

Задание к лекции 11. «Парсинг сайта»

Получение данных с сайта LENTA.RU

Произвести парсинг раздела «Главные новости» на главной странице сайта.

На основе парсинга должен быть сформирован массив данных следующей структуры:

1. Дата публикации статьи в формате YY-MM-DD-HH-mm, где YY – год, MM – месяц, DD – день, HH-час, mm – минуты.
2. Название статьи.
3. URL статьи.
4. Тег (название рубрики).
5. Количество слов в статье.
6. Ссылки на другие источники в статье в виде URL (если есть). Ссылки перечислить через запятую.

Полученные данные сохранить в ASCII-файле с названием **lenta.csv**. По каждой статье в файл должна быть записана отдельная строка данных. В каждой строке один элемент структуры от другого должен отделяться символом «точка с запятой». После последнего элемента структуры точка с запятой не ставится.