

Мощный информационный ресурс данных по состоянию ТЭК и отраслей народного хозяйства РФ и инновационный подход к его визуализации и анализу

А.С. Кашик

академик РАН, д.т.н., генеральный директор
cge@cge.ru

Т.Г. Шабельникова

начальник отдела поддержки и развития
геоинформационных технологий
fb@cge.ru

©ОАО «ЦГЭ», Москва, Россия

Описан новый метод визуализации и анализа больших объемов статистических данных. Метод позволяет эффективно работать с информацией и быстро получать результат для принятия решений.

Материалы и методы

Банк данных по разработке нефтяных месторождений РФ и статистика по отраслям ТЭК и отраслям народного хозяйства РФ по публикациям в открытой печати. Используются математические методы статистического анализа.

Ключевые слова

банк данных, динамическая визуализация, трехмерное пространство, статистическая обработка данных, широкий спектр преобразований

Начиная с 2001 года в ОАО «ЦГЭ» ведется Банк данных по разработке нефтяных месторождений РФ и состоянию основных отраслей ТЭК. Источниками информации являются официальные проектные документы по разработке нефтяных месторождений и публикации статистических данных в открытой печати.

Содержательно информация представлена геолого-геофизическими и промышленными данными по нефтяным и нефтегазовым месторождениям РФ (2044 месторождения и 13000 пластов), начиная с 1937 года и по 2010 год включительно, с общим количеством параметров 177, а также данными по нефтедобывающей, нефтеперерабатывающей, газовой промышленности, экономике и энергетике РФ, интегрированными по регионам, субъектам РФ, нефтяным компаниям и предприятиям, начиная с 1990 года и по 2010 год включительно, с общим количеством параметров 1200 и данными по отраслям народного хозяйства РФ, начиная с 2000 по 2011 год включительно, с общим количеством параметров 416. Вся информация имеет географическую привязку. Банк данных постоянно пополняется.

Единая Модель данных и созданная система стандартов описания и представления данных рассматриваемой прикладной области позволили организовать оптимальный и технологичный доступ к данным для последующей их сверхбыстрой и информативной визуализации с помощью разработанного в ОАО

«ЦГЭ» совместно с ООО «ПИК» (Петролеум Инжиниринг энд Консалтинг) программного комплекса MDV (MultiDimensional View).

В содержании Банка данных четко прослеживается 3-х мерная структура информационного пространства — Объекты, Параметры, Время, что позволяет компактно упаковать всю информацию в гипотетический Гиперкуб. В настоящее время общее количество точек пересечений в пространстве Гиперкуба — значений параметров, составляет более 15 миллионов. Объем и содержание Банка данных позволяет говорить о том, что анализ накопленного информационного ресурса требует нетрадиционных подходов и прежде всего высоких скоростей доступа, формирования любых запросов и разнообразной динамической визуализации данных. Необходимо с любой степенью детальности ориентироваться в этом огромном информационном пространстве и формировать различные тематические Кубы-проекты, удобные специалисту-аналитику для решения различных научно-производственных задач.

Чтобы оценить преимущества новой технологии анализа и визуализации данных, предлагаем кратко ознакомиться с основными функциональными возможностями и их особенностями. При этом внимание будет уделено главным достоинствам комплекса MDV — высокой скорости доступа к данным с любой степенью детализации и высокой информативности видеоряда при визуализации этих данных.

Чтобы обеспечить формирование из Гиперкуба различных тематических Кубов-проектов по запросам специалистов-аналитиков, создан процедурный язык программирования Th, позволяющий описывать неограниченное разнообразие таких проектов. Загрузка текущего состояния Баз данных из Банка данных выполняется в режиме реального времени и в полном объеме. Далее согласно запросам, описанным в Th, формируются Кубы-проекты, готовые к работе с данными в режиме их динамической визуализации. Метод динамической визуализации, который применен в данной технологии, основан на слайсировании информационного пространства по сечениям 3-х мерного Куба, где ось X — Объекты, ось Y — Параметры, ось Z — Время. В соответствии с этим сечение параллельное оси X будет отражать изменение во времени одного из параметров для всех объектов, сечение перпендикулярное оси X будет отражать изменение во времени всех параметров одного из объектов, сечение параллельное плоскости XY будет отражать изменение всех параметров всех объектов в заданное время. Скорость движения слайсов очень высокая, поэтому доступ к дозированному объему информации можно считать практически мгновенным.



Рис. 1

При подготовке к анализу и визуализации загруженного в Куб-проект информационного пространства большую роль играет возможность применения различных фильтров, позволяющих локализовать область работы специалиста-аналитика, максимально сократив время доступа к анализируемым объектам и их параметрам. Набор фильтров включает иерархический, параметрический, временной и геометрический фильтры. Иерархический фильтр позволяет в заданной системе иерархии — страны, Федеральные округа, Субъекты стран, компании и предприятия моментально на любом уровне конкретизировать область выбора объектов для визуализации и анализа. Параметрический фильтр позволяет задавать критерии для диапазона значений выбираемых параметров. Временной фильтр позволяет устанавливать периоды времени получения данных. Геометрический фильтр позволяет позиционировать отображение объектов и значений параметров на административных картах мира и Российской Федерации.

Кроме собственно динамической визуализации исходных данных, для углубленного анализа реализован широкий набор вычислительных функций и различных

преобразований — группирования, детализации, сортировки и др. для получения дополнительных параметров и агрегаций, необходимых при решении конкретных аналитических задач, и подключение этих параметров к процессу визуализации. Суммирование, группирование, детализация — все эти преобразования позволяют моментально получать количественные значения агрегаций параметров и качественные характеристики этих значений, благодаря их цветокодированию.

В программном комплексе MDV обеспечивается оконный интерфейс по принципу работы в Windows. Общий вид рабочего стола во время сеанса работы представляет собой экран с главным меню и набором открытых функциональных окон. Состояние окон отражает представление данных в виде объемов (тел), Кубов со слайсами сечений куба (3Д); двумерных цветокодированных таблиц, снятых со слайсов; двумерных графиков (2Д) в виде линий и гистограмм, позиционирования объектов на административных картах мира и Российской Федерации. Цветокодирование значений параметров применяется во всех видах визуализации данных для качествен-

ной оценки изменений величины параметра на уровне заданной нормировки. При перемещении слайса данные в открытых таблицах и на карте синхронно обновляются. За один сеанс можно работать со многими Кубами-проектами. Состояние рабочего стола сохраняется после его закрытия до следующего сеанса. Предусмотрена печать, как отдельных окон, так и экрана в целом.

Все вышеперечисленные функциональные возможности программного комплекса MDV обобщены и представлены на рис. 1, 2, 3.

Итоги

Основным результатом является принципиально новая технология работы с большим объемом данных в виде программного комплекса динамической визуализации информации и ее анализу.

Выводы

Создан новый высокотехнологичный метод визуализации и анализа больших объемов информации. Позволяет эффективно работать с информацией и быстро получать результат для принятия решений.



Рис. 2



Рис. 3

ENGLISH

INFORMATION TECHNOLOGIES

Powerful data resource showing the state of RF national fuel-and-energy complex and industry sectors — innovative visualization and analysis

UDC 004.92

Authors:

Alexey S. Kashik — scD (Tech.), member of Russian Academy of Natural Sciences, general director¹; cge@cge.ru

Tatiana G. Shabelnikova — head, department of geoinformation technologies maintenance and development¹; fb@cge.ru

¹ CGE, Moscow, Russian Federation

Abstract

New method is described to display and analyze large amounts of statistical data. The method allows effective treatment of data and rapid generation of results for use in decision-making.

Materials and methods

Databank storing published technical data on the development of hydrocarbon fields in Russia and published statistical data on the

national fuel-and-energy complex and other sectors on national economy. Mathematical methods of statistical analysis are used.

Results

The result is a software system featuring conceptually new technology of treating large amounts of data for dynamic visualization and analysis.

Conclusions

A new high-end method is created to display and analyze large amounts of data; the method allows effective treatment of data and rapid generation of results for use in decision-making.

Keywords

databank, dynamic visualization, 3D space, statistical processing of data, broad spectrum of transformations