Тема "Автоматизированная система мониторинга цен конкурентов"

1. Разбор исходных страниц магазинов.

1.1 Разбор каталога сайта.

1.1.1 Модуль разбора главного каталога.

Входные данные: магазин (идентификатор, название, url, кодировка), параметры разбора главного каталога (ассоциативный массив строк "имя параметра" -> "атрибуты в html странице"), html страница;

Выходные данные: главный каталог (идентификатор, название, url, идентификатор родительского каталога);

1.1.2 Рекурсивный обход каталогов сайта.

Входные данные: родительский каталог, параметры разбора страницы, html страница.

Выходные данные: каталог;

1.1.3 Обработка каталога с товарами.

Входные данные: родительский каталог, html страница, параметры разбора страницы, регион, дата разбора;

Выходные данные: товар (идентификатор, идентификатор каталога, имя товара, описание товара), цена-дата(идентификатор, цена, дата записи, наличие товара, идентификатор товара).

1.2 Разбор страниц товаров

1.2.1 Обработка страницы с товаром (при отсутствии каталога с необходимым описанием).

Входные данные: url, магазин, параметры разбора страницы.

Выходные данные: товар

2. Анализ продуктов

2.1 Авторизация пользователя с правами редактора.

Входные данные: логин, пароль.

Выходные данные: идентификатор сессии.

2.2 Интерфейс привязки пользователем с правами редактора товаров из магазина к номинальным товарам.

2.2.1 Вывод списка непривязанных товаров из магазинов.

Входные данные: идентификатор сессии пользователя, привязанные магазины.

Выходные данные: список не привязанных товаров, html страница.

2.2.2 Вывод списка номинальных товаров, похожих на товар из магазина.

Входные данные: непривязанный товар.

Выходные данные: список номинальных товаров, html страница.

2.3 Обработка товара из магазина

2.3.1 Использование метрики по названию товара в конкретном магазине для определения похожих номинальных товаров.

Входные данные: товар в магазине, список номинальных товаров.

Выходные данные: список похожих товаров.

2.3.2 Ранжирование списка по количеству совпадений в названии и цене номинального товара по убыванию от самого вероятного.

Входные данные: список похожих товаров.

Выходные данные: ранжированный список похожих товаров.

3. Пользовательский интерфейс.

3.1 Авторизация пользователя;

Входные данные: логин, пароль.

Выходные данные: идентификатор пользователя

3.2 Просмотр пользователем цены и наличия товара в различных магазинах на текущий момент;

Входные данные: идентификатор пользователя, дата.

Выходные данные: html страница