

**Руководство пользователя.
Сервис обезличивания
неструктурированных документов
DeepDocs.**

| | | |
|----------|---|----------|
| 1 | Инструкция по использованию сервиса через API | 4 |
| 1.1 | Как обезличить PDF документ с помощью API Anonymizer? | 4 |
| 1.1.1 | Пример в командной строке | 4 |
| 1.1.2 | Пример на Python | 4 |
| 1.2 | Как обезличить текстовый документ с помощью API Anonymizer? | 4 |
| 1.2.1 | Пример в командной строке | 4 |
| 1.2.2 | Пример на Python | 5 |
| 1.3 | Как указать, какие сущности обезличивать? | 5 |
| 1.3.1 | Пример в командной строке | 6 |
| 1.3.2 | Пример на Python | 6 |
| 2 | Инструкция по использованию сервиса через Web-сервис | 7 |
| 2.1 | Создание проекта | 7 |
| 2.2 | Загрузка документов | 7 |
| 2.3 | Функционал поиска и сортировки документов | 8 |
| 2.4 | Создание задания на обезличивание | 8 |
| 2.5 | Верификация и просмотр результатов | 9 |
| 2.6 | Экспорт данных | 9 |
| 2.7 | Просмотр дашбордов | 10 |

Сервис обезличивания неструктурированных документов DeerDocs можно использовать через API или через WEB-интерфейс.

1 Инструкция по использованию сервиса через API.

1.1 Как обезличить PDF документ с помощью API Anonymizer?

Чтобы обезличить pdf документ (или любой документ с конвертацией в pdf), необходимо послать POST запрос в формате multipart/form-data на http://host:port/anonymize_pdf (host и port сервера, где развернут сервис)

Ответом на запрос будет обезличенный документ в битовом представлении.

1.1.1 Пример в командной строке

```
# Общий вид
curl -F file=@file_to_anonymize http://host:port/anonymize\_pdf -o
output_file.pdf
# или
curl -F file=@file_to_anonymize http://host:port/anonymize\_pdf >>
output_file.pdf
```

1.1.2 Пример на Python

```
# Общий вид запроса
import requests
files = {'file': ('doc.pdf', open('file_to_anonymize', 'rb'))}
r = requests.post("http://host:port/anonymize\_pdf", files=files)
r.status_code # 200
# Запись результата
with open('output_file', 'wb') as fout:
    fout.write(r.content)
```

1.2 Как обезличить текстовый документ с помощью API Anonymizer?

Чтобы обезличить текстовый документ (в данный момент поддерживаются форматы txt, html и docx), необходимо послать POST запрос в формате multipart/form-data на <http://host:port/anonymize> (host и port сервера, где развернут сервис)

Ответом на запрос будет обезличенный документ в битовом представлении.

1.2.1 Пример в командной строке

```
# Общий вид
curl -F file=@file_to_anonymize http://host:port/anonymize -o
output_filename
# или
curl -F file=@file_to_anonymize http://host:port/anonymize >>
output_filename
```

1.2.2 Пример на Python

```
# Общий вид запроса
import requests
files = {'file': ('doc.docx', open('file_to_anonymize', 'rb'))} r =
requests.post("http://host:port/anonymize", files=files) r.status_code
# 200
# Запись результата
with open('output_file', 'wb') as fout:
    fout.write(r.content)
```

Также можно добавить параметры обезличивания - `type` и `anonymize_entities`. Как добавить `anonymize_entities` и что это такое описано ниже.

`type` - это тип обезличивания и он может принимать два значения: `label` (замена константами) и `word` (замена синтетическими данными). Этот параметр передается аналогично `anonymize_entities`.

1.3 Как указать, какие сущности обезличивать?

В своем запросе можно также указать сущности, которые вы хотите обезличить. Ниже представлен список сущностей, которые возможно указать.

- "Должность"
- "Орган"
- "Время"
- "ГосОрган"
- "Дата"
- "НЮЛ-Наименование"
- "НЮЛ-ОПФ"
- "Ссылка"
- "Сумма"
- "Сумма-сумма_прописью"
- "Цифры-цифры_прописью"
- "ФизЛицо"
- "Цифры"
- "Адрес"
- "Почта"
- "Цифры-цифры_номером"

Массив с выбранными сущностями нужно также передать как параметр `anonymize_entities`. Ниже представлены примеры.

По умолчанию (если параметр не передан) обезличиваются такие сущности как Цифры, Адрес, ФизЛицо и Почта.

Если передать в параметре пустой массив, то обезличивание закончится с ошибкой.

1.3.1 Пример в командной строке

```
# Общий вид
curl -F file=@file_to_anonymize http://host:port/anonymize_pdf -F
'anonymize_entities=entities_list' -o output_file.pdf
```

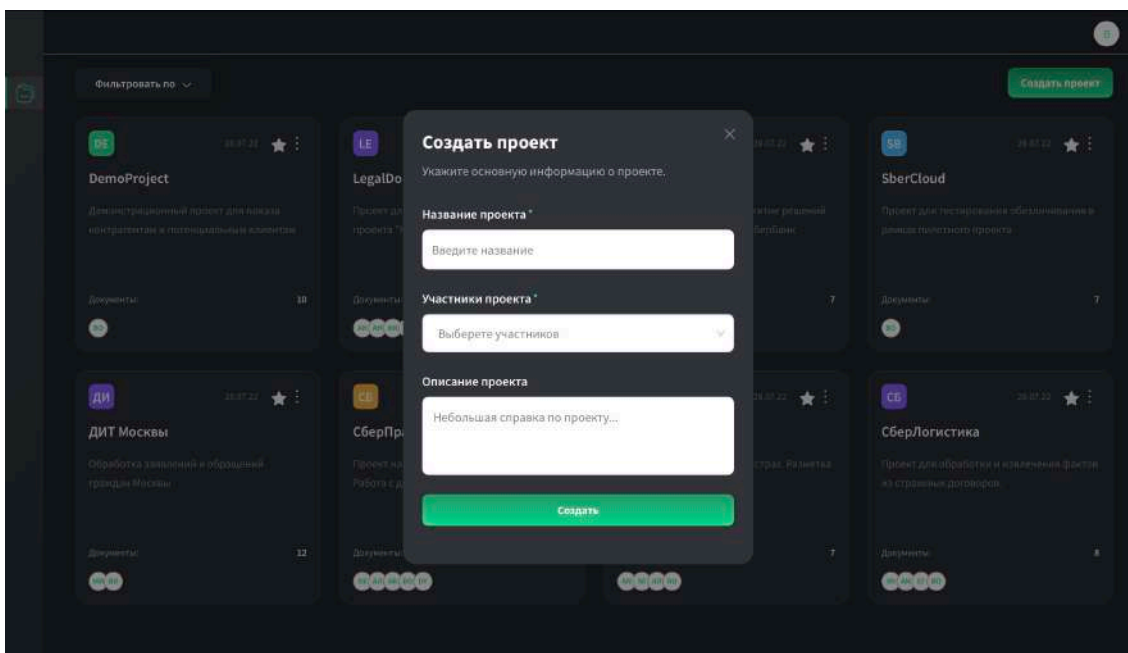
1.3.2 Пример на Python

```
# Общий вид запроса
import requests, json
files = {'file': ('doc.pdf', open('file_to_anonymize', 'rb'))}
entities_to_anonymize: List[str] = [...]
r = requests.post("http://host:port/anonymize_pdf", files=files,
                  data={'anonymize_entities':
                        json.dumps(entities_to_anonymize)})
r.status_code # 200
# Запись результата
with open('output_file', 'wb') as fout:
    fout.write(r.content)
```

2 Инструкция по использованию обезличивания через Web-сервис.

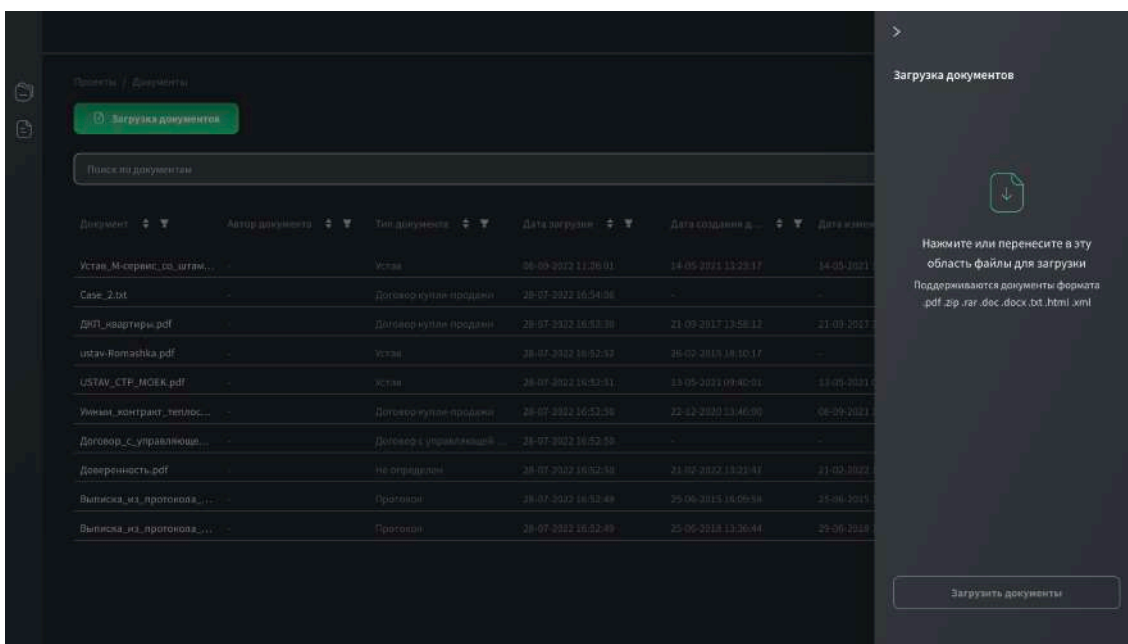
2.1 Создание проекта.

После авторизации/регистрации пользователя необходимо нажать на кнопку "Создать проект". Далее необходимо указать название проекта, предоставить доступ участникам проекта, а также оставить описание проекта.



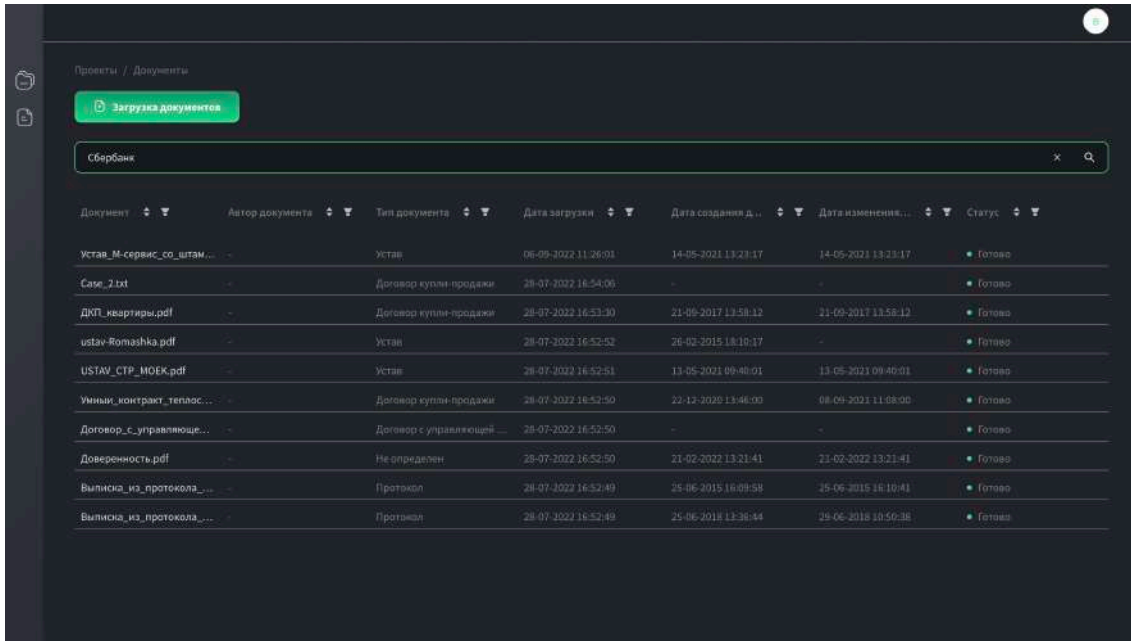
2.2 Загрузка документов

В созданном проекте необходимо нажать кнопку "Загрузка документов". Далее выбираете файлы любых форматов.



2.3 Функционал поиска и сортировки документов

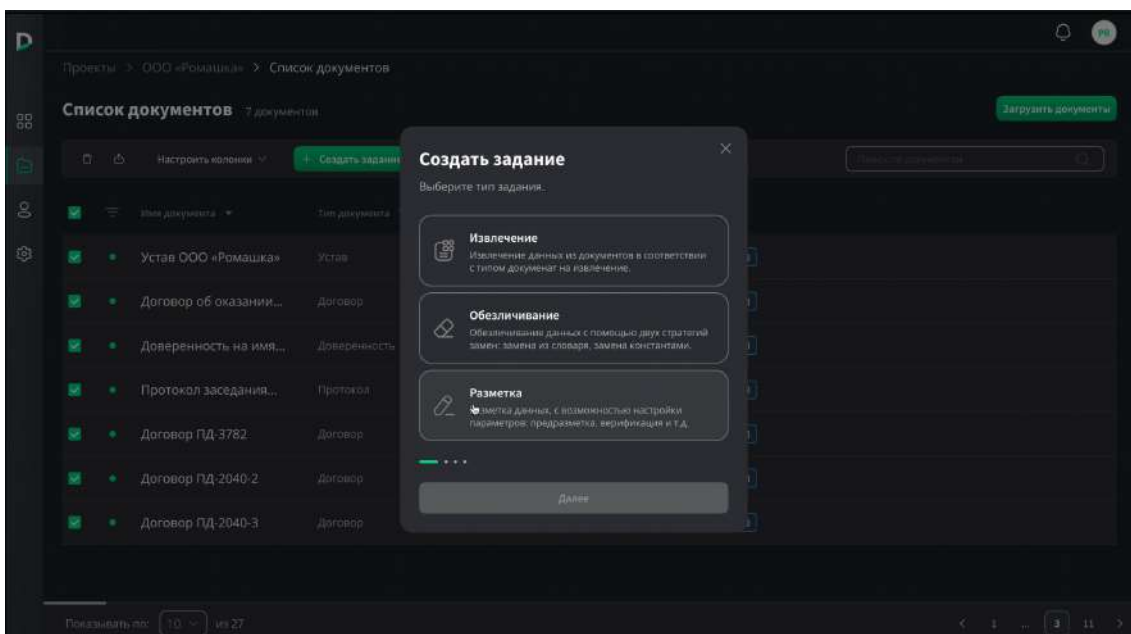
После загрузки будет доступен список документов. Для каждого документа доступны метаданные, а также результат классификации типа документов. Пользователь может отфильтровать файлы по различным параметрам. Также доступен полнотекстовый поиск по содержимому файлов.



| Документ | Автор документа | Тип документа | Дата загрузки | Дата создания д. . | Дата изменения... | Статус |
|---------------------------|-----------------|--------------------------|---------------------|---------------------|---------------------|--------|
| Устав_M-сервис_со_штам... | - | Устав | 06-09-2022 11:26:01 | 14-05-2021 13:23:17 | 14-05-2021 13:23:17 | Готово |
| Саме_2.txt | - | Договор купли-продажи | 28-07-2022 16:54:06 | - | - | Готово |
| ДКП_квартиры.pdf | - | Договор купли-продажи | 28-07-2022 16:53:30 | 21-09-2017 13:58:12 | 21-09-2017 13:58:12 | Готово |
| ustav_Romashka.pdf | - | Устав | 28-07-2022 16:52:52 | 26-02-2015 18:18:17 | - | Готово |
| USTAV_CTP_MOEK.pdf | - | Устав | 28-07-2022 16:52:51 | 13-05-2021 09:48:01 | 13-05-2021 09:48:01 | Готово |
| Учный_контракт_теплас... | - | Договор купли-продажи | 28-07-2022 16:52:50 | 22-12-2020 13:46:03 | 08-09-2021 11:08:08 | Готово |
| Договор_с_управляюще... | - | Договор с управляющей... | 28-07-2022 16:52:50 | - | - | Готово |
| Доверенность.pdf | - | Не определен | 28-07-2022 16:52:50 | 21-02-2022 13:21:41 | 21-02-2022 13:21:41 | Готово |
| Выписка_из_протокола_... | - | Протокол | 28-07-2022 16:52:49 | 25-06-2015 16:03:58 | 25-06-2015 16:10:41 | Готово |
| Выписка_из_протокола_... | - | Протокол | 28-07-2022 16:52:49 | 25-06-2018 13:38:44 | 29-06-2018 10:50:38 | Готово |

2.4 Создание задания на обезличивание.

Для выполнения задачи обезличивания с возможностью верификации пользователю необходимо выбрать документы, нажать на кнопку "Создать задание", выбрать "Обезличивание"



В задании "Обезличивание" нужно указать название, а также выбрать набор сущностей для обезличивания, исходя из бизнес-задачи.

Проекты > ООО «Ромашка» > Задания > Ромашка_Обезличивание 22

Ромашка_Обезличивание 22 Участники задания: PR, PB, AK, KK

7 документов · Обезличивание · Дата создания: 08.07.2022

Настроить колонки · Отправить на проверку Прогресс по заданию: 30%

| Имя документа | Тип документа | Дата создания | Автор | Проверяющие | Статус проверки | Тег |
|------------------------|---------------|---------------|--------------|-------------|-----------------|---------------|
| Устав ООО «Ромашка» | Устав | 11.01.2020 | Романенко П. | Боюка П. | Готово | Ромашка 22 Ф3 |
| Договор об оказани... | Договор | 11.01.2020 | Романенко П. | Боюка П. | На проверке | Ромашка 22 Ф3 |
| Доверенность на имя... | Доверенности | 11.01.2020 | Романенко П. | Боюка П. | На проверке | Ромашка 20 Ф3 |
| Протокол заседания... | Протокол | 11.01.2020 | Романенко П. | Боюка П. | На проверке | Ромашка 20 Ф3 |
| Договор ПД-3782 | Договор | 11.01.2020 | Романенко П. | Боюка П. | На проверке | Ромашка 22 Ф3 |
| Договор ПД-2040-2 | Договор | 11.01.2020 | Романенко П. | Боюка П. | На проверке | Ромашка 22 Ф3 |
| Договор ПД-2040-3 | Договор | 11.01.2020 | Романенко П. | Боюка П. | На проверке | Ромашка 22 Ф3 |

Показывать по: 10 из 27 < 1 ... 3 11 >

2.7 Просмотр дашбордов

Для пользователя доступна страница-дашборд, где отображается статистика по проекту и заданиям в нем.

