

Тестовое задание

## Задача

Вам предстоит собрать информацию с сайта [https://www.petsonic.com](https://www.petsonic.com/snacks-huesos-para-perros/)

Нас интересует та часть сайта, который занимается продажей товаров для животных - <https://www.petsonic.com/snacks-huesos-para-perros/>

Нужно собрать **все товары** из категории и записать в **csv** файл.

По каждому товару нужно собрать три параметра

1. Полное название товара
2. Цену
3. Изображение

**Важно!** На случай проблем с доступностью категории, воспользуйтесь альтернативной ссылкой: <https://www.petsonic.com/hobbit-half/>

### Пример разбора

Вот пример страницы товара <https://www.petsonic.com/hobbit-alf-galletas-granja-para-perros.html>



Такую страницу мы называем страницей мультипродукта. Это означает, что на одной странице находится информация о нескольких разновидностях продукта. В данном случае разная весовка: 200 gr, 430 gr, 4 kg.

С этой страницы в выходной файл должно быть добавлено 3 записи:

|  |  |  |
| --- | --- | --- |
| **Name** | **Price** | **Image** |
| Galletas Granja para Perro - 200 gr | 1.35 | http://www.petsonic.com/5830-large\_default/galletas-granja-para-perro.jpg |
| Galletas Granja para Perro - 430 gr | 2.34 | *http://...* |
| Galletas Granja para Perro - 4 kg | 16.83 | *http://...* |

### На что обратить внимание

1. Название сформировано, как общее название продукта *Galletas Granja para Perro* плюс весовка из таблицы *200gr*
2. Если у продуктов нет вариаций, то и весовка будет оказана только одна. В таком случае с одной страницы продукта получится одна запись в файле.
3. Список продуктов категории состоит из нескольких страниц

### Требования к оформлению кода

1. Должна быть написана программа/скрипт на Ruby. Программа получает на вход:
   1. ссылка на страницу категории (может передаваться любая категория сайта)
   2. имя файла в который будет записан результат
2. После отработки скрипта результаты записывает в выходной файл с заданными именем.
3. В тестовом задании нужно использовать **xpath** для получения содержимого html элементов, таких как цена, название и т.п. Возможно, понадобится использовать regexp для того чтобы достать отдельные части данных, если xpath недостаточно.
4. Для скачивания страниц нужно использовать библиотеку **curb** *(могут быть проблемы под Windows, лучше использовать Linux)*, для парсинга страниц **nokogiri**, для записи в csv файл – модуль **csv**.
5. Вместе с решением нужно прислать пример запуска скрипта.
6. Если в скрипте используются сторонние gem-ы, то приложить Gemfile.
7. Скрипт должен работать достаточно быстро и при этом не привлекать к себе внимание администраторов сайта.
8. Во время выполнения скрипта должно быть понятно, что он сейчас делает

## Теоретические вопросы

### Блок Q1

1. Какие сторонние библиотеки Вы используете чаще всего для разработки. Какие плюсы в них Вы выделяете для себя? (Ruby или тот язык на котором пишите)
2. C какими OС Вы работали? Работали ли с консолью? Какие консольные утилиты Linux Вы знаете и использовали?
3. Как Вы повышаете свою профессиональную квалификацию (названия книг, курсов, сайтов, скринкастов и т.д.)?

### Блок Q2

1. Какие инструменты для профайлинга и дебага Вы используете? Какие у них минусы?
2. Объясните почему происходит следующее:
   1. 1660 / 100 ≠ 16.6
   2. 24.0 \* 0.1 ≠ 2.4
3. С какими базами данных вы работали? Проектировали ли вы свою собственную БД? Сколько было в ней таблиц? С какой самой большой таблицей по количеству записей Вы работали?
4. Оптимизировали ли Вы запросы в SQL? Как Вы это делали? Как Вы понимаете что запрос выполняется оптимально?

### Блок Q3

Вы работаете разработчиком в компании, которая является крупной сетью магазинов, а также онлайн ритейлером, одним из крупнейших в своем регионе. В вашем онлайн магазине имеется порядка 100 тыс различных продуктов.  
На рынке кроме вас присутствуют несколько конкурирующих компаний, одна из которых (“Брусничка”) растет достаточно быстрыми темпами и уже фактически дышит вам в спину. Свою маркетинговую стратегию данная компания строит на демпинге цен, ее лозунг - “Дешевле чем везде”. Также вы знаете, что эта компания использую различные технические средства отслеживает цены своих конкурентов, чтобы предложить меньшую цену и привлечь покупателей к себе.  
Как разработчику Вам поставлена задача любой ценой блокировать доступ с целью сбора данных по вашим продуктам, но при этом сайт должен функционировать и ни в коем образе **не отталкивать текущих и новых покупателей**.   
Ваши действия?

## 

## Как все отправить

Перед тем, как отправить ваши результаты, проверьте по следующему чеклисту:

1. Ваше резюме
2. Ссылка на репозиторий с решением задачи (либо файл с кодом)
3. Команду на запуск скрипта
4. Скрин тестового запуска
5. Подробные ответы на все вопросы
6. Заархивируйте всё это в rar или zip архив - не все почтовые клиенты пропускают скрипты

До встречи на собеседовании!(но это не точно)