



119270, Москва, Лужнецкая наб., д. 6,  
стр.1, офис 214, ООО «ЭР СИ О»  
Тел: (495) 287-98-87  
E-mail: [info@rco.ru](mailto:info@rco.ru)  
<http://www.rco.ru>

# **Руководство администратора RCO Categorization Engine – библиотека для категоризации текстов**

(Версия для Astra Linux)

Москва, 2021

В содержание данного документа могут быть внесены изменения без предварительного уведомления. Названия организаций, имена и даты, используемые в качестве примеров, являются вымышленными, если не оговорено обратное.

© ООО «ЭР СИ О», 2021. Все права защищены.

ЭР СИ О, Russian Context Optimizer, RCO являются охраняемыми товарными знаками.

ООО «ЭР СИ О» может являться правообладателем патентов и заявок, поданных на получение патента, товарных знаков и объектов авторского права, которые имеют отношение к содержанию данного документа.

Предоставление вам данного документа не означает передачи какой-либо лицензии на использование данных патентов, товарных знаков и объектов авторского права, за исключением использования, явно оговоренного в лицензионном соглашении ООО «ЭР СИ О».

Все другие названия юридических лиц и изделий являются охраняемыми товарными знаками или товарными знаками, принадлежащими их владельцам.

## Содержание

Общие сведения .....	4
Назначение .....	4
Требования к квалификации.....	4
Системные требования .....	4
Подготовка к работе.....	5
Состав дистрибутива .....	5
Установка ПО .....	5
Проверка работоспособности .....	5
Дополнительные настройки.....	5
Удаление ПО .....	5
Регламентное обслуживание.....	7
Устранение неисправностей в ходе эксплуатации.....	7
Обновление ПО .....	7

# Общие сведения

---

## Назначение

Библиотека предназначена для категоризации документов по предварительно подготовленным профилям категорий.

Данный дистрибутив предназначен для работы с ОС Astra Linux (Common Edition) версии 2.12 и новее.

## Требования к квалификации

Администратор ПО должен иметь следующие навыки:

- Администрирования информационных систем;
- Технического обслуживания средств вычислительной техники, на которых устанавливается ПО;
- Работы с операционными системами Linux.

## Системные требования

Требования к характеристикам вычислительной техники зависят от количества и объёма обрабатываемых документов.

Минимальные требования: 1 ядро процессора частотой 2 ГГц, 4 Гб оперативной памяти (из которых 300 Мб непосредственно на запуск одного экземпляра), 300 Мб на жёстком диске.

Рекомендуемые требования, исходя из потока 50 тыс. документов в час общим объёмом 100 Мб: 2 ядра процессора частотой 3 ГГц (2 обработчика - по 1 ядру на каждый), 5 Гб оперативной памяти (по 500 Мб на каждый обработчик), 1 Гб на жёстком диске. При увеличении объёма потока документов необходимо пропорционально увеличить производительность ядра процессора или количество обработчиков, скорректировав соответствующим образом требования к памяти и количеству ядер процессора.

# Подготовка к работе

---

## Состав дистрибутива

Дистрибутив содержит пакеты с библиотеками и тестовыми приложениями, а также архив с профилями категорий, морфологическим словарём и тестовыми файлами.

- **gpcat\_4.0.0-1\_amd64.deb** пакет с основной библиотекой категоризатора, пакет для разработчика **gpcat-devel\_4.0.0-1\_amd64.deb** содержит заголовочные файлы библиотеки
- **GpcatTestStand.tar.bz2** – морфологический словарь, профили, тестовые примеры и скрипты.

## Установка ПО

- Перейти в каталог с пакетами Linux/pkgsg/:
  - Добавить каталог для устанавливаемых библиотек в путь для системной Linux утилиты ldconfig:

```
sudo cp ./x86_64-rh-compat.conf /etc/ld.so.conf.d/
```
  - Установить пакет с библиотекой и тестовым приложением:

```
sudo dpkg -i ./gpcat_4.0.0-1_amd64.deb
```
  - Если предполагается разработка приложения, вызывающего библиотеку, установить пакет разработчика:

```
sudo dpkg -i ./gpcat-devel_4.0.0-1_amd64.deb
```
- Перейти в каталог с данными Linux/data/:
  - Разархивировать каталог с данными:

```
tar jxvf ./GpcatTestStand.tar.bz2
```

## Проверка работоспособности

Исполнить командный файл GpcatTestStand/\_gpcat\_test.sh. В результате приложение gparitest загрузит профили из каталога ./Profiles/ и определит категории документов из ./imp/.

Результаты категоризации в html формате складываются в каталог ./out/.

## Дополнительные настройки

Путь к обрабатываемым файлам и к файлу с результатами анализа можно менять, редактируя командный файл «\_gpcat\_test.sh».

## Удаление ПО

Для удаления программы необходимо удалить используемые пакеты из системы:

```
sudo dpkg -r gpcat-devel
sudo dpkg -r gpcat
```

И удалить каталог «GpсatTestStand» с распакованными профилями и морфословарями.

# Регламентное обслуживание

---

## Устранение неисправностей в ходе эксплуатации

Отчёт об ошибках, возникающих при работе программы, печатается в консоль.

При разработке приложений, использующих библиотеку, текст ошибок может быть получен вызовом функции `CAQueryInfo` с параметрами `CA_Q_ERR_LOG_STR`, подробности в документации разработчика.

## Обновление ПО

Для обновления нужно остановить работу программы, перезаписать обновляемые файлы и снова запустить программу.