

Практическая работа № 3

Задание 1

Исследовать структуру html-файлов, чтобы произвести парсинг всех данных. В каждом файле содержится информация об одном объекте из случайной предметной области. Полученные данные собрать и записать в json. Выполните также ряд операций с данными:

- отсортируйте значения по одному из доступных полей
- выполните фильтрацию по другому полю (запишите результат отдельно)
- для одного выбранного числового поля посчитайте статистические характеристики (сумма, мин/макс, среднее и т.д.)
- для одного текстового поля посчитайте частоту меток

Задание 2

Исследовать структуру html-файлов, чтобы произвести парсинг всех данных. В каждом файле содержится информация об одном или нескольких объектах из случайной предметной области. Перечень всех характеристик объекта может меняться (у отдельного объекта могут отсутствовать некоторые характеристики). Полученные данные собрать и записать в json. Выполните также ряд операций с данными:

- отсортируйте значения по одному из доступных полей
- выполните фильтрацию по другому полю (запишите результат отдельно)
- для одного выбранного числового поля посчитайте статистические характеристики (сумма, мин/макс, среднее и т.д.)
- для одного текстового поля посчитайте частоту меток

Задание 3

Исследовать структуру xml-файлов, чтобы произвести парсинг всех данных. В каждом файле содержится информация об одном объекте из случайной предметной области. Полученные данные собрать и записать в json. Выполните также ряд операций с данными:

- отсортируйте значения по одному из доступных полей
- выполните фильтрацию по другому полю (запишите результат отдельно)
- для одного выбранного числового поля посчитайте статистические характеристики (сумма, мин/макс, среднее и т.д.)
- для одного текстового поля посчитайте частоту меток

Задание 4

Исследовать структуру xml-файлов, чтобы произвести парсинг всех данных. В каждом файле содержится информация об одном или нескольких объектах из случайной предметной области. Перечень всех характеристик объекта может меняться (у отдельного объекта могут отсутствовать некоторые характеристики). Полученные данные собрать и записать в json. Выполните также ряд операций с данными:

- отсортируйте значения по одному из доступных полей
- выполните фильтрацию по другому полю (запишите результат отдельно)
- для одного выбранного числового поля посчитайте статистические характеристики (сумма, мин/макс, среднее и т.д.)
- для одного текстового поля посчитайте частоту меток

Задание 5

Самостоятельно найти сайт, соответствующий следующим условиям:

- непопулярный, регионального уровня или из узкой области (с целью избежать дублирования)
- наличие страниц-каталогов, где есть информация сразу по нескольким объектам
- наличие страниц, посвященных отдельному объекту

Необходимо:

- спарсить нескольких страниц (минимум 10), посвященных только одному объекту;
- спарсить страницы-каталоги, где размещена информация сразу по нескольким объектам.

Данные можно скачать и сохранить локально в виде html, а можно организовать их получение напрямую через обращение к серверу сайта.

Результаты парсинга собрать отдельно по каждой подзадаче и записать в отдельный json.

Выполните произвольные операции с данными:

- отсортируйте значения по одному из доступных полей
- выполните фильтрацию по другому полю (запишите результат отдельно)
- для одного выбранного числового поля посчитайте показатели статистики
- для одного текстового поля посчитайте частоту меток.