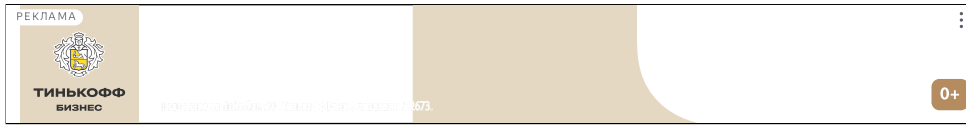


/ [Переводы документации](#) / [документация Scrapy](#) / Документация по Scrapy на русском языке



Документация по Scrapy на русском языке ¶

Scrapy – это быстрый высокоуровневый фреймворк сканирование сети и парсинга веб-страниц, используемый для сканирования веб-сайтов и извлечения структурированных данных с их страниц. Его можно использовать для самых разных целей, от интеллектуального анализа данных до мониторинга и автоматического тестирования.

Получение помощи ¶

Возникли проблемы? Хотим помочь!

- Прочитайте [ЧЗВ](#) – у него есть ответы на некоторые общие вопросы.
- Ищете конкретную информацию? Просмотрите [Алфавитный указатель](#) или [Состав модуля](#).
- Задайте или ищите вопросы на StackOverflow с scrapy тэгом.
- Задайте или ищите вопросы в Scrapy subreddit.
- Ищите вопросы по архивам в списке рассылки scrapy-users.
- Задайте вопрос в #scrapy IRC канале,
- Сообщайте об ошибках с Scrapy в issue tracker'е проекта.

Первые шаги ¶

[Краткий обзор Scrapy](#)

Узнайте, что такое Scrapy и чем он может вам помочь.

[Инструкция по установке](#)

Устанавливает Scrapy на свой компьютер.

[Scrapy Учебник](#)

Напишите свой первый проект Scrapy.

[Примеры](#)

Узнайте больше, играя с готовым проектом Scrapy.

Базовые концепты ¶

[Инструмент командной строки](#)

Узнайте об инструменте командной строки, который используется для управления вашим проектом Scrapy.

Пауки

Напишите правила сканирования ваших сайтов.

Селекторы

Извлекайте данные с веб-страниц с помощью XPath.

Оболочка Scrapy

Протестируйте свой код извлечения в интерактивной среде.

Элементы

Определяет данные, которые вы хотите сканировать.

Загрузчики элементов

Заполнить свои элементы извлечёнными данными.

Элемент конвейера

Обработать и сохранить сканированные данные.

Экспорт фидов

Вывести сканированные данные, используя разные форматы и хранилища.

Запросы и ответы

Разберитесь в классах, используемых для представления HTTP-запросов и ответов.

Извлечение ссылок

Удобные классы для извлечения ссылок со страниц.

Настройки

Узнайте, как настроить Scrapy и увидеть все [доступные настройки](#).

Исключения

Посмотрите все доступные исключения и их значение.

Встроенные сервисы ¶

Логирование

Узнайте, как использовать встроенное ведение журнала Python на Scrapy.

Сбор статистики

Соберите статистику о вашем парсере.

Отправка электронной почты

Отправлять уведомления по электронной почте при наступлении определенных событий.

Консоль Telnet

Проверяйте работающий поисковый робот с помощью встроенной консоли Python.

Веб-сервис

Отслеживайте и управляйте поисковым роботом с помощью веб-службы.

Решение конкретных проблем¶

Часто задаваемые вопросы

Получить ответы на наиболее часто задаваемые вопросы.

Отладка пауков

Узнайте, как отлаживать типичные проблемы вашего паука Scrapy.

Контракты пауков

Узнайте, как использовать контракты для тестирования ваших пауков.

Общие практики

Ознакомьтесь с некоторыми распространенными практиками Scrapy.

Широкие обходы

Настраивает Scrapy для параллельного сканирования большого количества доменов.

Использование инструментов разработчика вашего браузера для парсинга

Узнайте, как очистить парсинг с помощью инструментов разработчика вашего браузера.

Выбор динамически загружаемого контента

Чтение данных веб-страницы, которые загружаются динамически.

Устранение утечек памяти

Узнайте, как найти и избавиться от утечек памяти в вашем сканере.

Скачивание и обработка файлов и изображений

Загрузить файлы и/или изображения, связанные с вашими сканированными элементами.

Развёртывание пауков

Развертывание ваших пауков Scrapy и запуск их на удалённом сервере.

Расширение AutoThrottle

Регулируйте скорость сканирования динамически в зависимости от нагрузки.

Бенчмаркинг

Проверить, как Scrapy работает на вашем оборудовании.

Задания: приостановка и возобновление сканирования

Узнайте, как приостанавливать и возобновлять сканирование больших пауков.

Корутины

Используйте [синтаксис корутин](#).

[asyncio](#)

Использование библиотеки на базе [asyncio](#) и [asyncio](#).

Расширение Scrapy

Обзор архитектуры

Изучите архитектуру Scrapy.

Промежуточное ПО загрузчика

Настраивает способ запроса и загрузки страниц.

Промежуточное ПО Spider

Настраивает вход и выход ваших пауков.

Расширения

Расширьте Scrapy своими собственными функциями.

Базовый API

Используйте его в расширениях и промежуточном программном обеспечении для расширения функциональности Scrapy

Сигналы

Посмотреть все доступные сигналы и как с ними работать.

Планировщик

Изучите компонент планировщика.

Экспортёры элементов

Быстро экспортируйте скопированные элементы в файл (XML, CSV и т. д.).

Все остальные

Примечания к выпуску

Посмотрите, что изменилось в последних версиях Scrapy.

Внесение вклада в Scrapy

Узнайте, как внести свой вклад в проект Scrapy.

Управление версиями и стабильность API

Узнайте о версиях Scrapy и стабильности API.

| [следующий](#) »

Нашли ошибку? Выделите текст и нажмите **Shift+Enter**.

© Авторские права Digitology.tech

Последнее обновление: авг. 10, 2023