Добрый день, коллеги! Меня зовут Егор Корякин, а моего товарища — Андрей Толкунов. Мы студенты второго курса Московской академии TOP. Сегодня мы с удовольствием представим вам наш проект — парсер для Яндекс Карт.

### 

### **Цель проекта**

Цель проекта: Создание эффективного инструмента для автоматизированного сбора актуальной информации (названия компаний, адреса, телефоны, веб-сайты и страницы ВКонтакте) с целью оптимизации аналитики, маркетинговых исследований и работы с большими объемами данных.

### **Принцип работы**

Программа выполняет автоматический сбор данных с Яндекс.Карт, прокручивает страницу, обрабатывает дубликаты и сохраняет уникальные результаты. Логирование и проверка на ошибки сопровождают процесс.

### **Технологии в проекте**

Основные компоненты:

1. Selenium WebDriver — взаимодействие с браузером.
2. WebDriverWait и ExpectedConditions — ожидание элементов.
3. ActionChains — имитация кликов.
4. Pandas — сохранение данных в CSV.
5. Logging — логирование и отслеживание процесса.

### **Обработка данных**

Сбор и обработка данных: извлечение информации, исключение дубликатов и обработка ошибок.

### 

### 

### 

### 

### **Сохранение результатов**

Собранные данные сохраняются в текстовом и CSV форматах для удобства анализа и проверки. Используется библиотека Pandas для удобной обработки и сохранения.

### 

### **Скроллинг**

Программа автоматически прокручивает страницу, чтобы загрузить больше результатов, что увеличивает объем собираемых данных.

### **Извлечение информации**

Сбор информации с карточек бизнеса включает названия компаний, адреса, телефоны, веб-сайты и страницы ВКонтакте. Для исключения повторных записей используется множество seen\_entries, которое хранит уникальные подписи карточек.

### **Выбор веб-браузера**

Используется браузер Firefox с помощью Selenium WebDriver для стабильности и поддержки множества функций.

### **Логирование**

Процесс сопровождается логированием на каждом этапе. Логи фиксируют этапы выполнения, такие как инициализация браузера, обработка карточек, сохранение данных. Логирование помогает отслеживать ошибки и этапы работы приложения.

### **Уникальность проекта**

Уникальность проекта заключается в гибкости, масштабе и удобстве использования для анализа данных, необходимых для бизнеса и маркетинга.

### **Будущее развитие**

Планируется улучшение функционала, включая:

1. Добавление фильтров для отбора данных.
2. Интеграцию с другими платформами для анализа данных.
3. Создание интерфейса для более широкой аудитории.

### 

### 

### 

### **Обработка ошибок**

Программа обрабатывает ошибки Selenium, сети и другие возможные сбои. Ошибки логируются, а при критических проблемах работа корректно завершается.

### 

### **Автоматизация задач**

Процесс автоматизирован: запуск браузера, ввод запросов, прокрутка страниц, извлечение и сохранение данных, обработка ошибок.

### 

### **Тестирование и настройка**

Результаты тестирования:

1. Корректная отправка и обработка данных через POST-запросы.
2. Обработка ошибок Selenium и сети.
3. Программа завершает работу без критических ошибок.

### **Масштабируемость проекта**

Проект способен масштабироваться, охватывая большие объемы данных и длительные сессии, а также адаптироваться для интеграции с новыми функциональными возможностями.

### **Заключение**

Проект продемонстрировал свою эффективность и готовность к использованию в реальных задачах. Мы планируем его дальнейшее развитие и совершенствование для решения более сложных задач в области аналитики и маркетинга.