

Лабораторная работа № 7

Извлечение данных из веб-страниц.

Необходимо изучить материал и выполнить следующее:

- Найти на сайте Википедия минимум 20 страниц по информатике и программированию;
- Предусмотреть функцию сохранения базовых данных в текстовый файл (txt): заголовок и первый абзац статьи. Разрешается сохранять все данные в один файл.
- Предусмотреть функцию по парсингу списка внешних ссылок (External links) и сохранить их в виде CSV-файла с двумя столбцами: название и URL. Каждая страничка – отдельный CSV-файл, имя файла должно соответствовать страничке.
- Предусмотреть функцию вывода данных из таблицы "Инфокарты" (infobox) в консоль. Желательно использовать табличный способ вывода. Для этих целей можно применить табуляцию или специальные библиотеки, например, tabulate.
- Провести анализ данных, включая:
 - количество параграфов на странице,
 - количество ссылок,
- Построить график, показывающий количество параграфов и количество ссылок для каждой статьи. При значительном разбросе количества ссылок и количества параграфов рекомендуется построить два отдельных графика.