

Об одном из способов хранения и анализа пространственно-атрибутивных данных угледобывающего предприятия

On an optional way to store and analyze spatial attributive data on a coal mining enterprise

Степанов / Stepanov Y.

Юрий Александрович

(dambo290@yandex.ru)

кандидат технических наук, доцент.

Новокузнецкий институт (филиал)

ФГБОУ ВПО «Кемеровский государственный университет» (НФИ КемГУ), доцент.

г. Новокузнецк

Бурмин / Burmin L.

Леонид Николаевич

(Lnburmin@mail.ru)

НФИ КемГУ, аспирант.

г. Новокузнецк

Ключевые слова: OLAP; ГИС – GIS; визуализация – visualization; пространственно-атрибутивные данные – spatial attributive data; базы данных – databases; компьютерные технологии в горном деле – computer technologies in mining; компьютерная трехмерная графика – 3D computer graphics; геоинформатика – geo-informatic.

Управление данными в геоинформационных системах сопряжено с рядом задач: необходимо оперировать множеством атрибутивных характеристик, синхронизировать атрибутивные характеристики с пространственными данными, а в случае с динамическим моделированием – сохранять слой временных атрибутов. Для работы с данными необходим способ эффективной структуризации, хранения и анализа информации. В статье предлагается информационная модель, включающая технологии систематизации, анализа и хранения данных на основе современных компьютерных технологий: геопространственных СУБД, технологии OLAP и средств построения компьютерных трехмерных изображений.

Managing data in geo-information systems entails a number of problems: one needs to manipulate a multitude of attributive characteristics, synchronize attributive characteristics with spatial data and in case of dynamic modeling – preserve the layer of provisional attributes. To handle such data one needs a technique to structure them efficiently, store and analyze such information. The article proposes an information model that includes technologies of systematizing, analyzing and storing data based on modern computer technologies: geospatial DBMS, OLAP technology and plotting means for 3D computer images.

Использование высокоинтенсивных технологий выемки угля напрямую связано с проблемой обеспечения стабильной и безопасной высокопроизводительной работы комплексно-механизированных забоев, которая может быть решена на основе компьютерного моделирования и прогнозирования динамики состояния углепородного массива при ведении очистных работ.

Компьютерное моделирование геомеханических ситуаций с учетом положения секции механизированной крепи по длине выемочного столба возможно на основе исследования и прогнозирования динамических процессов разрушения горных пород в окрестности очистного забоя. В связи с этим возникает необходимость в мониторинге и управлении технологическими процессами, которые порождают большие массивы данных, объем которых обусловлен несколькими составляющими [2].

Во-первых, необходимо учитывать как можно больше факторов (согласно оценке влияния каждого из них) для построения адекватной модели. Зачастую это увеличивает атрибутивную составляющую информационной модели, поскольку необходимые характеристики относятся к разным сферам работы угольного предприятия: проблемам безопасности, экономической эффективности, геологической специфики и т.д. Разнородность в структурах данных, единицах измерения и сложная соотносимость данных порождают необходимость в разработке дополнительных способов хранения и обработки данных [1].

Во-вторых, при ведении работ на шахтах угольных предприятий географическая протяженность может измеряться сотнями и даже тысячами километров, причем как в ширину, так и в глубину. Это увеличивает пространственную составляющую информационной модели. Более того, сведения о пространстве могут быть представлены в различном виде: относительные координаты и географические координаты или параметрические значения глубин и высотных отметок, из которых вычисляются пространственные характеристики [6]. От способа хранения пространственных данных зависит методика построения геометрической модели горного массива или горных выработок угледобывающего предприятия.

В-третьих, для динамического моделирования процессов выемки угля необходимо сохранять несколько

статических состояний модели, снабжая каждую часть информационной модели темпоральным атрибутом. Только на основе построенного в результате моделирования хронологического ряда представляется возможным осуществлять прогноз состояния модели. Таким образом, и без того объемный массив данных увеличивается пропорционально количеству зафиксированных статических состояний модели.

На основе вышесказанного можно построить математическую модель, описывающую в общем случае пространство данных угледобывающего предприятия U :

$$U = \{X, Y, Z, L, t\}dt$$

где $\{X, Y, Z\}$ – множество координат трехмерного пространства;

L – атрибутивная составляющая, буквально описывающая множество сведений об изучаемой предметной области;

t – хронологическая составляющая, описывающая пространство времени, которое лежит в основе мониторинга и прогнозирования данных.

Каждая компонента пространства описана непрерывной величиной. Совокупное количество элементов этого пространства зависит от диапазона изменения по каждой из компонент системы: изменяется размер изучаемого пространства, время наблюдения, а с приращением времени изменяются атрибутивные характеристики. Соответственно, меняется и объем данных, необходимый для обработки. Для того чтобы зафиксировать это пространство целесообразно провести кван-

тование – разбить «поверхность» этого пространства на дискреты, более удобные для анализа и экстраполяции. Однако даже после дискретизации объем данных для обработки все еще остается довольно внушительным, что говорит о необходимости решения задачи управления большими объемами разнородных данных. Для решения этой задачи предлагается способ манипулирования данными посредством информационной модели системы, состоящий из трех частей: систематизации, анализа и хранения. Структура информационной модели данных для реализации предлагаемого способа представлена на рис. 1.

Систематизация сведений предметной области

Перед началом работы каждой информационной системы задействуются механизмы систематизации данных – подготовки сведений для использования внутри программных модулей. Задача заключается в систематизации массива данных. Полученная в результате систематизации структура данных должна обладать достаточным уровнем абстрактности и гибкости, но при этом конкретно описывать модель внутри предметной области. В данном случае речь идет о структурировании геологических данных, полученных на основе анализа образцов горных пород. Каждая скважина обладает рядом характеристик: географические координаты, высотная отметка устья, глубина скважины. Фиксируются данные о глубине залегания и мощности каждого пласта. Существуют различные способы структури-



Рис. 1. Структура информационной модели данных для системы угледобывающего предприятия

рования этих сведений. Например, можно сохранять высотную отметку кровли и почвы каждого пласта. Либо глубинную отметку начала залегания каждого пласта и мощность. Однако поскольку в будущем необходимо осуществлять трехмерную интерполяцию данных для получения из дискретных данных (скважин) непрерывной модели углепородного массива, было принято решение хранить данные следующим образом: сохраняется высотная координата середины каждого пласта в керне и мощность пласта. Дополнительно реализовать в модели данных вычисляемые свойства, с помощью которых из мощности пласта и координаты середины рассчитывается координаты границ каждого пласта в скважине. Поскольку для разработки приложения был выбран язык программирования C# (среда разработки Mono Develop), предлагается использовать в качестве базовой структуры данных List (T) из пространства имен System.Collections.Generic. Преимущества этой структуры заключаются в скорости поиска и сортировки (по сравнению ArrayList), динамическом размере (по сравнению с обычными массивами) и большей гибкости по сравнению с хеш-таблицей. Элементом этого списка будет являться ссылочный тип – экземпляр класса «Разведочная скважина» с полями «Номер», «X», «Y», «Z устья», «Глубина» и еще одним вложенным списком List (T) «Слой керна». В свою очередь, элементами этого вложенного списка являются экземпляры класса «Элемент керна» с полями «Z центра», «Мощность», «Тип породы», «Состояние во времени» и двумя вычисляемыми свойствами – «Z почвы» и «Z кровли». Описанный способ хранения данных, с одной стороны, достаточно абстрактен и масштабируем, что позволяет расширить существующую структуру новыми атрибутами при необходимости, но при этом адекватно описывает практическую сторону предметной области. Например, добавляя атрибуты к классу «Тип породы», которые описывают геомеханическое поведение горной породы, можно оценить напряженное состояние в нетронутом угольном массиве. Это не нарушит общую структуру хранения данных, однако расширит описание модели. Во время проведения трехмерной интерполяции заполняются «ниши» между разведочными скважинами, образуются элементы с такой же структурой, но описывающие интерполированные данные. Суммарно разведочные и интерполированные скважины образуют модель углепородного массива шахтного поля. После этого возникает задача обработки и анализа полученного массива данных.

Технология анализа и обработки данных

Анализ данных всегда проводится с целью извлечения полезной информации и принятия решений. В связи с этим, прежде чем выбирать способ обработки, необходимо определить конечную цель анализа. В данном случае в качестве цели можно выбрать проектирование горных выработок и оценку напряженного

состояния в нетронутом массиве. Сложность проектирования горных выработок заключается в их пространственном расположении – очистные выработки должны находиться параллельно пласту. В случае горизонтальных выработок необходимо учитывать наличие углов по падению и по простиранию пласта для облегчения транспортировки и обеспечения стока воды к водосборнику, соответственно необходим способ для оценки пространственного позиционирования выработки.

Изучение напряженно-деформированного состояния углепородного массива необходимо для оценки безопасности ведения горных работ. Однако помимо уровня напряжения горных пород существуют и другие факторы, влияющие на безопасность: физические (температура, влажность, уровень шума и вибрации), химические (раздражающие, сенсibiliзирующие, канцерогенные), биологические (микро и микроорганизмы) и психологические. Адекватно оценить психологические и биологические факторы безопасности довольно сложно, а зачастую и вовсе невозможно. В свою очередь, физические и химические свойства вполне поддаются оценке с помощью специализированного аппаратного обеспечения. Однако после сбора всего массива данных со всех исследуемых участков очень трудно проанализировать полученные сведения в совокупности. Для этого необходимо использовать специальные методы обработки: OLAP и Data Mining.

Технология OLAP позволяет соотнести разнородные данные, предоставив пользователю полную картину безопасности на выбранном участке. Одна из проблем, возникающая при сравнении и отображении показателей различных факторов, заключается в несоотносимых единицах измерения. Для решения этой проблемы предлагается на уровне СУБД организовать механизмы перехода от абсолютных шкал к относительным. Это позволит не только оценить степень опасности каждого выемочного участка, но и выявить причины возникновения чрезвычайных ситуаций. Механизм анализа Data Mining позволяет провести кластеризацию, выделяя группы факторов, совместно влияющих на общий уровень безопасности в исследуемом секторе, а также изучить динамику изменения состояния этого сектора и составить предположительный прогноз.

Зачастую в качестве результата работы OLAP понимается только таблично-текстовая форма представления [5]. Однако визуализировать результаты OLAP возможно с помощью технологий построения анаморфозного изображения [4]. Пример такой визуализации представлен на рис. 2. На данном изображении дискреты представляют собой прямоугольные «плиты», а значение их атрибутивной характеристики представлено геометрической высотой. Каждый атрибут обозначен отдельным цветом – синим, красным или желтым (на рис. 2 цвета показаны градациями серого). Дополнительно была проведена процедура кластери-

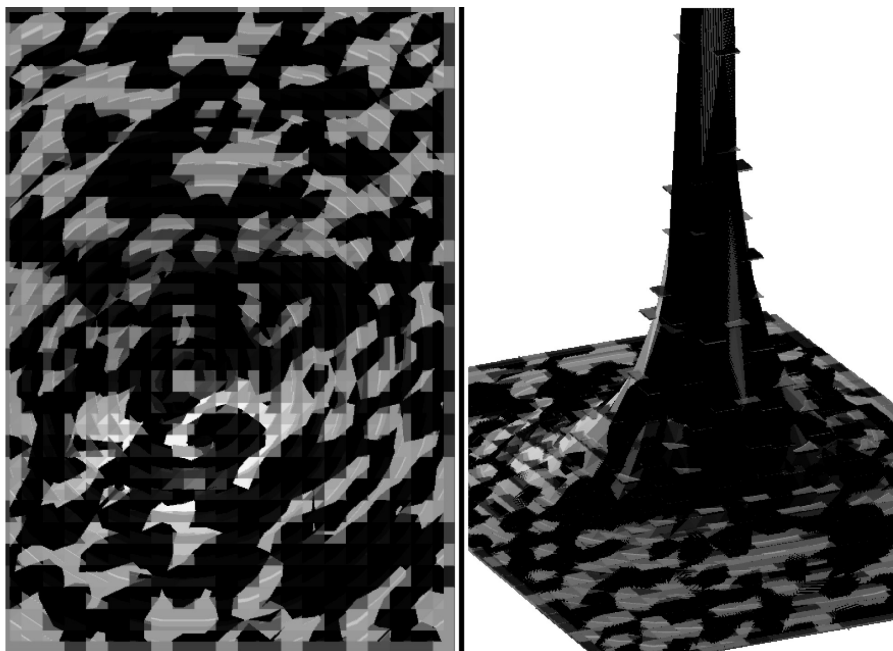


Рис. 2. Визуализация OLAP в виде анаморфозного изображения

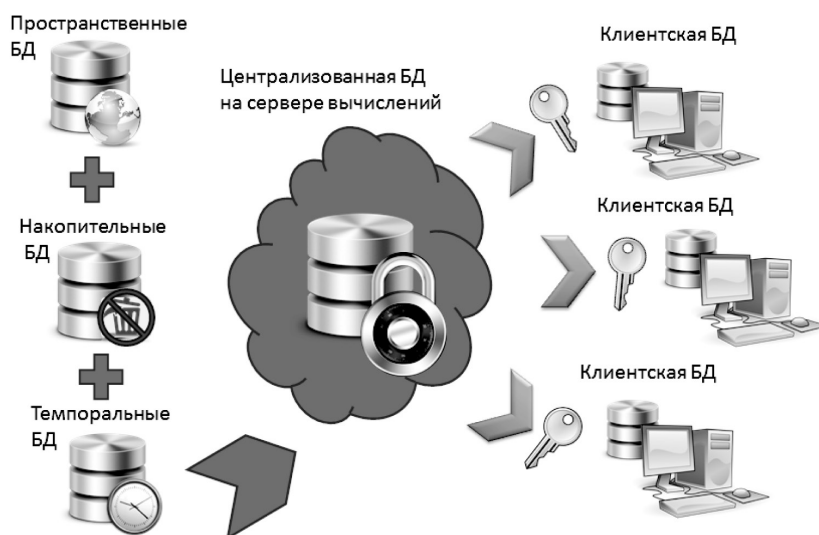


Рис. 3. Способ распределенного хранения данных

зации, с помощью которой были сформированы непрерывные зоны «доминирования» атрибута в определенном пространственном секторе.

Хранение данных посредством хранилищ и витрин данных

Очевидно, что такой большой объем данных предъявляет повышенные требования к носителям информации: объему и скорости работы жесткого диска. Следовательно, целесообразно использование аппаратного сервера с установленной СУБД и программным интерфейсом для подключения клиентских машин [3]. Один из способов распределенного хранения данных представлен на рис. 3.

На клиентской машине необходимо использование базы данных, в которой будут содержаться сведения для оперативной обработки – построения общей модели углепородного массива. В качестве серверной СУБД было принято использовать PostgreSQL. Серверная СУБД позволяет разграничить уровни доступа пользователей с помощью паролей. Еще одним плюсом PostgreSQL является наличие расширения PostGIS, которое позволяет реализовать пространственную составляющую базы данных [7]. Стоит также отметить, что база данных должна быть темпоральной (хранить хронологический атрибут) и реализовывать накопительный принцип.

На клиенте этого не требуется, следовательно, необходима база данных без серверной части, однако с более гибким программным интерфейсом для подключения и работы на разных операционных системах. Из существующих компактных встраиваемых реляционных баз данных больше всего в эту концепцию укладывается SQLite. Поскольку передаваемые данные могут представлять собой коммерческую или государственную тайну, на сервере и клиенте необходимы механизмы криптографической обработки данных.

В предложенном способе систематизации, хранения, анализа и преобразования пространственно-атрибутивных данных угледобывающего предприятия отображаемые данные по ходу работы детализируются и обобщаются в зависимости от действий пользователя. Все расчеты и анализ данных производится на сервере, а результаты обработки отправляются на клиентское приложение, задачей которого остается только преобразование полученных данных и отображение с помощью таблиц, диаграмм, изолиний на географической карте или средств трехмерной компьютерной визуализации.

Таким образом, задача систематизации данных была решена с помощью использования ссылочных структур данных, для хранения пространственно-атрибутивных данных угледобывающего предприятия было принято решение использовать распределенное хранилище данных, а для анализа и преобразования данных использовать технологии OLAP и

DataMining. Это поможет увеличить продуктивность работы программ, разгрузив клиентские машины от избыточного объема данных и вычислительных операций.

Литература

1. Дейт, К. Дж. Введение в системы баз данных / К.Дж. Дейт – 8 изд. – М.: Издательский дом "Вильямс", 2005. – 1328 с.
2. Степанов, Ю. А. Обеспечение информационной поддержки ведения горных работ с использованием ГИС-технологий / Ю.А. Степанов // Вестник КузГТУ. – 2014. – № 4. – С. 118–122.
3. Степанов, Ю. А. Структура региональной геоинформационной системы при ведении выемочных работ угледобывающих предприятий / Ю.А. Степанов // Геоинформатика. – 2012. – № 1. – С. 36–41.
4. Тикнов, В. С. Анаморфозы. Что это такое? / В.С. Тикун, С.М. Гусейн-Заде. – М.: ЛКИ, 2008. – 168 с.
5. Федоров, А. Введение в OLAP-технологии Microsoft / А. Федоров, Н. Елманова. – М.: Диалог-МИФИ, 2002. – 272 с.
6. Шаши, Ш. Основы пространственных баз данных / Чаула Санжей, Шекхар Шаши. – М.: Кудиц-образ, 2004. – 336 с.
7. PostGIS Cookbook / Paolo Corti [et al.]. – Birmingham: Packt Publishing, 2014. – 484 p.