Техническое задание на разработку модуля для парсинга цен. СMS OpenCart

По всем вопросам:

+7-906-265-79-24, Евгений (телефон, WhatsApp, Telegram)

epishkov.e.s@mail.ru

Пример того, как реализована часть функционала данного модуля, описан в данном документе: <https://drive.google.com/file/d/1iXADe-Ugk0STKH7nRsySLWOZ-HU5OpTt/view?usp=sharing>

## Общая информация:

Требуется разработать модуль для автоматического(через заданные промежутки времени) поиска и сравнения цен на определённые товары, который будет работать одновременно:

1. как самостоятельный инструмент, ищущий цены и вычисляющий средние значения для каждого товара
2. полностью интегрируемый модуль для CMS OpenCart версии …, который помимо поиска цен и вычисления среднего значения, сравнивает средние цены с нашим прайсом интернет магазина и выводит разницу в % в соответствующие карточки товаров.

Ресурсы для парсинга(список может незначительно меняться):

1. Перекресток (<https://www.perekrestok.ru>)
2. Лента (<https://sbermarket.ru/lenta>)
3. Окей (<https://www.okeydostavka.ru/spb>)
4. Дикси (<https://dostavka.dixy.ru>)
5. Магнит (<https://igooods.ru/magnit-dostavka-produktov-na-dom> - любой адрес в СПб)
6. Карусель (<https://igooods.ru/karusel-dostavka-produktov-na-dom> - любой адрес в СПб)
7. Ашан (<https://sbermarket.ru/auchan>)
8. Метро (<https://sbermarket.ru/metro>)
9. Призма (<https://igooods.ru/prisma-katalog-tovarov> - любой адрес в СПб)
10. Ряды (<https://ryady.ru>)

Также, интересует возможность парсинга мобильных приложений(*насколько изменится стоимость разработки*):

1. Чек.2Гис (<https://play.google.com/store/apps/details?id=ru.doublegis.cashback&hl=ru>) - средняя цена на товар
2. Пятёрочка.Доставка (<https://5-delivery.ru>; <https://play.google.com/store/apps/details?id=com.delivery_shop_5ka24.delivery_shop_5ka>)

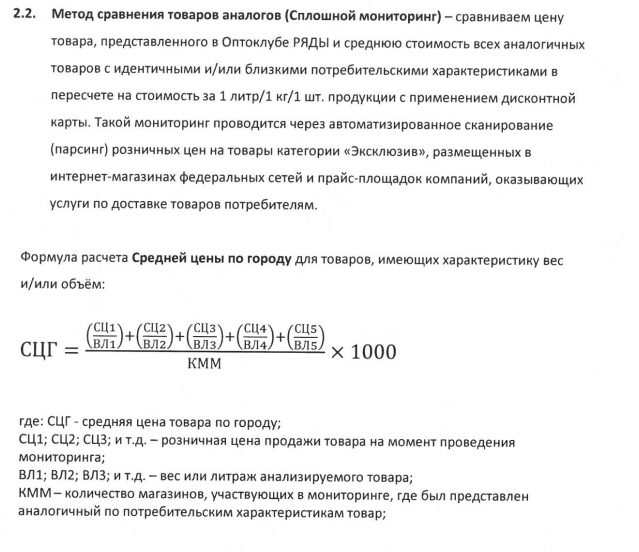
Что парсить:

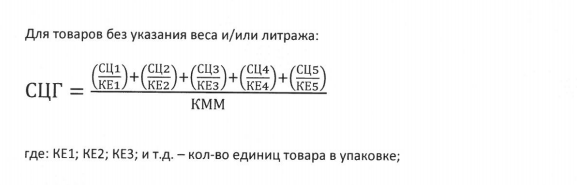
* Цена;
* Информация о скидке(да/нет; сколько %);
* Наличие(да/нет);

## Функциональные требования:

Изначальные задачи (как самостоятельного инструмента):

* Парсинг цен по загруженным спискам товаров на указанном списке интернет ресурсов.
* Автоматический запуск, через заданные промежутки.
* Расчёт средней цены на товар по окончании сессии. *(Цена в магазине 1 +...+ Цена в магазине N) / N.*
* Обработка информации и вывод в таблицу(*подробнее - ниже*).
* Если SKU искомой торговой марки не представлено ни на одном из указанных для поиска ресурсов и/или товар представлен только на одном ресурсе(выборка не репрезентативна), система должна собирать информацию о цене всех максимально одинаковых и/или схожих по потребительским характеристикам SKU в пересчёте на 1 кг/литр/шт.
  + *Необходимо разработать алгоритм, по которому это может работать. (Ключевые параметры: название; тип продукта(макароны спагетти, рожки, ракушки и т.д.) и др.). Интересны ваши мысли, как это может быть реализовано*.
  + В таблицу выводится только средняя цена на товар.
  + Рассчитанная средняя цена выводится со звёздочкой «\*», что означает, что средняя цена рассчитана по товарам не того же артикула, что и искомый, а по максимально похожим.
  + *Пример расчёта средней цены товаров аналогов:*

**

**

*Источник:*<https://drive.google.com/file/d/1iXADe-Ugk0STKH7nRsySLWOZ-HU5OpTt/view?usp=sharing>

* Вывод информации в виде онлайн-таблиц на каждую категорию.
* Регулярное автоматическое обновление информации в таблицах через заданные промежутки
* Предусмотреть обход защиты от парсинга, если она есть на каких-то из указанных ресурсов

Дополнительные задачи(как модуля для CMS OpenCart):

Помимо указанных выше функций:

* Интеграция данного инструмента в виде модуля в общую систему сервиса. Возможность настройки и корректная работа (чистый код, не конфликтующий с другими модулями; лёгкая интеграция; поддержка)
* Оптимизация нагрузки на сайт(запуск в определённое время; оптимизированный код и др.)
* Сравнение вычисленной средней цены на товар(*как в случае с прямым сравнением, так и в случае со средней ценой, вычисленной для товаров-аналогов*) с ценой из прайса (*будет два варианта цены:* как розницы, так и in bulk формата (одинаковое с одинаковым, либо подобное с подобным):

**% выгоды Розница =** (*Средняя Цена из поиска / Цена из прайса Розница – 1) \* 100;*

**% выгоды in bulk =** (*Средняя Цена из поиска / Цена из прайса in bulk – 1) \* 100;*

* + Если % выгоды **больше** нуля, то автоматический вывод его значения в карточку товара(Розница или in bulk).
  + Если % выгоды **меньше** нуля, то:
    - оставляем предыдущее значение % выгоды в карточке товара без изменения;
    - отправляем уведомления / письма на указанный email с текстом вида: «Цена на *Название товара(ссылка на карточку товара)* выше средней по рынку. Изменить(*ссылка в админку, где можно загрузить новый прайс)».*
    - После 2-3 повторения подряд, в систему отправляется автоматический запрос на то, чтобы информация о % выгоды убиралась из карточки товара.

Входная информация:

* Список ресурсов для парсинга. Настраивается очень редко.
* Список / прайс лист SKU для парсинга (*форматы xls/xml*); вся необходимая информация по ним для поиска таких же или подобных (название, характеристики, «тэги» и пр. *- в каком виде она должна быть - определяется изначально вместе с разработчиком*). Меняется постоянно.
* Расписание работы. Меняется не часто. Возможна настройка штатным разработчиков через код.

Выходная информация:

Для инструмента:

* Вывод в Гугл Таблицы(Google Sheets) / (*либо просто таблицы .xls*), где всё удобно собрано и отсортировано(цветовая разметка, сортировка по цене).
* Отдельная таблица для каждой категории. Список категорий совпадает с каталогом на сайте(*Макаронные изделия, Специи, Безалкогольные напитки и т.д.)*
* *Пример оформления таблицы:*

*Строка 1 - Название категории*

*Строка 2 -* название товара 1 (*выделение жирным*) | цена из нашего прайса Розница (*если есть*) | цена из нашего прайса in bulk (*если есть*) | акция у нас в прайсе(*если есть*)

*Строка 3 - Текст(«*средняя цена*»)* | цена средняя | \* - если средняя цена вычислена по аналогичным товарам(*в этом случае со следующей строки начинается другой товар*)

*/ далее, вначале отображаются товары с наименьшей ценой и дальше они сортируются по возрастанию цены*

*Строка 4 -* Название ресурса 1 | цена на ресурсе 1 | % акции, если есть | наличие(да/нет) | прямая ссылка на товар 1

...

*Строка (4 + n) -* Название ресурса n(*последнего*) | цена на ресурсе n | % акции, если есть | наличие(да/нет) | прямая ссылка на товар 1

*Строка (4 + n) + 1) -* название товара 2 (*выделение жирным*) | цена из нашего прайса(*если есть*) | цена из нашего прайса in bulk (*если есть*) | акция у нас в прайсе(*если есть*)

....

*/ и так с каждым товаром до конца категории*

Для модуля:

* Вывод информации о рассчитанном % выгоды в карточку товара
* Посылка уведомлений/письма на почту адресату, как было описано выше.

Дополнительные требования:

Для управления данным инструментом, а также добавления информации, крайне желательно предусмотреть минимально возможную графическую оболочку:

* чтобы поиск запускался кнопкой, а не из коммандной строки;
* для внесения информации, также, были поля ввода, либо он умел читать экселевские таблицы, оформленные по определённому образцу.

## Права пользователей:

- Администратор может:

· Запустить поиск вручную

· Изменить автоматическое расписание поиска

· Добавить списки/прайсы на товары/тэги/ключевые запросы для парсинга (*обсудить формат с разработчиком)*

· Загрузка таблицами

· Удалить товары/тэги/ключевые запросы для парсинга (*обсудить формат с разработчиком)*

· Скачать получившиеся результаты сравнения в формате таблиц

· Получать автоматические уведомления о:

* том, что % выгоды на какой-то из товаров меньше заданной величины

· Изменять текст уведомления(включая правку гиперссылок)

- Авторизованный/Неавторизованный пользователь может:

· Видеть % выгоды в карточке товара

## Примерные сроки разработки и тестирования модуля:

От 7 – 10 дней для версии «инструмент».

Ещё + 3 – 5 дней для версии «модуль».