УДК 004.652:004.932:004.056.5

А.Б. Кунгурцев канд. техн. наук, проф.,
Мунзер Аль Абдо, магистр,
Я.В. Поточняк, магистр,
Одес. нац. политехн. ун-т

ТЕХНОЛОГИЯ ЗАМЕНЫ ИЗОБРАЖЕНИЙ В МОДЕЛИ БАЗЫ ДАННЫХ ДЛЯ СОХРАНЕНИЯ КОНФИДЕНЦИАЛЬНОСТИ ДАННЫХ

О.Б. Кунгурцев, Мунзер Аль Абдо, Я.В. Поточняк. **Технологія заміни зображень в моделі бази даних для збереження конфіденціальності даних.** Розглядаються питання заміни зображень, які носять конфіденційний характер, перед розміщенням їх в імітаційній моделі. Визначено параметри оригінального зображення, які необхідно зберегти в замінюючому зображенні. Представлено алгоритм підготовки даних для моделі.

Ключові слова: реляційна база даних, оригінальне зображення, замінююче зображення, предметна область.

А.Б. Кунгурцев, Мунзер Аль Абдо, Я.В. Поточняк. **Технология замены изображений в модели базы данных для сохранения конфиденциальности данных.** Рассматриваются вопросы замены изображений, которые носят конфиденциальный характер, перед размещением их в имитационной модели. Определены параметры оригинального изображения, которые необходимо сохранить в заменяющем изображении. Представлен алгоритм подготовки данных для модели.

Ключевые слова: реляционная база данных, оригинальное изображение, заменяющее изображение, предметная область.

O.B. Kungurtsev, Al Abdo Munzer, I.V. Potochniak. **Image replacement technology in the database model for keeping data confidential.** The problems of replacement of images that are confidential before placing them in a simulation model are considered. The parameters of the original image, which are to be saved in the replacement image, are determined. An algorithm to generate data for the model is presented.

Keywords: relational database, original image, replacing image, object domain area.

Применение имитационных моделей реляционных баз данных (РБД) позволяет получить числовые оценки многих программных и структурных изменений, направленных на увеличение производительности или функциональности информационных систем: реструктуризация, настройка ключей, введение материализованных представлений, применение репликации данных и т.д. Известны способы построения моделей РБД [1]. Поскольку процесс моделирования обычно требует копирования данных из РБД в модель РБД (МРБД), то часто возникает проблема сохранения конфиденциальности данных, используемых в модели. Известны методы решения этой проблемы для числовых и текстовых данных [2,3]. Вопрос представления изображений в МРБД пока остается открытым.

Для формулирования требований к представлению изображений в модели необходимо проанализировать способы хранения изображений в наиболее распространенных системах управления базами данных (СУБД).

Большинство СУБД имеет возможность хранить изображения, а некоторые могут с ним работать [4]. Часто в БД хранится не само изображение, а ссылка на файл в файловой системе (ФС). Например, СУБД PostgresQL имеет специальный тип данных для ссылки на файловую систему [4]. Недостаток такого решения — необходимость удаления файла в ФС при удалении соответствующей записи из таблицы БД.

Во многих современных СУБД изображение хранится в виде бинарного потока (Binary Large Object — BLOB) [4]. Обычно поддерживаются следующие операции над BLOB данными:

ISSN 2076-2429 (print)
ISSN 2223-3814 (on line)

чтение из BLOB, вставка новой строки в BLOB, обновление данных в BLOB, удаление BLOB. Чаще всего BLOB используется для хранения целевых файлов — распространенный тип систем документооборота. При таком подходе весьма сложно синхронизировать ФС с БД: удалив запись, надо удалить и файл. В БД это решается каскадным удалением соответствующей записи из BLOB.

В СУБД Microsoft Access для сохранения изображения можно использовать тип данных объект (object linking and embedding — OLE). С точки зрения программного доступа, он напоминает BLOB, но если BLOB может хранить любые данные, то OLE сохраняет изображение собственного типа [4].

Для сохранения конфиденциальности предлагается использовать вместо оригинального изображения (ОИ) из РБД заменяющее изображение (ЗИ) в МРБД. Представлен алгоритм подготовки данных для МРБД. Определен набор требований к МРБД и ЗИ, которые позволяют сохранить конфиденциальность информации и обеспечивают адекватность модели оригиналу:

— ЗИ должно сохранять формат ОИ (GIF, PNG, JPEG и т.д.); в противном случае процедуры обработки изображений в РБД и МРБД могут различаться;

— предметная область (ПО) ЗИ и ОИ должны совпадать (портреты людей, схемы сооружений, карты местности, фотографии животных и др.); в противном случае ЗИ может не соответствовать названию поля таблицы, которому оно принадлежит, или контексту, в котором оно будет использовано при испытании модели, что позволит злоумышленнику, получившему доступ к данным модели, определить факт подмены ОИ;

— геометрические размеры ЗИ и ОИ должны быть идентичны; в противном случае процедуры просмотра одного изображения или в сочетании с другими данными будут различаться в РБД и МРБД;

— в РБД и МРБД должен быть сохранен способ доступа к изображению; в противном случае необходимо менять текст запросов для работы с ЗИ в МРБД по сравнению с РБД, что может нарушить адекватность поведения модели и оригинала и повлиять на оценки производительности;

— ЗИ должны занимать столько же места в памяти, как и соответствующие ОИ; в противном случае могут быть искажены значения производительности, получаемые при моделировании.

Сформулированные требования показывают, что известные способы шифрования (DES, AES*)* [5] для решения поставленной задачи неприменимы ввиду следующих обстоятельств:

— не содержат режимы испытания модели, для которых требуется восстановление ОИ на основе ЗИ, поэтому нет необходимости в ключах шифрования и дешифрования;

— вид зашифрованного изображения всегда указывает на примененную к нему процедуру шифрования;

— ряд алгоритмов шифрования может изменять характеристики шифруемого изображения, упомянутые в последних трех пунктах предлагаемого набора требований.

Для представления моделируемых данных следует исходить из отражения в модели сформулированных требований к ЗИ.

Пусть $M_t = \{t_i\}\overline{i=1,n}$ — множество всех таблиц t_i РБД.

Обозначим M_g группу таблиц $M_g \in M_t$, которая будет представлена в модели. Таким образом каждая таблица $t_{gj} \in M_g$, будет преобразована в таблицу $t_{mj} \in M_{gm}$, где M_{gm} — группа таблиц МРБД. Пусть получено множество полей из M_g, содержащих изображения, — $P = \{p_l \mid p_l \in t_{gj}\}\overline{l=1,q}$.

В соответствии с определением ПО можно считать, что все изображения, относящиеся к одному полю некоторой таблицы, относятся также к одной ПО, поскольку объединены названием и семантикой заголовка поля. Тогда значение q определяет количество различных предметных областей для ОИ и соответственно для ЗИ.

Если конфиденциальность определяется не собственно ОИ, а ОИ в контексте определенной записи в таблице, в которой оно приведено, то предлагается простейший вариант кодиро-

вания таблицы, заключающийся в перераспределении изображений в пределах столбца в случайном порядке.

Во всех остальных случаях необходимо формировать ЗИ.

Предлагается поле p_l для любой t_{gj} таблицы, содержащее ОИ, представить в виде

$$p_l = <pr_l, f_l, s_i, m_l, r_l, a_l>, \quad (1)$$

где pr_l — предметная область изображений поля;

f_l — формат изображений;

s_i — количество записей в таблице t_{gj};

m_l — размер памяти, занимаемый изображением;

r_l — размер изображений;

a_l — способ доступа к изображению.

Обычно параметры изображения m_l и r_l в пределах столбца неизменны, поскольку отдельные значения используются в одинаковых контекстах. Если это не так, то в качестве m_l и r_l можно использовать усредненные значения данного столбца.

Выражение (1) определяет полный набор требований к ЗИ для формирования соответствующего столбца в модели.

Возможны несколько способов создания множества ЗИ для ПО pr_l:

— выбор ЗИ из заранее составленных библиотек, являющихся приложением моделирующей системы;

— составление множества ЗИ из открытых материалов организации, для которой производится моделирование РБД, например, из рекламных материалов;

— формирование множества ЗИ в соответствии с конкретным заказом организации, для которой производится моделирование РБД, например, используя ресурсы Интернета.

Введем понятие библиотеки ЗИ — L_z, содержащей множество разделов,

$$M_p = \{m_{pj}\}, j = \overline{1, m},$$

где $m_{pj} =< p_j, m_{zj} >$, — раздел;

p_j — ПО раздела;

m_{zj} — множество ЗИ раздела m_{pj}.

Тогда для замены ОИ на ЗИ в некотором поле f таблицы t_{gj} необходимо получить из L_z множество ЗИ, отвечающее условию

$$M_{zfi} = \{z_k \mid z_k \in m_{pj} \wedge pr_i = p_j\}, k = \overline{1, s_i},$$

где z_k — некоторое ЗИ из L_z.

Перед размещением ЗИ в МРБД нужно определить его формат, что станет возможно при анализе ссылок, хранимых в таблице РБД.

Предлагается все изображения, которые хранятся в библиотеке моделирующей программы, представить в наиболее распространенном формате JPEG.

После определения ПО pr_l поля p_l и нахождения соответствующего множества ЗИ каждое изображение ЗИ преобразуется в формат соответствующего ОИ из РБД и размещается в МРБД.

В РБД могут храниться изображения разных размеров. Для сохранения адекватности ОИ и ЗИ необходимо сохранять размеры r_l ОИ в соответствующих ЗИ.

При работе с размером необходимо выполнять операции растяжения и сжатия изображения. Учитывая, что при увеличении изображений теряется резкость, а при уменьшении — детализация, рекомендуется изменять площади изображений в пределах 3…0,3.

Из (1) следует, что количество ОИ s_i для каждой ПО pr_l РБД определяется как количество записей соответствующей таблицы t_i. В оригинальной таблице могут содержаться тысячи записей. Создание библиотек ЗИ с таким количеством элементов для каждой ПО не представляется

ISSN 2076-2429 (print)
ISSN 2223-3814 (on line)

возможным. Решение проблемы — использование операций модификаций для множества ЗИ конкретной ПО. К возможным операциям модификации можно отнести:

— кадрирование;

— изменение ориентации;

— изменение цвета;

— инвертирование цветов (всех и некоторых);

— сжатие или растяжение по осям;

— зеркальное отражение.

Приведенные варианты манипуляции с изображениями реализованы в специальных графических пакетах для различных сред программирования — Java, C++, C# и др. [6].

Технология замены изображений представлена алгоритмом:

1. Определить множество таблиц M_{tm} РБД, которые будут моделироваться.

2. Выделить множества P полей p_l, содержащих ОИ, которые принадлежат выделенному множеству таблиц M_{tm}.

3. Удалить из множества P поля, которые не содержат конфиденциальной информации.

4. Выполнить случайное перераспределение ОИ в пределах столбца таблицы, для сохранения конфиденциальности; произвести соответствующее сокращение элементов множества P.

5. Определить предметную область pr_l каждого элемента множества P. Сформировать множество M_{lz} ЗИ с использованием библиотеки ЗИ L_z.

6. Если $M_{lz}\neq 0$, то пополнить библиотеку изображениями ЗИ из предметной области pr_l одним из предложенных способов.

7. Если количество изображений в M_{lz} недостаточно, то применить к элементам M_{lz} модификацию одним из предложенных способов.

8. Переформатировать изображения из M_{zi} в соответствии с форматом f_l.

9. Разместить ЗИ (ссылки на изображения) из M_{lz} в соответствующее поле p_l таблицы МРБД.

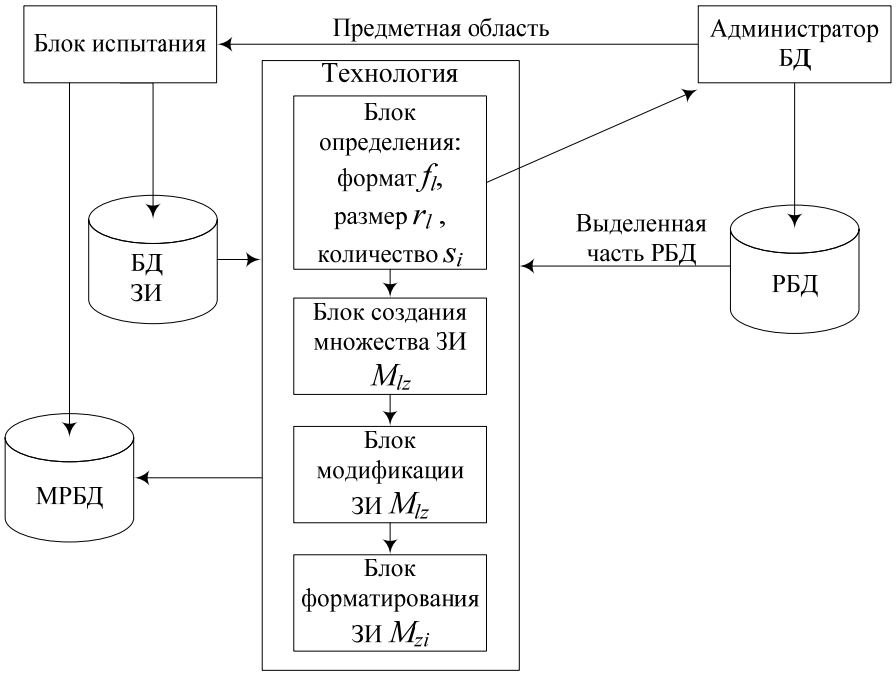
10. Повторить пункты 4…9 для каждого поля p_l из множества P.

На рисунке приведена структура программного обеспечения, реализующего предложенную технологию замены изображений в МРБД (см. рисунок). Получив сведения о ПО изображений, испытатель модели передает в распоряжение Администратора БД базу данных ЗИ и программное обеспечение для замены изображений. Формат f_l, размер r_l и количество s_i ЗИ определяются автоматически в соответствии с выделенной частью РБД. Выделенная часть РБД — основа для построения МРБД, в ней следует заменить изображения. После замены изображений МРБД передается для испытания.

Существуют различные тематики изображений, которые будут использоваться для создания БД ЗИ. Подбор ЗИ выполняется “ручным” способом. Проведенные эксперименты позволяют оценить трудоемкость “ручной” работы по подбору ЗИ. Например, для ПО “Портреты сотрудников” некоторой организации: количество сотрудников — 150 человек, возраст 23…60 лет, средний возраст — 30 лет, соотношение числа мужчин/ женщин — 5/3” было подобрано в Интернете необходимое число фотографий за 3 ч, что можно считать вполне приемлемым результатом.

Из выражения (1) следует, что параметры ОИ, используемые в ЗИ, не несут никакой информации о конкретном изображении, поэтому одному ЗИ можно поставить в соответствие бесконечное множество “подходящих” ОИ. Доступ к ОИ в РБД отсутствует как в процессе подготовки МРБД, так и в процессе ее испытания. В таких условиях процесс восстановления ОИ на основе ЗИ не имеет смысла.

Предложенная технология замены изображений позволяет расширить возможности имитационного моделирования РБД за счет включения в модель таблиц, содержащих конфиденциальную информацию. При этом испытании модели невозможно восстановить оригинальные данные. Для других лиц, случайно или сознательно получивших доступ к данным МРБД, значения изображений выглядят правдоподобно и не вызывают желания получить соответствующую информацию другим способом. Одновременно в МРБД сохраняются все параметры реальной РБД, которые могут влиять на результаты испытаний.



Структура программного обеспечения

Предполагается, что программное обеспечение для моделирования будет использоваться многократно. В этом случае объем библиотеки ЗИ будет постоянно увеличиваться, а трудоемкость подготовки данных для модели соответственно уменьшаться.

Литература

1. Кунгурцев, А.Б. Имитационная модель таблиц реляционной базы данных / А.Б. Кунгурцев, С.Л. Зиноватная, Аль Абдо Мунзер // Пр. Одес. політехн. ун-ту. Одеса, 2011. — Вип. 1 (35). — С. 90 — 96.
2. Кунгурцев, А.Б. Метод шифрования строк в имитационной модели реляционной базы данных / А.Б. Кунгурцев, С.Л. Зиноватная, Аль Абдо Мунзер // Восточ.-европ. журн. передовых технологий. — 2011. — № 6/ 2 (54). — С. 42 — 45.
3. Kungurtsev, O.B. Ensuring Data Confidentiality in Relational Database Modeling / O.B. Kungurtsev, S.L. Zinovatnaya, Al Abdo Munzer // Компьтер. науки и технологии. — 2009. — С. 57 — 61.
4. Гарсиа-Молина, Г. Системы баз данных: полный курс / Г. Гарсиа-Молина, Дж. Ульман, Дж. Уидом. — М. Издат. дом “Вильямс”, 2003. — 1088 с.
5. Венбо Мао. Современная криптография. Теория и практика / Венбо Мао, М. Вильямс // — М.: Издат. дом “Вильямс”, 2005. — 768 с.
6. Кей, Х. Java 2. Библиотека профессионала / Х. Кей, Г. Корнелл. — М.: Издат. дом “Вильямс”, 2009. — 815 с.

References

1. Kungurtsev, A.B. Imitatsionnaya model tablits relyatsionnoy bazy dannykh [Simulation model of relational database tables] / A.B. Kungurtsev, S.L. Zinovatnaya, Al Abdo Munzer // Pr. Od. politekh. un-tu [Proc. of Odesa Polytech. Univ.], 2011. — Issue 1 (35). — pp. 90 — 96.
2. Kungurtsev, A.B. Metod shifrovaniya strok v imitatsionnoy modeli relyatsionnoy bazy dannykh [The encryption method of lines in a simulation model of a relational database] / A.B. Kungurtsev, S.L. Zinovatnaya, Al Abdo Munzer // Vostochno-evropeyskiy zhurnal peredovykh tekhnologiy [East-European Journal of Advanced Technologies]. — 2011. — # 6/ 2 (54). — pp. 42 — 45.
3. Kungurtsev, A.B. [Ensuring Data Confidentiality in Relational Database Modeling] / A.B. Kungurtsev, S.L. Zinovatnaya, Al Abdo Munzer // Komp’yuter. nauki I tekhnologii [Computer Sciences and Technologies]. — 2009. — pp. 57 — 61.

ISSN 2076-2429 (print)
ISSN 2223-3814 (on line)

4. Garsia-Molina G. Sistemy baz dannykh: polnyy kurs [Database Systems: a Complete course] / G. Garsia-Molina, Dzh. Ulman, Dzh. Uidom. — Moscow, 2003. — 1088 pp.
5. Venbo Mao. Sovremennaya kriptografiya. Teoriya I praktika [Modern cryptography. Theory and practice] / Venbo Mao. — Moscow, 2005. — 768 pp.
6. Key, Kh. Java 2. Biblioteka professionala [Java 2. Library of a professional] / Kh. Key, G. Kornell. — Moscow, 2009. — 815 pp.

Рецензент д-р техн. наук, проф. Одес. нац. политехн. ун-та Антощук С.Г.

Поступила в редакцию 29 октября 2012 г.