Модуль BeautifulSoup

|  |  |
| --- | --- |
| Import bs4 | Парсер HTML кода |
| soup = bs4.BeautifulSoup(contents, 'lxml') | Создание обекта соуп. Контент это текстовый файл либо импортированная страница. Второй аргумент обязательный |
| print(soup.h2)      print(soup.head)      print(soup.li) | Вывод указанных тегов. Первых попавшихся |
| recursiveChildGenerator()  for child in soup.recursiveChildGenerator():            if child.name:                print(child.name) | Метод recursiveChildGenerator() позволяет перебрать содержимое HTML-документа. |
| print(soup.prettify()) | При помощи метода prettify() можно добиться того, чтобы HTML-код выглядел аккуратнее. |
| **find()**  print(soup.find("ul", id="mylist")) | При помощи метода find() можно найти элементы страницы, используя различные опорные параметры, id в том числе. |
| find\_all() for links in bs4.BeautifulSoup.find\_all(soup, "a"):      print(links["href"])  for tag in soup.find\_all("li"):          print("{0}: {1}".format(tag.name, tag.text)) | При помощи метода find\_all() можно найти все элементы, которые соответствуют заданным критериям. |
| Select() |  |
|  |  |
|  |  |
| with open("index.html", "r") as f:        contents = f.read()        soup = BeautifulSoup(contents, 'lxml')        print("HTML: {0}, name: {1}, text: {2}".format(soup.h2,          soup.h2.name, soup.h2.text)) |  |
| if links.get("href"): | Проверка наличия реквизита href в теге а |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **CSS селектор** | **Пример** | **Описание** | **CSS** |
| [.class](https://puzzleweb.ru/css/sel_class.php) | .myClass | Выбирает все элементы с классом myClass (class="myClass"). | 1 |
| [#id](https://puzzleweb.ru/css/sel_id.php) | #main | Выбирает элемент с идентификатором main (id="main"). | 1 |
| [\*](https://puzzleweb.ru/css/sel_all.php) | \* | Выбор всех элементов. | 2 |
| [элемент](https://puzzleweb.ru/css/sel_element.php) | span | Выбор всех элементов <span>. | 1 |
| [элемент,элемент](https://puzzleweb.ru/css/sel_element_comma.php) | div,span | Выбор всех элементов <div> и всех элементов <span>. | 1 |
| [[атрибут]](https://puzzleweb.ru/css/sel_attribute.php) | [title] | Выбирает все элементы с атрибутом title. | 2 |
| [[атрибут="значение"]](https://puzzleweb.ru/css/sel_attribute_value.php) | [title="cost"] | Выбирает все элементы с атрибутом title, значение которого в точности совпадает со значением указанным в селекторе (title="cost"). | 2 |
| [[атрибут~="значение"]](https://puzzleweb.ru/css/sel_attribute_value_contains.php) | [title~="один"] | Выбирает все элементы с атрибутом title, в значении которого (в любом месте) встречается подстрока (в виде отдельного слова) "один" (title="один и два"). | 2 |
| [[атрибут|="значение"]](https://puzzleweb.ru/css/sel_attribute_value_lang.php) | [lang|="ru"] | Выбор всех элементов с атрибутом lang, значение которого начинается с "ru". | 2 |
| [[атрибут^="значение"]](https://puzzleweb.ru/css/sel_attr_begin.php) | a[href^="https"] | Выбор каждого элемента <a> с атрибутом href, значение которого начинается с "https". | 3 |
| [[атрибут$="значение"]](https://puzzleweb.ru/css/sel_attr_end.php) | [src$=".png"] | Выбирает все элементы с атрибутом src, значение которого оканчивается на ".png" (src="some\_img.png"). | 3 |
| [[атрибут\*="значение"]](https://puzzleweb.ru/css/sel_attr_contain.php) | [title\*="один"] | Выбирает все элементы с атрибутом title, в значении которого (в любом месте) встречается подстрока (в виде отдельного слова или его части) "один" (title="один и два"). | 3 |
| **Комбинатор** | **Пример** | **Описание** | **CSS** |
| [элемент элемент](https://puzzleweb.ru/css/sel_element_element.php) | div span | Выбор всех элементов <span> внутри <div>. | 1 |
| [элемент>элемент](https://puzzleweb.ru/css/sel_element_gt.php) | div>span | Выбирает все дочерние элементы <span>, у которых родитель - элемент <div>. | 2 |
| [элемент+элемент](https://puzzleweb.ru/css/sel_element_pluss.php) | div+p | Выбирает все элементы <p>, которые расположены сразу после элементов <div>. | 2 |
| [элемент1~элемент2](https://puzzleweb.ru/css/sel_gen_sibling.php) | p~ol | Выбор всех элементов <ol>, которым предшествует элемент <p>. | 3 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Псевдо-класс** | **Пример** | **Описание** | **CSS** |
| [:link](https://puzzleweb.ru/css/sel_link.php) | a:link | Выбор всех не посещенных ссылок. | 1 |
| [:visited](https://puzzleweb.ru/css/sel_visited.php) | a:visited | Выбор всех посещенных ссылок. | 1 |
| [:active](https://puzzleweb.ru/css/sel_active.php) | a:active | Выбор активной ссылки. | 1 |
| [:hover](https://puzzleweb.ru/css/sel_hover.php) | a:hover | Выбор ссылки при наведении курсора мышки. | 1 |
| [:focus](https://puzzleweb.ru/css/sel_focus.php) | input:focus | Выбор элемента <input>, который находится в фокусе. | 2 |
| [:first-child](https://puzzleweb.ru/css/sel_firstchild.php) | p:first-child | Выбор каждого элемента <p>, который является первым дочерним элементом своего родителя. | 2 |
| [:lang(язык)](https://puzzleweb.ru/css/sel_lang.php) | p:lang(ru) | Выбор каждого элемента <p> с атрибутом lang, значение которого начинается с "ru". | 2 |
| [:first-of-type](https://puzzleweb.ru/css/sel_first-of-type.php) | p:first-of-type | Выбор каждого элемента <p>, который является первым из элементов <p> своего родительского элемента. | 3 |
| [:last-of-type](https://puzzleweb.ru/css/sel_last-of-type.php) | p:last-of-type | Выбор каждого элемента <p>, который является последним из элементов <p> своего родительского элемента. | 3 |
| [:only-of-type](https://puzzleweb.ru/css/sel_only-of-type.php) | p:only-of-type | Выбор каждого элемента <p>, который является единственным элементом <p> своего родительского элемента. | 3 |
| [:only-child](https://puzzleweb.ru/css/sel_only-child.php) | p:only-child | Выбор каждого элемента <p>, который является единственным дочерним элементом своего родительского элемента. | 3 |
| [:nth-child(n)](https://puzzleweb.ru/css/sel_nth-child.php) | p:nth-child(2) | Выбор каждого элемента <p>, который является вторым дочерним элементом своего родительского элемента. | 3 |
| [:nth-last-child(n)](https://puzzleweb.ru/css/sel_nth-last-child.php) | p:nth-last-child(2) | Выбор каждого элемента <p>, который является вторым дочерним элементом своего родительского элемента, считая от последнего дочернего элемента. | 3 |
| [:nth-of-type(n)](https://puzzleweb.ru/css/sel_nth-of-type.php) | p:nth-of-type(2) | Выбор каждого элемента <p>, который является вторым дочерним элементом <p> своего родительского элемента. | 3 |
| [:nth-last-of-type(n)](https://puzzleweb.ru/css/sel_nth-last-of-type.php) | p:nth-last-of-type(2) | Выбор каждого элемента <p>, который является вторым дочерним элементом <p> своего родительского элемента, считая от последнего дочернего элемента. | 3 |
| [:last-child](https://puzzleweb.ru/css/sel_last-child.php) | p:last-child | Выбор каждого элемента <p>, который является последним элементом своего родительского элемента. | 3 |
| [:root](https://puzzleweb.ru/css/sel_root.php) | :root | Выбор корневого элемента в документе. | 3 |
| [:empty](https://puzzleweb.ru/css/sel_empty.php) | p:empty | Выбор каждого элемента <p>, который не содержит дочерних элементов (включая текст). | 3 |
| [:target](https://puzzleweb.ru/css/sel_target.php) | :target | Выбор текущего целевого элемента на странице, то есть элемента, к которому был осуществлён переход по ссылке внутри страницы. | 3 |
| [:enabled](https://puzzleweb.ru/css/sel_enabled.php) | input:enabled | Выбор каждого включенного элемента <input>. | 3 |
| [:disabled](https://puzzleweb.ru/css/sel_disabled.php) | input:disabled | Выбор каждого выключенного элемента <input>. | 3 |
| [:checked](https://puzzleweb.ru/css/sel_checked.php) | input:checked | Выбор элемента <input>, выбранного по умолчанию или пользователем. | 3 |
| [:not(селектор)](https://puzzleweb.ru/css/sel_not.php) | :not(p) | Выбор всех элементов, кроме элемента <p>. | 3 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Псевдо-элемент** | **Пример** | **Описание** | **CSS** |
| [::first-letter](https://puzzleweb.ru/css/sel_firstletter.php) | p::first-letter | Выбор первой буквы каждого элемента <p>. | 1 |
| [::first-line](https://puzzleweb.ru/css/sel_firstline.php) | p::first-line | Выбор первой строки каждого элемента <p>. | 1 |
| [::before](https://puzzleweb.ru/css/sel_before.php) | p::before | Добавляет элемент с содержимым перед содержимым каждого элемента <p>. | 2 |
| [::after](https://puzzleweb.ru/css/sel_after.php) | p::after | Добавляет элемент с содержимым после содержимого каждого элемента <p>. | 2 |