

трехмерное R-дерево. Кроме того, TP-дерево может использоваться для эффективной обработки разных типов запроса о различных промежутках времени от прошлого до будущего. Оно может эффективно использоваться в транспортном контро-

ле и радиовещательных системах. Его внедрение в системы управления базами данных не вызывает сложности, т. к. строится на базе классического пространственного R-дерева, которое поддерживается различными коммерческими СУБД.

СПИСОК ЛИТЕРАТУРЫ

1. Guttman A. R-Trees: a Dynamic Index Structure for Spatial Searching // The ACM-SIGMOD Conference on the Management of Data. – ACM Press, New York, 1984. – P. 47–57.
2. Tao Y., Papadia D.S., Sun J. The TPR*-tree: An optimized spatio-temporal access method for predictive queries // Proc. Intern. Conf. on Very Large Data Bases. – VLDB Endowment, 2003. – P. 790–801.
3. Kollios G., Gunopulos D. Indexing Mobile Objects // The 18th ACM Symposium on Principles of Database Systems. – Philadelphia, 1999. – P. 261–272.
4. Theodoridis Y., Vazirgiannis M., Sellis T. Spatio-temporal indexing for large multimedia applications // The 3rd IEEE Intern. Conf. on Multimedia Computing and Systems. – Hiroshima, Japan, 1996. – P. 441–448.
5. Tzouramanis T., Vassilakopoulos M., Manolopoulos Y. Overlapping Linear Quad Trees: A Spatio-Temporal Access Method // The 6th Intern. Symp. on Advances in Geographic Information Systems. – Bethesda MD, 1998. – P. 1–7.
6. Sellis T., Roussopoulos N., Faloutsos C. The R+ Tree: A Dynamic Index for Multi-Dimensional Objects // In Proc. 13th Intern. Conf. on Very Large Data Bases. – Brighton, England, 1987. – P. 507–518.
7. Hadjieleftheriou M.G., Kollios V.J., Gunopulos D. Efficient indexing of spatiotemporal objects // The 8th Intern. Conf. on Extending Database Technology. – Prague, Czech Republic, 2002. – P. 251–268.
8. Pfoser D., Jensen C.S., Theodoridis Y. Novel approaches to the indexing of moving object trajectories // The 26th Intern. Conf. on Very Large Databases. – Cairo, Egypt, 2000. – P. 395–406.
9. Brinkhoff T. Generating Network-Based Moving Objects // Proc. of the 12th Intern. Conf. on Scientific and Statistical Database Management, SSDBM'00. – Berlin, Germany, 2000. – P. 253–255.

Поступила 18.03.2009 г.

УДК 004.021

РАСШИРЕНИЕ КОНТЕКСТА МОБИЛЬНОГО СЕРВИСА МАРШРУТОМ ПОЛЬЗОВАТЕЛЯ В ПЛАТФОРМЕ STREAMSPIN

Н.А. Шестаков, С.С. Jensen*

Томский политехнический университет

*Университет Ольборга (Aalborg University), Дания

E-mail: nicolai.shestakov@gmail.com

Представлена архитектура платформы мобильных сервисов StreamSpin, использующая открытые интерфейсы для интеграции мобильных сервисов, разработанных внешними провайдерами. Одной из важных особенностей является включение в контекст мобильного сервиса предсказанного маршрута пользователя. Описаны ключевые особенности реализации алгоритма предсказания маршрута пользователя.

Ключевые слова:

LBS-сервис, предсказание маршрута, предиктивная навигация, map matching, context-aware сервис.

Введение

Повсеместное внедрение беспроводных коммуникаций, таких как GPRS или WiFi, а также рост вычислительной мощности мобильных устройств постепенно приводят к росту доли мобильных пользователей в сети Интернет. В связи с этим растет и число мобильных сервисов – услуг, предоставляемых через сеть мобильным пользователям.

Сервисы, использующие информацию о местоположении пользователя, называются сервисами, основанными на местоположении (Location-Based Services, LBS), или LBS-сервисами. Актуальность таких сервисов возрастает с постоянным развитием технологий позиционирования: GSM, GPS/A-GPS,

WiFi и смешанных технологий. Согласно аналитическому докладу компании Berg Insight ожидаемый рост рынка LBS-сервисов составит 50 % в год [1].

Феномен Web 2.0 [2] определил новые концепции в создании и продвижении программных продуктов. Произошло смещение акцентов с десктоп-приложений к сервисам, доступным через сеть. Это видно на примерах развития бизнес-модели «Software-as-a-Service», вычислений в «облаке» («cloud computing») и многих других. При исследовании алгоритмов интересен вопрос, насколько применимы новые подходы и идеи в рамках новой парадигмы, как только эти работы выйдут за пределы тестовых сред и синтетических экспериментов. На-

пример, проблема конфиденциальности информации на практике может стать серьезным ограничением, которое не позволит использовать хорошие и эффективные алгоритмы просто потому, что пользователи не захотят доверять нужную информацию внешнему сервису. Поэтому создание экспериментальной, но работающей платформы, на которой можно опробовать новые идеи, является важной задачей.

Платформа StreamSpin (www.streamspin.com), разрабатываемая в Научно-исследовательском центре DAISY университета Ольборга (Дания), является экспериментальной площадкой для исследования контекстно-осведомленных (*context-aware*) мобильных сервисов [3]. Основными направлениями исследований являются интеграция мобильных и немобильных сервисов на единой платформе, интеграция социальных сетей и мобильных LBS-сервисов, изучение проблем, связанных с конфиденциальностью (*privacy*) информации, создание LBS-сервисов нового типа. Существенным нововведением, выполненным авторами, является расширение контекста сервиса информацией о текущем маршруте пользователя, что дает возможность LBS-сервисам знать, куда пользователь направляется (а не только, где он находится в настоящий момент). В данной работе описана архитектура StreamSpin, а также особенности реализации алгоритма предсказания маршрута пользователя. По результатам работы системы делаются выводы о достоинствах и недостатках выбранных проектных решений.

LBS-сервисы

Несмотря на то, что *Location-based services* (LBS-сервисы) используются в сфере мобильных коммуникаций уже несколько лет, терминология в этой области достаточно размыта. Зачастую термины *location-based service*, *location-aware service*, *location-related service* и *location service* используются как взаимозаменяемые [4]. Обзор литературных источников позволяет сделать вывод, что «*Location-based service*» употребляется чаще всего. Не удалось также найти широко используемого русскоязычного аналога; как правило, используется словосочетание «LBS-сервис», которое будет применяться далее.

Будем использовать следующее определение LBS-сервиса. LBS-сервис – это информационная услуга (*information service*), которая предоставляет информацию, созданную, скомпилированную, выбранную или отфильтрованную с учетом текущего местоположения пользователя (пользователей), местоположения других людей или мобильных объектов [4].

Многими исследователями LBS-сервисы относят к категории *context-aware services* (контекстно-осведомленные сервисы). Контекстно-осведомленный сервис можно определить как сервис, который автоматически подстраивает свое поведение (например, фильтрацию или представление информации) под один или несколько параметров,

отражающих контекст целевого объекта (контекстную информацию) [4].

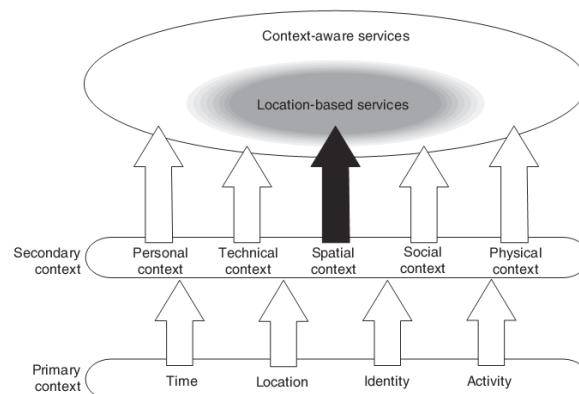


Рис. 1. Контекстно-осведомленные сервисы и LBS-сервисы

Как показано на рис. 1, контекст можно условно разделить на первичный (на рисунке: *primary context*) и вторичный (*secondary context*), который получается в результате обработки первичного [4]. Отличительная особенность LBS-сервисов – использование пространственного контекста (*spatial context*) наряду с другими типами контекста.

Расширение контекста LBS-сервисов маршрутами

Использование информации о маршруте, по которому движется пользователь, позволяет повысить качество информационных услуг, а также сделать систему более «интеллектуальной». Например, навигационная система может сообщить водителю о пробках на его пути без необходимости для водителя вручную указывать место назначения и путь [5, 6]. Информация о скидках на бензоколонке может быть сообщена заранее, а не в тот момент, когда водитель проезжает мимо на скорости 60 км/ч.

Несмотря на большое количество научных работ и патентов в области предиктивной навигации (*predictive navigation*), авторам неизвестно о существовании проектов, в которых предсказанное местоположение или текущий маршрут являлись бы частью контекста сервиса. Для того, чтобы включить маршрут в контекст, перед авторами стояли следующие задачи:

- разработать алгоритм предсказания маршрута пользователя и внедрить его на выбранной платформе;
- обеспечить сервисы информацией о текущем маршруте (через соответствующие программные интерфейсы).

Несмотря на то, что вторая задача является в большей мере технической, чем исследовательской, способ её решения может ограничить применимость тех или иных подходов, используемых для предсказания маршрута.

Основной целевой аудиторией описанных LBS-сервисов являются водители, а областью применения – навигационные приложения. В основном для определения местоположения используется

технология GPS, а устройства с включенным модулем GPS обладают повышенным энергопотреблением. Для того, чтобы накопить статистику перемещений, необходимо, чтобы модуль GPS был включен постоянно, поэтому без внешнего питания устройство быстро разряжает батарею. В автомобиле же есть возможность подключения устройств к бортовой сети.

Предсказание маршрута пользователя

Убежденность в том, что система будет способна предугадать, куда и по какому пути движется пользователь, основана на том, что большинство поездок (особенно, если речь идет о перемещениях на автомобиле) совершается по повторяющимся маршрутам. Накопив информацию о перемещениях, можно с определенной вероятностью предсказать маршрут уже в самом начале пути. В [6, 7] опубликованы результаты исследований, основанных на данных о перемещении 252 водителей на протяжении 15,1 сут. в среднем на одного водителя. В частности, по этим данным, на 15-й день наблюдений количество повторяющихся маршрутов достигает 40 %, через 40 дней – 60 %, после чего рост практически останавливается.

Для построения и хранения маршрутов используют следующие подходы:

- **Привязка местоположения к базовым станциям сотовой связи.** Обладает низкой точностью позиционирования (сотни метров), но не требует дополнительного оборудования, кроме сотового телефона. Маршрут является последовательностью идентификаторов станций (cell id).
 - **Привязка треков к графу дорожной сети** (задача «map matching») [5, 8]. Маршрут в этом случае определяется последовательностью сегментов дорожной сети (ДС).
 - **Группировка треков (траекторий) по пространственной схожести** [6]. Маршрут представляется как группа треков и усредненная по ним траектория. Не требует наличия данных ДС.
- Для предсказания маршрута можно использовать подходы:
- Зная часть пройденного пути, найти маршруты, содержащие часть этого пути и выбрать **наиболее часто используемый**.
 - Найти подходящие маршруты и выбрать с **учетом временных шаблонов** [9], сформированных заранее. Шаблон может определять время дня, день недели. Например, в воскресенье типичные маршруты отличаются от будничных, что позволяет отбросить часто используемые, но неправильные для выходного дня варианты.
 - Не предсказывать маршрут целиком, а **предсказывать несколько сегментов** [5], на основе накопленной статистики поворотов. Для этого случая подходит Марковская модель предсказания, наподобие Hidden Markov Model, предложенной в [5].
- Подходы, использующие привязку к ДС, предполагают, что пользователь является водителем транспортного средства. Для пешеходной навигации они подходят плохо, поскольку пешеходы часто перемещаются вне ДС и на них не действуют дорожные ограничения.

Архитектура платформы StreamSpin

StreamSpin – это экспериментальная платформа для внедрения инновационных технологий мобильных LBS-сервисов, позволяющая осуществлять быстрое внедрение и «доставку» LBS-сервисов до конечного пользователя. Основная идея StreamSpin – создать единый портал, подключив-

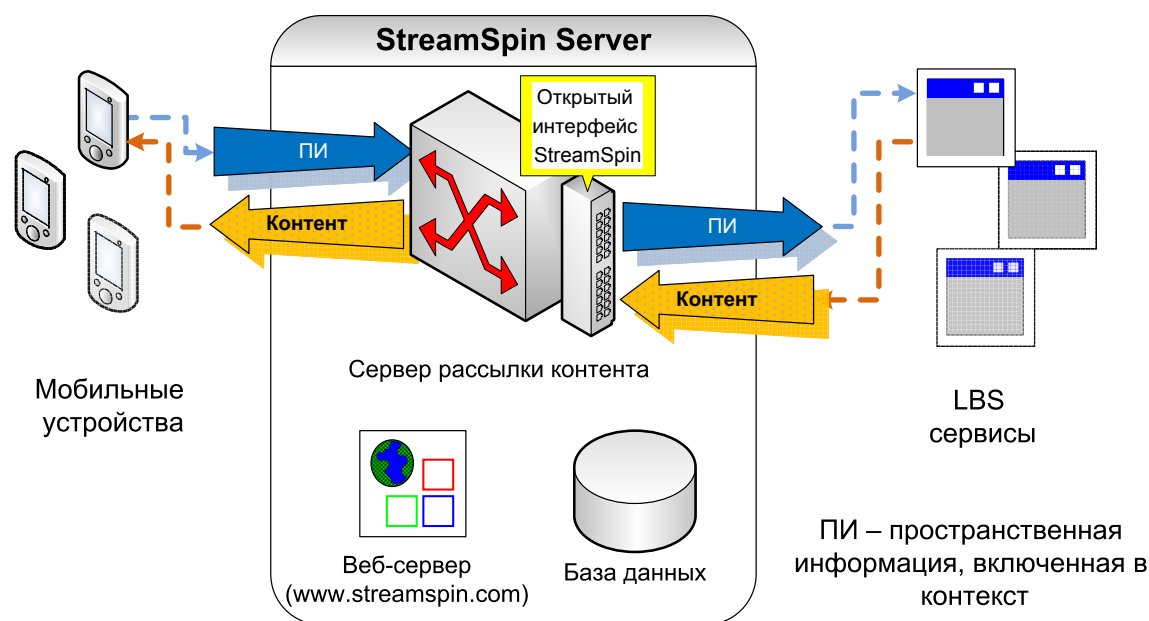


Рис. 2. Архитектура StreamSpin

шись к которому пользователь получил бы доступ к различным мобильным сервисам. При этом сервисы не реализуются внутри самой платформы: открытый интерфейс позволяет подключить сервисы, созданные сторонними разработчиками.

Структурно платформа StreamSpin может быть разделена на три составляющие (рис. 2):

- сервер StreamSpin;
- мобильные клиенты;
- внешние LBS-сервисы.

Сервер StreamSpin – главный компонент платформы. Он включает в себя:

- сервер рассылки контента (Content Distribution Server);
- веб-сервер;
- базу данных;
- открытый интерфейс (StreamSpin Open API), предоставляющий доступ к серверу со стороны внешних сервисов.

Сервер рассылки контента пересылает пространственную информацию от пользователей к сервисам и полученный от сервисов контент обратно пользователям (используется механизм подписки).

Веб-сервер обеспечивает функционирование сайта, дает возможность пользователям регистрироваться, настраивать свои учетные записи, подписываться на сервисы и отписываться от них, вести список «друзей» и пр.

Все компоненты сервера StreamSpin пользуются единой базой данных для хранения информации о пользователях, сервисах, подписках и т. п.

Взаимодействие сервисов и сервера StreamSpin осуществляется через открытый интерфейс, реализованный в виде Web-сервисов SOAP.

Мобильные клиенты – это наладонные компьютеры (PDA) с установленным приложением StreamSpin Mobile Client, которое осуществляет взаимодействие с сервером StreamSpin. Для работы приложения требуется операционная система Windows Mobile 5.0 или 6.0 и подключение к сети Интернет. Желательно наличие приемника GPS. При его отсутствии есть возможность задавать свое местоположение вручную, но предсказание маршрутов в этом случае работать не будет. Мобильный клиент передает на сервер пространственную информацию и иную информацию о пользователе, и отображает контент, полученный от LBS-сервисов.

Пространственная информация представляет не только информацию о текущем местоположении, но также о маршруте движения пользователя.

Внешние LBS-сервисы регистрируются на сервере и, если на них подписаны пользователи, то они могут отсылать контент пользователям и получать информацию о перемещениях пользователей и предполагаемом маршруте движения.

Маршрут в качестве контекста: реализация в StreamSpin

Основные идеи алгоритма предсказания маршрута взяты из работы А. Брилинггайте [8, 9]. В качестве технологии позиционирования выбрана технология GPS. Траектория движения пользователя (трек) в реальном времени отображается на графе ДС. Как правило, маршрут не имеет точного начала и конца: во время холодного запуска приемника GPS возможны существенные ошибки значений координат, места парковки могут быть разные. Чтобы дисперсия конечных точек треков не породила множество маршрутов, вводится понятие местоположения (*location*), которое занимает некоторую область на карте (например, круг радиусом 150...200 м). Координаты, попадающие внутрь местоположения, не учитываются, маршрут начинает строиться, когда пользователь покинет его пределы. Таким образом, маршрут включает начальное местоположение, конечное местоположение и путь между ними, определенный на графе ДС. Сам алгоритм можно разделить на две функциональные части: привязка текущего GPS трека к графу ДС (в результате получается некоторый путь на графе) и определение, к какому из сохраненных маршрутов можно отнести текущий путь (т. е., непосредственно предсказание). Проблема привязки координат к элементам ДС (*map matching*) нетривиальна, поскольку координаты местоположения неточно ложатся на геометрию ДС. В [6] и [8] описаны разные случаи, когда простые способы привязки к ближайшему элементу не работают. Для предсказания маршрута Брилинггайте [9] предлагает два метода: тривиальный, основанный на частоте использования, и метод с использованием временных шаблонов наподобие «каждый будний день утром», «по выходным» или «каждый вечер в пятницу».

Первоначально алгоритм был реализован Брилинггайте на сервере Oracle с поддержкой Spatial Cartridge и средств для работы с системой линейной привязки (использовались языки Java и PL/SQL). Численные эксперименты были поставлены в искусственной тестовой среде. В отличие от реализации алгоритма, выполненной Брилинггайте, StreamSpin является работающим прототипом реальной системы. В процессе воплощения в StreamSpin алгоритм претерпел существенные изменения. Эти изменения, и другие особенности реализации перечислены ниже.

1. Запись перемещений, хранение и предсказание маршрутов было перенесено на сторону клиента (т. е., приложения, запускаемого на мобильном устройстве пользователя). Основная причина этого – многие пользователи не желают хранить конфиденциальную информацию в сети. Вариант, когда вся статистика остается внутри их мобильного устройства, их устраивает больше. Даже пользователи, готовые доверить статистику своих перемещений сервису в сети

- Интернет, предпочитают на всякий случай иметь возможность переключиться в режим сохранения информации на клиенте. Т. о. для предсказания используется только персональная статистика перемещений пользователя, данные о перемещении других пользователей недоступны.
2. Улучшена процедура привязки в областях перекрестков, а также в областях близкого расположения элементов ДС друг к другу, поскольку алгоритм Брилингайте выдавал неверный результат на некоторых треках.
 3. Для хранения ДС, местоположений и маршрутов на мобильном устройстве была использована БД под управлением Microsoft SQL Server Compact Edition. Переход от серверной платформы Oracle на мобильную версию SQL Server обусловил основные различия в реализациях. Отсутствие поддержки системы линейной привязки привело к изменениям в методах доступа к данным ДС и усложнило написание пространственных запросов. Для повышения скорости выполнения запросов сегменты ДС были проиндексированы методом сетки.
 4. Была реализована поддержка программного интерфейса для включения маршрутов в контекст сервиса. Оповещения об изменении маршрута могут вызываться при условиях:
 - изменилось место назначения;
 - изменился маршрут, но место назначения осталось прежним;
 - маршрут остался прежним, но изменилась оценка времени.
 5. Была реализована поддержка нескольких ДС (для нескольких городов), задание параметров преобразования картографических проекций и импорт данных ДС из форматов *.shp и osm.xml.
 6. Поскольку выбранное проектное решение не подразумевает хранение ДС на сервере, то маршрут от клиента к сервису передается в виде последовательности географических координат, привязанных ко времени. Для хранения маршрутов в локальной БД на клиенте используется привязка к ДС с заданием координат в системе линейной привязки, т. е. сегмент маршрута задается ссылкой на сегмент ДС и начальной/конечной отметкой на сегменте ДС.
 7. В качестве алгоритма предсказания был использован тривиальный алгоритм (поиск наиболее часто используемого маршрута).

Апробация и планируемые исследования

Тестирование показало, что типичный коммуникатор под управлением Windows Mobile 5.0

(64 Мб оперативной памяти, встроенный модуль GPS) в состоянии строить и предсказывать маршруты в режиме реального времени. Задержки в процессе привязки к ДС возникают при пересечении сложных для алгоритма участков: перекрестков с большой плотностью сегментов ДС на единицу площади. Также возникают задержки, когда координаты с устройства GPS приходят с ошибками, или ошибки вызваны неточными данными ДС. При этом остается потенциал для оптимизации кода, поскольку разработка была нацелена на быстрое создание работающего прототипа.

Пожалуй, недостатком реализации является тот факт, что маршрут, при передаче его сервису, теряет привязку к ДС (фактически передается траектория движения). Для того, чтобы передавать маршруты, определенные на графе ДС, необходимо, чтобы ДС была свободно доступна с сервера StreamSpin, иначе копии ДС на стороне клиента и сервиса невозможно будет синхронизовать. Проблемой является условие использования данных ДС: цифровая карта Дании, использованная в StreamSpin, была предоставлена университету Ольборга для научных исследований без права свободного распространения, соответственно доступ к ней не может быть открыт для внешних сервисов. Возможным выходом из ситуации является использование внешних открытых источников данных ДС, например, OpenStreetMap [10].

В дальнейшем планируется улучшить алгоритм привязки, использовать временные шаблоны для предсказания маршрута. Также планируется сделать серверный вариант алгоритма предсказания.

Заключение

Взятый за основу алгоритм предсказания маршрута пользователя, описанный Брилингайте, был модифицирован и реализован в системе StreamSpin. При разработке алгоритм был существенно переделан: его пришлось адаптировать под мобильную платформу с ограниченными вычислительными ресурсами и под архитектуру платформы StreamSpin. Контекст мобильных сервисов был расширен информацией о текущем маршруте пользователя. Это значит, что любой сервис, подключенный к StreamSpin, может с согласия пользователя использовать информацию о его будущем местоположении. Авторам неизвестно о подобном использовании маршрутов в каком-либо из существующих мобильных сервисов. Для этого была модифицирована серверная часть и программный интерфейс StreamSpin. Реализованный алгоритм предсказания показал работоспособность на выбранной аппаратной и программной платформе.

СПИСОК ЛИТЕРАТУРЫ

1. Berg Insight: Strategic Analysis of the European Mobile LBS Market (Report in LBS Research Series) [Электронный ресурс]. – Режим доступа: http://www.berginsight.com/ShowReport.aspx?m_m=3&id=44. – 20.04.2009.
2. O'Reilly T. What Is Web 2.0 [Электронный ресурс]. – Режим доступа: <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>. – 20.04.2009.
3. Wind R., Jensen C., Pedersen K., Torp K. A Testbed for the Exploration of Novel Concepts in Mobile Service Delivery // Proc. of Mobile Data Management Int. Conf. – Mannheim, Germany, 2007. – P. 218–220.
4. Kupper A. Location-Based Services: Fundamentals and Operation. – Chichester: John Wiley & Sons Ltd, 2005. – P. 365.
5. Simmons R. et al. Learning to Predict Driver Route and Destination Intent. // Proc. of IEEE Intelligent Transportation Systems Conf. – Toronto, 17-20 Sept., 2006. – P. 127–132.
6. Froehlich J., Krumm J. Route Prediction from Trip Observations. // Society of Automotive Engineers World Congress. Paper 2008-01-0201. – Detroit, 22 April, 2008. – P. 103–117.
7. Froehlich J., Krumm J. The Microsoft Multiperson Location Survey. MSR-TR-2005-103 [Электронный ресурс]. – Режим доступа: <ftp://ftp.research.microsoft.com/pub/tr/TR-2005-103.doc>. – 20.04.2009.
8. Brilingaite A., Jensen C. Enabling Routes of Road Network Constrained Movements as Mobile Service Context // Geoinformatica. – 2007. – V. 11. – № 1. – P. 55–102.
9. Brilingaite A., Jensen C. Online Route Prediction for Automotive Applications // Proc. of the 13th World Congress and Exhibition on Intelligent Transport Systems and Services. – London, October, 2006. – P. 168–175.
10. Welcome to OpenStreetMap [Электронный ресурс]. – Режим доступа: http://wiki.openstreetmap.org/wiki/Main_Page. – 20.04.2009.

Поступила 21.04.2009 г.

УДК 004.89

ИСПОЛЬЗОВАНИЕ КРИ, ТЕХНОЛОГИЙ OLAP И DATA-MINING ПРИ ОБРАБОТКЕ ДАННЫХ

А.Р. Вахитов

Томский политехнический университет
E-mail: var-sasha@tpu.ru

Рассматривается способ обработки данных, основанный на совместном использовании аналитической обработки в реальном времени, а также ключевых индикаторов производительности и технологии извлечения данных. Обсуждаются принципы реализации способа, области применения, базовые термины, а также преимущества по сравнению с классическими способами решения подобных задач. Особое внимание уделяется практическому применению данного подхода в предметной области, связанной с НИРС в вузе.

Ключевые слова:

OLAP, обработка данных, data mining, ключевые индикаторы производительности.

В современном мире особую ценность приобретают эффективные способы обработки информации. Базы данных (БД), а также системы управления этими базами (СУБД) стали необходимыми в любой организации. Учебные заведения, банки, страховые, коммерческие и прочие компании собирают и хранят в своих базах гигабайты информации о сотрудниках, предоставляемых услугах, товарах и т. д. Ценность подобных сведений несомненна: они используются в различных целях (управление материально-техническими запасами, решение вопросов, связанных с перераспределением полномочий, отслеживание тенденций развития организации и другое).

Подобные БД называют операционными или транзакционными, поскольку они характеризуются огромным количеством небольших транзакций (операций записи-чтения). Компьютерные системы, осуществляющие учет операций и, собственно, доступ к транзакционным базам, принято называть системами оперативной обработки транзакций Online Transactional Processing (OLTP) или учетными системами [1].

Учетные системы настраиваются и оптимизируются для выполнения максимального количества транзакций за максимально короткое время. Показателем эффективности является количество транзакций, выполняемых за секунду. Обычно операции над отдельными записями очень просты и не связаны друг с другом. Однако совокупности записей можно использовать для получения качественно новой информации, а именно для создания отчетов и анализа деятельности организации.

Набор аналитических функций в учетных системах обычно весьма ограничен. Схемы, используемые в OLTP-приложениях, осложняют создание даже простых отчетов, так как данные чаще всего распределены по множеству таблиц, и для их агрегирования необходимо выполнять сложные операции объединения. Как правило, попытки создания комплексных отчетов требуют больших вычислительных мощностей и приводят к потере производительности [1].

Уместно также отметить, что в учетных системах хранятся постоянно изменяющиеся данные.