

УДК 004.021

Ю.Б. Гриценко, В.Ю. Вишняков, С.С. Ощепков

## Решение проблемы обновления пространственных данных в среде Oracle Spatial

Рассматривается технология хеширования пространственных данных, позволяющая создать однозначное соответствие геометрического и атрибутивного описания объектов в прикладных ГИС универсального назначения.

**Ключевые слова:** пространственные данные, хеширование, хеш-код.

На современном этапе развития геоинформационных систем широко развивается подход к хранению пространственных данных совместно с атрибутивными в клиент-серверных СУБД. Лидером на рынке программного обеспечения, предназначенного для хранения пространственных данных, можно смело назвать СУБД Oracle Spatial.

На этапе подготовки электронных чертежей присутствует постоянный характер обновления пространственных данных. При этом измененные пространственные данные записываются в базу данных Oracle Spatial.

Технология хранения пространственных данных не поддерживает механизм исключения дублирования геометрических объектов, что приводит к появлению в базе данных дублирующихся записей или объектов с одинаковой геометрией [1]. Несмотря на то, что все пространственные объекты в базе данных имеют свой уникальный код, это не препятствует появлению в хранилище объектов с одинаковой геометрией, но имеющих разные идентификаторы. Так происходит в силу того, что пространственные объекты часто копируются и вставляются, при этом встроенные механизмы геоинформационных систем, работающих с базой данных, не позволяют отследить объекты, имеющие одинаковую геометрию. Кроме того, дубликаты объектов появляются в результате неверных действий пользователей при обновлении пространственных данных. Например, объекты одного слоя могут быть дважды записаны в базу данных из файла чертежа. В таком случае загруженные из чертежа данные записываются «поверх» новых без их удаления.

Обычно процесс обновления данных в пространственном хранилище производится следующим образом:

1. Удаляются все объекты требуемого слоя из базы данных.
2. Из файла чертежа объекты вновь загружаются, включая новые или измененные.
3. После считывания все эти графические объекты записываются в БД.
4. Запускается процедура генерации атрибутивных описаний объектов (или информационных карточек) на основе графических объектов и их объектных данных. Таким образом, для каждого графического объекта создается своя собственная информационная карточка (описание), которая в дальнейшем может быть заполнена значениями свойств объекта.

Во всех современных ГИС к геометрическим объектам прикрепляются атрибутивные описания, характеризующие их свойства. Во многих ГИС атрибутивные описания хранятся непосредственно в тех же таблицах вместе с геометрией объекта. Такой подход оправдан в случае узкоспециализированной ГИС, ориентированной на картографирование и управление определенным видом инженерных сетей. В случае прикладной ГИС универсального назначения атрибутивные описания хранят отдельно от графической информации, что позволяет в любой момент времени модифицировать их структуру согласно изменениям в инфраструктуре предприятия.

Объект атрибутивной схемы данных представляет собой совокупность базовых и дополнительных свойств, характеризующих данных.

В то время как пространственные объекты удаляются, соответствующие им атрибутивные описания остаются в базе данных до тех пор, пока они явно не будут удалены администратором системы. При генерации информационных карточек каждый раз происходит их повторное создание вместо того, чтобы найти старые карточки и перепривязать к тем же самым графическим объектам. Это затруднено тем обстоятельством, что привязка графических и атрибутивных объектов осуществляется исключительно по идентификатору графического объекта, который при повторной записи такого объекта в базу

данных будет уже другим, и, соответственно, привязанная к нему информационная карточка будет «потерянной», но по-прежнему будет существовать в базе данных. Таким образом, выявляются две существенные проблемы: проблема поиска и удаления объектов, имеющих одинаковую геометрию, и проблема потери информационных карточек графических объектов при изменении у них идентификатора.

Предлагаемая авторами статьи технология хеширования пространственных данных позволяет найти и устранить дублирование одинаковых графических объектов и предотвратить генерацию избыточных атрибутивных описаний. Кроме того, «потерянные» информационные карточки всегда можно прикрепить к соответствующим графическим объектам даже при изменении у них идентификаторов.

Суть технологии состоит в следующем. Структура хранения атрибутивных и пространственных данных модифицирована таким образом, что помимо графической информации каждый пространственный объект имеет уникальный хеш-идентификатор, однозначно определяющий данный объект.

Ниже рассмотрим пространственную схему данных более подробно [2].

Обозначим множество всех графических объектов как  $\mathbf{G}$  ( $\mathbf{G} = \{g_1, g_2, \dots, g_n\}$ ). В пространственной схеме каждый графический объект состоит из множества точек  $\mathbf{P}_i$  ( $g_i \sim \mathbf{P}_i$ ,  $\mathbf{P}_i = \{p_1, p_2, \dots, p_m\}$ ).

Обозначим через  $\mathbf{F}$  – множество пространственных объектов, имеющих как минимум одно описание, а через  $\mathbf{N}$  – множество графических объектов, не имеющих атрибутивных описаний. Таким образом, все множество пространственных объектов является объединением двух множеств ( $\mathbf{G} = \mathbf{F} \cup \mathbf{N}$ ).

Обозначим множество всех атрибутивных объектов как  $\mathbf{A}$  ( $\mathbf{A} = \{a_1, a_2, \dots, a_k\}$ ). Объект атрибутивной схемы можно определить как

$$a_i = \{\mathbf{B}, h\},$$

где  $\mathbf{B} = \{b_1, b_2, \dots, b_l\}$  – множество свойств объекта;  $h$  – уникальный хеш-идентификатор.

Хеш-идентификатор представляет собой 32-байтный шестнадцатеричный код, вычисляемый алгоритмом MD5 [3] (Ronald L. Rivest, Massachusetts Institute of Technology Laboratory for Computer Science and RSA Data Security, Inc. April 1992, Request for Comments 1321), которому в качестве параметров передаются координаты всех точек, из которых состоит графический объект.

Функция хеширования совершает обход всех точек множества  $\mathbf{P}_i$ , из которых состоит объект, и по совокупному набору точек рассчитывает уникальный хеш-код  $h$ . При этом используется алгоритм MD5, входным параметром которого является множество координат объекта. Функция хеширования, примененная к множеству точек объекта, осуществляет отображение множества координат точек объекта  $\mathbf{P}_i$  в элемент  $h$ .

Хеширование обладает следующими свойствами:

1. Функция осуществляет отображение множества точек объектов в уникальных хеш-код, т.е. во множество из одного элемента:

$$f(\mathbf{N}_i) = \mathbf{H}_i,$$

где  $\mathbf{N}$  – множество графических объектов, не имеющих атрибутивных описаний;  $\mathbf{H}$  – множество уникальных хеш-идентификаторов.

2. Не существует обратной функции, способной по уникальному хеш-коду получить исходные данные, в данном случае координаты объекта:

$$\mathbf{F}^{-1}(\mathbf{H}) \neq \mathbf{N} \text{ или } \mathbf{F}^{-1}(\mathbf{F}(\mathbf{N})) \neq \mathbf{N}.$$

3. Процесс нахождения двух сообщений, которые дают два одинаковых  $h$ -значения, весьма трудоемок.

Таким образом, из третьего свойства вытекает, что если два графических объекта имеют одинаковый хеш-код, то это означает, что у них одинаковая геометрия, т.е. один из объектов является дубликатом второго. Технология использования хеш-кодов представлена на рис. 1.

Соответственно, после обновления пространственных данных, необходимо произвести выполнение процедуры, вычисляющей хеш-код, последовательно обходя все точки, из которых состоит объект. Рассчитанный по алгоритму MD5 хеш-код записывается в базу данных в ту же таблицу, где хранится геометрия объекта.

Выполнение привязки пространственных данных к атрибутивным описаниям производится по требованию администратора и осуществляется вызовом специальной сервисной

процедуры. Ниже приведен алгоритм работы процедуры привязки пространственных данных.

Для каждого графического объекта  $g_i$  необходимо выполнить шаги 1–7:

*шаг 1.* Определить, принадлежит ли графический объект множеству  $\mathbf{N}$ . Если да, то перейти к шагу 2, иначе – к шагу 7;

*шаг 2.* Получить хеш-значение  $h_i$  графического объекта и выполнить проверку наличия во множестве  $\mathbf{N}$  такого же хеш-идентификатора. Если хеш-значение  $h_i$  было найдено во множестве  $\mathbf{N}$ , перейти к шагу 3, иначе – к шагу 5;

*шаг 3.* Из множества атрибутивных объектов  $\mathbf{A}$  получить уникальный идентификатор  $ID$  объекта  $a_{id}$ , найденный по его хеш-значению;

*шаг 4.* Выполнить привязку графического объекта  $g_i$  к соответствующему атрибутивному объекту  $a_{id}$ , используя идентификатор  $ID$ , найденный на шаге 3. Перейти к шагу 7;

*шаг 5.* Сгенерировать новое атрибутивное описание (объект  $a_{id}$ );

*шаг 6.* Выполнить привязку графического объекта к сгенерированному на шаге 5 новому атрибутивному описанию;

*шаг 7.* Выход.

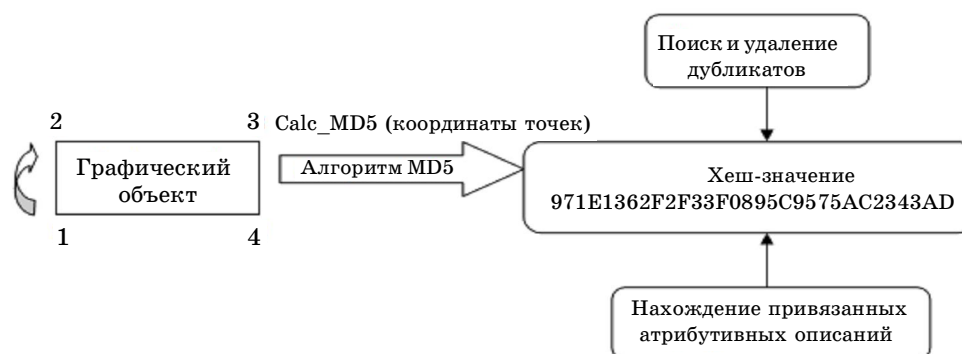


Рис. 1. Технология хеширования пространственных данных

Из множества геометрических объектов, не имеющих привязки, выполняется проверка существования атрибутивных объектов. Проверка производится по уникальному хеш-коду графического объекта. Таким образом, если во множестве атрибутивных описаний объектов находится объект, имеющий точно такой же хеш-идентификатор, то далее производится привязка этого графического объекта к его атрибутивному описанию. В противном случае для данного графического объекта вызывается функция генерации нового атрибутивного описания, после чего выполняется привязка только что созданного атрибутивного описания к графическому объекту.

Благодаря использованию хранимого хеш-значения каждого графического объекта, процедура нахождения дубликатов занимает на порядок меньше времени, чем обычные запросы на полную проверку геометрии объектов. Кроме того, благодаря двойной привязке атрибутивных описаний к графическим объектам становится возможным нахождение «потерянных» атрибутивных описаний объектов и прикрепление их к соответствующим графическим объектам.

Представленное решение проблемы обновления пространственных данных в среде СУБД Oracle Spatial легло в основу архитектуры построения электронного генерального плана ООО «Томскнефтехим», г. Томск.

#### Литература

1. Вишняков В.Ю. Проблема идентификации пространственных объектов / В.Ю. Вишняков, О.И. Жуковский, Ю.Б. Гриценко // Научная сессия ТУСУР-2007: Матер. докл. Всерос. науч.-техн. конф. студентов, аспирантов и молодых ученых. Тематический выпуск «Системная интеграция и безопасность» (Томск). – Томск: В-Спектр, 2007. – Ч. 1. – С. 347–350.

2. Vishnjakov V.Yu. The Data Model for the Topology Relations Representation in the Engineering Networks / V.Yu. Vishnjakov, Yu.B. Gritsenko, O.I. Zhukovsky // IEEE Inter-

national Siberian Conference on Control and Communications (SIBCON-2007): proceedings. – Tomsk: The Tomsk IEEE Chapter & Student Branch. Russia, Tomsk, 2007. – С. 89–93.

3. Репозиторий RFC на русском языке [Электронный ресурс]. – Режим доступа: <http://rfc.com.ru/rfc1321.htm>, свободный.

**Гриценко Юрий Борисович**

Канд. техн. наук, доцент кафедры автоматизации обработки информации ТУСУРа

Тел.: (3822) 41-44-70

Эл. почта: [ubg@muma.tusur.ru](mailto:ubg@muma.tusur.ru)

**Вишняков Вячеслав Юрьевич**

Аспирант кафедры автоматизации обработки информации ТУСУРа

Тел.: (3822) 41-44-70

Эл. почта: [slavik@gis.tusur.ru](mailto:slavik@gis.tusur.ru)

**Ощепков Сергей Сергеевич**

Аспирант кафедры автоматизации обработки информации ТУСУРа

Тел.: (3822) 41-44-70

Эл. почта: [oss@muma.tusur.ru](mailto:oss@muma.tusur.ru)

Yu.B. Gritsenko, V.Yu. Vishnjakov, S.S. Oschepkov

**Solution of the problem of spatial data updating in the Oracle Spatial development environment**

This paper considers the technology of hashing spatial data, which allows to create unambiguous conformity of the geometrical and attributive description of objects in the universal applied GIS.

**Key words:** spatial data, hashing, hash code.

---