Параметры для парсинга в Netpeak Spider

Google

Содержание Featured Snippet → XPath → //span[@class='ILfuVd']

Содержание маркированного списка → XPath → //ul[@class="i8Z77e"]/li

Содержание нумерованного списка → XPath → //ol[@class="X5LH0c"]/li

URL результата Featured Snippet → XPath → (//cite[@class="iUh30"])[1]

Источник изображения → XPath → //div[@class="rg_ilbg"]

Запросы из блока «Похожие запросы» (People also ask) → XPath →
(//div[1]/g-accordion-expander/div/div)[1]
(//div[2]/g-accordion-expander/div/div)[1]
(//div[3]/g-accordion-expander/div/div)[1]
(//div[4]/g-accordion-expander/div/div)[1]

Заголовки рекламной выдачи → CSS-селектор → [class="sA5rQ"]

URL результатов в рекламной выдаче → CSS-селектор → .V0MxL → Значение атрибута → href

Запросы из блока “Вместе с … часто ищут” (Searches related to) → XPath → //*[@class="brs_col"]

Amazon

Пример страницы для проверки: https://www.amazon.com/dp/B0799HPMFT/

Характеристики → XPath → //table[@class="a-keyvalue prodDetTable"]/tbody/tr/td
Цена → CSS-селектор → #priceblock_ourprice
Рейтинг → CSS-селектор → [data-hook="rating-out-of-text"]
Отзывы → CSS-селектор → [id^="customer_review"]

Wikipedia

Пример страницы для проверки:
https://ru.wikipedia.org/wiki/%D0%AD%D0%BA%D1%81%D0%BF%D0%B5%D1%80%D0%B8%D0%BC%D0%B5%D0%BD%D1%82_%D0%90%D1%88%D0%B0

Ссылки из разделов «Примечания», «Литература» и «Ссылки» → CSS-селектор → a.external.text[rel='nofollow'] → Значение атрибута → href

Запрос на уточнение информации → CSS-селектор → [class="plainlinks flaggedrevs_preview"]

Инфобокс → CSS-селектор → .infobox

SimilarWeb

Шаг 1. Сгенерировать URL вида *similarweb.com/website/ + site.com в MS Excel. Например:*

similarweb.com/website/netpeaksoftware.com
similarweb.com/website/bbc.com

Добавить список в Netpeak Spider.

Шаг 2. Задать настройки парсинга:

Global Rank → CSS-селектор → li[class~="js-globalRank"]>div[class~="js-websiteRanksValue"]

Country Rank → CSS-селектор → li[class~="js-countryRank"]>div[class~="js-websiteRanksValue"]

Category Rank → CSS-селектор → li[class~="js-categoryRank"]>div[class~="js-websiteRanksValue"]

Total Visits → CSS-селектор → [data-type="time"]>[class~="js-countValue"]

Avr. Visit Duration → CSS-селектор → [data-type="ppv"]>[class~="js-countValue"]

Bounce Rate → CSS-селектор → [data-type="bounce"]>[class~="js-countValue"]

Organic Search → CSS-селектор → [class~="searchPie-text--left"]>[class="searchPie-number"]

Paid Search → CSS-селектор → [class~="searchPie-text--right"]>[class="searchPie-number"]

Social Traffic → CSS-селектор → li[data-key="Social"]

Ad Traffic → CSS-селектор → [class="subheading-value display"]

Mail Traffic → CSS-селектор → li[data-key="Mail"]

Direct Traffic → CSS-селектор → li[data-key="Direct"]

Top Referring Sites → CSS-селектор → [class^="referrals-tree"]+ul>li
Similar Sites → CSS-селектор → ul[class~="js-similarSitesList"]>li → Значение атрибута → data-site

Шаг 3. Сохранить настройки в шаблон и запустить сканирование.

Reddit

Пример страницы для проверки: https://www.reddit.com/user/davidreiss666

Даты последней активности юзера → CSS-селектор → [class="_1sA-1jNHouHDpgCp1fCQ_F"]

Пример страницы для проверки: https://www.reddit.com/r/startups

Список всех подписчиков и подписчиков онлайн → CSS-селектор → [class="_3XFx6CfPlg-4Usgxm0gK8R"]

Список постеров → CSS-селектор → a[class="_2tbHP6ZydRpjI44J3syuqC _23wugcdiaj44hdfugIAlnX oQctV4n0yUb0uiHDdGnmE"] → Значение атрибута → href

Список постов → CSS-селектор → a[class="SQnoC3ObvgnGjWt90zD9Z _2INHSNB8V5eaWp4P0rY_mE"] → Значение атрибута → href

Пример страницы для проверки: https://www.reddit.com/r/startups/about/moderators/

Список модераторов сабреддита → CSS-селектор → a[class="_2Q3rLIRb_ij54AEsabVm9L"] → Значение атрибута → href

Пример страницы для проверки: https://www.reddit.com/r/science/comments/d91niz/ai_equal_with_human_experts_in_medical_diagnosis/

Заголовок поста → CSS-селектор → [class="_2SdHzo12ISmrC8H86TgSCp _29WrubtjAcKqzJSPdQqQ4h "]

Текст поста → CSS-селектор → [class="_3xX726aBn29LDbsDtzr_6E _1Ap4F5maDtT1E1YuCiaO0r D3IL3FD0RFy_mkKLPwL4"]

Профиль автора поста → CSS-селектор → a[class="_2tbHP6ZydRpjI44J3syuqC _23wugcdiaj44hdfugIAlnX oQctV4n0yUb0uiHDdGnmE"] → Значение атрибута → href

Содержание комментариев → CSS-селектор → [class="_1qeIAgB0cPwnLhDF9XSiJM"]

Список комментаторов → CSS-селектор → a[class="_23wugcdiaj44hdfugIAlnX "] → Значение атрибута → href

Quora

Пример страницы для проверки: https://www.quora.com/topic/Search-Engine-Marketing

Ссылки на вопросы → CSS-селектор → a[class="question_link"] → Значение атрибута → href

Заголовки вопросов → CSS-селектор → [class="ui_content_title ui_content_title--default ui_content_title--medium"]

Пример страницы для проверки: https://www.quora.com/search?q=seo

Ссылки на вопросы → CSS-селектор → a[class="question_link"] → Значение атрибута → href

Заголовки вопросов → CSS-селектор → [class="ui_content_title unstyled_ui_title"]

Пример страницы для проверки: https://www.quora.com/Which-form-of-marketing-is-best-PPC-SEO-or-digital-marketing-Why

Похожие вопросы → CSS-селектор → [class="related_question"] a[class="question_link"]

Наличие Answer Wiki → CSS-селектор → [class="AnswerWikiArea"]

Пример страницы для проверки:
https://www.quora.com/Which-form-of-marketing-is-best-PPC-SEO-or-digital-marketing-Why/log

Информация по вопросу → CSS-селектор → [class="QuestionStats"]>div>span

Пример страницы для проверки:
https://www.quora.com/topic/Search-Engine-Marketing/writers

Самые просматриваемые авторы → CSS-селектор → a[class="user"] → Значение атрибута → href

WikiHow

Пример страницы для проверки:
https://www.wikihow.com/Special:ListRequestedTopics?st_search=marketing&category=

Вопросы без ответа → CSS-селектор → [class="st_title"]

Пример страницы для проверки:
https://www.wikihow.com/wikiHowTo?search=marketing&start=50&beta=true

Ссылки на статьи → CSS-селектор → a[class="result_link"] → Значение атрибута → href

Заголовки статей → CSS-селектор → [class="result_title"]

Пример страницы для проверки: https://www.wikihow.com/Cram-for-Math

Статистика статьи → CSS-селектор → [class="sp_text_data"]

Рейтинг → CSS-селектор → [class="sp_helpful_rating_count"]

Yelp

Пример страницы для проверки:
https://www.yelp.com/search?find_desc=Roofers&find_loc=San+Francisco%2C+CA&ns=1

Ссылки на страницы бизнесов → CSS-селектор → [class="lemon--h3__373c0__sQmiG heading--h3__373c0__1n4Of alternate__373c0__1uacp"] a[class="lemon--a__373c0__IEZFH link__373c0__29943

link-color--blue-dark__373c0__1mhJo link-size--inherit__373c0__2JXk5"] → Значение атрибута → href

На Yelp есть два вида вёрстки, которые чередуются в зависимости от категории. Назовём их «старая» и «новая».

Старая вёрстка. Пример страницы для проверки:
https://www.yelp.com/biz/prosper-construction-development-san-francisco

Ссылки на страницы комментаторов → CSS-селектор → [class~="user-passport-info"] a[class~="lemon--a__373c0__IEZFH"] → Значение атрибута → href

Контакты и адрес → CSS-селектор → [class="lemon--div__373c0__1mboc island__373c0__3fs6U u-padding-t1 u-padding-r1 u-padding-b1 u-padding-l1 border--top__373c0__19Owr border--right__373c0__22AHO border--bottom__373c0__uPbXS border--left__373c0__1SjJs border-color--default__373c0__2oFDT background-color--white__373c0__GVEnp"]>div

Рейтинг → CSS-селектор → [class="lemon--div__373c0__1mboc u-space-b3 border-color--default__373c0__2oFDT"] [class~="i-stars__373c0__30xVZ"]

Кол-во отзывов → CSS-селектор → [class="lemon--p__373c0__3Qnnj text__373c0__2pB8f text-color--mid__373c0__3G312 text-align--left__373c0__2pnx_ text-size--large__373c0__1568g"]

Название бизнеса → XPath → //h1
Текст отзывов → CSS-селектор → [class="lemon--p__373c0__3Qnnj text__373c0__2pB8f comment__373c0__3EKjH text-color--normal__373c0__K_MKN text-align--left__373c0__2pnx_"] [class="lemon--span__373c0__3997G"]

Новая вёрстка. Пример страницы для проверки:
https://www.yelp.com/biz/seacliff-dental-san-francisco-3

Ссылки на страницы комментаторов → CSS-селектор → [class="review-sidebar-content"] a[class="user-display-name js-analytics-click"] → Значение атрибута → href

Кол-во отзывов → CSS-селектор → [class="biz-page-header clearfix"] [class="review-count rating-qualifier"]

Рейтинг → CSS-селектор → [class="biz-rating biz-rating-very-large clearfix"] [class~="i-stars"]

Адрес → CSS-селектор → [class="street-address"]

Ссылка на сайт → CSS-селектор → [class="biz-website js-biz-website js-add-url-tagging"] a → Значение атрибута → href

Бонусы посетителям → CSS-селектор → [class="check-in-offer-text"]

Название бизнеса → XPath → //h1

Текст отзывов → CSS-селектор → [class="review-content"] p

Better Business Bureau

Пример страницы для проверки: https://www.bbb.org/search?find_country=USA&find_entity=60289-000&find_id=1533_3900-100&find_latlng=41.867220%2C-87.666981&find_loc=Chicago%2C%20IL&find_text=Dentist&find_type=Category&page=1

Ссылки на бизнесы → CSS-селектор → [class="MuiTypography-root Name-sc-1srnbh5-0 gApSUX MuiTypography-h4 MuiTypography-gutterBottom"] a → Значение атрибута → href

Пример страницы для проверки: https://www.bbb.org/us/il/chicago/profile/dentist/chicago-all-care-dental-group-0654-88111467

Номер телефона → CSS-селектор → [class="dtm-phone"]

Адрес → CSS-селектор → [class="MuiCardContent-root Content-dqnwmu-0 dKkECV"]

Адрес сайта → CSS-селектор → [class="MuiCardContent-root Content-dqnwmu-0 dKkECV"] [class="MuiTypography-root MuiTypography-body1"] a → Значение атрибута → href

G2Crowd

Пример страницы для проверки: https://www.g2.com/categories/seo

Список продуктов → CSS-селектор → a[class="link flex fw-nw text-small ai-c"] → Значение атрибута → href

Пример страницы для проверки:
https://www.g2crowd.com/products/netpeak-spider/reviews

Текст отзывов → CSS-селектор → [itemprop="reviewBody"]

Ссылки на профили обзорщиков → CSS-селектор → [class="mr-1"] a → Значение атрибута → href

Stack Exchange

Пример страницы для проверки: https://vegetarianism.stackexchange.com/

Ссылки на вопросы → CSS-селектор → a[class="question-hyperlink"] → Значение атрибута → href

Заголовки вопросов → CSS-селектор → [class="question-hyperlink"]

Пример страницы для проверки:
https://vegetarianism.stackexchange.com/questions/1820/is-consensual-cannibalism-vegan

Заголовок вопроса → CSS-селектор → [id="question-header"]

Дата создания → CSS-селектор → [class="grid fw-wrap pb8 mb16 bb bc-black-2"] [itemprop="dateCreated"]

Дата последней активности → CSS-селектор → [class="grid fw-wrap pb8 mb16 bb bc-black-2"] [class="s-link s-link__inherit"]

Просмотры → CSS-селектор → [title~="Viewed"]

Рейтинг → CSS-селектор → [class="question"] [itemprop="upvoteCount"]

Количество ответов → CSS-селектор → [class="subheader answers-subheader"] [itemprop="answerCount"]

Schema, GA, GTM, OpenGraph, Twitter Cards

Использованные типы Schema → XPath → //*[@itemtype] → Значение атрибута → itemtype

Использованные аттрибуты Schema → XPath → //*[@itemprop] → Значение атрибута → itemprop

og:title → CSS-селектор → [property="og:title"] → Значение атрибута → content

twitter:title → CSS-селектор → [property="twitter:title"] → Значение атрибута → content

Код GA → RegExp → ["'](UA-.*?)["']

Код GTM → RegExp → ["'](GTM-\w+)["']

Google play store

Пример страницы для проверки: https://play.google.com/store/apps/collection/cluster?clp=ogouCAESBEdBTUUaHAoWcmVjc190b3BpY196R0FHTHI5M2FUQRA7GAMqAggCUgIIAg%3D%3D:S:ANO1ljKNL98&gsr=CjGiCi4IARIER0FNRRocChZyZWNzX3RvcGljX3pHQUdMeXkzYVRBEDsYAyoCCAJSAggC:S:ANO1ljKozJM&hl=en

Ссылки на продукты → CSS-селектор → [class="wXUyZd"] a → Значение атрибута → href

Пример страницы для проверки: https://play.google.com/store/apps/details?id=com.fingersoft.hillclimb&hl=en

Рейтинг → CSS-селектор → [class="BHMmbe"]

Ссылки на страницы комментаторов → CSS-селектор → [class="AYi5wd TBRnV"]

Информация о приложении → CSS-селектор → [class="hAyfc"]

Описание приложения → CSS-селектор → [jsname="sngebd"]

Facebook

Пример страницы для проверки: https://www.facebook.com/buro247ukraine

Информация о странице (количество лайков, подписчиков) → CSS-селектор → [class="_4bl9"]

Дата создания → CSS-селектор → [class="_3qn7 _61-0 _2fyi _3qnf _2pi9 _3-95"]

Instagram

Пример страницы для проверки: https://www.instagram.com/kasta.ua

Информация о странице из описания OG (количество постов, подписчиков, подписок) → CSS-селектор → [property="og:description"] → Значение атрибута → content

Pinterest

Пример страницы для проверки: https://www.pinterest.com/buro247ukraine/
Количество подписчиков/подписок → CSS-селектор → [class="tBJ dyH iFc SMy _S5 pBj DrD mWe"]

Twitter

Пример страницы для проверки: https://twitter.com/tsnua

Количество твитов → CSS-селектор → [class="ProfileNav-item ProfileNav-item--tweets is-active"] [class="ProfileNav-value"] → Значение атрибута → data-count

Количество подписок → CSS-селектор → [class="ProfileNav-item ProfileNav-item--following"] [class="ProfileNav-value"] → Значение атрибута → data-count

Количество подписчиков → CSS-селектор → [class="ProfileNav-item ProfileNav-item--followers"] [class="ProfileNav-value"] → Значение атрибута → data-count

Нравится → CSS-селектор → [class="ProfileNav-item ProfileNav-item--favorites"] [class="ProfileNav-value"] → Значение атрибута → data-count

Дата регистрации → CSS-селектор → [class="ProfileHeaderCard-joinDateText js-tooltip u-dir"]

Описание → CSS-селектор → [class="ProfileHeaderCard-bio u-dir"]