МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РФ
Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Пензенский государственный университет» (ПГУ)

А. Г. Петренко, С. Н. Катков

Парсинг веб-сайтов с Google Таблицами

Учебное пособие

Пенза
Издательство ПГУ
2026

УДК 004(075.8)
ББК 65ся7
П30

Р е ц е н з е н т ы :
кандидат экономических наук,
заведующий кафедрой «Информационные системы и цифровые технологии»
Пензенского казачьего института технологий (филиал)
Московского государственного университета технологий и управления
имени К. Г. Разумовского (ПКУ)»
А. А. Тусков;

кандидат технических наук, доцент кафедры
«Вычислительная техника»
Пензенского государственного университета
Н. В. Слепцов

Петренко, Александр Григорьевич.
П30 Парсинг веб-сайтов с Google Таблицами : учеб. пособие / А. Г. Петренко, С. Н. Катков. – Пенза : Изд-во ПГУ, 2026. – 56 с.

ISBN 978-5-908090-24-7

Описывается процесс парсинга с использованием языка запросов XPath и его синтаксиса. Приводятся правила составления запросов для выборки данных из HTML-документов. Подробно разбирается процесс парсинга с использованием соответствующих функций Google Sheets. Рассматриваются примеры для работы с сайтами Интернета.

Издание подготовлено на кафедре «Цифровая экономика» ПГУ и предназначено для обучающихся по направлению подготовки 38.03.05 «Бизнес-информатика», а также 38.03.01 «Экономика», 38.03.04 «Государственное и муниципальное управление», 41.03.05 «Международные отношения» и другим смежным направлениям.

УДК 004(075.8)
ББК 65ся7

ISBN 978-5-908090-24-7

© Пензенский государственный университет, 2026

СОДЕРЖАНИЕ

1. ПАРСИНГ ВЕБ-САЙТОВ И XPATH 4
2. ЯЗЫКИ РАЗМЕТКИ ДОКУМЕНТОВ XML И HTML 6
3. СТРУКТУРА ДОКУМЕНТА XML/HTML 8
3.1. Дерево документа 8
3.2. Отношения между узлами 10
4. ЯЗЫК ЗАПРОСОВ XPATH 13
5. ВЫБОР ЭЛЕМЕНТОВ С ИСПОЛЬЗОВАНИЕМ XPATH 14
6. СОЗДАНИЕ И ПРОВЕРКА СЕЛЕКТОРОВ
XPATH ДЛЯ ПАРСИНГА 28
6.1. Инструменты для исследования структуры веб-сайта 28
6.2. Проверка XPath-выражений 30
7. ПАРСИНГ С ИСПОЛЬЗОВАНИЕМ ТАБЛИЦ GOOGLE SHEETS 38
7.1. Функция IMPORTHTML 40
7.2. Функция IMPORTXML 42
СПИСОК ЛИТЕРАТУРЫ 54

Учебное издание

Петренко Александр Григорьевич,
Катков Сергей Николаевич

Парсинг веб-сайтов с Google Таблицами

Редактор *В. В. Чувашова*
Технический редактор *М. А. Борисов*
Компьютерная верстка *М. А. Борисова*
Дизайн обложки *О. И. Жаровой*

Подписано в печать 02.11.2026.
Формат 60×84 1/16. Усл. печ. л. 3,26
Тираж 38. Заказ № 512.

Издательство ПГУ
440026, г. Пенза, ул. Красная, 40
Тел.: (8412) 66-60-49, 66-67-77; e-mail: iic@pnzgu.ru