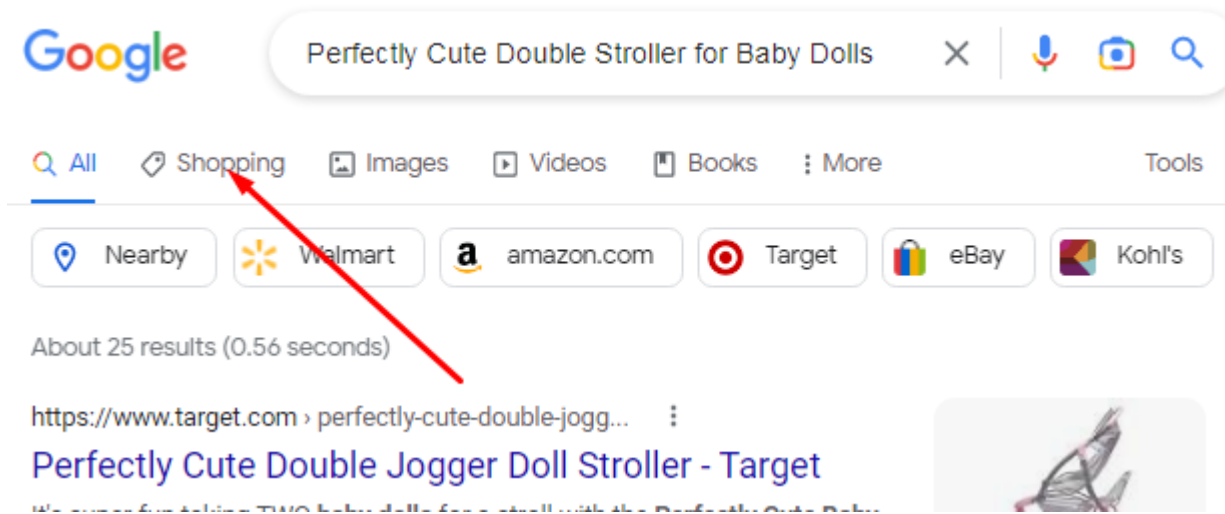


Парсер Google.com /shopping

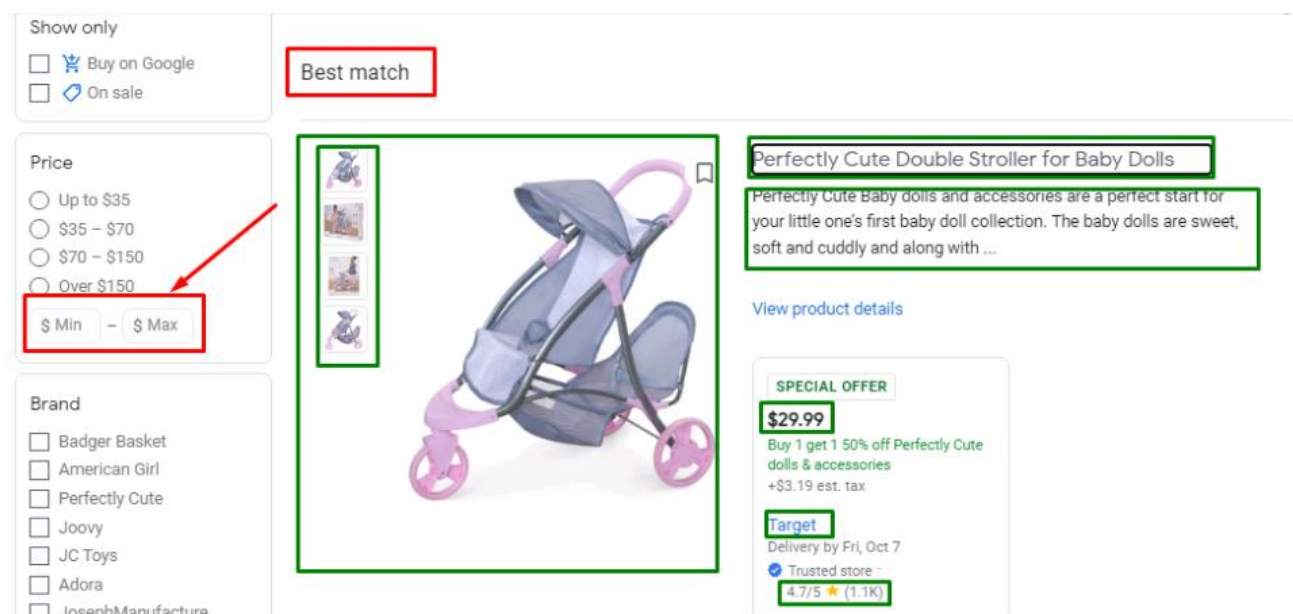
Парсинг производим через поиск на главной странице поиска (должен быть включен vpn US), по значению из поля title из json столбца info исходной таблицы parser_trademarkia_com (имя исходной таблицы может меняться и оно задается в конфиг файле config.inc.php веб-проекта). В результате выдачи, в первом запросе, при каждом запуске парсера, выбираем закладку Shopping :



После выбора shopping попадаем в данный раздел, в котором далее, при последующих запросах работы парсера и производим поиск и парсинг результатов выдачи, где задаем в фильтре значения:

- maxPrice берем из поля PriceMaxFind в json столбце info исходной таблицы ;
- minPrice берем из настроек, где задаем 2 поля: в % от значения PriceMaxFind, но не менее заданного абсолютного значения ;

Title_shop(заголовок) / Image_shop (урлы всех картинок) / Price (стоимость товара+est tax)/ Stores (имя магазина)/ Url_stores (ссылка страницы товара в магазине)/ URL_shop (ссылка страницы товара на google)/ Shop (пишем: google)/Other_Stores*/Prices_other_stores*/Urls_other_stores*/Categories_Root (значение из поля Categories: Root в json колонке info в исходной таблице parser_trademarkia_com той записи по которой был найден и записан результат)



*эти поля парсим после посещения страницы, и опцию посещения страницы задаем в настройках, в ссылке Url_stores отбрасываем лишнее, данные записываем только в том случае, если цена Prices_other_stores < PriceMaxFind

Compare buying options

Sold by	Details & special offers	Item price	Total price	
Amazon.com Trusted store · 4.5/5 ★ (3.8K)	Free delivery 30-day returns	GREAT PICK \$38.99	\$42.45	Visit site
Target Trusted store · 4.7/5 ★ (1.1K)	Free delivery 90-day returns Members earn exclusives & free shipping	\$38.99	\$42.45	Visit site
Walmart Trusted store · 4.5/5 ★ (1.4K)	Free delivery Free 90-day returns	\$38.99	\$42.45	Visit site

Пример выше по запросу: Perfectly Cute Double Stroller for Baby Dolls

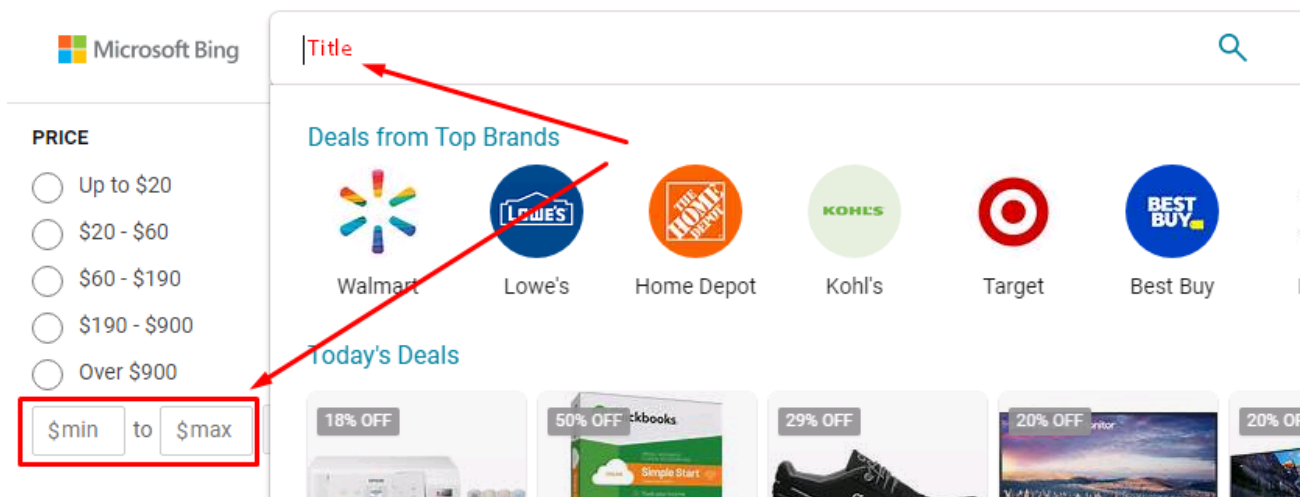
Парсер shopping.bing.com

<https://www.bing.com/shop> - для запуска выбрать страну US в настройках

Парсинг производим через поиск на странице <https://www.bing.com/shop> по значению из поля title из json столбца info исходной таблицы parser_trademarkia_com (имя исходной таблицы может меняться и оно задается в конфиг файле config.inc.php веб-проекта) и указанием диапазона цены, где

- maxPrice берем из поля PriceMaxFind в json столбце info исходной таблицы ;

- minPrice берем из настроек, где задаем 2 поля: в % от значения PriceMaxFind, но не менее заданного абсолютного значения ;



Полученный результат в виде списка, парсим в отдельную таблицу parser_shopping_results (таблица в бд уже имеется) , в столбец result (json) в следующие поля:



Title_shop(заголовок) / Image_shop (урл картинки) / Price/ Stores (имя магазина)/ Url_stores* (ссылка страницы товара в магазине)/ URL_shop (ссылка страницы товара на bing)/ Shop (пишем: bing)/Categories_Root (значение из поля Categories: Root в json колонке info в исходной таблице parser_trademarkia_com той записи по которой был найден и записан результат)/Other_Stores*/Prices_other_stores*/Urls_other_stores*

эти поля парсим после посещения страницы, и опцию посещения страницы задаем в настройках, в ссылке Url_stores отбрасываем лишнее, поля Other_Stores/Prices_other_stores*/Urls_other_stores* заполняем (возможность отключить в настройках) по клику See more только в том случае, если цена Prices_other_stores < PriceMaxFind

URL_bing_search (урл поиска который был сформирован для поиска записываем в исходную таблицу.

Парсер Shopping.Yahoo.com

Парсинг производим по прямым ссылкам которые формируем на основании данных из json столбца info исходной таблицы parser_trademarkia_com (имя исходной таблицы может меняться и оно задается в конфиг файле config.inc.php веб-проекта) , для каждой записи исходной таблицы по следующему алгоритму:

Пример ссылки:

<https://shopping.yahoo.com/category/888?maxPrice=20&minPrice=10&orderBy=DESC&p=Chevron%20Havoline%20Lifelong%20Full%20Synthetic%20Motor%20Oil>

где 888 – id категории который определяем согласно справочника ниже (справочник переносим в бд как отдельная таблица: categories_yahoo) путем сверки значений из одноименных полей Categories: Tree в json столбце info исходной таблицы и в таблице categories_yahoo приведенной ниже и ищем максимальное совпадение значений и далее по таблице ниже берем соответствующее значение Id_categories;

Пример: для значения Electronics › Car & Vehicle Electronics › Car Electronics › Car Audio › Subwoofers › Component Subwoofers из исходной таблицы будет соответствовать значению Electronics (id=222) в таблице ниже, а для значения Electronics › Camera & Photo › Video Surveillance › Surveillance Cameras › Bullet Cameras будет соответствовать значению Electronics › Camera & Photo (id=141) в таблице ниже

- 20 – значение maxPrice берем из поля PriceMaxFind в json столбце info исходной таблицы ;
- 10 – значение minPrice берем из настроек, где задаем 2 поля: в % от значения PriceMaxFind, но не менее заданного абсолютного значения ;
- Chevron%20Havoline%20Lifelong%20Full%20Synthetic%20Motor%20Oil - значение из поля title в json столбце info исходной таблицы, вместо пробелов в тексте вставляем %;

Категории (создаем таблицу в бд) и по этой таблице определяем id для каждой записи в исходной таблице

Categories: Tree	Id_categories	categories_yahoo*
Automotive	888	Automotive
Baby Products	537	Babies & Kids
Electronics › Camera & Photo	141	Cameras
Clothing, Shoes & Jewelry	2271	Clothing
Electronics › Computers & Accessories	278	Computers
Electronics › Television & Video	839	DVDs & Videos
Electronics	222	Electronics
Home & Kitchen › Furniture	436	Furniture
Grocery & Gourmet Food	422	Food
Clothing, Shoes & Jewelry › Luggage & Travel Gear	6551	Handbags & Wallets
Beauty & Personal Care	469	Health & Beauty
Patio, Lawn & Garden	536	Home & Garden
Health & Household › Household Supplies	604	Household Appliances
Home & Kitchen	730	Kitchen Appliances
Musical Instruments	855	Music
Office Products	922	Office Supplies
Pet Supplies	1	Pets
Sports & Outdoors	988	Sports & Outdoor
Toys & Games	1239	Toys & Games

*столбец categories_yahoo в таблице приведен для справки

По сформированной ссылке на сайте **Shopping.Yahoo.com** получаем результат в виде списка ниже который парсим в отдельную таблицу parser_shopping_results (таблица в бд уже имеется), в столбец result (json) следующие поля:

Title_shop(заголовок) / Image_shop (урл картинки) / Categories_shop (дерево всех категорий -их может быть несколько)/ Price/ Stores (имя магазина)/ Url_stores *(ссылка страницы товара в магазине)/ URL_shop

(ссылка страницы товара на yahoo)/ Shop (пишем: yahoo)/ Categories_Root (значение из поля Categories: Root в json колонке info в исходной таблице parser_trademarkia_com той записи по которой был найден и записан результат)

*Url_stores – парсим после посещения страницы, и опцию посещения страницы задаем в настройках, в ссылке отбрасываем лишнее

The image shows a screenshot of a Yahoo search results page. On the left, there are filter sections: 'Categories' with 'Vehicles & Parts' selected, 'Stores' with 'Walmart' checked, and 'Price' set between '\$10' and '\$20'. The main 'Results' section displays a product image of a 'Chevron Havoline Lifelong Full Synthetic Motor Oil' box. Below the image, the price '\$19.97' is shown, followed by the product title 'Chevron Havoline Lifelong Full Synthetic Motor Oil...' and the store name 'Walmart'. Red boxes and arrows highlight these elements with labels: 'categories', 'Image_yahoo', 'Price', 'Title', and 'Stores'.

URL_yahoo_search (урл поиска который был сформирован для поиска записываем в исходную таблицу

Настройки парсеров (общие для всех парсеров)

Отдельно в настройках создать текстовое поле (**Парсим только по:**) в которое можно внести один или несколько магазинов через разделитель | только по которым парсим, если поле пустое то по всем кроме игнорируемых в справочнике, при парсинге по заданному магазину слева в фильтре, отмечаем галками магазин (ы) если он(и) там есть, если нет, то пропускаем эту запись в бд.

В настройках задаем **глубину поиска**, количество результатов выдачи которые парсим, а если глубина поиска более чем количество результатов на странице выдачи результата (bing – 50/ yahoo -60), то парсер в конце поисковой выдачи должен раскрыть следующие результаты (по кнопке Load more - yahoo и по следующему номеру страницы выдачи результата - bing по следующему номеру страницы выдачи внизу страниц выдачи), парсер должен проверить магазины в результате с заблокированными магазинами в справочнике.

Настройки выносим на форму настроек как отдельный раздел и привязываем настройки к профилю – в заголовке полей в таблице настроек дописываем shopping.

В настройках предусмотреть 2 варианта парсинга, по сформированным ссылкам, согласно алгоритма выше (**Парсинг через поиск**), с посещением и без посещения страницы товара по ссылке (Url_stores), и **повторный парсинг** по прямым ссылкам из поля URL_shop в таблице результатов, для обновления цена и записи ссылки страницы товара в магазине в поле Url_stores.

Минимальная цена (minPrice): ____% от значения PriceMaxFind, но не менее ____\$

*Жирным выше отмечены имя поле в настройках.