

УДК 025.4.03:556.3

И.В. Сахаров, А.Л. Язвин

© И.В. Сахаров, А.Л. Язвин

ОБЩИЕ ТРЕБОВАНИЯ К ФАКТОГРАФИЧЕСКИМ ИНФОРМАЦИОННЫМ СИСТЕМАМ ОБЪЕКТНОГО УРОВНЯ ПРИ ИЗУЧЕНИИ И ИСПОЛЬЗОВАНИИ ПОДЗЕМНЫХ ВОД

Введение

Наблюдаемая тенденция увеличения сложности и трудоемкости задач, решаемых при обосновании использования недр, а также необходимость взаимоувязанного анализа и согласования большого объема разнообразных материалов приводят к все большему проникновению информационных технологий в практику проведения геологических исследований и в управление недропользованием.

Использование информационных систем (ИС) позволяет эффективно учитывать многообразие и сложность природно-географических, геолого-гидрогеологических и антропогенных факторов, определяющих состояние подземных вод и его изменения.

В нашем понимании, следуя за наиболее распространенным определением понятия «информационные технологии», информационная система – система, включающая организационно упорядоченную совокупность данных и средства управления ими, предназначенная для получения информации нового качества о состоянии объекта, процесса или явления. Таким образом, ИС должны обеспечивать хранение, обработку и представление данных в стандартных и произвольных форматах. Их основными составляющими являются специальным образом организованные данные (базы данных, БД) и программное обеспечение для управления ими.

Прикладные информационно-аналитические системы должны обладать функционалом информационно-поисковых систем (ИПС), систем обработки данных (СОД), систем поддержки принятия решений (СППР) на основе применения СУБД, ГИС и веб-технологий [4, 6].

Использование ИС в гидрогеологии, как и в геологии в целом, осложняется рядом существенных особенностей предметной области, таких как:

- наличие большого объема разнородных и разноуровневых данных, полученных из различных источников и накопленных в разных форматах;

- субъективность информации, что является следствием использования различных методов и способов получения первичных данных и их интерпретации (разрезы по скважинам, значения показателей и др.);
- разобщенность логической структуры информации, выражающаяся в отсутствии унификации классификаторов и справочников;
- слабая формализация задач, решаемых в процессе недропользования.

Наиболее развитые гидрогеологические ИС [1, 6] создавались для государственных нужд и предназначены в первую очередь для работы на территориальном и региональном уровнях в соответствии с методическими документами, регламентирующими представление материалов мониторинга, лицензирования и учета запасов подземных вод.

Вследствие указанных причин имеющиеся разработки не в полной мере отвечают требованиям их использования при решении задач объектного уровня (т.е. задач обоснования инженерных решений на конкретных участках недр), не обеспечивают эффективный ввод данных и получение выходной информации в необходимых формах и объемах. Об этом свидетельствует то обстоятельство, что практически все гидрогеологические организации в той или иной мере занимаются собственной разработкой ИС (баз данных) – несмотря на достаточно стандартный состав обрабатываемой информации.

При этом для информационных систем, применяемых при оценке запасов подземных вод, ведении мониторинга на локальных объектах, обосновании дренажных мероприятий и закачки вод в недра и т.п., характерен ряд недостатков, основные из которых сводятся к следующему:

1. Неоптимальная конфигурация структуры баз данных и отсутствие единой типизации объектов.
2. Недостаточно четко прописанные критерии идентификации гидрогеологических объектов

и информации, что обуславливает сложности импорта данных.

3. Фиксированное количество видов вводимых объектов и показателей.
4. Недостаточно развитые средства поиска гидрогеологических объектов и связанных данных.
5. Узкий спектр процедур, предназначенных для преобразования и обобщения информации.
6. Ограниченный набор выходных форм.
7. Отсутствие инструментов анализа состава и качества внесенной информации.

Рассмотрим далее возможные направления решения обозначенных выше проблем. В значительной степени предлагаемые подходы были реализованы при разработке программно-алгоритмического комплекса (ПАК) GeoCODE [5]. На наш взгляд, они являются общими при разработке аналогичных ИС.

Комплекс (свидетельство о государственной регистрации программы для ЭВМ № 2009610229 от 11.01.2009 г.) постоянно совершенствуется и показал свою эффективность на многочисленных объектах при решении таких задач, как оценка запасов пресных подземных вод (Московский регион, Красноярский край, Самарская, Волгоградская, Рязанская, Тверская, Амурская области); оценка запасов минеральных подземных вод (регион Кавказских Минеральных Вод); обоснование внутриводоносного дренажа (Западно-Сибирский металлургический комбинат); обоснование сброса промыслово-сточных вод и ликвидации очагов загрязнения (Саратовская и Волгоградская области) и др.

Итак, при создании гидрогеологических ИС следует выделить требования к разработке системы в целом (требования к проектированию) и требования к гидрогеологическим базам данных (требования к модели данных).

При проектировании гидрогеологических ИС [3] должны быть учтены следующие требования:

- адаптация программно-алгоритмического комплекса под потребности конечного пользователя;
- конструктивность – возможность дальнейшего развития, оперативного расширения как номенклатуры объектной структуры базы данных, так и коррекция (расширение) описаний отдельных объектов (в том числе при изменении нормативных требований);
- защита баз данных от несанкционированного доступа;
- защита от аппаратных сбоев и некорректных действий пользователя – обеспечение ИС средствами периодического сохранения БД (в ручном или автоматическом режиме)

и средствами восстановления БД из существующего временного списка.

Адаптация ПАК осуществляется путем ограничений перечня доступных процедур обработки и вывода информации, структура БД для всех пользователей неизменна.

Гидрогеологические базы данных (ГГБД) обычно строятся на основе модели данных – теории представления и обработки данных в системе управления базами данных [2]. В используемых моделях данных можно выделить следующие основные аспекты:

1. Структура базы данных – описания типов и логических связей данных;
2. Целостность – методы поддержки корректного состояния базы данных;
3. Обработка данных – методы ввода, модификации и извлечения данных.

1. Разработка структуры базы данных

1.1. Структура ГГБД должна быть основана на строгом разграничении видов информации, объектов базы данных, их свойств, перечней, классификаторов, входных и выходных форм. В ее составе целесообразно выделить три блока (раздела): блок информации об объектах базы данных, блок информации о состоянии природной среды, лексикографический блок. При разработке структуры ГГБД предлагается использовать типизацию объектов, представленную в табл. 1.

Блок информации об объектах базы данных является основным и содержит описание ее предметной области. При разработке структуры GeoCODE особое внимание было уделено требованию непротиворечивости информации в иерархии основных гидрогеологических объектов (гидрогеологическая структура – месторождение – участок месторождения – лицензионный участок), технических объектов и связанных документов (лицензии, протоколы, отчеты и др.) с учетом максимально близкого соответствия требованиям нормативных документов.

На рис. 1 в качестве примера приведены основные связи объекта «Лицензионный участок».

В информационной системе, в зависимости от ее назначения, реально существующие объекты должны быть разделены на две категории. К первой («информационные» объекты) относятся те, характеристики которых являются предметом хранения в БД и обработки посредством используемого программного обеспечения в конкретной ИС. Это сложные многопараметрические объекты, в базе данных для них предусматривается основная таблица с большим количеством полей для паспортных характеристик и связанные таблицы, в том числе для переменных показателей.

Таблица 1

Типизация объектов

Применение ГИС-технологий

№	Группа объектов	Тип объектов
1	Подземные водные объекты	– гидрогеологическая структура, – месторождение подземных вод, – участок месторождения, – лицензионный участок
2	Технические объекты (сооружения)	– пункт наблюдения (ПН): скважина, родник и т.д., – эксплуатационное сооружение (водозабор и т.д.)
3	Документальные объекты	– лицензии на право пользования недрами, – протоколы экспертизы запасов, – геологические отчеты
4	Производственные объекты (организации)	– гидрогеологические и другие предприятия; – недропользователи; – химико-аналитические лаборатории
5	Административно-территориальные объекты	– федеральные округа, – субъекты Федерации, – административные районы и др.
6	Объекты водохозяйственного районирования	– бассейновый округ; – речной бассейн; – водохозяйственный участок

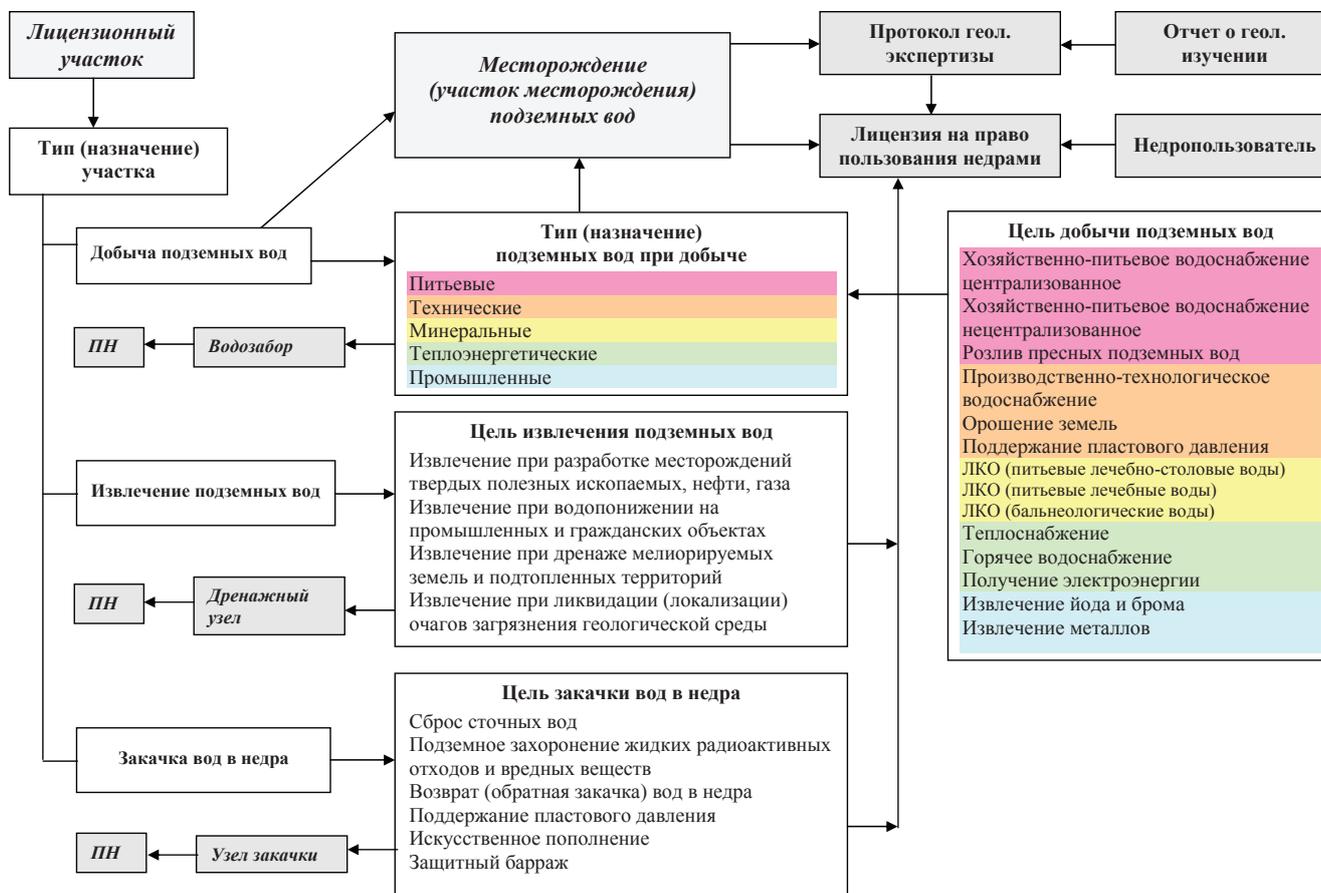


Рис. 1. Основные информационные связи объекта «Лицензионный участок»

Объекты второй категории представлены только их названиями, содержащимися в перечнях. Они связаны с информационными объектами и используются при обобщении данных и формировании запросов. Предназначенные для них таблицы имеют только одно содержательное поле (собственно наименование объекта).

Например, в ИС, создаваемой для оценки запасов подземных вод, «месторождение» является информационным объектом (для него вводятся данные по типу месторождения, группе сложности, водоносным горизонтам и др.), а «гидрогеологическая структура» – нет (какие-либо данные, в т.ч. паспортные, отсутствуют, поскольку не требуются для решения задачи).

Пространственные объекты в фактографической БД сопровождаются картографическими ГИС-объектами, хранящимися в связанных файлах в форматах ArcView и MapInfo. Картографические объекты создаются на основе топографических карт различного масштаба. При необходимости детализации возможно создание объектов на основе рисунков и схем.

Для иерархически связанных объектов (гидрогеологическая структура – месторождение – участок месторождения – лицензионный участок – эксплуатационное сооружение – ПН) предусмотрено использование (ссылки на соответствующие объекты) однотипных характеристик:

- пространственных текстовых: федеральные округа, субъекты Федерации, административные районы и др.;
- пространственных цифровых (проективные и географические координаты объектов).

Для обновляемых характеристик технических объектов предусмотрено хранение ретроспективных и текущих значений основных параметров (например, для ПН – назначение, техническое состояние, конструкция).

Помимо объектов и их паспортных (условно постоянных) характеристик, в структуре БД должны быть предусмотрены блок информации о состоянии природной среды и лексикографический блок.

Блок информации о состоянии природной среды включает разделы «наблюдения за режимом» и «результаты опробования». В данном блоке содержится пополняемая (динамическая) информация об объектах базы данных. Результаты исследований приурочены к датам их проведения, а также – для наблюдений за режимом – к временным периодам, на которые проведено обобщение срочных данных (сутки, месяц, год).

Лексикографический блок является вспомогательным, состоит из набора таблиц (перечней и

классификаторов), содержащих лексические единицы, используемые для описания свойств информационных объектов.

Перечни содержат названия объектов, не являющихся самостоятельными объектами БД. Общероссийские перечни (субъекты Федерации, административные районы и др.) содержат конечное количество записей и входят в базовый состав комплекса. Объектные перечни (водоносные горизонты, лаборатории, организации и др.) должны разрабатываться пользователями самостоятельно.

Классификаторы содержат наборы характеристик информационных объектов. Они представляют собой списки типов объектов БД, их свойств, видов наблюдаемых показателей (параметров), различных методов, устройств и т.д. Классификаторы составлены на основании различных литературных источников и введены в ИС при ее разработке.

1.2. Критерии идентификации объектов и данных. В структуре таблиц базы данных и импортируемых файлов должно быть предусмотрено наличие так называемых «ключевых (идентификационных) полей» – совокупности полей, по которым производится идентификация записи в таблице. Каждый объект БД, запись классификатора и запись в таблицах характеристик объектов являются уникальными, их дублирование невозможно.

1.3. Расширение состава информации. В ИС должна быть предусмотрена возможность создания:

- новых типов объектов, в том числе отражающих специальные виды исследований, например, геоботанические или геофизические.
- новых показателей, которые дополнительно вводятся в список динамических (переменных) параметров.

2. Целостность БД

Одним из важнейших вопросов является обеспечение устойчивого функционирования ИС (целостности БД). Создание ИС в любых областях деятельности предполагает, что внесенные данные должны быть унифицированы, взаимосвязаны и систематизированы. К составу и содержанию исходной информации предъявляются следующие основные требования:

- полнота, достоверность и достаточность (всестороннее, объективное и подробное описание предметной области);
- избыточность (сведение к минимуму дублирующих сведений в иерархической цепочке объектов БД);

- актуальность (свойство данных в любой момент времени адекватно отображать реальное состояние изучаемых объектов);
- непротиворечивость (совпадение одинаковых характеристик одних и тех же объектов);
- согласованность информации для связанных объектов;
- уникальность (отсутствие дублирования данных по объектам);
- динамическая непрерывность (накопление сведений об эволюции предметной области);
- устойчивость – возможность работы ИС при частично не введенных данных по объектам в целом или отдельным полям объектов.

Полнота и актуальность определяются наличием и своевременным вводом исходных данных и не зависят от структуры БД и программного обеспечения. Далее эти свойства не рассматриваются.

Целостность базы данных обеспечивается контролем уникальности вводимой информации (при необходимости для части информации хранится история проведенных изменений, например, конструкции скважины) и специальными процедурами сквозной проверки связанных основных объектов по общим параметрам (например, привязанность к объектам административно-территориального деления).

Наиболее критичным для целостности базы является ввод информации, так как здесь появляются как ошибки оператора подготовки информации, так и неточности исходных данных. Поэтому при вводе информации предусмотрена многоуровневая проверка ее корректности:

- 1) проверка наличия основной, целевой информации;
- 2) проверка достаточности информации для однозначной идентификации основного объекта ввода;
- 3) проверка достаточности информации для объектов-ссылок (например, месторождения для участка месторождения);
- 4) проверка уникальности вводимого объекта (для неуникального, но идентифицированного объекта предусматривается редакция соответствующих параметров);
- 5) проверка объектов-ссылок: для отсутствующих в базе объектов предусмотрен их ввод, для идентифицированных - ссылка на существующий объект;
- 6) проверка возможности преобразования текстовых описаний в числовые для соответствующих параметров.

После каждой проверки пользователю представляется log-файл с результатами проверки.

После редактирования вводной информации проверки могут быть повторены. После проведения всех проверок осуществляется загрузка новой информации в базу данных.

Входные проверки обеспечивают только предварительную оценку целостности базы данных. Для углубленной оценки предусмотрен набор дополнительных процедур:

1. Синхронизация информации о цифровых (проективных) координатах и географические координатах;
2. Синхронизация фактографической и картографической информации: ПН, лицензионные участки (центральная точка) и др.
3. Проверка пространственного взаимоположения информационно-связанных объектов (например, координат ПН в фактографической БД и ГИС-картографического образа связанного лицензионного участка);
4. Пересчеты технических характеристик объектов (например, глубина скважины по паспорту, по конструкции и по разрезу);
5. Проверка введенных средних значений и рассчитанных по режимным наблюдениям;
6. Удаление дублирующих значений в режимных наблюдениях;
7. Проверка соответствия горизонтов, назначений использования вод в лицензионных участках с лицензиями недропользователей и протоколах утверждения запасов, другие вспомогательные процедуры.

3. Обработка данных

При разработке методов обработки данных должны быть реализованы эффективные процедуры выборки объектов по комплексным признакам, обобщения данных по времени и между объектами, формирования стандартных и специальных выходных форм (каталоги, справки, таблицы, карты и др.).

3.1. Формирование выборки объектов. Выборка объектов для просмотра информации и построения выходных форм осуществляется на основе механизма, позволяющего учитывать различные типы критериев. При этом при выборке объектов может быть использован как один критерий, так и любая их совокупность. Например, при создании выборки скважин могут использоваться территориальные критерии (административный район, участок наблюдения), привязки к другим объектам базы (месторождение, водозабор и др.), текстовые (назначение, водоносный горизонт), числовые (глубина, год бурения) характеристики, а также критерии наличия связанных данных (данных

о конструкции, разреза и др.). Кроме того, для части критериев предусмотрен контекстный поиск по вхождениям выбранных символов в текстовое содержание критериев.

3.2. Обобщение данных. Обобщение переменных данных по времени (как при заполнении выходных форм, так и при создании дополнительных, комплексных «информационных объектов») производится в направлении «дата и время – дата – месяц – год». Для некоторых видов показателей, таких как осадки, объемы воды и др., определяются суммарные значения (сумма за месяц, сумма за год), для остальных – средние (среднемесячные, среднегодовые). Результаты расчетов могут сохраняться в БД.

Обобщение на уровне технических объектов производится по направлению ПН – эксплуатационное сооружение.

Пространственное обобщение объектов производится по направлениям:

- ПН – лицензионный участок;
- лицензионный участок – участок месторождения;
- участок месторождения – месторождение подземных вод;
- месторождение подземных вод – гидрогеологическая структура.

В качестве дополнительных предусмотрены связи:

- ПН – участок месторождения;
- ПН – месторождение подземных вод;
- лицензионный участок – месторождение подземных вод.

В GeoCODE предусмотрено также обобщение данных химических анализов при формировании выходных форм: определение типа воды, определение газового состава и газового фактора.

3.3. Формирование выходных форм. В качестве выходных предусмотрено формирование, для выбранных в соответствии с п. 2.1 объектов стандартных и специальных форм:

- каталоги,
- справки,
- таблицы,
- графики и диаграммы,
- карты,
- профили,
- геологические и геолого-технические разрезы.

При этом предусмотрено создание и сохранение отчетных материалов, установленных нормативными документами, а также в виде шаблонов дополнительных каталогов с содержанием, заданным пользователем.

Также предусмотрено создание комплексных рисунков большого формата («плакаты»), на которые пользователь может нанести одновременно ранее сформированные таблицы, графики, диаграммы, карты, профили, геологические и геолого-технические разрезы, штампы.

Все выходные формы создаются и сохраняются (с возможностью последующего редактирования) с использованием программ MS Word, MS Excel, специальных программных комплексов Surfer и Grapher (Golden Software), ГИС-приложений MapObjects (ESRI) и EasyMap VCL (Microolap Technologies LLC). В качестве дополнительного варианта возможно использование MapWinGIS AcvieX (MapWindow GIS Open Source).

Примеры использования комплекса GeoCODE для представления данных приведены на рис. 2.

Специальные ГИС-приложения используются также при построении профилей и геологических разрезов с использованием картографических объектов: подземные водные объекты, ПН, изогипсы, линейные водные объекты (реки, ручьи), площадные водные объекты (озера, заливы, широкие реки).

Кроме того, каталоги и справки можно сохранить в формате html для просмотра веб-браузером, а карты, графики, диаграммы, профили, геологические и геолого-технические разрезы – в виде рисунков.

Для минеральных водных источников предусмотрена оценка и классификация подземных вод в соответствии с нормативными документами по результатам химико-аналитических исследований.

С целью интеграции информационных ресурсов должен быть предусмотрен обмен данными с численными моделями и картографическими (географическими) информационными системами.

Выводы

1. Использование информационных технологий в гидрогеологии, несмотря на многолетний опыт разработки и постоянное совершенствование технических средств, сопряжено с рядом проблем, обусловленных, в первую очередь, особенностями предметной области. Помимо этого, наиболее развитые ИС создавались для государственных нужд и не в полной мере отвечают задачам обоснования инженерных решений на конкретных участках недр.

Предлагаемые в статье требования к структуре баз данных, процедурам поддержания их целостности и способам обработки информации, реализованные при создании программно-алгоритмического комплекса GeoCODE, являются общими при разработке аналогичных фактографических ИС объектного уровня.

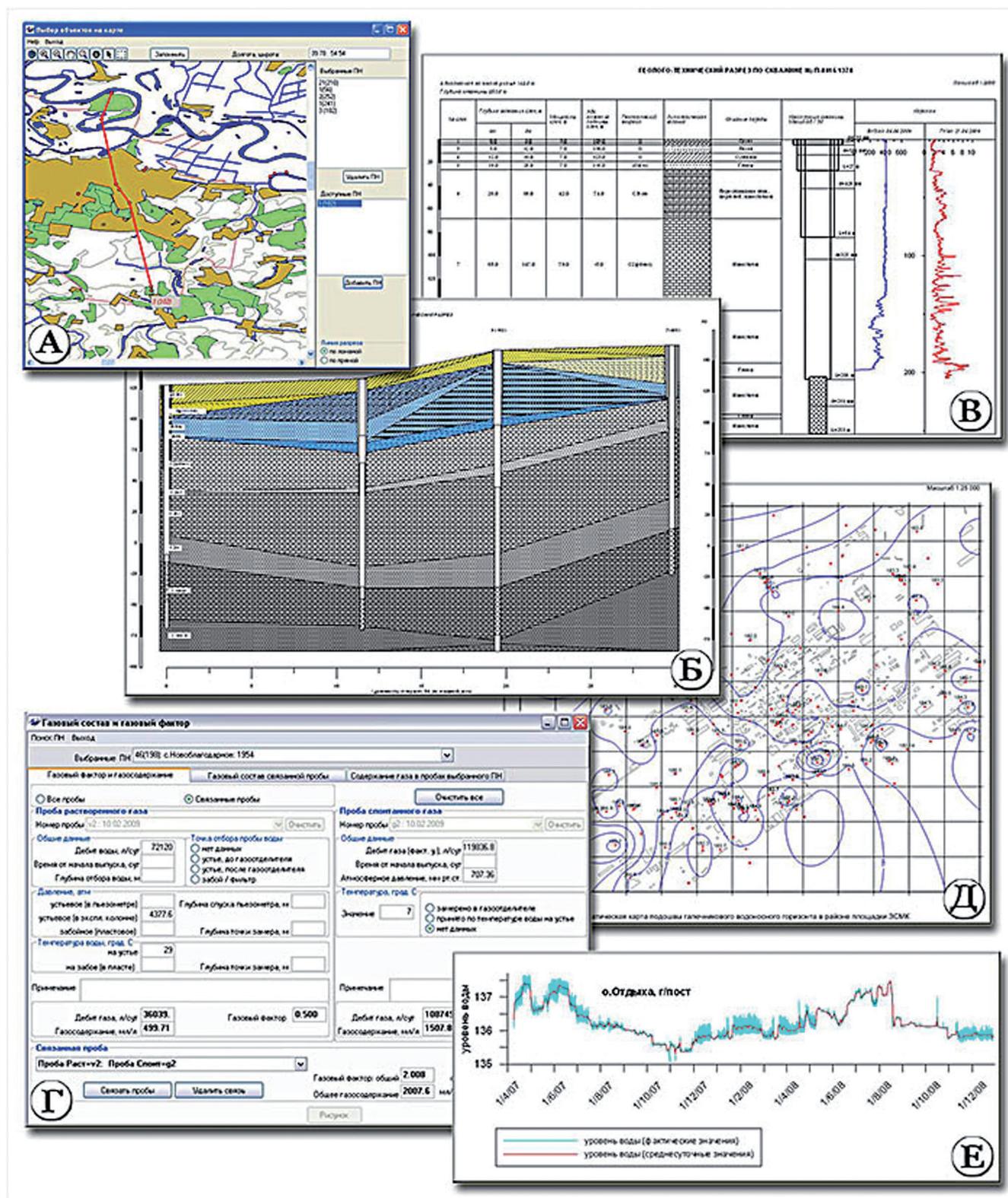


Рис. 2. Примеры использования комплекса GeoCODE: выбор скважин для построения разреза (А); геологический разрез (Б); геолого-технический разрез (В); расчет газового состава и газового фактора (Г); карта изолиний подошвы горизонта (Д); хронологический график уровня воды (Е)

2. Структура БД должна базироваться на строгом разграничении видов информации, объектов базы данных (информационных объектов), их свойств, перечней, классификаторов, входных и выходных форм. Под термином «информационный объект» понимаются объекты, характеристики которых являются предметом хранения и обработки средствами конкретной ИС. В ИС должна быть предусмотрена возможность создания пользователем новых объектов и новых показателей.

3. Важнейшим вопросом устойчивого функционирования ИС является обеспечение целостности БД. Наиболее критичным является ввод информации, в связи с чем требуется разработка критериев идентификации данных на основе так называемых «обязательных полей».

Для поддержания целостности БД предусмотрен набор дополнительных процедур, направленных на проверку уникальности, согласованности и непротиворечивости данных и их коррекцию (синхронизация пространственных характеристик, пересчеты цифровых показателей и др.).

4. Методы содержательной обработки данных должны включать эффективные процедуры выборки объектов по комплексным признакам, обобщения переменных данных по времени, пространственного обобщения информации, формирования стандартных и специальных выходных форм.

Следует предусмотреть возможность создания и сохранения отчетных материалов как по формам, заданным разработчиками ИС (в т.ч. – в соответствии с действующими нормативными документами), так и в виде шаблонов дополнительных каталогов с поддержанием, заданным пользователем.

Ключевые слова: фактографические информационные системы, гидрогеологические базы данных, требования к структуре баз данных, целостность, методы обработки.

ЛИТЕРАТУРА

1. Гомберг И.Г., Клюквин А.Н., Рошаль А.А. Интегрированная информационная технология поддержки принятия решений с сфере управления недропользованием // Разведка и охрана недр. – 2003. – № 9. – С. 38-44.
2. Дейт К.Дж. Введение в системы баз данных. – 8-е изд. – М. : Вильямс, 2006.
3. Зеегофер Ю.О., Клюквин А.Н., Пашковский И.С., Рошаль А.А. Постоянно действующие модели гидролитосферы территорий городских агломераций (на примере Московской агломерации). – М. : Наука, 1990. – 198 с.
4. Митракова О.В., Аракчеев Д.Б. Информационно-аналитическое обеспечение мониторинга состояния и использования природных ресурсов // Геоинформатика. – 2006. – № 3. – С. 25-29
5. Сахаров И.В., Язвин А.Л. Характеристика программно-алгоритмического комплекса GeoCODE // Разведка и охрана недр. – 2010. – № 10. – С. 42-47.
6. Чесалов Л.Е., Попов А.С., Аракчеев Д.Б. Создание территориально-распределенных информационных систем в сфере управления недропользованием // Разведка и охрана недр. – 2007. – № 11. – С. 45-50.