разделе «3.4 Работа с подсистемой».

## Классификатор корпоративной документации

Файлы могут быть загружены или переташены на параметры

Протокол или решение

+ Выбрать файл...

Начать загрузку

Произведено обучение модели классификатора RandomForest. Результаты обучения:

**Метрики для класса "Протокол"**  
Точность: 0,83  
Полнота: 1,00  
F-мера: 0,91

**Метрики для класса "Договор"**  
Точность: 1,00  
Полнота: 0,83  
F-мера: 0,91

Номер документа: 4  
Дата мероприятия: 20 . 08 . 1992 г.  
Название мероприятия: Решение  
Место продажи: Махачкала  
Вопросы занятости:

- Решение:** Освободить должность генерального директора ООО «ГРАМАТОР» - АХМЕДПАШАЕВА НАДЫРА ГАДЖИЕВИЧА.
- Решение:** Назначить новым генеральным директором ООО «ГРАМАТОН» - МУРТАЗАЛИЕБА ШАМИЛЯ МУРТАЗАЛИЕВИЧА.



Рисунок 60