

Создание интегрированных баз данных по товарам и ценам для Интернет-торговли

Арницанс Гунтис, Карнитис Гиртс
Латвийский университет
guntis.arnicans@lu.lv
girts.karnitis@lu.lv

Проблема интеграции информации

Развитие Е-бизнеса продолжается более 10 лет. Его популярность в разных странах различна, но независимо от этого число пользователей постоянно растет. Прогнозируется, что в ближайшее время будет продолжаться переход покупателей от традиционного вида покупок товаров в магазинах к покупкам в е-магазинах, используя интернет потому, что:

- уменьшается время поиска необходимого товара, т.к. это возможно сделать, не поднимаясь с кресла, легче сравнивать предложения и приобрести товар,
- цены в е-магазинах обычно значительно ниже, чем в обычных магазинах даже с учетом затрат на доставку товара и оплаты налогов. Исследования на западе показывают, что цена приблизительно меньше на 10%.

Однако при использовании е-магазинов появляется проблема как выбрать нужный товар по желаемой цене? Пока магазинов не так много и предлагаемый ассортимент не так велик, проблема не так существенна. Но если число магазинов больше 100 или 1000, а ассортимент товаров около 1000000, появляются проблемы с поиском товара и принятием решения. Психология человека: а вдруг в другом магазине товар лучше, цена ниже и обслуживание надежнее? Покупатель начинает метаться от одного магазина к другому, тратит время и нервы.

Существенный шаг вперёд в развитии е-магазинов был начат около 10 лет тому назад - это интеграция информации е-магазинов таким образом, чтобы предложить покупателю общий взгляд на доступный выбор товара. Во-первых это было и есть автоматическое сравнение цен товаров.

Покупателей обычно не интересуют технические нюансы интеграции е-магазинов, а именно, как получен общий взгляд на весь объем товаров и возможностях приобретения. По их понятию это большая информационная база данных. Поэтому и мы в этой статье будем считать, что интегрированная база данных е-магазинов по товарам и их ценам является базой данных, в которой собрана информация многих е-магазинов о предлагаемых товарах и услугах.

Настоящее состояние

И в мире, и в Белоруссии существуют много е-магазинов. В магазинах, которые предлагают товары одного и того же типа, обычно ассортимент очень похожий и можно один и тот же товар найти во многих магазинах. Покупатель, которого интересует более низкая цена, может потратить достаточно много времени, чтобы найти конкретный товар за приемлемую цену. Эта проблема осознана и имеет ее решение: сравнение цен одного и того же товара в различных магазинах. В мире существуют различные виды сервисов.

Сравнители цен для небольшого, специально отобранного круга магазинов. Эти системы поддерживают небольшое количество магазинов – для каждой группы товаров есть выбор в 20-50 магазинах. Ответственные за поддержку системы следят за тем, чтобы магазины успешно удовлетворяли желания клиентов, а именно, чтобы была хорошая

репутация, высокое качество обслуживания и приемлемые цены. В этих системах цена обычно не имеет решающей роли, здесь больший акцент приходится на репутацию магазина и качество обслуживания клиентов. Примерами таких сравнителей являются *Kelkoo*, *PriceGrabber.com*, *PriceRunner.com*.

Сравнители цен для широкого круга магазинов. В этих системах акцент делается на количество магазинов, меньше внимания уделяется качеству отдельных магазинов. Обычно в таких магазинах могут быть проблемы со сравнением товаров, т.к. объем предлагаемой продукции настолько велик, что система не предлагает достаточно квалифицированного инструмента для сравнения товаров. Обычно товар можно найти, просмотрев каталог по категориям товаров или проделать поиск по названию товара. Как пример такого инструмента можно назвать *Google products*.

Мола. Это место в Web-сети, где представляется возможность создания своего e-магазина, используя существующую инфраструктуру. Обычно предлагается имя домена третьего уровня, готовая платформа для создания интернет-магазина со стандартной функциональностью: категории продуктов, каталог продуктов, корзина товаров, поиск, оплата, используя кредитную карточку или инфраструктуру оплаты конкретного Мола (специфический вид оплаты). Такие мола популярны в Белоруссии, например, *onliner.by*, *shop.by*. В мире похожую услугу предлагает *yahoo.com*.

Искатели. Частично к интегрированным сервисам можно отнести любого искателя. Если у нас есть опыт работы с поисковыми инструментами и мы успешно сформулировали запрос, то мы можем отыскать магазины, которые предлагают необходимый товар. Однако мы не знаем сколько магазинов мы нашли и сколько вообще у нас есть магазинов, которые продают этот товар. Сравнение цен товара и услуг надо вести вручную, т.к. надо «зайти» в каждый магазин отдельно для получения нужной информации, чтобы принять решение.

Среда Е-бизнеса

Если мы рассматриваем только покупку в интернете, то модель e-бизнеса можно описать достаточно просто. Есть продавец товара, который создает доступную для покупателей информационную систему (обычно используя интернет и программу просмотра (браузер)). Покупатель, используя предложенную программу, отыскивает необходимый товар и высказывает желание его приобрести. Рассмотрение оплаты и системы функционального управления в рамках этой статьи не является существенным. Отношения между продавцом и покупателем «интимные», т.к. во время оказания/получения услуги третья персона не присутствует.

Однако, как только в игру включается посредник, которого впредь будем называть интегратором, ситуация существенно меняется. Создается треугольник отношений продавец-интегратор-покупатель. Пока интегратор маломощный, связь продавец-покупатель достаточно активна. Как только интегратор «нарастил мускулы», все больше покупателей выбирает возможность получить информацию, используя его услуги: формируются разные модели сотрудничества. В идеальной ситуации выигрывают все вовлеченные игроки, однако в реальной жизни какой-либо из игроков или пара игроков завоевывают ведущую роль и диктуют условия остальным игрокам.

На практике модель сотрудничества значительно сложнее, т.к. фактически имеются еще несколько игроков, которые влияют на ситуацию на рынке. В качестве примера можно привести поставщика системы расчетов, изготовителя товаров, поставщика услуг логистики как продавцам, так и покупателям. Надо отметить, что каждый из субъектов может быть во множественном числе. Конкуренция субъектов или же их объединение для

поддержки единой политики соответственно ослабляет или укрепляет позицию субъекта в этой игре купли/продажи.

Цель нашей статьи обратить внимание на интеграцию информации о товарах в единую базу данных. Рассмотрим, как это можно сделать и какое влияние это может произвести на рынок. Сразу отметим, что интеграторы существуют уже давно и есть возможность наблюдать различные модели, которые появились на практике.

Необходимость интегрированной базы данных

Как уже было отмечено, покупатели понемногу переходят к осуществлению покупок в е-магазинах потому, что это экономит затраченные ресурсы (время, деньги, нервы) для приобретения необходимого товара. С появлением большого числа е-магазинов эти выигрыши как будто исчезают, т.к. беготня по физическим магазинам заменяется рысканьем по виртуальным магазинам. Если есть проблема, тогда кто-то предложит ее решить; т.к. это есть способ как заработать деньги – в нашем случае это интеграторы, которые создают интегрированные базы данных и предлагают информационные услуги.

Важнейшая функциональность, которая создает условия для необходимости интеграторов, есть поиск и фильтрация информации. У покупателя должна быть возможность за короткий промежуток времени отыскать необходимые данные, используя для этого понятные ему термины, язык и приемы. Пока покупатель не нашел желаемого, не существенно насколько хорошо развиты другие сервисы. Так же как и магазину, интегратору надо приспособиться к желаниям клиента-покупателя, обеспечив различные взгляды на информацию.

Когда интегратору удастся привлечь (затянуть в свои сети) покупателей, он может начинать реализовывать свою политику. Поведение интегратора обязательно определяет источник необходимых ресурсов для поддержки сервиса и прибыль, если такая предусмотрена. Также существенным является вопрос о том, кто поддерживает существование этого сервиса, т.к. нейтральное отношение хорошего интегратора к игрокам довольно маловероятно (кто платит, тот и заказывает музыку). Нейтралитет чаще наблюдается в начальной стадии работы интегратора, пока происходит захват позиции на рынке.

Если интегратор становится на сторону покупателя, тогда необходимость интегрированной базы данных можно охарактеризовать следующими выигрышами:

- Рынок становится прозрачнее и объективнее. Непорядочным продавцам становится гораздо труднее использовать недобросовестные условия продажи. Появляются равносильные условия конкуренции;
- Доступен более широкий выбор товаров и есть возможность сравнения товаров между собой. Появляется возможность появиться на рынке новым производителям, которые заставляют большие компании снижать цены товаров;
- Дисперсия цен уменьшается, т.е. цены выравниваются и е-магазины вынуждены конкурировать в основном в улучшении качества услуг торговли (расчеты, доставка товара, гарантии);
- Более доступна становится информация о каждом продукте: изображение, спецификация, различные варианты комплектации продукта, обзоры экспертов о продуктах, отзывы покупателей о продукте или магазине и т.д.;
- Постоянно развивается поиск, фильтрация, подсоединяются новые е-магазины;
- Предлагаемая информация более объективна, т.к. главная цель не в том, чтобы покупатель что-либо купил, а в том, чтобы покупатель приобрел только нужный

товар по возможности более качественный, дешевый и услуги покупки были бы качественными и надежными;

- Производитель сближается с покупателем, т.к. покупатель получает более непосредственную информацию, производитель получает больше информации о желаниях клиента, уменьшается сумма платежа посредникам, что дает возможность покупателю приобретать больше и лучших товаров;
- Появляется возможность проводить аукционы среди продавцов за снижение цен по доставке товара;
- Обеспечивается свобода покупателю, т.к. предлагаемая покупателю информация не ограничена и не искажена, покупатель сам может свободно принять решение, что изучать, когда и что покупать.

В случае, если интегратор становится на сторону продавца (точнее группы продавцов), необходимость содержать интегрированную базу данных диктуют другие цели:

- Обеспечить, чтобы покупатель был привлечен к этой группе продавцов;
- Предлагать только ту информацию и только в таком виде, который выгоден продавцу;
- Поддерживать желаемый уровень цен, неафишируя, что фактически это не самые низкие цены, или вводить в заблуждение покупателя утверждая, что это выгодная покупка;
- Не допускать развитие конкурентов, стараться уменьшить их долю на рынке;
- Распознавать манеру поведения покупателя, чтобы незаметно заставлять его покупать еще и еще;
- Обеспечить, чтобы поиск и фильтрация первыми находили и «правильно» расставляли в списке те е-магазины, которые больше платят за такую услугу. Т.е. покупатель получает искаженный результат запроса;
- Автоматическое повышение цен при росте спроса;
- Накапливать информацию о зарегистрированных покупателях, платежеспособным клиентам указывать более высокие цены, чем покупателям с меньшими доходами;
- Новым покупателям предлагать более объективную информацию, пока они еще не убедились, что этот интегратор «самый лучший и самый объективный».

Очень похожая ситуация создается, если интегратора поддерживает производитель (группа производителей). Остается только вместо слова «продавец» вставить слово «производитель». Выше перечисленные цели это только примеры, чтобы проиллюстрировать ситуацию, которая появляется в том или ином случае.

Способы интеграции

Создание интегрированной базы данных возможно решать разными способами. Есть два генеральных подхода, как выполнить это задание: стандартизация и интеграция.

В случае стандартизации внедряются единые понятия, названия, приемы, протоколы взаимодействия, и т.д. Их акцептируют все или большая часть участников «игры». Обычно многие участники («игроки») об этом не могут договориться. Чаще всего может встетиться ситуация, когда интегратор достаточно могуществвен, чтобы ввести стандарты и заставить многих участников игры принять эти стандарты. Тот, кто не подчиняется,

обычно вынужден уйти или не вовлекается в сервисы интегратора. В нашем случае в основном речь идет о том подчиняется или нет e-магазин требованиям интегратора.

В случае интеграции, когда стандартизация на низком уровне, необходимо создать соответствие между многими компонентами (продукты, свойства продуктов, классификация продуктов). В основном здесь надо решать много неоднородных проблем. Часть из этих проблем рассмотрена ниже.

Все в одном

Это самый простой способ интеграции, который обычно используют создатели e-мола. Фактически это вариант использования стандартов. Все магазины e-мола созданы, используя единое программное обеспечение и единую структуру базы данных. В связи с этим практически не существуют проблемы технологического характера для объединения информации в одну большую систему. Для сбора информации существуют два варианта:

- Вариант А: У каждого магазина своя база данных (логическая или физическая) с обработкой запросов поступающих от клиентов
 - Online – каждый запрос собирает данные со всех баз во время выполнения;
 - Offline – регулярно через некоторое время создается большая интегрированная база данных со всех магазинов.
- Вариант Б: Данные всех магазинов находятся в одной базе данных и интеграция по существу не нужна.

Главная проблема, которую необходимо решить - это объединение семантически неоднородной информации – подробнее смотрите следующую главу.

Посылка информации ПОЛОЖИТЬ

В этом случае Магазин сам посылает свои данные интегратору. Магазин является инициатором обмена данных и может произвести обмен данных когда магазину удобно – например, когда изменились цены или ассортимент. Технологии и форматы записи данных могут быть различными. В этом случае надо решать ряд технологических проблем связанных с конвертацией данных.

Набор информации ВЗЯТЬ

В этом случае интегратор является активной стороной в коммуникации и обмен данными происходит по инициативе интегратора, когда это выгодно интегратору, а не магазину. Магазин предоставляет свои данные интегратору, интегратор сам берет данные. Технологии и форматы записи данных могут быть различными.

Может быть использована модель, когда магазин посылку специфических данных не предусматривает и Интегратор читает данные из интернет страницы магазина. В этих случаях также как и в предыдущем случае надо решать ряд технологических проблем.

Проблемы создания интегрированных БД

Интегрируя данные из разных источников, приходится столкнуться с рядом проблем, которые связаны с неоднородностью источников данных. Во-первых, это проблемы технического характера – источники данных созданы используя различные технологии, данные из этих источников получены в разных форматах (XML, HTML, XLS, DOC), использовано различное кодирование текста. В наши дни эти проблемы довольно легко решаются, т.к. существуют инструменты трансляции данных между различными форматами и различными таблицами кодирования.

Кодировка Данные могут быть закодированы, используя различные таблицы кодирования. Поэтому и необходима трансляция в одну таблицу кодирования, чтобы было возможно выполнить сравнение. Например, какой язык использован в данном случае – Английский или русский - *C, O, K, P, A, T, TOP?*

Один и тот же товар с разными названиями У нас нет возможности по названию легко определить, или это один и тот же это товар или это разные товары. Например - *Nokia 6270, Nokia 6270 мобильный телефон, Мобильный телефон Nokia 6270.*

Разные товары с одинаковым названием Может случиться, что есть несколько товаров, у которых одинаковое название, но товары отличаются. Например, кровать двуспальная. Например - *Кровать двуспальная.*

Одни и те же данные в разных форматах В разных системах одни и те же данные могут быть в различных форматах – могут отличаться форматы дат, в дате может быть или отсутствовать год, в качестве разделителя целой и десятичной части цены может быть как точка, так и запятая. Например - *26 апреля 2007г., 26.04.2007, 07/04/26.*

Разная интерпретация в зависимости от страны. В разных странах одни и те же данные могут быть интерпретированы по разному. Это могут быть различия в формате данных, в использованных единицах измерения (*м, мия, тысячи*), и в использованных валютах (*Белорусские рубли, Российские рубли, доллары*).

Структура информации. В разных источниках информации одни и те же данные могут быть структурированы по разному. Например,

- Текст: “Минск, Министерство торговли, ул. Кирова, 8”
- Структурированный:
 - Город: “Минск”
 - Улица: “Кирова”
 - Номер дома: “8”
 - Квартира: “”
 - Адресат: “Министерство торговли”

Количество свойств товаров. В различных источниках данных количество свойств у товара может отличаться – есть источники с широким описанием свойств товаров, есть источники данных, в которых почти совсем нет ни одного свойства товара.

Ошибочные значения свойств товаров. Может случиться, что в источнике данных указаны ошибочные значения свойств товаров.

Лишние символы. Источники данных могут присоединить дополнительные символы форматирования, которые не являются ошибочными, но не дают дополнительной информации. Например, *Nokia 6270 - “Nokia 6270”*.

Значения по умолчанию. Источник данных может предположить, что какая-то информация считается известной по умолчанию. Например, обычно в интернет-магазинах названия мобильных телефонов создаются по принципу «Производитель Модель», но в домашней страничке фирмы Nokia имеется только номер модели, т.к. производитель у всех один – Nokia. Например, *Nokia 6270, 3210.*

Данные из источников могут быть различной степени качества. В значениях полей могут быть как грамматические, так и ошибки содержательного характера. Выполняя интеграцию данных, надо знать, что делать с ошибочными данными.

Без уже указанных неоднородных проблем надо отметить, что источники данных, которыми в этом случае являются системы содержащие данные e-магазинов, со временем меняются и имеют свой взгляд на правильность организации хранения и отображения данных. У интегратора появляются следующие типичные проблемы:

- Ограниченные знания о структуре данных в источнике данных и о семантике данных. Неправильно поняв и интегрировав данные, покупателю может быть предложена дезинформация;
- Большое количество источников данных. Мануальная обработка каждого источника данных практически невозможна. Должны быть автоматические или полуавтоматические средства для получения и интеграции данных. Сервисы должны быть дешевыми, легко конфигурируемыми и поддерживаемыми;
- Резкая смена информации. Данные в источнике данных могут непрерывно меняться. Контроль изменения цен не самая трудная задача. Меняется состав товаров или же структура и свойства самого товара.

Использование базы данных

Интегрированные базы данных товаров можно использовать двумя главными способами. Во-первых, это поиск по свойствам товара или по ключевому слову, чаще всего по названию, но возможен также поиск по производителю и другим свойствам. Поиск по названию обычно производят, если известен товар, который желают купить, и цель поиска посмотреть и сравнить цены в разных интернет-магазинах. Поиск по свойствам производят если не известен конкретный товар, но известны кое-какие свойства. Типичный пример, известен желаемый размер фотоаппарата и границы цены, и производится поиск моделей, которые соответствуют указанным параметрам.

Второй способ использования базы данных – это просмотр предложений в одной или другой группе товаров. Этот способ обычно используется в случаях, когда хочется купить фотоаппарат, но в принципе нет никаких особых желаний – тогда пользователь смотрит, какие фотоаппараты имеются в продаже и каковы их свойства. В таком случае пользователю может очень помочь отзывы других пользователей и экспертов о конкретной модели.

Выигрыши и потери «игроков»

При появлении интегратора и его постоянной работе понемногу меняется рынок, что может существенно изменить ситуацию в использовании e-магазинов. Каждый из «игроков» может что-либо приобрести или потерять. Напомним еще раз как интегратор влияет на остальных «игроков» и какие перспективы открываются ему самому.

Покупатель в основном может выиграть в следующих вопросах:

- Один или несколько ресурсов, которые можно легко запомнить или найти и которые обеспечивают доставку интегрированной информации;
- Существенно экономится время на просмотр предложений и изучение их;
- Великолепный выбор товаров;
- Единый интерфейс пользователя, не надо овладевать различными интерфейсами других магазинов;
- Не надо искать новые e-магазины, они автоматически будут присоединены сами;
- Легче принять решение;

- Цены в магазинах выравниваются;
- Повышается качество услуг расчета, доставки, гарантий;
- Улучшается качество информации.

Покупатель рискует, что

- Визуальный интерфейс может быть хуже, чем в каком-либо конкретном магазине;
- Часть магазинов не будет присоединена к единому сервису;
- Часть товаров не будет включена в интегрированную базу данных;
- Количество информации может иногда утомить и затруднить принятие решения;
- Порядок поиска результата явно или скрыто определяет интегратор;
- Интегратор решает какую информацию показывать, и она может быть неполной и искаженной;
- Могут быть не видны специальные предложения, которые специфичны для каждого магазина (скидки, продажа комплектов и т.д.).

Неоднозначно для покупателя могут быть реализованы следующие вопросы:

- Количество рекламы и ее навязчивость;
- Интерфейс и его конфигурация в зависимости от желаний пользователя;
- Возможности поиска и фильтрации;
- Качество интеграции информации и ее количество;
- Группировка и классификация товаров;
- В результатах слишком много продавцов и информации о них.

Е-магазины в основном могут получить следующую пользу:

- У магазина дополнительная реклама;
- Небольшие магазины могут обойтись без собственной серьезной поддержки своего магазина, если интегратор предлагает обеспечение минимальной функциональности;

Е-магазин рискует, что:

- Его домашнюю страницу будут посещать только для оформления покупок, а не для поиска товаров. Нельзя организовать свою стратегию для приманивания покупателя и ознакомления его с другими товарами;
- Уменьшается количество посещений магазина. И это уменьшает интерес рекламодателей об этом магазине;
- Интегратор может диктовать свои условия игры. В качестве примера можно упомянуть: обязательную и регулярную доставку информации в определенном формате по определенному коммуникационному каналу, плату за использование услуг интегратора, т.е. абонентную плату за то, что интегратор вас не игнорирует.

Неоднозначно для магазина могут быть реализованы следующие вопросы, которые зависят от целей магазина и от прежней работы магазина:

- Уменьшаются возможности нечестной конкуренции;

- Интегратор сообщает другую информацию о конкретном товаре (информация может быть более качественная, подробная, точная, но «ошибки ради» может быть указана другая цена);
- В домашней странице интегратора другая функциональность, которая может быть как богаче, так и беднее, чем в системе e-магазина.

О выигрышах интегратора не хочется много говорить, т.к. их основу составляют политика деятельности, честность и интересы того, кого он представляет. Все таки отметим общие выигрыши:

- У него большое количество информации, которую можно превратить в деньги;
- Много посетителей, что дает возможность привлечь рекламодателей;
- Возможность предлагать своим клиентам другие сервисы.

И все таки хлеб интегратора имеет свою корочку или можно даже остаться совсем без хлеба:

- Очень рискованный бизнес, могут быть потери, наверняка будут много конкурентов, которые могут быть и за границей;
- Большие начальные капиталовложения, если ты не первый и у тебя нет поддержки от игроков на рынке;
- Большие расходы на содержание, если не удастся навязать свои условия игры (стандартизация и использование общего сервиса) остальным игрокам или создать очень интеллектуальное программное обеспечение;
- E-магазины могут не давать информацию и сопротивляться ее приобретению (если магазины чувствуют угрозу для себя), пока интегратор не стал популярным;
- Могут появиться недоброжелатели, которые будут применять экономическое или другого характера давление, т.к. раздел рынка редко происходит миролюбиво;
- Могут появиться проблемы с интеллектуальной собственностью, которая зависит от принципов работы интегратора и местного законодательства.

Выводы

Интегрированные базы данных товаров существуют и используются. С увеличением популярности торговли в интернете, увеличивается также популярность интегрированных баз данных и их использование постоянно становится более многообразно – это будет не только сравнение цен, но будет также служить помощником в выборе конкретного товара, руководствуясь различными другими критериями и свойствами. Пользователи смогут выбирать не только какой товар покупать, но и у какого продавца, что в свою очередь приумножит общее качество интернет-магазинов и уменьшит цены.